Treball Final de Màster

MÀSTER EN MATEMÀTICA AVANÇADA

Facultat de Matemàtiques
Universitat de Barcelona

VARIOUS EXTENSIONS OF THE MÜNTZ-SZÁSZ THEOREM

Universitat de Barcelona

Autor: Sergi Baena i Miret

Director: Dra. María Jesús Carro

Realitzat a: Departament de Matemàtiques i Informàtica

Data: Barcelona, September 10, 2017

“Polynomials pervade mathematics, and much that is beautiful in mathematics is related to polynomials. Virtually every branch of mathematics, from algebraic number theory and algebraic geometry to applied analysis, Fourier analysis, and computer science, has its corpus of theory arising from the study of polynomials. Historically, questions relating to polynomials, for example, the solution of polynomial equations, gave rise to some of the most important problems of the day. The subject is now much too large to attempt an encyclopedic coverage”.

Tamás Erdélyi

Abstract

The Müntz-Szász Classical Theorem characterizes increasing sequences $\{\lambda_j\}_{j=0}^{+\infty}$ with

$$0 = \lambda_0 < \lambda_1 < \lambda_2 < \cdots$$

for which the space $\langle 1, x^{\lambda_1}, x^{\lambda_2}, \dots \rangle$ is dense or not in $\mathcal{C}([0,1])$, depending on if the series $\sum_{j=1}^{+\infty} 1/\lambda_j$ diverges or not respectively.

In the book *Polynomials and Polynomials Inequalities* (see [7]), Tamás Erdélyi and Peter Borwein explain the tools needed in order to show a complete and extended proof of the Müntz-Szász Theorem. To do so, they use some techniques of complex analysis and also the algebraic properties of the zeros of some functions called Chebyshev functions.

On these notes we put together all these ideas, beginning with the well known Weierstrass Approximation Theorem, continuing with the development of the complex analysis results needed and giving a complete proof of an extended version of the Müntz-Szász Theorem. Such new version characterizes arbitrary sequences $\{\lambda_j\}_{j=0}^{+\infty}$ of different arbitrary positive real numbers (except for $\lambda_0 = 0$) for which the space of continuous functions spanned by the powers x^{λ_j} is dense or not in $\mathcal{C}([0,1])$. In that case, it depends on if the series $\sum_{j=1}^{+\infty} \lambda_j/(\lambda_j^2+1)$ diverges or not respectively. Moreover, pursuing in this direction, we also have studied an equivalent result for the Lebesgue spaces that characterizes arbitrary different sequences $\{\lambda_j\}_{j=1}^{+\infty}$ of real numbers greater than $-1/p$ for which the space $\langle x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \dots \rangle$ is dense or not in $L^p([0,1])$, which in that case depends on if the series $\sum_{j=1}^{+\infty} (\lambda_j + 1/p)/((\lambda_j + 1/p)^2 + 1)$ diverges or not respectively.

2010 Mathematics Subject Classification: Primary 41A30, 41A50; Secondary 28C05, 41A10, 41A17.

Keywords: Weierstrass Theorem; Riesz Representation Theorem; Hahn-Bannach Theorem; Müntz Theorem; Müntz Systems; Newman Inequalities; Bernstein Inequalities; Chebyshev Inequalities.

Contents

1 INTRODUCTION

In his seminal paper [4] of 1912, the Russian mathematician S. N. Bernstein asked under which conditions on an increasing sequence $\Lambda = \{\lambda_j\}_{j=0}^{+\infty}$ ($\lambda_0 = 0$) one can guarantee that the vector space

$$\Pi(\Lambda) := \langle x^{\lambda_j} \colon j = 0, 1, 2, \dots \rangle,$$

spanned by the polynomials x^{λ_j}, is a dense subset of the space of all continuous real valued functions defined on the interval $[0,1]$, denoted by $\mathcal{C}([0,1])$. He specifically proved that the condition

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

is necessary and the condition

$$\lim_{j \to +\infty} \frac{\lambda_j}{j \log j} = 0$$

is sufficient, and conjectured that a necessary and sufficient condition to have $\overline{\Pi(\Lambda)} = \mathcal{C}([0,1])$ is

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

This conjecture was proved by Müntz [15] in 1914 and by Szász [16] in 1916, which was only for distinct positive real sequences of exponents tending to infinity. After that, this result began to be called the Müntz-Szász Classical Theorem. Later works, see for example [1] and [5], include the original result as well as a treatment of the case when $\{\lambda_j\}_{j=0}^{+\infty}$ is a sequence of distinct positive real numbers (except for $\lambda_0 = 0$) such that $\inf_{j \geq 1} \lambda_j > 0$.

The beauty of the Müntz-Szász Classical Theorem lies on the fact that it connects a topological result (the density of a certain subset of a functional space) with an arithmetical one (the divergence of a certain harmonic series). Another reason to be interested on this theorem is that the original result not only solves a nice problem but also opens the door to many new interesting questions. For example, one is tempted to change the space of continuous functions $\mathcal{C}([0,1])$ to other spaces as $L^p([0,1])$, or to consider the analogous problem in several variables, on intervals away of the origin, for more general exponent sequences, for polynomials with integral coefficients, etc. As a consequence, many proofs (and generalizations) of the theorem have been done by many authors as, for example, Manfred Von Gloitschek [17] and Támas Érdelyi (see [7] and [6]).

On these notes, we concentrate our attention on the Müntz' problem in the univariate setting for the interval $[0,1]$ restricted to the uniform and the Lebesgue norms. Moreover, we provide proofs in great detail of all the results needed in order to show both necessary and sufficient conditions. We have structured theses notes chronologically and divided in three distinct parts where we develop different techniques respectively.

On the first part we have shown what condition is necessary in order to satisfy the Müntz-Szász Theorem in $\mathcal{C}([0,1])$ for sequences $\{\lambda_j\}_{j=0}^{+\infty}$ of distinct positive real numbers. To do so, we have begun by motivating the problem with a proof of S. N. Bernstein of the well known Weierstrass Approximation Theorem (see [3]) which is a particular case of the Müntz-Szász Theorem. The proof of Bernstein introduce a discrete function that approximates every continuous function in the interval $[0, 1]$ as close we desire. Then, we have studied some results in complex measure theory and functional analysis with the goal to show two relevant results for which the proof of the necessary condition in the Müntz-Szász Theorem is based on: the Riesz-Markov-Kakutani Theorem on $\mathcal{C}([0, 1])$ (see [1]) and a corollary of the Hahn Banach Theorem (see [5]). Finally, we have given the statement and the proof of the necessary condition in the Müntz-Szász Theorem.

On the second part our aim is to show that the necessary condition of the Müntz-Szász Theorem is also sufficient. To do so, we have introduced some finite vectorial subspaces of the continuous functions in $[0, 1]$ and we have related them with the space spanned by the powers x^{λ_j}, where $\lambda_0 = 0$ and $\{\lambda_j\}_{j=1}^{n}$ ($n \in \mathbb{N}$) is a sequence of different positive real values. To study them we have followed the book *Polynomials and Polynomials Inequalities* of Tamás Erdélyi and Peter Borwein (see [7]), where the material is often tersely presented, with much mathematics explored in the exercises, many of which are supplied with copious hints, some with complete proofs. Well over half the material in that book is presented in the exercises. Hence, together with [7], we also have taken use of the article *Müntz Type Theorems I* of J.M. Almira (see [18]). Finally, we have proved the sufficiency of the condition of the Müntz-Szász Theorem.

On the third part we have presented an extended version of the Müntz-Szász Theorem, but now for the Lebesgue spaces in $[0, 1]$ where $p \in [1, +\infty)$. However, in that case we have used sequences $\{\lambda_j\}_{j=1}^{+\infty}$ of distinct real numbers greater than $-1/p$. Even though, as in $\mathcal{C}([0, 1])$, we have proved the theorem on two steps: one for proving the necessary condition that satisfy the theorem (where we have seen a complete proof for any $p \in [1, +\infty)$), and then other for proving the sufficiency of such condition (where we have seen a complete proof for the case $p = 1$, but for the case $p > 1$ we restrict the sequence to satisfy $\inf_{j \geq 1} \lambda_j > -1/p$).

To finish, thanks are due to María Jesús Carro for leading me and watching that I did not digress much from the right way by giving me advices of the best method on each case. Moreover, she gave me a great range of bibliography and helped me every time I got stuck.

2 DENSITY ON MÜNTZ-SZÁSZ APPROXIMATION THEOREM

Müntz-Szász classical Theorem characterizes increasing sequences $\{\lambda_j\}_{j=0}^{+\infty}$ with

$$0 = \lambda_0 < \lambda_1 < \lambda_2 < \cdots$$

for which the space $\langle 1, x^{\lambda_1}, x^{\lambda_2}, \dots \rangle$ is dense or not in $\mathcal{C}([0,1])$, depending on if the series $\sum_{j=1}^{+\infty} 1/\lambda_j$ diverges or not respectively.

As a particular case, when $\lambda_n \in \mathbb{N} \cup \{0\}$, we have the well known Weierstrass Approximation Theorem, which says that any real valued continuous function defined on a real interval can be approximated arbitrarily well by polynomials. There are many different proofs of this result, however, on the following section we present a proof based on Bernstein polynomials [3].

2.1 Weierstrass Approximation Theorem

Before beginning, we note that either the Weierstrass Approximation Theorem holds in every interval $[a, b]$, it suffices to work with the interval $[0, 1]$. The reason is that the arbitrary real interval $a \leq y \leq b$ is mapped to $0 \leq x \leq 1$ by the polynomial $x = (a-y)/(a-b)$ and vice versa by $y = (b-a)x + a$. So, if g is continuous on $[a, b]$, then $f(x) = g((b-a)x + a)$ is continuous on $[0, 1]$. Therefore, if the polynomial P_n of degree n approximates f to within $\varepsilon > 0$ in $[0, 1]$, then the polynomial $Q_n(y) = P_n((a-y)/(a-b))$ of degree n approximates $g(y)$ to within ε in $[a, b]$.

First, we present some concepts needed before seeing the theorem, which follows the notation of [3].

Definition 2.1.1. Let $f \in \mathcal{C}([0,1])$ a continuous function and let $\delta > 0$. A modulus of continuity in $[0, 1]$ is a positive function defined as

$$w(f, \delta) = \sup_{\substack{|x-y|<\delta \\ x,y\in[0,1]}} \{|f(x) - f(y)|\}.$$

Observation 2.1.2. Since a continuous function is bounded in a compact set, it follows that $w(f, \delta) < +\infty$ for any $\delta > 0$. As a consequence, every $f \in \mathcal{C}([0,1])$ is uniformly continuous in $[0, 1]$. Hence,

$$\lim_{\delta\to 0} w(f, \delta) = 0, \quad (\forall f \in \mathcal{C}([0,1])).$$

Definition 2.1.3. A Bernstein Binomial is

$$P_{n,m}(x) = \binom{n}{m} x^m (1-x)^{n-m},$$

where $n, m \in \mathbb{N}$, $0 \leq m \leq n$ and $x \in [0, 1]$.

Lemma 2.1.4. *Let $n \in \mathbb{N}$, $x \in [0, 1]$,* *(i)* $\sum_{m=0}^{n} P_{n,m}(x) = 1$,

(ii) $\sum_{m=0}^{n} mP_{n,m}(x) = nx$,

(iii) $\sum_{m=0}^{n} m^2 P_{n,m}(x) = (n^2 - n)x^2 + nx$.

Proof. (i)

$$1 = 1^n = (1 - x + x)^n = \sum_{m=0}^{n} \binom{n}{m} x^m (1-x)^{n-m} = \sum_{m=0}^{n} P_{n,m}(x).$$

(ii)

$$\begin{aligned}\sum_{m=0}^{n} mP_{n,m}(x) &= \sum_{m=1}^{n} m\binom{n}{m} x^m (1-x)^{n-m} = nx \sum_{m=1}^{n} \binom{n-1}{m-1} x^{m-1}(1-x)^{(n-1)-(m-1)} \\ &= nx \sum_{m=0}^{n-1} \binom{n-1}{m} x^m (1-x)^{(n-1)-m} = nx.\end{aligned}$$

(iii)

$$\begin{aligned}\sum_{m=0}^{n} m^2 P_{n,m}(x) - nx &= \sum_{m=0}^{n} m^2 \binom{n}{m} x^m (1-x)^{n-m} - \sum_{m=0}^{n} m\binom{n}{m} x^m (1-x)^{n-m} \\ &= \sum_{m=2}^{n} m(m-1)\binom{n}{m} x^m (1-x)^{n-m} \\ &= n(n-1)x^2 \sum_{m=2}^{n} \binom{n-2}{m-2} x^{m-2}(1-x)^{(n-2)-(m-2)} \\ &= n(n-1)x^2 \sum_{m=0}^{n-2} \binom{n-2}{m} x^m (1-x)^{(n-2)-m} = n(n-1)x^2.\end{aligned}$$

□

Definition 2.1.5. A Bernstein Polynomial of degree $n \in \mathbb{N}$ for a function $f \in \mathcal{C}[0,1]$ is

$$B_{n,f}(x) = \sum_{m=0}^{n} P_{n,m}(x) f(x_m),$$

where $x_m = \frac{m}{n}$ and $m = 0, \ldots, n$.

The following result, given by Bernstein [3], would yield as a consequence the Weiesrtras Approximation Theorem.

Theorem 2.1.6 (Bernstein Approximation Theorem)**.** *Let f be a continuous function on $[0,1]$ and $n \geq 1$ a natural number. Then,*

$$|f(x) - B_{n,f}(x)| \leq \frac{9}{4} w(f, n^{-1/2}).$$

Proof. Let $\delta > 0$ and $x \in [0,1]$. Since $\sum_{m=0}^{n} P_{n,m}(x) = 1$, it follows that

$$f(x) - B_{n,f}(x) = \sum_{m=0}^{n} (f(x) - f(x_m))P_{n,m}(x) = \sum_{\substack{m=0 \\ |x-x_m|<\delta}}^{n} (f(x) - f(x_m))P_{n,m}(x) + \sum_{\substack{m=0 \\ |x-x_m|\geq\delta}}^{n} (f(x) - f(x_m))P_{n,m}(x).$$

Let's work each sum separately. First, using that $P_{n,m}(x) \geq 0$ for $x \in [0,1]$,

$$\left| \sum_{\substack{m=0 \\ |x-x_m|<\delta}}^{n} (f(x) - f(x_m))P_{n,m}(x) \right| \leq \sum_{\substack{m=0 \\ |x-x_m|<\delta}}^{n} |f(x) - f(x_m)|P_{n,m}(x) \leq w(f,\delta) \sum_{m=0}^{n} P_{n,m}(x) = w(f,\delta). \tag{2.1.1}$$

Next, we take $m \in \{0, \ldots, n\}$ such that $|x - x_m| \geq \delta$ and define

$$K_m := \left[\frac{|x - x_m|}{\delta} \right] \in \mathbb{Z}_{\geq 1}.$$

Now, we choose $y_0 < y_1 < \cdots < y_{K_m} < y_{K_m+1}$ uniformly spaced in the interval generated by x and x_m, so that each of the new $K_m + 1$ intervals have length $\frac{|x-x_m|}{K_m+1} < \delta$ and where $y_0 := \min(x, x_m)$ and $y_{K_m+1} := \max(x, x_m)$. So, since $|y_{i+1} - y_i| < \delta$,

$$|f(x) - f(x_m)| = |f(y_0) - f(y_{K_m+1})| \leq \sum_{i=0}^{K_m} |f(y_i) - f(y_{i+1})| \leq (K_m + 1)w(f,\delta) \leq \left(\frac{|x - x_m|}{\delta} + 1 \right) w(f,\delta).$$

Hence, we can bound the second sum by

$$\sum_{\substack{m=0 \\ |x-x_m|\geq\delta}}^{n} |f(x) - f(x_m)|P_{n,m}(x) \leq w(f,\delta) \left(\sum_{\substack{m=0 \\ |x-x_m|\geq\delta}}^{n} P_{n,m}(x) + \frac{1}{\delta} \sum_{\substack{m=0 \\ |x-x_m|\geq\delta}}^{n} |x - x_m|P_{n,m}(x) \right).$$

Since $\frac{|x-x_m|}{\delta} \geq 1$, then $\frac{(x-x_m)^2}{\delta^2} \geq \frac{|x-x_m|}{\delta}$ and

$$\sum_{\substack{m=0 \\ |x-x_m|\geq\delta}}^{n} P_{n,m}(x) + \frac{1}{\delta} \sum_{\substack{m=0 \\ |x-x_m|\geq\delta}}^{n} |x - x_m|P_{n,m}(x) \leq 1 + \frac{1}{\delta^2} \sum_{m=0}^{n} (x - x_m)^2 P_{n,m}(x) = 1 + \frac{1}{\delta^2} \left(\sum_{m=0}^{n} (x^2 - 2xx_m + x_m^2)P_{n,m}(x) \right).$$

Now, by Lemma 2.1.4, using that $x_m = \frac{m}{n}$,

$$\begin{aligned}\sum_{m=0}^{n}(x^2 - 2xx_m + x_m^2)P_{n,m}(x) &= x^2 - 2\left(\frac{x}{n}\right)nx + \frac{1}{n^2}((n^2-n)x^2 + nx) \\ &= x^2 - 2x^2 + x^2 + \frac{x(1-x)}{n} = \frac{x(1-x)}{n}.\end{aligned}$$

Finally, observe that for $x \in [0,1]$, the function $\frac{x(1-x)}{n}$ takes the maximum at $x = \frac{1}{2}$. Therefore,

$$\sum_{\substack{m=0 \\ |x-x_m|\geq\delta}}^{n} |f(x) - f(x_m)|P_{n,m}(x) \leq w(f,\delta)\left(1 + \frac{1}{4\delta^2 n}\right). \tag{2.1.2}$$

Thus, taking $\delta = n^{-1/2}$ and using the inequalities of (2.1.1) and (2.1.2) we get

$$|f(x) - B_{n,f}(x)| \leq w(f, n^{-1/2})\left(1 + 1 + \frac{1}{4}\right) = \frac{9}{4}w(f, n^{-1/2}).$$

□

Corollary 2.1.7 (Weierstrass Approximation Theorem). *Assume that $f \in \mathcal{C}([0,1])$. Given any $\varepsilon > 0$, there is a polynomial P_n with sufficiently high degree n such that*

$$|f(x) - P_n(x)| < \varepsilon, \quad \forall\, 0 \leq x \leq 1.$$

2.2 Previous Results in Functional and Complex Analysis

From now on, our aim is to extend the Müntz-Szász Classical Theorem to arbitrary sequences $\{\lambda_j\}_{j=0}^{+\infty}$ of distinct positive real numbers. To do so, we need some previous results on complex measure theory and functional analysis. In particular, this section consists on the study of some of the classical theorems in complex and functional analysis, with taking an special attention to the Riesz-Markov-Kakutani Representation Theorem on $\mathcal{C}([0,1])$, which basically says that any bounded linear functional T on $\mathcal{C}([0,1])$ is the same as integration against a complex measure μ, i.e.,

$$Tf = \int_0^1 f\, d\mu, \qquad (f \in \mathcal{C}([0,1])).$$

Since the only big result that we will show on this section that has not been studied neither on the bachelor's degree nor the master course is the Riesz-Markov-Kakutani Representation Theorem, we will deal first with it. So that, we will begin by showing all of the concepts and results that we will use in order to prove it. To do so, we have followed the Rudin's book *Real and Complex Analysis* [1], which begins with the Riesz Representation Theorem for positive measures, continues with the duality theorem on Lebesgue spaces, and ends with the Riesz-Markov-Kakutani Representation Theorem. For simplicity, we will consider the real interval I to be $[0,1]$ and we will denote the uniform norm on I by

$$\|f\|_\infty = \sup\{|f(x)|\colon x \in I\}.$$

2.2.1 Riesz Representation Theorem for Positive Measures on $\mathcal{C}([0,1])$

We will study the Riesz Representation Theorem on $\mathcal{C}(I)$ for any positive bounded linear functional T on $\mathcal{C}(I)$. In such cases, T is going to be the same as integration against a positive measure μ. On this section, we will denote by $\mathcal{C}_c(\mathbb{R})$ the space of real valued continuous functions with compact support on $\mathbb{R}$.

Definition 2.2.1. Let $\mathfrak{B}(I)$ be the smallest σ-algebra that contains the open sets of I; this is known as the σ-algebra of the Borel sets.

The following lemmas play an important role on the Riesz Representation Theorem for positive measures on $\mathcal{C}([0,1])$. Since both results are from the course of Functional Analysis and PDE's, we present them without proof. For more details, see [1].

Lemma 2.2.2 (Urysohn's Lemma)**.** *Let V an open set in $\mathbb{R}$, $K \subset \mathbb{R}$, and let K be compact. Then there exists an $f \in \mathcal{C}_c(\mathbb{R})$, such that $\chi_K \leq f \leq \chi_V$.*

For simplicity, we will say that f satisfies $K \prec f \prec V$.

Lemma 2.2.3 (Partition of Unity)**.** *Suppose $V_1, \ldots, V_n$ are open subsets of $\mathbb{R}$, K is compact, and*

$$K \subset V_1 \cup \cdots \cup V_n.$$

Then there exist functions $h_i \in \mathcal{C}_c(\mathbb{R})$ such that $h_i \prec V_i$ $(i = 1, \ldots, n)$ and

$$\sum_{i=1}^{n} h_i(x) = 1 \quad (x \in K).$$

Definition 2.2.4. The collection $\{h_1, \ldots, h_n\}$ is called a partition of the unity on K, subordinate to the cover $\{V_1, \ldots, V_n\}$.

Theorem 2.2.5 (Riesz Representation Theorem)**.** *Let Λ be a positive linear functional on $\mathcal{C}(I)$ (i.e., for any $f \in \mathcal{C}(I)$ such that $f \geq 0$, then $\Lambda f \geq 0$), there exists a σ-algebra $\mathfrak{M}$ in I which contains all Borel sets in I, and there exists a unique positive measure μ on $\mathfrak{M}$ which represents Λ in the sense that*

$$\Lambda f = \int_I f d\mu,$$

for every $f \in \mathcal{C}(I)$. Moreover, the following properties hold:

(a) $\mu(I) < +\infty$.

(b) For every $E \in \mathfrak{M}$,

$$\mu(E) = \inf\{\mu(V)\colon E \subset V, V \text{ open}\}.$$

(c) The relation

$$\mu(E) = \sup\{\mu(K)\colon K \subset E, K \text{ compact}\}$$

holds for every $E \in \mathfrak{M}$.

(d) If $E \in \mathfrak{M}$, $A \subset E$, and $\mu(E) = 0$, then $A \in \mathfrak{M}$ (i.e., $(I, \mathfrak{M}, \mu)$ is a complete measure).

Proof. Let us begin by proving the uniqueness of μ. Suppose that exist two positives measures μ_1 and μ_2 satisfying the hypothesis of the theorem. Given $E \in \mathfrak{M}$ and $\varepsilon > 0$, using properties (a), (b) and (c), we can find an open set V and a compact set K such that $K \subset E \subset V$ and

$$\mu_2(V) - \varepsilon/2 < \mu(E) < \mu_2(K) + \varepsilon/2$$

then, $\mu_2(V) < \mu_2(K) + \varepsilon$. Moreover, by Lemma 2.2.2, exists a continuous function f such that $K \prec f \prec V$. Hence,

$$\begin{aligned}\mu_1(K) = \int_I \chi_K d\mu_1 \leq \int_I f d\mu_1 = \Lambda f = \int_I f d\mu_2 \leq \\ \leq \int_I \chi_V d\mu_2 = \mu_2(V) < \mu_2(K) + \varepsilon.\end{aligned}$$

which holds for any arbitrary $\varepsilon > 0$. Hence, $\mu_1(K) \leq \mu_2(K)$.

Analogously, one can see that $\mu_2(K) \leq \mu_1(K)$. Since by property (b) these measures are completely determined by the compact sets, necessarily $\mu_1 = \mu_2$, and the uniqueness of μ is proved.

Now, let's see the existence of the σ-algebra $\mathfrak{M}$ and the measure μ.

(i) Construction of μ and $\mathfrak{M}$:

For any open set V in I, define

$$\mu(V) = \sup\{\Lambda f\colon\ f \prec V\}. \tag{2.2.1}$$

If $V_1 \subset V_2$, V_1, V_2 open sets, it is clear that (2.2.1) implies $\mu(V_1) \leq \mu(V_2)$. Hence, we can define

$$\mu(E) = \inf\{\mu(V)\colon\ E \subset V,\ V \text{ open}\} \tag{2.2.2}$$

for every $E \subseteq I$, and it is consistent with (2.2.1) to define $\mu(E)$ by (2.2.2) when E is open.

Now, let $\mathfrak{M}_F$ be the class of all $E \subset I$ which satisfies $\mu(E) < +\infty$ and

$$\mu(E) = \sup\{\mu(K)\colon\ K \subset E,\ K \text{ compact}\}. \tag{2.2.3}$$

Then, we define $\mathfrak{M}$ to be the class of all $E \subset I$ such that $E \cap K \in \mathfrak{M}_F$, for every compact K.

(ii) Proof that μ and $\mathfrak{M}$ have the required properties:

Observe that μ is monotone, since for $A \subset B \subset I$,

$$\mu(A) = \inf\{\mu(V)\colon\ A \subset V,\ V \text{ open}\} \leq \inf\{\mu(V)\colon\ B \subset V,\ V \text{ open}\} = \mu(B).$$

Hence, $\mu(E) = 0$ implies that $\mu(E \cap K) = \mu(K) = 0$ for every $K \subset E$ compact, so $E \in \mathfrak{M}_F$ and $E \in \mathfrak{M}$. Moreover, (d) holds, and so does (b) by definition on (2.2.2).

For the next properties, it will be convenient to divide them into several steps. First, observe that the positivity of Λ implies that Λ is monotone. This is clear, since $\Lambda g = \Lambda f + \Lambda(g - f) \geq \Lambda f$ if $g \geq f$.

STEP I: If $E_1, E_2, E_3, \dots$ are arbitrary subsets of I, then

$$\mu\left(\bigcup_{i=1}^{+\infty} E_i\right) \leq \sum_{i=1}^{+\infty} \mu(E_i). \tag{2.2.4}$$

Let V_1, V_2 open sets in I, and choose $g \prec V_1 \cup V_2$. By Lemma 2.2.3, there are functions h_1 and h_2 such that $h_i \prec V_i$ $(i = 1, 2)$ and $h_1(x) + h_2(x) = 1$ for all $x \in I$. Hence, $gh_i \prec V_i$, $g = gh_1 + gh_2$, and so, by the definition of μ,

$$\Lambda g = \Lambda(gh_1) + \Lambda(gh_2) \leq \mu(V_1) + \mu(V_2). \tag{2.2.5}$$

Now, let $\varepsilon > 0$. By the definition of supremum, exists $f \in \mathcal{C}(I)$, $f \prec V_1 \cup V_2$, such that $\mu(V_1 \cup V_2) < \Lambda f + \varepsilon$. Since (2.2.5) holds for any $g \prec V_1 \cup V_2$, then $\mu(V_1 \cup V_2) < \Lambda f + \varepsilon \leq \mu(V_1) + \mu(V_2) + \varepsilon$, and making ε tends to zero, we get

$$\mu(V_1 \cup V_2) \leq \mu(V_1) + \mu(V_2). \tag{2.2.6}$$

Now observe that if $\mu(E_i) = +\infty$ for some i, then (2.2.4) is trivially true. Suppose therefore that $\mu(E_i) < +\infty$ for every $i \in \mathbb{N}_{\geq 1}$ and choose $\varepsilon > 0$. By (2.2.2) there are open sets $V_i \supset E_i$ such that

$$\mu(V_i) < \mu(E_i) + 2^{-i}\varepsilon \quad (i = 1, 2, 3, \dots).$$

Put $V = \cup_{i=1}^{+\infty} V_i$, and choose $f \prec V$. Observe that since f is continuous and f is zero in V^c, then exists a compact set $K \subseteq I$ such that $\operatorname{supp}(f) \subset K$. Then, by the definition of compact, $f \prec V_1 \cup \cdots \cup V_n$ for some n. Iterating in (2.2.6) we obtain

$$\Lambda f \leq \mu(V_1 \cup \cdots \cup V_n) \leq \mu(V_1) + \cdots \mu(V_n) \leq \sum_{i=1}^{+\infty} \mu(E_i) + \varepsilon.$$

Since this holds for every $f \prec V$, and since $\cup E_i \subset V$, it follows that

$$\mu\left(\bigcup_{i=1}^{+\infty} E_i\right) \leq \mu(V) \leq \sum_{i=1}^{+\infty} \mu(E_i) + \varepsilon,$$

which proves (2.2.4) by making ε tend to zero.

STEP II: $\mathfrak{M}_F$ contains every compact set (observe that this implies property (a) because I is a compact subset of itself).

By (2.2.3), it is sufficient to see that $\mu(K)$ is finite for every compact set K. So fix some compact set K such that $K \prec f$ for some $f \in \mathcal{C}(I)$, and let $V = \{x\colon f(x) > \frac{1}{2}\}$. Then, $K \subset V$ and $g \leq 2f$ whenever $g \prec V$ (because $2f \geq \chi_V \geq g$). Hence,

$$\mu(K) \leq \mu(V) = \sup\{\Lambda g;\ g \prec V\} \leq \Lambda(2f) < +\infty.$$

Since K clearly satisfies (2.2.3), $K \in \mathfrak{M}_F$. Then, in particular we have that $\mu(E) < +\infty$, for every $E \subset I$.

STEP III: Every open set satisfies (2.2.3) *(then, $\mathfrak{M}_F$ contains every open set V, since we have that property (a) holds).*

Let V be an open set. Observe that the case $\mu(V) = 0$ is trivial. Then assume that $\mu(V) \neq 0$ and let α be a real number such that $0 < \alpha < \mu(V)$. So, there exists an $f \prec V$ with $\alpha < \Lambda f$. Now observe that if W is any open set which contains the support K of f, then $f \prec W$, hence $\Lambda f \leq \mu(W)$.

Given $\varepsilon > 0$, then exists $W_\varepsilon \supset K$ such that $\mu(W_\varepsilon) < \mu(K) + \varepsilon$, by the definition of infimum. Thus, $\Lambda f \leq \mu(W_\varepsilon) < \mu(K) + \varepsilon$, and by making ε tend to zero, we get that $\Lambda f \leq \mu(K)$. This exhibits a compact $K \subset V$ with $\alpha < \mu(K) \leq \mu(V)$ for any α satisfying $\alpha < \mu(V)$. Then taking $\alpha = \mu(V) - \delta$, for $\delta > 0$, we have that $\mu(V) - \delta < \mu(K) \leq \mu(V)$, and making δ tend to zero, we see that (2.2.3) holds for V.

STEP IV: Suppose $E = \cup_{i=1}^{+\infty} E_i$, where $E_1, E_2, E_3, \dots$ are pairwise disjoint members of $\mathfrak{M}_F$. Then

$$\mu(E) = \sum_{i=1}^{+\infty} \mu(E_i). \tag{2.2.7}$$

In addition, $E \in \mathfrak{M}_F$.

Let K_1 and K_2 be two disjoint compact sets. Then exist two disjoint open sets V_1 and V_2 such that $K_1 \subset V_1 \subset \overline{V_1} \subset K_2^c$ and $K_2 \subset V_2 \subset \overline{V_2} \subset K_1^c$. Choose $\varepsilon > 0$. By the definition of μ, there is an open set $W \supset K_1 \cup K_2$ such that $\mu(W) < \mu(K_1 \cup K_2) + \varepsilon/3$, and there are functions $f_i \prec W \cap V_i$ such that $\Lambda f_i > \mu(W \cap V_i) - \varepsilon/3$, $(i = 1, 2)$.

Since $K_i \subset W \cap V_i$ and since $V_1 \cap V_2 = \emptyset$, we have that $f_1 + f_2 \prec (W \cap V_1) \cup (W \cap V_2) \subset W$. So, we obtain

$$\begin{aligned}\mu(K_1) + \mu(K_2) &\leq \mu(W \cap V_1) + \mu(W \cap V_2) < \Lambda f_1 + \Lambda f_2 + \frac{2\epsilon}{3} \\ &\leq \mu(W) + \frac{2\epsilon}{3} < \mu(K_1 \cup K_2) + \varepsilon.\end{aligned}$$

Since ε was arbitrary, from *STEP I* it follows that

$$\mu(K_1 \cup K_2) = \mu(K_1) + \mu(K_2). \tag{2.2.8}$$

Now, due to $E \subset I$, we have that $\mu(E) \leq \mu(I) < +\infty$. So choose $\varepsilon > 0$. Since $E_i \in \mathfrak{M}_F$, there are compact sets $H_i \subset E_i$ with

$$\mu(H_i) > \mu(E_i) - 2^{-i}\varepsilon \quad (i = 1, 2, \dots).$$

Putting $K_n = H_1 \cup \cdots \cup H_n$, and applying induction on (2.2.8), we obtain

$$\mu(E) \geq \mu(K_n) = \sum_{i=1}^{n} \mu(H_i) > \sum_{i=1}^{n} \Big(\mu(E_i) - 2^{-i}\varepsilon\Big). \tag{2.2.9}$$

Since (2.2.9) is true for all n and every $\varepsilon > 0$, using *STEP I*, it follows (2.2.7).

Besides, since

$$\sum_{i=1}^{+\infty} \mu(E_i) = \mu(E) \leq \mu(I) < +\infty,$$

for every $\varepsilon > 0$ there exists some $N := N(\varepsilon) \in \mathbb{N}$ such that

$$\mu(E) \leq \sum_{i=1}^{N} \mu(E_i) + \varepsilon.$$

By (2.2.9), it follows that $\mu(E) \leq \mu(K_N) + 2\epsilon$, and this shows that $E \in \mathfrak{M}_F$.

STEP V: If $E \in \mathfrak{M}_F$ and $\varepsilon > 0$, there is a compact K and an open V such that $K \subset E \subset V$ and $\mu(V \setminus K) < \varepsilon$.

Our definition of μ shows that there exist a compact set K and an open set V so that

$$\mu(V) - \frac{\varepsilon}{2} < \mu(E) < \mu(K) + \frac{\varepsilon}{2}.$$

Since $V \setminus K$ is open, by *STEP III* we get that $V \setminus K \in \mathfrak{M}_F$. Hence, *STEP IV* implies that

$$\mu(K) + \mu(V \setminus K) = \mu(V) < \mu(K) + \varepsilon.$$

STEP VI: If $A \in \mathfrak{M}_F$ and $B \in \mathfrak{M}_F$, then $A \setminus B$, $A \cup B$, and $A \cap B$ belong to $\mathfrak{M}_F$.

If $\varepsilon > 0$, *STEP V* shows that there are compact sets K_i and open sets V_i $(i = 1, 2)$ such that $K_1 \subset A \subset V_1$, $K_2 \subset B \subset V_2$, and $\mu(V_i \setminus K_i) < \varepsilon$, for $i = 1, 2$.

Since

$$A \setminus B \subset V_1 \setminus K_2 \subset (V_1 \setminus K_1) \cup (K_1 \setminus V_2) \cup (V_2 \setminus K_2),$$

using *STEP I* we get

$$\mu(A \setminus B) \leq \mu(V_1 \setminus K_1) + \mu(K_1 \setminus V_2) + \mu(V_2 \setminus K_2) < \mu(K_1 \setminus V_2) + 2\epsilon. \tag{2.2.10}$$

Since $K_1 \setminus V_2$ is a compact subset of $A \setminus B$, (2.2.10) shows that $A \setminus B$ satisfies (2.2.3), hence $A \setminus B \in \mathfrak{M}_F$.

Now $A \cup B = (A \setminus B) \cup B$, so it follows (by *STEP IV*) that $A \cup B \in \mathfrak{M}_F$. Finally, $A \cap B = A \setminus (A \setminus B) \in \mathfrak{M}_F$.

STEP VII: $\mathfrak{M}$ is a σ-algebra in I containing all the Borel sets of I.

Let K be an arbitrary compact set in I (then $K \in \mathfrak{M}_F$). If $A \in \mathfrak{M}$, then $A^c \cap K = K \setminus (A \cap K)$, so that $A^c \cap K$ is a difference of two members of $\mathfrak{M}_F$. Hence, $A^c \cap K \in \mathfrak{M}_F$, and we conclude: $A \in \mathfrak{M}$ implies $A^c \in \mathfrak{M}$.

Next, suppose $A = \cup_{i \in \mathbb{N}} A_i$, where $A_i \in \mathfrak{M}$. Put $B_1 = A_1 \cap K$, and

$$B_n = (A_n \cap K) \setminus (B_1 \cup \cdots \cup B_{n-1}) \quad (n = 2, 3, \dots).$$

Then, $\{B_n\}_{n\in\mathbb{N}}$ is a disjoint sequence of members of $\mathfrak{M}_F$. Since $A \cap K = \cup_{n\in\mathbb{N}} B_n$, by *STEP IV*, it belongs to $\mathfrak{M}_F$. Hence, $A \in \mathfrak{M}$.

Finally, if C is closed, then $C \cap K$ is compact, hence $C \cap K \in \mathfrak{M}_F$, so $C \in \mathfrak{M}$ (in particular $I \in \mathfrak{M}$).

We have thus proved that $\mathfrak{M}$ is a σ-algebra in I which contains all closed subsets of I. Hence, $\mathfrak{M}$ contains all Borel sets in I.

STEP VIII: $\mathfrak{M}_F = \mathfrak{M}$ *(this implies assertion* (c) *of the theorem).*

If $E \in \mathfrak{M}_F$, *STEP II* and *STEP VI* imply that $E \cap K \in \mathfrak{M}_F$ for every compact K, hence, taking $K = I$ we see that $E \in \mathfrak{M}$.

Conversely, suppose $E \in \mathfrak{M}$, and choose $\varepsilon > 0$. By *STEP III* and *STEP V*, there is an open set $V \supset E$, and there is a compact set $K \subset E$ with $\mu(V \setminus K) < \varepsilon$. Since $E \cap K \in \mathfrak{M}_F$, by the definition of supremum, there is a compact $H \subset E \cap K$ with

$$\mu(E \cap K) < \mu(H) + \varepsilon.$$

Since $E \subset (E \cap K) \cup (V \setminus K)$, it follows that

$$\mu(E) \leq \mu(E \cap K) + \mu(V \setminus K) < \mu(H) + 2\epsilon,$$

which implies that $E \in \mathfrak{M}_F$.

STEP IX: μ *is a positive measure on* $\mathfrak{M}$.

It is a direct consequence of *STEP IV* and *STEP VIII.*

STEP X: For every $f \in \mathcal{C}(I)$, $\Lambda f = \int_I f d\mu$ *(this proves the main part of the theorem).*

Observe that if f is complex, then $f = u + iv$, so it is enough to prove this for a real f. Also, it is enough to prove the inequality

$$\Lambda f \leq \int_I f d\mu \tag{2.2.11}$$

for every real $f \in \mathcal{C}(I)$, since for once (2.2.11) is established, the linearity of Λ shows that

$$-\Lambda f = \Lambda(-f) \leq \int_I (-f) d\mu = -\int_I f d\mu,$$

which, together with (2.2.11), shows that equality holds.

So, let $[a, b]$ be an interval which contains the range of f. Choose $\varepsilon > 0$, and choose y_i, for $i = 1, \ldots, n := n(\varepsilon)$, so that $y_i - y_{i-1} < \varepsilon$ and $y_0 < a < y_1 < \cdots < y_n = b$. Put

$$E_i = \{x\colon\ y_{i-1} < f(x) \leq y_i\} \quad (i = 1, 2, \ldots, n).$$

Since f is continuous, f is Borel measurable, and the sets E_i are therefore disjoint Borel sets whose union is I. By the definition of μ, there are open sets $V_i \supset E_i$ such that

$$\mu(V_i) < \mu(E_i) + \frac{\varepsilon}{n} \quad (i = 1, \ldots, n)$$

and such that $f(x) < y_i + \varepsilon$ for all $x \in V_i$. By Lemma 2.2.3, there are functions $h_i \prec V_i$ such that $\sum_{i=1}^n h_i = 1$ on I. Hence, $f = \sum_{i=1}^n h_i f$. Since $h_i f \leq (y_i + \varepsilon)h_i$, and since $y_i - \varepsilon < f(x)$ on E_i, we have

$$\begin{aligned}
\Lambda f &= \sum_{i=1}^n \Lambda(h_i f) \leq \sum_{i=1}^n (y_i + \varepsilon)\Lambda(h_i) \leq \sum_{i=1}^n (y_i + \varepsilon)\mu(V_i) \\
&\leq \sum_{i=1}^n (y_i + \varepsilon)\mu(E_i) + \sum_{i=1}^n (y_i + \varepsilon)\frac{\varepsilon}{n} = \sum_{i=1}^n (y_i - \varepsilon + 2\varepsilon)\mu(E_i) + \sum_{i=1}^n (y_i + \varepsilon)\frac{\varepsilon}{n} \\
&\leq \sum_{i=1}^n (y_i - \varepsilon)\mu(E_i) + 2\epsilon\mu(I) + (b + \varepsilon)\varepsilon = \sum_{i=1}^n \int_{E_i} (y_i - \varepsilon) d\mu + \varepsilon[2\mu(I) + b + \varepsilon] \\
&\leq \sum_{i=1}^n \int_{E_i} f d\mu + \varepsilon[2\mu(I) + b + \varepsilon] = \int_I f d\mu + \varepsilon[2\mu(I) + b + \varepsilon].
\end{aligned}$$

Since ε was arbitrary, (2.2.11) is established, and the proof of the theorem is completed.

□

Observation 2.2.6. $\mu(I) = \|\Lambda\|$, since

$$\mu(I) \geq \sup\{\Lambda f;\ 0 \leq f \leq 1,\ f \in \mathcal{C}(I)\}$$

and $\Lambda(\chi_I) = \mu(I)$.

Definition 2.2.7. A positive measure μ defined on the σ-algebra of all Borel sets in an interval I is called a Borel measure on I. We say that a Borel set $E \subset I$ is outer regular or inner regular if E has property (b) or (c) of Theorem 2.2.5 respectively. If every Borel set in I is both outer and inner regular, μ is called regular.

Observation 2.2.8. The measure of the Riesz Representation Theorem (Theorem 2.2.5) is regular.

2.2.2 Complex Measures

Before talk about the Riesz-Markov-Kakutani Representation Theorem, we must give the definition of a complex measure and the total variation of a complex measure. On this section, we will study them and we will show that the total variation is indeed a positive measure. To do so, we will follow [1].

Definition 2.2.9. Let $\mathfrak{M}$ be a σ-algebra in a set X. We say that a countable collection $\{E_i\}_{i\in\mathbb{N}}$ of members of $\mathfrak{M}$ is a partition of E if

$$E = \bigcup_{i\in\mathbb{N}} E_i$$

and $E_i \cap E_j = \emptyset$ whenever $i \neq j$.

Definition 2.2.10. A complex measure μ on $\mathfrak{M}$ is a complex function on $\mathfrak{M}$ such that is σ-additive, i.e., for $E \in \mathfrak{M}$ we have that

$$\mu(E) = \sum_{i\in\mathbb{N}} \mu(E_i)$$

for any partition $\{E_i\}_{i\in\mathbb{N}}$ of E.

Definition 2.2.11. We define a set function $|\mu|$ on $\mathfrak{M}$ by

$$|\mu|(E) = \sup\left\{\sum_{i\in\mathbb{N}} |\mu(E_i)|\right\} \qquad (E \in \mathfrak{M})$$

where the supremum is being taken over all the partitions $\{E_i\}_{i\in\mathbb{N}}$ of E. This function is called the total variation of μ or the total variation measure.

Observe that if μ is positive, then $\mu = |\mu|$. Moreover, note that $|\mu(E)| \leq |\mu|(E)$. Our next step is to show that $|\mu|$ is in fact a positive finite measure.

Proposition 2.2.12. *The total variation $|\mu|$ of a complex measure μ on $\mathfrak{M}$ is a positive measure on $\mathfrak{M}$.*

Proof. Let $E \in \mathfrak{M}$, observe that clearly $|\mu|(E) \geq |\mu(E)| \geq 0$, then $|\mu|$ is a positive function on $\mathfrak{M}$.

Now let's fix $\{E_i\}_i$ a partition of E, we want to see that $|\mu|(E) = \sum_i |\mu|(E_i)$.

First, choose $t_i \in \mathbb{R}$ such that $|\mu|(E_i) \geq t_i \geq 0$ for each $i \geq 1$ (for example, $t_i = (1-\varepsilon)|\mu|(E_i)$, for $\varepsilon > 0$). Then, by the definition of $|\mu|$, we can find a partition $\{A_{ij}\}_j$ of E_i such that

$$\sum_j |\mu(A_{ij})| \geq t_i.$$

Hence, due to $\{A_{ij}\}_{ij}$ is a partition of E, it follows that

$$\sum_i t_i \leq \sum_i \sum_j |\mu(A_{ij})| \leq |\mu|(E).$$

Therefore, since

$$(1-\varepsilon)\sum_i |\mu|(E_i) = \sum_i t_i \leq |\mu|(E)$$

for any ε, making ε tend to zero, we see that

$$\sum_i |\mu|(E_i) \leq |\mu|(E).$$

To prove the opposite inequality, let $\{A_j\}_j$ be any partition of E, then for any fixed j, $\{E_i \cap A_j\}_i$ is a partition of A_j, and for any fixed i, $\{E_i \cap A_j\}_j$ is a partition of E_i. Hence, using that μ is a complex measure,

$$\sum_j |\mu(A_j)| = \sum_j \left|\sum_i \mu(A_j \cap E_i)\right| \leq \sum_j \sum_i |\mu(A_j \cap E_i)| = \sum_i \sum_j |\mu(A_j \cap E_i)| \leq \sum_i |\mu|(E_i).$$

Since this inequality works for any partition $\{A_j\}_j$ of E, taking the supremum over all of them we see that

$$|\mu|(E) \leq \sum_i |\mu|(E_i).$$

□

It reminds to see that $|\mu|(X) < +\infty$ for every complex measure μ on a set X. To do so, we will take use of the following technical lemma.

Lemma 2.2.13. *If $z_1, \dots, z_n \in \mathbb{C}$, $n \in \mathbb{N}$, there is $S \subseteq \{1, \dots, n\}$ such that*

$$\left|\sum_{j\in S} z_j\right| \leq \frac{1}{6}\sum_j |z_j|.$$

Proof. Put $w = |z_1| + \cdots + |z_n|$ and consider $\mathbb{C}$ as the union of four closed quadrants bounded by the lines $y = \pm x$. At least, there is one of them containing $z_{i_1}, \dots, z_{i_m}$ satisfying $|z_{i_1}| + \cdots + |z_{i_m}| \geq \frac{1}{4}w$, for $m \in \mathbb{N}$ and $m \leq n$. Let's consider $S = \{i_1, \dots, i_m\}$ and let Q be that quadrant. Since $|\sum_{k=1}^m z_{i_k} e^{i\theta}| = |\sum_{k=1}^m z_{i_k}|$, by an argument of rotation of the elements z_i's, we can assume without lose of generality that Q is the quadrant defined by $|y| \leq x$. Now observe that if $z = a + ib \in Q$, then

$$|z|^2 = a^2 + b^2 \leq 2a^2 \Rightarrow \mathrm{Re}(z) \geq \frac{|z|}{\sqrt{2}}.$$

Finally,

$$\left|\sum_{j\in S} z_j\right| = \left|\sum_{j\in S} (\mathrm{Re}(z_j) + i\,\mathrm{Im}(z_j))\right| \geq \sum_{j\in S} \mathrm{Re}(z_j) \geq \frac{1}{\sqrt{2}}\sum_{j\in S}|z_j| \geq \frac{1}{\sqrt{2}}\frac{w}{4} \geq \frac{w}{6}.$$

□

Proposition 2.2.14. *If μ is a complex measure on X, then $|\mu|(X) < +\infty$.*

Proof. We first show that if $|\mu|(E) = +\infty$ for some $E \in \mathfrak{M}$, then $E = A \cup B$, where $A, B \in \mathfrak{M}$ are disjoint and $|\mu(A)| > 1$, $|\mu|(B) = +\infty$.

Let's take $E \in \mathfrak{M}$. The definition of $|\mu|$ shows that for every $t < +\infty$, there corresponds a partition $\{E_j\}_j$ of E such that $\sum_j |\mu(E_j)| > t$. Let us take

$$t = 6(1 + |\mu(E)|) \leq 6(1 + |\mu(X)|) < +\infty.$$

Then, since $\sum_j |\mu(E_j)| > t$ for some $\{E_j\}_j$ partition of E, there exists some $n \in \mathbb{N}$ such that

$$\sum_{j=1}^n |\mu(E_j)| > t.$$

Let $z_j = \mu(E_j)$, then by the Lemma 2.2.13, exists some set $S \subseteq \{1, \dots, n\}$ such that

$$\left|\sum_{j\in S} z_j\right| \leq \frac{1}{6}\sum_j |z_j|.$$

Let $A = \cup_{j\in S} E_j$, then it follows that

$$A \subset E \text{ and } |\mu(A)| = \left|\sum_{j\in S} \mu(E_j)\right| \geq \frac{1}{6}\sum_{j=0}^n |\mu(E_j)| > \frac{t}{6} = 1 + |\mu(E)| \geq 1.$$

If we take $B = E \setminus A$, then

$$|\mu(B)| = |\mu(E) - \mu(A)| \geq |\mu(A)| - |\mu(E)| > \frac{t}{6} - |\mu(E)| = 1.$$

Since $|\mu|(E) = |\mu|(A) + |\mu|(B)$ and $|\mu|(E) = +\infty$, either $|\mu|(A) = +\infty$ or $|\mu|(B) = +\infty$ (or both). So, we get the claim.

Now assume that $|\mu|(X) = +\infty$. We define a sequence of sets $\{A_n\}_n$ and $\{B_n\}_n$ of X as follows:

Put $A_0 = \emptyset$ and $B_0 = X$. Then, to construct the following sets for $n \geq 1$, we apply the previous claim to B_{n-1} ($|\mu|(B_{n-1}) = +\infty$) choosing B_n as the set B and A_n as the set A of the claim. Then, we see that $A_n, B_n \subset B_{n-1}$, $A_n \cap B_n = \emptyset$, $|\mu(A_n)| > 1$ and $|\mu|(B_n) = +\infty$.

We does inductively obtain disjoint sets $A_1, A_2, A_3, \ldots$ with $|\mu(A_n)| > 1$ for every $n \geq 1$. If $C = \cup_n A_n$,

$$\mu(C) = \sum_{n\geq 1} \mu(A_n).$$

But this series can't converge, since $\mu(A_n)$ does not tend to zero. This contradiction shows that $|\mu|(X) < +\infty$ must hold. ☐

Definition 2.2.15. If μ and λ are complex measures on the same σ-algebra $\mathfrak{M}$, we define $\mu+\lambda$ and $c\mu$ by

$$\begin{aligned} (\mu + \lambda)(E) &= \mu(E) + \lambda(E), \\ (c\mu)(E) &= c\mu(E), \end{aligned} \qquad (E \in \mathfrak{M})$$

for any scalar $c \in \mathbb{C}$. It is then trivial to verify that $\mu + \lambda$ and $c\mu$ are complex measures. Then, the collection of all complex measures on $\mathfrak{M}$, denoted by M, is thus a vector space. If we put

$$\|\mu\| = |\mu|(X)$$

it is easy to verify that all axioms of a normed linear space are satisfied.

2.2.3 The Radon-Nikodym Theorem

We now turn to the Radon-Nikodym Theorem, which is probably one of the most important theorems in measure theory. This theorem concerns the concept of absolute continuity which gives a certain meaning of continuity of the complex measures.

Definition 2.2.16. Let μ be a positive measure on a σ-algebra $\mathfrak{M}$, λ an arbitrary measure on $\mathfrak{M}$ (positive or complex), we say that λ is absolutely continuous respect to μ, and write $\lambda << \mu$, if $\lambda(E) = 0$ for each $E \in \mathfrak{M}$ such that $\mu(E) = 0$.

If there is a set $A \in \mathfrak{M}$ with $\lambda(E) = \lambda(E \cap A)$ for all $E \in \mathfrak{M}$, we say that λ is concentrated on A, which is equivalent to $\lambda(E) = 0$ for each $E \cap A = \emptyset$.

Definition 2.2.17. Suppose λ_1, λ_2 are two measures on $\mathfrak{M}$ (positives or complexes), and suppose that exist $A, B \in \mathfrak{M}$ such that $A \cap B = \emptyset$, λ_1 is concentrated on A and λ_2 is concentrated on B. Then we say that λ_1 and λ_2 are mutually singular and write $\lambda_1 \perp \lambda_2$.

Proposition 2.2.18. *Suppose μ, λ, λ_1, λ_2 are measures on a σ-algebra $\mathfrak{M}$ and μ is positive:*

(a) If λ is concentrated on A, so is $|\lambda|$.

(b) If $\lambda_1 \perp \lambda_2$, then $|\lambda_1| \perp |\lambda_2|$.

(c) If $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 + \lambda_2 \perp \mu$.

(d) If $\lambda_1 << \mu$ and $\lambda_2 << \mu$, then $\lambda_1 + \lambda_2 << \mu$.

(e) If $\lambda << \mu$, then $|\lambda| << \mu$.

(f) If $\lambda_1 << \mu$ and $\lambda_2 \perp \mu$, then $\lambda_1 \perp \lambda_2$.

(g) If $\lambda << \mu$ and $\lambda \perp \mu$, then $\lambda = 0$.

Proof. (a) If $E \cap A = \emptyset$ and $\{E_j\}_j$ is a partition of E, then since λ is concentrated on A, $\lambda(E_j) = 0$ for all j, so $|\lambda|(E) = 0$.

(b) Let $A, B \in \mathfrak{M}$ such that $A \cap B = \emptyset$, λ_1 is concentrated on A and λ_2 is concentrated on B. Then from (a), $|\lambda_1|$ is concentrated on A and $|\lambda_2|$ is concentrated on B. Thus, $|\lambda_1| \perp |\lambda_2|$.

(c) Let $A_1, B_1 \in \mathfrak{M}$ such that $A_1 \cap B_1 = \emptyset$, λ_1 is concentrated on A_1 and μ is concentrated on B_1, and let $A_2, B_2 \in \mathfrak{M}$ such that $A_2 \cap B_2 = \emptyset$, λ_2 is concentrated on A_2 and μ is concentrated on B_2. Taking $A = A_1 \cup A_2$ and $B = B_1 \cap B_2$, we have that for all $E \in \mathfrak{M}$ such that $E \cap A = \emptyset$ (then $E \cap A_1 = E \cap A_2 = \emptyset$)

$$(\lambda_1 + \lambda_2)(E) = \lambda_1(E) + \lambda_2(E) = 0,$$

and if $E \cap B = \emptyset$,

$$\mu(E) = \mu(E \cap B_1 \cap B_2) = 0.$$

Moreover, $(A_1 \cup A_2) \cap (B_1 \cap B_2) = \emptyset$. Therefore, $\lambda_1 + \lambda_2 \perp \mu$.

(d) Obvious.

(e) Let $\{E_j\}_j$ be a partition of $E \in \mathfrak{M}$ and $\mu(E) = 0$. Since μ is positive, $\mu(E_j) = 0$ for each j. So, $\lambda(E_j) = 0$ for all j. Hence, $|\lambda|(E) = 0$.

(f) Since $\lambda_2 \perp \mu$, there is a set A where $\mu(A) = 0$ and where λ_2 is concentrated. Then, $\lambda_1(A) = 0$ ($\lambda_1 << \mu$). Hence, for any $E \subset A$, $\lambda_1(E) = 0$, and so λ_1 is concentrated on $B \subseteq A^c$.

(g) By (f), we have that $\lambda \perp \lambda$. Hence, $\lambda = 0$. □

The following proposition, gives us a continuity sense for measures that are absolutely continuous.

Proposition 2.2.19. *Suppose μ and λ are measures on a σ-algebra $\mathfrak{M}$, μ is positive and λ is complex. Then the following two conditions are equivalent:*

(a) $\lambda << \mu$,

(b) $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $|\lambda(E)| < \varepsilon$ for all $E \in \mathfrak{M}$ with $\mu(E) < \delta$.

Proof. If (b) holds, let $E \in \mathfrak{M}$ such that $\mu(E) = 0$. Then, for any $\varepsilon > 0$, $|\lambda(E)| < \varepsilon$ (because $\mu(E) < \delta$ for any $\delta > 0$). Hence, $\lambda(E) = 0$.

If (b) is false, then $\exists\varepsilon > 0$ and there exist sets $E_n \in \mathfrak{M}$ ($n = 1, 2, \dots$) such that $\mu(E_n) < 2^{-n}$ but $|\lambda(E_n)| \geq \varepsilon$. Then, in particular, $|\lambda|(E_n) \geq \varepsilon$.

Let $A_n = \cup_{i\geq n} E_i$, and define $A = \cap_{n\geq 1} A_n$. Then $\mu(A_n) \leq \sum_{i\geq n} \mu(E_i) < \sum_{i\geq n} 2^i = 2^{1-n}$ and $A_{n+1} \subseteq A_n$. Hence,

$$\mu(A) = \lim_{n\to+\infty} \mu(A_n) \leq \lim_{n\to+\infty} 2^{1-n} = 0.$$

Moreover, $|\lambda|(A) \geq \lim_{n\to+\infty} |\lambda|(A_n) \geq \varepsilon > 0$, since $|\lambda|(A_n) \geq |\lambda|(E_n)$.

Thus, $|\lambda|$ is not absolute continuous respect to μ. Thus, by Proposition 2.2.18 (e), λ is not absolute continuous respect to μ.

□

The following two lemmas, which both are concerned with measurable functions, will be used many times on what follows in this chapter.

Lemma 2.2.20. *Suppose μ is a complex measure on a σ-algebra $\mathfrak{M}$, $f \in L^1(\mu)$, S is a closed set in the complex plane, and the averages*

$$A_E(f) = \frac{1}{\mu(E)} \int_E f d\mu$$

lie in S for all $E \in \mathfrak{M}$ with $\mu(E) > 0$. Then, $f(x) \in S$ a.e. $x \in X$.

Proof. We will see that if $E \in f^{-1}(S^c)$, then $\mu(E) = 0$. Let Δ be a closed circular disc, $\Delta = \overline{D(\alpha, r)}$ for some $\alpha \in S^c$ and $r > 0$, such that $\Delta \subset S^c$. Since $S^c \subseteq \mathbb{C}$ is open, S^c is the union of countable many such discs. So, it is enough to prove that $\mu(E) = 0$ where $E = f^{-1}(\Delta)$.

If we have that $\mu(E) > 0$, then

$$|A_E(f) - \alpha| = \frac{1}{\mu(E)} \left| \int_E (f - \alpha) d\mu \right| \leq \frac{1}{\mu(E)} \int_E |f - \alpha| \, d\mu \leq r \frac{1}{\mu(E)} \mu(E) = r,$$

which the last inequality holds because $f(E) \subset \Delta$. But this means that $A_E(f) \in \Delta \subset S^c$, which is impossible (by hypothesis, $A_E(f) \in S$). Thus, $\mu(E) = 0$.

□

Lemma 2.2.21. *Let μ be a positive measure on a σ-algebra $\mathfrak{M}$ in a set X:*

(a) Suppose $f : X \to [0, +\infty]$ is measurable, $E \in \mathfrak{M}$, and $\int_E f d\mu = 0$. Then $f = 0$ a.e. on E.

(b) Suppose $f \in L^1(\mu)$ and $\int_E f d\mu = 0$ for every $E \in \mathfrak{M}$. Then, $f = 0$ a.e. on X.

Proof. Let's first see (a). Take $E \in \mathfrak{M}$. If $A_n = \{x \in E;\ f(x) > 1/n\}$, $n \in \mathbb{N}$, then

$$\frac{1}{n}\mu(A_n) = \frac{1}{n} \int_{A_n} d\mu \leq \int_{A_n} f d\mu = 0,$$

so that $\mu(A_n) = 0$. Since

$$\{x \in E;\ f(x) > 0\} = \cup_n A_n,$$

(a) follows.

Now, we see (b). Put $f = u+iv$, and let $E = \{x;\, u(x) \geq 0\}$. Let u^+ be the positive real part of f, then since $\int_E f d\mu = 0$, we get that $0 = \mathrm{Re}\,(\int_E f d\mu) = \int_E u^+ d\mu$, and (a) implies that $u^+ = 0$ a.e. on X.

We conclude similar that $u^- = v^+ = v^- = 0$ a.e. on X. Thus, $f = 0$ a.e. on X.

□

Now we present the Lebesgue Decomposition Theorem, from which the Radon-Nikodym Theorem follows directly.

Proposition 2.2.22 (Lebesgue Decomposition Theorem). *Let λ be a positive finite measure and let μ be a σ-finite positive measure, both on a σ-algebra $\mathfrak{M}$ in a set X. Then, there is a unique pair of measures λ_a and λ_s on $\mathfrak{M}$ such that*

$$\lambda = \lambda_s + \lambda_a, \quad \lambda_a << \mu, \quad \lambda_s \perp \mu. \tag{2.2.12}$$

Proof. First, let's see the unicity. If λ'_a and λ'_s is another pair that satisfies (2.2.12), then $\lambda_s + \lambda_a = \lambda'_s + \lambda'_a$. So,

$$\lambda_s - \lambda'_s = \lambda'_a - \lambda_a,$$

and by Proposition 2.2.18 (c) and (d),

$$\lambda'_a - \lambda_a << \mu \text{ and } \lambda_s - \lambda'_s \perp \mu.$$

Thus, by Proposition 2.2.18 (g),

$$\lambda'_a = \lambda_a \text{ and } \lambda_s = \lambda'_s.$$

Now, let's see the existence. First we deal with the case when μ is a finite positive measure.

Put $\varphi = \lambda + \mu$, then φ is a positive and finite measure on $\mathfrak{M}$, and we have that

$$\int_X f d\varphi = \int_X f d\lambda + \int_X f d\mu.$$

Observe that if $f \in L^2(\varphi)$, then

$$\left|\int_X f d\lambda\right| \leq \int_X |f| d\lambda \leq \int_X |f| d\varphi \leq \left(\int_X |f|^2 d\varphi\right)^{1/2} \varphi(X)^{1/2} < +\infty,$$

where in the last inequality we have used the *Cauchy-Schwarz inequality.* Then the mapping

$$f \longmapsto \int_X f d\lambda$$

defines a bounded linear operator in the Hilbert space $L^2(\varphi)$. So, since every Hilbert space is isomorphic to its dual, there exists a unique $g \in L^2(\varphi)$ such that

$$\int_X f d\lambda = \int_X f g d\varphi \tag{2.2.13}$$

for every $f \in L^2(\varphi)$.

Put $f = \chi_E$ in (2.2.13), for $E \in \mathfrak{M}$ with $\varphi(E) > 0$. Then it follows that $\lambda(E) = \int_E g d\varphi$. Now, since $0 \leq \lambda \leq \varphi$,

$$0 \leq \frac{\lambda(E)}{\varphi(E)} = \frac{1}{\varphi(E)} \int_X g d\varphi \leq 1.$$

Hence, by Lemma 2.2.20, $g(x) \in [0, 1]$ a.e.x $[\varphi]$.

Since $g \in L^2(\varphi)$, we can assume then that $0 \leq g \leq 1$ for all $x \in X$, without affecting (2.2.13). So we rewrite (2.2.13) in the form

$$\int_X (1 - g) f d\lambda = \int_X g f d\mu \quad (f \in L^2(\varphi)). \tag{2.2.14}$$

Put $A = \{x;\, 0 \leq g(x) < 1\}$ and $B = \{x;\, g(x) = 1\}$, and define

$$\lambda_a(E) = \lambda(A \cap E),\ \lambda_s(E) = \lambda(B \cap E) \quad (E \in \mathfrak{M}).$$

If we take $f = \chi_B$ in (2.2.14), we see that $\mu(B) = 0$. Then $\lambda_s \perp \mu$.

Now, observe that since g is bounded, (2.2.14) holds if we replace f by $(1 + g + \cdots + g^n)\chi_E$, for $n \in \mathbb{N}$ and $E \in \mathfrak{M}$. We then obtain

$$\int_E (1 - g^{n+1}) d\lambda = \int_X (1 - g) f d\lambda = \int_X g f d\mu = \int_E g(1 + g + \cdots + g^n) d\mu. \tag{2.2.15}$$

At every point of B, $g(x) = 1$, hence $1 - g^{n+1}(x) = 0$. However, at every point of A, g^{n+1} converges to 0 monotonically on n. Then the left side of (2.2.15) converges, therefore, by the *monotone convergence theorem*, to $\lambda(A \cap E) = \lambda_a(E)$ as $n \to +\infty$. The integrand on the right side of (2.2.15) increases monotonically to a positive measurable limit

$$h = \frac{g}{1 - g} \chi_A,$$

and the *monotone convergence theorem* shows that the right side tends to $\int_E h d\mu$ as $n \to +\infty$.

Then, we have proved that

$$\lambda_a(E) = \int_E h d\mu \tag{2.2.16}$$

for every $E \in \mathfrak{M}$. In particular, for $E = X$, we see that $h \in L^1(\mu)$, since $\lambda(X) < +\infty$.

Therefore, if $\mu(E) = 0$ for some $E \in \mathfrak{M}$, then

$$0 = \int_E h d\mu = \int_{A \cap E} d\lambda = \lambda_a(E).$$

Thus, $\lambda_a << \mu$ and the theorem follows when μ is a finite positive measure.

Now, if μ is a σ-finite positive measure on $\mathfrak{M}$, then X is the union of countably many sets X_n such that $\mu(X_n) < +\infty$ $(n \geq 1)$. We may assume that the X_n are disjoint, for if not, we replace $\{X_n\}_n$ by $\{Y_n\}_n$, where $Y_1 = X_1$ and $Y_n = X_n \setminus (Y_1 \cup \cdots \cup Y_{n-1})$ for $n \geq 2$. Then, we can apply the same argument to the measures μ and λ_n for each X_n, where λ_n is defined by $\lambda_n(E) := \lambda(E \cap X_n)$ for every $E \in \mathfrak{M}$. Hence, the decomposition for each λ_n and μ add up to a decomposition of λ and μ, since $\lambda(E) = \sum_n \lambda_n(E)$.

□

Definition 2.2.23. The pair λ_a and λ_s is called the Lebesgue decomposition of λ relative to μ.

Finally, we present the Radon-Nikodym Theorem. The point of this theorem is the converse. However, on this section we will not go such further for this theorem.

Theorem 2.2.24 (Radon-Nikodym Theorem). *Let λ be a positive finite measure and let μ be a σ-finite positive measure, both on a σ-algebra $\mathfrak{M}$ in a set X, such that $\lambda << \mu$. Then, there exists a unique $h \in L^1(\mu)$ such that*

$$\lambda(E) = \int_E h d\mu \quad (E \in \mathfrak{M}). \tag{2.2.17}$$

Proof. First, let's see the unicity. If $h' \in L^1(\mu)$ is another function satisfying (2.2.17), then

$$\int_E (h - h')d\mu = 0$$

for every $E \in \mathfrak{M}$. Hence, by Lemma 2.2.21 (b), $h = h'$ a.e. $[\mu]$.

Now, let's see the existence. By Proposition 2.2.22, we can get the Lebesgue decomposition of λ relative to μ

$$\lambda = \lambda_s + \lambda_a, \quad \lambda_a << \mu, \quad \lambda_s \perp \mu.$$

Since $\lambda << \mu$ and $\lambda_s \perp \mu$, then $\lambda_s = 0$. Therefore, $\lambda = \lambda_a$. Thus, we have seen in the proof of Proposition 2.2.22 in (2.2.16) that there exists an $h \in L^1(\mu)$ such that

$$\lambda = \lambda_a(E) = \int_E h d\mu$$

for every $E \in \mathfrak{M}$. This ends the proof.

□

Definition 2.2.25. The function h is called the Radon-Nikodym derivative of λ respect to μ. We may express it in the form $d\lambda = hd\mu$ (or $h = \frac{d\lambda}{d\mu}$).

Now, we introduce the real measures for a σ-algebra $\mathfrak{M}$ on a set X.

Definition 2.2.26. A signed measure μ on $\mathfrak{M}$ is a function defined as

$$\mu : \mathfrak{M} \to \mathbb{R} \cup \{+\infty\}$$

such that is σ-additive, i.e., for $E \in \mathfrak{M}$ we have that

$$\mu(E) = \sum_i \mu(E_i)$$

for any partition $\{E_i\}_i$ of E.

Observation 2.2.27. When a signed measure μ is finite is, in particular, a complex measure. Therefore, we can define its total variation $|\mu|$ as we did for the complex measures.

Definition 2.2.28. Let μ be a finite signed measure on a σ-algebra $\mathfrak{M}$. The positive and negative variations of μ are the positive measures on $\mathfrak{M}$ defined as

$$\mu^+ = \frac{1}{2}\left(|\mu| + \mu\right), \quad \mu^- = \frac{1}{2}\left(|\mu| - \mu\right).$$

Then, observe that μ^+ and μ^- are both bounded and they satisfy

$$\mu = \mu^+ - \mu^-, \quad |\mu| = \mu^+ + \mu^-.$$

This representation of μ as the difference of the positive measures μ^+ and μ^- is known as the Jordan decomposition of μ.

Now, let μ be a σ-finite measure and let λ be a complex measure on $\mathfrak{M}$. Then $\lambda = \lambda_1 + i\lambda_2$ where λ_1 and λ_2 are finite signed measures. So, applying the Lebesgue decomposition and the Radon-Nykodim Theorem to the positive and negative variations of λ_1 and λ_2, we also have the following corollary.

Corollary 2.2.29. *The Lebesgue decomposition and the Radon-Nikodym theorem are valid if μ is a positive σ-finite measure on $\mathfrak{M}$ and if λ is a complex measure on $\mathfrak{M}$.*

2.2.4 Consequences of the Radon-Nikodym Theorem

On this section we see some of the consequences of the Radon-Nikodym Theorem. As an interesting one, we will study the Hahn Decomposition Theorem, which characterizes the positive and the negative variations of a finite signed measure.

Corollary 2.2.30 (The polar representation of μ)**.** *Let μ be a complex measure on a σ-algebra $\mathfrak{M}$ in a set X. There is a measurable function h such that $|h(x)| = 1$ for every $x \in X$ and such that*

$$d\mu = hd|\mu|.$$

Proof. It is obvious that $\mu << |\mu|$ and, therefore, the *Radon-Nikodym Theorem* guarantees the existence of $h \in L^1(|\mu|)$ which satisfies $d\mu = hd|\mu|$. So, let $A_r := \{x \colon |h(x)| < r\}$ where $r > 0$ and let $\{E_j\}_j$ a partition of A_r. Then,

$$\sum_j |\mu(E_j)| = \sum_j \left| \int_{E_j} d\mu \right| = \sum_j \left| \int_{E_j} hd|\mu| \right| \leq r \sum_j |\mu|(E_j) = r|\mu|(A_r).$$

So that, taking the supremum over all the partitions, it follows that $|\mu|(A_r) \leq r|\mu|(A_r)$. Then, if $r < 1$, clearly $|\mu|(A_r) = 0$, so $|h| \geq 1$ a.e. $[|\mu|]$.

Now, let $E \in \mathfrak{M}$ such that $|\mu|(E) > 0$, then

$$\left| \frac{1}{|\mu|(E)} \int_E hd|\mu| \right| = \left| \frac{1}{|\mu|(E)} \int_E d\mu \right| = \left| \frac{\mu(E)}{|\mu|(E)} \right| \leq 1.$$

So, using Lemma 2.2.20 and the fact that $|h| \geq 1$ a.e. $[|\mu|]$, we deduce that $|h| = 1$ a.e. $[|\mu|]$.

Finally, take $B = \{x \in X;\ |h(x)| \neq 1\}$. We have seen that $|\mu|(B) = 0$, so defining $h(x) = 1$ for $x \in B$, we have the desired function.

□

Corollary 2.2.31. *Suppose μ is a positive measure on $\mathfrak{M}$, $g \in L^1(\mu)$ and*

$$\lambda(E) = \int_E g d\mu, \quad (E \in \mathfrak{M}).$$

Then,

$$|\lambda|(E) = \int_E |g| d\mu, \quad (E \in \mathfrak{M}).$$

Proof. First observe that if g is a positive real function a.e. in $[\mu]$, then we are done. Hence, suppose the contrary. Then λ is a complex (and may also real) measure. So, by Corollary 2.2.30, $\exists h \in L^1(\mu)$ with $|h| = 1$ such that $d\lambda = h d|\lambda|$ (i.e., $\lambda(E) = \int_E h d|\lambda|$ for each $E \in \mathfrak{M}$). By hypothesis, $d\lambda = g d\mu$. Therefore, $h d|\lambda| = g d\mu$ as measures. Hence,

$$\int_E d|\lambda| = \int_E \overline{h} h d|\lambda| = \int_E \overline{h} g d\mu \quad (E \in \mathfrak{M}),$$

and we get that $d|\lambda| = \overline{h} g d\mu$ as measures. Since $|\lambda| \geq 0$ and $\mu \geq 0$, it follows that $\overline{h} g \geq 0$ a.e. $[\mu]$. Since h is a complex measurable function, necessarily $\overline{h} g = |g|$ a.e. $[\mu]$. Thus,

$$|\lambda|(E) = \int_E |g| d\mu \quad (E \in \mathfrak{M}).$$

□

Theorem 2.2.32 (The Hahn Decomposition Theorem)**.** *Let μ be a finite signed measure on a σ-algebra $\mathfrak{M}$ in a set X. Then, there exist disjoint sets A and B in $\mathfrak{M}$ such that $A \cup B$, and such that the positive and negative variations of μ satisfy*

$$\mu^+(E) = \mu(A \cap E), \quad \mu^-(E) = -\mu(B \cap E) \quad (E \in \mathfrak{M}).$$

The pair A and B is called the Hahn decomposition of X, induced by μ.

Proof. By Corollary 2.2.30, $d\mu = h d|\mu|$, where $|h| = 1$. Since μ is real, it follows that h is real a.e. in X. Hence, we can redefine h to be real everywhere. Then, $h = \pm 1$. Put

$$A = \{x : h(x) = 1\}, \quad B = \{x : h(x) = -1\}.$$

Since $\mu^+ = \frac{1}{2}(|\mu| + \mu)$, and since

$$\frac{1}{2}(1 + h) = \begin{cases} h & \text{on } A, \\ 0 & \text{on } B, \end{cases}$$

we have, for any $E \in \mathfrak{M}$,

$$\mu^+(E) = \frac{1}{2} \int_E (1 + h) d|\mu| = \int_{E \cap A} h d|\mu| = \mu(E \cap A).$$

Since $\mu(E) = \mu(E \cap A) + \mu(E \cap B)$ and since $\mu = \mu^+ - \mu^-$, it follows that for any $E \in \mathfrak{M}$,

$$\mu^-(E) = -\mu(B \cap E).$$

□

Finally, we can see that the Hahn-Bannach decomposition of a finite signed measure μ is the smallest decomposition in the sense that we state on the following corollary.

Corollary 2.2.33. *If $\mu = \lambda_1 - \lambda_2$, where λ_1 and λ_2 are positive measures, then $\lambda_1 \geq \mu^+$ and $\lambda_2 \geq \mu^-$.*

Proof. Since $\mu \leq \lambda_1$, we have for any $E \in \mathfrak{M}$

$$\mu^+(E) = \mu(E \cap A) \leq \lambda_1(E \cap A) \leq \lambda_1(E).$$

On the other side,

$$\lambda_2 = \lambda_1 - \mu = \lambda_1 - (\mu^+ - \mu^-) = \mu^- + (\lambda_1 - \mu^+) \geq \mu^-.$$

□

2.2.5 Riesz-Markov-Kakutani Representation Theorem on $\mathcal{C}([0,1])$

Now, we are in conditions to study the Riesz-Markov-Kakutani Representation Theorem. To do so, we first introduce some results about Borel complex measures.

Definition 2.2.34. A Borel complex measure is any measure μ defined on $\mathfrak{B}(I)$ (the σ-algebra of the Borel sets). If the total variation of a Borel complex measure is both inner regular and outer regular, it is called a regular Borel complex measure. We denote the space of regular Borel complex measure on the interval I by $M(I)$.

Proposition 2.2.35. *Let μ_1 and μ_2 be two regular Borel complex measures on the interval I, then $\mu_1 - \mu_2$ is also a regular Borel complex measure on I.*

Proof. The difference of Borel complex measures is clearly a Borel complex measure, so that we just have to see that the measure $\mu_1 - \mu_2$ is regular.

First observe that since $|\mu_2| = |-\mu_2|$, the function $-\mu_2$ is a regular Borel complex measure. Hence, let's take $\mu_3 = -\mu_2$. Then, we have to see that the complex Borel measure $\mu_1 + \mu_3$ is regular. So, take an $\varepsilon > 0$ and a measurable set $E \in \mathfrak{B}(I)$. Since both μ_1 and μ_3 are inner regular, there exist compacts sets K_1 and K_3 in I such that $|\mu_i|(E) - |\mu_i|(K_i) < \varepsilon/2$ $(i = 1, 3)$. Hence let $K = K_1 \cup K_3$. Clearly, K is a compact set in I and

$$\begin{aligned}|\mu_1 + \mu_3|(E \setminus K) &\leq (|\mu_1| + |\mu_3|)(E \setminus K) = (|\mu_1| + |\mu_3|)(E) - (|\mu_1| + |\mu_3|)(K) \\ &= (|\mu_1| + |\mu_3|)(E) - |\mu_1|(K_1) - |\mu_3|(K_3) < \varepsilon.\end{aligned}$$

Moreover, since both μ_1 and μ_3 are outer regular, there exist open sets V_1 and V_3 such that $|\mu_i|(V_i) - |\mu_i|(E) < \varepsilon/2$ $(i = 1, 3)$. Hence let $V = V_1 \cap V_3$, then V is also an open set and

$$\begin{aligned} |\mu_1 + \mu_3|(V \setminus E) &\leq (|\mu_1| + |\mu_3|)(V \setminus E) = (|\mu_1| + |\mu_3|)(V) - (|\mu_1| + |\mu_3|)(E) \\ &\leq |\mu_1|(V_1) + |\mu_3|(V_3) - (|\mu_1| + |\mu_3|)(E) < \varepsilon. \end{aligned}$$

Thus, $\mu_1 + \mu_3$ is a regular Borel complex measure.

□

The following result is another of the consequences of the Radon-Nikodym Theorem, and states that the dual of the Lebesgue space L^p (when $1 \leq p < \infty$) is the Lebesgue space L^q, where q is the conjugate exponent of p (that is, $1/p + 1/q = 1$).

Corollary 2.2.36 (Riesz-Representation Theorem for Lebesgue spaces)**.** *Suppose* $1 \leq p < \infty$, *μ is a finite positive measure on a σ-algebra $\mathfrak{M}$ in the interval I, and Φ is a bounded linear functional on $L^p(\mu)$. Then, there is a unique $g \in L^q(\mu)$ such that*

$$\Phi(f) = \int_I fg d\mu \qquad (f \in L^p(\mu)), \tag{2.2.18}$$

where $\frac{1}{p} + \frac{1}{q} = 1$*. Moreover, if Φ and g are related as* (2.2.18)*, we have*

$$\|\Phi\| = \|g\|_{L^q(\mu)} := \|g\|_q \, .$$

In other words, $L^q(\mu)$ is the dual space of $L^p(\mu)$, under the stated conditions.

Proof. The uniqueness of g is clear, for if g and g' satisfy the properties, then taking $f = \chi_E$, the integral of $g - g'$ over any measurable set E is zero, then by Lemma 2.2.21 (b), $g = g'$ a.e.

Moreover, if $\Phi(f) = \int_I fg d\mu$ for any $f \in L^p(\mu)$, by the Hölder's inequality, we get that $\|\Phi\| \leq \|g\|_\infty$. So it remains to prove that the other inequality holds and that such g exists.

First, observe that if $\|\Phi\| = 0$, taking $g = 0$ we are done. So assume $\|\Phi\| > 0$.

For any measurable set $E \subset I$, define

$$\lambda(E) := \Phi(\chi_E).$$

Since Φ is linear, and since $\chi_{A \cup B} = \chi_A + \chi_B$ if A and B are disjoint, we deduce that λ is additive. To prove countable additivity, suppose E is the union of countably many disjoint measurable sets E_i, put $A_k = E_1 \cup \cdots \cup E_k$, and note that

$$\|\chi_E - \chi_{A_k}\|_p = [\mu(E - A_k)]^{1/p} \to 0$$

as $k \to \infty$ (since $p < +\infty$). The continuity of Φ now shows that $\lambda(E - A_k) \to 0$ as $k \to \infty$. So λ is a complex measure.

Observe now that if $\mu(E) = 0$, then $\lambda(E) = \Phi(\chi_E) = 0$ (since $\chi_E = 0$ a.e. in μ), so $\lambda << \mu$. Hence, the *Radon-Nikodym Theorem* ensures the existence $g \in L^1(\mu)$ such that, for every measurable $E \subset I$,

$$\Phi(\chi_E) = \lambda(E) = \int_E g d\mu = \int_I \chi_E g d\mu.$$

By linearity, it follows that

$$\Phi(f) = \int_I fg d\mu \tag{2.2.19}$$

holds for every simple measurable f, and so for every $f \in L^1(\mu)$.

We want to conclude that $g \in L_q(\mu)$ and that $\|\Phi\| = \|g\|_q$; it is best to split the argument into two cases:

(a) Case 1 ($p = 1$): We have that

$$\left|\int_E g d\mu\right| = |\Phi(\chi_E)| \leq \|\Phi\| \, \mu(E)$$

for every $E \in \mathfrak{M}$. By Lemma 2.2.20, $|g(x)| \leq \|\Phi\|$ a.e. Thus, $\|g\|_\infty \leq \|\Phi\|$.

(b) Case 2 ($1 < p < +\infty$): Observe that since μ is a finite positive measure, then $L^p(\mu) \subset L^1(\mu)$ (for every $1 < p \leq +\infty$), then (2.2.19) is well defined for any $f \in L^p(\mu)$, $1 < p < +\infty$.

Let α be a measurable function such that $|\alpha| = 1$ and $\alpha g = |g|$. Let $E_n = \{x \colon |g(x)| \leq n\}$, and put $f = \chi_{E_n}|g|^{q-1}\alpha$. Then, $|f|^p = |g|^{(q-1)p} = |g|^q$ on E_n, $f \in L^p(\mu)$, and (2.2.19) gives

$$\begin{aligned} \int_{E_n} |g|^q d\mu &= \int_I \chi_{E_n} \alpha g |g|^{q-1} \, d\mu = \int_I fg d\mu = \Phi(f) \\ &\leq \|\Phi\| \, \|f\|_p = \|\Phi\| \left(\int_{E_n} |g|^q\right)^{1/p}, \end{aligned}$$

so that

$$\int_I \chi_{E_n}|g|^q d\mu \leq \|\Phi\|^q,$$

for every $n = 1, 2, 3, \ldots$. Applying the monotone convergence to $h_n := \chi_{E_n}|g|^q \in L^1(\mu)$ ($0 \leq h_n \leq h_{n+1}$), we obtain $\|g\|_q \leq \|\Phi\|$.

Thus, for any $1 \leq p < +\infty$, $\|g\|_q = \|\Phi\|$ and $g \in L^q(\mu)$.

□

Finally, we are in conditions to state and proof the most important result in this section.

Theorem 2.2.37 (Riesz-Markov-Kakutani Representation Theorem). *To each bounded linear functional Φ on $\mathcal{C}(I)$ there corresponds a unique complex regular Borel measure μ such that*

$$\Phi(f) = \int_I f d\mu = \langle f, \mu \rangle \quad (f \in \mathcal{C}(I)). \tag{2.2.20}$$

Moreover, if Φ and μ are related as in (2.2.20)*, then $\|\Phi\| = \|\mu\|$.*

Proof. We first settle the uniqueness question. Suppose $\mu \in M(I)$ and $\int_I f d\mu = 0$ for every $f \in \mathcal{C}(I)$. By Corollary 2.2.30, there is a Borel function h (measurable in $d|\mu|$), with $|h| = 1$ such that $d\mu = hd|\mu|$. Then, for any sequence $\{f_n\}_n$ in $\mathcal{C}(I)$ we have that

$$\begin{aligned}|\mu|(I) = \int_I d|\mu| = \int_I |h|^2 d|\mu| + \int_I f_n d\mu = \int_I |h|^2 d|\mu| + \int_I f_n h d|\mu| = \\ = \int_I (\overline{h} - f_n) h d|\mu| = \left| \int_I (\overline{h} - f_n) h d|\mu| \right| \leq \int_I |\overline{h} - f_n| d|\mu|,\end{aligned} \tag{2.2.21}$$

and since $\mathcal{C}(I)$ is dense in $L^1(|\mu|)$ (indeed, the continuous functions with compact support in $\mathbb{R}$ are dense in $L^1(|\mu|)$), $\{f_n\}_n$ can be so chosen such that the last expression in (2.2.21) tends to zero as $n \to +\infty$. Then we get that $|\mu|(I) = 0$. So, $|\mu(E)| \leq |\mu|(E) \leq |\mu|(I) = 0$ for any $E \in \mathfrak{B}(I)$. Hence, $\mu = 0$.

Finally, since the difference of two regular Borel complex measures on I is also regular, we get the unicity.

Now consider a given bounded linear functional Φ on $\mathcal{C}(I)$. Observe that if $\Phi = 0$, taking $\mu = 0$ we are done. Then, without lose of generality assume $\|\Phi\| = 1$ (otherwise take $\Phi/\|\Phi\|$). We will construct a positive linear functional Λ on $\mathcal{C}(I)$, such that

$$|\Phi(f)| \leq \Lambda(|f|) \leq \|f\|_\infty \quad (f \in \mathcal{C}(I)), \tag{2.2.22}$$

where $\|f\|_\infty$ denotes the supremum norm.

Let $\mathcal{C}^+(I)$ the set of positive continuous functions and let

$$\Lambda f = \sup\{|\Phi(h)|;\, h \in \mathcal{C}(I),\, |h| \leq f\} \quad (f \in \mathcal{C}^+(I)).$$

Then, clearly $\Lambda f \geq 0$ for any $f \in \mathcal{C}^+(I)$. Moreover,

$$|\Phi(f)| \leq \sup\{|\Phi(h)|;\, h \in \mathcal{C}(I),\, |h| \leq |f|\} = \Lambda(|f|),$$

and

$$\Lambda(|f|) \leq \sup\{\|\Phi\| \, \|h\|_\infty;\, h \in \mathcal{C}(I),\, |h| \leq |f|\} \leq \|f\|_\infty.$$

Hence, Λ satisfies (2.2.22) and it just remains to see the linearity.

Observe that for any $c \in \mathbb{R}_{>0}$ and $f \in \mathcal{C}^+(I)$,

$$\begin{aligned}\Lambda(cf) &= \sup\{|\Phi(h)|;\, h \in \mathcal{C}(I),\, |h| \leq cf\} \\ &= \sup\{|\Phi(ch)|;\, ch \in \mathcal{C}(I),\, |ch| \leq cf\} \\ &= c\Lambda f\end{aligned}$$

(if $c = 0$, then $\Lambda(0) = 0$ since Φ is linear).

Hence, to prove the linearity we just have to show that

$$\Lambda(f + g) = \Lambda f + \Lambda g, \quad f \text{ and } g \in \mathcal{C}^+(I). \tag{2.2.23}$$

To do so, observe that $0 \leq f_1 \leq f_2$ implies $\Lambda f_1 \leq \Lambda f_2$. Then, fix f and g in $\mathcal{C}^+(I)$. By the definition of supremum, if $\varepsilon > 0$, there exists $|h_1|$ and $|h_2|$ in $\mathcal{C}(I)$ such that $|h_1| \leq f$, $|h_2| \leq g$, and

$$\Lambda f \leq |\Phi(h_1)| + \varepsilon, \quad \Lambda g \leq |\Phi(h_2)| + \varepsilon.$$

Moreover, there are complex numbers α_i, $|\alpha_i| = 1$, so that $\alpha_i \Phi(h_i) = |\Phi(h_i)|$, $(i = 1, 2)$. Then, since $|\alpha_1 h_1 + \alpha_2 h_2| \leq |h_1| + |h_2| \leq f + g$,

$$\Lambda f + \Lambda g \leq |\Phi(h_1)| + \varepsilon + |\Phi(h_2)| + \varepsilon = \Phi(\alpha_1 h_1 + \alpha_2 h_2) + 2\varepsilon \leq$$

$$\leq \Lambda(|h_1| + |h_2|) + 2\varepsilon \leq \Lambda(f + g) + 2\varepsilon,$$

and this inequality holds for any ε. Therefore, $\Lambda f + \Lambda g \leq \Lambda(f + g)$.

For the other inequality, choose $h \in \mathcal{C}(I)$ subject only to the condition $|h| \leq f + g$, and let $V = \{x \colon f(x) + g(x) > 0\}$ (which is an open set in I since f and g are continuous functions). We define

$$h_1(x) := \frac{f(x)h(x)}{f(x) + g(x)}, \quad h_2(x) := \frac{g(x)h(x)}{f(x) + g(x)} \quad (\text{if } x \in V),$$

and $h_1(x) = h_2(x) = 0$ if $x \notin V$. It is clear that h_1 is continuous at every point $x_0 \in V$ (since V is open). If $x_0 \notin V$, then $h(x_0) = 0$ and due to h is continuous and $|h_1(x)| \leq |h(x)|$ for all point $x \in I$, it follow that x_0 is a point of continuity of h_1 (for any $\varepsilon > 0$, $\exists \delta > 0$ such that if $|x - x_0| < \delta$, then $|h_1(x)| \leq |h(x)| < \varepsilon$). Thus, $h_1 \in \mathcal{C}(I)$, and the same holds for h_2.

Since $h = h_1 + h_2$ and $|h_1| \leq f$, $|h_2| \leq g$ (because we are supposing $|h| \leq f + g$), we have

$$|\Phi(h)| = |\Phi(h_1) + \Phi(h_2)| \leq |\Phi(h_1)| + |\Phi(h_2)| \leq \Lambda f + \Lambda g.$$

Finally, taking the supremum over all $h \in \mathcal{C}(I)$ such that $|h| \leq f + g$, we get (2.2.23). Thus, Λ is linear.

If f is now a real function, $f \in \mathcal{C}(I)$, then $2f^+ = |f| + f$ so that $f^+ \in \mathcal{C}^+(I)$ and $2f^- = |f| - f$ so that $f^- \in \mathcal{C}^+(I)$. Since $f = f^+ - f^-$, it is natural to define

$$\Lambda f = \Lambda f^+ - \Lambda f^- \quad (f \in \mathcal{C}(I),\ f \text{ real}).$$

Moreover, we can define

$$\Lambda(u + iv) = \Lambda u + i\Lambda v.$$

By simple algebraic manipulations, it is easy to show that our extended function Λ is linear on $\mathcal{C}(I)$.

Then, by the Theorem 2.2.5, we can associate that linear operator with a regular positive Borel measure as

$$\Lambda f = \int_I f d\lambda,$$

such that

$$\lambda(I) = \|\Lambda\|.$$

Since $|\Lambda f| \leq 1$ if $\|f\|_\infty \leq 1$, we see that actually $\lambda(I) \leq 1$. Using (2.2.22), we also deduce that

$$|\Phi(f)| \leq \Lambda(|f|) = \int_I |f| d\lambda = \|f\|_1 \tag{2.2.24}$$

for any $f \in \mathcal{C}(I)$. Thus, Φ is a linear functional on $\mathcal{C}(I)$ of norm at most 1 with respect to the $L^1(\lambda)$-norm on $\mathcal{C}(I)$. Then, since the set $\mathcal{C}(I)$ is dense in $L^1(\lambda)$, we can define an operator $\tilde{\Phi}$

on $L^1(\lambda)$ by taking $\tilde{\Phi}(f) = \lim_n \Phi(f_n)$, for any $f \in L^1(\lambda)$ and $\{f_n\}_n \subset \mathcal{C}(I)$ such that f_n tends to f in $L^1(\lambda)$ as n goes to $+\infty$. Observe that if $\{g_n\}_n \subset \mathcal{C}(I)$ is another sequence that tends to f in $L^1(\lambda)$, then by the linearity of Φ,

$$|\Phi(f_n) - \Phi(g_n)| = |\Phi(f_n - g_n)| \leq \int_I |f_n - g_n| d\lambda = \|f_n - g_n\|_{L^1(\lambda)} .$$

Moreover,

$$\|f_n - g_n\|_{L^1(\lambda)} \leq \|f_n - f\|_{L^1(\lambda)} + \|f - g_n\|_{L^1(\lambda)} \to 0 \quad (n \to +\infty),$$

so $\lim_n \Phi(f_n) = \lim_n \Phi(g_n)$, and the operator $\tilde{\Phi}$ is well defined. Moreover, is linear by the linearity of Φ, and by (2.2.24) is bounded.

Hence, there is a norm-preserving extension of Φ to a linear functional $\tilde{\Phi}$ on $L^1(\lambda)$, and therefore Corollary 2.2.36 gives a measurable Borel function g, with $|g| \leq 1$, such that

$$\tilde{\Phi}(f) = \int_I fg d\lambda \quad (f \in L^1(\lambda)),$$

and then

$$\Phi(f) = \int_I fg d\lambda \quad (f \in \mathcal{C}(I)). \tag{2.2.25}$$

By hypothesis, Φ is a continuous functional, and also we have that

$$\left| \int_I fg d\lambda \right| \leq \|f\|_\infty \qquad (f \in \mathcal{C}(I)).$$

So, each side of (2.2.25) is a continuous functional on $\mathcal{C}(I)$. Hence, we obtain the representation that we want with $d\mu = g d\lambda$.

Observe that (2.2.25) shows that for $f \in \mathcal{C}(I)$ with $\|f\|_\infty \leq 1$,

$$|\Phi(f)| \leq \int_I |fg| d\lambda \leq \int_I |g| d\lambda.$$

Since $\|\Phi\| = 1$,

$$\int_I |g| d\lambda \geq \sup \{|\Phi(f)|\colon\ f \in \mathcal{C}(I),\ \|f\|_\infty \leq 1\} = 1.$$

Due to $|g| \leq 1$, one can see that $\lambda(I) \geq 1$, so we get that $\lambda(I) = 1$. Moreover,

$$0 = \int_I (1-1) d\lambda \leq \int_I (1 - |g|) d\lambda = \lambda(I) - \int_I |g| d\lambda \leq 1 - 1 = 0.$$

Then, using Lemma 2.2.21 (a), we get that $|g| = 1$ a.e. $[\lambda]$. Thus, by Corollary 2.2.31, $d|\mu| = |g| d\lambda = d\lambda$, and

$$\|\mu\| = |\mu|(I) = \lambda(I) = 1 = \|\Phi\| ,$$

which proves the theorem.

□

Remark 2.2.38. In its original form by F. Riesz [20], the theorem states that every continuous linear functional Φ over the space $\mathcal{C}([0,1])$ of continuous functions in the interval $[0,1]$ can be represented in the form

$$\Phi(f) = \int_0^1 f(x)d\mu(x), \qquad (f \in \mathcal{C}([0,1]))$$

where μ is a function of bounded variation (i.e., such that $|\mu|([0,1]) < +\infty$), and the integral is a Riemann-Stieltjes integral. See [19], for a historical discussion.

2.2.6 Classical Results in Functional and Complex Analysis

Finally, we show some classical results in functional and complex analysis. We first begin by seeing one of the most important theorems on functional analysis, the Hahn-Banach Theorem. The point of this result will be a corollary which will play an important role on the proof of the Müntz-Szász Theorem. We will continue by presenting some results in complex analysis but without proving them, since these results have been seen either on the bachelor's degree or during the master course.

Definition 2.2.39. Let E be a vectorial space over a field K. We say that the function $p : E \to \mathbb{R}$ is a convex functional if $p(x+y) \leq p(x)+p(y)$ and $p(\alpha x) = \alpha p(x)$ for $\alpha > 0$, $\alpha \in K$, and $x, y \in E$.

Theorem 2.2.40 (Hahn-Banach Theorem)**.** *Let p be a convex functional over the normed vectorial space X, and u be a linear functional over a subspace M of X. If $u(z) \leq p(z)$ for every $z \in M$, then u can be extended to a linear functional v over X such that $v(z) \leq p(z)$ for every $z \in X$.*

Corollary 2.2.41. *Let M be a vectorial subspace of a normed vectorial space X, and let $z_0 \in X$. Then, $z_0 \in \overline{M}$ if and only if there is not any linear bounded functional T over X such that $T(z) = 0$ for every $z \in M$, but $T(z_0) \neq 0$.*

Proof. First suppose that $z_0 \in \overline{M}$, and let T be a linear bounded functional on X such that $T(z) = 0$ for every $z \in M$. Then, given $\{z_n\}_n \subset M$ a convergent sequence to z_0, by the continuity of T,

$$T(z_0) = T\left(\lim_{n\to\infty} z_n\right) = \lim_{n\to\infty} T(z_n) = 0.$$

Conversely, suppose that $z_0 \notin \overline{M}$. Then, $\exists\delta > 0$ such that $\|z_0 - z\|_X > \delta$ for all $z \in M$. So, we define over the vectorial subspace $M' = \langle M, x_0\rangle \subset X$ the functional $T(z + \lambda z_0) = \lambda$, where $z \in M$ and λ is an scalar. Since M is a vectorial space, if $z \in M$ and $\lambda \neq 0$, then $-\lambda^{-1}z \in M$, and therefore,

$$\delta|\lambda| \leq |\lambda| \left\|z_0 + \lambda^{-1}z\right\|_X = \|\lambda z_0 + z\|_X = \|z + \lambda z_0\|_X .$$

Then, we see that T is a linear functional bounded by δ^{-1}, since

$$\|T\| = \sup_{\|z+\lambda z_0\|_X=1} |T(z + \lambda z_0)| = \sup_{\|z+\lambda z_0\|_X=1} |\lambda| \leq \sup_{\|z+\lambda z_0\|_X=1} \frac{\|z + \lambda z_0\|_X}{\delta} = \delta^{-1}.$$

Moreover, $T(z) = 0$ over M and $T(z_0) = 1$. Since T is bounded (then is continuous), exists $C > 0$ such that $|T(z + \lambda z_0)| \leq C \|z + \lambda z_0\|_X$, and $C \|\cdot\|_X$ is a convex functional. Hence, by the Theorem 2.2.40, we can extend T from M' to X.

□

Now, let's see some results about holomorphic functions. We will denote by $\mathcal{H}(\Omega)$ the set consisting on holomorphic functions in an open set $\Omega \subset \mathbb{C}$, and by $H^\infty := H^\infty(\mathbb{D})$ the set of all bounded holomorphic functions in the unit disk.

The first result that we present is about the zero set of functions in the space H^∞. For simplicity on the statement, we introduce first the concept of Blaschke condition.

Definition 2.2.42. Given a sequence $\{a_j\}_j \subset \mathbb{D}$ of complex numbers. We say that $\{a_j\}_j$ satisfies the Blaschke condition if

$$\sum_j (1 - |a_j|) = \sum_j d(a_j, \partial\mathbb{D}) < +\infty.$$

In this context, we have the following result about the zero set of some holomorphic functions.

Theorem 2.2.43. *If $f \in H^\infty$ has the zero set $Z(f) = \{a_j\}_j$ in $\mathbb{D}$ and if $Z(f)$ does not satisfy the Blaschke condition, then $f(z) = 0$ for any $z \in \mathbb{D}$.*

The previous theorem is a particular case of Theorem 15.23 in [1]. The following result that we are introducing, now for the set $\mathcal{H}(\Omega)$ (where Ω is a domain), gives a condition for which an infinite product of holomorphic functions converges to an holomorphic function. For the proof, see Theorem 15.6 of [1].

Theorem 2.2.44. *Let $\Omega \subset \mathbb{C}$ be a domain. Assume that $f_n \in \mathcal{H}(\Omega)$ for every $n \in \mathbb{N}$ such that*

$$\sum_{n \geq 1} |1 - f_n(z)|$$

converges uniformly over compact sets in Ω. Then, the product

$$f := \prod_{n \geq 1} f_n(z)$$

converges uniformly over compact sets in Ω. In particular, $f \in \mathcal{H}(\Omega)$.

Now we state the well-known theorem *Arzelá-Ascoli Theorem*. For more details see [2].

Theorem 2.2.45 (The Arzelá-Ascoli Theorem)**.** *Let $K \subset \mathbb{C}$ be compact and let $(f_n)_{n\in\mathbb{N}}$ be a sequence of continuous functions defined on K. If $\mathcal{F} = \{f_n;\, n \in \mathbb{N}\}$ is uniformly bounded and equicontinuous on K, then there exists a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ that converges uniformly to a function $f \in \mathcal{C}(K)$.*

The following result is a basic theorem on Complex Analysis and says that an holomorphic function in a domain Ω is determined by its values in any set that contains an accumulation point of Ω (see [1], page 211).

Theorem 2.2.46 (Identity Principle)**.** *If f and g are two holomorphic functions in a complex domain Ω and if $f(z) = g(z)$ for every z in some subset with an accumulation point in Ω, then $f(z) = g(z)$ for every $z \in \Omega$.*

On the proof of the Müntz-Szász Theorem we will need to see what functions are holomorphic. The following classic results of the complex analysis will be very helpful for this purpose.

Theorem 2.2.47 (Morera's Theorem)**.** *Given a continuous function f defined in a complex domain Ω. If*

$$\oint_C f(z)\, dz,$$

for every closed path C and piece-wise $\mathcal{C}^1$ with compact support in Ω, then f is holomorphic in Ω.

Theorem 2.2.48 (Cauchy's Theorem)**.** *If f is an holomorphic function in a simple connected domain Ω, then*

$$\oint_C f(z)\, dz,$$

for every rectifiable closed path C in Ω.

Once checked the analicity of a function, the Cauchy's representation formula will allow us to rewrite our function in a specific way.

Theorem 2.2.49 (Cauchy's Formula)**.** *If f is an holomorphic function inside and on the boundary C of a simple connected domain Ω, then for every z_0 in Ω,*

$$f(z_0) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z - z_0}\, dz.$$

Finally we introduce the Möbius transformation, an holomorphic function that will play an important role on proving the density of the Müntz-Szász Theorem.

Definition 2.2.50. In complex analysis, we can define the Möbius transformation as a rational holomorphic function from the complex unit disc onto the right complex half-plane

$$\begin{aligned} \mathbb{D} &\to \mathbb{H}_0 := \{z \in \mathbb{C} \colon \operatorname{Re} z > 0\}, \\ z &\mapsto \frac{1+z}{1-z}. \end{aligned} \tag{2.2.26}$$

Observe that this function is well defined, since

$$\operatorname{Re}\left(\frac{1+z}{1-z}\right) = \operatorname{Re}\left(\frac{(1+z)(1-\overline{z})}{|1-z|^2}\right) = \frac{1-|z|^2}{|1-z|^2} > 0, \quad \text{for every } z \in \mathbb{D}.$$

The inverse Möbius transformation of (2.2.26) is

$$\begin{aligned} \mathbb{H}_0 &\to \mathbb{D}, \\ w &\mapsto \frac{w-1}{w+1}. \end{aligned}$$

For our interest, we will also consider the definition of inverse Möbius transformation

$$\begin{aligned} &\mathbb{H}_{-1} \to \mathbb{D}, \\ &w \mapsto \frac{a-1-w}{a+1+w} = -\frac{\frac{w+1}{a}-1}{\frac{w+1}{a}+1}, \end{aligned} \tag{2.2.27}$$

for an arbitrary $a > 0$, and which is the inverse of the function

$$z \mapsto a\left(\frac{1-z}{1+z}\right) - 1$$

(hence, $\mathrm{Re}\left(a\left(\frac{1-z}{1+z}\right)-1\right) = a\mathrm{Re}\left(\frac{1-z}{1+z}\right) - 1 = \frac{a}{|1+z|^2}\mathrm{Re}\left[(1-z)(1+\overline{z})\right] - 1 > -1$).

2.3 Density on Müntz-Szász Theorem on $\mathcal{C}([0,1])$

With all this previous results in functional and complex analysis, we are in conditions to see an interesting extension of the Müntz-Szász Theorem on $\mathcal{C}([0,1])$ which allows us to use a sequence $\{\lambda_j\}_{j=0}^{+\infty}$ ($\lambda_0 = 0$) of distinct real positive numbers without any more restriction. Consequently, instead of working with the series $\sum_{j=1}^{+\infty} 1/\lambda_j$, we will have to deal with the series $\sum_{j=1}^{+\infty} \lambda_j/(\lambda_j^2+1)$, since the sequence $\{\lambda_j\}_{j=1}^{+\infty}$ may has a subsequence that converges to zero.

However, we think that it could be interesting and clarifying to see first a proof for the original theorem. So before going ahead, the first part on this chapter is focused on the proof of the Müntz-Szász Classical Theorem.

2.3.1 Müntz-Szász Classical Theorem

We will show a constructive proof of the Müntz-Szász Classical Theorem given by M. Von Golitschek [18], who gives a simple argument to show that $\langle 1, x^{\lambda_1}, x^{\lambda_2}, \dots \rangle$ is dense in $\mathcal{C}([0,1])$ when $\{\lambda_j\}_{j=0}^{+\infty}$ is an increasing sequence of distinct positive real numbers tending to infinity.

Theorem 2.3.1 (Müntz-Szász Classical Theorem)**.** *Let $0 < \lambda_1 < \lambda_2 < \lambda_3 < \dots$ such that $\lim_n \lambda_n = +\infty$. If $\sum_n 1/\lambda_n = +\infty$ then the set $\langle 1, x^{\lambda_1}, x^{\lambda_2}, \dots \rangle$ is dense in $\mathcal{C}([0,1])$.*

Proof. Assume that $m \neq \lambda_k$ for every $k = 1, 2, 3, \dots$, and $m \in \mathbb{Z}_+$. Let's define the functions inductively:

$$\begin{aligned} &Q_0(x) := x^m, \\ &Q_n(x) := (\lambda_n - m)\, x^{\lambda_n} \int_x^1 Q_{n-1}(t) t^{-1-\lambda_n}\, dt, \qquad n = 1, 2, 3, \dots. \end{aligned}$$

We first claim that for $n \geq 1$

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

where $a_{i,n} \in \mathbb{R}$ for every $i = 1, \dots, n$.

Let's see it by induction. For $n=1$,

$$Q_1(x) = (\lambda_1 - m)x^{\lambda_1}\int_x^1 t^{m-1-\lambda_1}dt = (\lambda_1 - m)x_1^{\lambda}\left[\frac{1}{m-\lambda_1}t^{m-\lambda_1}\right]_x^1 = x^m - x^{\lambda_1}.$$

Hence, assuming that $Q_{n-1}(x) = x^m - \sum_{i=1}^{n-1} a_{i,n-1}x^{\lambda_i}$, we have

$$\begin{aligned}
Q_n(x) &= (\lambda_n - m)x^{\lambda_n}\int_x^1 \left(t^{m-1-\lambda_n} - \sum_{i=1}^{n-1} a_{i,n-1}t^{\lambda_i-1-\lambda_n}\right)dt\\
&= x^m - x^{\lambda_n} + (\lambda_n - m)x^{\lambda_n}\left(-\sum_{i=1}^{n-1} a_{i,n-1}\left[\frac{1}{\lambda_i-\lambda_n}t^{\lambda_i-\lambda_n}\right]_x^1\right)\\
&= x^m - x^{\lambda_n} + (\lambda_n - m)x^{\lambda_n}\left(-\sum_{i=1}^{n-1} a_{i,n-1}\left(\frac{1}{\lambda_i-\lambda_n} - \frac{1}{\lambda_i-\lambda_n}x^{\lambda_i-\lambda_n}\right)\right)\\
&= x^m - \sum_{i=1}^{n-1}\frac{\lambda_n - m}{\lambda_n-\lambda_i}a_{i,n-1}x^{\lambda_i} - \left(1 + \sum_{i=1}^{n-1}\frac{\lambda_n - m}{\lambda_i-\lambda_n}a_{i,n-1}\right)x^{\lambda_n}.
\end{aligned}$$

So denoting by $a_{i,n} = \frac{\lambda_n - m}{\lambda_n - \lambda_i}a_{i,n-1}$ for $i = 1,\dots,n-1$ and $a_{n,n} = 1 - \sum_{i=1}^{n-1} a_{i,n}$, we get

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

Now, observe that

$$\|Q_0\|_\infty = \sup_{x\in[0,1]} |x^m| = 1.$$

Moreover,

$$\begin{aligned}
\|Q_n\|_\infty &= \sup_{x\in[0,1]}\left|(\lambda_n - m)\,x^{\lambda_n}\int_x^1 Q_{n-1}(t)t^{-1-\lambda_n}\,dt\right|\\
&\leq |\lambda_n - m|\,\|Q_{n-1}\|_\infty \sup_{x\in[0,1]} x^{\lambda_n}\int_x^1 t^{-1-\lambda_n}\,dt\\
&= |\lambda_n - m|\,\|Q_{n-1}\|_\infty \sup_{x\in[0,1]} x^{\lambda_n}\left(\frac{-1+x^{-\lambda_n}}{\lambda_n}\right) = \left|1 - \frac{m}{\lambda_n}\right|\|Q_{n-1}\|_\infty.
\end{aligned}$$

Hence, by iteration we have that

$$\|Q_n\|_\infty \leq \prod_{i=1}^{n}\left|1 - \frac{m}{\lambda_i}\right|.$$

Finally, since $\lim_n \lambda_n = +\infty$, then exists an integer $N > 0$ such that $\lambda_n > m$ for every $n \geq N$ and using the inequality $1 - x \leq e^{-x}$ for $x \geq 0$, we have that for every $n \geq N$

$$\begin{aligned}
\|Q_n\|_\infty &\leq \prod_{i=1}^{n}\left|1 - \frac{m}{\lambda_i}\right| = \prod_{i=1}^{N-1}\left|1 - \frac{m}{\lambda_i}\right|\prod_{i=N}^{n}\left(1 - \frac{m}{\lambda_i}\right)\\
&\leq \prod_{i=1}^{N-1}\left|1 - \frac{m}{\lambda_i}\right|\exp\left(-\sum_{i=N}^{n}\frac{m}{\lambda_i}\right) \to 0 \quad (n \to +\infty),
\end{aligned}$$

due to $\sum_n 1/\lambda_n = +\infty$.

Thus, $\|Q_n\|_\infty \to 0$ as $n \to +\infty$. Hence, $x^m - Q_n(x)$ converges uniformly to x^m.

Finally, the proof follows by the Weierstrass Approximation Theorem (Corollary 2.1.7).

□

Observation 2.3.2. Let $n \in \mathbb{N}$ and take $i \in \{1, \dots, n\}$. Let's denote by

$$L_{i,n}(m) = \prod_{\substack{j=1 \\ j \neq i}}^{n} \frac{\lambda_j - m}{\lambda_j - \lambda_i}$$

the Legendre polynomial of degree $n-1$ evaluated at $m \in \mathbb{N}$. Then one can see that

$$Q_n(x) = -\sum_{i=1}^{n} L_{i,n}(m) x^{\lambda_i} + x^m.$$

2.3.2 Density on Full Müntz-Szász Theorem on $\mathcal{C}([0,1])$

Finally, we present an extension of the density part for the Müntz-Szász Classical Theorem on $\mathcal{C}([0,1])$, which involves arbitrary sequences $\{\lambda_j\}_{j=0}^{\infty}$ ($\lambda_0 = 0$) of distinct nonnegative real numbers.

Before showing this Müntz-Szász Theorem, we see that for the density case, we will be able to split our sequence in three different subsequences, and then work with them separately instead of doing it with the original sequence.

Lemma 2.3.3. *Let $\{\lambda_j\}_{j=1}^{+\infty}$ be a sequence of distinct real positive numbers such that*

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

and let $\gamma > 0$ be a real number. Then, there is a subsequence $\{\lambda_j\}_{k=1}^{+\infty}$ such that

$$\sum_{k=1}^{+\infty} \frac{\lambda_{j_k}}{\lambda_{j_k}^2 + 1} = +\infty$$

and it belongs in one of these three cases:

(i) ***Case 1****: $\lambda_{j_k} \geq \gamma$ for each $k = 1, 2, \dots$.*

(ii) ***Case 2:*** *$0 < \lambda_{j_k} < \gamma$ for each $k = 1, 2, \dots$ and $\lim_{j_k} \lambda_{j_k} = \alpha > 0$.*

(iii) ***Case 3:*** *$0 < \lambda_{j_k} < \gamma$ for each $k = 1, 2, \dots$ and $\lim_{j_k} \lambda_{j_k} = 0$.*

Proof. Let $J = \{j \in \mathbb{N};\, 0 < \lambda_j < \gamma\}$. Since the terms of the series are all positive, it follows that

$$+\infty = \sum_{j=1}^{+\infty} \frac{\lambda_j}{\lambda_j^2 + 1} = \sum_{j \notin J} \frac{\lambda_j}{\lambda_j^2 + 1} + \sum_{j \in J} \frac{\lambda_j}{\lambda_j^2 + 1}.$$

Hence, we have two possibilities: either

$$\sum_{j\notin J} \frac{\lambda_j}{\lambda_j^2+1} = +\infty$$

and therefore **Case 1** holds by taking $\{\lambda_j\}_{j\notin J}$, or either

$$\sum_{j\in J} \frac{\lambda_j}{\lambda_j^2+1} = +\infty. \tag{2.3.1}$$

Then, suppose that it occurs (2.3.1). Observe that in J, all the λ_j are bounded by γ. Since every bounded sequence has a convergent subsequence, then there exists $\{\lambda_{j_k}\}_{k=1}^{+\infty} \subseteq \{\lambda_j\}_{j\in J}$ such that $\lim_k \lambda_{j_k} = \alpha \in [0,\gamma]$.

Assume that $\alpha > 0$. Then, since $\lambda_{j_k} \to \alpha$ as $k \to +\infty$, it follows that

$$\lim_{k\to+\infty} \frac{\lambda_{j_k}}{\lambda_{j_k}^2+1} = \frac{\alpha}{\alpha^2+1} \neq 0,$$

and therefore,

$$\sum_{k=1}^{+\infty} \frac{\lambda_{j_k}}{\lambda_{j_k}^2+1} = +\infty.$$

Hence, **Case 2** holds.

Now we claim that if there is not any convergent subsequence $\{\lambda_{j_k}\}_{k=1}^{+\infty}$ of $\{\lambda_j\}_{j\in J}$ such that $\lim_k \lambda_{j_k} = \alpha > 0$, then the sequence $\{\lambda_j\}_{j\in J}$ converges to zero.

Suppose the contrary. Then there exists an $\varepsilon > 0$ such that for every $N := N(\varepsilon) \in \mathbb{N}$, there is a $j_1 > N$ so that $j_1 \in J$ and $\lambda_{j_1} > \varepsilon$. Then, replacing j_1 instead of N, there exists $j_2 > j_1$, such that $j_2 \in J$ and $\lambda_{j_2} > \varepsilon$.

Iterating, we get a sequence $(j_k)_{k\in\mathbb{N}} \subseteq J$ such that $j_1 < j_2 < j_3 < \cdots$ and $\lambda_{j_k} > \varepsilon$ for every $k \in \mathbb{N}$. Hence, we have a subsequence $\{\lambda_{j_k}\}_{k\in\mathbb{N}}$ of $\{\lambda_j\}_{j\in J}$ such that is bounded in $(\varepsilon,\gamma]$. Therefore, this subsequence must have a subsequence that converges to some $\alpha \in [\varepsilon,\gamma]$. In particular, $\alpha > 0$. But this is a contradiction, due to this subsequence is also a subsequence of $\{\lambda_j\}_{j\in J}$.

Hence, the sequence $\{\lambda_j\}_{j\in J}$ converges to zero and by assumption

$$\sum_{j\in J} \frac{\lambda_j}{\lambda_j^2+1} = +\infty,$$

so **Case 3** holds.

□

Here we present a technical Lemma from which the extension of the Müntz-Szász Theorem will follow directly. Our proof will be split up in the three cases that we have already seen in Lemma 2.3.3. For the first and the second cases we have followed the proof of [5]. However, there are equivalent proofs using similar arguments in [1] and [18]. For the third case, we have argue similarly as in [6] but reducing us to the real valued continuous functions $\mathcal{C}([0,1])$, since it such article they work in the Lebesgue spaces $L^p([0,1])$ for $1 \leq p < +\infty$. We want to point that is here where Section 2.2.6 takes place.

Lemma 2.3.4. *Let* $\{\lambda_j\}_{j=0}^{+\infty}$ $(\lambda_0 = 0)$ *be a sequence of nonnegative real numbers, and let* $\lambda_0 = 0$. *If*

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

and μ *is a Borel complex measure in* $[0,1]$ *such that*

$$\int_0^1 t^{\lambda_j} d\mu(t) = 0, \quad \forall j = 0, 1, 2, \dots, \tag{2.3.2}$$

then

$$\int_0^1 t^k d\mu(t) = 0, \quad \forall k = 0, 1, 2, \dots. \tag{2.3.3}$$

Proof. Observe that the integrands in (2.3.2) and (2.3.3) vanishes at $t = 0$. Hence, we can assume that the measure μ is concentrated in $I = (0, 1]$.

Then, consider the function

$$f(z) = \int_I t^z d\mu(t) = \int_I e^{z \log t} d\mu(t),$$

which is well defined in the right complex half-plane $\mathbb{H}_0$, since if $\mathrm{Re}(z) > 0$ then $|t^{\mathrm{Re}(z)}| \leq 1$ for $t \in I$ and, therefore,

$$\begin{aligned} |f(z)| &\leq \int_I |e^{z \log t}|\, d|\mu|(t) = \int_I e^{\mathrm{Re}(z) \log t} d|\mu|(t) \\ &= \int_I t^{\mathrm{Re}(z)} d|\mu|(t) \leq |\mu|(I) = \|\mu\| < +\infty. \end{aligned} \tag{2.3.4}$$

Now, let's see that f is holomorphic in $\mathbb{H}_0$. To do so, we will see first that f is continuous. Due to the difference

$$f(z) - f(z_0) = \int_I t^z d\mu(t) - \int_I t^{z_0} d\mu(t) = \int_I (t^z - t^{z_0}) d\mu(t),$$

then

$$|f(z) - f(z_0)| \leq \int_I |t^z - t^{z_0}|\, d|\mu|(t).$$

Fix $\varepsilon > 0$. Since $(t, z) \mapsto t^z$ is a continuous function in $I \times \mathbb{H}_0$ (uniformly in t, because I is compact), exists $\delta := \delta(\varepsilon) > 0$ such that, if $|z - z_0| < \delta$, then $|t^z - t^{z_0}| < \varepsilon$, for every $t \in I$. Hence,

$$|f(z) - f(z_0)| \leq \varepsilon \int_I d|\mu|(t) = \varepsilon \|\mu\|,$$

which proves the continuity of f.

Now, let γ be a closed piece-wise $\mathcal{C}^1$ path in $\mathbb{H}_0$. Then,

$$\oint_\gamma f(z) dz = \oint_\gamma \int_I t^z d\mu(t) dz. \tag{2.3.5}$$

Observe that

$$\left|\oint_\gamma f(z)dz\right| \leq \oint_\gamma \int_I |t^z|\, d|\mu|(t)d|z| = \oint_\gamma \int_I t^{\mathrm{Re}(z)} d|\mu|(t)d|z|$$
$$\leq \oint_\gamma \int_I d|\mu|(t)d|z| = \|\mu\|\, L(\gamma) < +\infty,$$

where $L(\gamma)$ denotes the length of the curve γ. Hence, we can apply the Fubini's Theorem to (2.3.5), and we get

$$\oint_\gamma f(z)dz = \int_I \oint_\gamma t^z dz d\mu(t) = 0,$$

where the last inequality follows from the fact that t^z is an holomorphic function, which allows us to apply the Cauchy Theorem (Theorem 2.2.48). Therefore, by the Morera's Theorem (Theorem 2.2.47) we conclude that f is holomorphic in $\mathbb{H}_0$. Moreover, we have proved on (2.3.4) that f is bounded.

Now, we will see that f vanishes in all $\mathbb{H}_0$. Without loss of generality, taking a subsequence if necessary, we will suppose that $\{\lambda_j\}_{j\in\mathbb{N}}$ is in one of the three cases of Lemma 2.3.3 when $\gamma = 1$.

Observe that in **Case 1** and in **Case 2** happen that $\inf_{j\in\mathbb{N}} \lambda_j > 0$. So that, in this two cases we will consider the function

$$g(z) = f\left(\frac{1+z}{1-z}\right) \qquad z \in \mathbb{D}.$$

Observe that g is the composition of a Möbius transformation from the disk to the right half-plane (see (2.2.26)) and our function f. Hence,

- $g \in \mathcal{H}(\mathbb{D})$,
- g is bounded on $\mathbb{D}$ (since f is bounded).

This means that $g \in H^\infty(\mathbb{D})$. Moreover, from (2.3.2), we have that

$$f(\lambda_j) = T(t^{\lambda_j}) = 0,\ \forall j = 1, 2, \ldots,$$

so, $g(\alpha_j) = 0$, where $\alpha_j = \frac{\lambda_j - 1}{\lambda_j + 1}$.

Now we claim that

$$\sum_{j\geq 1} \frac{\lambda_j}{\lambda_j^2 + 1} = +\infty \Rightarrow \sum_{j\geq 1} (1 - |\alpha_j|) = +\infty.$$

Indeed,

$$\sum_{j\geq 1} (1 - |\alpha_j|) = \sum_{j\geq 1} 1 - \left|\frac{\lambda_j - 1}{\lambda_j + 1}\right| = \sum_{j\geq 1} \frac{\lambda_j + 1 - |\lambda_j - 1|}{\lambda_j + 1}.$$

Hence, depending on if we are in **Case 1** or in **Case 2**, we have two different possibilities:

- Case 1: $0 < \lambda_j < 1$ for every $j \in \mathbb{N}$ and $\lim_j \lambda_j = \beta > 0$: In this case,

$$\lambda_j + 1 - |\lambda_j - 1| = 2\lambda_j$$

for every $j \in \mathbb{N}$, so

$$\sum_{j\geq 1}(1-|\alpha_j|) \geq \sum_{j\geq 1}\frac{2\lambda_j}{\lambda_j+1} = +\infty$$

since $\frac{2\lambda_j}{\lambda_j+1} \nrightarrow 0$, when $j \to +\infty$.

- Case 2: $\lambda_j \geq 1$ for every $j \in \mathbb{N}$: In this case,

$$\lambda_j + 1 - |\lambda_j - 1| = 2.$$

Thus,

$$\sum_{j\geq 1}(1-|\alpha_j|) = \sum_{j\geq 1}\frac{2}{\lambda_j+1} = +\infty,$$

since when $\inf_{j\in\mathbb{N}} \lambda_j > 0$, the series

$$\sum_{j\geq 1}\frac{1}{\lambda_j} \quad \text{and} \quad \sum_{j\geq 1}\frac{\lambda_j}{\lambda_j^2+1}$$

are equivalent.

Therefore, applying Theorem 2.2.43, we deduce that $g(z) = 0$ for every $z \in \mathbb{D}$. In particular, $f \equiv 0$ in $\mathbb{H}_0$.

Now suppose that we are in **Case 3**, that is $0 < \lambda_j < 1$ for every $j \in \mathbb{N}$ and $\lim_j \lambda_j = 0$. Let's consider

$$g(z) := f\left(z+1\right), \quad (z \in \mathbb{D}).$$

Hence, g is holomorphic in the unit disk and bounded (f is bounded), so $g \in H^\infty(\mathbb{D})$. Moreover, since f vanishes in λ_j for every $j = 1, 2, \ldots$, then g vanishes in $\lambda_j - 1 \in \mathbb{D}$.

Now, since $0 < \lambda_j < 1$ for every j,

$$\sum_{j=1}^{+\infty} 1 - |\lambda_j - 1| = \sum_{j=1}^{+\infty}\lambda_j \geq \sum_{j=1}^{+\infty}\frac{\lambda_j}{\lambda_j^2+1} = +\infty.$$

Therefore, Theorem 2.2.43 yields that $g = 0$ on the open disk. Therefore, $f(z) = 0$ on the open disk with diameter $[0, 2]$. Now observe that f is analytic on $\mathbb{H}_0$; hence, by the Identity Principle (Theorem 2.2.46) $f \equiv 0$ whenever $\mathrm{Re}(z) > 0$.

Thus, in all the cases, $f \equiv 0$ in $\mathbb{H}_0$. In particular,

$$T(t^k) = \int_I t^k\, d\mu(t) = f(k) = 0, \quad k = 0, 1, 2, \ldots.$$

This concludes the proof of the Lemma.

□

As a consequence of Lemma 2.3.4, we have the following Müntz-Szász Theorem extension in the dense case:

Theorem 2.3.5 (Full Müntz-Szász Theorem)**.** *Let* $\{\lambda_j\}_{j=1}^{\infty}$ *be a sequence of different positive real numbers and* X *the closure in* $\mathcal{C}([0,1])$ *of the set generated by the finite linear combinations of the functions* $1, x^{\lambda_1}, x^{\lambda_2}, \dots$ *. If*

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

then $X = \mathcal{C}([0,1])$*.*

Proof. By the Weierstrass Approximation Theorem (Corollary 2.1.7), it is enough to see that every function x^k, with $k \in \mathbb{N}$, belongs to X. Suppose, on the contrary, that exists $k_0 \in \mathbb{N}$, $k_0 \neq 0$, such that $x^{k_0} \notin X$ (that is, $X \subsetneq \mathcal{C}([0,1])$). Clearly, $x^{k_0} \in \mathcal{C}([0,1])$, and due to the Corollary 2.2.41, exists a linear and bounded functional $T : \mathcal{C}([0,1]) \to \mathbb{R}$ such that

$$T(x^{k_0}) \neq 0 \quad \text{and} \quad T\Big|_X \equiv 0.$$

Since T satisfies the hypothesis of the Riesz-Markov-Kakutani Representation Theorem (Theorem 2.2.37), there exists a unique regular Borel complex measure μ such that

$$T(\varphi) = \int_0^1 \varphi(t)d\mu(t), \quad \forall \varphi \in \mathcal{C}([0,1]),$$

satisfying also

(i) $T(t^{k_0}) = \int_0^1 t^{k_0} d\mu(t) \neq 0$,

(ii) $T(t^{\lambda_j}) = \int_0^1 t^{\lambda_j} d\mu(t) = 0$, $\forall j = 1, 2, \dots$.

By Lemma 2.3.4, since T satisfies (2.3.2), we have that $T(t^{k_0}) = 0$, which contradicts the fact that $T(t^{k_0}) \neq 0$. Thus, $t^{k_0} \in X$ and $X = \mathcal{C}([0,1])$.

□

3 RECIPROCAL ON FULL MÜNTZ-SZÁSZ APPROXIMATION THEOREM

On the previous chapter we have seen that a sufficient condition for the Full Müntz-Szász Theorem (Theorem 2.3.5) is that $\sum_{j=1}^{+\infty} \lambda_j/(\lambda_j^2+1) = +\infty$ for an arbitrary sequence $\{\lambda_j\}_{j=0}^{\infty}$ ($\lambda_0 = 0$) of distinct nonnegative real numbers. Our aim in this chapter is to prove that this condition is also a necessary condition, i.e., that the reciprocal of the Full Müntz-Szász Theorem also holds.

On this chapter, we present different vectorial subspaces of the real valued continuous functions space, which have some interesting properties about the zeros of its functions. To do so, we will study some inequalities due to Newman, S.N. Bernstein and P. Chebishev (see [7] and [18]) which are related with that spaces.

Moreover, these vectorial spaces will also be very interesting for us due to it will turn out that the subspace of the continuous functions $\langle x^{\lambda_0}, x^{\lambda_1}, \ldots, x^{\lambda_n}\rangle$ for the real values $0 \leq \lambda_0 < \lambda_1 < \cdots < \lambda_n$ is a particular case of all of them for every $n \in \mathbb{N}$.

Finally, we will see that such properties can be used in order to prove the reciprocal of the Full Müntz-Szász Theorem arguing by contradiction.

3.1 Müntz Systems

On that section we begin by introducing the vectorial spaces. The order that we have chosen to show them is from the biggest one to the smallest, since it turns out that each vectorial space that we are going to study is contained in the previous one.

Even though in the previous chapter we have worked with $\mathcal{C}([0,1])$, on this chapter we may work in some sections with $\mathcal{C}([a,b])$ for the real values $a < b$.

3.1.1 Chebyshev Systems

The first vectorial space that we study is the Chebyshev system. The ubiquitous of such system lie at the heart of many analytic problems, particularly problems on $\mathcal{C}([a,b])$, the space of real valued continuous functions equipped with the uniform norm

$$\|f\|_{[a,b]} = \sup_{x\in[a,b]} |f(x)|.$$

Although we will not see it in detail (since it is not the aim of these notes) the Chebyshev systems will generalize the idea of the vectorial space generated by the orthogonal Chebyshev polynomials $T_n(x) = \cos(n \text{ arcos } x)$ for $n \geq 0$ and $x \in [-1,1]$ (see [7] and [11]).

Then, on this section we will study this vectorial space, giving its properties and also presenting an important result that will play an important role on the proof of the reciprocal of the Full Müntz-Szász Theorem.

Definition 3.1.1. A sequence of functions $(f_k)_{k=0}^n \subseteq \mathcal{C}([a,b])$ is called a Haar system on $[a,b]$ if

$$\dim\left(\langle f_0, f_1, \ldots, f_n\rangle\right) = n+1. \tag{3.1.1}$$

A special type of Haar systems are the Chebyshev systems, which are those which also satisfies

$$\det\begin{pmatrix} f_0(x_0) & f_1(x_0) & \cdots & f_n(x_0) \\ f_0(x_1) & f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_n(x_n) \end{pmatrix} > 0$$

whenever $x_0 < x_1 < \cdots < x_n$, $\{x_i\}_{i=0}^n \subseteq [a,b]$.

We will say that the Chebyshev system $(f_k)_{k=0}^n$ is complete if $(f_k)_{k=0}^m$ is a Chebyshev system for every $0 \le m \le n$.

Remark 3.1.2. On this section, we will consider a Chebyshev system to be a complete Chebyshev system.

Now, we will see a characterization of the Chebyshev systems. However, we need first an important definition.

Definition 3.1.3. We call the point $x_0 \in [a,b]$ a double zero of $f \in \mathcal{C}([a,b])$ if $f(x_0) = 0$ and

$$f(x_0 - \varepsilon) \cdot f(x_0 + \varepsilon) > 0$$

for all sufficiently small $\varepsilon > 0$ (in other words, if f vanishes without changing sign at x_0). Otherwise, we call x_0 a simple zero of f.

Proposition 3.1.4 (Zeros of functions in Chebyshev Spaces)**.** *Let $(f_k)_{k=0}^n \subseteq \mathcal{C}([a,b])$ be a Chebyshev system. Then, every $0 \neq p \in \langle f_0, f_1, \ldots, f_n\rangle$ has at most n distinct zeros in $[a,b]$. Moreover, p has at most n zeros in $[a,b]$ even if each double zero is counted twice.*

Proof. Suppose that p has $n+1$ distinct zeros in $[a,b]$, namely $a \le x_0 < x_1 < \cdots < x_n \le b$, and assume that $p = \sum_{i=0}^n \mu_i f_i$ for some $\mu_i \in \mathbb{R}$. Then, we have the homogeneous linear system

$$\begin{pmatrix} f_0(x_0) & f_1(x_0) & \cdots & f_n(x_0) \\ f_0(x_1) & f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_n(x_n) \end{pmatrix} \begin{pmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Since the determinant of such homogeneous linear system is different from zero implies that the only solution is $\mu_i = 0$ for every i, which contradicts the fact that $p \neq 0$. Thus, p has at most n distinct zeros.

Now we assume that p has at least one double zero and p has at least $n+1$ zeros if each double zero is counted twice. We denote the distinct zeros of p by $a \le t_1 < \cdots < t_k \le b$ and add to these points the point $t_i + \varepsilon$ for each double zero t_i and also $t_i - \varepsilon$ for the first double zero.

Observe that we can take ε small enough such that the additional points are different from $t_1, \ldots, t_k$ and are contained in $[a, b]$, for example

$$\varepsilon = \min\{(t_1 - a)/2, (t_2 - t_1)/2, \ldots, (t_k - t_{k-1})/2, (b - t_k)/2\},$$

since in a and b we can not have a double zero. Furthermore, the resulting set contains at least $n+2$ points. This can be seen due to we are adding to the set of points $t_1, \ldots, t_k$ the point $t_i + \varepsilon$ for each double zero (so a total of at least $n+1$ points) and we also add the point $t_i - \varepsilon$ for the first double zero.

We arrange these in increasing order and relabel the first $n+2$ of these points as $s_0, s_1, \ldots, s_{n+1}$. We claim that the values $p(s_i)$ must then alternate in sign in the sense that $p(s_i) \geq 0$ for i odd and $p(s_i) \leq 0$ for i even or vice-versa.

To see this, observe that in this arrangement there is some i such that $p(s_i) \neq 0$ (otherwise p would not have a double zero). So take

$$j_0 = \min\{i\colon\ 0 \leq i \leq n+1 \text{ and } p(s_i) \neq 0\}$$

and consider s_{j_0}. Then, for $k < j_0$, $p(s_k) = 0$ and so $p(s_k)$ alternates in sign. By the definition of j_0, necessarily $p(s_{j_0+1}) = 0$ and $\operatorname{sign}(p(s_{j_0})) = \operatorname{sign}(p(s_{j_0+2}))$ (s_{j_0+1} is the first double zero). Now suppose that j_0 is even since the case that j_0 is odd is completely analogous. Then, $j_1 = j_0 + 2$ is also even. Therefore, if there is no more double zeros, we are done since this would mean that the other points are all simple zeros. Otherwise, take

$$j_2 = \min\{i\colon\ j_1 + 1 \leq i \leq n+1 \text{ and } p(s_i) \neq 0\}.$$

Hence for all $j_1 + 1 \leq i < j_2$, $p(s_i) = 0$, then p alternates its sign

$$(j_2 - 1) - (j_1 + 1) = (j_2 - j_1) - 2$$

times. Observe that s_{j_2} "keeps" the sign of s_{j_2-2} since s_{j_2-1} is a double zero. So, if j_2 is even, then

$$\operatorname{sign}(p(s_{j_0})) = \operatorname{sign}(p(s_{j_2}))$$

otherwise

$$\operatorname{sign}(p(s_{j_0})) = -\operatorname{sign}(p(s_{j_2})),$$

and then the alternation property of sign of the values $p(s_i)$ holds.

In either case, observe that if $p(s) = \sum_{i=0}^{n} \lambda_i f_i(s)$, then clearly

$$\begin{vmatrix} p(s_0) & p(s_1) & \cdots & p(s_{n+1}) \\ f_0(s_0) & f_0(s_1) & \cdots & f_0(s_{n+1}) \\ \vdots & \vdots & \ddots & \vdots \\ f_n(s_0) & f_n(s_1) & \cdots & f_n(s_{n+1}) \end{vmatrix} = 0$$

since the first row is a linear combination of the following rows. Upon expanding the determinant along the first row and using that

$$\begin{vmatrix} f_0(x_0) & f_1(x_0) & \cdots & f_n(x_0) \\ f_0(x_1) & f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_n(x_n) \end{vmatrix} > 0$$

whenever $x_0 < x_1 < \cdots < x_n$, $\{x_i\}_{i=0}^n \subseteq [0,1]$, we obtain

$$\sum_{i=0}^{n+1} (-1)^i p(s_i) a_i = 0$$

where a_i are the $n \times n$ minors which are strictly positive.

Hence, since the $p(s_i)$ alternate in sign, we can consider $(-1)^i p(s_i) a_i \geq 0$ for every i (otherwise multiply p by -1), and this implies $(-1)^i p(s_i) a_i = 0$, or what is the same, $p(s_i) = 0$ for every $i \in \{0, \ldots, n+1\}$, which yields a contradiction.

□

Definition 3.1.5. Let $(f_k)_{k=0}^n \subseteq \mathcal{C}([a,b])$ be a Chebyshev system. If $g \in \mathcal{C}([a,b])$ and $p \in \langle f_0, f_1, \ldots, f_n \rangle$ satisfy

$$\|g - p\|_{[a,b]} = \inf_{q \in \langle f_0, \ldots, f_n \rangle} \|g - q\|_{[a,b]},$$

then p is said to be a best approximation to g from $\langle f_0, f_1, \ldots, f_n \rangle$.

The following result ensures the existence of such best approximations.

Proposition 3.1.6. *Let $(f_k)_{k=0}^n \subseteq \mathcal{C}([a,b])$ be a Chebyshev system and let $g \in \mathcal{C}([a,b])$, then there exists a best approximation to g from $\langle f_0, f_1, \ldots, f_n \rangle$.*

Proof. Take $q \in \langle f_0, f_1, \ldots, f_n \rangle$. If

$$\|q - g\|_{[a,b]} = \inf_{h \in \langle f_0, \ldots, f_n \rangle} \|g - h\|_{[a,b]},$$

we are done. Otherwise, consider

$$T := \{p \in \langle f_0, f_1, \ldots, f_n \rangle \colon \ \|p - q\|_{[a,b]} \leq \|g - q\|_{[a,b]} + 1\}.$$

Since $\dim(\langle f_0, f_1, \ldots, f_n \rangle) = n + 1 < +\infty$, the set T is a compact subset of $\langle f_0, f_1, \ldots, f_n \rangle$.

Now, by the definition of infimum, there is a sequence $(p_j)_{j=1}^n \subset T$ such that

$$\|g - p_j\|_{[a,b]} \leq j^{-1} + \inf_{h \in \langle f_0, \ldots, f_n \rangle} \|g - h\|_{[a,b]}, \quad \text{for all } j \geq 1.$$

Therefore, since T is compact, $(p_j)_{j=1}^n$ has a convergent subsequence with limit in $T \subseteq \langle f_0, f_1, \ldots, f_n \rangle$ and this limit is so a best approximation to g from $\langle f_0, f_1, \ldots, f_n \rangle$.

□

Definition 3.1.7. Let $x_0 < \cdots < x_n$ be $n+1$ points of $[a,b]$. Then, $(x_0, \ldots, x_n)$ is said to be an alternation sequence of length $n+1$ for a real valued $f \in \mathcal{C}([a,b])$ if

$$|f(x_i)| = \|f\|_{[a,b]}, \ \ i = 0, 1, \ldots, n$$

and

$$\operatorname{sign}(f(x_{i+1})) = -\operatorname{sign}(f(x_i)), \ \ i = 0, 1, \ldots, n-1.$$

Lemma 3.1.8 (Functions in a Chebyshev Space with Prescribed Sign Changes). *Let $(f_k)_{k=0}^n \subseteq \mathcal{C}([a,b])$ be a Chebyshev system on $[a,b]$, and let*

$$a < z_1 < z_2 < \cdots < z_m < b, \;\; 0 \leq m \leq n.$$

Then, there is a function $p^* \in \langle f_0, \ldots, f_n \rangle$ such that

(i) $p^*(x) = 0$ if and only if $x = z_i$ for $i = 1, 2, \ldots, m$,

(ii) $p^*(x)$ changes sign at each z_i, $i = 1, 2, \ldots, m$.

Furthermore, if $m = n$ then p^* is unique (up to a constant).

Proof. If $m = n$, take

$$p^*(x) = \begin{vmatrix} f_0(x) & f_1(x) & \cdots & f_n(x) \\ f_0(z_1) & f_1(z_1) & \cdots & f_n(z_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(z_n) & f_1(z_n) & \cdots & f_n(z_n) \end{vmatrix} \in \langle f_0, \ldots, f_n \rangle.$$

Then, clearly $p^*(z_i) = 0$ for every $i = 1, \ldots, n$. Moreover, recall that a function in a Chebyshev Space has at most n distinct zeros (Proposition 3.1.4) which implies that $z_1 < z_2 < \cdots < z_n$ are the only zeros of p^*.

Finally, we have to see that $p^*(x)$ changes sign at each z_i, $i = 1, 2, \ldots, m$. If $y_i \in (z_i, z_{i+1})$ for some i, then

$$p^*(y_i) = \begin{vmatrix} f_0(y_i) & f_1(y_i) & \cdots & f_n(y_i) \\ f_0(z_1) & f_1(z_1) & \cdots & f_n(z_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(z_n) & f_1(z_n) & \cdots & f_n(z_n) \end{vmatrix} = (-1)^i \begin{vmatrix} f_0(z_1) & f_1(z_1) & \cdots & f_n(z_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(z_i) & f_1(z_i) & \cdots & f_n(z_i) \\ f_0(y_i) & f_1(y_i) & \cdots & f_n(y_i) \\ f_0(z_{i+1}) & f_1(z_{i+1}) & \cdots & f_n(z_{i+1}) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(z_n) & f_1(z_n) & \cdots & f_n(z_n) \end{vmatrix},$$

and since $z_1 < \cdots < z_i < y_i < z_{i+1} < \cdots < z_n$ and the fact that $(f_k)_{k=0}^n$ is a Chebyshev system imply that $\text{sign}(p^*(y_i)) = (-1)^i$. Similarly, if $a \leq y_0 < z_1$ and $z_n < y_n \leq b$, $\text{sign}(p^*(y_0)) = 1$ and $\text{sign}(p^*(y_n)) = (-1)^n$ respectively. Thus, p^* is our desired function.

If $m < n$, take

$$p^*(x) = \begin{vmatrix} f_0(x) & f_1(x) & \cdots & f_m(x) \\ f_0(z_1) & f_1(z_1) & \cdots & f_m(z_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(z_m) & f_1(z_m) & \cdots & f_m(z_m) \end{vmatrix} \in \langle f_0, \ldots, f_m \rangle.$$

Since we are considering that $(f_k)_{k=0}^m \subset (f_k)_{k=0}^n$ is also a Chebyshev system, then for what we have seen above, p^* is our desired function.

Finally we have to show the unicity when $m = n$. Let $q^* = \sum_{i=0}^{n} a_i f_i$ and $p^* = \sum_{j=0}^{n} b_j f_j$ two Chebyshev functions satisfying (i) and (ii). If $a_n = 0$, then since $(f_k)_{k=0}^{n-1}$ is also a Chebyshev system, it implies that q^* is a Chebyshev function on $(f_k)_{k=0}^{n-1}$ with n distinct zeros, which is impossible unless $q^* = 0$ (since it has at most $n-1$ zeros).

If $a_n \neq 0$, then consider

$$p^* - (b_n/a_n)q^* = \sum_{j=0}^{n-1} \left(b_j - \frac{a_j b_n}{a_n} \right) f_j \in \langle f_0, \dots, f_{n-1} \rangle,$$

and again, $p^* - (b_n/a_n)q^*$ has n zeros, which only holds if $p^* = (b_n/a_n)q^*$.

□

Proposition 3.1.9 (Alternation of Best Approximations)**.** *Suppose $(f_k)_{k=0}^{n} \subseteq \mathcal{C}([a,b])$ is a Chebyshev system. Then, $p \in \langle f_0, f_1, \dots, f_n \rangle$ is a best approximation to $g \in \mathcal{C}[a,b]$ from $\langle f_0, f_1, \dots, f_n \rangle$ in the uniform norm on $[a,b]$ if and only if there exists an alternation sequence of length $n+2$ for $g - p$ on $[a,b]$.*

Moreover, in this conditions such p is unique.

Proof. Assume first that p is a best approximation of required type and suppose an alternation sequence of maximal length for $g - p$ is $(x_0, \dots, x_m)$ where $x_i \in [a,b]$ and where $m < n+1$. Suppose, without loss of generality, that $g(x_0) - p(x_0) > 0$ (otherwise multiply $g - p$ by -1). Now let

$$Y := \{x \in [a,b] \colon \; |g(x) - p(x)| = \|g - p\|_{[a,b]}\}. \tag{3.1.2}$$

Note that Y is compact and clearly $x_i \in Y$ for every i (by the definition of alternation sequence). Since $(x_0, \dots, x_m)$ is an alternation sequence of maximal length, we can divide Y into $m+1$ disjoint compact subsets $Y_0, \dots, Y_m$ with $x_0 \in Y_0, \dots, x_m \in Y_m$ so that

$$\operatorname{sign}(g(x) - p(x)) = -\operatorname{sign}(g(y) - p(y)) \neq 0, \;\; x \in Y_i, y \in Y_{i+1},$$

for $i = 0, \dots, m-1$.

Now choose m points $z_1 < z_2 < \dots < z_m$ such that

$$\max Y_{i-1} < z_i < \min Y_i, \;\; i = 1, 2, \dots, m,$$

where

$$\max Y_{i-1} := \max_{y \in Y_{i-1}} y \quad \text{and} \quad \min Y_i := \min_{y \in Y_i} y.$$

Then applying Lemma 3.1.8, there exists a unique (up to a constant) Chebyshev function $p^* \in \langle f_0, f_1, \dots, f_n \rangle$ such that $p^*(x) = 0$ if and only if $x = z_i$ for $i = 1, \dots, m$, and p^* changes sign at each z_i, $i = 1, \dots, m$. Therefore, since

$$\max Y_{i-1} < z_i < \min Y_i < \max Y_i < z_{i+1} < \min Y_{i+1}$$

for $i = 1, \dots, m-1$, we can assume

$$\operatorname{sign}_{x \in Y_i}(p^*(x)) = (-1)^i = \operatorname{sign}_{x \in Y_i}((g-p)(x)), \;\; i = 0, 1, \dots, m.$$

We now claim that, for $\delta > 0$ sufficiently small,

$$\|g - (p + \delta p^*)\|_{[a,b]} < \|g - p\|_{[a,b]}, \tag{3.1.3}$$

which contradicts the fact that p is a best approximation (since $p + \delta p^* \in \langle f_0, \ldots, f_n \rangle$), and so there must exist an alternation set of length $n + 2$ for $g - p$ on $[a, b]$. To verify (3.1.3) we proceed as follows.

First recall that $g(x_0) - p(x_0) > 0$ (then $g(x) - p(x) > 0$ for every $x \in Y_0$) and $g - p$ alternates its sign in the compact sets Y_i. Hence, since the sets Y_i are all compact sets, and by the definition of Y in (3.1.2), for each $i = 0, 1, \ldots, m$ we can choose an open set $O_i \subset [a, b]$ (in the usual metric topology relative to $[a, b]$) containing Y_i so that for every $x \in \overline{O}_i$,

$$\operatorname{sign}(g(x) - p(x)) = \operatorname{sign}(p^*(x)) \tag{3.1.4}$$

and

$$|g(x) - p(x)| \geq \frac{1}{2} \|g - p\|_{[a,b]}. \tag{3.1.5}$$

Observe that (3.1.5) holds due to $Y_i \subseteq Y$. Now pick a $\delta_1 > 0$ such that for every $x \in B := [a, b] \setminus \bigcup_{i=0}^m O_i$ and $\delta \in (0, \delta_1)$,

$$|g(x) - (p(x) + \delta p^*(x))| \leq |g(x) - p(x)| + \delta |p^*(x))| < \|g - p\|_{[a,b]},$$

which can be done since B is compact and by construction $Y_i \cap B = \emptyset$, so we have that

$$\|g - p\|_B < \|g - p\|_{[a,b]}.$$

For example, we can take

$$\delta_1 = \frac{\|g - p\|_{[a,b]} - \|g - p\|_B}{2\|p^*\|_B} < \frac{\|g - p\|_{[a,b]} - \|g - p\|_B}{\|p^*\|_B}.$$

Now note that (3.1.4) and (3.1.5) allow us to pick a $\delta_2 > 0$ such that for $x \in A := \bigcup_{i=0}^m \overline{O}_i$ and $\delta \in (0, \delta_2)$,

$$|g(x) - (p(x) + \delta p^*(x))| < \|g - p\|_{[a,b]}. \tag{3.1.6}$$

For example, we can take

$$\delta_2 = \frac{\|g - p\|_{[a,b]}}{2\|p^*\|_A} < \frac{\|g - p\|_{[a,b]}}{\|p^*\|_A}.$$

So if $x \in \bigcup_{i=0}^m \overline{O}_i$, we have that $x \in \overline{O}_i$ for some i. Then, (3.1.5) first implies that $g(x) - p(x) \neq 0$. Moreover, (3.1.4) yields that we can consider the following cases:

- Case 1: $g(x) - p(x) < 0$, then $p^*(x) < 0$. So we have the inequalities

$$(g(x) - p(x)) - \delta p^*(x) > g(x) - p(x) = -|g(x) - p(x)| \geq -\|g - p\|_{[a,b]}$$

  and

$$\begin{aligned} (g(x) - p(x)) - \delta p^*(x) &= (g(x) - p(x)) + \delta |p^*(x)| \\ &< (g(x) - p(x)) + \frac{|p^*(x)|}{\|p^*\|_A} \|p - g\|_{[a,b]} \\ &\leq \|p - g\|_{[a,b]} - (p(x) - g(x)) \leq \|p - g\|_{[a,b]}. \end{aligned}$$

- Case 2: $g(x) - p(x) > 0$, then $p^*(x) > 0$. So we have the inequalities

$$(g(x) - p(x)) - \delta p^*(x) < g(x) - p(x) = |g(x) - p(x)| \leq \|p - g\|_{[a,b]}$$

and

$$\begin{aligned}(g(x) - p(x)) - \delta p^*(x) &= (g(x) - p(x)) - \delta |p^*(x)| \\ &> (g(x) - p(x)) - \frac{|p^*(x)|}{\|p^*\|_A} \|p - g\|_{[a,b]} \\ &\geq (g(x) - p(x)) - \|p - g\|_{[a,b]} \geq - \|p - g\|_{[a,b]} .\end{aligned}$$

Thus, all together clearly implies (3.1.6). Therefore, taking $\delta \in (0, \min(\delta_1, \delta_2))$ verifies (3.1.3) and finishes the first part of the proof.

The proof of the conversely is simple. Suppose that there is an alternation sequence of length $n+2$ for $g - p$ on $[0, 1]$, and suppose there exists a p^* with

$$\|g - p^*\|_{[a,b]} < \|g - p\|_{[a,b]} .$$

Let $x_0 < \cdots < x_{n+1}$ be the alternation sequence for $g - p$ on $[a, b]$, then

$$|(g - p)(x_i)| = \|g - p\|_{[a,b]}$$

for $i = 0, \ldots, n+1$, and

$$\text{sign}((g - p)(x_i)) = -\text{sign}((g - p)(x_{i+1}))$$

for $i = 0, \ldots, n$.

Now fix $i \in \{0, \ldots, n+1\}$ and suppose first that $\text{sign}((g - p)(x_i)) = 1$. Hence,

$$(g - p)(x_i) = \|g - p\|_{[a,b]} > \|g - p^*\|_{[a,b]} \geq (g - p^*)(x_i).$$

Therefore,

$$(p^* - p)(x_i) = (g - p)(x_i) - (g - p^*)(x_i) > 0.$$

Moreover, $\text{sign}((g - p)(x_{i+1})) = -1$ and

$$(g - p)(x_{i+1}) = - \|g - p\|_{[a,b]} < - \|g - p^*\|_{[a,b]} \leq (g - p^*)(x_{i+1}).$$

Therefore,

$$(p^* - p)(x_{i+1}) = (g - p)(x_{i+1}) - (g - p^*)(x_{i+1}) < 0.$$

If $\text{sign}((g - p)(x_i)) = -1$, we similarly see that

$$(p^* - p)(x_i) = (g - p)(x_i) - (g - p^*)(x_i) < 0$$

and

$$(p^* - p)(x_{i+1}) = (g - p)(x_{i+1}) - (g - p^*)(x_{i+1}) > 0.$$

Then $p^* - p$ alternates its sign at least $n+1$ times, one between two consecutive alternation points of $g-p$ on $[a,b]$. So, it has at least $n+1$ distinct zeros on (a,b) and, by Proposition 3.1.4, $p^* - p$ must be the zero function.

Now, if g has another best approximation $p_1 \in \langle f_0, f_1, \ldots, f_n \rangle$, then $\|g - p_1\|_{[a,b]} = \|g - p\|_{[a,b]}$. So, by the alternation characterization, as we argue above, $p_1 - p$ has at least $n+1$ zeros on (a,b). Finally, Proposition 3.1.4 implies that $p_1 = p$.

□

Proposition 3.1.9 allows us to define the Chebyshev polynomial function for a Chebyshev system $(f_k)_{k=0}^n$, following the notation of [18].

Definition 3.1.10. Let $(f_k)_{k=0}^n \subseteq \mathcal{C}([a,b])$ be a Chebyshev system, recall that then $(f_k)_{k=0}^{n-1} \subseteq \mathcal{C}([a,b])$ is also a Chebyshev system. So, there exists a best approximation P_n to f_n from $\langle f_0, f_1, \ldots, f_{n-1} \rangle$ which, by Proposition 3.1.9, P_n is unique. We say that

$$T_n := \frac{f_n - P_n}{\|f_n - P_n\|_{[a,b]}}$$

is the Chebyshev polynomial associated with the Chebyshev system $(f_k)_{k=0}^n$.

One can easily see from the results above that T_n satisfy the following properties.

(i) $T_n \in \langle f_0, f_1, \ldots, f_n \rangle$,

(ii) there exists an alternation sequence $(x_0, x_1, \ldots, x_n)$ for T_n on $[a,b]$, and

(iii) $\|T_n\|_{[a,b]} = 1$.

Observation 3.1.11. Since any function in $\langle f_0, f_1, \ldots, f_n \rangle$ has at most n distinct zeros, T_n has exactly n distinct zeros which are not double zeros, one between two consecutive alternation points of T_n.

Now, let's see a technical lemma that will be very useful on these chapter.

Lemma 3.1.12. *Let $f, g \in \mathcal{C}([a,b])$ such that $\|f\|_{[a,b]} = \|g\|_{[a,b]} \neq 0$. Suppose that f has $n+1$ alternation points in $[a,b]$. Then, $f \pm g$ has at least n zeros, where we are counting each double zero twice.*

Proof. First let's see that between any two consecutive alternation points of f, of which there are $n+1$, there is at least one zero of $f \pm g$, where may some of them coincide with an alternation point of f.

Let $x_0 < x_1 < \cdots < x_n$ be the $n+1$ alternation points of f and take some $i \in \{0, \ldots, n-1\}$. Assume that $\|f\|_{[a,b]} = 1 = \|g\|_{[a,b]}$ (otherwise divide f and g by their norm). So we have the following cases:

(i) $f(x_i) = \mp g(x_i)$, then $f(x_i) \pm g(x_i) = 0$,

(ii) $f(x_{i+1}) = \mp g(x_{i+1})$, then $f(x_{i+1}) \pm g(x_{i+1}) = 0$,

(iii) $f(x_i) \neq \mp g(x_i)$, $f(x_{i+1}) \neq \mp g(x_{i+1})$ and $f(x_i) = -1 = -f(x_{i+1})$. Then,

$$f(x_i) = -1 < \mp g(x_i) \quad \text{and} \quad f(x_{i+1}) = 1 > \mp g(x_{i+1}).$$

Hence, $f(x_i) \pm g(x_i) < 0$ and $f(x_{i+1}) \pm g(x_{i+1}) > 0$, which implies that $f \pm g$ vanishes at least in one point in the interval (x_i, x_{i+1}).

(iv) $f(x_i) \neq \mp g(x_i)$, $f(x_{i+1}) \neq \mp g(x_{i+1})$ and $f(x_i) = 1 = -f(x_{i+1})$. Then

$$f(x_i) = 1 > \mp g(x_i) \quad \text{and} \quad f(x_{i+1}) = -1 < \mp g(x_{i+1}).$$

Hence $f(x_i) \pm g(x_i) > 0$ and $f(x_{i+1}) \pm g(x_{i+1}) < 0$, which implies that $f \pm g$ vanishes at least in one point in the interval (x_i, x_{i+1}).

Thus, the claim follows. Now, observe that if we are in either case (i) or (ii), this zero of $f \pm g$ is at an internal alternation point of f. In that case, when $(f \pm g)(x_i) = 0$ for some $i \in \{1, \dots, n-1\}$, we claim that either x_i is a double zero of $f \pm g$ or there is at least another zero of $f \pm g$ in $[x_{i-1}, x_{i+1}] \setminus \{x_i\}$.

Assume that x_i is not a double zero. So, we can consider the following:

- Case 1: $f(x_{i-1}) \pm g(x_{i-1}) = 0$ or $f(x_{i+1}) \pm g(x_{i+1}) = 0$, and the claim follows.
- Case 2: $f(x_{i-1}) \pm g(x_{i-1}) \neq 0$ and $f(x_{i+1}) \pm g(x_{i+1}) \neq 0$ and $(f \pm g)(x) \neq 0$ for every $x \in [x_{i-1}, x_i)$. If $f(x_{i-1}) = -1 < \mp g(x_{i-1})$ (resp. $f(x_{i-1}) = 1 > \mp g(x_{i-1})$) then $f(x_{i+1}) = -1 < \mp g(x_{i+1})$ (resp. $f(x_{i+1}) = 1 > \mp g(x_{i+1})$). Now take $\varepsilon > 0$ small enough such that
$$(f \pm g)(x_i - \varepsilon)(f \pm g)(x_i + \varepsilon) < 0$$
and $(f \pm g)(x_i + \delta) \neq 0$ for all $0 < \delta \leq \varepsilon$. Then, by the continuity of $f \pm g$ and since $(f \pm g)(x) \neq 0$ for every $x \in [x_{i-1}, x_i)$, it yields that
$$f(x_i - \varepsilon) \pm g(x_i - \varepsilon) < 0, \text{ (resp. } f(x_i - \varepsilon) \pm g(x_i - \varepsilon) > 0)$$
and
$$f(x_i + \varepsilon) \pm g(x_i + \varepsilon) > 0, \text{ (resp. } f(x_i + \varepsilon) \pm g(x_i + \varepsilon) < 0).$$
Therefore, we have that in particular $f(x_i + \varepsilon) \pm g(x_i + \varepsilon) > 0$ and $f(x_{i+1}) \pm g(x_{i+1}) < 0$ (resp. $f(x_i + \varepsilon) \pm g(x_i + \varepsilon) < 0$ and $f(x_{i+1}) \pm g(x_{i+1}) > 0$), which implies that there is at least one zero in the interval $(x_i + \varepsilon, x_{i+1})$.
- Case 3: $f(x_{i-1}) \pm g(x_{i-1}) \neq 0$ and $f(x_{i+1}) \pm g(x_{i+1}) \neq 0$ and $(f \pm g)(x) \neq 0$ for every $x \in (x_i, x_{i+1}]$. Similarly as in the Case 2, we see that there is at least one zero in the interval $(x_{i-1}, x_i - \varepsilon)$, for some small enough $\varepsilon > 0$.

This proves the claim. Now, in counting the zeros that $f \pm g$ has between two alternation points of f, we see that it must have at least n zeros if we are counting each double zeros twice. This occurs because if we have three consecutive alternation points $x_{i-1} < x_i < x_{i+1}$ $(1 \leq i \leq n-1)$ then either we have at least two different zeros y_1 and y_2 such that $x_{i-1} \leq y_1 \leq x_i$

and $x_i \leq y_2 \leq x_{i+1}$; or we have a double zero at x_i (which is counted twice). All in all, $f \pm g$ has at least as many zeros as the number of pairs of alternating points

$$x_0 < x_1, x_1 < x_2, \ldots, x_{n-1} < x_n, \tag{3.1.7}$$

which are exactly n.

This ends the proof of the lemma.

□

Finally we introduce the most important result on this section, which characterizes the zeros of a Chebyshev polynomial. We want to remark that this result will take an important part on the proof of the Full Müntz-Szász Theorem.

Proposition 3.1.13 (Zeros of a Chebyshev Polynomial)**.** *Let $n \geq 4$ and let*

$$\mathcal{T} = (f_0, \ldots, f_{n-1}, f_n) \ \textit{and}\ \mathcal{S}_\sigma = (f_{\sigma_0}, \ldots, f_{\sigma_k}) \quad (0 \leq \sigma_0 < \cdots < \sigma_k \leq n)$$

be Chebyshev systems on $[a, b]$. Consider $T_n = T_{n,\mathcal{T}}$ and $S_n = S_{n,\mathcal{S}_\sigma}$ the associated Chebyshev polynomials of the Chebyshev systems $\mathcal{T}$ and $\mathcal{S}_\sigma$ respectively. Then between two consecutive zeros of S_n there is at least one zero of T_n.

Proof. First observe that since $(f_0, \ldots, f_{n-1}, f_n)$ is a Chebyshev system on $[a, b]$, $T_n \pm S_n \in \langle f_0, \ldots, f_{n-1}, f_n \rangle$ have at most n zeros (Proposition 3.1.4). So, by Lemma 3.1.12, $T_n \pm S_n$ must have exactly n zeros if we are counting each double zero twice.

Our aim is to proof that between two consecutive zeros of S_n there is at least one zero of T_n. To do so, we will suppose the contrary and we will reach a contradiction by the properties of the zeros of the functions $T_n \pm S_n$. Hence, suppose that there are two consecutive zeros of S_n without any zero of T_n between them, i.e., S_n has at least two consecutive zeros between two consecutive zeros of T_n. Namely

$$z_1 < y_1 < y_2 < z_2$$

such that

$$T_n(z_j) = S_n(y_j) = 0, \ \ (j = 1, 2).$$

Let $x_{i-1} < x_i < x_{i+1}$ be the three consecutive alternating points of T_n such that

$$x_{i-1} < z_1 < x_i < z_2 < x_{i+1},$$

for some $i \in \{1, \ldots, n-1\}$. We claim that in these conditions, either $T_n + S_n$ or $T_n - S_n$ has four zeros in $[x_{i-1}, x_{i+1}]$ (where we are counting each double zero twice).

For proving the claim, our first step is to see that in the interval $(y_1, y_2) \subset [x_{i-1}, x_{i+1}]$, either $T_n + S_n$ or $T_n - S_n$ has two zeros (where we are counting the double zeros twice).

So observe that we can assume $T_n(y_1) > 0$, since $T_n \pm S_n$ and $-(T_n \mp S_n)$ have the same number of zeros. By the continuity of T_n it follows that $T_n(y_2) > 0$. Moreover, we assume that $S_n(y) > 0$ if $y \in (y_1, y_2)$ (otherwise, multiply S_n by -1 since we are working with $T_n + S_n$ and $T_n - S_n$). Therefore, we are going to work with $T_n - S_n$.

Observe that there is a point $\hat{y} \in (y_1, y_2)$ such that $S_n(\hat{y}) = 1$ (since between two consecutive zeros of a Chebyshev polynomial there is an alternation point). Hence, necessarily $T_n(\hat{y}) \leq S_n(\hat{y}) = 1$.

First, if $T_n(\hat{y}) < S_n(\hat{y})$, then $T_n - S_n$ has at least one zero in the intervals $(y_1, \hat{y})$ and $(\hat{y}, y_2)$ respectively (since $T_n(y_j) - S_n(y_j) = T_n(y_j) > 0$, for $j = 1, 2$).

On the other side, if $T_n(\hat{y}) = S_n(\hat{y}) = 1$, since T_n is determined by the $n+1$ alternation points (due to in the definition of a Chebyshev polynomial, the determinant (3.1.1) is greater than zero and the matrix is invertible), necessarily $\hat{y} = x_2$. If x_2 is a double zero of $T_n - S_n$, then it will have at least two zeros in the interval (y_1, y_2). Otherwise, there will exists an $\varepsilon > 0$ small enough such that

$$(T_n - S_n)(\hat{y} - \varepsilon) \cdot (T_n - S_n)(\hat{y} + \varepsilon) < 0$$

and $y_1 < \hat{y} - \varepsilon < \hat{y} < \hat{y} + \varepsilon < y_2$. So if $(T_n - S_n)(\hat{y} - \varepsilon) < 0$, since $(T_n - S_n)(y_1) = T_n(y_1) > 0$, $T_n - S_n$ has at least a zero in the interval $(y_1, \hat{y} - \varepsilon)$. Conversely, if $(T_n - S_n)(\hat{y} + \varepsilon) < 0$, since $(T_n - S_n)(y_2) = T_n(y_2) > 0$, $T_n - S_n$ has at least a zero in the interval $(\hat{y} + \varepsilon, y_2)$.

All in all, $T_n - S_n$ has at least two zeros in the interval (y_1, y_2).

To end the claim, observe that since $T_n(y_j) - S_n(y_j) = T_n(y_j) > 0$ $(j = 1, 2)$ and $T_n(x_{i-1}) - S_n(x_{i-1}) = -1 - S_n(x_{i-1}) \leq 0$ and $T_n(x_{i+1}) - S_n(x_{i+1}) = -1 - S_n(x_{i+1}) \leq 0$ we have that $T_n - S_n$ has at least one zero in the intervals $[x_{i-1}, y_1)$ and $(y_2, x_{i+1}]$ respectively.

Therefore, either $T_n + S_n$ or $T_n - S_n$ has four zeros in $[x_{i-1}, x_{i+1}]$ (where we are counting each double zero twice). Our final step is to see that it is not possible the existence of the four zeros.

Assume that $T_n - S_n$ has four zeros in $[x_{i-1}, x_{i+1}]$ (where we are counting each double zero twice) since the case $T_n + S_n$ is completely analogous. For simplicity in counting the zeros of $T_n - S_n$ we will relate them with the pairs of alternation points (3.1.7) by saying that $T_n - S_n$ has n zeros, one for each pair.

So, if these zeros are different from x_{i-1} and x_{i+1}, and since $T_n - S_n$ has a zero for each pair of alternation points, for the $n-2$ pairs

$$x_0 < x_1, x_1 < x_2, \ldots, x_{i-2} < x_{i-1}, x_{i+1} < x_{i+2}, x_{n-1} < x_n$$

in addition to the four zeros in the interval (x_{i-1}, x_{i+1}), we have that $T_n - S_n$ has at least $n - 2 + 4 = n + 2$ zeros, but $T_n - S_n$ has n zeros.

If either x_{i-1} or x_{i+1} is a zero of $T_n - S_n$, but not both, then $T_n - S_n$ will have one less pair of consecutive alternation points (either the pair $x_{i-2} < x_{i-1}$ or the pair $x_{i+1} < x_{i+2}$). All in all, $T_n - S_n$ will have at least $n - 3 + 4 = n + 1$ zeros, but again we know that $T_n - S_n$ has exactly n zeros.

Finally, if x_{i-1} and x_{i+1} are zeros of $T_n - S_n$, then they must be not double zeros, since if we are counting them once, we have at least the zeros of the $n - 4$ pairs (without considering the pairs $x_{i-2} < x_{i-1}$ and $x_{i+1} < x_{i+2}$) plus the four zeros in the interval $[x_{i-1}, x_{i+1}]$, which are a total of at least $n - 4 + 4 = n$ zeros for $T_n - S_n$. So, if we count them twice, we will get $n + 2$ zeros for $T_n - S_n$, which is not possible. Then, suppose that x_{i-1} and x_{i+1} are both simple zeros of $T_n - S_n$.

Now observe that since $n \geq 4$, we can consider either the alternation point x_{i-2} or x_{i+2} (depending on if $i \geq 2$ or $i \leq n - 2$). Assume that $i \geq 2$, since the other case is completely analogous. As before, we can assume that $(T_n - S_n)(y_1) = T_n(y_1) > 0$.

Then, if there is no zero in the interval (x_{i-1}, y_1), by the continuity of $T_n - S_n$, we have that for any $\delta \in (0, z_1 - x_{i-1})$

$$(T_n - S_n)(x_{i-1} + \delta) > 0.$$

Moreover, since x_{i-1} is not a double zero, for a small enough $\delta > 0$, $(T_n - S_n)(x_{i-1} - \delta) < 0$ and, since $(T_n - S_n)(x_{i-2}) = 1 - S_n(x_{i-2}) \geq 0$, there is at least a zero of $T_n - S_n$ in $[x_{i-2}, x_{i-1})$.

If the zero is different from x_{i-2}, then we have to add one more zero to $T_n - S_n$ (which is not possible). Otherwise, x_{i-2} must be a simple zero and we can argue analogously with x_{i-3} (if $i \geq 3$). Iterating, until we get that the alternation point x_0 is a zero of $T_n - S_n$, we finally reach the same conclusion of adding another zero to $T_n - S_n$ (which is the x_0) and again we know that it is not possible.

Therefore, there are not two consecutive zeros of S_n between two consecutive zeros of T_n. Thus, there exists at least one zero of T_n between any two consecutive zeros of S_n.

□

3.1.2 Descartes Systems

Another vectorial space about we will talk is the Descartes system. This system results to be a particular case of a Chebyshev system. We have seen that the Chebyshev systems capture some of the essential properties of polynomials. We will see that the Descartes systems capture some additional properties.

For this vectorial space, we have followed the notation of [7] and [18].

Definition 3.1.14. We say that a Haar system $(f_0, \ldots, f_n)$ is a Descartes system on $[a, b]$ if for every $m \leq n$,

$$\det \begin{pmatrix} f_{i_0}(x_0) & f_{i_1}(x_0) & \cdots & f_{i_m}(x_0) \\ \vdots & \vdots & \ddots & \vdots \\ f_{i_0}(x_m) & f_{i_1}(x_m) & \cdots & f_{i_m}(x_m) \end{pmatrix} > 0$$

holds whenever $0 \leq i_0 < i_1 < \cdots < i_m \leq n$ and $a \leq x_0 < x_1 < \cdots < x_m \leq b$.

Observe that when we take $m = n$ we have exactly the definition of a Chebyshev system. The following version of the Descartes' rule of signs holds for Descartes systems.

Proposition 3.1.15 (Descartes' Rule of Signs)**.** *If $(f_0, \ldots, f_n)$ is a Descartes system on $[a, b]$, then the number of distinct zeros of any*

$$0 \neq f = \sum_{i=0}^{n} a_i f_i, \quad a_i \in \mathbb{R}$$

is bounded by the number of sign changes in $(a_0, \ldots, a_n)$, where we are considering a sign change between a_i and a_{i+k} when $a_i a_{i+k} < 0$ and $a_{i+1} = a_{i+2} = \cdots = a_{i+k-1} = 0$.

Proof. Suppose that $(a_0, \dots, a_n)$ has p sign changes. Then, we can partition $\{a_0, \dots, a_n\}$ into exactly $p+1$ blocks so that each block is of the form

$$a_{n_k+1}, a_{n_k+2}, \dots, a_{n_{k+1}}, \quad k = 0, 1, \dots, p$$

($n_0 := -1$, $n_{p+1} := n$), where all of the coefficients in each of the blocks are of the same sign, not all the coefficients in a block vanish and the last coefficient in a block is different from zero. Assume without loss of generality that the first block $a_0, \dots, a_{n_1}$ is a "positive block", that is $a_0, a_1, \dots, a_{n_1-1} \geq 0$ and $a_{n_1} > 0$ (otherwise, consider $-f$).

Now let

$$g_k := \sum_{i=n_k+1}^{n_{k+1}} |a_i| f_i, \quad k = 0, 1, \dots, p.$$

Then, for $0 \leq x_0 < x_1 < \dots < x_p \leq 1$,

$$\det \begin{pmatrix} g_0(x_0) & g_1(x_0) & \cdots & g_p(x_0) \\ g_0(x_1) & g_1(x_1) & \cdots & g_p(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ g_0(x_p) & g_1(x_p) & \cdots & g_p(x_p) \end{pmatrix} = \det \begin{pmatrix} \sum_{i=n_0+1}^{n_1} |a_i| f_i(x_0) & \cdots & \sum_{i=n_p+1}^{n_{p+1}} |a_i| f_i(x_0) \\ \sum_{i=n_0+1}^{n_1} |a_i| f_i(x_1) & \cdots & \sum_{i=n_p+1}^{n_{p+1}} |a_i| f_i(x_1) \\ \vdots & \ddots & \vdots \\ \sum_{i=n_0+1}^{n_1} |a_i| f_i(x_p) & \cdots & \sum_{i=n_p+1}^{n_{p+1}} |a_i| f_i(x_p) \end{pmatrix}$$

$$= \sum_{i_0=0}^{n_1} \cdots \sum_{i_p=n_p}^{n} |a_{i_0}| \cdots |a_{i_p}| \det \begin{pmatrix} f_{i_0}(x_0) & f_{i_1}(x_0) & \cdots & f_{i_p}(x_0) \\ f_{i_0}(x_1) & f_{i_1}(x_1) & \cdots & f_{i_p}(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ f_{i_0}(x_p) & f_{i_1}(x_p) & \cdots & f_{i_p}(x_p) \end{pmatrix} > 0,$$

since $0 \leq i_0 < i_1 < \dots < i_p \leq n$ and each of the determinants in the sum is positive (we have that $(f_j)_{j=0}^n$ is a Descartes system). Thus, $\{g_0, \dots, g_p\}$ is a $(p+1)$-dimensional Chebyshev system on $[a, b]$, and hence

$$f = g_0 - g_1 + \dots + (-1)^p g_p$$

has at most p zeros. This finishes the proof.

□

Now, we present a comparison theorem due to A.Pinkus [10] and, independently, P.W. Smith [12]. Before showing it, we will see first a technical lemma.

Lemma 3.1.16. *Let $0 \leq \delta_0 < \delta_1 < \dots < \delta_s \leq n$ and let $(f_{\delta_0}, \dots, f_{\delta_s})$ be a Descartes system in $[a, b]$. Take $a < x_1 < \dots < x_s < b$. Then there exists a unique $p = f_{\delta_s} + \sum_{i=0}^{s-1} a_i f_{\delta_i}$ such that $p(x) = 0$ if and only if $x = x_i$ for $i = 1, 2, \dots, s$. Moreover, such p has the following properties:*

(a) $p(x)$ changes sign at each x_i, $i = 1, 2, \dots, s$,

(b) $a_i a_{i+1} < 0$, for $i = 0, 1, \dots, s-1$, where $a_s := 1$,

(c) $p(x) > 0$, for $x \in (x_s, b]$.

Proof. Since $(f_{\delta_0}, \dots, f_{\delta_s})$ is also a Chebishev system, by Proposition 3.1.8 there exists a unique (up to a constant) $q = \sum_{i=0}^{s} q_i f_{\delta_i}$ such that

(i) $q(x) = 0$ if and only if $x = x_i$ for $i = 1, 2, \ldots, s$,

(ii) $q(x)$ changes sign at each x_i, $i = 1, 2, \ldots, s$.

Moreover, the fact that $(f_{\delta_0}, \ldots, f_{\delta_{s-1}})$ being also a Chebishev system implies that $q_s \neq 0$, since q has exactly s zeros. So, consider

$$p = \frac{q}{q_s} = \sum_{i=0}^{s} \frac{q_i}{q_s} f_{\delta_i} = f_{\delta_s} + \sum_{i=0}^{s-1} a_i f_{\delta_i}.$$

We claim that such p satisfies the desired properties.

First observe that we just have to see properties (b) and (c) (since the others are clearly satisfied).

Observe that (b) is a direct consequence of Proposition 3.1.15, since p has s zeros and s coefficients, then there must be exactly s sign changes on the coefficients. To see (c), observe that for $x \in [a, b]$ we have for some constant μ

$$\begin{vmatrix} f_{\delta_0}(x_1) & f_{\delta_1}(x_1) & \cdots & f_{\delta_s}(x_1) \\ f_{\delta_0}(x_2) & f_{\delta_1}(x_2) & \cdots & f_{\delta_s}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_{\delta_0}(x_s) & f_{\delta_1}(x_s) & \cdots & f_{\delta_s}(x_s) \\ f_{\delta_0}(x) & f_{\delta_1}(x) & \cdots & f_{\delta_s}(x) \end{vmatrix} = \mu p(x), \tag{3.1.8}$$

since the determinant above satisfies (i) and (ii). Hence,

$$\begin{aligned} \mu p(x) &= \sum_{i=0}^{s} (-1)^{s+i} f_{\delta_i}(x) \begin{vmatrix} f_{\delta_0}(x_1) & f_{\delta_1}(x_1) & \cdots & f_{\delta_{i-1}}(x_1) & f_{\delta_{i+1}}(x_1) & \cdots & f_{\delta_s}(x_1) \\ f_{\delta_0}(x_2) & f_{\delta_1}(x_2) & \cdots & f_{\delta_{i-1}}(x_2) & f_{\delta_{i+1}}(x_2) & \cdots & f_{\delta_s}(x_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{\delta_0}(x_s) & f_{\delta_1}(x_s) & \cdots & f_{\delta_{i-1}}(x_s) & f_{\delta_{i+1}}(x_s) & \cdots & f_{\delta_s}(x_s) \end{vmatrix} \\ &:= \sum_{i=0}^{s} (-1)^{s+i} f_{\delta_i}(x) b_i \end{aligned}$$

where $b_i > 0$ for $i = 0, \ldots, s$, due to $(f_{\delta_0}, \ldots, f_{\delta_s})$ is a Descartes system. Observe now that since the coefficient on p of f_{δ_s} is 1, we have that

$$\mu = (-1)^{s+s} b_s = b_s > 0.$$

Thus, since $\mu > 0$ and $(f_{\delta_0}, \ldots, f_{\delta_s})$ is a Descartes system, by (3.1.8), $p(x) > 0$ for $x \in (x_s, b]$.

□

Proposition 3.1.17. *Let us assume that $(f_0, \ldots, f_n)$ is a Descartes system on $[a, b]$, and let*

$$p = f_n + \sum_{i=1}^{m} a_i f_{k_i},\ q = f_n + \sum_{i=1}^{m} b_i f_{t_i} \quad \textit{with } a_i, b_i \in \mathbb{R} \textit{ and } m \leq n$$

be chosen such that $0 \le t_i \le k_i < n$ *for all* $i \in \{1, \dots, m\}$ *with strict inequality for at least one of the indexes* $i \in \{1, \dots, m\}$. *If* $p(x_i) = q(x_i) = 0$ *for the distinct points* $x_i \in [a, b]$, $i = 1, \dots, m$, *then*

$$|p(x)| \le |q(x)|, \ \forall x \in [a, b].$$

Furthermore, the inequality is strict for all $x \in [a, b] \setminus \{x_i\}_{i=1}^m$.

Proof. First suppose that there is an index j such that

$$t_j < k_j \quad \text{and} \quad t_i = k_i \text{ whenever } i \ne j.$$

So we assume

$$p = f_n + a_j f_{k_j} + \sum_{i=1\, i\ne j}^{m} a_i f_{k_i}$$

and

$$q = f_n + b_j f_{t_j} + \sum_{i=1\, i\ne j}^{m} b_i f_{k_i},$$

where $0 \le k_1 < k_2 < \cdots < k_m < n$ and $0 \le k_{j-1} < t_j < k_j$ (of course the inequality $k_{j-1} < t_j$ holds only if $j > 1$). Then

$$p - q = a_j f_{k_j} - b_j f_{t_j} + \sum_{i=1,\, i\ne j}^{m} (a_i - b_i) f_{k_i} \in \langle f_{k_1}, \dots, f_{k_{j-1}}, f_{t_j}, f_{k_j}, \dots, f_{k_m} \rangle$$

has at most m zeros on $[a, b]$. Since $(p - q)(x_i) = 0$ for $i = 1, \dots, m$, then $p - q$ has exactly m zeros on $[a, b]$ at $x_1, \dots, x_m$. Moreover, this implies that $a_i \ne b_i$ for $i \in \{1, \dots, m\} \setminus \{j\}$.
Applying Lemma 3.1.16 (c) to p and q, we have respectively

$$p(x) > 0 \quad \text{and} \quad q(x) > 0, \qquad x \in (x_m, b]. \tag{3.1.9}$$

Now we consider $\mu(p - q)$, where μ is chosen so that the lead coefficient of $\mu(p - q)$ is 1. So applying Lemma 3.1.16 (c) to $\mu(p - q)$, we have that

$$\mu(p(x) - q(x)) > 0, \qquad x \in (x_m, b]. \tag{3.1.10}$$

Observe that $p - q$ and p have the same coefficient for f_{k_j}. So by Lemma 3.1.16 (b) applied to $\mu(p - q)$ and p we have that between the sequence

$$\mu a_j, \mu(a_{j+1} - b_{j+1}), \dots, \mu(a_{m-1} - b_{m-1}), 1$$

there are $m - j$ sign changes for $\mu(p - q)$ (if $j = m$, then there are no sign changes) and between the sequence

$$a_j, a_{j+1}, \dots, a_m, 1$$

there are $m - j + 1$ sign changes for p. This means that

$$\text{sign}(a_j) = -\text{sign}(\mu a_j).$$

Therefore, $\mu < 0$ and by (3.1.10) we have that

$$p(x) - q(x) < 0, \quad x \in (x_m, b]. \tag{3.1.11}$$

Now, Lemma 3.1.16 (a) implies that $p-q$, p and q only changes sign at $x_1, \ldots, x_m$. Then observe that (3.1.9) means that p and q has the same sign in $[a,b]$, but (3.1.11) implies that $p-q$ alternates in sign with p and q in $[a,b]$. Thus, when $p(x) > 0$ and $q(x) > 0$, $p(x) < q(x)$ and when $p(x) < 0$ and $q(x) < 0$, $-p(x) < -q(x)$. Therefore,

$$|p(x)| \leq |q(x)|, \ \forall x \in [a,b].$$

Furthermore, since $p-q$ just vanishes at $x_1, \ldots, x_m$, then the inequality is strict for all $x \in [a,b] \setminus \{x_i\}_{i=1}^m$.

Finally, if there is more than one index such that $t_i < k_i$, then consider

$$j := \max\{i\colon\ 1 \leq i \leq m \text{ and } t_i < k_i\}.$$

Then, arguing analogously for j, we get the same result. This ends the proof.

□

3.1.3 Müntz Systems

Finally, we present the vectorial space that is related with the powers x^{λ_k} that we use on the Full Müntz-Szász Theorem. This powers will be of distinct nonnegative real numbers. We will see that these spaces are a particular case of Descartes systems, and therefore, of Chebyshev systems. Then, since they are smallest spaces, they will have more interesting properties than the Descartes systems and the Chebyshev systems have.

As in the Descartes systems and the Chebyshev systems, we have followed the notation on the references [7] and [18]. Moreover, on this section we will show some results that will be very useful for the next section (Section 3.2) in order to state some inequalities that holds in such systems.

Definition 3.1.18. Let $(x^{\lambda_k})_{k=0}^n \subseteq \mathcal{C}([a,b])$, where $0 \leq a < b < +\infty$. We call Müntz systems of order n the vectorial spaces

$$M(\Lambda_n) := M(\{\lambda_k\}_{k=0}^n) := \langle x^{\lambda_0}, \ldots, x^{\lambda_n} \rangle,$$

where $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_n < +\infty$ and $\Lambda_n = \{\lambda_k\}_{k=0}^n$.

We have two important properties of the Müntz systems. The first one is the following proposition.

Proposition 3.1.19. *Let us assume that $0 \leq \lambda_0 < \lambda_1 < \cdots < \lambda_n < +\infty$. Then $(x^{\lambda_k})_{k=0}^n$ is a Chebyshev system on $[a,b] \subset (0,+\infty)$, for every $0 < a < b < +\infty$.*

Proof. Since $\lambda_j \neq \lambda_i$ for every $i \neq j$, clearly $\dim\left(\langle x^{\lambda_0}, x^{\lambda_1}, \ldots, x^{\lambda_n}\rangle\right) = n+1$. Now let

$$\Delta = \{(\alpha_0, \alpha_1, \ldots, \alpha_n) \in \mathbb{R}^{n+1} \colon \exists i \neq j \text{ such that } \alpha_i = \alpha_j\}.$$

Then we claim that

$$D(\rho_0, \ldots, \rho_n) := \det \begin{pmatrix} x_0^{\rho_0} & x_0^{\rho_1} & \cdots & x_0^{\rho_n} \\ x_1^{\rho_0} & x_1^{\rho_1} & \cdots & x_1^{\rho_n} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{\rho_0} & x_n^{\rho_1} & \cdots & x_n^{\rho_n} \end{pmatrix} \neq 0$$

whenever $(\rho_0, \ldots, \rho_n) \in \mathbb{R}^{n+1} \setminus \Delta$ and $a \leq x_0 < \cdots < x_n \leq b$.

We will show it by induction on n. First, $D(\rho_0) = x_0^{\rho_0} \neq 0$ for every $x_0 \in [a,b]$ and $\rho_0 \in \mathbb{R}$. Moreover, for every $(\rho_0, \rho_1) \in \mathbb{R}^2 \setminus \Delta$ and $0 < a \leq x_0 < x_1 \leq b$,

$$\begin{aligned} D(\rho_0, \rho_1) = \begin{vmatrix} x_0^{\rho_0} & x_0^{\rho_1} \\ x_1^{\rho_0} & x_1^{\rho_1} \end{vmatrix} &= x_0^{\rho_0} x_1^{\rho_1} - x_0^{\rho_1} x_1^{\rho_0} = 0 \\ &\Leftrightarrow x_0^{\rho_0} x_1^{\rho_1} \left(1 - x_0^{\rho_1 - \rho_0} x_1^{\rho_0 - \rho_1}\right) = 0 \\ &\Leftrightarrow \left(\frac{x_0}{x_1}\right)^{\rho_1 - \rho_0} = 1, \end{aligned}$$

which is not possible since $\rho_0 \neq \rho_1$ and $x_0 < x_1$.

Hence, suppose that $D(\rho_0, \ldots, \rho_k) \neq 0$ whenever $0 \leq k < n$ and observe that

$$\begin{aligned} D(\rho_0, \ldots, \rho_n) &= \det \begin{pmatrix} x_0^{\rho_0} & x_0^{\rho_1} & \cdots & x_0^{\rho_n} \\ x_1^{\rho_0} & x_1^{\rho_1} & \cdots & x_1^{\rho_n} \\ \vdots & \vdots & \ddots & \vdots \\ x_n^{\rho_0} & x_n^{\rho_1} & \cdots & x_n^{\rho_n} \end{pmatrix} \\ &= x_0^{\rho_0} \cdots x_n^{\rho_0} \det \begin{pmatrix} 1 & x_0^{\rho_1 - \rho_0} & \cdots & x_0^{\rho_n - \rho_0} \\ 1 & x_1^{\rho_1 - \rho_0} & \cdots & x_1^{\rho_n - \rho_0} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^{\rho_1 - \rho_0} & \cdots & x_n^{\rho_n - \rho_0} \end{pmatrix}. \end{aligned} \tag{3.1.12}$$

Since $x_0^{\rho_0} \cdots x_n^{\rho_0} \neq 0$, we must check that the determinant on the right side of (3.1.12) is not zero. Take $\gamma_i = \rho_i - \rho_0 > 0$ for $i = 1, \ldots, n$ and suppose that the determinant above is zero. Then this means that the first column of the matrix is a linear combination of the others n columns. Hence, there exist $\mu_1, \ldots, \mu_n$ not all zero such that

$$\begin{cases} 1 &= \mu_1 x_0^{\gamma_1} + \cdots + \mu_n x_0^{\gamma_n}, \\ \vdots & \\ 1 &= \mu_1 x_n^{\gamma_1} + \cdots + \mu_n x_n^{\gamma_n}. \end{cases}$$

Consider

$$p(x) = 1 - \sum_{i=1}^{n} \mu_i x^{\gamma_i} \in \langle 1, x^{\gamma_1}, \ldots, x^{\gamma_n} \rangle.$$

Clearly, $p(x_i) = 0$ for $i = 0, \dots, n$. Now take

$$p'(x) = -\sum_{i=1}^{n} \mu_i \gamma_i x^{\gamma_i - 1} \in \langle x^{\gamma_1 - 1}, \dots, x^{\gamma_n - 1} \rangle,$$

which is continuous in $[a, b]$. Since $p(x_i) = p(x_{i+1}) = 1$ for $i = 0, \dots, n-1$, by the Rolle's Theorem, p' has at least n zeros, one in each interval (x_i, x_{i+1}) $(0 \le i \le n-1)$. But by the induction hypothesis, $(x^{\gamma_k - 1})_{k=1}^{n}$ is a Chebyshev system, which implies that p' is a Chebyshev function, and therefore, that p' has at most $n-1$ zeros. This gives a contradiction on supposing that the determinant in (3.1.12) is zero.

Now we take $\tau : [0, 1] \to \mathbb{R}^{n+1} \setminus \Delta$ a continuous path defined by

$$\tau(t) = ((1-t)\lambda_0, 1 + (1-t)\lambda_1, \dots, n + (1-t)\lambda_n)$$

such that $\tau(1) = (\lambda_0, \dots, \lambda_n)$ and $\tau(0) = (0, 1, \dots, n)$. This is possible since $0 \le \lambda_0 < \cdots < \lambda_n$ and for $i \ne j$,

$$i + (1-t)\lambda_i = j + (1-t)\lambda_j \Leftrightarrow 1 - t = -\frac{i-j}{\lambda_i - \lambda_j} < 0$$

but $1 - t \ge 0$. The continuity of τ and the continuity of the determinant implies that

$$\operatorname{sign}(D(\tau(0))) = \operatorname{sign}(D(\tau(1))) = +1,$$

since the last determinant $D(\tau(1))$ is the well known Vandermonde determinant.

□

Observation 3.1.20. In particular, when $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_n < +\infty$, $(x^{\lambda_k})_{k=0}^{n}$ is a Chebyshev system on $[a, b] \subset [0, +\infty)$, for every $0 \le a < b < +\infty$, since

$$\det \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & x_1^{\lambda_1} & \cdots & x_1^{\lambda_n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^{\lambda_1} & \cdots & x_n^{\lambda_n} \end{pmatrix} = \det \begin{pmatrix} x_1^{\lambda_1} & \cdots & x_1^{\lambda_n} \\ \vdots & \ddots & \vdots \\ x_n^{\lambda_1} & \cdots & x_n^{\lambda_n} \end{pmatrix} \ne 0$$

whenever $0 < x_1 < \cdots < x_n < +\infty$.

Moreover, as a consequence of Proposition 3.1.19, we have the second property of the Müntz systems.

Proposition 3.1.21. *Let* $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_n$, *then* $(1, x^{\lambda_1}, \dots, x^{\lambda_n})$ *is a Descartes system on each interval* $[a, b] \subset (0, +\infty)$.

Therefore, we can define the Chebyshev polynomial for the system $(x^{\lambda_k})_{k=0}^{n}$ on the interval $[a, b] \subset (0, +\infty)$. Now, we show some of its properties.

Properties 3.1.22. Let $\Lambda_n = \{\lambda_k\}_{k=0}^{n}$, $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_n$ and let $T_{n,\lambda}$ be the corresponding Chebyshev polynomial for the Chebyshev system $(x^{\lambda_k})_{k=0}^{n}$ on the interval $[a, b] \subset (0, +\infty)$. Then, the following properties hold

(i) $T_{n,\lambda} \in M(\Lambda_n)$,

(ii) there exists an alternation sequence $(x_0, x_1, \ldots, x_n)$ for $T_{n,\lambda}$ in $[a, b]$,

(iii) $T_{n,\lambda}$ has n simple distinct zeros (i.e., with sign change), one between two consecutive alternation points of $T_{n,\lambda}$,

(iv) $\|T_{n,\lambda}\|_{[a,b]} = 1$,

(v) $T'_{n,\lambda}$ has $n-1$ simple distinct zeros (one between two consecutive zeros of $T_{n,\lambda}$),

(vi) $x_0 = a$ and $x_n = b$, so $|T_{n,\lambda}(a)| = |T_{n,\lambda}(b)| = 1$.

Proof. Properties (i), (ii), (iii) and (iv) follows by the definition of the Chebyshev polynomial. Let's show (v). Recall that there exist n simple distinct zeros $y_1, \ldots, y_n$ such that $T_{n,\lambda}(y_i) = 0$ for $1 \leq i \leq n$. Hence, by the Rolle's Theorem, $T'_{n,\lambda}$ has at least $n-1$ simple distinct zeros in (a, b), one in each interval (y_i, y_{i+1}) $(1 \leq i \leq n-1)$.

Since $T'_{n,\lambda} \in \langle x^{\lambda_1 - 1}, \ldots, x^{\lambda_n - 1} \rangle$ and $(x^{\lambda_1 - 1}, \ldots, x^{\lambda_n - 1})$ is a Chebyshev system in $[a, b]$, $T'_{n,\lambda}$ hast at most $n-1$ zeros. Therefore, $T'_{n,\lambda}$ has exactly $n-1$ simple distinct zeros.

Finally, it remains to see (vi). Let $x_0, x_1, \ldots, x_n$ be the alternation points of $T_{n,\lambda}$. If $a < x_0$, there exists some small enough $\varepsilon > 0$ such that $a < x_0 - \varepsilon$ and $|T_{n,\lambda}(x_0 - \varepsilon)| < 1$, which implies that in x_0 there is a change of monotony of $T_{n,\lambda}$. Hence, there is another zero of $T'_{n,\lambda}$ at $x_0 < y_0$, which is impossible since $T'_{n,\lambda}$ has exactly $n-1$ simple distinct zeros. Similarly, we have the same result for b. This finishes the proof.

□

Observation 3.1.23. Indeed, the $n-1$ zeros of $T'_{n,\lambda}$ coincide with the alternation points $x_1, \ldots, x_{n-1}$ of $T_{n,\lambda}$.

We will use these results with the Müntz systems

$$M(\Lambda_n) = M(\{\lambda_k\}_{k=0}^n) \quad \text{and} \quad M(\Gamma_n) = M(\{\gamma_k\}_{k=0}^n)$$

taken in $[0, 1]$, where we assume that

$$0 = \lambda_0 < \lambda_1 < \cdots < \lambda_n, \quad 0 = \gamma_0 < \gamma_1 < \cdots < \gamma_n$$

and $\lambda_k \geq \gamma_k$ for all k. With this idea in mind, we take $s \in (0, 1)$ and denote by $T_{n,\lambda}$ and $T_{n,\gamma}$ the Chebyshev polynomials associated with $M(\Lambda_n)$ and $M(\Gamma_n)$ respectively on the interval $[1-s, 1]$.

The following lemma study the monotony of the Chebyshev polynomials $T_{n,\lambda}$ and $T_{n,\gamma}$.

Lemma 3.1.24. *The continuous functions $|T_{n,\lambda}(x)|$ and $|T_{n,\gamma}(x)|$ are monotone decreasing functions on the interval $[0, 1-s]$. Furthermore, if $\lambda_1 = \gamma_1 = 1$, then also $|T'_{n,\lambda}(x)|$ and $|T'_{n,\gamma}(x)|$ are monotone decreasing on the interval $[0, 1-s]$.*

Proof. Recall that by the properties of the Chebyshev polynomials on the Müntz systems we have that $T_{n,\lambda}$ has n zeros in $(s-1,1)$ (one between two consecutive alternation points of $T_{n,\lambda}$) and $T'_{n,\lambda}$ has $n-1$ in $(s-1,1)$ (one between two consecutive zeros of $T_{n,\lambda}$). Then, if $y_1 < \cdots < y_n$ are the zeros of $T_{n,\lambda}$, then either $T'_{n,\lambda}(y) > 0$ for $y \in [0, y_1)$ or $T'_{n,\lambda}(y) < 0$ for $y \in [0, y_1)$ (otherwise, $T'_{n,\lambda}$ would have at least another zero in $[0, y_1)$). Thus, since $T_{n,\lambda}(y_1) = 0$, it follows that $|T_{n,\lambda}|$ is monotone decreasing on the interval $[0, 1-s] \subset [0, y_1)$. Analogously, we see that $|T_{n,\gamma}(x)|$ is monotone decreasing.

For the second case, since $T'_{n,\lambda}(x)$ has exactly $n-1$ simple zeros in $(1-s,1)$ and $\lambda_1 = 1$, then $T''_{n,\lambda}(x)$ has exactly $n-2$ simple zeros in $(1-s,1)$ (one between two consecutive zeros of $T'_{n,\lambda}$ by the Rolle's Theorem). Therefore, arguing similarly as above, we have that $|T'_{n,\lambda}(x)|$ is monotone decreasing. Analogously, we see that $|T'_{n,\gamma}(x)|$ is monotone decreasing.

□

Now, we present a comparison result between the Müntz systems $M(\Lambda_n)$ and $M(\Gamma_n)$. Before showing it, we first will see a technical lemma.

Lemma 3.1.25. *With the hypothesis and notation just introduced, the following claims hold:*

(a) Let $y \in [0, 1-s)$. Then the maximum values of the expressions

$$\max_{0 \neq p \in M(\Lambda_n)} \frac{|p(y)|}{\|p\|_{[1-s,1]}} \quad \textit{and} \quad \max_{0 \neq p \in M(\Lambda_n)} \frac{|p'(y)|}{\|p\|_{[1-s,1]}}$$

are both attained by $p = \pm T_{n,\lambda}$, where in the second case we are assuming that $\lambda_1 \geq 1$.

(b) $|T_{n,\lambda}(0)| \leq |T_{n,\gamma}(0)|$. Furthermore, if $\lambda_1 = \gamma_1 = 1$ then also $|T'_{n,\lambda}(0)| \leq |T'_{n,\gamma}(0)|$.

Proof. Let's prove (a).

Suppose that there is $y \in [0, 1-s)$ such that exists some $p \in M(\Lambda_n)$ with

$$\frac{|p(y)|}{\|p\|_{[1-s,1]}} > |T_{n,\lambda}(y)|.$$

We will work with the Chebyshev functions $T_{n,\lambda} \pm p/\|p\|_{[1-s,1]}$ and we will reach a contradiction on counting their zeros.

Recall that in $[1-s, 1]$ the Chebyshev polynomial $T_{n,\lambda}$ has $n+1$ alternation points

$$x_0 < x_1 < \cdots < x_n \quad \text{such that} \quad |T_{n,\lambda}(x_i)| = 1,$$

for $i = 0, \ldots, n$. Therefore, by Lemma 3.1.12, the Chebyshev functions $T_{n,\lambda} \pm p/\|p\|_{[1-s,1]}$ have at least n zeros in $[1-s, 1]$, where we are counting each double zero twice. Hence, $T_{n,\lambda} \pm p/\|p\|_{[1-s,1]}$ has exactly n zeros in $[1-s, 1]$, where we are counting each double zero twice.

First, suppose that $T_{n,\lambda}(x_0) = 1$ (otherwise, multiply $T_{n,\lambda}$ by -1). Moreover, assume that $\text{sign}(T_{n,\lambda}(y)) = \text{sign}(p(y))$ and let's work with $T_{n,\lambda} - p/\|p\|_{[1-s,1]}$ (otherwise, consider $-p/\|p\|_{[1-s,1]}$

and work with $T_{n,\lambda}+p/\|p\|_{[1-s,1]}$). Then, by the continuity of $T_{n,\lambda}$ we have that $\text{sign}(T_{n,\lambda}(y)) = \text{sign}(T_{n,\lambda}(x_0)) = 1$, and it follows that

$$\left(T_{n,\lambda}(y) - \frac{p(y)}{\|p\|_{[1-s,1]}}\right) \cdot \left(T_{n,\lambda}(x_0) - \frac{p(x_0)}{\|p\|_{[1-s,1]}}\right) \le 0, \tag{3.1.13}$$

since $(T_{n,\lambda} - p/\|p\|_{[1-s,1]})(y) < 0$ and $(T_{n,\lambda} - p/\|p\|_{[1-s,1]})(x_0) \ge 0$.

So let's study the case when $T_{n,\lambda}(x_0) - p(x_0)/\|p\|_{[1-s,1]} = 0$ (hence $p(x_0)/\|p\|_{[1-s,1]} = 1$).

If x_0 is a double zero, then we will have that $T_{n,\lambda}(x_0) - p(x_0)/\|p\|_{[1-s,1]}$ has at least $n+1$ zeros in $[1-s,1]$, yielding that $T_{n,\lambda} = p/\|p\|_{[1-s,1]}$.

Otherwise, there exists an $\epsilon_0 > 0$ small enough such that

$$\text{sign}\left(\left(T_{n,\lambda} - p/\|p\|_{[1-s,1]}\right)(x_0 - \epsilon_0)\right) = \text{sign}\left(\left(T_{n,\lambda} - p/\|p\|_{[1-s,1]}\right)(y)\right) = -1$$

and

$$\text{sign}\left(\left(T_{n,\lambda} - p/\|p\|_{[1-s,1]}\right)(x_0 + \epsilon_0)\right) = 1.$$

Since $T_{n,\lambda}(x_1) - p(x_1)/\|p\|_{[1-s,1]} \le 0$, there is a zero of $T_{n,\lambda} - p/\|p\|_{[1-s,1]}$ in $(x_0+\epsilon_0, x_1]$. If this zero is less than x_1, then $T_{n,\lambda} - p/\|p\|_{[1-s,1]}$ will have at least $n+1$ zeros in $[1-s,1]$, yielding again that $T_{n,\lambda} = p/\|p\|_{[1-s,1]}$. So, we can assume that the zero is located at x_1. Arguing similarly, we have that if x_1 is a double zero, then $T_{n,\lambda} = p/\|p\|_{[1-s,1]}$. Otherwise, there is an small enough $\epsilon_1 > 0$ such that

$$\text{sign}\left(\left(T_{n,\lambda} - p/\|p\|_{[1-s,1]}\right)(x_1 + \epsilon_1)\right) = -1.$$

However, $T_{n,\lambda}(x_2) - p(x_2)/\|p\|_{[1-s,1]} \ge 0$, so it follows that there is a zero of $T_{n,\lambda} - p/\|p\|_{[1-s,1]}$ in $(x_1+\epsilon_1, x_2]$.

Working inductively, we see that for all $i \in \{0, \dots, n-1\}$, there is an $\epsilon_i > 0$ such that

$$\text{sign}\left(\left(T_{n,\lambda} - p/\|p\|_{[1-s,1]}\right)(x_i + \epsilon_i)\right) = (-1)^i$$

and

$$(-1)^i\left(T_{n,\lambda} - p/\|p\|_{[1-s,1]}\right)(x_{i+1}) \ge 0.$$

Therefore, there is a zero of $T_{n,\lambda} - p/\|p\|_{[1-s,1]}$ in $(x_i+\epsilon_i, x_{i+1}]$.

Thus, we deduce that the zeros of $T_{n,\lambda} - p/\|p\|_{[1-s,1]}$ are located in the alternation points and they must be simple zeros (otherwise, $T_{n,\lambda} = p/\|p\|_{[1-s,1]}$). Since there are $n+1$ alternation points, this implies again that $T_{n,\lambda} = p/\|p\|_{[1-s,1]}$.

Now, if $|T_{n,\lambda}(x_0)| - |p(x_0)|/\|p\|_{[1-s,1]} > 0$, therefore, (3.1.13) yields that $T_{n,\lambda} - p/\|p\|_{[1-s,1]}$ has at least one more zero between y and x_0, which implies that $T_{n,\lambda}(x) - p(x)/\|p\|_{[1-s,1]}$ has at least $n+1$ zeros in $[y,1]$. Thus $p = T_{n,\lambda}$. This proofs the first part of (a).

To prove the second part of (a), we do the following. Recall that $\lambda_1 \ge 1$. So first observe that if $f \in \langle 1, x^{\lambda_1}, \dots, x^{\lambda_n}\rangle$ is a differentiable Chebyshev function with n zeros in $[1-s,1]$ (counting each double zero twice), then by the Rolle's Theorem f' will have at least $n-1$ simple zeros in $(1-s,1)$ (where we are adding also to the zeros of f' the double zeros of f, but now counted once). Since $f' \in \langle x^{\lambda_1-1}, x^{\lambda_2-1}, \ldots, x^{\lambda_n-1}\rangle$ with $\lambda_i - 1 \geq 0$, for $i=1,\ldots,n$, it follows that f' is a Chebyshev function in $(x^{\lambda_1-1}, x^{\lambda_2-1}, \ldots, x^{\lambda_n-1})$, so it has exactly $n-1$ distinct simple zeros in $(1-s,1)$.

So suppose that there is $y \in [0,1-s)$ such that exists some $p \in M(\Lambda_n)$ with

$$\frac{|p'(y)|}{\|p\|_{[1-s,1]}} > |T'_{n,\lambda}(y)|.$$

Assume at this point that $T_{n,\lambda}(x_0) = 1$ (otherwise, multiply $T_{n,\lambda}$ by -1). Then, since all the zeros of $T_{n,\lambda}$ lies in $(1-s,1)$, we deduce that $T_{n,\lambda}(y) > 0$. Moreover, for the observation made before, $T'_{n,\lambda}(x) \pm p'(x)/\|p\|_{[1-s,1]}$ has exactly $n-1$ distinct simple zeros in $(1-s,1)$. Besides, by Lemma 3.1.24, $T'_{n,\lambda}(y) < 0$.

Now assume that $p'(y) < 0$ (otherwise multiply p by -1). Therefore,

$$0 < \frac{|p'(y)|}{\|p\|_{[1-s,1]}} - |T'_{n,\lambda}(y)| = -\frac{p'(y)}{\|p\|_{[1-s,1]}} + T'_{n,\lambda}(y).$$

Let $y_0 \in (x_0, x_1]$ be the closest zero to x_0 of $T_{n,\lambda} - p(y)/\|p\|_{[1-s,1]}$. Then, since the zeros of $T'_{n,\lambda} - p'/\|p\|_{[1-s,1]}$ lie between two consecutive zeros of $T_{n,\lambda} - p/\|p\|_{[1-s,1]}$,

$$-\frac{p'(z)}{\|p\|_{[1-s,1]}} + T'_{n,\lambda}(z) > 0 \quad (\forall z \in [0, y_0)).$$

By the first part of (a), we have that $T_{n,\lambda}(y) - p(y)/\|p\|_{[1-s,1]} > 0$, and since $p/\|p\|_{[1-s,1]}$ decreases faster than $T_{n,\lambda}$ in $[0,y_0)$, necessary $T_{n,\lambda}(y_0) - p(y_0)/\|p\|_{[1-s,1]} > 0$, and we reach a contradiction.

Let us now proof (b). Let $0 \neq p \in M(\Lambda_n)$ be such that it interpolates $T_{n,\lambda}$ at its zeros (which are exactly n and all of them are simple distinct zeros), and such that $|p(0)| = |T_{n,\lambda}(0)| > 0$ in $[0,1]$ (recall that $T_{n,\lambda}$ has its zeros in $[1-s,1]$). This can be done since p is a Chebyshev function and a function in a Chebyshev system is determined by $n+1$ points.

Therefore, it follows from Proposition 3.1.17 that $|p(x)| \leq |T_{n,\lambda}(x)|$ for all $x \in [0,1]$. In particular, $\|p\|_{[1-s,1]} \leq \|T_{n,\lambda}\|_{[1-s,1]} = 1$ and, taking into account part (a) of this lemma, we get

$$|T_{n,\lambda}(0)| = |p(0)| \leq \frac{|p(0)|}{\|p\|_{[1-s,1]}} \leq \frac{|T_{n,\gamma}(0)|}{\|T_{n,\gamma}\|_{[1-s,1]}} = |T_{n,\gamma}(0)|,$$

which proves the first part of (b).

To prove the second part of (b), the argument is similar. We take $0 \neq p \in M(\Lambda_n)$ such that it interpolates $T_{n,\lambda}$ at its zeros in $[s-1,1]$ and we normalize $p'(0) = T'_{n,\lambda}(0)$. Observe that there is no problem in normalize $p'(0)$ since is different from zero; otherwise, the fact that $\gamma_1 = 1$

would imply that $p \in \langle 1, x^{\gamma_2}, \ldots, x^{\gamma_n} \rangle$, which is also a Chebyshev system of dimension $n-1$, and then p would have at most $n-1$ zeros, contradicting the fact that it has exactly n.

Then, it follows from Proposition 3.1.17 that $|p(x)| \leq |T_{n,\lambda}(x)|$ for all $x \in [0,1]$. Hence, in particular we have that $\|p\|_{[1-s,1]} \leq \|T_{n,\lambda}\|_{[1-s,1]} = 1$ and it follows again from part (a) of this lemma that

$$|T'_{n,\lambda}(0)| = |p'(0)| \leq \frac{|p'(0)|}{\|p\|_{[1-s,1]}} \leq \frac{|T'_{n,\gamma}(0)|}{\|T_{n,\gamma}\|_{[1-s,1]}} = |T'_{n,\gamma}(0)|,$$

which proves the second part of (b).

□

Proposition 3.1.26 (Comparison Theorem)**.** *The inequality*

$$\max_{0 \neq p \in M(\Lambda_n)} \frac{\|p\|_{[0,1]}}{\|p\|_{[1-s,1]}} \leq \max_{0 \neq p \in M(\Gamma_n)} \frac{\|p\|_{[0,1]}}{\|p\|_{[1-s,1]}}$$

holds. Furthermore, if $\lambda_1 = \gamma_1 = 1$ *then*

$$\max_{0 \neq p \in M(\Lambda_n)} \frac{\|p'\|_{[0,1-s]}}{\|p\|_{[1-s,1]}} \leq \max_{0 \neq p \in M(\Gamma_n)} \frac{\|p'\|_{[0,1-s]}}{\|p\|_{[1-s,1]}}.$$

Proof. Let $y \in [0, 1-s)$. Observe that from Lemma 3.1.24 and Lemma 3.1.25 (b), we have that $|T_{n,\lambda}(y)| \leq |T_{n,\lambda}(0)| \leq |T_{n,\gamma}(0)|$. Then,

$$\begin{aligned}
\max_{0 \neq p \in M(\Lambda_n)} \frac{|p(y)|}{\|p\|_{[1-s,1]}} &= \frac{|T_{n,\lambda}(y)|}{\|T_{n,\lambda}\|_{[1-s,1]}} = |T_{n,\lambda}(y)| \leq |T_{n,\gamma}(0)| \\
&= \frac{|T_{n,\gamma}(0)|}{\|T_{n,\gamma}\|_{[1-s,1]}} = \max_{y \in [0,1-s]} \frac{|T_{n,\gamma}(y)|}{\|T_{n,\gamma}\|_{[1-s,1]}} \\
&\leq \max_{0 \neq p \in M(\Gamma_n)} \frac{\|p\|_{[0,1-s]}}{\|p\|_{[1-s,1]}} \leq \max_{0 \neq p \in M(\Gamma_n)} \frac{\|p\|_{[0,1]}}{\|p\|_{[1-s,1]}}.
\end{aligned}$$

On the other hand, if $y \in [1-s, 1]$, then

$$\max_{0 \neq p \in M(\Lambda_n)} \frac{|p(y)|}{\|p\|_{[1-s,1]}} \leq 1 \leq \max_{0 \neq p \in M(\Gamma_n)} \frac{\|p\|_{[0,1]}}{\|p\|_{[1-s,1]}}.$$

Hence,

$$\max_{0 \neq p \in M(\Lambda_n)} \frac{\|p\|_{[0,1]}}{\|p\|_{[1-s,1]}} \leq \max_{0 \neq p \in M(\Gamma_n)} \frac{\|p\|_{[0,1]}}{\|p\|_{[1-s,1]}},$$

which is what we wanted to prove. By analogous arguments, we have that for $y \in [0, 1-s]$

$$\begin{aligned}
\max_{0 \neq p \in M(\Lambda_n)} \frac{|p'(y)|}{\|p\|_{[1-s,1]}} &= \frac{|T'_{n,\lambda}(y)|}{\|T_{n,\lambda}\|_{[1-s,1]}} = |T'_{n,\lambda}(y)| \leq |T'_{n,\lambda}(0)| \leq |T'_{n,\gamma}(0)| \\
&= \frac{|T'_{n,\gamma}(y)|}{\|T_{n,\gamma}\|_{[1-s,1]}} \leq \max_{0 \neq p \in M(\Gamma_n)} \frac{\|p'\|_{[0,1-s]}}{\|p\|_{[1-s,1]}},
\end{aligned}$$

which is the second claim of the theorem.

□

3.2 Müntz Systems Inequalities

Let $\Lambda_n = \{\lambda_j\}_{j=0}^n$ ($\lambda_0 = 0$) be a sequence of nonnegative real distinct numbers, there are some inequalities that holds for the Müntz systems $M(\Lambda_n)$, for each $n \in \mathbb{N}$, when we constrain Λ_n under certain conditions. Such inequalities relates the behavior of a function in $M(\Lambda_n)$ with its derivative.

For simplicity, we will denote the uniform norm on the real interval $[0, 1]$ by $\|\cdot\|_\infty$.

3.2.1 Newman's Inequality

The first inequality to deal with is called Newman's Inequality (see [18]). Before showing it, we see a technical lemma.

Lemma 3.2.1 (Landau Inequality). *Let $f \in \mathcal{C}^2(\mathbb{R})$, then*

$$\|f'\|_\infty^2 \leq 4 \|f\|_\infty \|f''\|_\infty .$$

Proof. Let $a \in \mathbb{R}$, we have the following Taylor expansion

$$f(t) = f(a) + (t-a)f'(a) + \int_a^t (t-s)f''(s)ds.$$

So,

$$f'(a) = \frac{f(t) - f(a)}{t-a} - \frac{1}{t-a}\int_a^t (t-s)f''(s)ds.$$

Taking absolute values on both sides,

$$\begin{aligned} |f'(a)| &\leq \left|\frac{f(t)-f(a)}{t-a}\right| + \frac{1}{|t-a|}\left|\int_a^t (t-s)|f''(s)|ds\right| \\ &\leq \left|\frac{f(t)-f(a)}{t-a}\right| + \frac{\|f''\|_\infty}{|t-a|}\left|\int_a^t (t-s)ds\right| \leq \frac{2\|f\|_\infty}{|t-a|} + \frac{1}{2}\|f''\|_\infty |t-a|. \end{aligned} \tag{3.2.1}$$

Set $s = t - a$, then (3.2.1) is true for every s. In particular,

$$|f'(a)| \leq \inf_{s\in\mathbb{R}} \left(\frac{2\|f\|_\infty}{s} + s\frac{\|f''\|_\infty}{2} \right).$$

Now observe that the function

$$g(s) = \frac{2\|f\|_\infty}{s} + s\frac{\|f''\|_\infty}{2}$$

has minimum value at

$$s = 2\sqrt{\|f\|_\infty / \|f''\|_\infty}.$$

So, for any $a \in \mathbb{R}$,

$$\begin{aligned} |f'(a)| &\leq \frac{2\|f\|_\infty}{2\sqrt{\|f\|_\infty / \|f''\|_\infty}} + 2\sqrt{\|f\|_\infty / \|f''\|_\infty}\frac{\|f''\|_\infty}{2} \\ &= \sqrt{\|f\|_\infty / \|f''\|_\infty}\,(\|f''\|_\infty + \|f''\|_\infty) = \sqrt{4\|f\|_\infty \|f''\|_\infty}. \end{aligned}$$

Thus,

$$\|f'\|_\infty^2 \leq 4\,\|f\|_\infty\,\|f''\|_\infty\,.$$

□

Proposition 3.2.2 (Newman's Inequality)**.** *Let $\{\lambda_j\}_{j\in\mathbb{N}}$ be a sequence of different real positive numbers. Then, the inequality*

$$\|xp'(x)\|_\infty \leq 11 \cdot \left(\sum_{j=1}^{n} \lambda_j\right) \|p(x)\|_\infty$$

holds for all $p \in \langle 1, x^{\lambda_1}, \ldots, x^{\lambda_n}\rangle$ and all $n \in \mathbb{N}$.

Proof. Fix $n \in \mathbb{N}$ and set $\lambda_0 = 0$. We may assume, without loss of generality, that $M_n := \sum_{j=1}^{n} \lambda_j = 1$ since we may make the change of variable

$$x \to x^{1/M_n}.$$

Hence, $\lambda_j \in (0,1]$ for every $1 \leq j \leq n$.
Set $x = e^{-t}$. If $p(x) = \sum_{j=0}^{n} a_j x^{\lambda_j}$ and $q(t) = p\left(e^{-t}\right) = \sum_{j=0}^{n} a_j e^{-\lambda_j t}$ then

$$xp'(x) = x\sum_{j=0}^{n} \lambda_j a_j x^{\lambda_j - 1} = \sum_{j=0}^{n} \lambda_j a_j x^{\lambda_j} = \sum_{j=0}^{n} \lambda_j a_j e^{-\lambda_j t} = -q'(t),$$

so that we have changed our problem to one of estimating the uniform norm, on the interval $[0,+\infty)$, of the derivatives of functions of the form

$$\sum_{j=0}^{+\infty} a_j e^{-\lambda_j t}, \tag{3.2.2}$$

in terms of their uniform norm in the same interval, i.e., we have changed the problem to prove that

$$\|q'(t)\|_\infty \leq 11\,\|q(t)\|_\infty\,,$$

where now by $\|\cdot\|_\infty$ we mean the uniform norm in $[0,+\infty)$.
Let

$$B(z) := \prod_{j=1}^{n} \frac{z-\lambda_j}{z+\lambda_j} \in \mathcal{H}(\mathbb{H}_0)$$

and define

$$T(t) := \frac{1}{2\pi i}\int_\Gamma \frac{e^{-zt}}{B(z)}dz \in \mathcal{C}^\infty(\mathbb{R}), \qquad \text{where } \Gamma := \{z;\ |z-1| = 1\}.$$

Since $B(z)$ has zeros of order 1 in $\lambda_1, \ldots, \lambda_n$, if we let

$$b_k = \prod_{\substack{j=1\\ j\neq k}}^{n} \frac{\lambda_k + \lambda_j}{\lambda_k - \lambda_j} \neq 0,$$

for every $k \in \{1, \ldots, n\}$, it follows from the residue theorem that

$$T(t) = \sum_{k=1}^{n} b_k e^{-\lambda_k t}$$

and therefore, T is of the form of (3.2.2).

To prove Newman's inequality we first show the following estimate

$$|B(z)| \geq \frac{1}{3} \text{ for all } z \in \Gamma. \tag{3.2.3}$$

To do so, we will see that

$$\frac{|z-\lambda|}{|z+\lambda|} \geq \frac{2-\lambda}{2+\lambda},$$

when $z \in \Gamma$.

Take $z = 1 + e^{it}$ for $t \in [0, 2\pi]$. Then, it is easy to check that

$$\left(\frac{|z-\lambda|}{|z+\lambda|}\right)^2 = \frac{1+(1-\lambda)^2+2(1-\lambda)\cos(t)}{1+(1+\lambda)^2+2(1+\lambda)\cos(t)}. \tag{3.2.4}$$

If we derivate the expression of (3.2.4), we get

$$\frac{4\lambda^3 \sin(t)}{(2+2\lambda+\lambda^2+2(1+\lambda)\cos(t))^2}, \tag{3.2.5}$$

and by the expression of (3.2.5) we get that the function of (3.2.4) has a minimum value at $t = 0$. Therefore,

$$\left(\frac{|z-\lambda|}{|z+\lambda|}\right)^2 \geq \frac{1+(1-\lambda)^2+2(1-\lambda)\cos(0)}{1+(1+\lambda)^2+2(1+\lambda)\cos(0)} = \left(\frac{2-\lambda}{2+\lambda}\right)^2,$$

and it holds for every $z \in \Gamma$. Thus,

$$|B(z)| \geq \prod_{j=1}^{n} \frac{2-\lambda_j}{2+\lambda_j} \text{ for all } z \in \Gamma.$$

To estimate the above product, we take into consideration the fact that for all $x, y \geq 0$, the inequality

$$\left(\frac{1-x}{1+x}\right) \cdot \left(\frac{1-y}{1+y}\right) \geq \frac{1-(x+y)}{1+x+y} \tag{3.2.6}$$

holds. Iterating (3.2.6), leads us to the inequality

$$|B(z)| \geq \prod_{j=1}^{n} \frac{2-\lambda_j}{2+\lambda_j} = \prod_{j=1}^{n} \frac{1-\lambda_j/2}{1+\lambda_j/2} \geq \frac{1-\frac{1}{2}\sum_{j=1}^{n}\lambda_j}{1+\frac{1}{2}\sum_{j=1}^{n}\lambda_j} = \frac{1}{3}$$

which proves (3.2.3).

Moreover, it follows from the definition of T and the continuity and derivation under the integral sign theorem, that

$$T''(t) = \frac{1}{2\pi i}\int_\Gamma \frac{z^2 e^{-zt}}{B(z)}dz.$$

So taking $z = 1 + e^{i\theta}$ as a parametrization of Γ, the Fubini's Theorem yields that

$$\begin{aligned}\int_0^{+\infty} |T''(t)| \leq & \frac{3}{2\pi}\int_0^{+\infty}\int_0^{2\pi} |1+e^{i\theta}|^2 \left|e^{-t(1+e^{i\theta})}\right| d\theta dt \\ = & \frac{3}{2\pi}\int_0^{+\infty}\int_0^{2\pi} 2(1+\cos(\theta))e^{-(1+\cos(\theta))t} d\theta dt \\ = & \frac{3}{\pi}\int_0^{2\pi} (1+\cos(\theta))\frac{1}{(1+\cos(\theta))}d\theta = 6.\end{aligned}$$

Now, we will compute integrals of the form

$$\int_0^{+\infty} e^{-\lambda_k t}T''(t)dt$$

in terms of the scalars λ_k. So fix $k \in \{1, \ldots, n\}$ and note that by the Fubini's Theorem,

$$\begin{aligned}\int_0^{+\infty} e^{-\lambda_k t}T''(t)dt = & \frac{1}{2\pi i}\int_\Gamma \int_0^{+\infty} \frac{z^2 e^{-(z+\lambda_k)t}}{B(z)}dtdz \\ = & \frac{1}{2\pi i}\int_\Gamma \frac{z^2}{B(z)(z+\lambda_k)}dz,\end{aligned} \tag{3.2.7}$$

and taking into consideration the fact that

$$\frac{z^2}{B(z)(z+\lambda_k)}$$

has no poles in the exterior of Γ (since $0 < \lambda_k \leq 1$), the above integral (3.2.7) depends only on its residue at ∞. Now, for every z in the exterior of Γ,

$$\frac{z^2}{z+\lambda_k} = z \cdot \sum_{j=0}^{+\infty}\left(\frac{-\lambda_k}{z}\right)^j = z - \lambda_k + \frac{\lambda_k^2}{z} - \cdots$$

and taking the Taylor expansion around the ∞ we get

$$\frac{1}{B(z)} = \prod_{j=1}^n \frac{z+\lambda_j}{z-\lambda_j} = 1 + \frac{2\left(\sum_{j=1}^n \lambda_j\right)}{z} + \frac{\left(2\sum_{j=1}^n \lambda_j\right)^2}{2z^2} + \cdots = 1 + \frac{2}{z} + \frac{2}{z^2} + \cdots$$

so that

$$\frac{z^2}{B(z)(z+\lambda_k)} = z + (2-\lambda_k) + \frac{\lambda_k^2 - 2\lambda_k + 2}{z} + \cdots .$$

This, together with (3.2.7), leads us to the formula

$$\int_0^{\infty} e^{-\lambda_k t}T''(t)dt = \lambda_k^2 - 2\lambda_k + 2, \tag{3.2.8}$$

by its residue at ∞.

Now let q be an exponential polynomial of the form (3.2.2). Then, if we take into consideration (3.2.8), we conclude that for every $a \in [0, +\infty)$

$$\begin{aligned}\int_0^{+\infty} q(t+a)T''(t)dt &= \int_0^{+\infty} \left(\sum_{k=0}^{n} a_k e^{-\lambda_k(t+a)} T''(t)\right) dt \\ &= \sum_{k=0}^{n} a_k e^{-\lambda_k a} \int_0^{+\infty} e^{-\lambda_k t} T''(t) dt \\ &= \sum_{k=0}^{n} a_k e^{-\lambda_k a} \left(\lambda_k^2 - 2\lambda_k + 2\right) \\ &= q''(a) + 2q'(a) + 2q(a).\end{aligned}$$

Hence,

$$\begin{aligned}|q''(a) + 2q'(a) + 2q(a)| &= \left|\int_0^{+\infty} q(t+a)T''(t)dt\right| \\ &\leq \int_0^{+\infty} |q(t+a)T''(t)|\, dt \leq \|q\|_\infty \int_0^{+\infty} |T''(t)|\, dt \\ &\leq 6\,\|q\|_\infty\,.\end{aligned}$$

Therefore,

$$|q''(a)| - |2q'(a) + 2q(a)| \leq |q''(a) + 2q'(a) + 2q(a)| \leq 6\,\|q\|_\infty\,. \tag{3.2.9}$$

And (3.2.9) holds for every $a \geq 0$, so that

$$\|q''\|_\infty \leq 2\,\|q'\|_\infty + 8\,\|q\|_\infty\,.$$

Now, Lemma 3.2.1 yields that

$$\|f'\|_\infty^2 \leq 4\,\|f\|_\infty\,\|f''\|_\infty$$

holds for all functions $f \in \mathcal{C}^2([0, +\infty))$, so that

$$\|q'\|_\infty^2 \leq 4\,\|q\|_\infty\,\|q''\|_\infty \leq 4\,\|q\|_\infty\,(2\,\|q'\|_\infty + 8\,\|q\|_\infty)$$

and

$$\left(\frac{\|q'\|_\infty}{\|q\|_\infty}\right)^2 \leq 8\left(\frac{\|q'\|_\infty}{\|q\|_\infty}\right) + 32.$$

Now consider the equation

$$x^2 - 8x - 32 = (x - 4(1-\sqrt{3})) \cdot (x - 4(1+\sqrt{3})).$$

Then, since $4(1-\sqrt{3}) < 0$, it follows that $x^2 - 8x - 32 \leq 0$ if and only if

$$x \leq 4(1+\sqrt{3}) < \left[4(1+\sqrt{3})\right] + 1 = 11.$$

Thus, $\|q'\|_\infty \leq 11\,\|q\|_\infty$ for all expressions of the form (3.2.2).

$\square$

3.2.2 Gram Matrix and Determinant

In some cases we can give an explicit computation of the best approximation of some given space. For example, using what is called the Gram Matrix, which we present here, we can compute the best approximation to elements of a given Hilbert space (see [8]). This results will be used in the next section, where we will work with another inequality for the Müntz spaces.

Then, given H a Hilbert space over a field K (so we have an inner product inside it) and let $x_1, x_2, \dots, x_n$ independent elements of H. Consider $V = \langle x_1, x_2, \dots, x_n\rangle$ a vectorial subspace of H. Then we can construct an orthonormal system in H by a sequence of orthonormal elements $x_1^*, x_2^*, \dots, x_n^*$. That is,

$$(x_i, x_j^*) = \begin{cases} 0, & i \neq j, \\ 1, & i = j, \end{cases}$$

where $(\cdot, \cdot)$ denotes the inner product in H.

Moreover, observe that if $w \in V$, then $w = a_1x_1 + \cdots + a_nx_n$ for some $a_i \in K$. Since we can get an orthonormal system from the independent elements $x_1, x_2, \dots, x_n$, then $w = b_1x_1^* + \cdots + b_nx_n^*$ for some $b_i \in K$, which implies that every element of V can be written by its Fourier expansion respect the orthonormal system $x_1^*, x_2^*, \dots, x_n^*$.

Now, before introducing the Gram matrix, let's see a few technical results of this orthonormal basis.

Proposition 3.2.3 (Bessel's Inequality)**.** *Let $x_1^*, x_2^*, \dots, x_n^*$ be an orthonormal system and let y be arbitrary. Then,*

$$\left\| y - \sum_{i=1}^{N} (y, x_i^*)x_i^* \right\| \leq \min_{(a_1,\dots,a_N)} \left\| y - \sum_{i=1}^{N} a_i x_i^* \right\|$$

for every $N \in \mathbb{N}$, $N \leq n$.

Proof.

$$\begin{aligned}
\left\| y - \sum_{i=1}^{N} a_i x_i^* \right\|^2 &= \left(y - \sum_{i=1}^{N} a_i x_i^*, y - \sum_{i=1}^{N} a_i x_i^* \right) \\
&= (y, y) - \sum_{i=1}^{N} a_i(x_i^*, y) - \sum_{i=1}^{N} \overline{a_i}(y, x_i^*) + \sum_{i,j=1}^{N} a_i\overline{a_j}(x_i^*, x_j^*) \\
&= (y, y) - \sum_{i=1}^{N} a_i(x_i^*, y) - \sum_{i=1}^{N} \overline{a_i}(y, x_i^*) + \sum_{i=1}^{N} |a_i|^2 \\
&+ \sum_{i=1}^{N} |(y, x_i^*)|^2 - \sum_{i=1}^{N} |(y, x_i^*)|^2 \\
&= (y, y) - \sum_{i=1}^{N} |(y, x_i^*)|^2 + \sum_{i=1}^{N} |a_i - (y, x_i^*)|^2.
\end{aligned}$$

Since the first two terms of the last member are independent of the a_i's and the last term is greater or equal than zero, it is clear that the minimum of

$$\left\| y - \sum_{i=1}^{N} a_i x_i^* \right\|^2$$

is achieved if and only if $a_i = (y, x_i^*)$, for $i = 1, \ldots, N$. This ends the proof.

□

As a direct consequence, we have the following result.

Corollary 3.2.4. *Let $x_1, x_2, \ldots, x_n$ be independent elements and let $x_1^*, x_2^*, \ldots, x_n^*$ be an orthonormal system respect to the x_i's. Then, for any element y,*

$$y - \sum_{k=1}^{n} (y, x_k^*) x_k^*$$

is orthogonal to x_i^*, for every $i \in \{1, \ldots, n\}$.

Moreover, another result that follows from Proposition 3.2.3 is the following.

Corollary 3.2.5. *Let $a_1x_1 + a_2x_2 + \cdots + a_nx_n$ be the best approximation to y from among the linear combinations of $x_1, \ldots, x_n$ (assumed independent). Then the coefficients a_i are the solution of the following system of equations:*

$$\begin{cases} a_1(x_1, x_1) + a_2(x_2, x_1) + \cdots + a_n(x_n, x_1) = (y, x_1), \\ \vdots \\ a_1(x_1, x_n) + a_2(x_2, x_n) + \cdots + a_n(x_n, x_n) = (y, x_n). \end{cases}$$

Proof. First, we can write $a_1x_1 + a_2x_2 + \cdots + a_nx_n$ in the orthonormal basis as $b_1x_1^* + \cdots + b_nx_n^*$. Observe that since $b_1x_1^* + \cdots + b_nx_n^*$ is the best approximation to y, by Proposition 3.2.3, $b_i = (y, x_i^*)$. Moreover, we also can write x_j in the orthonormal basis as $x_j = c_1x_1^* + \cdots + c_nx_n^*$. Therefore, this theorem follows directly by the fact that for every $1 \leq j \leq n$

$$(y - a_1x_1 - \cdots - a_nx_n, x_j) = (y - (y, x_1^*)x_1^* - \cdots - (y, x_n^*)x_n^*, c_1x_1^* + \cdots + c_nx_n^*) = 0$$

which the last equality is a consequence of Corollary 3.2.4.

□

Now we are in conditions to introduce the Gram matrix and the Gram determinant.

Definition 3.2.6. Given a sequence of elements $x_1, \ldots, x_n$ in an inner product space. The $n \times n$ matrix

$$G(x_1, \ldots, x_n) := \begin{pmatrix} (x_1, x_1) & (x_1, x_2) & \cdots & (x_1, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ (x_n, x_1) & (x_n, x_2) & \cdots & (x_n, x_n) \end{pmatrix}$$

is known as the Gram matrix of $x_1, \ldots, x_n$. Its determinant is known as the Gram determinant of $x_1, \ldots, x_n$ and denoted by $g(x_1, \ldots, x_n)$.

The following result gives us a formula for computing the best approximation in a Hilbert space using the Gram Determinant. The proof is due to [8], but there are also distinct proofs of it, for example, [14].

Proposition 3.2.7. *Let $x_1, \ldots, x_n$ be independent. If*

$$\delta = \min_{(a_i)} \|y - (a_1x_1 + a_2x_2 + \cdots + a_nx_n)\|,$$

then

$$\delta^2 = \frac{g(x_1, \ldots, x_n, y)}{g(x_1, \ldots, x_n)}.$$

Proof. Let $s = a_1x_1 + a_2x_2 + \cdots + a_nx_n$ be the best approximation to y. Then,

$$\delta^2 = \|y - s\|^2 = (y - s, y - s) = (y - s, y) - (y - s, s).$$

By Corollary 3.2.5, $(y - s, s) = 0$ so that

$$\delta^2 = (y - s, y) = (y, y) - (s, y).$$

Therefore, we have the following system of equations

$$\begin{cases} a_1(x_1, x_1) + a_2(x_2, x_1) + \cdots + a_n(x_n, x_1) - (y, x_1) = 0, \\ \vdots \\ a_1(x_1, x_n) + a_2(x_2, x_n) + \cdots + a_n(x_n, x_n) - (y, x_n) = 0, \\ a_1(x_1, y) + a_2(x_2, y) + \cdots + a_n(x_n, y) + \delta^2 - (y, y) = 0. \end{cases} \tag{3.2.10}$$

If we introduce the value $a_{n+1} = 1$ as a coefficient of the elements of the last column, then (3.2.10) becomes a system of $n+1$ homogeneous linear equations in $n+1$ variable $a_1, \ldots, a_n, a_{n+1}$, which possesses a nontrivial solution. Thus, the determinant of this system must therefore vanish:

$$\begin{vmatrix} (x_1, x_1) & (x_2, x_1) & \cdots & (x_n, x_1) & 0 - (y, x_1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (x_1, x_n) & (x_2, x_n) & \cdots & (x_n, x_n) & 0 - (y, x_n) \\ (x_1, y) & (x_2, y) & \cdots & (x_n, y) & \delta^2 - (y, y) \end{vmatrix} = 0.$$

Therefore, by the properties of the determinants

$$\delta^2 g(x_1, \ldots, x_n) = \begin{vmatrix} (x_1, x_1) & (x_2, x_1) & \cdots & (x_n, x_1) & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (x_1, x_n) & (x_2, x_n) & \cdots & (x_n, x_n) & 0 \\ (x_1, y) & (x_2, y) & \cdots & (x_n, y) & \delta^2 \end{vmatrix}$$

$$= \begin{vmatrix} (x_1, x_1) & (x_2, x_1) & \cdots & (x_n, x_1) & (y, x_1) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ (x_1, x_n) & (x_2, x_n) & \cdots & (x_n, x_n) & (y, x_n) \\ (x_1, y) & (x_2, y) & \cdots & (x_n, y) & (y, y) \end{vmatrix} = g(x_1, \ldots, x_n, y).$$

This ends the proof.

□

3.2.3 Bernstein-Chebyshev's Inequality

On this section, we will prove the Bernstein and Chebyshev's Inequality Theorem (see [7] and [9]) for the arbitrary Müntz system $\Pi(\Lambda) := \langle 1, x^{\lambda_1}, x^{\lambda_2}, \dots \rangle$ for an increasing sequence $\Lambda = \{\lambda_k\}_{k=0}^{+\infty}$ ($\lambda_0 = 0$) of different positive real numbers. However, we first will show a reduced version of it, where we suppose the sequences of exponents that satisfy the following jump condition:

$$\inf_{k\in\mathbb{N}} (\lambda_k - \lambda_{k-1}) > 0.$$

Before seeing it, let's see a result about error computations that follows from the results that we have already seen on the Gram Matrix section.

Lemma 3.2.8. *For all $m \in \mathbb{N}$*

$$\begin{aligned} E(x^{\lambda_m}, \Pi(\Lambda \setminus \{\lambda_m\})) &:= \min_{p\in\Pi(\Lambda\setminus\{\lambda_m\})} \left\| x^{\lambda_m} - p(x) \right\|_{L^2([0,1])} \\ &= \frac{1}{2\lambda_m + 1} \prod_{k\geq 0,\, k\neq m} \left| \frac{\lambda_m - \lambda_k}{\lambda_m + \lambda_k + 1} \right|. \end{aligned}$$

Proof. By Proposition 3.2.7, we have that for every $n \in \mathbb{N}$, $n > m$,

$$E(x^{\lambda_m}, M_n(\Lambda \setminus \{\lambda_m\})) = \sqrt{\frac{g(1, x^{\lambda_1}, \dots, x^{\lambda_{m-1}}, x^{\lambda_{m+1}}, \dots, x^{\lambda_n}, x^{\lambda_m})}{g(1, x^{\lambda_1}, \dots, x^{\lambda_{m-1}}, x^{\lambda_{m+1}}, \dots, x^{\lambda_n})}},$$

where g denotes the Gram determinant of the Gram matrix respect to the Hilbert space $L^2([0,1])$.

Now observe that

$$(x^{\lambda_i}, x^{\lambda_j}) = \int_0^1 x^{\lambda_i+\lambda_j}\, dx = \frac{1}{\lambda_i + \lambda_j + 1}.$$

Therefore,

$$\begin{aligned} &g(1, x^{\lambda_1}, \dots, x^{\lambda_{m-1}}, x^{\lambda_{m+1}}, \dots, x^{\lambda_n}) \\ &= \det \begin{pmatrix} (1,1) & (1,x^{\lambda_1}) & \cdots & (1,x^{\lambda_{m-1}}) & (1,x^{\lambda_{m+1}}) & \cdots & (1,x^{\lambda_n}) \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ (x^{\lambda_{m-1}},1) & (x^{\lambda_{m-1}},x^{\lambda_1}) & \cdots & (x^{\lambda_{m-1}},x^{\lambda_{m-1}}) & (x^{\lambda_{m-1}},x^{\lambda_{m+1}}) & \cdots & (x^{\lambda_{m-1}},x^{\lambda_n}) \\ (x^{\lambda_{m+1}},1) & (x^{\lambda_{m+1}},x^{\lambda_1}) & \cdots & (x^{\lambda_{m+1}},x^{\lambda_{m-1}}) & (x^{\lambda_{m+1}},x^{\lambda_{m+1}}) & \cdots & (x^{\lambda_{m+1}},x^{\lambda_n}) \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ (x^{\lambda_n},1) & (x^{\lambda_n},x^{\lambda_1}) & \cdots & (x^{\lambda_n},x^{\lambda_{m-1}}) & (x^{\lambda_n},x^{\lambda_{m+1}}) & \cdots & (x^{\lambda_n},x^{\lambda_n}) \end{pmatrix} \\ &= \det \begin{pmatrix} 1 & \frac{1}{\lambda_1+1} & \cdots & \frac{1}{\lambda_{m-1}+1} & \frac{1}{\lambda_{m+1}+1} & \cdots & \frac{1}{\lambda_n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\ \frac{1}{\lambda_{m-1}+1} & \frac{1}{\lambda_{m-1}+\lambda_1+1} & \cdots & \frac{1}{2\lambda_{m-1}+1} & \frac{1}{\lambda_{m-1}+\lambda_{m+1}+1} & \cdots & \frac{1}{\lambda_{m-1}+\lambda_n+1} \\ \frac{1}{\lambda_{m+1}+1} & \frac{1}{\lambda_{m+1}+\lambda_1+1} & \cdots & \frac{1}{\lambda_{m+1}+\lambda_{m-1}+1} & \frac{1}{2\lambda_{m+1}+1} & \cdots & \frac{1}{\lambda_{m+1}+\lambda_n+1} \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\lambda_n+1} & \frac{1}{\lambda_n+\lambda_1+1} & \cdots & \frac{1}{\lambda_n+\lambda_{m-1}+1} & \frac{1}{\lambda_n+\lambda_{m+1}+1} & \cdots & \frac{1}{2\lambda_n+1} \end{pmatrix}. \end{aligned}$$

If we take $a_i = \lambda_i$ and $b_i = \lambda_i + 1$, where $i = 0, \ldots, m-1, m+1, \ldots, n$ and $\lambda_0 = 0$, then the determinant above is the well known Cauchy determinant of order n, and we get that

$$\begin{aligned} g(1, x^{\lambda_1}, \ldots, &x^{\lambda_{m-1}}, x^{\lambda_{m+1}}, \ldots, x^{\lambda_n}) \\ &= \frac{\prod_{1\le i<j\le m-1}(a_j - a_i)(b_j - b_i)\prod_{m+1\le i<j\le n}(a_j - a_i)(b_j - b_i)}{\prod_{1\le i,\, j\le m-1}(a_i + b_j)\prod_{m+1\le i,\, j\le n}(a_i + b_j)}. \end{aligned}$$

Similarly,

$$\begin{aligned} g(1, &x^{\lambda_1}, \ldots, x^{\lambda_{m-1}}, x^{\lambda_{m+1}}, \ldots, x^{\lambda_n}, x^{\lambda_m}) \\ &= g(1, x^{\lambda_1}, \ldots, x^{\lambda_{m-1}}, x^{\lambda_{m+1}}, \ldots, x^{\lambda_n}) \frac{\prod_{i=1,\, i\ne m}^{n}(a_m - a_i)\prod_{i=1,\, i\ne m}^{n}(b_m - b_i)}{\prod_{i=1}^{n}(a_i + b_m)\prod_{i=1}^{n}(a_m + b_i)}. \end{aligned}$$

Thus,

$$\begin{aligned} E(x^{\lambda_m}, \Pi(\Lambda \setminus \{\lambda_m\}))^2 &= \frac{\prod_{i=1,\, i\ne m}^{n}(a_m - a_i)\prod_{i=1,\, i\ne m}^{n}(b_m - b_i)}{\prod_{i=1}^{n}(a_i + b_m)\prod_{i=1}^{n}(a_m + b_i)} \\ &= \frac{\prod_{i=1,\, i\ne m}^{n}(\lambda_m - \lambda_i)\prod_{i=1,\, i\ne m}^{n}(\lambda_m - \lambda_i)}{\prod_{i=1}^{n}(\lambda_i + \lambda_m + 1)\prod_{i=1}^{n}(\lambda_m + \lambda_i + 1)} \\ &= \frac{\prod_{i=1,\, i\ne m}^{n}(\lambda_m - \lambda_i)^2}{(2\lambda_m + 1)^2\prod_{i=1\, i\ne m}^{n}(\lambda_m + \lambda_i + 1)^2}. \end{aligned} \tag{3.2.11}$$

By taking the square root of (3.2.11) the theorem follows.

□

Proposition 3.2.9 (Bounded Bernstein's Inequality for Special Sequences)**.** *Let us assume that $\Lambda = \{\lambda_k\}_{k=0}^{+\infty}$ is an increasing sequence of nonnegative real numbers that satisfies the jump condition $\inf_{k\in\mathbb{N}}(\lambda_k - \lambda_{k-1}) > 0$, and $\sum_{k=1}^{+\infty} 1/\lambda_k < \infty$, $\lambda_0 = 0$, $\lambda_1 \ge 1$. Then for all $\varepsilon \in (0,1)$ there exists a constant $c_\varepsilon = c(\varepsilon, \Lambda) > 0$ such that the inequalities*

$$\|p'\|_{[0,1-\varepsilon]} \le c_\varepsilon \|p\|_{L^2(0,1)}, \quad \|p'\|_{[0,1-\varepsilon]} \le c_\varepsilon \|p\|_{[0,1]}$$

hold for all $p \in \langle 1, x^{\lambda_1}, x^{\lambda_2}, \ldots \rangle$.

Proof. It follows from Lemma 3.2.8 that for all $m \in \mathbb{N}$ and all $p \in \Pi(\Lambda \setminus \{\lambda_m\})$ the inequality

$$\begin{aligned} \left\| x^{\lambda_m} - p(x) \right\|_{L^2([0,1])} &\ge \frac{1}{2\lambda_m + 1}\prod_{k\ge 0,\, k\ne m}\left|\frac{\lambda_m - \lambda_k}{\lambda_m + \lambda_k + 1}\right| \\ &= \frac{1}{2\lambda_m + 1}\prod_{k\ge 0,\, k\ne m}\left|\frac{(\lambda_k + 1/2) - (\lambda_m + 1/2)}{(\lambda_k + 1/2) + (\lambda_m + 1/2)}\right| \end{aligned} \tag{3.2.12}$$

holds. Hence it is of interest to study products of the form

$$\prod_{k\ge 0,\, k\ne m}\left|\frac{\alpha_k + \alpha_m}{\alpha_k - \alpha_m}\right|$$

for sequences $\{\alpha_k\}_{k=0}^{+\infty}$ such that $\inf_{k\in\mathbb{N}}(\alpha_k - \alpha_{k-1}) > 0$, $\alpha_k \geq 1/2$, and $\sum_{k=1}^{+\infty} 1/\alpha_k < +\infty$ (note that we have, for ease of exposition, reversed the quotients).

Now, we decompose

$$\prod_{k\geq 0,\, k\neq m} \left|\frac{\alpha_k+\alpha_m}{\alpha_k-\alpha_m}\right| = \prod_{k\geq 0,\, \alpha_k<\alpha_m} \left|\frac{\alpha_k+\alpha_m}{\alpha_k-\alpha_m}\right| \prod_{k\geq 0,\, \alpha_m<\alpha_k<2\alpha_m} \left|\frac{\alpha_k+\alpha_m}{\alpha_k-\alpha_m}\right| \prod_{k\geq 0,\, \alpha_k\geq 2\alpha_m} \left|\frac{\alpha_k+\alpha_m}{\alpha_k-\alpha_m}\right|. \tag{3.2.13}$$

Let's see that the decomposition (3.2.13) can be done. First, we can write the third product as

$$\left|\frac{\alpha_k+\alpha_m}{\alpha_k-\alpha_m}\right| = \prod_{k\geq 0,\, \alpha_k\geq 2\alpha_m} \left|1+\frac{2\alpha_m}{\alpha_k-\alpha_m}\right|,$$

since for all k such that $\alpha_k \geq 2\alpha_m$ we have that $\alpha_k/2 \geq \alpha_m$. Hence, $\alpha_k-\alpha_m \geq \alpha_k-\alpha_k/2 = \alpha_k/2$, which yields

$$\begin{aligned} 1 \leq \prod_{k\geq 0,\, \alpha_k\geq 2\alpha_m} \left|1+\frac{2\alpha_m}{\alpha_k-\alpha_m}\right| &= \prod_{k\geq 0,\, \alpha_k\geq 2\alpha_m} \left(1+\frac{2\alpha_m}{\alpha_k-\alpha_m}\right) \\ \leq \prod_{k\geq 0,\, \alpha_k\geq 2\alpha_m} \exp\left(\frac{2\alpha_m}{\alpha_k-\alpha_m}\right) &= \exp\left(\sum_{k\geq 0,\, \alpha_k\geq 2\alpha_m}\left(\frac{2\alpha_m}{\alpha_k-\alpha_m}\right)\right) \\ = \exp\left(\sum_{k\geq 0,\, \alpha_k\geq 2\alpha_m}\frac{2\alpha_m}{\alpha_k-\alpha_m}\right) &\leq \exp\left(4\alpha_m\sum_{k\geq 0,\, \alpha_k\geq 2\alpha_m}\frac{1}{\alpha_k}\right) < +\infty, \end{aligned}$$

where for the inequalities we have used that $1+x \leq e^x$ for $x \geq 0$. Moreover, the surjectivity of the exponential implies that there exist a constant $\xi_{1,m} > 0$ such that

$$\xi_{1,m} \leq 4\sum_{k\geq 0,\, \alpha_k\geq 2\alpha_m}\frac{1}{\alpha_k} \quad \text{and} \quad \lim_{m\to+\infty}\xi_{1,m} = 0$$

satisfying

$$\prod_{k\geq 0,\, \alpha_k\geq 2\alpha_m} \left|1+\frac{2\alpha_m}{\alpha_k-\alpha_m}\right| = \exp\left(\alpha_m\xi_{1,m}\right).$$

Now, it remains to bound the products

$$\prod_{k\geq 0,\, \alpha_k<\alpha_m} \left|\frac{\alpha_k+\alpha_m}{\alpha_k-\alpha_m}\right| \tag{3.2.14}$$

and

$$\prod_{k\geq 0,\, \alpha_m<\alpha_k<2\alpha_m} \left|\frac{\alpha_k+\alpha_m}{\alpha_k-\alpha_m}\right|. \tag{3.2.15}$$

We begin with (3.2.14). To do so, first we claim that

$$\lim_{k\to+\infty}\frac{\alpha_k}{k} = +\infty. \tag{3.2.16}$$

Using that $\{\alpha_k\}_{k=0}^{+\infty}$ is monotonically increasing and $\sum_{k=1}^{+\infty} 1/\alpha_k < +\infty$, we have the following inequality

$$\sum_{k=n}^{2n} \frac{1}{\alpha_k} \geq \frac{2n+1}{\alpha_{2n}} \geq \frac{2n}{\alpha_{2n}} \geq 0, \quad (n \in \mathbb{N})$$

and tending n to $+\infty$ it yields that

$$0 \leq \lim_{n\to+\infty} \frac{2n}{\alpha_{2n}} \leq \lim_{n\to+\infty} \sum_{k=n}^{2n} \frac{1}{\alpha_k} = 0.$$

Now, let $r := \inf_{k\in\mathbb{N}} (\alpha_k - \alpha_{k-1})$ and observe that for $0 \leq k < m$ we have the inequality

$$\alpha_m - \alpha_k \geq (m-k)r.$$

Hence, for $0 \leq k < m$ it holds that

$$\frac{\alpha_k + \alpha_m}{\alpha_m - \alpha_k} \leq \frac{2\alpha_m}{(m-k)r}.$$

Therefore,

$$\prod_{k\geq 0,\, \alpha_k<\alpha_m} \left|\frac{\alpha_k + \alpha_m}{\alpha_k - \alpha_m}\right| \leq \prod_{k\geq 0,\, \alpha_k<\alpha_m} \frac{2\alpha_m}{(m-k)r} = \left(\frac{2}{r}\right)^m \frac{\alpha_m^m}{m!}. \tag{3.2.17}$$

Using into (3.2.17) the Stirling formula $n! \geq n^n e^{-n}$, which is valid for any $n \geq N$ (where $N \in \mathbb{N}$ is big enough), we get that

$$\prod_{k\geq 0,\, \alpha_k<\alpha_m} \left|\frac{\alpha_k + \alpha_m}{\alpha_k - \alpha_m}\right| \leq \left(\frac{2}{r}\right)^m \left(\frac{\alpha_m}{m}\right)^m e^m \quad (m \geq N). \tag{3.2.18}$$

Finally, applying that $\lim_{x\to+\infty} x^{1/x} = 1$ into (3.2.18) and using (3.2.16), we get that

$$1 \leq \prod_{k\geq 0,\, \alpha_k<\alpha_m} \left(\left|\frac{\alpha_k + \alpha_m}{\alpha_k - \alpha_m}\right|\right)^{\frac{1}{\alpha_m}} \leq \left(\frac{2}{r}\right)^{\frac{m}{\alpha_m}} \left(\frac{\alpha_m}{m}\right)^{\frac{m}{\alpha_m}} e^{\frac{m}{\alpha_m}} \to 1, \quad (m \to +\infty).$$

Thus, since

$$\lim_{m\to+\infty} \log\left(\left(\frac{2}{r}\right)^{\frac{m}{\alpha_m}} \left(\frac{\alpha_m}{m}\right)^{\frac{m}{\alpha_m}} e^{\frac{m}{\alpha_m}}\right) = 0$$

and by the surjectivity of the exponential, there exist a constant $\xi_{2,m} > 0$ such that

$$\lim_{m\to+\infty} \xi_{2,m} = 0 \quad \text{and} \quad \xi_{2,m} \leq \log\left(\left(\frac{2}{r}\right)^{\frac{m}{\alpha_m}} \left(\frac{\alpha_m}{m}\right)^{\frac{m}{\alpha_m}} e^{\frac{m}{\alpha_m}}\right)$$

satisfying

$$\prod_{k\geq 0,\, \alpha_k<\alpha_m} \left|\frac{\alpha_k + \alpha_m}{\alpha_k - \alpha_m}\right| = \exp(\alpha_m \xi_{2,m}).$$

Now it remains to bound (3.2.15). Let $L_m := \#\{k\colon \alpha_m < \alpha_k < 2\alpha_m\}$ and observe that since $\sum_{k=0}^{+\infty} 1/\alpha_k < +\infty$, it follows that

$$\sum_{\alpha_k<\alpha_m} \frac{1}{\alpha_k} + \frac{L_m}{2\alpha_m} + \sum_{\alpha_k\geq 2\alpha_m} \frac{1}{\alpha_k} \leq \sum_{\alpha_k<\alpha_m} \frac{1}{\alpha_k} + \sum_{\alpha_m<\alpha_k<2\alpha_m} \frac{1}{\alpha_k} + \sum_{\alpha_k\geq 2\alpha_m} \frac{1}{\alpha_k}$$
$$\leq \sum_{k=0}^{+\infty} \frac{1}{\alpha_k} < +\infty,$$

which implies that $L_m/\alpha_m \to 0$ as $m \to +\infty$ (since $\alpha_m \to +\infty$ as $m \to +\infty$).

Hence, arguing analogously as when we bound (3.2.14), we have that

$$1 \leq \left(\prod_{k\geq 0,\, \alpha_m<\alpha_k<2\alpha_m} \left| \frac{\alpha_k+\alpha_m}{\alpha_k-\alpha_m} \right| \right)^{\frac{1}{\alpha_m}} \leq \left(\frac{3}{r}\right)^{\frac{L_m-m}{\alpha_m}} \left(\frac{\alpha_m}{L_m-m} \right)^{\frac{L_m-m}{\alpha_m}} e^{\frac{L_m-m}{\alpha_m}}.$$

Thus, as before, this implies that there exist a constant $\xi_{3,m} > 0$ such that

$$\lim_{m\to+\infty} \xi_{3,m} = 0 \quad \text{and} \quad \prod_{k\geq 0,\, \alpha_m<\alpha_k<2\alpha_m} \left| \frac{\alpha_k+\alpha_m}{\alpha_k-\alpha_m} \right| = \exp\left(\alpha_m \xi_{3,m}\right).$$

Finally, consider $2\alpha_m \geq 1$ and take $\xi_{4,m} = \log(2\alpha_m)/\alpha_m \to 0$ as $m \to +\infty$. Therefore,

$$1 \leq 2\alpha_m = \exp\left(\log(2\alpha_m)\right) = \exp\left(\alpha_m \xi_{4,m}\right).$$

Thus, taking $\gamma_m = \xi_{1,m} + \xi_{2,m} + \xi_{3,m} + \xi_{4,m}$ we get that there exists a sequence of constants $\{\gamma_m\}_{m=0}^{+\infty}$ such that

$$2\alpha_m \prod_{k\geq 0,\, k\neq m} \left| \frac{\alpha_k+\alpha_m}{\alpha_k-\alpha_m} \right| = \exp(\alpha_m\gamma_m), \qquad \lim_{m\to+\infty} \gamma_m = 0.$$

Hence,

$$\frac{1}{2\alpha_m} \prod_{k\geq 0,\, k\neq m} \left| \frac{\alpha_k-\alpha_m}{\alpha_k+\alpha_m} \right| = \exp(-\alpha_m\gamma_m), \qquad \lim_{m\to+\infty} \gamma_m = 0,$$

and taking into consideration the formula (3.2.12) together with $\alpha_k = \lambda_k + 1/2$ for every $k \geq 0$, we obtain that

$$\left\| x^{\lambda_m} - p(x) \right\|_{L^2([0,1])} \geq \exp(-(\lambda_m + 1/2)\gamma_m)$$

where $\lim_{m\to+\infty} \gamma_m = 0$ and $p \in \Pi(\Lambda \setminus \{\lambda_m\})$.

This clearly implies that for every function $p = \sum_{k=0}^{n} a_k x^{\lambda_k} \in \Pi(\Lambda)$, the inequality

$$\|p\|_{L^2([0,1])} = \left\| a_m x^{\lambda_m} - (a_m x^{\lambda_m} - p(x)) \right\|_{L^2[(0,1)]} \geq |a_m| \exp(-(\lambda_m + 1/2)\gamma_m)$$
$$= |a_m| \exp(-\lambda_m\gamma_m) \exp(-\gamma_m/2)$$

holds for every $0 \leq m \leq n$. Hence, for a given $\varepsilon > 0$ there is an $m := m(\varepsilon) \in \mathbb{N}$ such that $\exp(\gamma_k) < 1 + \varepsilon$ for any $k \geq m$. So take

$$s_k = \begin{cases} \gamma_k \lambda_k, & \text{if } 1 \leq k < m, \\ 0, & \text{if } k \geq m, \end{cases}$$

and define $c_{k,\varepsilon} := \exp(s_k)$. Then it holds that

$$|a_m| \leq \exp(\gamma_m/2)\exp(\gamma_m)^{\lambda_m} \|p\|_{L^2[(0,1)]} \leq (1+\varepsilon)c_{m,\varepsilon}(1+\varepsilon)^{\lambda_m} \|p\|_{L^2[(0,1)]}. \tag{3.2.19}$$

Observe that $c_{m,\varepsilon}$ just depends on ε and Λ. Taking into consideration that $\lambda_1 \geq 1$ and $\inf_k(\lambda_k - \lambda_{k-1}) > 0$, we have that there exists a strictly increasing sequence of natural numbers m_k, $k \geq 0$, such that $\{[\lambda_k]\}_{k=0}^{+\infty} = \{m_k\}_{k=0}^{+\infty}$, where $[\lambda_k]$ denotes the integer part of λ_k for each $k \in \mathbb{N}$, and furthermore,

$$M := M(\Lambda) := \max_{j\geq 0} \#\{k\colon\ [\lambda_k] = m_j\} < +\infty.$$

Thus, for all $\alpha \in (0,1)$,

$$\begin{aligned}
\sum_{k=0}^{n} \lambda_k \alpha^{\lambda_k - 1} &\leq \sum_{k=0}^{n} ([\lambda_k]+1)\alpha^{[\lambda_k]-1} = \sum_{k=0}^{n} [\lambda_k]\alpha^{[\lambda_k]-1} + \sum_{k=0}^{n} \alpha^{[\lambda_k]-1} \\
&\leq M\left(\sum_{k=0}^{n} m_k \alpha^{m_k-1} + \sum_{k=0}^{n} \alpha^{m_k-1}\right) \leq M\left(\sum_{k=0}^{+\infty} m_k \alpha^{m_k-1} + \sum_{k=0}^{+\infty} \alpha^{m_k-1}\right) \\
&\leq M\left(\sum_{k=0}^{+\infty} k\alpha^{k-1} + \sum_{k=0}^{+\infty} \alpha^{k-1}\right) = \frac{M}{(1-\alpha)^2\alpha} := C(\alpha, M) < +\infty.
\end{aligned} \tag{3.2.20}$$

We can use (3.2.19) and (3.2.20) to estimate the norm $\|p'\|_{[0,1-\varepsilon]}$ as follows,

$$\begin{aligned}
\|p'\|_{[0,1-\varepsilon]} &\leq \sum_{k=0}^{n} \lambda_k |a_k| (1-\varepsilon)^{\lambda_k-1} \leq \sum_{k=0}^{n} \lambda_k c_{k,\varepsilon}(1+\varepsilon)^{\lambda_k+1}(1-\varepsilon)^{\lambda_k-1} \|p\|_{L^2([0,1])} \\
&= \left(\max_{k\geq 0} c_{k,\varepsilon}\right)(1+\varepsilon)^2 \|p\|_{L^2([0,1])} \sum_{k=0}^{n} \lambda_k (1-\varepsilon^2)^{\lambda_k-1} \\
&\leq \left(\max_{k\geq 0} c_{k,\varepsilon}\right)(1+\varepsilon)^2 \|p\|_{L^2([0,1])} C\left(1-\varepsilon^2, M\right) \\
&\leq c(\varepsilon,\Lambda) \|p\|_{\mathcal{C}([0,1])},
\end{aligned}$$

where $c(\varepsilon,\Lambda) = (\max_{k\geq 0} c_{k,\varepsilon})(1+\varepsilon)^2 C(1-\varepsilon^2, M)$, which is what we wanted to prove.

□

The next theorem, proved by Laurent Schwartz (see [13] and [9]), is an important consequence of the inequality (3.2.19).

Lemma 3.2.10. *Let us assume that $\Lambda = \{\lambda_k\}_{k=0}^{+\infty}$ is an increasing sequence of nonnegative real numbers such that $\inf_{k\in\mathbb{N}}(\lambda_k - \lambda_{k-1}) > 0$ and $\sum_{k=1}^{+\infty} 1/\lambda_k < +\infty$, $\lambda_0 = 0$, $\lambda_1 \geq 1$. Then the functions that belong to the closure of $\Pi(\Lambda)$ can be analytically extended to $\mathbb{D} \setminus [-1,0]$.*

Proof. Let f in the closure of $\Pi(\Lambda)$. Assume that $\|f\|_\infty = 1$ (otherwise, take $f/\|f\|_\infty$). Then, there exists $(q_n)_{n=0}^{+\infty} \subseteq \Pi(\Lambda)$ such that

$$\|f - q_n\|_\infty < \frac{1}{n} \quad \text{and} \quad q_n = \sum_{k=0}^{s_n} a_{n,k} x^{\lambda_k}.$$

Then the sequence of functions $(q_n)_{n=0}^{+\infty}$ is a Cauchy sequence in $\mathcal{C}([0,1])$. It follows that for each $\delta > 0$ and all $n, m \in \mathbb{N}$, applying (3.2.19)

$$\begin{aligned} |a_{n,k} - a_{m,k}| &\leq c_{k,\delta}(1+\delta)^{\lambda_k} \|q_n - q_m\|_\infty \\ &\leq c_{k,\delta}(1+\delta)^{\lambda_k} \left(\frac{1}{n} + \frac{1}{m}\right) \to 0, \qquad (n, m \to +\infty). \end{aligned}$$

This means that for every $k \in \mathbb{N}$, there are numbers $a_k \in \mathbb{R}$ such that $\lim_{n\to+\infty} a_{n,k} = a_k$. Let $h(x) := \sum_{k=0}^{+\infty} a_k x^{\lambda_k}$. Then, for all $\delta > 0$ we can write

$$|a_k| = \lim_{n\to+\infty} |a_{n,k}| \leq \lim_{n\to+\infty} c_{k,\delta}(1+\delta)^{\lambda_k} \|q_n\|_\infty = c_{k,\delta}(1+\delta)^{\lambda_k} \|f\|_\infty = c_{k,\delta}(1+\delta)^{\lambda_k}.$$

We claim that the series $h(x) = \sum_{k=0}^{+\infty} a_k x^{\lambda_k}$ is absolutely convergent for all $x < 1$. First observe that for $x = 0$ is trivial. So take $0 < x < 1$. Hence, there exists $\delta := \delta_x > 0$ such that

$$0 < \delta < \frac{1}{x} - 1 \Rightarrow (1+\delta)x < 1.$$

Now observe that there is some $k_1 \in \mathbb{N}$ such that for every $k \geq k_1$, we have that $\lambda_k/k \geq r$ for some $r \in \mathbb{R}$ (see (3.2.16)). Moreover, there is some $k_2 \in \mathbb{N}$ so that for every $k \geq k_2$, we have $c_{\delta,k} = 1$. So take $k_3 := \max\{k_1, k_2\}$. Therefore, for every $k \geq k_3$,

$$\begin{aligned} \left|a_k x^{\lambda_k}\right|^{1/k} &\leq \left(c_{k,\delta}(1+\delta)^{\lambda_k} x^{\lambda_k}\right)^{1/k} = ((1+\delta)x)^{\lambda_k/k} \\ &\leq ((1+\delta)x)^r < 1 \end{aligned}$$

holds. Thus, the series

$$\sum_{k=0}^{+\infty} |a_k x^{\lambda_k}| = \sum_{k=0}^{k_3-1} |a_k x^{\lambda_k}| + \sum_{k=k_3}^{+\infty} \left(|a_k x^{\lambda_k}|^{1/k}\right)^k \leq \sum_{k=0}^{k_3-1} |a_k x^{\lambda_k}| + \sum_{k=k_3}^{+\infty} \left(((1+\delta)x)^r\right)^k$$

converges absolutely for $x < 1$.

Now, it is clear that h coincides with the function f since

$$\|f - h\|_\infty \leq \|f - q_n\|_\infty + \|q_n - h\|_\infty \to 0 \ \ (n \to +\infty).$$

If we consider the branch of logarithm that is defined on the complex plane cut along $(-\infty, 0]$ and that is positive for values greater than 1, this defines a branch of $z^\mu = \exp(\mu \log z)$, for any μ. Now, if $z \in \mathbb{D} \setminus [-1, 0]$ then

$$\sum_{k=0}^{+\infty} \left|a_k z^{\lambda_k}\right| \leq f(|z|) < +\infty.$$

This proves that $f(z) = \sum_{k=0}^{+\infty} a_k z^{\lambda_k}$ is analytic on $\mathbb{D} \setminus [-1, 0]$.

□

We now will prove (for "special sequences") the Chebyshev's inequality which claims that the norms of the elements of non dense Müntz spaces essentially depend on the behavior of the elements near $x = 1$.

Proposition 3.2.11 (Bounded Chebyshev Inequality for Special Sequences). *Let us assume that $\Lambda = \{\lambda_k\}_{k=0}^{+\infty}$ is an increasing sequence of nonnegative real numbers that satisfies the jump condition $\inf_{k\in\mathbb{N}}(\lambda_k - \lambda_{k-1}) > 0$, and $\sum_{k=1}^{+\infty} 1/\lambda_k < +\infty$, $\lambda_0 = 0$, $\lambda_1 \geq 1$. Then for all $\varepsilon \in (0,1)$ there exists a constant $c_\varepsilon = c(\varepsilon, \Lambda) > 0$ such that $\|p\|_{[0,1]} \leq c_\varepsilon \|p\|_{[1-\varepsilon,1]}$ for all $p \in \Pi(\Lambda)$.*

Proof. Let us assume that there exists a sequence of polynomials $(p_n)_{n=0}^{+\infty} \subseteq \Pi(\Lambda)$ such that $\lim_{n\to+\infty} \|p_n\|_{[0,1]} = +\infty$ but $\|p_n\|_{[1-\varepsilon,1]} = 1$ for all n. Then, $q_n := p_n / \|p_n\|_{[0,1]}$ satisfies $\|q_n\|_{[0,1]} = 1$ for all n and $\lim_{n\to+\infty} \|q_n\|_{[1-\varepsilon,1]} = 0$.

It follows from the bounded Bernstein Inequality that for each $\delta \in (0,1)$ there exists a constant c_δ such that

$$\|q_n'\|_{[0,1-\delta]} \leq c_\delta \|q_n\|_{[0,1]} = c_\delta,$$

for all n. The mean Value Theorem implies that the family $(q_n)_{n=0}^{+\infty}$ is equicontinuous at $[0, 1-\delta]$.

Let's take $\varepsilon \in (0,1)$. We may use the Arzelá-Ascoli theorem (Theorem 2.2.45) in the interval $[0, 1-\varepsilon/2]$ to obtain from $(q_n)_{n=0}^{+\infty}$ a subsequence that converges uniformly to a certain $f \in \mathcal{C}([0, 1-\varepsilon/2])$. By Lemma 3.2.10, f can be analytically extended on $(0, 1-\varepsilon/2)$. But $\lim_{n\to+\infty} \|q_n\|_{[1-\varepsilon,1]} = 0$ implies that $f\Big|_{(1-\varepsilon,1-\varepsilon/2)} \equiv 0$, which by the Identity principle, f must be the zero function.

Therefore, $\|q_n\|_{[0,1-\varepsilon/2]}$ converges to zero as n goes to infinity for every $\varepsilon \in (0,1)$. Making ε tend to zero, contradicts the fact that $\|q_n\|_{[0,1]} = 1$.

□

Finally we see the complete Bounded Bernstein's and Chebyshev's Inequalities for general sequence, i.e, where we remove the condition gap $\inf_{k\in\mathbb{N}}(\lambda_k - \lambda_{k-1}) > 0$. This is the most important theorem in this section and it will play an important role on the proof of the reciprocal of the Full Müntz-Szász Theorem.

Proposition 3.2.12 (Bounded Bernstein's and Chebyshev's Inequalities). *Let us assume that $0 \leq \lambda_0 < \lambda_1 < \lambda_2 < \cdots$, $\lambda_0 = 0$, $\lambda_1 \geq 1$ and $\sum_{k=1}^{+\infty} 1/\lambda_k < +\infty$. Then for each $\varepsilon > 0$ there are constants $c_\varepsilon, c_\varepsilon^* > 0$ such that*

$$\|p\|_{[0,1]} \leq c_\varepsilon^* \|p\|_{[1-\varepsilon,1]} \ \text{ and } \ \|p'\|_{[0,1-\varepsilon]} \leq c_\varepsilon^* c_\varepsilon \|p\|_{[1-\varepsilon,1]} \leq c_\varepsilon^* c_\varepsilon \|p\|_{[0,1]}$$

for all $p \in \langle 1, x^{\lambda_1}, x^{\lambda_2}, \dots \rangle$.

Proof. Observe first that $\lim_{k\to+\infty} \lambda_k / k = +\infty$ (see (3.2.16)). So, there exists some $m \in \mathbb{N}$ such that $\lambda_k > 2k$ for all $k \geq m$. Fix such m and take the sequence $\Gamma := \{\gamma_k\}_{k=0}^{+\infty}$ defined by

$$\gamma_k := \begin{cases} \min\{\lambda_k, k\}, & \text{if } k \in \{0, 1, \dots, m\}, \\ \frac{1}{2}\lambda_k + k, & \text{if } k > m. \end{cases}$$

Then, $0 \leq \gamma_0 < \gamma_1 < \cdots$, and

$$\sum_{k=1}^{+\infty} \frac{1}{\gamma_k} = \sum_{k=1}^{m} \frac{1}{\gamma_k} + \sum_{k=m+1}^{+\infty} \frac{1}{\gamma_k} = \sum_{k=1}^{m} \frac{1}{\gamma_k} + 2 \sum_{k=m+1}^{+\infty} \frac{1}{2m + \lambda_k} \leq \sum_{k=1}^{m} \frac{1}{\gamma_k} + 2 \sum_{k=m+1}^{+\infty} \frac{1}{\lambda_k} < +\infty.$$

Moreover, we have that

$$\gamma_k - \gamma_{k-1} = \begin{cases} \min\{\lambda_k, k\} - \min\{\lambda_{k-1}, k-1\}, & \text{if } k \in \{0, 1, \ldots, m\}, \\ \frac{1}{2}\lambda_{m+1} + m + 1 - \min\{\lambda_m, m\}, & \text{if } k = m+1, \\ \frac{1}{2}(\lambda_k - \lambda_{k-1}) + 1, & \text{if } k > m+1. \end{cases}$$

So observe that if $k > m + 1$, then

$$\gamma_k - \gamma_{k-1} = \frac{1}{2}(\lambda_k - \lambda_{k-1}) + 1 \geq 1,$$

and if $k = m + 1$, since $\lambda_m > 2m$,

$$\gamma_{m+1} - \gamma_m = \frac{1}{2}\lambda_{m+1} + m + 1 - \min\{\lambda_m, m\} = \frac{1}{2}\lambda_{m+1} + 1 \geq 1.$$

Finally, if $k \leq m$,

$$\begin{aligned} \gamma_k - \gamma_{k-1} &= \min\{\lambda_k, k\} - \min\{\lambda_{k-1}, k-1\} \\ &= \begin{cases} k - (k-1) = 1 > 0, & \text{if } k \leq \lambda_k \text{ and } k-1 \leq \lambda_{k-1}, \\ \lambda_k - (k-1) > 0, & \text{if } k \geq \lambda_k \text{ and } k-1 \leq \lambda_{k-1}, \\ k - \lambda_{k-1} \geq 1 > 0, & \text{if } k \leq \lambda_k \text{ and } k-1 \geq \lambda_{k-1}, \\ \lambda_k - \lambda_{k-1} > 0, & \text{if } k \geq \lambda_k \text{ and } k-1 \geq \lambda_{k-1}. \end{cases} \end{aligned}$$

and also we have that $\gamma_0 = 0$ and $\gamma_1 = \min\{\lambda_1, 1\} = 1$.

So, we have an increasing sequence Γ of nonnegative real numbers that satisfies the jump condition $\inf_{k\in\mathbb{N}} (\gamma_k - \gamma_{k-1}) > 0$, and $\sum_k 1/\gamma_k < +\infty$, $\gamma_0 = 0$ and $\gamma_1 \geq 1$. Therefore, by Proposition 3.2.9 and Proposition 3.2.11 it follows that for all $\varepsilon \in (0, 1)$ there exist constants $c_\varepsilon, c_\varepsilon^* > 0$ such that

$$\|p'\|_{[0,1-\varepsilon]} \leq c_\varepsilon \|p\|_{[0,1]}$$

and

$$\|p\|_{[0,1]} \leq c_\varepsilon^* \|p\|_{[1-\varepsilon,1]}$$

for all $p \in \langle 1, x^{\gamma_1}, x^{\gamma_2}, \ldots \rangle$.

Furthermore, observe that $\gamma_k \leq \lambda_k$ for all k. Thus, denoting by $\Lambda_n = \{\lambda_k\}_{k=0}^n$ and $\Gamma_n = \{\gamma_k\}_{k=0}^n$, by Proposition 3.1.26,

$$\max_{0\neq p\in M(\Lambda_n)} \frac{\|p\|_{[0,1]}}{\|p\|_{[1-\varepsilon,1]}} \leq \max_{0\neq p\in M(\Gamma_n))} \frac{\|p\|_{[0,1]}}{\|p\|_{[1-\varepsilon,1]}} \leq c_\varepsilon^*$$

and since $\lambda_1 = \gamma_1 = 1$ then

$$\max_{0\neq p\in M(\Lambda_n)} \frac{\|p'\|_{[0,1-\varepsilon]}}{\|p\|_{[1-\varepsilon,1]}} \leq \max_{0\neq p\in M(\Gamma_n)} \frac{\|p'\|_{[0,1-\varepsilon]}}{\|p\|_{[1-\varepsilon,1]}} \leq c_\varepsilon^* \max_{0\neq p\in M(\Gamma_n)} \frac{\|p'\|_{[0,1-\varepsilon]}}{\|p\|_{[0,1]}} \leq c_\varepsilon^* c_\varepsilon.$$

Therefore, for any $p \in \Pi(\Lambda)$, there is some $n \in \mathbb{N}$ such that $p \in M(\Lambda_n)$ satisfying

$$\|p\|_{[0,1]} \leq c_\varepsilon^* \|p\|_{[1-\varepsilon,1]} \quad \text{and} \quad \|p'\|_{[0,1-\varepsilon]} \leq c_\varepsilon^* c_\varepsilon \|p\|_{[1-\varepsilon,1]} \leq c_\varepsilon^* c_\varepsilon \|p\|_{[0,1]}.$$

□

3.3 Reciprocal on Full Müntz-Szász Theorem on $\mathcal{C}([0,1])$

Finally, we will show the necessary part of the Full Müntz-Szász Theorem for arbitrary sequences $\{\lambda_j\}_{j=0}^{+\infty}$ ($\lambda_0 = 0$) of nonnegative distinct real values. For simplicity, we will split up our proof in three cases:

(a) $\inf_{j\in\mathbb{N}} \lambda_j > 0$,

(b) $\lim_{j\to+\infty} \lambda_j = 0$ and

(c) $\{\lambda_j\}_{j\in\mathbb{N}} = \{\mu_j;\ \lim_j \mu_j = 0\} \cup \{\gamma_j;\ \lim_j \gamma_j = +\infty\}$.

For the first case, which is based in complex analysis, we have followed [5], but you can find equivalent proofs on, for example, [1]. For the other two cases, which are based on properties of the zeros of the Chebyshev polynomials, we have followed [7] and [18].

Theorem 3.3.1 (Full Müntz-Szász Theorem)**.** *Let $\{\lambda_j\}_{j=1}^{+\infty}$ be a sequence of different real positive numbers, and let X be the closure in $\mathcal{C}([0,1])$ of the set generated by the finite linear combinations of the functions $1, x^{\lambda_1}, x^{\lambda_2}, \dots$. If*

$$\sum_{j=1}^{+\infty} \frac{\lambda_j}{\lambda_j^2+1} < +\infty,$$

then $X \subsetneq \mathcal{C}([0,1])$.

Proof. First suppose that $\inf_j \lambda_j > 0$. Set $\lambda_0 = 0$. We will construct a linear bounded functional $T = \langle \cdot, \mu\rangle$ on $\mathcal{C}([0,1])$ such that $T(t^{\lambda_j}) = 0$ for every $j = 0, 1, 2, \dots$, but with $T(t^{\lambda}) \neq 0$, for any $\lambda > 0$ with $\lambda \notin \{\lambda_j\}_{j=0}^{+\infty}$. If we get such functional, we will be able to apply the Corollary of the Hahn-Bannach Theorem (Corollary 2.2.41), which would finish the proof.

To do so, observe that by the Riesz-Markov-Kakutani Representation Theorem (Theorem 2.2.37), it is equivalent to find a Borel complex measure μ such that

$$T(\varphi) = \int_0^1 \varphi(t)d\mu(t), \qquad (\varphi \in \mathcal{C}([0,1]))$$

satisfying

(a) $T(t^{\lambda}) = \int_0^1 t^{\lambda} d\mu(t) \neq 0$,

(b) $T(t^{\lambda_j}) = \int_0^1 t^{\lambda_j} d\mu(t) = 0$, $\forall j = 0, 1, 2, \dots$.

Hence, we can reduce our problem to find a bounded and holomorphic function f in

$$\mathbb{H}_{-1} = \{z \in \mathbb{C};\ \operatorname{Re} z > -1\}$$

for which exists a Borel complex measure μ such that

(i) $f(z) = \int_0^1 t^z d\mu(t)$, for every $z \in \mathbb{H}_{-1}$,

(ii) $f(\lambda_j) = 0$, $\forall j = 0, 1, 2, \ldots$, and

(iii) $f(\lambda) \neq 0$.

One can think of taking the infinite product of holomorphic functions $f_j(z)$ in $\mathbb{H}_{-1}$ such that $f_j(\lambda_j) = 0$ for every $j \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, and could be in the right way, but this product may not converge. So, we do something similar but choosing such f_j so that the product converge.

Thus, the function that we consider is

$$f(z) = \frac{z}{(2+z)^3} \prod_{j=1}^{+\infty} \frac{\lambda_j - z}{2 + \lambda_j + z}, \qquad z \in \mathbb{H}_{-1}.$$

Observe that each term

$$f_j(z) = \frac{\lambda_j - z}{2 + \lambda_j + z}$$

is holomorphic in $\mathbb{H}_{-1}$ and satisfies $f_j(\lambda_j) = 0$. Moreover, we have added and additional term,

$$\frac{z}{(2+z)^3},$$

which is also holomorphic in $\mathbb{H}_{-1}$ and vanishes at $z = 0$. For the last term, we could choose any other function, but we choose that one in order to guarantee an integrability property which we will see later.

So, let's begin by proving that f is holomorphic in $\mathbb{H}_{-1}$, which by Theorem 2.2.44, it is enough to see that the series with terms

$$\left| 1 - \frac{\lambda_j - z}{2 + \lambda_j + z} \right| = \left| \frac{2z + 2}{2 + \lambda_j + z} \right|$$

converges absolutely and uniformly over compact sets in the domain $\mathbb{H}_{-1}$. So, let's fix K, a compact subset of $\mathbb{H}_{-1}$ and let $C_K := \sup_{z \in K} |2z + 2|$. Observe that for every $z \in K$ it holds that

$$\begin{aligned} \left| \frac{2z+2}{2 + \lambda_j + z} \right| &\leq \frac{C_K}{\inf_{w \in K} |2 + \lambda_j + w|} \leq \frac{C_K}{\inf_{w \in \mathbb{H}_{-1}} |2 + \lambda_j + w|} \\ &\leq \frac{C_K}{\inf_{w \in \mathbb{H}_{-1}} \operatorname{Re}(2 + \lambda_j + w)} = \frac{C_K}{1 + \lambda_j} \leq \frac{C_K}{\lambda_j}. \end{aligned}$$

Thus, by the convergence of the series $\sum_j 1/\lambda_j$ and the M Weierstrass criteria, we can conclude that the series with therms

$$\frac{2z + 2}{2 + \lambda_j + z}$$

converges absolutely and uniformly over compact sets in the domain $\mathbb{H}_{-1}$.

Let's see now that f is bounded by 1 in $\mathbb{H}_{-1}$. Since each term

$$\frac{\lambda_j - z}{2 + \lambda_j + z}$$

of the infinite product is the inverse of a Möbius transformation from $\mathbb{H}_{-1}$ to $\mathbb{D}$ (see (2.2.27) and take $a = \lambda_j + 1 > 0$), then the infinite product is bounded by 1 in $\mathbb{H}_{-1}$. Moreover, the term outside the infinite product, $z/(2+z)^3$, is also bounded in $\mathbb{H}_{-1}$, since

$$\frac{z}{(2+z)^3} = \left(\frac{z}{z+2}\right) \cdot \left(\frac{1}{(2+z)^2}\right),$$

and $\frac{z}{z+2}$ is the inverse of a Möbius transformation from $\mathbb{H}_{-1}$ to $\mathbb{D}$, and

$$\left|\frac{1}{(2+z)^2}\right| \leq 1, \ \forall z \in \mathbb{H}_{-1}$$

due to $|2+z| \geq |\mathrm{Re}(2+z)| \geq 1$ in $\mathbb{H}_{-1}$.

Now, let's see that f is in L^1 when we restrict to $\mathrm{Re}(z) = -1$. Since when $z = -1 + ir$, $r \in \mathbb{R}$ we have

$$\left|\frac{\lambda_j - z}{2 + \lambda_j + z}\right| = \frac{|(\lambda_j + 1) - ir|}{|(\lambda_j + 1) + ir|} = 1,$$

then, the norm of the infinite product of f is equal to 1, hence

$$\int_{\mathbb{R}} |f(-1+ir)| \, dr = \int_{\mathbb{R}} \frac{|-1+ir|}{|1+ir|^3} dr = \int_{\mathbb{R}} \frac{1}{1+r^2} dr = \pi < +\infty,$$

and this implies that $f \in L^1(\{\mathrm{Re}(z) = -1\})$. Observe that here f would not be integrable if we had missed the term $z/(2+z)^3$.

The next step is to consider a fixed z_0, with $\mathrm{Re}(z_0) > -1$, and to apply the Cauchy's Formula (Theorem 2.2.49) to $f(z_0)$, through the semicircumference centered at $(-1, 0)$ and with radius $R > 1 + |z_0|$, taken from $-1 - iR$ to $-1 + R$, and unto $-1 + iR$, and then continued by the segment from $-1 + iR$ to $-1 - iR$, as is shown in Figure 1.

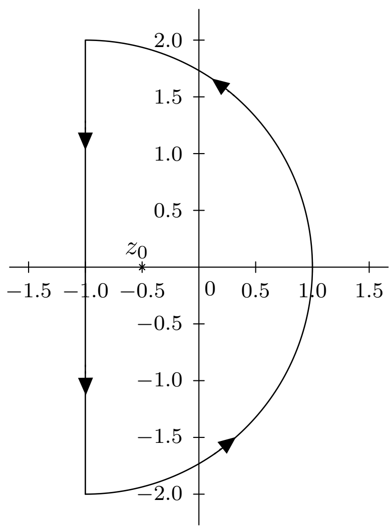

Figure 1: Smooth path where we apply the Cauchy's Formula (with R = 2 and $z_0 = -0.5$).

Calling this curve C, once it is parametrized, we have that

$$\begin{aligned} f(z_0) &= \frac{1}{2\pi i}\int_C \frac{f(z)}{z-z_0}dz \\ &= -\frac{1}{2\pi i}\int_{-R}^{R} \frac{f(-1+ir)}{-1+ir-z_0} i dr + \frac{1}{2\pi i}\int_{-\pi/2}^{\pi/2} \frac{f(-1+Re^{i\theta})}{-1+Re^{i\theta}-z_0} Rie^{i\theta} d\theta. \end{aligned} \tag{3.3.1}$$

We want to see that the second term of the integral through the semicircumference, which we will denote by I_R, tends to 0 as $R \to +\infty$. Using again the bound given by

$$|f(z)| \leq \frac{|z|}{|2+z|^3},$$

we can write

$$\begin{aligned} |I_R| &\leq \frac{R}{2\pi}\int_{-\pi/2}^{\pi/2} \frac{|-1+Re^{i\theta}|}{|1+Re^{i\theta}|^3|-1+Re^{i\theta}-z_0|} d\theta \\ &\leq \frac{R}{2} \sup_{\theta\in(-\pi/2,\pi/2)} \frac{|-1+Re^{i\theta}|}{|1+Re^{i\theta}|^3|-1+Re^{i\theta}-z_0|}. \end{aligned}$$

The triangular inequality implies that

$$|-1+Re^{i\theta}| \leq 1+R \quad \text{and} \quad |1+Re^{i\theta}| \geq R-1.$$

Moreover,

$$|-1+Re^{i\theta}-z_0| \geq R-|1+z_0| \geq R-1-|z_0| > 0,$$

so that

$$\frac{|-1+Re^{i\theta}|}{|1+Re^{i\theta}|^3|-1+Re^{i\theta}-z_0|} \leq \frac{R+1}{(R-1)^3(R-|1+z_0|)}.$$

Since z_0 is fixed, the term $R-|1+z_0|$ grows with R, so it can be bounded inferiorly by 1, if $R := R(z_0)$ is big enough. Then,

$$|I_R| \leq \frac{R+1}{(R-1)^3(R-|1+z_0|)} \longrightarrow 0,$$

when $R \to +\infty$. Applying this, and the dominated convergence theorem over the integrability of f in $\{\mathrm{Re}(z) = -1\}$, making $R \to +\infty$ in (3.3.1), we obtain

$$f(z_0) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{f(-1+ir)}{1+z_0-ir} dr. \tag{3.3.2}$$

Now observe that since $\mathrm{Re}\, z > -1$,

$$\int_0^1 t^{z-ir} dt = \left[\frac{1}{z-ir+1} t^{z-ir+1}\right]_{t=0}^{t=1} = \frac{1}{z-ir+1}.$$

Then, (3.3.2) can be written as

$$f(z_0) = \int_0^1 t^{z_0} \left[\frac{1}{2\pi}\int_{\mathbb{R}} f(-1+ir)e^{-ir\log t} dr\right] dt. \tag{3.3.3}$$

The change of the integration order is legitimate since if the integrand in (3.3.3) is replaced by its absolute value, appears a finite integral due to the fact that the restriction of f to the line $\{\mathrm{Re}(z) = -1\}$ belongs to L^1:

$$\begin{aligned}\frac{1}{2\pi}\int_0^1\int_{\mathbb{R}}\left|t^{z_0}f(-1+ir)e^{-ir\log t}\right| dr\,dt &= \frac{1}{2\pi}\int_0^1 t^{\mathrm{Re}(z_0)}\int_{\mathbb{R}}|f(-1+ir)|\,dr\,dt \\ &= \frac{M}{2\pi}\int_0^1 t^{\mathrm{Re}(z_0)}\,dt < +\infty \quad (\mathrm{Re}(z_0) > -1),\end{aligned}$$

where $M = \|f\|_{L^1(\{\mathrm{Re}(z)=-1\})}$.

Let's take $g(r) = f(-1+ir)$. Then,

$$f(z_0) = \int_0^1 t^{z_0}\left[\frac{1}{2\pi}\int_{\mathbb{R}} g(r)e^{-ir\log t}dr\right] dt = \int_0^1 t^{z_0}\hat{g}(\log t)\,dt,$$

where $\hat{g}$ denotes the Fourier transform of g. Since $g \in L^1([0,1])$, then $\hat{g}$ is bounded and continuous in $[0,1]$. Moreover, since the integral vanishes when $t = 0$, taking

$$d\mu(t) = \hat{g}(\log t)dt,$$

in $(0,1]$ (considering μ concentrated in $(0,1]$) we obtain a Borel complex measure that represents f in the desired way, i.e.,

$$f(z) = \int_I t^z d\mu(t), \qquad (\mathrm{Re}\, z > -1),$$

which, by construction, it vanishes at $z = \lambda_j$ for all $j \in \mathbb{N}\cup\{0\}$ but $f(\lambda) \neq 0$ for $\lambda \notin \{\lambda_j\}_{j=0}^{+\infty}$. Equivalently, we have a bounded functional $T = \langle\cdot,\mu\rangle$ that vanishes at t^{λ_j}, and therefore, it does on every linear combination of such powers. Thus, by Corollary 2.2.41, since $t^\lambda \notin X$ when $\lambda \neq \lambda_j$, it follows that $X \subsetneq \mathcal{C}([0,1])$.

It remains to see what happens when $\inf_j \lambda_j = 0$. Observe that in that case,

$$\{\lambda_j\}_{j=1}^{+\infty} = \{\mu_j\colon\ \lim_j \mu_j = 0\} \cup \{\gamma_j\colon\ \lim_j \gamma_j = +\infty\},$$

were we are allowing $\{\gamma_j\colon\ \lim_j \gamma_j = +\infty\} = \emptyset$.

Assume first that $\{\gamma_j\colon\ \lim_j \gamma_j = +\infty\} = \emptyset$ (i.e., $\lim_j \lambda_j = 0$). Then,

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

implies that $M := \sum_{j=1}^{+\infty}\lambda_j < +\infty$. Hence, Proposition 3.2.2 yields that

$$\|xp'(x)\|_\infty \le 11\cdot M\cdot\|p(x)\|_\infty$$

holds for all $p \in X$.

Suppose that $X = \mathcal{C}([0,1])$ and take $f(x) = (1-x)^{1/2} \in \mathcal{C}([0,1])$. Then, for every $m \in \mathbb{N}$, there exists $p_m \in X$ such that

$$\|p_m - f\|_\infty \le \frac{1}{m^2}.$$

Take $m \geq 2$. Hence for every $x \in [0,1]$,

$$|f(x)| - \frac{1}{m^2} \leq |p_m(x)| \text{ and } -\left(|f(x)| + \frac{1}{m^2}\right) \leq -|p_m(x)|.$$

So it follows that

$$\begin{aligned}\left|p_m(1-1/m^2) - p_m(1)\right| \geq & |p_m(1-1/m^2)| - |p_m(1)| \geq \left|f(1-1/m^2)\right| - 1/m^2 \\ & - \left(|f(1)| + 1/m^2\right) = 1/m - 1/m^2 - 1/m^2 \\ = & \frac{1}{m} - \frac{2}{m^2}.\end{aligned}$$

Now, by the Mean Value Theorem, for a certain $\xi \in (1-1/m^2, 1)$ we have that

$$\begin{aligned}|\xi p'_m(\xi)| = & \xi \cdot \frac{|p_m(1-1/m^2) - p_m(1)|}{1/m^2} \geq \left(1 - \frac{1}{m^2}\right) \frac{1/m - 2/m^2}{1/m^2} \\ = & \left(1 - \frac{1}{m^2}\right)(m-2) \geq \frac{m-2}{2}.\end{aligned} \tag{3.3.4}$$

Thus, for every $m \geq 2$,

$$\frac{m-2}{2} \leq \|xp'_m(x)\|_\infty \leq 11 \cdot M \cdot \|p_m(x)\|_\infty \leq 11 \cdot M \cdot \left(\|f(x)\|_\infty + \frac{1}{m^2}\right),$$

which is clearly a contradiction since $f \in \mathcal{C}([0,1])$ and the left side of (3.3.4) increases with m while the right side decreases with m.

Finally, we just have one case left. So, we will assume now that $\{\gamma_j\colon \lim_j \gamma_j = +\infty\} \neq \emptyset$. First observe that

$$\sum_{j=1}^{+\infty} \frac{\lambda_j}{\lambda_j^2+1} = \sum_{j=1}^{+\infty} \frac{\mu_j}{\mu_j^2+1} + \sum_{j=1}^{+\infty} \frac{\gamma_j}{\gamma_j^2+1} < +\infty$$

is equivalent to

$$\sum_{j=1}^{+\infty} \mu_j < +\infty \text{ and } \sum_{j=1}^{+\infty} \frac{1}{\gamma_j} < +\infty.$$

Now, observe that we can assume that $\mu_j \leq 1$ and $\gamma_j \geq 1$. So take $n \in \mathbb{N}$ and relabel $\mu_1, \dots, \mu_n$ and $\gamma_1, \dots, \gamma_n$ such that

$$\mu_1 < \mu_2 < \cdots < \mu_n < \gamma_1 < \cdots < \gamma_{n-1} < \gamma_n.$$

Hence, by Proposition 3.1.19, $(1, x^{\mu_k})_{k=1}^n$, $(1, x^{\gamma_k})_{k=1}^n$ and $(1, x^{\mu_1}, \dots, x^{\mu_n}, x^{\gamma_1}, \dots, x^{\gamma_n})$ are Chebyshev Systems. Let us use the following notation:

- $T_{n,\mu}$ denotes the Chebyshev polynomial associated to the system $(1, x^{\mu_k})_{k=1}^n$,
- $T_{n,\gamma}$ denotes the Chebyshev polynomial associated to the system $(1, x^{\gamma_k})_{k=1}^n$, and

- $T_{2n,\mu,\gamma}$ denotes the Chebyshev polynomial associated to the system

$$(1, x^{\mu_1}, \dots, x^{\mu_n}, x^{\gamma_1}, \dots, x^{\gamma_n}).$$

It follows from Newman's inequality (Proposition 3.2.2) that

$$\left\| x T'_{n,\mu}(x) \right\|_\infty \leq 11M(\{\mu_k\}) \left\| T_{n,\mu} \right\|_\infty = 11M(\{\mu_k\}) < +\infty.$$

Now observe that if $x_i < x_{i+1}$ are two consecutive alternation points of $T_{n,\mu}$, for $i \in \{0, \dots, n-1\}$, by the Mean Value Theorem, there exists $\xi \in (x_i, x_{i+1})$ such that

$$\xi \left| \frac{T_{n,\mu}(x_{i+1}) - T_{n,\mu}(x_i)}{x_{i+1} - x_i} \right| = |\xi T'_{n,\mu}(\xi)| \leq 11M(\{\mu_k\}).$$

Therefore, since $|T_{n,\mu}(x_i)| = |T_{n,\mu}(x_{i+1})| = 1$ and $\operatorname{sign}(T(x_i)) = -\operatorname{sign}(T(x_{i+1}))$,

$$\frac{2x_i}{x_{i+1} - x_i} = x_i \left(\frac{2}{x_{i+1} - x_i} \right) \leq \xi \left| \frac{T_{n,\mu}(x_{i+1}) - T_{n,\mu}(x_i)}{x_{i+1} - x_i} \right| \leq 11M(\{\mu_k\})$$

which implies

$$(2 + 11M(\{\mu_k\}))\, x_i \leq 11M(\{\mu_k\}) x_{i+1} \Rightarrow x_i \leq \left(\frac{11M(\{\mu_k\})}{2 + 11M(\{\mu_k\})} \right) x_{i+1}.$$

Iterating and using that $x_i \leq 1$ for every $i = 0, \dots, n$, we obtain that

$$\begin{aligned} x_i \leq \left(\frac{11M(\{\mu_k\})}{2 + 11M(\{\mu_k\})} \right) x_{i+1} \leq \dots &\leq \left(\frac{11M(\{\mu_k\})}{2 + 11M(\{\mu_k\})} \right)^{n-i} x_n \\ &\leq \left(\frac{11M(\{\mu_k\})}{2 + 11M(\{\mu_k\})} \right)^{n-i}. \end{aligned}$$

Moreover, observe that

$$\left(\frac{11M(\{\mu_k\})}{2 + 11M(\{\mu_k\})} \right)^m \to 0 \quad (m \to +\infty).$$

Hence, for a given $\varepsilon > 0$, take

$$N := \min \left\{ m \in \mathbb{N}\colon \ \left(\frac{11M(\{\mu_k\})}{2 + 11M(\{\mu_k\})} \right)^m < \varepsilon \right\}.$$

Then,

$$\left(\frac{11M(\{\mu_k\})}{2 + 11M(\{\mu_k\})} \right)^N < \varepsilon$$

which means that in the interval $[\varepsilon, 1]$ there are at most N alternation points of $T_{n,\mu}$.

Now, recall that the zeros of a Chebyshev polynomial lie between two consecutive alternation points of it. So for such given $\varepsilon > 0$, there exists a constant $k_1(\varepsilon) := N - 1$ which only depends on ε and $M = M(\{\mu_k\})$ (but does not depend on n) such that $T_{n,\mu}$ has at most $k_1(\varepsilon)$ distinct zeros in $[\varepsilon, 1]$ and at least $n - k_1(\varepsilon)$ distinct zeros in $[0, \varepsilon)$.

On the other hand, it follows from the bounded Bernstein's inequality (Proposition 3.2.12) applied to $T_{n,\gamma}$ that for a given $\varepsilon > 0$,

$$\left\|T'_{n,\gamma}\right\|_{[0,1-\varepsilon]} \leq c^*_\varepsilon \left\|T_{n,\gamma}\right\|_\infty = c^*_\varepsilon < +\infty.$$

Then, observe that if $y_{i-1} < y_i$ are two consecutive alternation points of $T_{n,\gamma}$ in $[0, 1-\varepsilon]$, for $i \in \{1, \ldots, n\}$, by the Mean Value Theorem there exists $\xi \in (y_{i-1}, y_i)$ such that

$$\left|\frac{T_{n,\gamma}(y_i) - T_{n,\gamma}(y_{i-1})}{y_i - y_{i-1}}\right| = |T'_{n,\gamma}(\xi)| \leq c^*_\varepsilon.$$

Hence, since $|T_{n,\gamma}(y_i) - T_{n,\gamma}(y_{i-1})| = 2$, it follows that

$$\frac{2}{c^*_\varepsilon} \leq y_i - y_{i-1} \Rightarrow y_i \geq y_{i-1} + \frac{2}{c^*_\varepsilon}.$$

Iterating, we get that

$$y_i \geq y_0 + i\left(\frac{2}{c^*_\varepsilon}\right) \geq i\left(\frac{2}{c^*_\varepsilon}\right).$$

So, take

$$N' := \min\left\{m \in \mathbb{N}\colon\ m\left(\frac{2}{c^*_\varepsilon}\right) > 1 - \varepsilon\right\}.$$

Therefore, there are at most N' alternation points of $T_{n,\gamma}$ in $[0, 1-\varepsilon]$. Thus, arguing as above, we have that $T_{n,\gamma}$ has at most $k_2(\varepsilon) := N' - 1$ zeros in $[0, 1-\varepsilon)$ and at least $n - k_2(\varepsilon)$ zeros in $[1-\varepsilon, 1]$ (where $k_2(\varepsilon)$ only depends on ε and $M = M(\{\mu_k\})$, but does not depend on n).

Now, if we take into account the fact that the system $(1, x^{\mu_k}, x^{\gamma_k})_{k=1}^n$ is an extension of both systems $(1, x^{\mu_k})_{k=1}^n$ and $(1, x^{\gamma_k})_{k=1}^n$, it follows by the interlacing properties of the zeros of Chebyshev polynomials (Proposition 3.1.13) that between two consecutive zeros of $T_{n,\mu}$ there is at least a zero of $T_{2n,\mu,\gamma}$ (and the same for $T_{n,\gamma}$).

Therefore, $T_{2n,\mu,\gamma}$ has at least $n - k_1(\varepsilon) - 1$ zeros on $[0, \varepsilon)$ and at least $n - k_2(\varepsilon) - 1$ zeros on $[1-\varepsilon, 1]$. Hence, we conclude that there exists a certain constant $k := k(\varepsilon) = k_1(\varepsilon) + k_2(\varepsilon) + 2$ (which only depends on the sequence $\{\lambda_k\}_{k=1}^{+\infty}$ and ε) such that $T_{2n,\mu,\gamma}$ has at most

$$2n - (n - k_1(\varepsilon) - 1) - (n - k_2(\varepsilon) - 1) = k(\varepsilon)$$

zeros in the interval $(\varepsilon, 1-\varepsilon)$.

Set $k = k(1/4)$ and let us take a set of points

$$1/4 < t_0 < t_1 < \cdots < t_{k+3} < 3/4$$

and a function $f \in \mathcal{C}([0,1])$ such that $f(x) = 0$ for all $x \in [0, 1/4] \cup [3/4, 1]$ and $f(t_i) = (-1)^i 2$ for all $0 \leq i \leq k+3$ (see Figure 2).

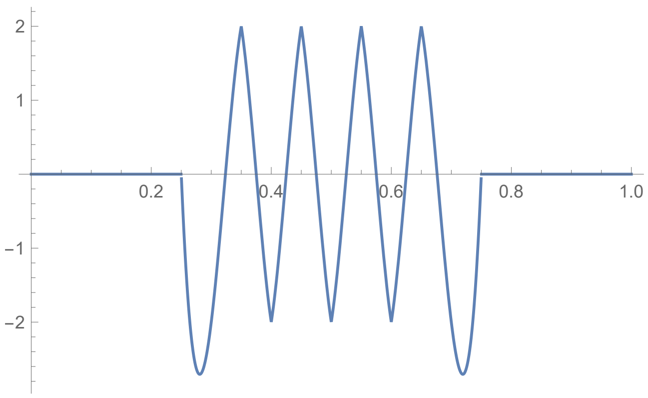

Figure 2: Graphic of a continuous function f in $[0, 1]$
(with $k = 3$).

Let us assume that $X = \mathcal{C}([0, 1])$. Then there exists $p \in \langle 1, x^{\lambda_1}, x^{\lambda_2}, \dots \rangle$ of degree n such that $\|f - p\|_\infty < 1$. We claim that $p - T_{2n,\mu,\gamma}$ has at least $2n + 1$ zeros in the interval $[0, 1]$.

First observe that since $f(x) = 0$ for all $x \in [0, 1/4] \cup [3/4, 1]$, then $|p(x)| < 1$ in such intervals. Moreover, $T_{2n,\mu,\gamma}$ has at least $(n - k_1(\varepsilon) - 1) + (n - k_2(\varepsilon) - 1) = 2n - k$ zeros in $(0, 1/4] \cup [3/4, 1)$, so it follows that it has at least $2n - k$ alternation points in such intervals (since the zeros of a Chebyshev polynomial lie between two consecutive alternation points). Recall that in the alternation points $T_{2n,\mu,\gamma}$ takes the values 1 and -1. Since $|p(x)| < 1$ in $(0, 1/4] \cup [3/4, 1)$, we deduce that $p - T_{2n,\mu,\gamma}$ has at least $2n - k - 2$ zeros in $(0, 1/4] \cup [3/4, 1)$, each one located on the alternation points of $T_{2n,\mu,\gamma}$ that lies in $(0, 1/4] \cup [3/4, 1)$.

Moreover, we have that $f(t_i) = (-1)^i 2$ for all $0 \le i \le k + 3$. Hence,

$$|p(t_i)| = 2 > 1 = \|T_{2n,\mu,\gamma}\|_\infty$$

for every $0 \le i \le k + 3$ and $\operatorname{sign}(p(t_i)) = -\operatorname{sign}(p(t_{i+1}))$. Therefore, since

$$|T_{2n,\mu,\gamma}(t_i)| \le 1 < |p(t_i)|,$$

in $(1/4, 3/4)$ the function $p - T_{2n,\mu,\gamma}$ changes its sign at least $k + 3$ times (since there are $k + 4$ points t_i). Thus, $p - T_{2n,\mu,\gamma}$ has at least $k + 3$ zeros in $(1/4, 3/4)$.

If we put all together, we get that $p - T_{2n,\mu,\gamma}$ has at least $2n - k - 2 + (k + 3) = 2n + 1$ zeros in $[0, 1]$. This is a contradiction to the fact that $p - T_{2n,\mu,\gamma} \in \langle 1, x^{\lambda_1}, x^{\lambda_2}, \dots, x^{\lambda_{2n}} \rangle$ for all n large enough, since then necessary $p - T_{2n,\mu,\gamma}$ has at most $2n$ zeros. This ends the proof.

□

4 EXTENSIONS ON MÜNTZ-SZÁSZ APPROXIMATION THEOREM

It is well known that the space $\mathcal{C}([0,1])$ of the real valued continuous functions defined on $[0,1]$ is dense on the Lebesgue space $L^p([0,1])$ for every $1 \le p \le +\infty$. Therefore, one could be tempted to ask what would happen if we take $L^p([0,1])$ instead of $\mathcal{C}([0,1])$ on the hypothesis of the Full Müntz-Szász Theorem.

Certainly, the theorem turns out to be true when $1 \le p < +\infty$. Moreover, if we take the series

$$\sum_{j=1}^{+\infty} \frac{\lambda_j + 1/p}{(\lambda_j + 1/p)^2 + 1}$$

instead of

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

for sequences $\{\lambda_j\}_{j=1}^{+\infty}$ of distinct real numbers greater than $-1/p$, we also can get a new version of the Full Müntz-Szász Theorem for the Lebesgue spaces.

On this chapter, we study the Full Müntz-Szász Theorem in $L^p([0,1])$ for $1 \le p < +\infty$. However, there are results that extends the theorem for $0 < p < 1$, but we will not pursue further on this way. For more details, see [6].

4.1 Density on the Full Müntz-Szász Theorem on $L^p([0,1])$

As with the continuous function space section, we first deal with the case when the density holds. If we try to extend the Müntz-Szász Theorem to the Lebesgue spaces $L^p([0,1])$, with $1 \le p \le +\infty$, we realize that the case $p = +\infty$ is not true. Indeed, we have already seen that

$$\overline{\langle 1, x^{\lambda_1}, x^{\lambda_2}, \dots \rangle}^{L^\infty([0,1])} = \mathcal{C}([0,1]) \subsetneq L^\infty([0,1]).$$

However, when $1 \le p < +\infty$, it is possible to obtain an analogous result of approximation. First, let's see what happen if we work with the series $\sum_{j=1}^{+\infty} \lambda_j/(\lambda_j^2+1)$ for sequences $\{\lambda_j\}_{j=1}^{+\infty}$ of distinct positive real numbers.

Theorem 4.1.1. *Let $\{\lambda_j\}_{j=1}^{+\infty}$ be a sequence of distinct positive real numbers and X the closure in $L^p([0,1])$, with $1 \le p < +\infty$, of the set generated by the finite linear combinations of the functions*

$$1, x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \dots.$$

If

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

then $X = L^p([0,1])$.

Proof. Let's call $A = \langle 1, x^{\lambda_1}, x^{\lambda_2}, \dots \rangle \subset L^p([0,1])$. So, trivially

$$X = \overline{A}^{L^p([0,1])} \subset L^p([0,1]).$$

Now, remember that the uniform convergence is stronger than the convergence in $L^p([0,1])$, i.e., $\|f\|_{L^p([0,1])} \leq \|f\|_{L^\infty([0,1])}$ for every $f \in L^\infty([0,1])$. Then, the opposite inclusion is obtained by the Full Müntz-Szász Theorem for continuous functions (Theorem 2.3.5) and the density of $\mathcal{C}([0,1])$ in $L^p([0,1])$, since

$$L^p([0,1]) = \overline{\mathcal{C}([0,1])}^{L^p([0,1])} = \overline{\overline{A}^{L^\infty([0,1])}}^{L^p([0,1])} \subset \overline{\overline{A}^{L^p([0,1])}}^{L^p([0,1])} = X.$$

□

Observe that Lemma 4.1.1 gives an equivalent result for the Full Müntz-Szász Theorem for $\mathcal{C}([0,1])$. However, as we have introduced before, we can get an extension of the Müntz-Szász Theorem in $L^p([0,1])$ for $1 \leq p < +\infty$, which will allow us to work with sequences of distinct real numbers $\{\lambda_j\}_{j=1}^{+\infty}$ greater than $-1/p$. Before show it, we need some previous results. First, let's see a result about density in $L^p([0,1])$ by continuous functions vanishing at 0.

Proposition 4.1.2. *Let $\mathcal{F}_0([0,1]) = \{f \in \mathcal{C}[0,1]$ such that $f(0) = 0\}$. Then,*

$$\overline{\mathcal{F}_0([0,1])}^{L^p([0,1])} = L^p([0,1]).$$

Proof. Let $g \in \mathcal{C}([0,1])$, and take $g_n \in \mathcal{F}_0([0,1])$ defined by

$$g_n(x) = \begin{cases} nxg(1/n), & x \in [0,1/n], \\ g(x), & x \in [1/n,1]. \end{cases} \tag{4.1.1}$$

Then,

$$\begin{aligned} \|g_n - g\|_p^p &= \int_0^1 |g_n(x) - g(x)|^p dx = \int_0^{1/n} |nxg(1/n) - g(x)|^p dx \\ &\leq \int_0^{1/n} (|nxg(1/n)| + |g(x)|)^p \, dx \leq \int_0^{1/n} (|g(1/n)| + |g(x)|)^p dx \\ &\leq (2\,\|g\|_\infty)^p \int_0^{1/n} dx \to 0, \end{aligned}$$

as n tends to $+\infty$.

Now, let $h \in L_p([0,1])$ and $\varepsilon > 0$. Since $\overline{\mathcal{C}([0,1])}^{L^p([0,1])} = L^p([0,1])$, there is a continuous function $h_\epsilon \in \mathcal{C}([0,1])$ such that $\|h_\epsilon - h\|_p < \varepsilon/2$. Moreover, there exists an $n \in \mathbb{N}$ such that $\|h_\epsilon - h_{\varepsilon,n}\|_p < \varepsilon/2$, where $h_{\varepsilon,n} \in \mathcal{F}_0([0,1])$ defined as (4.1.1), but now for the continuous function h_ϵ. Thus,

$$\|h_{\varepsilon,n} - h\|_p \leq \|h_{\varepsilon,n} - h_\varepsilon\|_p + \|h_\varepsilon - h\|_p < \varepsilon.$$

□

Now, we see a result which claims that the vectorial space $\langle x^{\lambda_1}, x^{\lambda_2}, \dots \rangle$ is dense on $L^p([0,1])$ for $1 \le p < +\infty$ when the sequence $\{\lambda_j\}_{j=1}^{+\infty}$ of distinct real numbers greater that $-1/p$ converges to some real number $-(1/p) + \alpha$, where $0 < \alpha \le 2$. Observe that in this case, clearly

$$\sum_{j=1}^{+\infty} \frac{\lambda_j + 1/p}{(\lambda_j + 1/p)^2 + 1} = +\infty.$$

Proposition 4.1.3. *Let $p \in [1, +\infty)$. Suppose $\{\lambda_j\}_{j=1}^{+\infty}$ is a sequence of distinct real numbers greater than $-(1/p)$ tending to $-(1/p) + \alpha$, where $0 < \alpha \le 2$. Then $\langle x^{\lambda_1}, x^{\lambda_2}, \dots \rangle$ is dense in $L^p([0,1])$.*

Proof. Let $\mu_j = \lambda_j + \frac{1}{p} - \frac{\alpha}{2}$. Then, μ_j is a sequence of distinct real numbers greater than $-(\alpha/2)$ tending to $\alpha/2$. Hence, there exists $j_0 \in \mathbb{N}$ such that for every $j \ge j_0$, $|\mu_j - \alpha/2| < \alpha/4$, and then $\mu_j > \alpha/4 > 0$ for every $j \ge j_0$. Moreover, since $\mu_j \to \alpha/2 > 0$, it follows that

$$\sum_{j=j_0}^{+\infty} \frac{\mu_j}{\mu_j^2 + 1} = +\infty,$$

and the Full Müntz-Szász Theorem in $\mathcal{C}([0,1])$ (Theorem 2.3.5) implies that $\langle 1, x^{\mu_{j_0}}, x^{\mu_{j_0+1}}, \dots \rangle$ is dense in $\mathcal{C}([0,1])$.

Now, let $m \in \mathbb{N}_{\ge 1}$, and observe that since

$$m + \frac{1}{p} - \frac{\alpha}{2} \ge 1 + \frac{1}{p} - \frac{\alpha}{2} \ge \frac{1}{p} > 0,$$

then $x^{m+1/p-\alpha/2} \in \mathcal{C}([0,1])$. Hence, since $x^{m+1/p-\alpha/2}$ vanishes at 0, fixed $\varepsilon > 0$ there exists $Q_\varepsilon \in \langle x^{\mu_{j_0}}, x^{\mu_{j_0+1}}, \dots \rangle$ such that

$$\left\| x^{m+1/p-\alpha/2} - Q_\varepsilon \right\|_\infty < \varepsilon.$$

Let

$$R_\varepsilon = x^{-(1/p)+\alpha/2} Q_\varepsilon(x) \in \langle x^{\lambda_{j_0}}, x^{\lambda_{j_0+1}}, \dots \rangle,$$

then we have the inequality

$$\begin{aligned}
\int_0^1 |x^m - R_\varepsilon(x)|^p \, dx &= \int_0^1 \left| x^{-(1/p)+\alpha/2} \left(x^{m+(1/p)-\alpha/2} - Q_\varepsilon(x) \right) \right|^p dx \\
&\le \left(\int_0^1 x^{-1+p(\alpha/2)} dx \right) \left\| x^{m-(\alpha/2)+(1/p)} - Q_\varepsilon \right\|_\infty^p \\
&\le \frac{\varepsilon^p}{p(\alpha/2)}.
\end{aligned}$$

Hence, the monomials x^m are in the $L^p([0,1])$ closure of $\langle x^{\lambda_{j_0}}, x^{\lambda_{j_0+1}}, \dots \rangle$ for all $m \ge 1$. Therefore, by the Weierstrass Approximation Theorem and using that the the uniform convergence is stronger than the convergence in $L^p([0,1])$, we have that $\langle x^{\lambda_{j_0}}, x^{\lambda_{j_0+1}}, \dots \rangle$ is dense in $\mathcal{F}_0([0,1])$ with the $L^p([0,1])$.

Thus, by Proposition 4.1.2, in $L^p([0,1])$ we have that

$$\overline{\mathcal{F}_0([0,1])}^{L^p([0,1])} = L^p([0,1]),$$

and since

$$\langle x^{\lambda_{j_0}}, x^{\lambda_{j_0+1}}, \dots \rangle \subseteq \langle x^{\lambda_1}, x^{\lambda_2}, \dots \rangle \subseteq L^p([0,1])$$

the theorem follows.

□

Finally, before seeing the extension of the Müntz-Szász Theorem for Lebesgue spaces with $p \in [1,\infty)$, we present a last result about the $L^p([0,1])$ functions.

Proposition 4.1.4. *Suppose $1 \leq p \leq +\infty$. Let $h \in L^p([0,1])$ such that*

$$\int_0^1 u(t)h(t)dt = 0,$$

for every $u \in \mathcal{C}([0,1])$. Then, $h(x) = 0$ a.e. x on I.

Proof. First observe that since $u \equiv 1 \in \mathcal{C}([0,1])$, then

$$\int_0^1 h(t)dt = 0.$$

Let $\alpha \in [0,1)$ and let $\delta > 0$ such that $\alpha + \delta < 1$. We define $u_{\alpha,\delta} \in \mathcal{C}([0,1])$ by

$$u_{\alpha,\delta}(x) = \frac{d\left(x, [\alpha+\delta, 1]\right)}{d\left(x, [\alpha+\delta, 1]\right) + d\left(x, [0,\alpha]\right)},$$

where $d(x,U) = \inf_{y\in U} |x-y| \in \mathcal{C}([0,1])$ for $U \subseteq I$. Then, $u_{\alpha,\delta}(x) = 0$ for $x \in [\alpha+\delta, 1]$ and $u_{\alpha,\delta}(x) = 1$ for $x \in [0,\alpha]$. Hence, $u_{\alpha,\delta} \in \mathcal{C}([0,1])$ and it follows that

$$\int_0^1 u_{\alpha,\delta}(t)h(t)dt = 0.$$

Moreover, $|u_{\alpha,\delta}| \leq 1$ and

$$\lim_{\delta\to 0} u_{\alpha,\delta}(x) = \frac{d\left(x, [\alpha, 1]\right)}{d\left(x, [\alpha, 1]\right) + d\left(x, [0,\alpha]\right)} = \begin{cases} 1 & x \in [0,\alpha] \\ 0 & x \in [\alpha, 1] \end{cases} = \chi_{[0,\alpha]}(x),$$

for every $x \in I$.

Thus, by the Dominated Convergence Theorem, for every $\alpha \in [0,1)$,

$$\int_0^\alpha h(t)dt = \int_0^1 \chi_{[0,\alpha]}(t)h(t)dt = \int_0^1 \lim_{\delta\to 0} u_{\alpha,\delta}(t)h(t)dt$$
$$= \lim_{\delta\to 0}\int_0^1 u_{\alpha,\delta}(t)h(t)dt = 0.$$

Therefore, for every $0 \leq \alpha_1 \leq \alpha_2 \leq 1$,

$$\int_{\alpha_1}^{\alpha_2} h(t)dt = \int_0^{\alpha_1} h(t)dt + \int_{\alpha_1}^{\alpha_2} h(t)dt = \int_0^{\alpha_2} h(t)dt = 0.$$

Thus, for every open set $U \subseteq I$,

$$\int_U h(t)dt = 0,$$

and it follows from Lemma 2.2.21 (b), $h(x) = 0$ a.e. x on I.

□

We are now in conditions to state and prove the extension of the Full Müntz-Szász Theorem in $L^p([0,1])$ in the dense case. The proof is based on the Riesz Representation Theorem for continuous linear functionals in $L^p([0,1])$, valid for $p \in [1,+\infty)$, so that the assumption of $p < +\infty$ is essential.

Theorem 4.1.5 (Full Müntz-Szász Theorem for Lebesgue Spaces)**.** *Let $p \in [1,+\infty)$. Suppose that $\{\lambda_j\}_{j=1}^{+\infty}$ is a sequence of distinct real numbers greater than $(-1/p)$. If*

$$\sum_{j=1}^{+\infty} \frac{\lambda_j + (1/p)}{(\lambda_j + (1/p))^2 + 1} = +\infty,$$

then the closure X of the set generated by the powers $(x^{\lambda_j})_{j=1}^{+\infty}$ is dense in $L_p([0,1])$.

Proof. Assume that $X \subsetneq L^p([0,1])$. Taking $\mu_j = \lambda_j + 1/p$ and $\gamma = 2$ in Proposition 2.3.3, and choosing a subsequence if necessary, without loss of generality we may assume that one of the following three cases occurs:

(i) **Case 1:** $\lambda_j \geq 2 - 1/p$ for each $j = 1, 2, \ldots$.

(ii) **Case 2:** $-1/p < \lambda_j < 2 - 1/p$ for each $j = 1, 2, \ldots$ and $\lim_j \lambda_j = \alpha - 1/p$ with $\alpha \in (0, 2]$.

(iii) **Case 3:** $-1/p < \lambda_j < 2 - 1/p$ for each $j = 1, 2, \ldots$ and $\lim_j \lambda_j = -1/p$.

In **Case 1**, since $\lambda_j \geq 2 - 1/p \geq 1$, we have that $\inf_j \lambda_j \geq 1 > 0$. Hence,

$$+\infty = \sum_{j=1}^{+\infty} \frac{\lambda_j + (1/p)}{(\lambda_j + (1/p))^2 + 1} \leq \sum_{j=1}^{+\infty} \frac{1}{\lambda_j + (1/p)} \leq \sum_{j=1}^{+\infty} \frac{1}{\lambda_j}.$$

Therefore, since the series $\sum_{j=1}^{+\infty} 1/\lambda_j$ and $\sum_{j=1}^{+\infty} \lambda_j/(\lambda_j^2 + 1)$ are equivalent when $\inf_j \lambda_j > 0$, in particular,

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

Thus, **Case 1** follows by Lemma 4.1.1.

In **Case 2**, Proposition 4.1.3 implies that $\langle x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \ldots \rangle$ is dense in $L^p([0,1])$.

In **Case 3**, we argue as follows. Let's call $A = \langle x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \dots \rangle$. By Corollary 2.2.41, and since the uniform convergence is stronger than the convergence in $L^p([0,1])$, we get that $X \subsetneq L^p([0,1])$ if and only if there exists a linear bounded functional T such that $T(x^{\lambda_j}) = 0$ for every $j = 1, 2, \dots$ but $T(t^\lambda) \neq 0$ for any $t^\lambda \notin X$, $\lambda > -(1/p)$.

Moreover, by Corollary 2.2.36 (Riesz-Representation Theorem for $L^p([0,1])$), we can restate it as follows: $X \subsetneq L^p([0,1])$ if and only if there exists a $0 \neq h \in L^q([0,1])$ satisfying

$$\int_0^1 t^{\lambda_j} h(t)dt = 0, \quad j = 1, 2, \dots,$$
$$\int_0^1 t^{\lambda} h(t)dt \neq 0, \quad \lambda \notin \{\lambda_j\}_{j=1}^{+\infty}, \ \lambda > -(1/p),$$

where q is the conjugate exponent of p defined by $p^{-1} + q^{-1} = 1$.

Assume there exists such $h \neq 0$. Let

$$f(z) := \int_0^1 t^z h(t)dt, \quad \mathrm{Re}(z) > -\frac{1}{p}.$$

Observe that the integrand vanishes at $t = 0$, hence, we can assume that the measure $h(t)dt$ is concentrated in $I = (0, 1]$.

Then,

$$f(z) = \int_I t^z h(t)dt = \int_I e^{z \log t} h(t)dt.$$

By the Hölder's inequality,

$$|f(z)| \leq \left(\int_0^1 t^{p\mathrm{Re}(z)} \right)^{1/p} \|h\|_q = \frac{\|h\|_q}{(p\mathrm{Re}(z) + 1)^{1/p}} < +\infty$$

since $p\,\mathrm{Re}(z) > -1$.

Moreover, using an analogous argument as used in the proof of the analicity of f in Proposition 2.3.4, we have that f is holomorpic in $\mathrm{Re}(z) > -1/p$. Now, let's define

$$g(z) := f\left(1 + z - \frac{1}{p}\right) = f \circ \tau_{1/p-1},$$

where $\tau_{1/p-1}(z) = z - 1/p + 1$ is a translation from $\mathbb{H}_{-1}$ to $\mathbb{H}_{-1/p}$. Hence, g is holomorphic in the unit disk and bounded (f is bounded), thus $g \in \mathcal{H}^\infty(\mathbb{D})$.

Now, let

$$\alpha_j = \lambda_j + \frac{1}{p} - 1 \in \left(-\frac{1}{p} + \frac{1}{p} - 1, \, 2 - \frac{1}{p} + \frac{1}{p} - 1 \right) = (-1, 1) \subseteq \mathbb{D}$$

so that

$$g(\alpha_j) = f(\alpha_j + 1 - 1/p) = f(\lambda_j) = 0,$$

for every $j \in \mathbb{N}$. Therefore, since $\lim_j \lambda_j = -1/p$, there exists $j_0 \in \mathbb{N}$ such that $|\lambda_j + 1/p| = \lambda_j + 1/p < 1$, for every $j \geq j_0$. Thus,

$$\sum_{j=1}^{+\infty} 1 - |\alpha_j| = \sum_{j=1}^{+\infty} 1 - \left| \lambda_j + \frac{1}{p} - 1 \right| = \sum_{j=1}^{j_0} \left(1 - \left| \lambda_j + \frac{1}{p} - 1 \right| \right) + \sum_{j=j_0+1}^{+\infty} \left(\lambda_j + \frac{1}{p} \right) = +\infty,$$

due to the inequality

$$+\infty = \sum_{j=j_0}^{+\infty} \frac{\lambda_j + (1/p)}{(\lambda_j + (1/p))^2 + 1} \leq \sum_{j=j_0+1}^{+\infty} \left(\lambda_j + \frac{1}{p}\right).$$

Hence, Theorem 2.2.43 yields that $g = 0$ on the open disk. Therefore, $f(z) = 0$ on the open disk with diameter $[-1/p, 2 - 1/p]$. Now observe that f is analytic on $\mathbb{H}_{-1/p}$; hence, by the Identity Principle (Theorem 2.2.46), $f = 0$ whenever $\mathrm{Re}(z) > -1/p$. So in particular,

$$f(n) = \int_0^1 t^n h(t) dt = 0, \qquad \text{for every } n = 0, 1, 2, \ldots.$$

Now the Weierstrass Approximation Theorem and the Dominated Convergence Theorem yields that

$$\int_0^1 u(t)h(t)dt = 0,$$

for every $u \in \mathcal{C}[0,1]$. By Proposition 4.1.4, this implies $h(x) = 0$ a.e. x on I, which contradicts the fact that $0 \neq h$. Thus, $X = L^p([0,1])$.

□

4.2 Reciprocal on the Full Müntz-Szász Theorem on $L^p([0,1])$

As in $\mathcal{C}([0,1])$, we also have the reciprocal of the Full Müntz-Szász Theorem for $L^p([0,1])$ when $p \in [1, +\infty)$. However, on these notes we will just see a complete proof for the case $p = 1$. However, for $p > 1$ we will show that the reciprocal also holds when the sequence $\{\lambda_j\}_{j=1}^{+\infty}$ satisfies the condition $\inf_j \lambda_j > -1/p$. For more details, see [7].

Theorem 4.2.1 (Full Müntz-Szász Theorem for integrable functions)**.** *Let $\{\lambda_j\}_{j=1}^{+\infty}$ be a sequence of distinct numbers greater that -1, and X the closure in $L^1([0,1])$ of the set generated by the finite linear combinations of the functions*

$$x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \ldots.$$

If

$$\sum_{j=1}^{+\infty} \frac{\lambda_j + 1}{(\lambda_j + 1)^2 + 1} < +\infty$$

then $X \subsetneq L^1([0,1])$.

Proof. Assume that $X = L^1([0,1])$. Let $m \in \mathbb{Z}_{>0}$ and $\varepsilon > 0$. Choose a $p \in \langle x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \ldots \rangle$ such that

$$\|x^m - p(x)\|_{L^1([0,1])} < \varepsilon.$$

Now let

$$q(x) := \int_0^x p(t)\, dt \in \langle 1, x^{\lambda_1+1}, x^{\lambda_2+1}, \ldots \rangle.$$

Then,

$$\begin{aligned}
\left\|\frac{x^{m+1}}{m+1} - q(x)\right\|_{[0,1]} &= \sup_{x\in[0,1]} \left|\frac{x^{m+1}}{m+1} - q(x)\right| \\
&\leq \sup_{x\in[0,1]} \int_0^x |t^m - q(t)|\, dt \\
&= \|x^m - p(x)\|_{L^1([0,1])} < \varepsilon.
\end{aligned}$$

So the Weierstrass Approximation Theorem yields that

$$\langle 1, x^{\lambda_1+1}, x^{\lambda_2+1}, \dots \rangle$$

is dense in in $\mathcal{C}([0,1])$.

Therefore,

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

This contradiction ends the proof.

□

For the case when $1 < p < +\infty$ we present the following result. The proof is similar as for the reciprocal of the Full Müntz-Szász Theorem (Theorem 3.3.1).

Theorem 4.2.2 (Full Müntz-Szász Theorem for Lebesgue Spaces)**.** *Let $\{\lambda_j\}_{j=1}^{+\infty}$ be a sequence of distinct real numbers greater than $-1/p$ such that $\inf_j \lambda_j > -1/p$, and X the closure in $L^p([0,1])$, with $1 < p < +\infty$, of the set generated by the finite linear combinations of the functions*

$$x^{\lambda_1}, x^{\lambda_2}, x^{\lambda_3}, \dots.$$

If

$$\sum_{j=1}^{+\infty} \frac{\lambda_j + \frac{1}{p}}{\left(\lambda_j + \frac{1}{p}\right)^2 + 1} < +\infty$$

then $X \subsetneq L^p([0,1])$.

Proof. We will construct a linear bounded functional $T = \langle \cdot, \mu \rangle$ on $L^p([0,1])$ such that $T(t^{\lambda_j}) = 0$ for every $j = 1, 2, 3, \dots$, but with $T(t^{\lambda}) \neq 0$, for any $\lambda > -1/p$ with $\lambda \notin \{\lambda_j\}_{j=1}^{+\infty}$. If we get such functional, we will be able to apply the Corollary of the Hahn-Bannach Theorem (Corollary 2.2.41), which would finish the proof.

To do so, observe that by the Riesz Representation Theorem for $L^p([0,1])$ (Theorem 2.2.36), it is equivalent to find a nonzero function $h \in L^q([0,1])$, where $1/p + 1/q = 1$, such that

$$T(\varphi) = \int_0^1 \varphi(t)h(t)dt, \quad (\varphi \in L^p([0,1]))$$

satisfying

(a) $T(t^\lambda) = \int_0^1 t^\lambda h(t)dt \neq 0$,

(b) $T(t^{\lambda_j}) = \int_0^1 t^{\lambda_j} h(t)dt = 0$, $\forall j = 1, 2, 3, \ldots$.

Recall that in the proof of the reciprocal of the Full Müntz-Szász Theorem (Theorem 3.3.1), we find a bounded and holomorphic function f in

$$\mathbb{H}_{-1} = \{z \in \mathbb{C};\ \mathrm{Re}\, z > -1\}$$

defined as

$$f(z) = \frac{z}{(2+z)^3} \prod_{j\geq 1} \frac{\lambda_j - z}{2 + \lambda_j + z}, \qquad z \in \mathbb{H}_{-1}$$

such that

(i) $f(z) = \int_0^1 t^z \hat{g}(\log t)\, dt$, for every $z \in \mathbb{H}_{-1}$, where $g(r) = f(-1 + ir) \in L^1([0,1])$,

(ii) $f(0) = 0$,

(iii) $f(\lambda_j) = 0$, $\forall j = 1, 2, 3, \ldots$, and

(iv) $f(\lambda) \neq 0$.

Observe that by (i), necessary $\hat{g}$ is bounded and continuous in $[0, 1]$. Hence, using also that the integral vanishes when $t = 0$, we can take the nonzero function

$$h(t) = \hat{g}(\log t)\chi_{\{t>0\}}(t) \in L^\infty([0,1]) \subseteq L^q([0,1]).$$

So, the functional in $L^p([0,1])^*$,

$$\begin{array}{rccl} \langle \cdot, \overline{h} \rangle : & L^p([0,1]) & \to & \mathbb{R} \\ & f & \mapsto & \langle f, \overline{h} \rangle = \int_0^1 f(t)h(t)dt, \end{array}$$

vanishes in X, but not in x^λ for every $\lambda > -1/p$, $\lambda \notin \{\lambda_j\}_j$. Thus, by Corollary 2.2.41, $X \subsetneq L^p([0,1])$.

□

5 CONCLUSIONS

When I asked María Jesús Carro to be my final master project advisor, my goal was to learn more techniques in analysis as well as to strengthen the acquired knowledges on the Master in the subject *Functional Analysis and PDE's*. So we mark ourselves the aim of studying the Müntz-Szász Theorem, which deals with increasing sequences $\{\lambda_j\}_{j=0}^{+\infty}$, of positive real numbers except for $\lambda_0 = 0$. Therefore, we start by motivating the problem with the well known Weierstrass Approximation Theorem, a particular case of the Müntz-Szász Theorem. Although our idea was to find a constructive proof of it, we came across with an interesting proof of Bernstein, which was at least different of what we had already studied. Then, we continue by introducing the main ideas of the complex measure theory where, during this process, we had to see some important results such as the *Radon-Nikodym Theorem* and *Riesz-Markovi-Kakutani Representation Theorem*. Moreover, we have recalled some classical results on complex analysis and on functional analysis, as the *Cauchy's Theorem* and the *Hahn-Banach Theorem*. Therefore, making retrospective from this point to the beginning, we could say that a great part of our goal was completed, since the proof that we have studied of the Müntz-Szász Theorem requires all this theory.

Studying on the same line, we saw an extension that holds with the same hypothesis but now with the spaces $L^p([0,1])$ $(1 \leq p < +\infty)$ instead of $\mathcal{C}([0,1])$, whose proof was based on the density of $\mathcal{C}([0,1])$ into $L^p([0,1])$ together with the Lebesgue norm and the *Riesz Representation Theorem on* $L^p([0,1])$. This stimulated us to do research in that way, so we begun to study generalizations of the Müntz-Szász Theorem involving the spaces $L^p([0,1])$, and we found a new version where the hypothesis dealt with general sequences of exponents, with no more restriction to be different between them and greater to $-1/p$. Therefore, this inspired us to seek for a Müntz-Szász Theorem in $\mathcal{C}([0,1])$ but for general sequences of exponents greater than 0. Fortunately, we succeed in the search. Moreover, we realized that we had to divide the proof of this extension in two parts: the part of the density result, which it could be easily extended to general sequences by using the complex analysis background on these notes; and the reciprocal, for which we had to introduce some vectorial subspaces of the continuous functions in order to complete the proof.

All in all, I can say that working on these notes has overcome my expectations. Not only for all the background that I have won, but also for all the research I had to do and all the bibliography I had to read and to understand. Moreover, I had the opportunity to face an important result with many different proofs and extensions but relatively new, since the most recently progress are from the beginnings of the XXI century. However, I will say that what I really regret is not having had more time in order to deepen in some aspects of these notes. As a result of everything studied, we tried to study the Muntz-Szász Theorem with another Banach space, since as one could see on these notes, the proof of the Muntz-Szász Theorem is based on the Riesz representation and the Hahn Banach. In particular, we have tried to study the case of the Lorentz spaces $L^{p,q}([0,1])$ for $1 \leq p \leq q \leq +\infty$, since we have not found any bibliography in that sense which could mean that it is an original research to do. However, even though I had the opportunity of introducing myself to these spaces and also to deepen in them, we have not found any result in that sense.

6 BIBLIOGRAPHY

References

[1] Walter Rudin, *Real and Complex Analysis*, International Student Edition, Third Edition (1970), pp. 40-47, 117-133.

[2] Walter Rudin, *Principles of mathematical analysis*, International Student Edition, Third Edition (1976), pp. 154-165.

[3] D. Estep, *Practical Analysis in One Variable*, Hardcover (2002), pp. 509-521.

[4] S.N. Bernstein, *Sur les recherches récentes relatives à la meilleure approximation des fonctions continues par les polynômes*, in Proc. of 5th International Mathematical Congress Vol. 1 (1912), pp. 256-266.

[5] Daniel Eceizabarrena, Alejandro Mas Mas, Francisco Mengual Bretón and María Soria Carro, *El Teorema de Müntz-Szász sobre la aproximación de funciones continuas*, TeMat, Vol. 1 (2017), pp. 31-44.

[6] Tamás Erdélyi and William B. Johnson, *The "Full Müntz Theorem" in* $L^p[(0,1)]$ *for* $0 < p < \infty$, Journal D'Analyse Mathématic, Vol. 84 (2001), pp. 145-172.

[7] Tamás Erdélyi and Peter Borwein, *Polynomials and Polynomials Inequalities*, Graduate Texts in Mathematics, Springer, Vol. 161 (1991), pp. 91-100, 171-218.

[8] Philip J. Davis, *Interpolation and Approximation*, Dover Publications, New York, Brown University (1975), pp. 176-184.

[9] Gurariy, Vladimir I. and Lusky, Wolfgang: *Geometry of Müntz Spaces and Related Questions*, Springer (2005), pp. 78-81.

[10] Allan Pinkus, *Density Methods and Results in Approximation Theory*, Banach Center Publications, Vol. 64 (2004).

[11] G. G. Lorentz, K. Jetter and S. D. Riemenschneider, *Encyclopedia of Mathematics and its Applications: Birkhoff Interpolation*, Longman Higher Education (1983) pp. 131-132.

[12] Philip W. Smith, *An Improvement Theorem for Descartes Systems*, American Mathematical Society, Vol. 70 (1978).

[13] Laurent Schwartz, *Étude des sommes d'exponentielles réelles*, Thèses de l'entre-deux-guerres (1943).

[14] J. A. Clarkson, P. Erdös, *Approximation by Polynomials*, Duke Math. J. 10 (1943), no. 1, pp. 5-11.

[15] Ch. H. Müntz, *Über den Approximationssatz von Weierstrass*, in H. A. Schwarz's Festschrift, Berlin, 1914, pp. 303-312.

[16] O. Szász, *Über die Approximation stetiger Funktionen durch lineare Aggregate von Potenzen*, Math.Ann. Vol. 77 (1916), pp. 482-496.

[17] Le Baron O. Ferguson and Manfred Von Golitschek, *Müntz-Szász Theorem with Integral Coefficients II*, Transactions of the American Mathematical Society, Vol. 213 (1975), pp. 115-126.

[18] J.M. Almira, *Müntz Type Theorems I*, Surveys in Approximation Theory, Vol. 3 (2007), pp. 152-194.

[19] J. D. Gray, *The shaping of the Riesz representation theorem: A chapter in the history of analysis*, Archive for History in the Exact Sciences, Vol. 31(2) (1984-85), pp. 127-187.

[20] F. Riesz, *Sur les opérations fonctionnelles linéaires*, C. R. Académie des Sciences Paris 149 (1909), pp. 974-977.