DENSITY AND APPROXIMATION PROPERTIES OF MARKOV SYSTEMS [1]

A. L. González, Universidad Nacional de Mar del Plata, Argentina
R. A. Zalik, Auburn University

In memory of Oved Shisha, a dear friend, a respected colleague, and an honest and compassionate man

[1]Author's addresses: A. L. González, Departmento de Matemática, Universidad Nacional de Mar del Plata, Funes 3350, 7600 Mar del Plata, Argentina, e–mail: algonzal@mdp.edu.ar
R. A. Zalik, Department of Mathematics, Auburn University, AL 36849–5310, e-mail: zalik@math.auburn.edu

Abstract. Borwein, Bojanov, Erdélyi, and others, have studied density and approximation properties of those Markov systems of differentiable functions on a closed interval $[a,b]$, having the property that the system of derivatives is a Markov system on the open interval (a,b).

In this paper we show that several of these results are valid for any Markov system of continuous functions on a closed interval.

Keywords: Chebychev and Markov systems. Density. Relative Differentiation.

1. Introduction

This article is organized as follows: in this section we introduce the basic definitions and some basic results from the theory of Chebychev and Markov systems. Section 2 is the key section: in it we introduce a generalized differentiation operator that we use extensively in the sequel. In section 3 we generalize the results of [2], and in sections 4 and 5 we generalize the results of [3] and [1], respectively. In sections 3 and 4 we bypass the assumption of differentiability by using the operator introduced in section 2. We also show that all the other restrictions imposed in [1, 2, 3] are unnecessary.

The theory of Chebychev systems is discussed, from different points of view, in the books [4, 5, 8, 9, 16], and in the survey papers [6, 13]. Here we mention only the basic definitions and essential properties.

Let A be a set of real numbers, let $F(A)$ denote the set of all real–valued functions defined on A, let $U_n := \{u_0, \ldots, u_n\}$ be a sequence of functions, or *system*, and let $S(U_n)$ denote the linear span of $\{u_0, \ldots, u_n\}$.

A system of functions $G_n \subset F(A)$ is called a *Chebychev system* or *T–system* if A contains at least $n+1$ points, and all the determinants of the square collocation matrices

$$\bigcup \begin{pmatrix} g_0, \cdots, g_n \\ t_0, \cdots, t_n \end{pmatrix} := \det(g_j(t_i);\ 0 \le i, j \le n)$$

with $t_0 < \ldots < t_n$ in A, are positive. If all these determinants are merely nonnegative, and, moreover, the functions in G_n are linearly independent on A, then G_n is called a *weak Tchebychev system* or *WT–system*. Tchebychev systems are also called *Haar systems*. A system G_n is called a *Markov system* (*weak Markov system*) if $G_k = \{g_0, \ldots, g_k\}$ is a T–system (weak T– system) for each $k = 0, 1, \ldots, n$. Markov systems are also called *Complete Chebychev systems* or *CT–systems*. If $g_0 = 1$, we say that G_n is *normalized*. If $G := \{g_0, g_1, g_2, \ldots\} \subset F(A)$ and G_n is a (normalized) Markov system for all $n \ge 0$, we say that G is a (normalized) infinite Markov system.

The definition and properties of weak Markov systems will only be used in the proof of Lemma 3 below.

Let $f(t)$ be a real valued function defined on a set A of $n \ge 2$ elements. A sequence $x_0 < \cdots < x_{n-1}$ of elements of A is called a strong alternation (weak alternation) of f of length n, if either $(-1)^i f(x_i)$ is positive (nonnegative) for all i, or $(-1)^i f(x_i)$ is negative (nonpositive) for all i.

It is well known that if G_n is a Tchebychev system on A, then the only function in $S(G_n)$ that has a weak alternation of length $n+2$ on A is the zero function, and that if G_n is a weak Tchebychev system on A, then no function in $S(G_n)$ has a strong alternation of length $n+2$ on A. It is also well known that the only function in $S(G_n)$ that has more than $n+1$ zeros is the zero function (cf., e. g. [9, 15, 16]).

Let $I(A)$ denote the convex hull of A (thus, for example, if $A := (1,2] \cup (3,\infty)$, then $I(A) = (1,\infty)$). We call $Z_n \subset F(A)$ *representable* if for all $c \in A$ there is a basis U_n of $S(Z_n)$, obtained from Z_n by a triangular transformation (i. e., $u_0(x) = z_0(x)$ and $u_i - z_i \in S(Z_{i-1}), 1 \le i \le n$); a strictly increasing function h (an "embedding function") defined on A, with $h(c) = c$; and a set $P_n := \{p_1, \ldots, p_n\}$ of continuous, increasing functions defined on $I(h(A))$, such that for every $t \in A$ and $1 \le k \le n$,

$$u_k(x) = u_0(x) \int_c^{h(x)} \int_c^{t_1} \cdots \int_c^{t_{k-1}} dp_k(t_k) \cdots dp_1(t_1). \tag{1}$$

In this case we say that (h, c, P_n, U_n) is a *representation* of Z_n. A linear space $\mathcal{S}$ is called *representable*, if it has a representable basis, and (h, c, P_n, U_n) will be called a *representation for* $\mathcal{S}$, if it is a representation for some basis of $\mathcal{S}$.

The main result of [17] implies that a Markov system on an *open* interval is representable. However, not every Markov system on a *closed* interval is representable. The problem of the representability of Markov systems is studied in, e. g., [7, 10, 11] and references thereof.

Example 1. Let the functions u_k be given by (1) with $u_0(x) = 1$,

$$h(x) := \begin{cases} 2x - 1 & \text{if } x \le 0 \\ 2x & \text{if } x > 0, \end{cases} \qquad p_1(x) := \begin{cases} (x+1)/2 & \text{if } x < -1 \\ 0 & \text{if } -1 \le x < 0 \\ x/2 & \text{if } x \ge 0, \end{cases}$$

and $p_k(x) := x$ for $k \ge 2$. Since $h(-\infty, \infty) = (-\infty, -1] \cup (0, \infty)$ and the functions p_k are strictly increasing on $h(-\infty, \infty)$, applying [11, Corollary 1] we deduce that, for any $n \ge 0$, U_n is a Markov system on $(-\infty, \infty)$. Thus U is an infinite Markov system on $(-\infty, \infty)$.

2. Generalized Derivatives

Proposition 1. *Let G_n be a Markov system on a set A, $n > 0$, and let (h, c, P_n, U_n) be a representation for G_n. Then u_1 depends only on g_0, g_1, and c. If, moreover, $p_1(c) = 0$, then also $p_1 \circ h$ depends only on g_0, g_1, and c.*

Proof. Since U_n is obtained from G_n by a triangular transformation, we have:

$$u_0 = g_0 \quad \text{and} \quad u_1 = g_1 + \alpha g_0.$$

However,

$$u_1(x) = u_0(x) \int_c^{h(x)} dp_1(t).$$

This implies that $u_1(c) = 0$ (whence thus $\alpha = -g_1(c)/g_0(c)$), and

$$u_1(x) = u_0(x)[(p_1 \circ h)(x) - (p_1 \circ h)(c)], \tag{2}$$

and therefore

$$(p_1 \circ h)(x) = u_1(x)/u_0(x) + (p_1 \circ h)(c) = g_1(x)/g_0(x) + \alpha + (p_1 \circ h)(c). \tag{3}$$

Since $h(c) = c$ and $p_1(c) = 0$, we conclude that $(p_1 \circ h)(c) = 0$, and the assertion readily follows. □

Proposition 2. *Let G_n be a Markov system of continuous functions on a closed interval $[a, b]$ with $n \ge 2$, and let (h, c, P_n, U_n) be a representation for the restriction to (a, b) of the functions in G_n. Then, for $x \in [a, b]$,*

$$u_1(x) = g_0(x) \int_c^x d(g_1/g_0)(t)$$

and

$$u_k(x) = g_0(x) \int_c^x \int_c^{h(t_1)} \cdots \int_c^{t_{k-1}} dp_k(t_k) \cdots dp_2(t_2)\, d(g_1/g_0)(t_1), \qquad 2 \le k \le n.$$

Proof. By continuity, it suffices to assume that $a < x < b$. The proof for $k = 1$ follows from (3). Assume therefore that $k \ge 2$.

Let

$$f(x) := \int_c^x \int_c^{t_2} \cdots \int_c^{t_{k-1}} dp_k(t_k) \cdots dp_2(t_2).$$

It is easy to see that the inverse function of h can be extended to an increasing (but not necessarily strictly increasing) function g, continuous on $(h(a+), h(b-))$: if $h(\alpha-) \le y \le$

$h(\alpha+)$, set $g(y) := \alpha$. Since $h(c) = c$, applying the standard change of variables theorem for Riemann–Stieltjes integrals we obtain

$$\int_c^x (f \circ h)(t)\, d(p_1 \circ h)(t) = \int_c^{h(x)} (f \circ h \circ g)(t)\, d(p_1 \circ h \circ g)(t).$$

Let $I(x)$ denote the closed interval with endpoints c and $h(x)$, let $\{\alpha_k; k \in \Lambda \subset \mathbb{Z}^+\}$ be the set of points of discontinuity of h in (a, b), $A := h(a, b) \setminus \{h(\alpha_k\}; k \in \Lambda\}$, and

$$B := \bigcup_{k \in \Lambda} [h(\alpha_k -), h(\alpha_k +)).$$

Then

$$I(x) = [A \cap I(x)) \cup (B \cap I(x))\,.$$

From (2) we see readily see that $p_1(x)$ must be constant on each nonempty interval of the form $[h(\alpha_k -), h(\alpha_k +))$. Thus $B \cap I(x)$ can be represented as the union of intervals where both p_1 and g are constant. Hence,

$$\int_{B \cap I(x)} (f \circ h \circ g)(t)\, d(p_1 \circ h \circ g)(t) = 0,$$

and

$$\int_{B \cap I(x)} f(t)\, dp_1(t) = 0.$$

Since $h \circ g = 1$ on A and $u_0 = g_0$, we thus conclude that

$$g_0(x) \int_c^x (f \circ h)(t)\, d(p_1 \circ h)(t) = u_0(x) \int_c^{h(x)} f(t)\, dp_1(t) = u_k(x),$$

and the assertion follows from (3). □

As mentioned in the preceding section, we know that if G_n is a normalized Markov system on an open interval (a, b), then it is representable. The question of under what conditions an infinite Markov system is representable is still open. In other words, given an infinite Markov system $G \subset F(A)$, we don't know under what conditions for all $c \in A$ there is a basis U of $S(G)$, obtained from G by a triangular transformation, a strictly increasing function $h \subset F(A)$ with $h(c) = c$, and a set $P := \{p_1, p_2, \dots\}$ of continuous, increasing functions defined on $I(h(A))$, such that (1) holds for every $t \in A$ and $k \geq 1$. This is because, if (h, c, P_n, U_n) is a representation for G_n, it is not yet known under what conditions there exist functions p_{n+1} and u_{n+1} such that, if $P_{n+1} := \{p_1, \dots, p_{n+1}\}$ and $U_{n+1} := (u_0, \dots, u_{n+1})$, then (h, c, P_{n+1}, U_{n+1}) is a representation for G_{n+1}.

If $A = [a, b]$, u_1/u_0 is continuous, and h is discontinuous at α then, as remarked in the proof of the preceding proposition, (2) implies that p_1 must be constant on $(h(\alpha-), h(\alpha+))$. Thus we may assume, without essential loss of generality, that h is left–continuous on (a, b). A representation (h, c, P_n, U_n) such that h is left–continuous and $p_1(c) = 0$ will be called *standard*. Since $d(p_1 - p_1(c)) = dp_1$, every (finite) Markov system of continuous functions on an open interval (a, b) has a standard representation for every $c \in (a, b)$.

For the functions h, p_k and u_k of Example 1, $(h, 0, P_n, U_n)$ is a standard representation of U_n for every $n > 0$; thus U is representable.

In the sequel, I will denote an interval and g a *continuous* real–valued strictly increasing function on I. If f is a real–valued function on I and $x \in I$ then, provided the limit exists, we define:

$$Df(x) := \lim_{h \to 0} \frac{f(x + h) - f(x)}{g(x + h) - g(x)}$$

The operator D is called the *relative derivative* with respect to g. Some applications of relative differentiation in the theory of Markov systems can be found in, e. g., [12, 14, 17].

Note that f is differentiable with respect to g at x if and only if $f \circ g^{-1}$ is differentiable in the usual sense at $g(x)$. Moreover,

$$Df(x) = \left(f \circ g^{-1}\right)' (g(x)),$$

i. e., $Df(x)$ equals the derivative of $f \circ g^{-1}$ evaluated at $g(x)$. In particular, if g is differentiable at x and $g'(x) \neq 0$, then f is differentiable with respect to g at x if and only if f is differentiable in the usual sense; in this case $Df(x) = f'(x)/g'(x)$.

The preceding displayed formula implies the following:

Lemma 1. *Let g be strictly increasing and continuous on I, and assume that f is integrable with respect to dg in I. If $c \in I$ and*

$$F(x) := \int_c^x f(t)\, dg(t),$$

then the relative derivative DF of F with respect to g exists a. e. in I, and $DF(x) = f(x)$. If f is continuous on I, then $DF(x) = f(x)$ for every $x \in I$.

Proof. Let $x \in I$. Then

$$(F \circ g^{-1})(x) = \int_c^{g^{-1}(x)} f(t)\, dg(t) = \int_{g(c)}^x \left(f \circ g^{-1}\right)(t)\, dt \tag{4}$$

is of bounded variation on every closed subinterval of $g(I)$, and thus differentiable a. e. in $g(I)$. If f is continuous, then $f \circ g^{-1}$ is continuous, and therefore $(F \circ g^{-1})(x)$ is differentiable everywhere in $g(I)$. □

For the functions u_k of Example 1, a straightforward computation shows that $u_1(x) = x$ and

$$u_2(x) = \begin{cases} x^2 - x & \text{if } x \leq 0 \\ x^2 & \text{if } x > 0. \end{cases}$$

Thus, differentiation with respect to u_1 coincides with ordinary differentiation, and we have:

$$Du_2(x) = \begin{cases} 2x - 1 & \text{if } x < 0 \\ 2x & \text{if } x > 0, \end{cases} \tag{5}$$

and $Du_2(0)$ does not exist.

Given a representation (h, c, P_n, U_n) of a Markov system G_n on (a, b) with $n \geq 1$, we define the linear operator H_n on $S(G_n)$ as follows:

$$H_n u_0 := 0, \quad H_n u_1 := 1.$$

If $n \geq 2$,

$$H_n u_2(x) := \int_c^{h(x)} dp_2(t_2).$$

If $n \geq 3$,

$$H_n u_k(x) := \int_c^{h(x)} \int_c^{t_2} \cdots \int_c^{t_{k-1}} dp_k(t_k) \cdots dp_2(t_2),\ 3 \leq k \leq n.$$

And for every $f \in S(G_n)$ by linearity.

The operator H_n is called the *weak relative derivative* with respect to (h, c, P_n, U_n).

Proposition 3. *Let $G := \{g_0, g_1, \cdots\}$ be a finite or infinite Markov system on a closed interval $[a, b]$, let $c \in (a, b)$, let $n \geq 1$, assume that $G_n \subset G$, and let H_n be the weak relative derivative with respect to some standard representation (h, c, P_n, U_n) of the restriction to*

(a,b) of the functions in G_n. Then $\{H_n u_1, H_n u_2, \cdots\}$ is a normalized Markov system on (a,b). If, moreover, the functions g_k are continuous on $[a,b]$, then

$$u_k(x) = u_0(x) \int_c^x H_n u_k(t)\, d(g_1/g_0)(t), \quad x \in [a,b], \quad 1 \le k \le n, \tag{6}$$

and the operator H_n depends neither on n nor on c, nor on the representation, but only on g_0 and g_1.

Proof. We will assume that $n \ge 2$; the proof for the special case $n = 1$ is trivial.

From [17, Corollary 1] we know that the functions p_k, $1 \le k \le n$, must be strictly increasing on $h(a,b)$. Applying the Lemma of [11], we deduce that if

$$r_1(x) := \int_c^x dp_2(t_2)$$

and

$$r_k(x) = \int_c^x \int_c^{t_2} \cdots \int_c^{t_k} dp_{k+1}(t_{k+1}) \cdots dp_2(t_2),\ 2 \le k \le n-1,$$

then $\{1, r_1, r_2, \cdots\}$ is a Markov system on $h(a,b)$. Since $H_n u_k(x) = (r_{k-1} \circ h)(x)$ and h is strictly increasing, we deduce that $\{H_n u_1, H_n u_2, \cdots\}$ is a Markov system on (a,b).

Assume now that the functions g_k are continuous. Then (6) is a consequence of Proposition 2, and g_1/g_0 is strictly increasing and continuous. If D denotes the relative derivative with respect to (g_1/g_0), from Lemma 1 and (6) we deduce that if $G_n \subset G$ and $f \in S(G_n)$, then $H_n f = Df$ a. e. on (a,b). Thus, if (u, c, P_n, U_n) and (v, c, Q_n, V_n) are two standard representations of the restrictions to (a,b) of the functions in G_n, the left–continuity of u and v imply that both the weak relative derivative with respect to (u, c, P_n, U_n) and the weak relative derivative with respect to (v, c, Q_n, V_n) of a function in $S(G_n)$ are left–continuous, and each one coincides with D a.e., and therefore they coincide with each other on (a,b), and the conclusion follows. □

Combining Lemma 1 and Proposition 3, and bearing in mind the argument used to prove the latter part of Proposition 3, we obtain the following two theorems:

Theorem 1. *Let $G := \{g_0, g_1, g_2, \cdots\}$ be a finite or infinite Markov system of continuous functions on a closed interval $[a,b]$. Then there is a unique linear operator $\widetilde{D}$ defined on the restriction of $S(G)$ to the open interval (a,b) and depending only on g_0 and g_1, such that for every $n \ge 1$, if $G_n \subset G$ and (h, c, P_n, U_n) is a standard representation of the restriction to (a,b) of the functions in G_n, with associated operator H_n, then $\widetilde{D} = H_n$ on $S(G_n)$.*

$\widetilde{D}$ will be called the *generalized derivative* associated with the system G.

Theorem 2. *Let $G := \{g_0, g_1, g_2, \cdots\}$ be an infinite (resp. finite) Markov system of continuous functions on a closed interval $[a,b]$, and let (h, c, P_n, U_n) be any standard representation of the restriction to (a,b) of the functions in G_n, Then the generalized derivative $\widetilde{D}$ associated with G has the following properties:*

(a) *The functions $\widetilde{D}g_k$ are left–continuous on (a,b), and if D denotes the relative derivative with respect to g_1/g_0, and $f \in S(G)$, then $\widetilde{D}f(x) = Df(x)$ a. e. in (a,b).*

(b)
$$u_k(x) = u_0(x) \int_c^x \widetilde{D}u_k(t)\, d(g_1/g_0)(t), \quad x \in [a,b], \quad k \ge 1.$$

(c) $\{\widetilde{D}g_1, \widetilde{D}g_2, \widetilde{D}g_3 \cdots\}$ *is an infinite (resp. finite) normalized Markov system on* (a,b).

(d) *For any $n > 0$, if $G_n \subset G$, then,*

$$\widetilde{D}u_0 = 0, \quad \widetilde{D}u_1 = 1.$$

If $n \geq 2$,

$$\widetilde{D}u_2(x) = \int_c^{h(x)} dp_2(t_2).$$

If $n \geq 3$,

$$\widetilde{D}u_k(x) = \int_c^{h(x)} \int_c^{t_2} \cdots \int_c^{t_{k-1}} dp_k(t_k) \cdots dp_2(t_2),\ 3 \leq k \leq n.$$

Since $\widetilde{D}f(x) = Df(x)$ a. e., and $\widetilde{D}u_2$ must be left–continuous, from (5) we deduce that, if u_1 and u_2 are defined as in Example 1, then

$$\widetilde{D}u_2(x) := \begin{cases} 2x - 1 & \text{if } x \leq 0 \\ 2x & \text{if } x > 0. \end{cases} \tag{7}$$

Proposition 4. *Let G_n be a normalized Markov system of continuous functions on $[a, b]$, and let $f \in S(G_n)$. Then, for every $x_0, x_1 \in [a, b]$,*

$$f(x_1) = f(x_0) + \int_{x_0}^{x_1} \widetilde{D}f(t)\, dg_1(t).$$

Proof. Let (h, c, P_n, U_n) be any standard representation of the restriction to (a, b) of the functions in G_n (this implies that $c \in (a, b)$). Since $f = \alpha_0 u_0 + \cdots + \alpha_n u_n$ for some n, from the definition of $\widetilde{D}$ we have

$$\widetilde{D}f = \alpha_1 + \alpha_2 \widetilde{D}u_2 + \cdots + \alpha_n \widetilde{D}u_n.$$

Integrating the preceding identity term by term and applying Theorem 2(b), we see that for every $x \in [a, b]$

$$u_0(x) \int_c^x \widetilde{D}f(t)\ d(g_1/g_0)(t) = \sum_{k=1}^{n} u_0(x) \int_c^x \widetilde{D}u_k\ d(g_1/g_0(t) = f(x) - \alpha_0 u_0(x).$$

Since $f(c) = \alpha_0 u_0(c)$ and $u_0 = g_0 = 1$, we conclude that

$$f(x) = f(c) + \int_c^x \widetilde{D}f(t)\, dg_1(t), \quad x \in [a, b],$$

and the assertion follows. □

3. Density of Infinite Markov Systems and Zeros of Chebychev Polynomials

The main result of this section is Theorem 3. A similar theorem is proved in [2], but only for normalized infinite Markov systems G that satisfy the so–called *Assumption 1*, i. e., each g_i is differentiable in (a, b), and if $f \in S(G)$ and f' has $n - 1$ zeros on (a, b) then f is identically constant. Here we prove the same result but only with the assumption of continuity, thus achieving maximum generality. The key to our approach is that, although the functions in a Markov system are not necessarily differentiable in the ordinary sense, they are, nevertheless, in the domain of the operator $\widetilde{D}$ defined in the preceding section. This allows us to adapt the original proof. However, the range of $\widetilde{D}$ is not contained in the set of continuous functions, and this will introduce some complications.

Given an infinite Markov system G of continuous functions defined on a closed interval $I = [a, b]$, the *generalized Chebychev polynomials* T_n associated with this system are defined, for $n > 0$, by the following properties:

(a) $T_n \in S(G_n)$.

(b) There is a sequence $y_0 < y_1 < \cdots < y_n$ in I, such that

$$\operatorname{sign} T_n(y_{k+1}) = -\operatorname{sign} T_n(y_k) = \pm\|T_n\|, \quad 0 \le k \le n. \tag{8}$$

(c) $\|T_n\| = 1 \quad \text{and} \quad T_n(b) > 0,$

where $\|\cdot\|$ denotes the norm of the supremum. The sequence $\{y_0, y_1, \cdots, y_n\}$ will be called an *alternation* for T_n. We also set $T_0 := g_0$. As is well known [4], applying the alternation and the uniqueness of best approximation theorems for Chebychev systems [4, 5], we obtain:

Proposition 5. *Let G be an infinite Markov system of continuous functions defined on a closed interval $[a, b]$, and, for $n > 0$, let $v_n := \sum_{k=0}^{n-1} a_k^n g_k$ be the element of best approximation in the norm of the supremum to g_n from $S(G_{n-1})$. Then*

$$T_n = C(g_n - v_n),$$

where C is chosen so that (c) above is satisfied.

We say that a function f has weakly constant sign on an interval I, if it is either nonnegative on I or nonpositive on I. We have:

Corollary 1. *Under the hypotheses of Proposition 5:*

(a) *$S(G_n) = S(T_0, \cdots, T_n)$.*
(b) *If $g_0 = 1$, then $T_0 = 1$ and, if $n \ge 1$ and $y_0 < y_1 < \cdots < y_n$ is an alternation for T_n, then T_n is strictly monotonic on each interval $[y_{k-1}, y_k]$, $1 \le k \le n$.*
(c) *If $g_0 = 1$, and $y_0 < y_1 < \cdots < y_n$ is an alternation for T_n, then $y_0 = a$ and $y_n = b$.*
(d) *If $g_0 = 1$, then for each n there is a unique alternation for T_n.*
(e) *If $g_0 = 1$, then $\widetilde{D}T_n$ has weakly constant sign on each interval (y_{k-1}, y_k), $1 \le k \le n$.*

Proof.
(a) Trivial.
(b) Assume T_n is not strictly monotonic on some interval $[y_{k-1}, y_k]$, and assume for example that $T_n(y_{k-1}) = -1 = -T_n(y_k)$. Then there exist points $\xi, \eta \in (y_{k-1}, y_k)$ such that $\xi < \eta$ and $-1 < T_n(\eta) \le T_n(\xi) < 1$. But $T_n(\eta) \ne T_n(\xi)$, otherwise $T_n(x) - T_n(\xi)$ would have $n+1$ zeros in $[a, b]$. Choosing now some point $\delta \in (T_n(\eta), T_n(\xi))$ we see that $y_0, \cdots, y_{k-1}, \xi, \eta, y_k, \cdots, y_n$ is a strong (and therefore weak) alternation of length $n+2$ for $T_n - \delta$, which is contradiction.
(c) Assume $a < y_0$. If $T_n(a) = -T_n(y_0)$, then $a, y_0, \cdots, y_n$ is a strong alternation of length $n+2$ for T_n, which is a contradiction. If $T_n(a) = T_n(y_0)$ then $a, y_1, \cdots, y_n$ is an alternation for T_n, and (b) implies that T_n must be strictly monotonic on $[a, y_1]$, which is a contradiction. We have therefore established that $|T_n(a)| < 1$. This implies that $g := T_n - T_n(a)$ has a weak alternation of length $n+2$ on $[a, b]$, which is also a contradiction. The proof that $y_n = b$ is similar.
(d) This follows from (b) and (c).
(e) Assume for example that $T_n(y_{k-1}) = -1 = -T_n(y_k)$. Then (b) implies that T_n must be strictly increasing on $[y_{k-1}, y_k]$. We claim that $\widetilde{D}T_n$ is nonnegative on (y_{k-1}, y_k). Indeed, were it not so, there would be a point $x_1 \in (y_{k-1}, y_k)$ such that $\widetilde{D}T_n(x_1) < 0$. Since $\widetilde{D}T_n$ is left–continuous it must be strictly negative on some subinterval $[x_0, x_1]$ of (y_{k-1}, y_k) . Applying Proposition 4 we conclude that

$$T_n(x_1) - T_n(x_0) = \int_{x_0}^{x_1} \widetilde{D}T_n(t)\, dg_1(t) < 0.$$

Thus T_n is not strictly increasing on $[y_{k-1}, y_k]$, and we have a contradiction. □

Let G be an infinite Markov system on $[a,b]$, let T_n denote the generalized Chebychev polynomials associated with G, and let

$$M_n := \max\{|x_i - x_{i-1}| : 1 \le i \le n+1\},$$

where $x_1 \cdots < x_n$ are the zeros of T_n, $x_0 = a$, and $x_{n+1} = b$.

The following proposition generalizes [2, Lemma 1]. The method of proof is similar, using $\widetilde{D}$ instead of ordinary differentiation to bypass the assumption of differentiability. However, there is an important difference between the two operators: whereas a differentiable function must have a vanishing derivative at a relative extremum, this is not necessarily the case for $\widetilde{D}$. Indeed, assume for example that f has a relative minimum at a point $x_0 \in (a,b)$, but that the relative derivative Df does not exist at x_0. Since $Df(x) \le 0$ a. e. on an interval of the form $(x_0 - \delta, x_0)$, and $\widetilde{D}f$ equals Df wherever the latter is defined, the left–continuity of $\widetilde{D}f$ implies that $\widetilde{D}f(x_0) \le 0$. Since $\widetilde{D}f$ is not necessarily right–continuous at x_0, this is all we can say, i. e., $\widetilde{D}f$ may not vanish at x_0. This introduces a complication, since the proof of [2, Lemma 1] is based on the number of zeros of the derivatives of functions in $S(G)$ (see the definition of Assumption 1 at the beginning of this section). We get around this difficulty by considering alternations instead of zeros.

Lemma 2. *Let $G \subset C[a,b]$ be a normalized infinite Markov system, let $a < a_1 < b_1 < b$, let f be the function defined in $A := [a,a_1] \cup [b_1,b]$ by*

$$f(x) := \begin{cases} 0 & \text{if } x \in [a,a_1] \\ 1 & \text{if } x \in [b_1,b], \end{cases}$$

and let $S_n \subset S(G_n)$ denote the element of best approximation to f in A. Then

(a) *S_n is increasing on $[a_1,b_1]$.*

(b) *Assume that $\lim_{n\to\infty} M_n = 0$. Then $\|f - S_n\|_A < 8M_n/(b_1 - a_1)$, where $\|\cdot\|_A$ denotes the norm of the supremum in A.*

Proof. Since f is continuous on A, applying the alternation theorem for Chebychev systems ([5]), we deduce that for each n there is an element of best approximation S_n and an alternation $y_0 < y_1 < \cdots y_{n+1}$ for $f - S_n$.

We have:

$$\varepsilon_n := \|f - S_n\|_A \le \|f - 1/2\|_A \le 1/2. \tag{9}$$

Since f vanishes on $[a,a_1]$, the $y_j \in [a,a_1]$ are local extrema of S_n; since $f = 1$ on $[b_1,b]$, the $y_j \in [b_1,b]$ are local extrema of $S_n - 1$. Note that

$$\widetilde{D}(S_n - 1) = \widetilde{D}S_n \tag{10}$$

on $[a,b]$, and, for every $x, y \in [a,b]$

$$[S_n - 1](y) - [S_n - 1](x) = S_n(y) - S_n(x). \tag{11}$$

We now show that, for $1 \le j < n+1$, there is a point $\xi_j \in (y_{j-1}, y_j)$ such that

$$\operatorname{sign}\left[\widetilde{D}S_n(\xi_j)\right] = \operatorname{sign}[S_n(y_j) - S_n(y_{j-1})]. \tag{12}$$

To verify this assertion first note that, since y_{j-1} and y_j belong to an alternation for $f - S_n$, we must have $S_n(y_j) - S_n(y_{j-1}) \ne 0$. Assume for instance that $S_n(y_j) - S_n(y_{j-1} > 0$. Then, if $\widetilde{D}S_n \le 0$ in (y_{j-1}, y_j), the identity

$$S_n(y_j) - S_n(y_{j-1}) = \int_{y_{j-1}}^{y_j} \widetilde{D}S_n \, dg_1(t),$$

(which follows from Proposition 4), would yield a contradiction.

It is now easy to see that if $y_{j-1}, y_j \in [a, a_1]$ or $y_{j-1}, y_j \in [b_1, b]$, then

$$\operatorname{sign}[\widetilde{D}S_n(\xi_j)] = \operatorname{sign}\left[(S_n - f)(y_j) - (S_n - f)(y_{j-1})\right]. \tag{13}$$

Indeed, if $y_{j-1}, y_j \in [a, a_1]$, this follows from (12),whereas for $y_{j-1}, y_j \in [b_1, b]$, this follows from (10), (11), and (12). Equation (13) implies that if all the y_j are contained in $[a, a_1]$ or all the y_j are contained in $[b_1, b]$, then $\{\xi_j\}_{j=1}^{n+1}$ must be a strong alternation of length $n+1$ for $\widetilde{D}S_n$. Since $\widetilde{D}S_n \in S(\widetilde{D}g_1, \widetilde{D}g_2, \cdots \widetilde{D}g_n\}$, and Theorem 2(c) implies that $\{\widetilde{D}g_1, \widetilde{D}g_2, \cdots \widetilde{D}g_n\}$ is a Markov system, and therefore a weak Tchebychev system, this would be a contradiction. Thus there is an m such that $y_m \le a_1 < b_1 \le y_{m+1}$.

Note that if $m = 0$, then ξ_{m-1} is undefined, whereas if $m = n-1$, ξ_{m+1} is undefined. If $0 < m < n-1$, (13) implies that $\operatorname{sign}\left[\widetilde{D}S_n(\xi_{m-1})\right] = \operatorname{sign}\left[\widetilde{D}S_n(\xi_{m+1})\right]$.

Assume S_n is not monotonic in $[y_m, y_{m+1}]$. This implies that there are points $d_1 < d_2 < d_3$ in $[y_m, y_{m+1}]$ such that $S_n(d_1) > S_n(d_2) < S_n(d_3)$ or $S_n(d_1) < S_n(d_2) > S_n(d_3)$. Repeating the argument used to establish (12), we deduce that in either case there are points $\zeta_0 \in (y_m, d_2)$ and $\zeta_1 \in (d_2, y_{m+1})$ such that

$$\operatorname{sign}\left[\widetilde{D}S_n(\zeta_0)\right] \ne \left[\operatorname{sign}\widetilde{D}S_n(\zeta_1)\right].$$

Thus, redefining ξ_m to equal ζ_0 or ζ_1, as appropriate, we see that $\{\xi_j\}_{j=0}^{n}$ is a strong alternation of length $n+1$ for $\widetilde{D}S_n$, and again we have a contradiction. Since $[a_1, b_1] \subset [y_m, y_{m+1}]$ and $S_n(a_1) < \frac{1}{2} < S_n(b_1)$, we conclude that S_n is increasing on $[a_1, b_1]$.

To prove (b) we proceed as follows:

(i) Let $Q_n = \frac{3}{4}\varepsilon_n T_n + S_n$. Then, for $y_j \in [a, a_1]$ we have:

$$Q_n(y_j) = \begin{cases} \frac{3}{4}\varepsilon_n T_n(y_j) + \varepsilon_n \ge \frac{1}{4}\varepsilon_n & \text{if} \quad [S_n - f](y_j) = \varepsilon_n \\ \\ \frac{3}{4}\varepsilon_n T_n(y_j) - \varepsilon_n \le -\frac{1}{4}\varepsilon_n & \text{if} \quad [S_n - f](y_j) = -\varepsilon_n, \end{cases}$$

which implies that

$$\operatorname{sign}[Q_n(y_j)] = \operatorname{sign}[S_n - f](y_j). \tag{14}$$

Similarly, we also see that if $y_j \in [b_1, b]$,

$$\operatorname{sign}[Q_n(y_j) - 1] = \operatorname{sign}[S_n - f](y_j). \tag{15}$$

Since

$$(S_n - f)(y_j) = \delta(-1)^j \varepsilon_n, \quad 0 \le j \le n,$$

where $\delta = \pm 1$, from (14) and (15) we deduce that

$$\operatorname{sign}\left[(S_n - f)(y_j) - (S_n - f)(y_{j-1})\right] = \operatorname{sign}\left[Q_n(y_j) - Q_n(y_{j-1})\right], \quad 1 \le j \le n,$$

and by an argument similar to the one employed in (a), we deduce that there is a strictly increasing sequence of points ξ_j, with $0 \le j \le n$, $j \ne m$, $0 \le m \le n-1$, such that

$$\xi_j \in (y_j, y_{j+1}) \subseteq [a, a_1] \quad \text{if } j < m, \qquad \xi_j \in (y_j, y_{j+1}) \subseteq [b_1, b] \quad \text{if } j \ge m+1,$$

$$\operatorname{sign}\left[\widetilde{D}Q_n(\xi_j)\right] = \operatorname{sign}\left([S_n - f](y_j) - [S_n - f](y_{j-1})\right), \ 0 \le j \le n, j \ne m,$$

and

$$\operatorname{sign}\left[\widetilde{D}Q_n(\xi_{m-1})\right] = \operatorname{sign}\left[\widetilde{D}Q_n(\xi_{m+1})\right] \quad \text{if;} \, m < n.$$

Since $0 \le m \le n$, ξ_{m-1} is undefined if $m = 0$, and ξ_{m+1} is undefined if $m = n-1$.

(ii) Assume $[\alpha, \beta] \subset [a_1, b_1]$ is such that T_n has three alternation points $\nu_1 < \nu_2 < \nu_3$, as in (8). We will show that $S_n(\beta) - S_n(\alpha) > \varepsilon_n$.

Since S_n is increasing on $[a_1, b_1]$, we have

$$S_n(\alpha) \le S_n(x) \le S_n(\beta), \quad x \in [\alpha, \beta].$$

If $0 \le S_n(\beta) - S_n(\alpha) \le \varepsilon_n$ then, for $\alpha \le x \le \beta$ we would have

$$S_n(x) - \frac{S_n(\alpha)+S_n(\beta)}{2} \le \frac{S_n(\beta)-S_n(\alpha)}{2} \le \frac{\varepsilon_n}{2}$$

and

$$S_n(x) - \frac{S_n(\alpha)+S_n(\beta)}{2} \ge -\frac{S_n(\beta)-S_n(\alpha)}{2} \ge -\frac{\varepsilon_n}{2}.$$

Therefore, for $i = 1, 2, 3,$

$$Q_n(\nu_i) - \frac{S_n(\alpha)+S_n(\beta)}{2} = \frac{3}{4}\varepsilon_n T_n(\nu_i) + \left[S_n(\nu_i) - \frac{S_n(\alpha)+S_n(\beta)}{2}\right]$$

is positive if $T_n(\nu_i) = 1$ and negative if $T_n(\nu_i) = -1$. This implies the existence of points $\eta_i \in (\nu_i, \nu_{i+1})$, $i = 1, 2$, such that

$$\operatorname{sign}\left[\widetilde{D}Q_n(\eta_1)\right] \ne \operatorname{sign}\left[\widetilde{D}_1 Q_n(\eta_2)\right].$$

Setting $\xi_m := \eta_i$, where $i = 1$ or $i = 2$ is chosen in such a way that

$$\operatorname{sign}\left[\widetilde{D}Q_n(\xi_m)\right] \ne \operatorname{sign}\left[\widetilde{D}Q_n(\xi_{m-1})\right]$$

if ξ_{m-1} is defined (i. e. if $m > 0$), and

$$\operatorname{sign}\left[\widetilde{D}Q_n(\xi_m)\right] \ne \operatorname{sign}\left[\widetilde{D}Q_n(\xi_{m+1})\right]$$

if ξ_{m-1} is undefined, we see that $\{\xi_j\}_{j=0}^n$ is a strong alternation of length $n+1$ for $\widetilde{D}Q$ in $[a, b]$, which is a contradiction. Thus $S_n(\beta) - S_n(\alpha) > \varepsilon_n$.

(iii) Let $n \ge 4$. If T_n vanishes at α and at β and has exactly two additional zeros in $[\alpha, \beta]$, then $\beta - \alpha \le 3M_n$, since M_n is the largest distance between consecutive zeros of T_n.

Since $M_n \to 0$, there is a number n_0 such that, for $n \ge n_0$, $[a_1, b_1]$ contains at least four zeros of T_n. For such an n, let α_1 denote the smallest of the zeros of T_n that is larger or equal to a_1, and let α_2 be the zero of T_n satisfying the condition that (α_1, α_2) contains exactly two zeros of T_n. Once α_i has been chosen, let α_{i+1} be the zero of T_n satisfying the condition that (α_i, α_{i+1}) contains exactly two zeros of T_n. We continue in this fashion until we find the largest possible α_{m-1} that is smaller or equal to b_1 (note that $m-1 \ge 2$). Now let $\alpha_0 := a_1$ and $\alpha_m := b_1$; then

$$b_1 - a_1 \le \sum_{i=1}^{m} |\alpha_i - \alpha_{i-1}| \le 3M_n m. \tag{16}$$

Since Corollary 1 implies that each interval (α_i, α_{i+1}) must contain three alternation points and, as we showed in (a) above, S_n is increasing on $[a_1, b_1]$, from (ii) we deduce that

$$S_n(b_1) - S_n(a_1) = \sum_{i=1}^{m} S_n(\alpha_i) - S_n(\alpha_{i-1}) \ge \sum_{i=2}^{m-1} S_n(\alpha_i) - S_n(\alpha_{i-1}) > (m-2)\varepsilon_n\ . \tag{17}$$

From (9) and (17) we obtain

$$2 \ge 1 + 2\varepsilon_n \ge S_n(b_1) - S_n(a_1) \ge (m-2)\varepsilon_n\ , \tag{18}$$

whereas from (16)

$$m - 2 \ge \frac{b_1 - a_1}{3M_n} - 2. \tag{19}$$

Moreover, for n sufficiently large we have

$$\frac{b_1 - a_1}{3\mathrm{M}_n} - 2 > \frac{b_1 - a_1}{4\mathrm{M}_n}. \tag{20}$$

Finally, combining (18), (19), and (20), we obtain

$$2 \geq (m-2)\varepsilon_n > \frac{b_1 - a_1}{4M_n}\varepsilon_n,$$

i. e.

$$\|f - S_n\|_A = \varepsilon_n < \frac{8M_n}{b_1 - a_1}.$$

□

The following theorem generalizes part of [2, Theorem 1]. It uses Lemma 2 instead of [2, Lemma 1], but otherwise the proof is identical. Since the details were omitted in [2], we include them here for the sake of completeness.

Theorem 3. *Let G be an infinite normalized Markov system of continuous functions on a closed interval $[a, b]$. Then the following propositions are equivalent:*

(a) *$S(G)$ is dense in $C[a, b]$, in the norm of the supremum.*
(b) $\lim_{n\to\infty} M_n = 0$.

Proof. (a)⇒(b) is [2, Theorem 1(A)].
(b)⇒(a): Assume $S(G)$ is not dense in $C[a, b]$. Then there exists a nonzero Borel measure μ such that for every function $g \in S(G)$

$$\int_a^b g(t)\, d\mu(t) = 0.$$

Let $b_1 \in (a, b)$ and $0 < \varepsilon < 1/2$ be arbitrary but fixed, let $a_1 \in (a, b_1)$ be such that

$$\mu([a_1, b_1]) < \frac{\varepsilon}{6}\mu([a, b]),$$

$A := [a, a_1] \cup [b_1, b]$, f the function defined in Lemma 2, and S one of the functions S_n defined in Lemma 2, such that $\|S - f\|_A < \frac{\varepsilon}{3}$. Since $S \in S(G)$,

$$0 = \int_a^b S(t)\, d\mu(t) = \int_a^{a_1} S(t)\, d\mu(t) + \int_{a_1}^{b_1} S(t)\, d\mu(t) + \int_{b_1}^{b} S(t)\, d\mu(t).$$

Thus,

$$\begin{aligned} \mu([b_1, b]) &= \int_{b_1}^b d\mu(t) = \int_{b_1}^b S(t)\, d\mu(t) - \int_{b_1}^b [S(t) - 1]\, d\mu(t) \\ &= -\int_a^{a_1} S(t)\, d\mu(t) - \int_{a_1}^{b_1} S(t)\, d\mu(t) - \int_{b_1}^b [S(t) - 1]\, d\mu(t). \end{aligned}$$

Since S is increasing on $[a_1, b_1]$,

$$\|S\|_{[a_1,b_1]} = \max\{|S(a_1)|, |S(b_1)|\}.$$

But

$$|S(a_1)| = |S(a_1) - f(a_1)| \leq \|S - f\|_{[a,a_1]} \leq \frac{\varepsilon}{3},$$

and

$$|S(b_1)| \leq |S(b_1) - f(b_1)| + |f(b_1)| \leq \|S - f\|_{[b_1,b]} + \|f\|_{[b_1,b]} \leq 1 + \frac{\varepsilon}{3} \leq 2.$$

Thus

$$\|S\|_{[a_1,b_1]} \leq 2.$$

Moreover,

$$\|S\|_{[a,a_1]} = \|S - f\|_{[a,a_1]} < \frac{\varepsilon}{3} \quad \text{and} \quad \|S - 1\|_{[b_1,b]} = \|S - f\|_{[b_1,b]} < \frac{\varepsilon}{3},$$

and we conclude that

$$\mu([b_1, b]) \leq \frac{\varepsilon}{3}\mu([a, a_1]) + 2\mu([a_1, b_1]) + \frac{\varepsilon}{3}\mu([b_1, b]) < \varepsilon\mu([a, b]).$$

This implies that $\mu = 0$, which is a contradiction. □

4. Density of Infinite Markov Systems and Unbounded Bernstein Inequalities

The main result of this section is Theorem 5. A similar result is proved in [3], but only for infinite Markov systems G that satisfy the additional hypotheses that each function g_i is two times differentiable in (a, b) and $(g_1/g_0)'$ does not vanish in (a, b). Although here we make the assumption that $g_0 = 1$, dividing if necessary all the functions in G by g_0, it is easy to see that this does not imply any essential loss of generality.

Let G be a normalized Markov system of continuous functions in $[a, b]$, and let $V \subset S(G)$. We say that V has an *everywhere unbounded Bernstein inequality* if for any real number γ and every $[\alpha, \beta] \subset [a, b]$, there is a function $f \in V$ such that

$$\gamma\|f\|_{[a,b]} \leq \|\widetilde{D}f\|_{[\alpha,\beta]}.$$

Example 2. Let

$$0 < \lambda_1 < \cdots < \lambda_k < \lambda_{k+1} < \cdots,$$

and $G^0 := \{1, x^{\lambda_1}, \cdots, x^{\lambda_k}, \cdots\}$. Then G^0 is a normalized infinite Markov system on $[0, 1]$ (cf., e. g., [3, p. 96], [13, p. 305]). Moreover, by a classical theorem of Müntz (cf. [5]), $S(G^0)$ is dense in $\mathcal{C}[0, 1]$ if and only if

$$\sum_{k=1}^{\infty} \frac{1}{\lambda_k} = \infty. \tag{21}$$

Taking

$$u_0 := 1, \quad u_1(x) := x^{\lambda_1}, \quad u_2(x) := x^{\lambda_2},$$

we have:

$$u_1(x) = \int_0^x d(t_1^{\lambda_1}), \quad \text{and} \quad u_2(x) = \int_0^x \int_0^{t_1} d((\lambda_2/\lambda_1)t_2^{\lambda_2-\lambda_1})\, d(t_1^{\lambda_1}),$$

Thus, we have a representation with

$$h : [0, 1] \to [0, 1], \quad h(x) := x, \quad c := 0, \quad p_1(t) := t^{\lambda_1}, \quad p_2(t) := (\lambda_2/\lambda_1)t^{\lambda_2-\lambda_1}.$$

If $p \in S(G^0)$, then p is differentiable in the usual sense. Thus, if D denotes the relative derivative with respect to u_1, we have:

$$Dp(x) = \frac{\left(\dfrac{dp(x)}{dx}\right)}{\left(\dfrac{du_1(x)}{dx}\right)} = \lambda_1^{-1}x^{1-\lambda_1}p'(x),$$

which is continuous, and by Theorem 2(a) we deduce that $\widetilde{D} \equiv D$. In particular, if $\lambda_k := \frac{k}{2}$ for $k \in \mathbb{Z}^+$, and $p \in S(G^0)$, then $\widetilde{D}p(x) = 2\sqrt{x}p'(x)$.

Note that for this choice of the λ_k, Müntz's theorem implies that $S(G^0)$ is dense in $C[0, 1]$. Let us verify that $S(G^0)$ has an everywhere unbounded Bernstein inequality.

If $P_n(x)$ is the classical Chebyshev polynomial of degree n, we know that $P_n(x) \in S(G^0)$ and that $P_n(x)$ has exactly one alternation of length $n+1$ such that the ξ_k are extrema of $P_n(x)$:

$$\xi_k = \cos(\frac{\pi k}{n}), \quad k = 0, ..., n$$

(cf. Corollary 1 and [5]).

Given $0 \le \alpha < \beta \le 1$ and $\gamma > 0$, we can find an n such that $P_n(x)$ has consecutive extrema ξ_k, ξ_{k+1} of $P_n(x)$, such that

$$\alpha \le \xi_k < \xi_{k+1} \le \beta \quad \text{and} \quad \xi_{k+1} - \xi_k < \frac{4\sqrt{\alpha}}{\gamma}.$$

Replacing if necessary $P_n(x)$ by $-P_n(x)$ so that $P_n(\xi_{k+1}) = -P_n(\xi_k) = 1$, and applying the Mean Value Theorem, we deduce the existence of a point $\zeta \in [\xi_k, \xi_{k+1}]$ such that

$$P_n'(\zeta) = \frac{2}{\xi_{k+1} - \xi_k}.$$

Since

$$\|\widetilde{D}P_n\|_{[\alpha,\beta]} \ge 2\sqrt{\alpha}P_n'(\zeta) = \frac{4\sqrt{\alpha}}{\xi_{k+1} - \xi_k} > \gamma,$$

and $\|P_n\|_{[0,1]} = 1$, the assertion follows.

The next theorem generalizes Theorem [3, Theorem 3]. It uses the operator $\widetilde{D}$ instead of ordinary differentiation. Otherwise the proof is the same.

Theorem 4. *Suppose $G := \{1, g_1, g_2 \cdots\}$ is a normalized infinite Markov system of continuous functions on $[a,b]$, and let $f \in S(G_n)$ for some $n \in \mathbb{Z}^+$. Then for every $z_0 \in [a,b]$ such that $|T_n(z_0)| \ne 1$,*

$$\frac{|\widetilde{D}f(z_0)|}{\|f\|_{[a,b]}} \le \frac{2}{1 - |T_n(z_0)|}|\widetilde{D}T_n(z_0)|.$$

Proof. The hypotheses imply that z_0 is not an extremum of T_n. Thus, if $a = y_0 \le \cdots \le y_n = b$ is the alternation for T_n, then

$$y_k < z_0 < y_{k+1}.$$

for some $0 \le k \le n-1$.

Without essential loss of generality, we may assume that

$$\widetilde{D}f(z_0) \ne 0, \quad \|f\|_{[a,b]} = 1, \quad \text{and} \quad \text{sign}\left[\widetilde{D}f(z_0)\right] = \text{sign}\left[\widetilde{D}T_n(z_0)\right], \tag{22}$$

because, if $\widetilde{D}f(z_0) = 0$ there is nothing to prove, and the other two conditions are readily obtained dividing by a suitable constant.

We may also assume that $T_n(y_k) = -1$, otherwise we may apply the argument that follows to the function $-T_n$.

Let $\delta := |T_n(z_0)| < 1$, let ε, $0 < \varepsilon < 1$, be arbitrary but constant, and

$$\eta := T_n(z_0) - \frac{1-\delta}{2}(1-\varepsilon)f(z_0).$$

Then

$$|\eta| \le \delta + \frac{1-\delta}{2} = \frac{\delta + 1}{2}.$$

If

$$g(x) := \eta + \frac{1-\delta}{2}(1-\varepsilon)f(x)$$

it is clear that $g \in S(G_n)$, and

$$g(z_0) = T_n(z_0). \tag{23}$$

Moreover, from (22) we obtain

$$|g(x)| \le \eta + \frac{1-\delta}{2}(1-\varepsilon)|f(x)| < \eta + \frac{1-\delta}{2} \le 1, \tag{24}$$

and

$$\operatorname{sign}\left[\widetilde{D}g(z_0)\right] = \operatorname{sign}\left[\widetilde{D}f(z_0)\right] = \operatorname{sign}\left[\widetilde{D}T_n(z_0)\right]. \tag{25}$$

If

$$|\widetilde{D}f(z_0)| > \frac{2}{1-|T_n(z_0)|}|\widetilde{D}T_n(z_0)| = \frac{2}{1-\delta}|\widetilde{D}T_n(z_0)|,$$

then, for ε sufficiently small,

$$|\widetilde{D}g(z_0)| = \frac{1-\delta}{2}(1-\varepsilon)|\widetilde{D}f(z_0)| > |\widetilde{D}T_n(z_0)| \tag{26}$$

Since Corollary 1(e) implies that $\widetilde{D}T_n \ge 0$ on (y_k, y_{k+1}), from (25) and (26) we conclude that

$$0 \le \widetilde{D}T_n(z_0) < \widetilde{D}g(z_0).$$

Moreover, since Theorem 2(a) implies that $\widetilde{D}g$ and $\widetilde{D}T_n$ are left–continuous, setting

$$r := (1/3)[\widetilde{D}g(z_0) - \widetilde{D}T_n(z_0)]$$

there must exist a number $\rho > 0$ such that if $y_k < z_0 - \rho < x < z_0 < y_{k+1}$, then

$$|\widetilde{D}T_n(x) - \widetilde{D}T_n(z_0)| < r \quad \text{and} \quad |\widetilde{D}g(x) - \widetilde{D}g(z_0)| < r,$$

which implies that for every $x \in (z_0 - \rho, z_0)$ we have

$$\widetilde{D}T_n(x) < \widetilde{D}T_n(z_0) + r < \widetilde{D}g(z_0) - r < \widetilde{D}g(x).$$

Thus,

$$T_n(z_0) - T_n(z_0 - \rho) = \int_{z_0-\rho}^{z_0} \widetilde{D}T_n(t)\, dg_1(t) < \int_{z_0-\rho}^{z_0} \widetilde{D}g(t)\, dg_1(t) = g(z_0) - g(z_0 - \rho)$$

and therefore, applying (23),

$$g(z_0 - \rho) < T_n(z_0 - \rho).$$

If

$$q := g - T_n$$

we see that

$$q(z_0) = 0 \quad \text{and} \quad q(z_0 - \rho) < 0.$$

It is also easy to see that

$$q(y_k) = g(y_k) - T_n(y_k) > -1 + 1 = 0,$$

since (24) implies that $|g| < 1$. Thus $q(x)$ must have at least two zeros in (y_k, y_{k+1}). But (24) also implies that $q(x)$ must have at least one zero in each interval (y_i, y_{i+1}), $0 \le i \le n$. Thus $q(x)$ will have $n+2$ zeros in $[a, b]$. Since $q \in S(G_n)$ and G_n is a Markov system, this is a contradiction. □

Although a real–valued function defined and bounded on a closed interval may not have extrema (unless it is continuous) its domain has, nevertheless, points that behave locally as if they were:

Proposition 6. *Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then there are points $\nu_0, \nu_1 \in [a, b]$, such that*

$$\inf\{f(x) : x \in [a, b]\} = \lim_{r\to 0^+}(\inf\{f(x) : |x - \nu_0| < r\})$$

and

$$\sup\{f(x) : x \in [a, b]\} = \lim_{r\to 0^+}(\sup\{f(x) : |x - \nu_0| < r\}).$$

Proof. Let $s_0 := \inf\{f(x) : x \in [a, b]\}$. Then, for every $n \in \mathbb{Z}^+$ there is a point $x_n \in [a, b]$ such that $f(x_n) - s_0 < 1/n$. Let ν_0 be an accumulation point of $\{x_n; n \in \mathbb{Z}^+\}$, and $\{x_{n_j}; j \in \mathbb{Z}^+\}$ a subsequence that converges to ν_0.

Given $\varepsilon, r > 0$, there is a $j_0 > 1/\varepsilon$ such that $|x_{n_j} - \nu_0| < r$ for $j > j_0$. Thus,

$$s_0 \le \inf\{f(x); |x - \nu_0| < r\} \le f(x_{n_j}) < s_0 + 1/n_j.$$

Therefore

$$|\inf\{f(x); |x - \nu_0| < r\} - s_0| < \varepsilon,$$

i. e.,

$$\inf\{f(x); x \in [a, b]\} = \lim_{r\to 0^+}\{\inf\{f(x); |x - \nu_0| < r\}\}.$$

The corresponding statement for ν_1 is established by a similar argument. □

The points ν_0 and ν_1 are respectively called a *pseudominimum* and a *pseudomaximum* of $f(x)$.

Using Proposition 6 we will establish the following generalization of [3, Lemma 1]:

Lemma 3. *Suppose $G := \{1, g_1, g_2 \cdots\}$ is a normalized infinite Markov system of continuous functions on $[a, b]$, and the sequence of associated Chebyshev polynomials $\{T_n\}$ has a subsequence $\{T_{n_i}\}$ with no zeros on some subinterval $[\alpha, \beta] \subset [a, b]$. Then there exists another subinterval $[a_1, b_1]$ of $[a, b]$ and another infinite subsequence $\{T_{n_k}\}$ such that for some $\delta > 0$, $\gamma > 0$, and for each n*

$$\|T_{n_k}\|_{[a_1,b_1]} < 1 - \delta \tag{27}$$

and

$$\|\widetilde{D}T_{n_k}\|_{[a_1,b_1]} < \gamma \tag{28}$$

Proof.
(i) We will find an interval $[a_2, b_2] \subset [a, b]$ and a sequence $\{n_{i,1}\} \in \mathbb{Z}^+$ such that either all the $T_{n_{i,1}}$ are increasing on $[a_2, b_2]$ or all the $T_{n_{i,1}}$ are decreasing on $[a_2, b_2]$.

Since, by hypothesis, the T_{n_i} do not vanish on $[\alpha, \beta]$, each alternation may have at most one point in $[\alpha, \beta]$. The set of such points is either finite or has an accumulation point $\xi \in [\alpha, \beta]$; in either case there is a closed interval I that contains ξ and is strictly contained in $[\alpha, \beta]$, and a subsequence $\{T_{n_j}\}$ of $\{T_{n_i}\}$, such that each T_{n_j} has an alternation point in I. Therefore, we may choose a subinterval $[a_2, b_2] \subset [\alpha, \beta]$, disjoint with I, such that no T_{n_j} has an alternation point in $[a_2, b_2]$. Applying Corollary 1(b) we deduce that each T_{n_j} is either increasing or decreasing on $[a_2, b_2]$. We will show that if the sequence of those T_{n_j} that are increasing on $[a_2, b_2]$ is infinite, then this will be the subsequence $\{T_{n_{i,1}}\}$, whereas if the sequence of the T_{n_j} that are increasing on $[a_2, b_2]$ is finite, then $\{T_{n_{i,1}}\}$ will be the sequence of those T_{n_j} that are decreasing on $[a_2, b_2]$.

(ii) Proceeding exactly as in the proof of the first inequality in [3, Lemma 1], we conclude that there is a subinterval $[a_3, b_3] \subset [a_2, b_2]$, $\delta > 0$, and a subsequence $\{T_{n_{i,2}}\}$ of $\{T_{n_{i,1}}\}$, such that (27) is satisfied. Without loss of generality, we may assume that all the elements of the sequence $\{n_{i,2}\}$ are strictly larger than 3.

(iii) Let n be any term of the sequence $\{n_{i,2}\}$, and let $a = y_0 < y_1 < \cdots < y_{n-1} < y_n = b$ be the alternation points of T_n, as in (8). By Proposition 4 we know that for $x \in [a, b]$

$$T_n(x) = T_n(c) + \int_c^x \widetilde{D}T_n(t)\, dg_1(t). \tag{29}$$

Since $T_n(x)$ is strictly monotonic on (y_k, y_{k+1}), from (29) we deduce that if

$$y_k \leq z_1 < z_2 \leq y_{k+1}, \quad 0 \leq k \leq n-1,$$

then there is a point $z \in (z_1, z_2)$ such that

$$\operatorname{sign}\left[\widetilde{D}T_n(z)\right] = \operatorname{sign}\left[T_n(y_{k+1}) - T_n(y_k)\right]$$

(z can be chosen to be an interior point because T_n is continuous).

In particular, there is a $\xi_k \in (y_k, y_{k+1})$ such that

$$\operatorname{sign}\left[\widetilde{D}T_n(\xi_k)\right] = \left[\operatorname{sign}[T_n(y_{k+1}) - T_n(y_k)]\right].$$

Thus, we have found a sequence $a < \xi_0 < \cdots < \xi_{n-1} < b$, with $\xi_k \in (y_k, y_{k+1})$, and a constant η, where $\eta = \pm 1$, such that

$$\operatorname{sign}\left[(-1)^k \widetilde{D}T_n(\xi_k)\right] = \eta \tag{30}$$

Let (h, c, P_n, U_n) be a standard representation for the restrictions to (a, b) of the functions in G_n, $\widetilde{G}_n := \{1, \widetilde{D}g_2, \cdots, \widetilde{D}g_n\}$, $\widetilde{U}_n := \{1, \widetilde{D}u_2, \cdots, \widetilde{D}u_n\}$, and $\widetilde{P}_n := \{p_2, \cdots, p_n\}$. Since U_n is obtained from G_n by a triangular transformation, it follows that $\widetilde{U_n}$ is obtained from $\widetilde{G_n}$ by a triangular transformation. This implies that $\widetilde{U_n}$ is a basis for $S(\widetilde{G}_n)$.

Let H_n^* be the linear operator defined on $S(\widetilde{G}_n)$ as follows:

$$H_n^* \widetilde{D}u_1 = H_n^* 1 := 0, \quad H_n^* \widetilde{D}u_2 := 1,$$
$$H_n^* \widetilde{D}u_3 = \int_c^x dp_3(t_3),$$

and

$$H_n^* \widetilde{D}u_k(x) := \int_c^x \int_c^{t_3} \cdots \int_c^{t_{k-1}} dp_k(t_k) \cdots dp_3(t_3),\ 4 \leq k \leq n,$$

and for every $f \in S(\widetilde{G}_n)$ by linearity.

Applying Theorem 2(d) and proceeding as in the proof of Proposition 4, we deduce that for every $f \in S(\widetilde{G}_n)$ and every $x \in (a, b)$,

$$f(x) = f(c) + \int_c^{h(x)} H_n^* f(t)\, dp_2(t).$$

In particular,

$$\widetilde{D}T_n(x) = \widetilde{D}T_n(c) + \int_c^{h(x)} H_n^* \widetilde{D}T_n(t)\, dp_2(t). \tag{31}$$

From (30) and (31) we see that there are points $\eta_k \in (h(\xi_k), h(\xi_{k+1}))$, $0 \leq k \leq n-2$, such that

$$\operatorname{sign}\left[H_n^* \widetilde{D}T_n(\eta_k)\right] = \operatorname{sign}\left[\widetilde{D}T_n(\xi_{k+1}) - \widetilde{D}T_n(\xi_k)\right] = -\operatorname{sign}\left[\widetilde{D}T_n(\xi_k)\right] \tag{32}$$

(η_k can be chosen to be an interior point because $H_n^* T_n$ is continuous).

(iv) Since, by construction, no T_{n_j} has an alternation point in $[a_2, b_2]$ and since $[a_3, b_3] \subset [a_2, b_2]$, there is an integer k_0 (that depends on n, as do the alternation points y_k), such that $[a_3, b_3] \subset (y_{k_0}, y_{k_0+1})$. Assume for example that $T_n(y_{k_0}) = -T_n(y_{k_0+1}) = T_n(y_{k_0+2}) = 1$. From Corollary 1(b) we know that $T_n(x)$ must be strictly decreasing on $[y_{k_0}, y_{k_0+1}]$, and strictly increasing on $[y_{k_0+1}, y_{k_0+2}]$, and in particular on $[a_3, b_3]$. From Corollary 1(e) and (31) we deduce that $\widetilde{D}T_n(x) \le 0$ on $[a_3, b_3]$ and strictly negative on at least one point of this interval. Thus $\widetilde{D}T_n$ will have a pseudominimum $e < 0$ at some point $\nu \in [a_3, b_3]$ (which will depend on n), although $\widetilde{D}T_n$ may not have an absolute minimum on this interval (cf. Proposition 6). Moreover, since $\xi_k \in (y_k, y_{k+1})$, applying (32) we also conclude that $\operatorname{sign}\left[H_n^* \widetilde{D}T_n(\eta_{k_0})\right] > 0$.

(v) We now show that $\widetilde{D}T_n$ is monotonic on each of the intervals $[a_3, \nu]$ and $[\nu, b_3]$. Since $\widetilde{D}T_n$ has a pseudominimum at ν, we have to show that this function is decreasing on $[a_3, \nu]$ and increasing on $[\nu, b_3]$.

First note that $\widetilde{D}T_n$ cannot be constant on an interval I, for otherwise we would infer from Proposition 4 that there are constants α and β such that $T_n = \alpha + \beta g_1$ on I, i. e. $T_n - \alpha - \beta g_1 = 0$ on I. But the definition of T_n implies that $T_n \in S(G_n) \setminus S(G_{n-1})$, and we would have a contradiction.

Assume that $\nu < b_3$ and that $\widetilde{D}T_n$ is not increasing on $[\nu, b_3]$. In view of the remarks made in the previous paragraphs, we conclude that there exist points $\nu < z_0 < z_1 < z_2 \le b_3$ such that

$$e \le \widetilde{D}T_n(z_0) < \widetilde{D}T_n(z_1) \le 0 \quad \text{and} \quad e \le \widetilde{D}T_n(z_2) < \widetilde{D}T_n(z_1) \le 0. \tag{33}$$

We first consider the case $0 < k_0 < n-1$.

Since T_n is strictly increasing on $[y_{k_0+1}, \xi_{k_0+1}]$, from (29) we deduce that there is a point $z_3 \in (y_{k_0+1}, \xi_{k_0+1})$ such that $\widetilde{D}T_n(z_3) > 0$, i. e., $\widetilde{D}T_n(z_3) > \widetilde{D}T_n(z_2)$. Thus, applying (31) we deduce that there are points $\kappa_1, \kappa_2, \kappa_3$, such that

$$h(\xi_{k_0-1}) < h(y_{k_0}) \le h(a_3) \le h(\nu) < z_0 < \kappa_1 < z_1 < \kappa_2 < z_2 < \kappa_3 < z_3 < h(\xi_{k_0+1}),$$

and

$$1 = \operatorname{sign}\left[H_n^* \widetilde{D}T_n(\kappa_1)\right] = -\operatorname{sign}\left[H_n^* \widetilde{D}T_n(\kappa_2)\right] = \operatorname{sign}\left[H_n^* \widetilde{D}T_n(\kappa_3)\right] = \operatorname{sign}\left[H_n^* \widetilde{D}T_n(\eta_{k_0})\right]. \tag{34}$$

Since $\xi_{k_0} \in (y_{k_0}, y_{k_0+1})$, we have $\widetilde{D}T_n(\xi_{k_0}) \le 0$, and from (30) we deduce that $\widetilde{D}T_n(\xi_{k_0}) < 0$ and $\widetilde{D}T_n(\xi_{k_0-1}) > 0$. Since $\widetilde{D}T_n(z_1) \le 0$ and $\eta_{k_0-1} \in (h(\xi_{k_0-1}), h(\xi_{k_0}))$, using (31) we may redefine η_{k_0-1}, if necessary, so that $\eta_{k_0-1} \in (h(\xi_{k_0-1}), z_0)$. Similarly, we may redefine η_{k_0+1}, if necessary, so that $\eta_{k_0+1} \in (z_3, h(\xi_{k_0+1}))$.

Thus $\{\eta_0, \cdots, \eta_{k_0-1}, \kappa_1, \kappa_2, \kappa_3, \eta_{k_0+1}, \cdots, \eta_{n-2}\}$ is a strong alternation of length $n+1$ for $H_n^* \widetilde{D}T_n$. But the Lemma of [11] implies that $\{1, H_n^* \widetilde{D}u_2, \cdots, H_n^* \widetilde{D}u_n\}$ is a weak Markov system, and we have a contradiction.

If $k_0 = 0$, proceeding exactly as in the preceding case we see that there are points $\kappa_1, \kappa_2, \kappa_3$, such that

$$\nu < z_0 < \kappa_1 < z_1 < \kappa_2 < z_2 < \kappa_3 < z_3 < h(\xi_2),$$

and (34) is satisfied with $k_0 = 0$. Redefining η_1, if necessary, so that $\eta_1 \in (z_3, h(\xi_2))$, we see that $\{\kappa_1, \kappa_2, \kappa_3, \eta_1, \cdots, \eta_{n-2}\}$ is a strong alternation of length $n+1$ for $H_n^* \widetilde{D}T_n$, which yields a contradiction.

Finally, if $k_0 = n-2$, and bearing in mind the definition of pseudominimum, we see that there are points z_{-1}, z_0, z_1, z_2, with $h(\xi_{n-3}) < h(y_{n-2}) < h(a_3) \le z_{-1} < h(\nu) < z_0 < z_1 < z_2 \le h(b_3)$, such that (33) holds and $\widetilde{D}T_n(z_{-1}) > \widetilde{D}T_n(z_0)$. Using (31) we conclude that there are points $\kappa_0, \kappa_1, \kappa_2$, such that

$$h(\xi_{n-3}) < \kappa_0 < z_0 < \kappa_1 < z_1 < \kappa_2 < z_3 \le h(b_3)$$

and

$$\begin{aligned} 1 &= -\text{sign}\left[H_n^*\widetilde{D}T_n(\kappa_0)\right] = \text{sign}\left[H_n^*\widetilde{D}T_n(\kappa_1)\right] \\ &= -\text{sign}\left[H_n^*\widetilde{D}T_n(\kappa_2)\right] = \text{sign}\left[H_n^*\widetilde{D}T_n(\eta_{n-2})\right]. \end{aligned}$$

In this case $\{\eta_0, \cdots, \eta_{n-2}, \kappa_0, \kappa_1, \kappa_2\}$ is a strong alternation of length $n+1$ for $H_n^*\widetilde{D}T_n$, which also yields a contradiction.

The proof that $\widetilde{D}T_n$ is monotonic on $[a_3, \nu]$ is similar.

(vi) The set of pseudoextrema in $[a_3, b_3]$ of the functions $\widetilde{D}T_{n_{i,2}}$ is either finite or has a point of accumulation. Repeating the argument used in (i), we conclude that we may choose a subinterval $[a_4, b_4] \subset [a_3, b_3]$ having the property that either an infinite number of the $\widetilde{D}T_{n_{i,2}}$ are increasing thereon, or an infinite number of the $\widetilde{D}T_{n_{i,2}}$ are decreasing thereon. We will see that this is the desired subsequence $\{T_{n_k}\}$.

Let a_1, b_1, with $a_4 < a_1 < b_1 < b_4$, be arbitrary but fixed. Let us see that the functions $\widetilde{D}T_{n_k}$ are uniformly bounded in $[a_1, b_1]$. Assume the contrary. Then, for each integer $q > 0$ there is a $n_{k,q}$ such that

$$\|\widetilde{D}T_{n_{k,\,q}}\|_{[a_1,b_1]} \geq q.$$

If the functions $\widetilde{D}T_{n_k}$ are increasing, then either

$$\widetilde{D}T_{n_{k,\,q}}(x) \geq q, \quad x \in [b_1, b_4],$$

or

$$\widetilde{D}T_{n_{k,\,q}}(x) \leq -q, \quad x \in [a_4, a_1].$$

Since $\|T_{n_{k,\,q}}\| = 1$, applying Proposition 4 we see that, either

$$q[g_1(b_4) - g_1(b_1)] = \int_{b_1}^{b_4} q\, dg_1(t) \leq \int_{b_1}^{b_4} \widetilde{D}T_{n_{k,\,q}}(t)\, dg_1(t) = T_{n_{k,\,q}}(b_4) - T_{n_{k,\,q}}(b_1) \leq 2,$$

or

$$-2 \leq T_{n_{k,\,q}}(a_1) - T_{n_{k,\,q}}(a_4) = \int_{a_4}^{a_1} \widetilde{D}T_{n_{k,\,q}}(t)\, dg_1(t) \leq -\int_{a_4}^{a_1} q\, dg_1(t) = -q[g_1(a_1) - g_1(a_4)].$$

Since q is arbitrary and $g_1(x)$ is strictly increasing, this is a contradiction.

A similar argument is applied if the functions $\widetilde{D}T_{n_j}$ are decreasing □

The next theorem generalizes [3, Theorem 1]. Apart from using Lemma 3 instead of [3, Lemma 1], the proofs are similar.

Theorem 5. *Let G be an infinite normalized Markov system of continuous functions on a closed interval $[a, b]$. Then $S(G)$ is dense in $C[a, b]$ if and only if $S(G)$ has an everywhere unbounded Bernstein inequality.*

Proof. If $S(G)$ is not dense in $C[a, b]$, then Theorem 3, implies that $lim_{n\to\infty} M_n \neq 0$. Thus there is a sequence $\{T_{n_i}\}_{i\in\mathbb{Z}^+}$ of generalized Chebychev polynomials and a number $r > 0$, such that $M_{n_i} > r$.

Let $\alpha_{n_i}, \beta_{n_i}$ be two consecutive zeros of T_{n_i} such that $M_{n_i} = \beta_{n_i} - \alpha_{n_i}$; then $\{\alpha_{n_i}\}_{i\in\mathbb{Z}^+}$ has a point of accumulation in $[a, b]$. We may therefore choose a subinterval $[\alpha, \beta] \subset [a, b]$, and a subsequence $\left\{T_{n_i'}\right\}$ of $\{T_{n_i}\}$, that has no zeros in $[\alpha, \beta]$.

Applying Lemma 3 we thus conclude that there is an interval $[a_1, b_1] \subset [a, b]$, a sequence $\left\{T_{n_j}\right\}_{j\in\mathbb{Z}^+}$ of generalized Chebychev polynomials, and constants $\delta > 0$ and $\gamma > 0$, such that for any n_j

$$\|T_{n_j}\|_{[a_1,b_1]} < 1 - \delta$$

and

$$\|\widetilde{D}T_{n_j}\|_{[a_1,b_1]} < \gamma.$$

Let $f \in S(G)$; then $f \in S(G_n)$ for some n. If $n_j \geq n$ we have:
(i) For every $x \in [a_1, b_1]$

$$|T_{n_j}(x)| \leq \|T_{n_j}\|_{[a_1,b_1]} < 1 - \delta < 1.$$

Thus,

$$|T_{n_j}(x)| \neq 1 \tag{35}$$

and

$$\frac{1}{1 - |T_{n_j}(x)|} < \frac{1}{\delta}.$$

(ii) Since $f \in S(G_{n_j})$ and (35) holds, we may apply Theorem 4 to conclude that for every $x \in [a_1, b_1]$

$$\frac{|\widetilde{D}f(x)|}{\|f\|_{[a,b]}} \leq \frac{2}{1 - |T_{n_j}(x)|}|\widetilde{D}T_{n_j}(x)| \leq \frac{2}{\delta}\|\widetilde{D}T_{n_j}\|_{[a_1,b_1]} < \frac{2\gamma}{\delta}.$$

Thus $S(G)$ does not satisfy an unbounded Bernstein inequality.

Conversely, given $[\alpha, \beta] \subset [a, b]$ and $\gamma > 0$, let $a_1 = \alpha < b_1 \leq \beta$ be such that

$$g_1(b_1) - g_1(a_1) < \frac{1}{\gamma}, \tag{36}$$

and let $g \in C[a, b]$ be such that $g(t) = 0$ if $t \notin (a_1, b_1)$, $g((a_1 + b_1)/2) = 1$, and extended linearly to the rest of $[a, b]$. The density implies that for each $\varepsilon > 0$ there is a function $f \in S(G)$ such that $\|f - g\|_{[a,b]} < \varepsilon$.

Since $f(x) - f(a_1) \in S(G)$, we may assume without essential loss of generality that $f(a_1) = 0$. Thus, for ε sufficiently small there exists a point $x_0 \in (a_1, b_1)$ such that

$$f(x_0) = \|f\|_{[\alpha,\beta]} = \|f\|_{[a,b]}.$$

Therefore

$$\|f\|_{[a,b]} = f(x_0) = f(x_0) - f(a_1) = \int_{a_1}^{x_0} \widetilde{D}f(t)\, dg_1(t) \leq$$

$$\leq [g_1(x_0) - g_1(a_1)]\|\widetilde{D}f\|_{[\alpha,\beta]} \leq [g_1(b_1) - g_1(a_1)]\|\widetilde{D}f\|_{[\alpha,\beta]},$$

and applying (36) we conclude that

$$\|f\|_{[a,b]} \leq \frac{\|\widetilde{D}f\|_{[\alpha,\beta]}}{\gamma}.$$

□

5. A Jackson Type Theorem for finite Markov systems

The main result of this section is Theorem 6. It generalizes [1, Theorem 1], which relies on the additional assumptions that the functions in the Markov system are differentiable, and that the system of derivatives is also a Markov system. Here we simplify Bojanov's proof in such a way that the differentiation operator is not used at all.

Given a Markov system G_n in $C[a, b]$ and a function $f \in C[a, b]$, $E_n(f)$ and $P_n(f, \cdot)$ will respectively denote the distance from f to $S(G_n)$ and the element of best approximation from $S(G_n)$ to f, in the norm of the supremum.

Let ω be an increasing function on $[a, b]$, $\mathcal{A}$ a set of square–integrable functions with respect to $d\omega$ on $[a, b]$, and ψ a square–integrable function with respect to $d\omega$ on $[a, b]$.

We will say that ψ is ω*–orthogonal* to $\mathcal{A}$, and write $\psi \perp \mathcal{A}$, if ψ is orthogonal to $\mathcal{A}$ in the inner product space defined by $d\omega$, i. e,. if for every $\varphi \in \mathcal{A}$

$$\int_a^b \psi(t)\varphi(t)\, d\omega(t) = 0.$$

We will say that a function $f \in F[a,b]$ has a sign change at x_0 if there is a $\delta > 0$ with $a < x_0 - \delta < x_0 + \delta < b$ such that, if $x_0 - \delta < t_1 < x_0 < t_2 < x_0 + \delta$, then $\operatorname{sign}[f(t_1)f(t_2)] = -1$.

We begin with two preliminary propositions:

Lemma 4. *Let $\{g_0, \ldots g_n\}$ be a Markov system on a bounded interval $[a,b]$, ω a strictly increasing function on $[a,b]$, and $\psi \in F[a,b]$. Assume, moreover, that $\psi(x) \neq 0$ a. e., and that ψ is ω-orthogonal to $S(G_n)$. Then ψ has at least n sign changes in $[a,b]$.*

Proof. Assume ψ has $p < n$ sign changes. Then there is a partition

$$a = t_0 < \ldots < t_p < t_{p+1} = b$$

such that $(-1)^k \psi(x)$ has weakly constant sign on each interval $(t_k, t_{k+1}), 0 \leq k \leq p$.

Let

$$\varphi(t) := \bigcup \begin{pmatrix} g_0 & \ldots & g_p & g_{p+1} \\ t_0 & \ldots & t_p & t \end{pmatrix} = \det \begin{pmatrix} g_0(t_0) & \ldots & g_0(t_p) & g_0(t) \\ \vdots & \ddots & \vdots & \vdots \\ g_p(t_0) & \ldots & g_p(t_p) & g_p(t) \\ g_{p+1}(t_0) & \ldots & g_{p+1}(t_p) & g_{p+1}(t) \end{pmatrix}.$$

Then

$$\varphi = \sum_{k=0}^{p+1} \alpha_k g_k,$$

and therefore:

(i) $\varphi \in S_n$.

(ii) $\alpha_{p+1} = \bigcup \begin{pmatrix} g_0 & \ldots & g_p \\ t_0 & \ldots & t_p \end{pmatrix} > 0$ (whence $\varphi \neq 0$).

(iii) $(-1)^{p-k}\varphi(t) > 0$ if $0 \leq k \leq p$ and $t_k < t < t_{k+1}$.

Thus $\psi\varphi$ has weakly constant sign on (a,b). Since $\psi\varphi \neq 0$ a. e., we conclude that

$$\int_a^b \psi(t)\varphi(t) d\,\omega(t) \neq 0,$$

which is a contradiction. □

Lemma 5. *Assume that G_n is a* normalized *Markov system on $C[a,b]$, let*

$$a \leq x_0 < \ldots < x_{n+1} \leq b, \tag{37}$$

and let $\mathbf{D}$ denote the array

$$\{g_i(x_{j+1}) - g_i(x_j); 1 \leq i \leq n, 0 \leq j \leq n\}.$$

Let $\mathbf{D}_k$ be obtained from $\mathbf{D}$ by deleting the k^{th} column, and $d_k := \det \mathbf{D}_k$. Then $d_k \geq 0$, $0 \leq k \leq n-1$, and $d_n > 0$.

Proof. First note that

$$d_n = \bigcup \begin{pmatrix} 1 & g_1 & \ldots & g_n \\ x_0 & x_1 & \ldots & x_n \end{pmatrix} > 0.$$

Since $G_n \subset C[a,b]$, to complete the proof it will suffice to assume that $a < x_0$ and $x_n < b$. We will prove that under these conditions all the d_k are strictly positive.

If $f \in S(G_n)$, then clearly

$$\bigcup \begin{pmatrix} 1 & g_1 & \dots & g_n & f \\ x_0 & x_1 & \dots & x_n & x_{n+1} \end{pmatrix} = 0$$

We now used an argument that was employed in [8, Chapter 1]: subtracting from each column in this determinant the preceding one and then developing by the first row, we thus obtain:

$$\begin{aligned} 0 &= \det \begin{pmatrix} 1 & 0 & \dots & 0 \\ g_1(x_0) & g_1(x_1)-g_1(x_0) & \dots & g_1(x_{n+1})-g_1(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ g_n(x_0) & g_n(x_1)-g_n(x_0) & \dots & g_n(x_{n+1})-g_n(x_n) \\ f(x_0) & f(x_1)-f(x_0) & \dots & f(x_{n+1})-f(x_n) \end{pmatrix} = \\ &= \det \begin{pmatrix} g_1(x_1)-g_1(x_0) & \dots & g_1(x_{n+1})-g_1(x_n) \\ \vdots & \ddots & \vdots \\ g_n(x_1)-g_n(x_0) & \dots & g_n(x_{n+1})-g_n(x_n) \\ f(x_1)-f(x_0) & \dots & f(x_{n+1})-f(x_n) \end{pmatrix} = \qquad (38) \\ &= (-1)^n \sum_{k=0}^{n} (-1)^k \left[f(x_{k+1}) - f(x_k) \right] d_k. \end{aligned}$$

Setting

$$\psi(t) := \begin{cases} (-1)^k d_k & \text{if} \quad x_k \le t < x_{k+1}, \quad 0 \le k \le n-1, \\ (-1)^n d_n & \text{if} \quad x_n \le t \le x_{n+1}, \end{cases}$$

and taking into account that

$$f(x_{k+1}) - f(x_k) = \int_{x_k}^{x_{k+1}} \widetilde{D} f(t)\, dg_1(t),$$

we obtain

$$\int_{x_0}^{x_{n+1}} \psi(t) \widetilde{D} f(t)\, dg_1(t) = \sum_{k=0}^{n} (-1)^k \left[f(x_{k+1}) - f(x_k) \right] d_k = 0.$$

Thus, if $S(\widetilde{G}_n, x_0, x_n)$ denotes the restrictions to $[x_0, x_n]$ of the functions in $S(\widetilde{G}_n)$, we have shown that

$$f \perp S(\widetilde{G}_n, x_0, x_n).$$

Since ψ has constant sign on each interval (x_{j-1}, x_j), and moreover, it cannot have a sign change neither at x_0 nor at x_{n+1}, applying Lemma 4 we conclude that the d_k all have the same sign. Since $d_n > 0$, the assertion follows.

□

In view of the preceding lemma, we may define

$$a_k := \frac{d_k}{2 \sum_{j=0}^{n} d_j}, \quad 0 \le k \le n. \qquad (39)$$

If

$$\delta_n := \sup \left\{ \sum_{k=0}^{n} a_k \left[g_1(x_{k+1}) - g_1(x_k) \right] \right\},$$

where the supremum is taken over the set of all sequences of the form (37), we have

Theorem 6. *Let G_n be a normalized Markov system on $C[a,b]$. Then, for any function $f \in C([a,b])$,*

$$E_n(f) \leq \frac{3}{2}\omega(f \circ g_1^{-1}; \delta_n).$$

Proof. Let $x_0 < \ldots < x_{n+1}$ be the alternation points for $f - P_n(f;\cdot)$, i. e.

$$f(x_i) - P_n(f;x_i) = \varepsilon(-1)^i E_n(f), \quad 0 \leq i \leq n+1,$$

where ε is a constant and $\varepsilon = \pm 1$, and let $c_k := -\varepsilon a_k$, $0 \leq k \leq n$. Then, proceeding exactly as in the proof of [1, Theorem 1], we see that

$$E_n(f) = \sum_{k=0}^{n}(-1)^k c_k[f(x_{k+1}) - f(x_k)] = -\varepsilon \sum_{k=0}^{n}(-1)^k a_k[f(x_{k+1}) - f(x_k)]. \tag{40}$$

But, since $d_k \geq 0$, we see that also $a_k \geq 0$. Moreover, it readily follows from the definition of the a_k, that

$$\sum_{k=0}^{n} a_k = \frac{1}{2}.$$

Since $\omega(f;\lambda\delta) \leq (\lambda+1)\omega(f;\delta)$, $\lambda > 0$, and, moreover, $g_1(x)$ is strictly increasing, (40) implies that for each $\delta > 0$ we have:

$$\begin{aligned} E_n(f) &\leq \sum_{k=0}^{n} a_k |f(x_{k+1}) - f(x_k)| \\ &\leq \sum_{k=0}^{n} a_k \omega(f \circ g_1^{-1}; g_1(x_{k+1}) - g_1(x_k)) \\ &\leq \sum_{k=0}^{n} a_k \left[\frac{g_1(x_{k+1}) - g_1(x_k)}{\delta} + 1\right] \omega(f \circ g_1^{-1}; \delta)) \\ &= \left\{\frac{1}{\delta}\sum_{k=0}^{n} a_k \left[g_1(x_{k+1}) - g_1(x_k)\right] + \frac{1}{2}\right\} \omega(f \circ g_1^{-1}; \delta). \end{aligned}$$

Setting $\delta = \delta_n$, the assertion follows. □

Lemma 6. *Let G_n, $n \geq 1$ be a normalized Markov system of continuous functions on $[a,b]$, and let*

$$\sigma_t(s) := \begin{cases} 0 & \text{if } \ a \leq s \leq t \\ g_1(s) - g_1(t) & \text{if } \ t < s \leq b. \end{cases}$$

Then $E_n(\sigma_t)$ is a continuous real–valued function of t, with domain $[a,b]$.

Proof. Let $\theta_n(t;\cdot) := P_n(\sigma_t;\cdot)$. Then

$$E_n(\sigma_t) = \|\sigma_t - \theta_n(t;.)\|.$$

If $t_1, t_2 \in [a,b]$ we have

$$\|\sigma_{t_1} - \theta_n(t_1;.)\| \leq \|\sigma_{t_1} - \theta_n(t_2;.)\| \leq \|\sigma_{t_1} - \sigma_{t_2}\| + \|\sigma_{t_2} - \theta_n(t_2;.)\|,$$

whence

$$E_n(\sigma_{t_1}) - E_n(\sigma_{t_2}) \leq \|\sigma_{t_1} - \sigma_{t_2}\|.$$

By a similar argument we obtain

$$E_n(\sigma_{t_2}) - E_n(\sigma_{t_1}) \leq \|\sigma_{t_2} - \sigma_{t_1}\|.$$

Therefore

$$|E_n(\sigma_{t_1}) - E_n(\sigma_{t_2})| \leq \|\sigma_{t_1} - \sigma_{t_2}\|.$$

Assume that $t_1 < t_2$ and let $s \in [a,b]$. Then, if $t_2 < s$,

$$\sigma_{t_1}(s) - \sigma_{t_2}(s) = g_1(s) - g_1(t_1) - [g_1(s) - g_1(t_2)] = g_1(t_2) - g_1(t_1).$$

On the other hand, if $t_1 < s \leq t_2$,

$$\sigma_{t_1}(s) - \sigma_{t_2}(s) = g_1(s) - g_1(t_1) \leq g_1(t_2) - g_1(t_1).$$

Finally, if $s \leq t_1$,

$$\sigma_{t_1}(s) - \sigma_{t_2}(s) = 0 \leq g_1(t_2) - g_1(t_1).$$

Therefore

$$\|\sigma_{t_1} - \sigma_{t_2}\| \leq \|g_1(t_1) - g_1(t_2)\|.$$

Since g_1 is continuous, the assertion follows. □

In view of the preceding lemma, we may define

$$\Delta_n := \max_{t\in[a,b]} E_n(\sigma_t) = \max_{t\in[a,b]} \|\sigma_t - \theta_n(t;.)\|.$$

Using the operator $\widetilde{D}$ we obtain an estimate for δ_n:

Theorem 7. *Let G_n, $n \geq 1$, be a normalized Markov system of continuous functions on $[a,b]$. Then*

$$\delta_n \leq \sqrt{[g_1(b) - g_1(a)]\,\Delta_n}$$

Proof. For $a \leq x_0 < \ldots < x_{n+1} \leq b$, let the coefficients a_k be defined as in (39), and let $\psi : [a,b] \to I\!R$ be defined by

$$\psi(t) = \begin{cases} (-1)^k a_k & \text{if} \quad t \in (x_k, x_{k+1}], \quad 0 \leq k \leq n \\ 0 & \text{if} \quad t \in [a, x_0] \cup (x_{n+1}, b] \end{cases},$$

where we adopt the convention that $(\alpha,\alpha] = \emptyset$. If $f \in S(G_n)$, proceeding as in the proof of Lemma 5 we have:

$$0 = \bigcup \begin{pmatrix} 1 & g_1 & \ldots & g_n & f \\ x_0 & x_1 & \ldots & x_n & x_{n+1} \end{pmatrix} = (-1)^n \sum_{k=0}^{n} (-1)^k [f(x_{k+1}) - f(x_k)]\; d_k,$$

and an application of Proposition 4 yields

$$\begin{aligned} \int_a^b \psi(s)\widetilde{D}f(s)\,dg_1(s) &= \int_{x_0}^{x_{n+1}} \psi(s)\widetilde{D}f(s)\,dg_1(s) \\ &= -\frac{\varepsilon}{2\sum_{k=0}^{n} d_k} \sum_{k=0}^{n} (-1)^k [f(x_{k+1}) - f(x_k)]\; d_k = 0. \end{aligned}$$

In particular, for every $t \in [a,b]$

$$\int_a^b \psi(s)\widetilde{D}\theta_n(t;s)\,dg_1(s) = 0.$$

We now extend the definition of $\widetilde{D}$ to the linear space generated by G_n and the functions $\sigma_t(s)$, $s \in (a,b)$, by stipulating that it should be a linear operator and defining

$$\widetilde{D}\sigma_t(s) := \begin{cases} 0 & \text{if} \quad s \leq t \\ 1 & \text{if} \quad s > t \end{cases}.$$

Since

$$\int_{x_1}^{x_2} \widetilde{D}\sigma_t(s)\,dg_1(s) = \sigma_t(x_2) - \sigma_t(x_1),$$

we see that the conclusions of Proposition 4 are valid for this extended operator. Thus, setting $f_t(s) := \sigma_t(s) - \theta_n(t;s)$, we obtain

$$\begin{aligned} \int_a^b \psi(s)\widetilde{D}\sigma_t(s)\,dg_1(s) &= \int_a^b \psi(s)\widetilde{D}[\sigma_t(s) - \theta_n(t;s)]\,dg_1(s) = \\ \sum_{k=0}^{n}(-1)^k a_k \int_{x_k}^{x_{k+1}} \widetilde{D}f_t(s)\ dg_1(s) &= \sum_{k=0}^{n}(-1)^k a_k[f_t\,(x_{k+1}) - f_t\,(x_k)]. \end{aligned}$$

Therefore

$$\left|\int_a^b \psi(s)\widetilde{D}\sigma_t(s)\,dg_1(s)\right| \le \sum_{k=0}^{n} |a_k|\,|f_t\,(x_{k+1}) - f_t\,(x_k)| \le 2\Delta_n \sum_{k=0}^{n} |a_k| = \Delta_n.$$

On the other hand, if we define

$$a'_0 := a_0, \quad a'_{n+1} := (-1)^{n+1}a_n, \quad a'_k := (-1)^k(a_{k-1} + a_k), \quad 1 \le k \le n,$$

we obtain the following representation for ψ:

$$\psi(s) = \sum_{k=0}^{n+1} a'_k\widetilde{D}\sigma_{x_k}(s).$$

This implies that

$$\begin{aligned} \int_a^b [\psi(s)]^2\ dg_1(s) &= \int_a^b \psi(s)\left[\sum_{k=0}^{n+1} a'_k\widetilde{D}\sigma_{x_k}(s)\right]\ dg_1(s) = \\ &= \sum_{k=0}^{n+1} a'_k \int_a^b \psi(s)\widetilde{D}\sigma_{x_k}(s)\,dg_1(s). \end{aligned}$$

Therefore

$$\begin{aligned} \int_a^b [\psi(s)]^2\ dg_1(s) &\le \sum_{k=0}^{n+1} |a'_k| \left|\int_a^b \psi(s)\widetilde{D}\sigma_{x_k}(s)\,dg_1(s)\right| \le \\ &\le \Delta_n \sum_{k=0}^{n+1} |a'_k| \le \Delta_n \left[2\sum_{k=0}^{n} |a_k|\right] = \Delta_n, \end{aligned}$$

and applying the Cauchy–Schwarz inequality we obtain:

$$\begin{aligned} \sum_{k=0}^{n} a_k[g_1(x_{k+1}) - g_1(x_k)] &= \left|\int_a^b \psi(s)\,dg_1(s)\right| \\ &\le \left[\int_a^b dg_1(s)\right]^{\frac{1}{2}} \left[\int_a^b [\psi(s)]^2\ dg_1(s)\right]^{\frac{1}{2}} \\ &\le \sqrt{[g_1(b) - g_1(a)]\,\Delta_n}. \end{aligned}$$

Taking the supremum over all sequences of the form (37), the conclusion follows. □

Acknowledgements: The second author is grateful to Professors Peter Borwein and Borislav Bojanov, for kindly making available reprints and preprints of their work.

References

[1] Bojanov, B., A Jackson type theorem for Tchebycheff systems, *Math. Balkanika (NS)* 10, 73–82 (1996).

[2] Borwein, P., Zeros of Chebyshev Polynomials in Markov Systems, *J. Approximation Theory* 63, 56-64 (1990).

[3] Borwein, P. and Erdélyi, T., Dense Markov Spaces and Unbounded Bernstein Inequalities, *J. Approximation Theory* 83, 66-77 (1995).

[4] Borwein, P. and Erdélyi, T., *Polynomials and Polynomial Inequalities*, Springer-Verlag, NY, 1995.

[5] E. W. Cheney, *Introduction to Approximation Theory*, McGraw–Hill, NY, 1966.

[6] J. M. Carnicer and J. M. Peña, Tchebycheff spaces and total positivity, in *Proceedings of the International Workshop on Total Positivity and its Applications* (M. Gasca and C. A. Micchelli eds.), Kluwer Academic Publishers, Dordrecht, 1996, pp. 277–300.

[7] J. M. Carnicer, J. M. Peña, and R. A. Zalik, Strictly totally positive systems, *J. Approximation Theory* 92, 411–441 (1998).

[8] S. Karlin and W. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics*, Interscience, NY, 1966.

[9] L. L. Schumaker, *Spline Functions: Basic Theory,* Wiley, New York, 1981.

[10] R. A. Zalik, Integral representation and embedding of weak Markov systems, *J. Approx. Theory* 58, 1–11 (1989).

[11] R. A. Zalik, Integral representation of Markov systems and the existence of adjoined functions for Haar spaces, *J. Approx. Theory* 65, 22–31 (1991).

[12] R. A. Zalik and D. Zwick, Some properties of Markov systems, *J. Approx. Theory* 65, 32–45 (1991).

[13] R. A. Zalik, Čebyšev and weak Čebyšev systems, in *Proceedings of the International Workshop on Total Positivity and its Applications* (M. Gasca and C. A. Micchelli eds.), Kluwer Academic Publishers, Dordrecht, 1996, pp. 300–332.

[14] R. A. Zalik, Interpolation, the Budan–Fourier theorem, and Descartes' rule of signs for Markov systems, in *Approximation Theory: In Memory of A.K. Varma*, (N.K. Govil, R.N. Mohapatra, Z. Nashed, A. Sharma and J. Szabados, eds.), Marcel Dekker, New York, May 1998, pp. 499–511.

[15] R. Zielke, Alternation properties of Tchebyshev–systems and the existence of adjoined functions, *J. Approx. Theory* 10, 172–184 (1974).

[16] R. Zielke, *Discontinuous Čebyšev systems*, Lecture Notes in Mathematics Vol. 707, Springer–Verlag, NY, 1979.

[17] R. Zielke, Relative differentiability and integral representation of a class of weak Markov systems, *J. Approx. Theory* 44, 30–42 (1985).