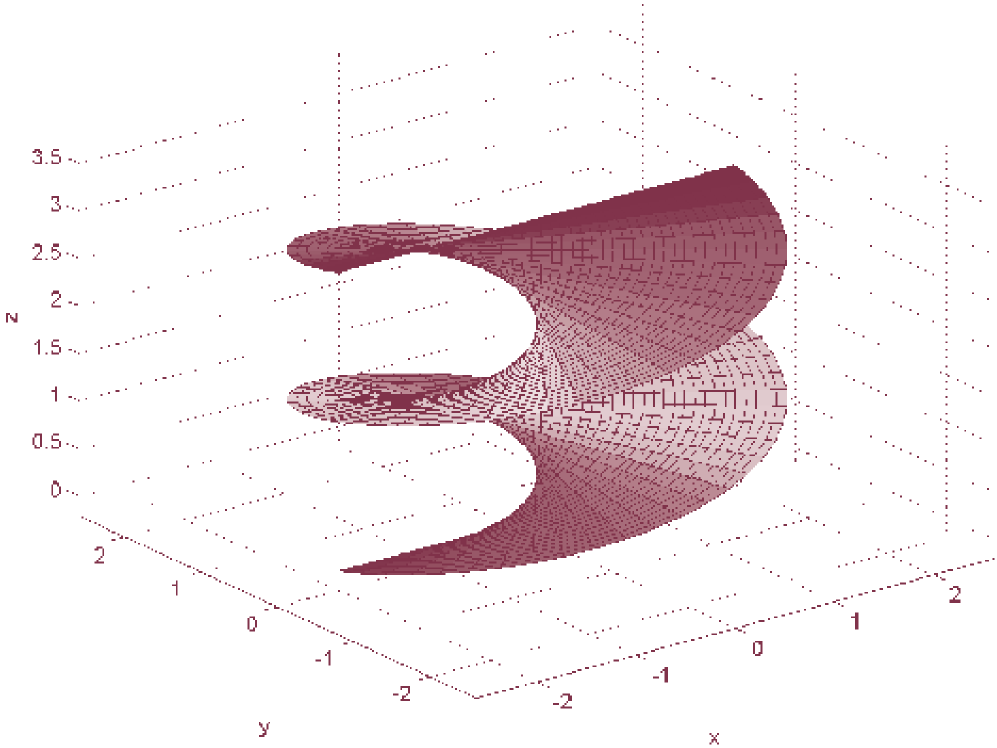
STUDIA UNIVERSITATIS
BABEŞ-BOLYAI



MATHEMATICA

2/2019

STUDIA
UNIVERSITATIS BABEŞ-BOLYAI
MATHEMATICA

2/2019

YEAR	**(LXIV) 2019**
MONTH	**JUNE**
ISSUE	**2**

S T U D I A
UNIVERSITATIS BABEŞ-BOLYAI

MATHEMATICA
2

Redacţia: 400084 Cluj-Napoca, str. M. Kogălniceanu nr. 1
Telefon: 0264 405300

CONTENTS

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 139–150
DOI: 10.24193/subbmath.2019.2.01

Possibly infinite generalized iterated function systems comprising φ-max contractions

Silviu-Aurelian Urziceanu

Abstract. One way to generalize the concept of iterated function system was proposed by R. Miculescu and A. Mihail under the name of generalized iterated function system (for short GIFS). More precisely, given $m \in \mathbb{N}^*$ and a metric space (X, d), a generalized iterated function system of order m is a finite family of functions $f_1, \ldots, f_n : X^m \to X$ satisfying certain contractive conditions. Another generalization of the notion of iterated function system, due to F. Georgescu, R. Miculescu and A. Mihail, is given by those systems consisting of φ-max contractions. Combining these two lines of research, we prove that the fractal operator associated to a possibly infinite generalized iterated function system comprising φ-max contractions is a Picard operator (whose fixed point is called the attractor of the system). We associate to each possibly infinite generalized iterated function system comprising φ-max contractions $\mathcal{F}$ (of order m) an operator $H_{\mathcal{F}} : \mathcal{C}^m \to \mathcal{C}$, where $\mathcal{C}$ stands for the space of continuous and bounded functions from the shift space on the metric space corresponding to the system. We prove that $H_{\mathcal{F}}$ is a Picard operator whose fixed point is the canonical projection associated to $\mathcal{F}$.

Mathematics Subject Classification (2010): 28A80, 37C70, 41A65, 54H25.

Keywords: Possibly infinite generalized iterated function system, φ-max contraction, attractor, canonical projection.

1. Introduction

One way to generalize the concept of iterated function system was proposed by R. Miculescu and A. Mihail (see [6] and [8]) under the name of generalized iterated function system. More precisely, given $m \in \mathbb{N}^*$ and a metric space (X, d), a generalized iterated function system (for short a GIFS) of order m is a finite family of functions $f_1, \ldots, f_n : X^m \to X$ satisfying certain contractive conditions.

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

They proved that there exists a unique attractor of a GIFS, studied some of its properties and provided examples showing that GIFSs are real generalizations of iterated function systems. In addition, F. Strobin (see [13]) proved that, for any $m \in \mathbb{N}$, $m \geq 2$, there exists a Cantor subset of the plane which is the attractor of some GIFS of order m, but is not the attractor of any GIFS of order $m-1$. This kind of iterated function system was generalized in several ways (see [1], [2], [10], [12], [14] and [15]). In addition, the Hutchinson measure associated with a GIFS was studied in [7] (for GIFS with probabilities), in [4] (for generalized iterated function systems with place dependent probabilities) and in [11]

Another generalization of the notion of iterated function system in given by those systems consisting of φ-max-contractions (see [3]).

Combining these lines of research, we prove that the fractal operator associated to a possibly infinite generalized iterated function system comprising φ-max contractions is a Picard operator (whose fixed point is called the attractor of the system).

The main tool in the study of topological properties of the attractor of an iterated function system is the canonical projection. Paper [9] inspired us to associate to each possibly infinite generalized iterated function system comprising φ-max contractions $\mathcal{F}$ (of order m) an operator $H_{\mathcal{F}} : \mathcal{C}^m \to \mathcal{C}$, where $\mathcal{C}$ stands for the space of continuous and bounded functions from the shift space on the metric space corresponding to the system. We prove that $H_{\mathcal{F}}$ is a Picard operator whose fixed point is the canonical projection associated to $\mathcal{F}$.

2. Preliminaries

For a metric space (X,d) and $m \in \mathbb{N}^*$, we consider:

- $P_{b,cl}(X)$ the set of all non-empty, bounded and closed subsets of X;
- the Hausdorff-Pompeiu metric $h : P_{b,cl}(X) \times P_{b,cl}(X) \to [0,\infty)$ given by

$$h(A,B) = \max\{d(A,B), d(B,A)\},$$

for every $A, B \in P_{b,cl}(X)$, where $d(A,B) = \sup_{x\in A} \inf_{y\in B} d(x,y)$;

- the Cartesian product X^m endowed with the maximum metric $d_{\max}$ defined by

$$d_{\max}((x_1,\ldots,x_m),(y_1,\ldots,y_m)) = \max\{d(x_1,y_1),\ldots,d(x_m,y_m)\},$$

for all $(x_1,\ldots,x_m),(y_1,\ldots,y_m) \in X^m$;

- the spaces X_1, X_2, $\ldots$, X_k, $\ldots$, defined inductively in the following way:

$$X_1 = \underbrace{X \times X \times \ldots \times X}_{m \text{ times}} = X^m$$

and

$$X_{k+1} = \underbrace{X_k \times X_k \times \ldots \times X_k}_{m \text{ times}}$$

for every $k \in \mathbb{N}^*$. We endow X_k with the maximum metric for every $k \in \mathbb{N}^*$. Note that X_k is isometric to X^{m^k} with the maximum metric for every $k \in \mathbb{N}^*$ and that we will identify X_k and X^{m^k};

- $\mathcal{F}_i^p = \{\sigma : \{1,2,\ldots,m^i\} \to \{1,2,\ldots,m^p\}\}$, where $p \in \mathbb{N}^*$ and $i \in \{0,1,\ldots,p-1\}$

• $x_\sigma = (x_{\sigma(1)}, \ldots, x_{\sigma(m^i)})$ and $y_\sigma = (y_{\sigma(1)}, \ldots, y_{\sigma(m^i)})$, where $x = (x_1, x_2, \ldots, x_{m^p})$, $y = (y_1, y_2, \ldots, y_{m^p}) \in X^{m^p}$, $p \in \mathbb{N}^*$, $i \in \{0, 1, \ldots, p-1\}$ and $\sigma \in \mathcal{F}_i^p$.

Definition 2.1. A possibly infinite generalized iterated function system of order $m \in \mathbb{N}^*$ is a pair $\mathcal{F} = ((X,d),(f_i)_{i \in I})$, where (X,d) is a metric space, $f_i : X^m \to X$ is continuous for every $i \in I$ and the family of functions $(f_i)_{i \in I}$ is bounded (i.e. $\underset{i \in I}{\cup} f_i(B)$ is bounded for each bounded subset B of X^m).

The function $F_{\mathcal{F}} : (P_{b,cl}(X))^m \to P_{b,cl}(X)$, described by

$$F_{\mathcal{F}}(B_1, \ldots, B_m) = \overline{\underset{i \in I}{\cup} f_i(B_1 \times \ldots \times B_m)},$$

for all $(B_1, \ldots, B_m) \in (P_{b,cl}(X))^m$, is called the fractal operator associated to $\mathcal{F}$.

If there exists a unique $A \in P_{b,cl}(X)$ such that $F_{\mathcal{F}}(A, \ldots, A) = A$, then we say that $\mathcal{F}$ has attractor and A, which is denoted by $A_{\mathcal{F}}$, is called the attractor of $\mathcal{F}$.

Now we recall the concept of code space associated to a possibly infinite generalized iterated function system which was considered by A. Mihail and F. Strobin & J. Swaczyna.

Let us consider $m \in \mathbb{N}^*$ and a set I. One can define inductively the sets $\Omega_1, \Omega_2, \ldots, \Omega_k, \ldots$ in the following way:

$$\Omega_1 = I \text{ and } \Omega_{k+1} = \underbrace{\Omega_k \times \Omega_k \times \ldots \times \Omega_k}_{m \text{ times}},$$

for every $k \in \mathbb{N}^*$.

We are also dealing in the sequel with the following sets:

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_k \times \ldots$$

and

$${}_k\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_k,$$

where $k \in \mathbb{N}^*$.

For $i \in \{1, 2, \ldots, m\}$, $k \in \mathbb{N}$, $k \geq 2$ and $\alpha = \alpha^1 \alpha^2 \ldots \alpha^k \in {}_k\Omega$, where

$$\alpha^2 = \alpha_1^2 \alpha_2^2 \ldots \alpha_m^2 \in \Omega_2, \ldots, \alpha^k = \alpha_1^k \alpha_2^k \ldots \alpha_m^k \in \Omega_k,$$

we consider

$$\alpha(i) = \alpha_i^2 \alpha_i^3 \ldots \alpha_i^k \in {}_{k-1}\Omega.$$

For $\alpha \in \Omega$ and $i \in \{1, 2, \ldots, m\}$, $\alpha(i) \in \Omega$ could be similarly defined in a similar manner.

Definition 2.2. Ω is called the Mihail-Strobin&Swaczyna generalized code space.

Ω becomes a complete metric space if it is furnished with the metric d given by

$$d(\alpha, \beta) = \sum_{k \in \mathbb{N}} C^k d(\alpha^k, \beta^k),$$

for every $\alpha = \alpha^1 \alpha^2 \ldots \alpha^i \alpha^{i+1} \ldots$, $\beta = \beta^1 \beta^2 \ldots \beta^i \beta^{i+1} \ldots \in \Omega$, where

$$d(\alpha^k, \beta^k) = \begin{cases} 1, & \alpha^k \neq \beta^k \\ 0, & \alpha^k = \beta^k \end{cases}$$

and $C \in (0,1)$. Moreover, the metric space (Ω, d) is compact provided that I is finite.

To a possibly infinite generalized iterated function system $\mathcal{F} = ((X,d),(f_i)_{i\in I})$ of order m, one can associate, for every $k \in \mathbb{N}^*$, a family of functions

$$\{f_\alpha : X_k \to X \mid \alpha \in_k \Omega\}$$

defined inductively in the following way:

i) For $k = 1$, the family is $(f_i)_{i\in I}$.

ii) If the functions f_α, where $\alpha \in\ _k\Omega$, have been defined, then, we set

$$f_\alpha(x_1, x_2, \dots, x_m) = f_{\alpha^1}(f_{\alpha(1)}(x_1), \dots, f_{\alpha(m)}(x_m))$$

for every $\alpha = \alpha^1\alpha^2 \dots \alpha^k\alpha^{k+1} \in\ _{k+1}\Omega$,

$$(x_1, x_2, \dots, x_m) \in X_{k+1} = \underbrace{X_k \times X_k \times \dots \times X_k}_{m \text{ times}}.$$

Note that if $m = 1$, then $_k\Omega = I^k$ and if $\omega = \omega^1\omega^2 \dots \omega^k \in\ _k\Omega$, then

$$f_\omega = f_{\omega^1} \circ \dots \circ f_{\omega^k}.$$

Hence the above introduced families of functions are natural generalizations of compositions of functions.

Given a set X, $m \in \mathbb{N}^*$ and a function $f : X^m \to X$, we define inductively a family of functions $f^{[k]} : X^{m^k} \to X$, $k \in \mathbb{N}^*$, in the following way:

i) $f^{[1]} = f$;

ii) assuming that we have defined $f^{[k]}$, then

$$f^{[k+1]}(x_1, \dots, x_m) = f(f^{[k]}(x_1), \dots, f^{[k]}(x_m)),$$

for every $(x_1, \dots, x_m) \in \underbrace{X^{m^k} \times \dots \times X^{m^k}}_{m \text{ times}} = X^{m^{k+1}} = X_{k+1}$.

Note that for $m = 1$, we have $f^{[k]} = \underbrace{f \circ \dots \circ f}_{k \text{ times}}$. We remark that maps $f^{[k]}$ are special cases of f_α defined earlier.

Definition 2.3. Given a set X, $m \in \mathbb{N}^*$ and a function $f : X^m \to X$, an element x of X such that $f(x, \dots, x) = x$ is called a fixed point of f.

Definition 2.4. Given a metric space (X,d) and $m \in \mathbb{N}^*$, a function $f : X^m \to X$ is called contraction if there exists $C \in [0,1)$ such that $d(f(x), f(y)) \le C d_{\max}(x,y)$ for all $x, y \in X^m$.

Definition 2.5. A function $\varphi : [0,\infty) \to [0,\infty)$ is called comparison function provided that it satisfies the following properties:

i) it is nondecreasing;

ii) it is right-continuous;

iii) $\varphi(t) < t$ for every $t > 0$.

Definition 2.6. a) Given a metric space (X,d), $m \in \mathbb{N}^*$ and a comparison function φ, a function $f : X^m \to X$ is called φ-contraction if $d(f(x), f(y)) \le \varphi(d_{\max}(x,y))$ for all $x, y \in X^m$.

b) Given a metric space (X,d), a comparison function φ and $m \in \mathbb{N}^*$, a function $f : X^m \to X$ is called φ-max generalized contraction if there exists $p \in \mathbb{N}^*$ such that

$$d(f^{[p]}(x), f^{[p]}(y)) \le \varphi(\max\{\max_{\sigma\in\mathcal{F}_i^p} d(f^{[i]}(x_\sigma), f^{[i]}(y_\sigma)) \mid i \in \{0,1,2,\dots,p-1\}\}),$$

for all $x, y \in X^{m^p}$.

Now let us introduce an important tool that will be used in this paper, namely the operator $H_{\mathcal{F}}$ associated to a generalized possibly infinite generalized iterated function system $\mathcal{F}$.

To a possibly infinite generalized iterated function system $\mathcal{F} = ((X,d),(f_i)_{i\in I})$ of order m, we associate the operator $H_{\mathcal{F}} : \mathcal{C}^m \to \mathcal{C}$ described by

$$H_{\mathcal{F}}(g_1,\dots,g_m)(\alpha) = f_{\alpha^1}(g_1(\alpha(1)),\dots,g_m(\alpha(m))),$$

for every $g_1,\dots,g_m \in \mathcal{C}$ and every $\alpha = \alpha^1\alpha^2\dots\alpha^k\dots \in \Omega$, where the metric space $(\mathcal{C},d_u)$ is described by $\mathcal{C} = \{f : \Omega \to X \mid f \text{ is continuous and bounded}\}$ and

$$d_u(f,g) = \sup_{\alpha\in\Omega} d(f(\alpha),g(\alpha))$$

for every $f,g \in \mathcal{C}$.

Remark 2.7. i) $H_{\mathcal{F}}(g_1,\dots,g_m)$ is continuous for all $g_1,\dots,g_m \in \mathcal{C}$. This results from the following facts: the maps $\alpha \to \alpha(i)$ are continuous, $\Omega = \bigcup_{i\in I}\Omega^i$, where

$$\Omega^i = \{\alpha = \alpha^1\alpha^2\dots\alpha^i\alpha^{i+1}\dots \in \Omega \mid \alpha^1 = i\},$$

and the restriction of $H_{\mathcal{F}}(g_1,\dots,g_m)$ to the open set Ω^i is continuous for every $i \in I$.

ii) $H_{\mathcal{F}}(g_1,\dots,g_m)$ is bounded for all $g_1,\dots,g_m \in \mathcal{C}$. This results from the boundedness of the family of functions $(f_i)_{i\in I}$, the boundedness of the functions $g_1,\dots,g_m$ and from the fact that

$$H_{\mathcal{F}}(g_1,\dots,g_m)(\Omega) = H_{\mathcal{F}}(g_1,\dots,g_m)(\bigcup_{i\in I}\Omega^i)$$

$$= \bigcup_{i\in I} H_{\mathcal{F}}(g_1,\dots,g_m)(\Omega^i) = \bigcup_{i\in I} f_i(g_1(\Omega)\times\dots\times g_m(\Omega)).$$

iii) $H_{\mathcal{F}}$ *is well defined*. This results from i) and ii).

Remark 2.8. $(\mathcal{C},d_u)$ is complete provided that (X,d) is complete.

Finally we introduce the canonical projection associated to a possibly infinite generalized iterated function system $\mathcal{F}$.

Definition 2.9. A possibly infinite generalized iterated function system

$$\mathcal{F} = ((X,d),(f_i)_{i\in I})$$

of order $m \in \mathbb{N}^*$ admits canonical projection if has attractor (denoted by $A_{\mathcal{F}}$) and for every $\alpha = \alpha^1\dots\alpha^n\dots \in \Omega$ the set $\bigcap_{n\in\mathbb{N}} \overline{f_{\alpha^1\dots\alpha^n}(A_{\mathcal{F}},\dots,A_{\mathcal{F}})}$ consists of a single element denoted by $\pi(\alpha)$. In this case the function $\pi : \Omega \to X$ is called the canonical projection associated to $\mathcal{F}$.

3. Main results

Theorem 3.1. *Let (X,d) be a complete metric space, $\varphi : [0,\infty) \to [0,\infty)$ a comparison function, $m,p \in \mathbb{N}^*$ and a continuous function $f : X^m \to X$ such that*

$$d(f^{[p]}(x),f^{[p]}(y)) \le \varphi(\max\{\max_{\sigma\in\mathcal{F}_i^p} d(f^{[i]}(x_\sigma),f^{[i]}(y_\sigma)) \mid i \in \{0,1,2,\dots,p-1\}\}),$$

for all $x,y \in X^{m^p}$.

Then:

a) There exists a unique $\alpha \in X$ such that $f(\alpha, \ldots, \alpha) = \alpha$.

b) If f is bounded on bounded subsets of X^m, then, for every $B \in P_{b,cl}(X)$ and every $x_k \in B^{m^k}$, $\lim_{k\to\infty} f^{[k]}(x_k) = \alpha$, the convergence being uniform with respect to x_k.

Proof. a) Note that the continuous function $g : X \to X$ given by $g(x) = f(x, \ldots, x)$ satisfies the inequality

$$d(g^{[p]}(x), g^{[p]}(y)) \leq \varphi(\max\{d(g^{[i]}(x), g^{[i]}(y)) \mid i \in \{0, 1, \ldots, p-1\}\}), \tag{3.1}$$

for all $x, y \in X$. Then, based on (3.1), using Theorem 3.1 from [5], we infer that there exists a unique $\alpha \in X$ such that $g(\alpha) = \alpha$ and $\lim_{n\to\infty} g^{[n]}(x) = \alpha$ for every $x \in X$. Hence there exists a unique $\alpha \in X$ such that $f(\alpha, \ldots, \alpha) = \alpha$.

b) In the sequel, for $B \in P_{b,cl}(X)$ and $k \in \mathbb{N}$, we shall use the following notations:

$$M_k(B) \overset{not}{=} \sup_{x\in B^{m^k}} d(\alpha, f^{[k]}(x))$$

and

$$N_k(B) \overset{not}{=} \max\{M_{k+i}(B) \mid i \in \{0, 1, \ldots, p-1\}\}.$$

As

$$M_n(f(B)) = \sup_{y\in(f(B))^{m^n}} d(\alpha, f^{[k]}(y)) = \sup_{x\in B^{m^{n+1}}} d(\alpha, f^{[n+1]}(x)) = M_{n+1}(B)$$

for all $B \in P_{b,cl}(X)$ and all $n \in \mathbb{N}$, the mathematical induction method leads us to the following conclusion:

$$M_m(f^{[n]}(B)) = M_{m+n}(B), \tag{3.2}$$

for every $B \in P_{b,cl}(X)$, $m, n \in \mathbb{N}$.
Moreover, we have

$$M_{n+p}(B) \leq \varphi(\max\{M_{n+i}(B) \mid i \in \{0, 1, \ldots, p-1\}\}), \tag{3.3}$$

for every $B \in P_{b,cl}(X)$ and $n \in \mathbb{N}$.
Indeed,

$$\begin{aligned} M_{n+p}(B) &\overset{(3.2)}{=} M_n(f^{[p]}(B)) = \sup_{x\in B^{m^{n+p}}} d(\alpha, f^{[m+p]}(x)) \\ &\leq \sup_{x\in B^{m^{n+p}}} \varphi(\max\{d(\alpha, f^{[n+i]}(x)) \mid i \in \{0, 1, \ldots, p-1\}\}) \\ &\leq \varphi(\max\{\sup_{x\in B^{m^{n+i}}} d(\alpha, f^{[n+i]}(x)) \mid i \in \{0, 1, \ldots, p-1\}\}) \\ &= \varphi(\max\{M_{n+i}(B) \mid i \in \{0, 1, \ldots, p-1\}\}). \end{aligned}$$

In addition, from (3.3), we have $N_{n+1}(B) \leq N_n(B) \leq \ldots \leq N_0(B) < \infty$ and $N_{n+p}(B) \leq \varphi(N_n(B))$ for every $n \in \mathbb{N}$.
Hence $N_n(B) \leq \varphi^{[\frac{n}{p}]}(\max\{M_i(B) \mid i \in \{0, 1, \ldots, p-1\}\})$ and consequently $\lim_{n\to\infty} N_n(B) = \lim_{n\to\infty} M_n(B) = 0$ for every $B \in P_{b,cl}(X)$. □

Theorem 3.2. *Let $\mathcal{F} = ((X, d), (f_i)_{i\in I})$ be a possibly infinite generalized iterated function system of order $m \in \mathbb{N}^*$ and $p \in \mathbb{N}$ such that*

$$d(f_\alpha(x), f_\alpha(y)) \leq \varphi(\max\{\max_{\sigma\in\mathcal{F}_q^p} d(f_\beta(x_\sigma), f_\beta(y_\sigma)) \mid \beta \in_q \Omega, q \in \{0, 1, \ldots, p-1\}),$$

for all $x, y \in X^{m^p}$, *where* φ *is a comparison function. Then:*

a) There exists a unique $A_{\mathcal{F}} \in P_{b,cl}(X)$ *such that* $F_{\mathcal{F}}(A_{\mathcal{F}}, \ldots, A_{\mathcal{F}}) = A_{\mathcal{F}}$, *i.e.* $\mathcal{F}$ *has attractor.*

b) $\lim\limits_{n\to\infty} F_{\mathcal{F}}^{[n]}(B_n) = A_{\mathcal{F}}$ *for all* $B \in P_{b,cl}(X)$ *and* $B_n = (B_1^n, \ldots, B_{m^n}^n) \subseteq B^{m^n}$ *with* $B_i^n \in P_{b,cl}(X)$ *for all* $i \in \{1, 2, \ldots, m^n\}$.

c) For all $\alpha = \alpha^1 \ldots \alpha^n \ldots \in \Omega$, *the set* $\bigcap\limits_{n\in\mathbb{N}} \overline{f_{\alpha^1\alpha^2\ldots\alpha^n}(A_{\mathcal{F}}, \ldots, A_{\mathcal{F}})}$ *has only one element denoted by* a_α, *so* $\mathcal{F}$ *admits canonical projection.*

d) For all $\alpha = \alpha^1 \ldots \alpha^n \ldots \in \Omega$, $B \in P_{b,cl}(X)$ *and* $B_n = (B_1^n, \ldots, B_{m^n}^n) \subseteq B^{m^n}$ *with* $B_i^n \in P_{b,cl}(X)$ *for each* $i \in \{1, 2, \ldots, m^n\}$, *we have* $\lim\limits_{n\to\infty} f_{\alpha^1\alpha^2\ldots\alpha^n}(B_n) = \{a_\alpha\}$ *and the convergence is uniform with respect to* α *and the sets* B.

Proof. a) The function $F : P_{b,cl}(X) \to P_{b,cl}(X)$ given by $F(B) = F_{\mathcal{F}}(B, \ldots, B)$ for every $B \in P_{b,cl}(X)$ has the property that

$$h(F^{[p]}(B_1), F^{[p]}(B_2)) \le \varphi(\max\{h(F^{[i]}(B_1), F^{[i]}(B_2)) \mid i \in \{0, 1, \ldots, p-1\}\}),$$

for all $B_1, B_2 \in P_{b,cl}(X)$. Theorem 3.1 assures the existence and the uniqueness of a set $A_{\mathcal{F}} \in P_{b,cl}(X)$ such that $F(A_{\mathcal{F}}) = A_{\mathcal{F}}$ (i.e. $F_{\mathcal{F}}(A_{\mathcal{F}}, \ldots, A_{\mathcal{F}}) = A_{\mathcal{F}}$) and $\lim\limits_{n\to\infty} F^{[n]}(B) = A_{\mathcal{F}}$ for every $B \in P_{b,cl}(X)$.

b) For $B_1, B_2 \in P_{b,cl}(X)$, $p, n \in \mathbb{N}$ and $\alpha \in {}_p\Omega$, in the sequel, we shall use the following notations:

$$M_\alpha(B_1, B_2) = \sup_{x\in B_1^{m^p}, y\in B_2^{m^p}} d(f_\alpha(x), f_\alpha(y)),$$

$$M_p(B_1, B_2) = \sup_{\alpha\in {}_p\Omega} M_\alpha(B_1, B_2)$$

and

$$N_n(B_1, B_2) = \max\{M_n(B_1, B_2), \ldots, M_{n+p-1}(B_1, B_2)\}.$$

Then, we have

$$\begin{aligned} d(f_\alpha(x), f_\alpha(y)) &\le \varphi(\max\{\max_{\sigma\in\mathcal{F}_q^p} d(f_\beta(x_\sigma), f_\beta(y_\sigma)) \mid \beta \in {}_q\Omega,\ q \in \{0, 1, \ldots, p-1\}) \\ &\le \varphi(\max\{\max_{\omega\in {}_i\Omega} M_\omega(B_1, B_2)) \mid i \in \{0, 1, \ldots, p-1\}) \\ &\le \varphi(\max\{M_i(B_1, B_2)) \mid i \in \{0, 1, \ldots, p-1\}), \end{aligned}$$

for all $B_1, B_2 \in P_{b,cl}(X)$ and $x \in B_1^{m^p}, y \in B_2^{m^p}$, so

$$M_\alpha(B_1, B_2) \le \varphi(\max\{M_0(B_1, B_2), \ldots, M_{p-1}(B_1, B_2)\})$$

and

$$M_p(B_1, B_2) \le \varphi(\max\{M_0(B_1, B_2), \ldots, M_{p-1}(B_1, B_2)\}), \tag{3.4}$$

for all $B_1, B_2 \in P_{b,cl}(X)$ and $\alpha \in {}_p\Omega$. Moreover

$$M_{i+j}(B_1, B_2) = M_j(F_{\mathcal{F}}^{[i]}(B_1, \ldots, B_1), F_{\mathcal{F}}^{[i]}(B_2, \ldots, B_2)), \tag{3.5}$$

for all $B_1, B_2 \in P_{b,cl}(X)$, $i, j \in \mathbb{N}$. By replacing, in (1), the set B_1 by $F_{\mathcal{F}}^{[n]}(B_1, \ldots, B_1)$ and the set B_2 by $F_{\mathcal{F}}^{[n]}(B_2, \ldots, B_2)$, we get

$$M_{n+p}(B_1, B_2) \le \varphi(\max\{M_n(B_1, B_2), \ldots, M_{n+p-1}(B_1, B_2)), \tag{3.6}$$

for all $B_1, B_2 \in P_{b,cl}(X)$, $n \in \mathbb{N}$. From (3.6) we infer that

$$N_{n+1}(B_1, B_2) \leq N_n(B_1, B_2) \text{ and } N_{n+p}(B_1, B_2) \leq \varphi(N_n(B_1, B_2)),$$

for all $B_1, B_2 \in P_{b,cl}(X)$ and $n \in \mathbb{N}$. Therefore

$$N_n(B_1, B_2) \leq \varphi^{[\frac{n}{p}]}(\max\{M_0(B_1, B_2), \ldots, M_{p-1}(B_1, B_2)\}),$$

for all $B_1, B_2 \in P_{b,cl}(X)$ and $n \in \mathbb{N}$, so

$$\lim_{n\to\infty} N_n(B_1, B_2) = \lim_{n\to\infty} M_n(B_1, B_2) = \lim_{n\to\infty} h(F_{\mathcal{F}}^{[n]}(B_1, \ldots, B_1), \quad (3.7)$$

$$F_{\mathcal{F}}^{[n]}(B_2, \ldots, B_2)) = 0,$$

for all $B_1, B_2 \in P_{b,cl}(X)$. In particular, for $B_2 = A_{\mathcal{F}}$, we obtain that

$$\lim_{n\to\infty} h(F_{\mathcal{F}}^{[n]}(B_1, \ldots, B_1), A_{\mathcal{F}}) = 0, \text{ i.e. } \lim_{n\to\infty} F_{\mathcal{F}}^{[n]}(B_1, \ldots, B_1) = A_{\mathcal{F}},$$

for each $B_1 \in P_{b,cl}(X)$. Moreover, we have

$$M_\alpha(B_1, B_2) \leq M_\alpha(C_1, C_2) \text{ and } M_n(B_1, B_2) \leq M_n(C_1, C_2),$$

for all $B_1, B_2, C_1, C_2 \in P_{b,cl}(X)$, $B_1 \subseteq C_1$, $B_2 \subseteq C_2$, $n \in \mathbb{N}$ and $\alpha \in {}_n\Omega$.
If for $B, C \in P_{b,cl}(X)$ and $n \in \mathbb{N}$, $B_n = (B_1^n, \ldots, B_{m^n}^n)$, $C_n = (C_1^n, \ldots, C_{m^n}^n) \subseteq B^{m^n}$, with $B_i^n, C_i^n \in P_{b,cl}(X)$ and $B_i^n \subseteq B, C_i^n \subseteq C$ for all $i \in \{1, \ldots, m^n\}$, then

$$\lim_{n\to\infty} F_{\mathcal{F}}^{[n]}(B_n) = A_{\mathcal{F}}.$$

Indeed, we have only to take into account (3.7) and the inequality

$$h(F_{\mathcal{F}}^{[n]}(B_n), F_{\mathcal{F}}^{[n]}(C_n)) \leq M_n(B, C),$$

which is valid for all $n \in \mathbb{N}$, for $C = A_{\mathcal{F}}$.

c) Let us note that, as $h(f_\alpha(B_n), f_\alpha(C_n)) \leq M_n(B, C)$ for all $\alpha \in {}_n\Omega$, taking into account (3.7), we infer that $\lim_{n\to\infty} h(f_\alpha(B_n), f_\alpha(C_n)) = 0$ for all $B, C \in P_{b,cl}(X)$ and $\alpha \in {}_n\Omega$.

In the sequel, for $\alpha = \alpha^1 \ldots \alpha^n \ldots \in \Omega$, we shall use the following notation:

$$f_{\alpha^1 \ldots \alpha^n}(A_{\mathcal{F}}, \ldots, A_{\mathcal{F}}) \overset{not}{=} A_{\alpha^1 \ldots \alpha^n}.$$

Note that $\operatorname{diam}(A_{\alpha^1 \ldots \alpha^n}) = M_{\alpha^1 \ldots \alpha^n}(A_{\mathcal{F}}, A_{\mathcal{F}})$ for all $\alpha = \alpha^1 \ldots \alpha^n \ldots \in \Omega$ and $n \in \mathbb{N}$. As $A_{\alpha^1 \ldots \alpha^n \alpha^{n+1}} \subseteq A_{\alpha^1 \ldots \alpha^n}$, we obtain that

$$\operatorname{diam}(A_{\alpha^1 \ldots \alpha^n \alpha^{n+1}}) \leq \operatorname{diam}(A_{\alpha^1 \ldots \alpha^n}) \leq M_n(A_{\mathcal{F}}, A_{\mathcal{F}})$$

for all $\alpha = \alpha^1 \ldots \alpha^n \ldots \in \Omega$ and $n \in \mathbb{N}$ and, based on (3.7), we conclude that the set $\bigcap_{n\in\mathbb{N}} \overline{f_{\alpha^1\alpha^2 \ldots \alpha^n}(A_{\mathcal{F}}, \ldots, A_{\mathcal{F}})}$ has only one element denoted by a_α.

Let us note that

$$h(f_{\alpha^1 \ldots \alpha^n}(A_{\mathcal{F}}, \ldots, A_{\mathcal{F}}), \{a_\alpha\}) \leq \operatorname{diam}(f_{\alpha^1 \ldots \alpha^n}(A_{\mathcal{F}}, \ldots, A_{\mathcal{F}})) \leq M_n(A_{\mathcal{F}}, A_{\mathcal{F}})$$

for all $\alpha = \alpha^1 \ldots \alpha^n \ldots \in \Omega$ and $n \in \mathbb{N}$. Therefore, using (3.7), we get

$$\lim_{n\to\infty} f_{\alpha^1\alpha^2 \ldots \alpha^n}(A_{\mathcal{F}}, \ldots, A_{\mathcal{F}}) = \{a_\alpha\}.$$

d) Because $\lim_{n\to\infty} f_{\alpha^1\ldots\alpha^n}(A_{\mathcal{F}},\ldots,A_{\mathcal{F}}) = \{a_\alpha\}$ and

$$\lim_{n\to\infty} f_{\alpha^1\ldots\alpha^n}(A_{\mathcal{F}},\ldots,A_{\mathcal{F}}) = \lim_{n\to\infty} f_{\alpha^1\ldots\alpha^n}(B_1^n,\ldots,B_{m^n}^n),$$

we conclude that

$$\lim_{n\to\infty} f_{\alpha^1\alpha^2\ldots\alpha^n}(B_1^n,\ldots,B_{m^n}^n) = 0$$

for all $\alpha = \alpha^1\ldots\alpha^n\ldots \in \Omega$, $B \in P_{b,cl}(X)$ and $B_n = (B_1^n,\ldots,B_{m^n}^n) \subseteq B^{m^n}$ with $B_i^n \in P_{b,cl}(X)$ for each $i \in \{1,2,\ldots,m^n\}$.

Concerning the rate of the convergence we have the following estimation:

$$h(f_{\alpha^1\alpha^2\ldots\alpha^n}(B_n),\{a_\alpha\}) \leq h(f_{\alpha^1\alpha^2\ldots\alpha^n}(B_n), A_{\alpha^1\ldots\alpha^n}) + h(A_{\alpha^1\ldots\alpha^n},\{a_\alpha\})$$

$$\leq M_n(A_{\mathcal{F}},B) + M_n(A_{\mathcal{F}},A_{\mathcal{F}}) \leq 2\varphi^{[\frac{n}{p}]}(\max\{M_i(A_{\mathcal{F}},B) \mid i \in \{0,1,\ldots,p-1\}\}),$$

for all $\alpha = \alpha^1\ldots\alpha^n\ldots \in \Omega$ and $n \in \mathbb{N}$. □

Theorem 3.3. *Let $\mathcal{F} = ((X,d),(f_i)_{i\in I})$ be a possibly infinite generalized iterated function system of order $m \in \mathbb{N}^*$ and $p \in \mathbb{N}$ such that*

$$d(f_\alpha(x),f_\alpha(y)) \leq \varphi(\max\{\max_{\sigma\in\mathcal{F}_q^p} d(f_\beta(x_\sigma),f_\beta(y_\sigma)) \mid \beta \in_q \Omega, q \in \{0,1,\ldots,p-1\}),$$

for all $x,y \in X^{m^p}$, where φ is a comparison function. Then there exists a unique $\pi \in \mathcal{C}$ such that:

a) $H_{\mathcal{F}}(\pi,\ldots,\pi) = \pi$ and $\overline{\pi(\Omega)} = A_{\mathcal{F}}$.

b) $\lim_{n\to\infty} H_{\mathcal{F}}^{[n]}(f_n) = \pi$ for all $B \in P_{b,cl}(X)$ and $f_n = (f_1^n,\ldots,f_{m^n}^n) \in \mathcal{C}_B^{m^n}$, where $\mathcal{C}_B = \{f : \Omega \to B \mid f$ is continuous$\}$ is endowed with the uniform metric, the convergence being uniform with respect to B.

c) π is the canonical projection associated to $\mathcal{F}$.

Proof. a) Using the mathematical induction method, one can easily prove that

$$H_{\mathcal{F}}^{[n]}(g_1,\ldots,g_{m^n})(\alpha) =$$

$$= f_{\alpha^1\alpha^2\ldots\alpha^n}(g_1(\alpha(11\ldots1)),\ldots,g_m(\alpha(11\ldots m)),\ldots,g_{m^n}(\alpha(mm\ldots m))), \quad (3.8)$$

for all $\alpha = \alpha^1\ldots\alpha^n\ldots \in \Omega$ and all $n \in \mathbb{N}$, where we adopted the following notation:

$$\alpha(i_1)(i_2)\ldots(i_k) \overset{not}{=} \alpha(i_1i_2\ldots i_k).$$

For a fixed $n \in \mathbb{N}$, for each $l \in \{1,\ldots,m^n\}$ there exists a unique ordered subset $\{l_1,\ldots,l_n\}$ of $\{1,2\ldots,m\}$ such that $l-1 = l_1m^{n-1} + l_2m^{n-2} + \ldots + l_n$, so we can consider the function $u : \{1,2,\ldots,m^n\} \to \{1,2,\ldots,m\}^n$ given by

$$u(l) = (l_1+1,l_2+1,\ldots,l_n+1)$$

for all $l \in \{1,2,\ldots,m^n\}$ and rewrite (3.8) in the following form:

$$H_{\mathcal{F}}^{[n]}(g_1,\ldots,g_{m^n})(\alpha) = f_{\alpha^1\alpha^2\ldots\alpha^n}(g_1(\alpha(u(1))),\ldots,g_{m^n}(\alpha(u(m^n)))),$$

for all $\alpha = \alpha^1\ldots\alpha^n\ldots \in \Omega$ and all $n \in \mathbb{N}$.

Claim. $H_{\mathcal{F}}$ is a φ-max generalized contraction.

Justification of the claim. Indeed, we have

$$d_u(H_{\mathcal{F}}^{[p]}(g_1,\ldots,g_{m^p}),H_{\mathcal{F}}^{[p]}(h_1,\ldots,h_{m^p}))$$

$$= \sup_{\alpha\in\Omega} d(H^{[p]}_{\mathcal{F}}(g_1,\ldots,g_{m^p})(\alpha), H^{[p]}_{\mathcal{F}}(h_1,\ldots,h_{m^p})(\alpha))$$
$$\leq \sup_{\alpha\in\ _m\Omega}\ \sup_{\alpha(1),\ldots,\alpha(m^p)\in\Omega}$$
$$\varphi(\max_{i\in\{0,1,\ldots,p-1\}}\ \max_{\beta\in\ _i\Omega}\ \max_{\sigma\in\mathcal{F}^p_i} d(f_\beta(g_{\sigma(i)}(\alpha(\sigma(u(i))))), f_\beta(h_{\sigma(i)}(\alpha(\sigma(u(i)))))))$$
$$\leq \varphi(\sup_{\alpha\in\ _m\Omega}\ \max_{i\in\{0,1,\ldots,p-1\}}\ \max_{\sigma\in\mathcal{F}^p_i}\ \max_{\beta\in\ _i\Omega}$$
$$\sup_{\alpha(1),\ldots,\alpha(m^p)\in\Omega} d(f_\beta(g_{\sigma(i)}(\alpha(\sigma(u(i))))), f_\beta(h_{\sigma(i)}(\alpha(\sigma(u(i)))))))$$
$$\leq \varphi(\max\{\max_{\sigma\in\mathcal{F}^p_i} d_u(H^{[i]}_{\mathcal{F}}(g_\sigma), H^{[i]}_{\mathcal{F}}(h_\sigma)) \mid i\in\{0,1,\ldots,p-1\}\}),$$

for all $g_1,\ldots,g_{m^p},h_1,\ldots,h_{m^p}\in\mathcal{C}$.
The Claim and Theorem 3.1 assure us that there exists a unique $\pi\in\mathcal{C}$ such that

$$H_{\mathcal{F}}(\pi,\ldots,\pi)=\pi.$$

Moreover, we have $\overline{\pi(\Omega)}=A_{\mathcal{F}}$. Indeed,

$$\overline{\pi(\Omega)}=\overline{H_{\mathcal{F}}(\pi,\ldots,\pi)(\Omega)}$$
$$=\overline{\bigcup_{i\in I}\ \bigcup_{\alpha_1,\ldots,\alpha_m\in\Omega} f_i(\pi(\alpha_1),\ldots,\pi(\alpha_m))}=\overline{\bigcup_{i\in I} f_i(\pi(\Omega)\times\ldots\times\pi(\Omega))}$$
$$\overset{f_i\text{ continuous}}{=}\overline{\bigcup_{i\in I} f_i(\overline{\pi(\Omega)}\times\ldots\times\overline{\pi(\Omega)})}=F_{\mathcal{F}}(\overline{\pi(\Omega)}\times\ldots\times\overline{\pi(\Omega)})$$

and $\overline{\pi(\Omega)}\in P_{b,cl}(X)$ (since $\pi\in\mathcal{C}$). In view of Theorem 3.2, a), we conclude that $\overline{\pi(\Omega)}=A_{\mathcal{F}}$.

b) Let us consider $B\in P_{b,cl}(X)$ and $f_n=(f^n_1,\ldots,f^n_{m^n})\in\mathcal{C}^{m^n}_B$, $n\in\mathbb{N}$. Note that the family of function $(f^n_i)_{i\in\{1,2,\ldots,m^n\}}$ is bounded (as $\bigcup_{i\in\{1,2,\ldots,m^n\}} f^n_i(\Omega)\subseteq B$) for all $n\in\mathbb{N}$.

Claim 1. $H_{\mathcal{F}}(\mathcal{C}_1\times\ldots\times\mathcal{C}_1)$ is bounded for every bounded subset $\mathcal{C}_1$ of $\mathcal{C}$.
Justification of Claim 1. Let us consider $\mathcal{C}_1$ a bounded (with respect to d_u) subset of $\mathcal{C}$. Then there exists $g\in\mathcal{C}$ and $r>0$ such that $\mathcal{C}_1\subseteq B(g,r)$. It follows that

$$\overline{\bigcup_{f\in\mathcal{C}_1} f(\Omega)}\subseteq\overline{B(g(\Omega),r)}$$

and we shall use the following notation: $B\overset{not}{=}\overline{\bigcup_{f\in\mathcal{C}_1} f(\Omega)}\in P_{b,cl}(X)$. The inclusion

$$H_{\mathcal{F}}(\mathcal{C}_1,\ldots,\mathcal{C}_1)\subseteq C(\Omega,F_{\mathcal{F}}(B,\ldots,B))=\{f:\Omega\to F_{\mathcal{F}}(B,\ldots,B)\mid f\text{ is continuous }\}$$

is valid as

$$H_{\mathcal{F}}(f_1,\ldots,f_m)(\Omega)=\bigcup_{i\in I}\ \bigcup_{\alpha(1),\ldots,\alpha(m)\in\Omega} f_i(f_1(\alpha(1)),\ldots,f_m(\alpha(m)))$$
$$\subseteq\bigcup_{i\in I} f_i(f_1(\Omega),\ldots,f_m(\Omega))\subseteq\bigcup_{i\in I} f_i(B,\ldots,B)\subseteq F_{\mathcal{F}}(B,\ldots,B),$$

for all $f_1,\ldots,f_m\in\mathcal{C}_1$. Hence

$$d_u(H_{\mathcal{F}}(f_1,\ldots,f_m),H_{\mathcal{F}}(g_1,\ldots,g_m))\leq\operatorname{diam}(F_{\mathcal{F}}(B,\ldots,B))$$

for all $f_1,\ldots,f_m,g_1,\ldots,g_m\in\mathcal{C}_1$, so $H_{\mathcal{F}}(\mathcal{C}_1\times\ldots\times\mathcal{C}_1)$ is bounded for every bounded subset $\mathcal{C}_1$ of $\mathcal{C}$. The justification of the claim is done.

Let $\mathcal{C}_1$ be a bounded subset of $\mathcal{C}$. Since

$$d_u(H_{\mathcal{F}}^{[n]}(g_1^n,\dots,g_{m^n}^n),H_{\mathcal{F}}^{[n]}(h_1^n,\dots,h_{m^n}^n))\leq \operatorname{diam}(F_{\mathcal{F}}^{[n]}(B,\dots B))$$

for all $n\in\mathbb{N}$ and $g_1^n,\dots,g_{m^n}^n,h_1^n,\dots,h_{m^n}^n\in\mathcal{C}_1\cup\{\pi\}$, using Theorem 3.1, b), we conclude that $\lim\limits_{n\to\infty}H_{\mathcal{F}}^{[n]}(f_n)=\pi$.

c) Note that

$$\pi(\alpha)=H_{\mathcal{F}}(\pi,\dots,\pi)(\alpha)=f_{\alpha^1}(\pi(\alpha(1)),\dots,\pi(\alpha(m))), \tag{3.9}$$

for all $\alpha\in\Omega$.

Claim 2.

$$\pi(F_{\alpha^1\alpha^2\dots\alpha^n}(\Lambda_1,\dots,\Lambda_{m^n}))=f_{\alpha^1\alpha^2\dots\alpha^n}(\pi(\Lambda_1)\times\dots\times\pi(\Lambda_{m^n})), \tag{3.10}$$

for all $n\in\mathbb{N}^*$, $\alpha^1\in I,\alpha^2\in\Omega_2,\dots,\alpha^n\in\Omega_n$ and $\Lambda_1,\dots,\Lambda_{m^n}\subseteq\Omega$.

Justification of Claim 2. We are going to use the mathematical induction method.

Using (3.9), we get Claim 2 for $n=1$.

Let us suppose that (3.10) is valid for n. We shall prove that it is also true for $n+1$. We have

$$\begin{aligned}
&\pi(F_{\alpha^1\alpha^2\dots\alpha^n\alpha^{n+1}}(\Lambda_1,\dots,\Lambda_{m^{n+1}}))\\
&=\pi((F_{\alpha(1)}(\Lambda_1,\dots,\Lambda_{m^n}),\dots,F_{\alpha(m)}(\Lambda_{m^{n+1}-m^n+1},\dots,\Lambda_{m^{n+1}})))\\
&\overset{(3.9)}{=} f_{\alpha^1}(\pi(F_{\alpha(1)}(\Lambda_1,\dots,\Lambda_{m^n})),\dots,\pi(F_{\alpha(m)}(\Lambda_{m^{n+1}-m^n+1},\dots,\Lambda_{m^{n+1}}))))\\
&= f_{\alpha^1}(f_{\alpha(1)}(\pi(\Lambda_1),\dots,\pi(\Lambda_{m^n})),\dots,f_{\alpha(m)}(\pi(\Lambda_{m^{n+1}-m^n+1}),\dots,\pi(\Lambda_{m^{n+1}})))\\
&\overset{\text{Claim 2 for } n}{=} f_{\alpha^1\alpha^2\dots\alpha^n\alpha^{n+1}}(\pi(\Lambda_1)\times\dots\times\pi(\Lambda_{m^{n+1}})),
\end{aligned}$$

for all $\Lambda_1,\dots,\Lambda_{m^{n+1}}\subseteq\Omega$, where $\alpha=\alpha^1\alpha^2\dots\alpha^n\dots$.

Finally, we have

$$\begin{aligned}
\pi(\alpha)&\in\pi(\bigcap_{n\in\mathbb{N}^*}F_{\alpha^1\alpha^2\dots\alpha^n}(\Omega,\dots,\Omega))\subseteq\bigcap_{n\in\mathbb{N}^*}\pi(F_{\alpha^1\alpha^2\dots\alpha^n}(\Omega,\dots,\Omega))\\
&\overset{\text{Claim 2}}{=}\bigcap_{n\in\mathbb{N}^*}f_{\alpha^1\alpha^2\dots\alpha^n}(\pi(\Omega),\dots,\pi(\Omega))\subseteq\bigcap_{n\in\mathbb{N}^*}\overline{f_{\alpha^1\alpha^2\dots\alpha^n}(A_{\mathcal{F}},\dots,A_{\mathcal{F}})},
\end{aligned}$$

for all $\alpha=\alpha^1\alpha^2\dots\alpha^n\dots\in\Omega$, so, based on Theorem 3.2, b), π is the canonical projection associated to $\mathcal{F}$. □

References

[1] Dumitru, D., *Generalized iterated function systems containing Meir-Keeler functions*, An. Univ. Bucur., Mat., **58**(2009), 109-121.

[2] Dumitru, D., *Contraction-type functions and some applications to GIIFS*, An. Univ. Spiru Haret, Ser. Mat.-Inform., **12**(2016), 31-44.

[3] Georgescu, F., Miculescu, R., Mihail, A., *A study of the attractor of a φ-max-IFS via a relatively new method*, J. Fixed Point Theory Appl., (2018) 20-24.

[4] Miculescu, R., *Generalized iterated function systems with place dependent probabilities*, Acta Appl. Math., **130**(2014), 135-150.

[5] Miculescu, R., Mihail, A., *A generalization of Matkowski's fixed point theorem and Istrăţescu's fixed point theorem concerning convex contractions*, J. Fixed Point Theory Appl., **19**(2017), 1525-1533.

[6] Mihail, A., Miculescu, R., *Applications of Fixed Point Theorems in the Theory of Generalized IFS*, Fixed Point Theory, Appl. Volume 2008. Article ID 312876, 11 pages.

[7] Mihail, A., Miculescu, R., *A generalization of the Hutchinson measure*, Mediterr. J. Math., **6**(2009), 203–213.

[8] Mihail, A., Miculescu, R., *Generalized IFSs on Noncompact Spaces*, Fixed Point Theory Appl., Volume 2010, Article ID 584215, 11 pages.

[9] Mihail, A., *The canonical projection between the shift space of an IIFS and its attractor as a fixed point*, Fixed Point Theory Appl., 2015, Paper No. 75, 15 p.

[10] Oliveira, E., Strobin, F., *Fuzzy attractors appearing from GIFZS*, Fuzzy Set Syst., **331**(2018), 131-156.

[11] Secelean, N.A., *Invariant measure associated with a generalized countable iterated function system*, Mediterr. J. Math., **11**(2014), 361-372.

[12] Secelean, N.A., *Generalized iterated function systems on the space* $l^\infty(X)$, J. Math. Anal. Appl., **410**(2014), 847-858.

[13] Strobin, F., *Attractors of generalized IFSs that are not attractors of IFSs*, J. Math. Anal. Appl., **422**(2015), 99-108.

[14] Strobin, F., Swaczyna, J., *On a certain generalisation of the iterated function system*, Bull. Aust. Math. Soc., **87**(2013), 37-54.

[15] Strobin, F., Swaczyna, J., *A code space for a generalized IFS*, Fixed Point Theory, **17**(2016), 477-493.

Silviu-Aurelian Urziceanu
Department of Mathematics and Computer Science
University of Piteşti
Târgul din Vale 1, 110040, Piteşti, Argeş, Romania
e-mail: `fmi_silviu@yahoo.com`

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 151–160
DOI: 10.24193/subbmath.2019.2.02

On a stochastic arc furnace model

Hans-Jörg Starkloff, Markus Dietz and Ganna Chekhanova

Abstract. One of the approaches in modeling of electric arc furnace is based on the power balance equation and results in a nonlinear ordinary differential equation. In reality it can be observed that the graph of the arc voltage varies randomly in time, in fact it oscillate with a random time-varying amplitude and a slight shiver. To get a more realistic model, at least one of the model parameters should be modeled as a stochastic process, which leads to a random differential equation.

We propose a stochastic model using the stationary Ornstein-Uhlenbeck process for modeling stochastic influences. Results, gained by applying Monte Carlo method and polynomial chaos expansion, are given here.

Mathematics Subject Classification (2010): 34F05.

Keywords: Electric arc furnace, random differential equation, Ornstein-Uhlenbeck process, Monte Carlo method, polynomial chaos expansion.

1. Introduction

An electric arc furnace (EAF) is used for melting metals in steel industry. One deterministic model of the EAF energy system is based on the instantaneous power balance of the system, which results in the following nonlinear ordinary differential equation

$$k_1 r^n(t) + k_2 r(t) \frac{\mathrm{d}r(t)}{\mathrm{d}t} = \frac{k_3}{r^{m+2}(t)} i^2(t), \quad t \in \mathcal{I}. \tag{1.1}$$

This equation describes how the arc radius r depends on the arc current i, both are functions on a given time interval $\mathcal{I}$ (cf. [1]). The model coefficients k_1, k_2 and k_3 are positive. The parameters m and n belong to the set $\{0, 1, 2\}$ and reflect different working conditions of the arc furnace (cf. [4] or [5]). Equation (1.1) suggests, that the arc radius function r is positive and should be prevented from getting zero. In case

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

the arc radius takes on the value zero one has to deal with a differential equation with singularities which requires additional investigations.

Between the arc voltage u, the arc radius r and the arc current i it holds the relationship

$$u(t) = \frac{k_3}{r^{m+2}(t)} i(t), \quad t \in \mathcal{I}. \tag{1.2}$$

In reality it can be observed that the arc voltage varies randomly in time, it oscillates with a random time-varying amplitude and the graph of the function shows a slight shivering (cf. [4] or [6]). To take this into account, it is better to model the arc voltage as a stochastic process. Here it is proposed to model the coefficient k_3 of equation (1.1) as a stochastic process and then solve the corresponding random differential equation.

Section 3 presents the model in more detail and gives some first results gained by the Monte Carlo method. Section 4 describes how polynomial chaos expansions can be applied.

The stochastic model is based on a deterministic one, for which for certain parameters explicit solutions of (1.1) can be given. This deterministic model was investigated for example in [6] and is recapped here in section 2.

2. A deterministic model

In this section assume the time interval is $\mathcal{I} = \mathbb{R}$. In this paper we want to consider the case $n = 2$, for which (1.1) can be solved explicitly. For this parameter n the equation (1.1) with the substitution $y = r^{m+4}$ results in the following linear ordinary differential equation

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = -\beta y(t) + f(t), \quad t \in \mathcal{I} \tag{2.1}$$

with $f(t) := \frac{(m+4)k_3}{k_2} i^2(t)$ and $\beta := \frac{(m+4)k_1}{k_2} > 0$.

If we assume, that the arc current i is a continuous function and that the initial value condition

$$y(t_0) = y_0 > 0 \tag{2.2}$$

holds ($t_0 \in \mathbb{R}$), then (2.1) has the unique continuously differentiable positive solution

$$y(t) = y(t; t_0, y_0) = y_0 e^{-\beta(t-t_0)} + \int_{t_0}^{t} f(s) e^{-\beta(t-s)}\, \mathrm{d}s, \quad t \in \mathcal{I}. \tag{2.3}$$

If we additionally assume, that i is a p-periodic (with the period p), it is bounded and there exists a unique p-periodic solution of the differential equation (2.1). We get a formula of this p-periodic solution y_{per} also by applying the pullback limit of t_0 on the initial value solution (2.3). It holds

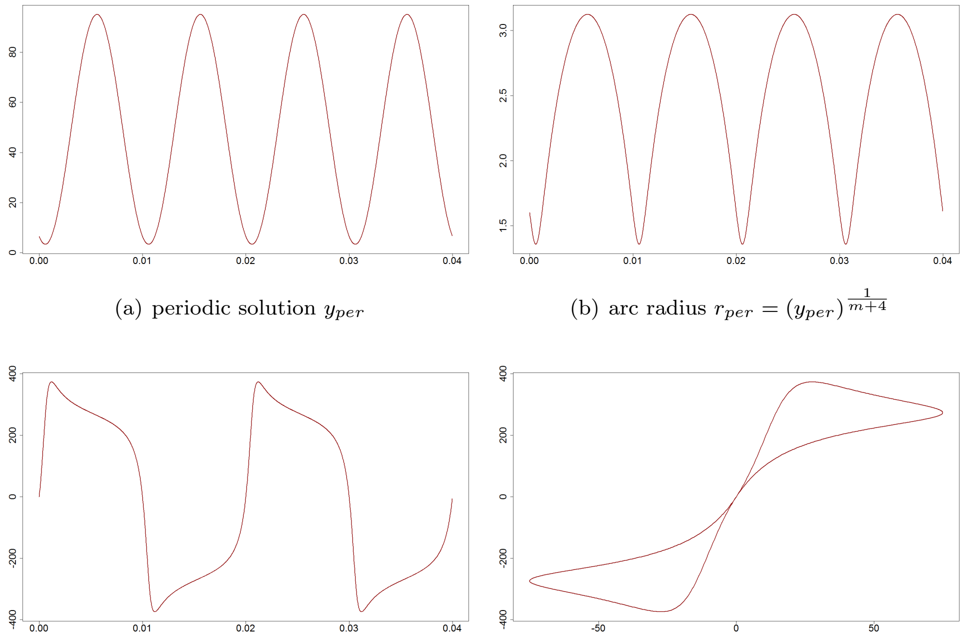
$$y_{per}(t) = \lim_{t_0 \to -\infty} y(t; t_0, y_0) = \int_0^\infty e^{-\beta s} f(t-s)\, \mathrm{d}s, \quad t \in \mathcal{I}. \tag{2.4}$$

In [6] this periodic solution is referred to as a steady state solution of the system. Sometimes it is also called the equilibrium solution.

According to the real world data the graph of the arc current has a sinusoidal shape. To satisfy this behavior the arc current function is chosen as

$$i(t) = a\sin(\omega t), \quad t \in \mathcal{I}. \tag{2.5}$$

Here a (amplitude) and ω (corresponding frequency) are positive parameters. If we apply (2.5), from (2.4) an explicit formula for the periodic solution can be derived.



(a) periodic solution y_{per}

(b) arc radius $r_{per} = (y_{per})^{\frac{1}{m+4}}$

(c) arc voltage u_{per}

(d) voltage-current characteristic

FIGURE 1. Graphs of characteristics of the deterministic model with the following parameters

a	ω	m	k_1	k_2	k_3
75	100π	0	2 000	5	35

Then also an explicit formula for the associated voltage function can be calculated. It holds

$$y_{per}(t) = b\left[1 - c\sin(2\omega t + \psi)\right], \quad t \in \mathcal{I} \tag{2.6}$$

and

$$u_{per}(t) = d\left[1 - c\sin(2\omega t + \psi)\right]^{-\frac{m+2}{m+4}}\sin(\omega t), \quad t \in \mathcal{I} \tag{2.7}$$

with the constants

$$b = \frac{k_3 a^2}{2k_1}, \qquad c = \frac{1}{\sqrt{1 + \left(\frac{2\omega k_2}{(m+4)k_1}\right)^2}}, \tag{2.8}$$

$$d = 2^{\frac{m+2}{m+4}} k_1^{\frac{m+2}{m+4}} k_3^{\frac{2}{m+4}} a^{-\frac{m}{m+4}}, \qquad \psi = \arctan\left(\frac{(m+4)k_1}{2\omega k_2}\right), \tag{2.9}$$

depending on the model parameters m, k_1, k_2, k_3, a and ω.

Figure 1 shows graphs of the periodic solution y_{per}, the arc voltage u_{per}, the arc radius $r_{per} = (y_{per})^{\frac{1}{m+4}}$ and the voltage-current characteristic, i.e., of the curve $((i(t), u_{per}(t)) : t \in \mathcal{I})$.

3. A stochastic model

Here we want to propose a stochastic model, in which the coefficient k_3 from the former deterministic equation (1.1) is now a stochastic process. One of the modeling challenges is to make sure, that the inhomogeneity in equation (2.1) stays positive, which provides that the solution y is always strictly positive and prevents the arc radius r from getting negative. We take this into account by considering a stochastic process $(X_t)_{t\in\mathcal{I}}$ and multiplying k_3 with the non-negative stochastic process $\left((1+X_t)^2\right)_{t\in\mathcal{I}}$. By inserting the stochastic process

$$\left(k_3(1+X_t)^2\right)_{t\in\mathcal{I}} \tag{3.1}$$

into equation (2.1) instead of the deterministic coefficient k_3, the differential equation (2.1) turns into the random ordinary differential equation

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = -\beta Y_t + F_t, \quad t \in \mathcal{I} \tag{3.2}$$

with $F_t = f(t)\,(1+X_t)^2$. $(Y_t)_{t\in\mathcal{I}}$ and $(F_t)_{t\in\mathcal{I}}$ are now stochastic processes.

Let $\mathcal{I} = \mathbb{R}$ be the considered time interval and $(X_t)_{t\in\mathcal{I}}$ be a stochastic process with continuous paths. If we assume additionally that holds $Y_{t_0} = y_0$ with a deterministic initial value $y_0 > 0$ and a deterministic initial time $t_0 \in \mathbb{R}$, equation (3.2) has the unique pathwise random solution

$$Y_t = y_0 e^{-\beta(t-t_0)} + \int_{t_0}^{t} e^{-\beta(t-s)} F_s \,\mathrm{d}s, \quad t \in \mathcal{I}. \tag{3.3}$$

As a stochastic process $(X_t)_{t\in\mathbb{R}}$ we choose the stationary Ornstein-Uhlenbeck process (OUP). This is a Gaussian process with mean function constant zero and autocovariance function $\mathrm{Cov}(X_t, X_{t+h}) = \frac{\sigma^2}{2\theta}\exp(-\theta|h|)$. σ and θ are positive parameters. If the time interval is restricted to $\mathcal{I} = [0,\infty)$, the OUP can be also considered as a solution of the stochastic differential equation

$$\mathrm{d}X_t = -\theta X_t \mathrm{d}t + \sigma \mathrm{d}W_t \tag{3.4}$$

with a standard Wiener process $W = (W_t)_{t\geq 0}$ and an initial value X_0, which is independent of W and follows a centred normal distribution with variance $\frac{\sigma^2}{2\theta}$. This also allows to consider the equations (3.3) and (3.4) as a coupled system of stochastic differential equations

$$\mathrm{d}\begin{pmatrix} Y_t \\ X_t \end{pmatrix} = \begin{pmatrix} -\beta Y_t + f(t)\,(1+X_t)^2 \\ -\theta X_t \end{pmatrix} \mathrm{d}t + \begin{pmatrix} 0 \\ \sigma \end{pmatrix} \mathrm{d}W_t. \tag{3.5}$$

Such coupled systems are investigated in many ways and special methods are developed for them, but this path is not followed here. Instead the random differential equation (3.2) is investigated.

With the choice of X_t as Ornstein-Uhlenbeck process, there are explicit representations for certain characteristics of the solution (3.3) available. For example it holds for the mean function

$$\mathrm{E}\,[Y_t] = y_0 e^{-\beta(t-t_0)} + \left(1 + \frac{\sigma^2}{2\theta}\right) \cdot \int_{t_0}^{t} e^{-\beta(t-s)} f(s)\,\mathrm{d}s, \quad t \in \mathcal{I}. \tag{3.6}$$

More challenging is the question of how to determine characteristics for the arc radius function R and the arc voltage function U. The reason lies in the nonlinear relationship between the functions R and Y, and respectively U, Y and X. So it holds

$$R_t = Y_t^{\frac{1}{m+4}}, \quad U_t = k_3\,(1+X_t)^2\,Y_t^{-\frac{m+2}{m+4}} i(t), \quad t \in \mathcal{I}. \tag{3.7}$$

The Monte Carlo method can be applied by simulating paths of the OUP and computing paths of the functions U and R numerically. The red line in Figures 2(a) and 2(c) show the estimated mean function of U, which was gained by 1000 simulations of the OUP. The distance between the analytical mean function (3.6) and the estimated mean function gained by the Monte Carlo method in relation to the analytical mean function is always less than 2.6% in the considered interval (see figure 3).

Figures 2(b) and 2(d) show a single path of the arc voltage U. The graph shows a slight trembling, but not as strong as it can be observed in real world data. Figure 2(b) shows a random time varying amplitude of the arc voltage, similar as it can be observed in reality (see, e.g. [4] or [6]).

4. Series expansions of the pathwise solution

In this section let the considered time interval be a bounded interval $\mathcal{I} = [t_0, T]$ with $0 \leq t_0 < T$. According to the Karhunen-Loève theorem the Ornstein-Uhlenbeck process restricted to a bounded time interval can be expanded as

$$X_t = \sum_{k\in\mathbb{N}} \sqrt{\lambda_k}\,\phi_k(t)\xi_k, \quad t \in \mathcal{I}, \tag{4.1}$$

where $(\xi_k)_{k\in\mathbb{N}}$ is a sequence of independent standard normally distributed random variables, $(\lambda_k)_{k\in\mathbb{N}}$ are the eigenvalues and $(\phi_k)_{k\in\mathbb{N}}$ are the associated eigenfunctions of the covariance operator of X (cf. e.g. [7], chapter 11). The eigenvalues $(\lambda_k)_{k\in\mathbb{N}}$ are positive and converge to zero.

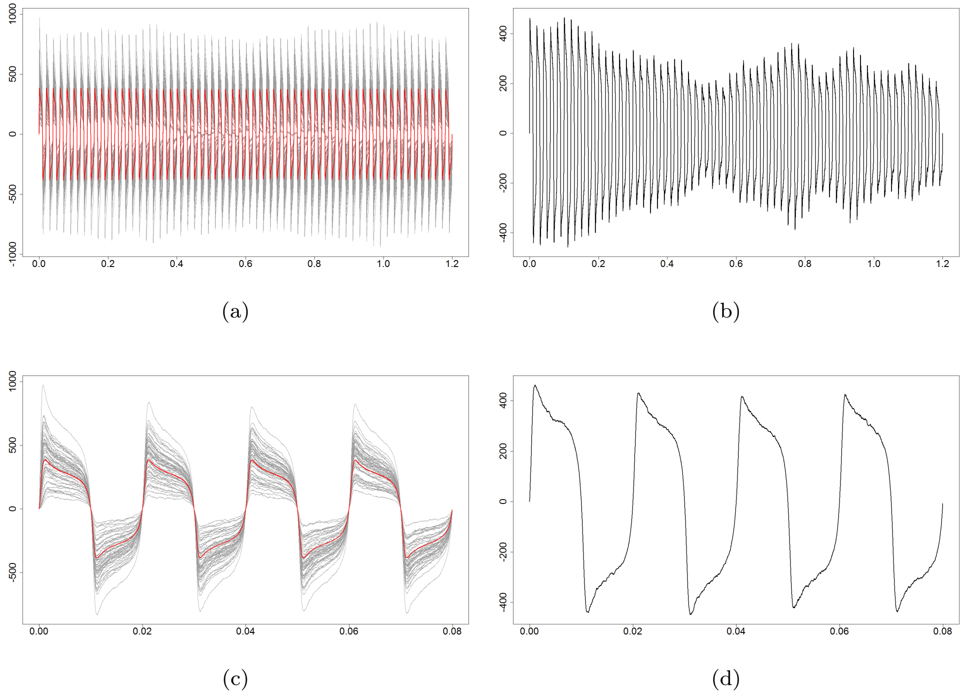


FIGURE 2. *Left (a and c)*: estimated mean function of the arc voltage U_t from 1000 simulated paths using Monte-Carlo method (red line) and 50 paths of U_t (grey lines). *Right (b and d)*: one single path of U_t simulations using Monte Carlo method. The following parameters were used.

a	ω	m	k_1	k_2	k_3	θ	σ	t_0	y_0
75	100π	0	2 000	5	35	0.5	0.5	0	6.58

The series (4.1) converges in the function space $\mathrm{L}^2([t_0, T])$ as well as in the space $\mathrm{C}([t_0, T])$ almost surely and in the p-th mean for all $1 \le p < \infty$ (cf. e.g. [2], chapter 3). For the OUP the eigenvalues can be calculated numerically and the eigenfunctions for given eigenvalues can be calculated analytically (cf. [3]).

The representation (4.1) can be applied to the random differential equation (3.2), which results in

$$\frac{\mathrm{d}Y_t}{\mathrm{d}t} = -\beta Y_t + f(t)\left(1 + \sum_{k\in\mathbb{N}} \sqrt{\lambda_k}\phi_k(t)\xi_k\right)^2, \quad t \in \mathcal{I}. \tag{4.2}$$

Expanding the square leads to an equation with a simple linear structure. This can be used to get a representation of the pathwise solution Y_t in terms of the random variables $(\xi_k)_{k\in\mathbb{N}}$ and their products $(\xi_k\xi_j)_{k,j\in\mathbb{N}}$.

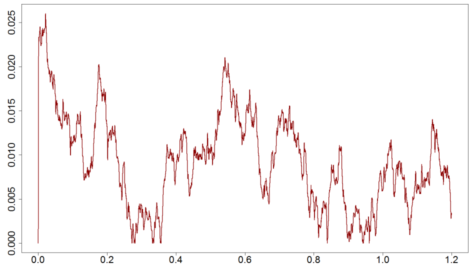


FIGURE 3. Graph of the function $t \mapsto \frac{|\mathrm{E}[Y_t]-\hat{\mathrm{E}}[Y_t]|}{\mathrm{E}[Y_t]}$, where $\hat{\mathrm{E}}[Y_t]$ is the estimated mean function estimated by the Monte Carlo method and $\mathrm{E}[Y_t]$ is the analytical mean function (3.6) with the same parameters as in figure 2.

Notice that these random variables do not form an orthogonal system. It holds

$$Y_t = y_0(t) + \sum_{k\in\mathbb{N}} y_k(t)\xi_k + \sum_{k\in\mathbb{N}} y_{kk}(t)\xi_k^2 + \sum_{\substack{k,j\in\mathbb{N},\\ k<j}} y_{kj}(t)\xi_k\xi_j, \quad t\in\mathcal{I} \tag{4.3}$$

where the deterministic functions y_k and y_{kj} are solutions of deterministic linear ordinary differential equations with corresponding initial values. It holds

$$\frac{\mathrm{d}y_0(t)}{\mathrm{d}t} = -\beta y_0(t) + f(t), \qquad y_0(t_0) = y_0, \tag{4.4}$$

$$\frac{\mathrm{d}y_k(t)}{\mathrm{d}t} = -\beta y_k(t) + 2\sqrt{\lambda_k}\phi_k(t)f(t), \qquad y_k(t_0) = 0, \tag{4.5}$$

$$\frac{\mathrm{d}y_{kk}(t)}{\mathrm{d}t} = -\beta y_{kk}(t) + \lambda_k\phi_k^2(t)f(t), \qquad y_{kk}(t_0) = 0 \tag{4.6}$$

for $k\in\mathbb{N}$ and

$$\frac{\mathrm{d}y_{kj}(t)}{\mathrm{d}t} = -\beta y_{kj}(t) + 2\sqrt{\lambda_k\lambda_j}\phi_k(t)\phi_j(t)f(t), \qquad y_{kj}(t_0) = 0 \tag{4.7}$$

for k, $j\in\mathbb{N}$ with $k<j$ and $t\in\mathcal{I}$. Explicit solutions can be given for these equations. It holds

$$y_0(t) = y_0 e^{-\beta(t-t_0)} + \int_{t_0}^{t} f(s)e^{-\beta(t-s)}\,\mathrm{d}s, \qquad t\in\mathcal{I}, \tag{4.8}$$

$$y_k(t) = 2\sqrt{\lambda_k}\int_{t_0}^{t} \phi_k(s)f(s)e^{-\beta(t-s)}\,\mathrm{d}s, \qquad t\in\mathcal{I}, \tag{4.9}$$

$$y_{kk}(t) = \lambda_k\int_{t_0}^{t} \phi_k^2(s)f(s)e^{-\beta(t-s)}\,\mathrm{d}s, \qquad t\in\mathcal{I} \tag{4.10}$$

for $k\in\mathbb{N}$ and

$$y_{kj}(t) = 2\sqrt{\lambda_k\lambda_j}\int_{t_0}^{t} \phi_k(s)\phi_j(s)f(s)e^{-\beta(t-s)}\,\mathrm{d}s, \qquad t\in\mathcal{I}, \tag{4.11}$$

for $k,\, j \in \mathbb{N}$ with $k < j$. For numerical computations the sums in equation (4.3) have to be truncated, which gives an approximation of the solution Y. The representation (4.3) could be achieved due to the simple structure of the random differential equation (3.2).

In general one can use an expansion of an already approximated solution in terms of orthogonal random variables, which is also known as polynomial chaos expansion. Therefore, let Y^N denote the pathwise solution (3.3) of our initial value problem, where the OUP $(X_t)_{t\in\mathcal{I}}$ is replaced by the truncated sum

$$X_t^N := \sum_{k=1}^{N} \sqrt{\lambda}_k \phi_k(t) \xi_k, \quad t \in \mathcal{I}, \tag{4.12}$$

with $N \in \mathbb{N}$. $\left(X_t^N\right)_{t\in\mathcal{I}}$ can be considered as a smoothed version of the OUP $(X_t)_{t\in\mathcal{I}}$. Then Y^N can be represented as polynomial chaos expansion

$$Y_t^N = \sum_{k=0}^{M} \tilde{y}_k(t) \Psi_k, \quad t \in \mathcal{I}, \tag{4.13}$$

with

$$M = \frac{(N+2)\,(N+1)}{2} - 1.$$

The $(\tilde{y}_k)_{0\leq k\leq M}$ are deterministic functions and the $(\Psi_k)_{0\leq k\leq M}$ are orthogonal random variables, for which applies that for every Ψ_k exists a N-variate polynomial p_k, such that

$$\Psi_k = p_k\left(\xi_1,\, \xi_2,\, \dots\, ,\, \xi_N\right), \tag{4.14}$$

in particular $\Psi_0 = 1$ and $\Psi_k = \xi_k$ for $k \in \{1, \dots, N\}$. In general the polynomial chaos expansion does not consist of a finite number of summands. But from the representation (4.3) follows, that one has to consider only polynomials up to degree two. It is convenient to choose the sequence $(\Psi_k)_{0\leq k\leq M}$, such that the degrees of the associated polynomials $(p_k)_{0\leq k\leq M}$ are increasing. In the considered case the sequence of polynomials can be determined by using the Hermite polynomials (cf. e.g. [7]). From the orthogonality of the random variables $(\Psi_k)_{0\leq k\leq M}$ follows, that holds

$$\mathrm{E}\left[Y^N(t)\Psi_k\right] = \tilde{y}_k(t) \cdot \mathrm{E}\left[\Psi_k^2\right]$$

for all $k \in \{0, 1, \dots, M\}$ and $t \in [t_0, T]$. This results in representations of the coefficient functions $(\tilde{y}_k)_{0\leq k\leq M}$. It holds

$$\tilde{y}_k(t) = \begin{cases} y_0 e^{-\beta(t-t_0)} + \int_{t_0}^{t} e^{-\beta(t-s)} f(s)\,\mathrm{d}s + \int_{t_0}^{t} e^{-\beta(t-s)} g_0(s)\,\mathrm{d}s, & \text{if } k = 0 \\ \int_{t_0}^{t} e^{-\beta(t-s)} g_k(s)\,\mathrm{d}s, & \text{if } k \geq 1 \end{cases} \tag{4.15}$$

with $k \in \{0, 1, \dots, M\}$, where the functions $(g_k)_{0\leq k\leq M}$ are defined by

$$g_k(t) := \begin{cases} f(t) \sum_{j=1}^{N} x_j^2(t) & \text{if } k = 0 \\ 2 f(t) x_k(t) & \text{if } 1 \leq k \leq N \\ f(t) \sum_{i=1}^{N} \sum_{j=1}^{N} x_i(t) x_j(t) M_{ijk} & \text{else} \end{cases} \tag{4.16}$$

with

$$x_j(t) := \sqrt{\lambda_j}\phi_j(t) \text{ for } j \in \{1, 2, \ldots, N\}$$

and

$$M_{ijk} := \frac{\mathrm{E}\left[\Psi_i \Psi_j \Psi_k\right]}{\mathrm{E}\left[\Psi_k^2\right]} \text{ for } i, j \in \{0, 1, \ldots, M\}\,.$$

5. Conclusion

One modeling approach of the power system of electric arc furnaces leads to a nonlinear ordinary differential equation, which in some important cases can be solved with a linear differential equation for an auxiliary quantity. Real data show partly an irregular behaviour so that a stochastic modeling seems to be advisable.

In the present work one such stochastic model is investigated. Thereby one coefficient of the differential equation is replaced by a stochastic process, leading to a random differential equation and hence also to a stochastic voltage process. The input stochastic process is modelled with the help of a stationary Ornstein-Uhlenbeck process, for which many properties and results are known. The random differential equation is investigated with the help of Monte Carlo method, but also the usage of polynomial chaos expansions is explained.

The results show a relatively good agreement with real data. In the future we plan to investigate further methods and models and we also plan to investigate methods for statistical inference from real data for the considered models.

References

[1] Acha, E., Semlyen, A., Rajaković, N., *A harmonic domain computational package for nonlinear problems and its application to electric arcs*, IEEE Transactions on Power Delivery, **5**(1990), no. 3, 1390-1397.

[2] Adler, R.J., Taylor, J.E., *Random Fields and Geometry*, Springer, New York, 2007.

[3] Corlay, S., Pagès, G., *Functional quantization-based stratified sampling methods*, Monte Carlo Methods Appl., De Gruyter, **21**(2015), no. 1, 1-32.

[4] Grabowski, D., *Selected Applications of stochastic approach in circuit theory*, Wydawnictwo Politechniki Ślaskiej, Gliwice, 2015.

[5] Grabowski, D., Walczak, J., *Analysis of deterministic model of electric arc furnace*, 10th International Conference on Environment and Electrical Engineering, 1-4.

[6] Grabowski, D., Walczak, J., Klimas, M., *Electric arc furnace power quality analysis based on stochastic arc model*, 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe, 1-6.

[7] Sullivan, T.J., *Introduction to Uncertainty Quantification*, Springer, Cham, 2012.

Hans-Jörg Starkloff
Technische Universität Bergakademie Freiberg
Faculty of Mathematics and Computer Science
Prüferstraße 9, 09599 Freiberg, Germany
e-mail: Hans-Joerg.Starkloff@math.tu-freiberg.de

Markus Dietz
Technische Universität Bergakademie Freiberg
Faculty of Mathematics and Computer Science
Prüferstraße 9, 09599 Freiberg, Germany
e-mail: Markus.Dietz@math.tu-freiberg.de

Ganna Chekhanova
Technische Universität Bergakademie Freiberg
Faculty of Mathematics and Computer Science
Prüferstraße 9, 09599 Freiberg, Germany
e-mail: Anna.Chekhanova@math.tu-freiberg.de

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 161–172
DOI: 10.24193/subbmath.2019.2.03

Finite difference scheme for a high order nonlinear Schrödinger equation with localized damping

Marcelo M. Cavalcanti, Wellington J. Corrêa, Mauricio A. Sepúlveda C. and Rodrigo Véjar Asem

Abstract. In this work we present a finite difference scheme used to solve a High order Nonlinear Schrödinger Equation. These equations can model the propagation of solitons travelling in fiber optics ([3], [11]). The scheme is designed to preserve the numerical energy at L^2 level, and control the energy at H^1 level for a suitable choose on the equation's parameters. We show numerical results displaying conservation properties of the schemes using solitons as initial conditions.

Mathematics Subject Classification (2010): 35Q55, 65-06, 65M06, 65Z05.

Keywords: High order, nonlinear Schrödinger equation, localized damping, dissipation, finite difference methods.

1. Introduction

We will study a numerical solution of a Higher order Non-Linear Schrödinger (HNLS) equation over an interval $\Omega := [0, L]$:

$$\begin{cases} iu_t + a_1 u_{xx} + ia_2 u_{xxx} + a_3|u|^2 u + ia_4|u|^2 u_x + ia_5 u|u|_x^2 + ia(x)u = 0, \ \Omega \times (0,T) \\ u(0,t) = u(L,t) = 0, \quad u_x(L,t) = 0, \ t \geq 0 \\ u(x,0) = u_0 \text{ in } \Omega \end{cases} \tag{1.1}$$

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.
Research of Marcelo M. Cavalcanti partially supported by the CNPq Grant 300631/2003-0.
Research of Wellington J. Corrêa partially supported by the CNPq Grant 438807/2018-9.
Research of Mauricio Sepúlveda C. was supported FONDECYT grant no. 1180868, and by CONICYT-Chile through the project AFB170001 of the PIA Program: Concurso Apoyo a Centros Científicos y Tecnológicos de Excelencia con Financiamiento Basal.
Rodrigo Véjar Asem acknowledges support by CONICYT-PCHA/Doctorado Nacional/2015-21150799.

where $a_1, a_2, a_3, a_4, a_5 \in \mathbb{R}$, $a_1, a_3 \geq 0$, and $u = u(x,t)$ is a complex valued function. The damping function $a(x)$ belongs to $L^\infty(\Omega)$ and assumed to be such that

$$a(x) \geqslant a_0 > 0 \text{ a.e. in an open, non-empty subset } \omega \text{ of } (0, L), \tag{1.2}$$

where it is acting effectively. This equation plays an important rule in soliton theory. It has applications in the propagation of femtosecond optical pulses in a monomode optical fiber, accounting for additional effects such as third order dispersion, self-steeping of the pulse, and self-frequency shift [11]. When $a(x) = 0$, we can also consider equation (1.1) as a generalization of the classical Nonlinear Schrödinger (NLS) equation

$$iu_t + a_1 u_{xx} + a_2 |u|^2 u = 0 \tag{1.3}$$

which can be obtained using $a_3 = a_4 = a_5 = 0$ in (1.1). This equation describes the electric field envelope of a laser beam in a medium with Kerr nonlinearity [7]. It is also known in plasma physics, where it describes Langmuir waves in a plasma with non-homogeneous density [10]. If in (1.1) we also take $a_1 = a_2 = 0$, $a_3 = 1$, $a_5 = 0$ and $a_4 = 6$, we can obtain a modified Korteweg-de Vries (KdV) equation which studies, for example, surface waves on conducting nonviscous incompressible liquid under the presence of a transverse electric field [16]. The KdV equation has also great importance in the study of surface water waves [12]. In this sense, numerically solving (1.1) can also solve many subproblems derived from it.

When considering $a(x) = 0$, Carvajal proved in [4] for $a_3 a_5 \neq 0$ the global well-posedness of the Cauchy Problem (1.1) in $H^s(\mathbb{R}), s > \frac{1}{4}$ when $3a_2a_3 = a_1a_4$. Meanwhile, Takaoka proved in [22], for $a_3 = 1$, the local well-posedness for the Cauchy Problem (1.1) in $H^s(\mathbb{T})$, $s > \frac{1}{2}$, where $\mathbb{T}$ is a unidimensional torus. Similar conclusions were obtained also by Takaoka in [21] for $a_3 = 0$, where the well-posedness is over $H^{\frac{1}{2}}(\mathbb{R})$. Regularity properties were studied by Alves et al. [2] when $a_4 = a_5 = 0$.

Exact solutions for (1.1) can be found using the Inverse Scattering Transform (IST) [1], proposed originally in Zakharov et al. [23]. Its integration depends on the values of a_3, a_4 and a_5. In particular: for $a_1 = \frac{1}{2}$, $a_2 = 1$, and rewriting equation (1.1) as

$$iu_t + \frac{1}{2} u_{xx} + |u|^2 u + i\varepsilon(\beta_1 u_{xxx} + \beta_2 |u|^2 u_x + \beta_3 |u|^2_x u) = 0 \tag{1.4}$$

where $\beta_1, \beta_2, \beta_3, \varepsilon$ are real constants, then exact solutions can be obtained via IST for the following cases:

- For the derivative NLS equation of type I: $\beta_1 : \beta_2 : \beta_3 = 0 : 1 : 1$ [3].
- For the derivative NLS equation of type II: $\beta_1 : \beta_2 : \beta_3 = 0 : 1 : 0$ [5].
- For the Hirota equation: $\beta_1 : \beta_2 : \beta_3 = 1 : 6 : 0$ [9].
- For the Sasa-Satsuma equation: $\beta_1 : \beta_2 : \beta_3 = 1 : 6 : 3$ [18].

Exact solutions are all of solitonic form. N-soliton solutions can also be obtained [9]. Potasek [17] shows some particular solutions that has been proven experimentally. But even when continuous solutions can be found for some specific initial conditions and some values for the real constans in (1.1), numerical solutions can prescinde from those requirements when computed. We can even use non-solitonic initial conditions in order to obtain a result. One way to compute numerical solutions is using the Finite

Difference Method, whose computational implementation can be done in an fast and efficient way.

Other ways to obtain numerical solutions for (1.1) has been studied by different authors in the recents years. One of the first scheme were proposed by Delfour, Fortin and Payre [6], which solves the NLS equation (1.3) proposing a rule to discretize powers of the nonlinearity multiplying the a_2 term. Their method has a strong property: it preserves the discrete versions of both the L^2 norm and the energy of the numerical solution, where their continuous versions are given by the following relations:

$$\begin{aligned} ||u||^2_{L^2(\Omega)}(t) &= \int_\Omega |u(x,t)|^2 dx \\ E(t) &:= \frac{\alpha}{2}\int_\Omega |\nabla u(x,t)|^2 dx - \frac{\gamma}{4}\int_\Omega |u(x,t)|^4 dx \end{aligned}$$

for $u = u(x,t) \in \Omega \subset \mathbb{R} \times \mathbb{R}^+ \longmapsto \mathbb{C}$ the exact solution of (1.1). The convergence of the numerical method is proved in Matsuo and Furihata [8]. Pazoto et al [15] proposed a finite difference scheme which solves the critical generalilzed Kortewetg-de Vries equation (GKdV-4) in a bounded domain. The higher-power term $u^4 u_x$ was rewritten as a linear combination of other derivatives in order to obtain specific conservation properties. Smadi and Bahloul [19] [20] combined a Compact Padé Finite Difference scheme [13] with a fourth order Runge-Kutta (RK4) scheme. They splitted the problem in two parts: a linear one which is solved using the finite difference scheme; and a nonlinear, which is solved using the RK4 scheme. The method was implemented with an interesting success, but no analysis of the error, convergence, or preserved quantities was made. We will compare their proposal with our method later on.

The structure of this work is as follows: Section 2 shows the numerical scheme we propose, along with some notation and its properties; Section 3 will present results for some experiments; and Section 4 contains conclusions regarding the results obtained.

2. Numerical scheme

2.1. Notation

Because of the boundary conditions given in Problem (1.1), and for the sake of the following analysis, we will introduce the vector space for a $M \in \mathbb{N}$ given:

$$X_M := \left\{ u = [u_0\, u_1\, \dots\, u_M]^T \in \mathbb{C}^{M+1} : u_0 = u_{M-1} = u_M = 0 \right\}$$

Let us introduce the classical finite differences operators for complex-valued arrays:

$$\begin{aligned} &\left[\boldsymbol{D}^+ u\right]_j := \frac{u_{j+1} - u_j}{\Delta x},\ \left[\boldsymbol{D}^- u\right]_j := \frac{u_j - u_{j-1}}{\Delta x} \\ &\boldsymbol{D} u := \frac{1}{2}\Big(\boldsymbol{D}^+ u + \boldsymbol{D}^- u\Big), \\ &\boldsymbol{D}^2 u := \boldsymbol{D}^+ \boldsymbol{D}^- u,\ \boldsymbol{D}^3 u := \boldsymbol{D}\boldsymbol{D}^+ \boldsymbol{D}^- u \end{aligned} \tag{2.1}$$

For $u, v \in X_M$, and $\Delta x := \frac{L}{M+1}$, let us consider the discrete space $L^2(0,L)_\Delta$ of complex-valued vectors endowed with the inner product

$$(u,v)_2 := \sum_{j=0}^{M} u_j \overline{v}_j \Delta x \tag{2.2}$$

this induces a discrete version of the L^2 norm:

$$||u||_2^2 := (u,u)_2.$$

and hence, at a timestep n, we define a numerical energy at L^2 level as follows:

$$E_{L^2}(u^n) := \frac{1}{2}||u^n||_2^2$$

Here we will write the functions that will approximate Problem (1.1). The linear terms will be approximated using the classical finite differences operators given in (2.1). The nonlinear terms are approximated as follows: for the term multiplied by a_3, we will use the method proposed in Delfour et al. [6]; this is:

$$|u(x_j,t_n)|^2 u(x_j,t_n) \approx |u_j^{n+\frac{1}{2}}|^2 \left(u_j^{n+\frac{1}{2}}\right)$$

For the term multiplied by a_4, we define

$$\begin{aligned} F_{a_4} : \mathbb{C}^M &\longrightarrow \mathbb{C}^M \\ u_j^{(p)} &\longrightarrow [F_{a_4}(u^{(p)})]_j := \frac{1}{2}\left|\frac{u_j^p + u_j^n}{2}\right|^2 D\left(\frac{u_j^p + u_j^n}{2}\right) \\ &\qquad + \frac{1}{2}\boldsymbol{D}\left(\left|\frac{u_j^p + u_j^n}{2}\right|^2 \frac{u_j^p + u_j^n}{2}\right) \\ &\qquad - \frac{1}{2}\frac{u_j^p + u_j^n}{2} D\left(\left|\frac{u_j^p + u_j^n}{2}\right|^2\right) \end{aligned} \tag{2.3}$$

The a_5 term will be discretized directly; this is, we define

$$\begin{aligned} F_{a_5} : \mathbb{C}^M &\longrightarrow \mathbb{C}^M \\ u_j^{(p)} &\longrightarrow [F_{a_5}(u^{(p)})]_j := \left(\frac{u_j^p + u_j^n}{2}\right) D\left(\left|\frac{u_j^p + u_j^n}{2}\right|^2\right) \end{aligned}$$

The given functions were defined in such a way that the numerical energy at L^2 level is conserved when $a(x) = 0$, $x \in \Omega$. Our proposal then reads as follows: $\forall j \in \{1, \ldots, M\}$, $\forall n \in \mathbb{N}$, and for a given $u^0 \in X_M$ the numerical scheme will be given component-wise by

$$\begin{cases} iD_t u_j^n + a_1 D^2(u_j^{n+\frac{1}{2}}) + ia_2 D^3(u_j^{n+\frac{1}{2}}) + a_3|u_j^{n+\frac{1}{2}}|^2 u_j^{n+\frac{1}{2}} \\ \quad + ia_4[F_{a_4}(u^{(n+1)})]_j + ia_5[F_{a_5}(u^{(n+1)})]_j + a(x_j)u_j^{n+\frac{1}{2}} = 0 \\ u^n \in X_M, \quad \forall n \in \mathbb{N} \\ u^0 \in X_M \ \text{ given.} \end{cases} \tag{2.4}$$

where $u_j^n \approx u(x_j, t_n)$ is the approximation of the exact solution $u(x,t)$ at the time $t_n = n\Delta t$ and at the coordinate $x_j = j\Delta x$; and $u_j^{n+\frac{1}{2}} := \frac{1}{2}\left(u_j^{n+1} + u_j^n\right)$. At each timestep, the scheme leads us to solve a linear system of equations, where the matrix involved is pentadiagonal; and a nonlinear problem solved using a Picard fixed-point iteration, a procedure similar to the one proposed in Delfour, Fortin, and Payre [6], which in turn induces us to use a small Δt in order to guarantee the contraction of the operator involved. The numerical scheme (2.4) has the following main properties:

Theorem 2.1. *Let $u^0 \in X_M : ||u^0||_2^2 < \infty$. If $a(x) = 0$, $\forall x \in \Omega$, then, $\forall n \in \mathbb{N}$, and for $u^n \in X_M$, we have*

$$E_{L^2}(u^{n+1}) = E_{L^2}(u^n) \tag{2.5}$$

On the other hand, we will consider the energy at H^1-level of the numerical solution at a timestep n as follows

$$E_{H^1}(u^n) := \frac{a_1}{2}||D^+u^n||_2^2 - \frac{a_3}{4}||u^n||_4^4 \tag{2.6}$$

where

$$\Delta E_{H^1} := \max_{n,m\in\mathbb{N}} \Big| E_{H^1}(u^n) - E_{H^1}(u^m) \Big|$$

Then, we have:

Theorem 2.2. *Let $u^n \in X_M$ the numerical solution of (1.1) using scheme (2.4) using $a(x) = 0$, $\forall x \in \Omega$. If in (1.1) $3a_2a_3 = a_1(a_4 + 2a_5)$, and if $C = \frac{1}{2}$ in (2.3), then the following property holds*

$$E_{H^1}(u^{n+1}) = E_{H^1}(u^n) + \mathcal{O}\Big(\Delta t(\Delta t^2 + \Delta x^2)\Big) \tag{2.7}$$

When a damping function $a(x)$ is present, then we have the following property:

Theorem 2.3. *Consider a sequence $\{u^n\}_{n\in\mathbb{N}} \subset X_M$ induced by the numerical scheme (2.4), and consider the function $a(x)$ and the set ω as defined in (1.2). If $u^0 \in L^2(0,L)_\Delta$, $6a_3 \geq |3a_4 + 2a_5|$ or $3a_4 + 2a_5 = 0$, and for $T_0 = n\Delta t > 0$, there exist a positive constant $C = C(T_0)$ and $\mu = \mu(T_0)$, both independent of Δx and Δt, such that the inequality*

$$E_{L^2}^n \leq C||u^0||_2^2 e^{-\mu n\Delta t}$$

holds for all $n > 0$.

3. Numerical results

In this section, we will start by comparing our scheme with previous works. We will then show some numerical experiments performed using a FORTRAN code which implements scheme (2.4). Given the particular interest on physical applications, our simulations will be performed over a domain $\Omega = [a,b]$, $a, b \in \mathbb{R}$ instead of the interval $[0,L]$. The results proved on this paper will still hold if we use a suitable coordinates change of the kind of $x = \tilde{x} - a$, $\tilde{x} \in [a,b]$, where the variable change is such that $L = b - a$.

3.1. Comparison with previous works

The only known work dealing with numerical methods for Problem (1.1) was proposed by Smadi and Bahloul [19] (from now on: SB scheme), which was implemented again in [20] with other examples. Their scheme splits Problem (1.1) in two parts: a linear one, which is solved using a Compact Finite Difference scheme [13]; and a non-linear part, which solves the problem through an explicit fourth-order Runge-Kutta method.

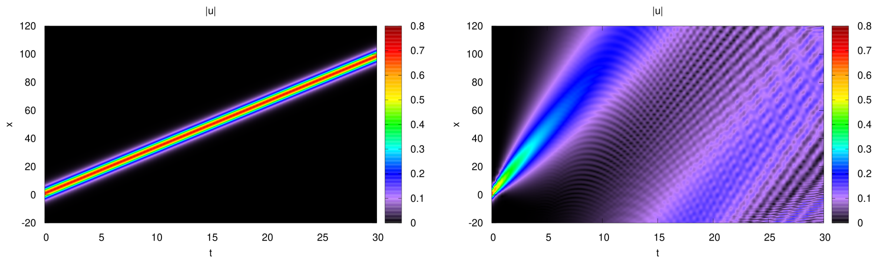


FIGURE 1. Numerical approximation when using (3.1) with $t = 0$ as initial condition, obtained by: at left, numerical scheme (2.4); at right, SB scheme.

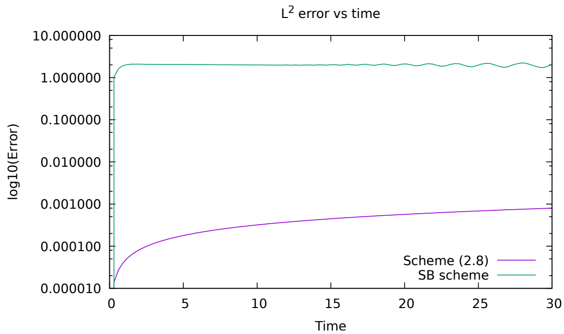


FIGURE 2. Semi-log plot of the comparison of the numerical error produced by both schemes.

In this subsection, we will compare the results computed by our proposal and the SB scheme for a particular case. To this end, we will solve equation (1.1) without damping term; this is, using $a(x) = 0$, $\forall x \in \Omega$. We will use the following solution obtained by Li [14]:

$$u(x,t) = \frac{i\rho}{cosh_q(\eta(x-\chi t))} e^{i(kt-\tilde{\Omega}x)}, \quad x \in \Omega := [-40,150], \quad t \in [0,30] \qquad (3.1)$$

where, in (1.1), $a_1 = 3$, $a_2 = 3$, $a_3 = 1$, $a_4 = 3$; while in (3.1), $\alpha_3 = -a3$, $\alpha_4 = -a4$,

$$a_5 = -(\alpha_5 + \alpha_4) = 0,\ \eta = \tfrac{1}{2},\ q = 3,\ \tilde{\Omega} = a_2/\alpha_4,\ k = -2\tilde{\Omega}^3\alpha_3,\ \chi = \alpha_3(\eta^2 + 3\tilde{\Omega}^2),$$

$$\rho = \sqrt{\frac{6\alpha_3\eta^2 q}{\alpha_4}} \text{ and } cosh_q(\xi) := \frac{e^{\xi} - qe^{-\xi}}{2}.$$

Figure 1 shows the results using the numerical scheme given in (2.4) using $\Delta t = 10^{-3}$ and $\Delta x = \frac{190}{2^{14}} \approx 0.0116$. Because there is no damping function, we expect the numerical preservation of the energy at levels L^2 and H^1. For our numerical scheme, we have that $\Delta E_{H^1} = 7.176423 \cdot 10^{-8}$, while $\Delta E_{L^2} = 6.883855 \cdot 10^{-10}$. The computation time was 450.835999s. Meanwhile, Figure 2 shows what is obtained using the SB scheme for the same mesh and timestep used for the results in Figure 1. Here, we have that $\Delta E_{H^1} = 175.938612$, $\Delta E_{L^2} = 9.723801 \cdot 10^{-4}$. The computation time was 271.823999s.

Remark 3.1. The paper of Smadi and Bahloul [19] does not give details regarding the discretization of the first-order derivatives in non-linear terms (the description of the Runge-Kutta method is presented in [19] in terms of a $f(u)$, and not of a $f(u, u_x)$). We assume in our simulations a finite deference centered as an approximation of u_x, however, it is worth noting that we tested with other more efficient approaches obtaining very similar results for the SB method that we show here.

3.2. Some numerical experiments

In this subsection of numerical examples we show some original results that test the adaptability of our numerical scheme for different situations.

3.2.1. First example: Effects of a strong damping. We assume the following initial condition $u(x, 0) = u_0 sech(kx)$, where $k = 1$ and $u_0 = \sqrt{6}$. We consider additionally, that, $a_1 = 3$, $a_2 = 1$, $a_3 = 0.03$, $a_4 = 0.1$, $a_5 = -0.05$, and $a(x) \equiv 0$ (that is, without damping term). Then an exact solution of (1.1) is obtained, which corresponds to a soliton of a hyperbolic secant squared pulses often referred to as "bright" pulses (see for more details Potasek and Tabor [17]). Now, the effect that we want to show in this example is what happens with this solution when adding a strong damping term. For that, we introduce a damping function concentrated in a neighborhood of the boundary of the spatial interval, given by

$$a(x) = \begin{cases} 1000, & x \in (-15, -10) \cup (10, 15) \\ 0, & \text{in other case.} \end{cases}$$

In our computations, $t \in [0, 1000]$, $x \in [-15, 15]$, $\Delta t = 0.00001$ and $\Delta x = \frac{30}{2^{13}} \approx 0.00366$. The form of the travelling soliton can be found in Figure 4. First, we observe that in the first times the wave propagates as the hyperbolic secant soliton predicted in Potasek and Tabor [17], while does not touch the support of the damping function. However, once the soliton approaches the area of influence (approximately at t = 180), the damping function is so high that the soliton gets reflected instead of proceeding with his original path. In each reflection the soliton loses energy at L^2 level following the exponential rate predicted in the previous theorems, and illustrated in Figure 3 left. The energy at H^1 level also decays at an exponential rate in each reflection.

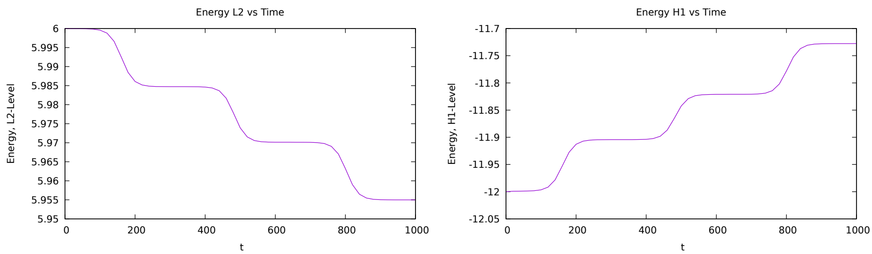


FIGURE 3. First case results. Left: time evolution of the L^2 energy. Right: time evolution of the H^1 energy.

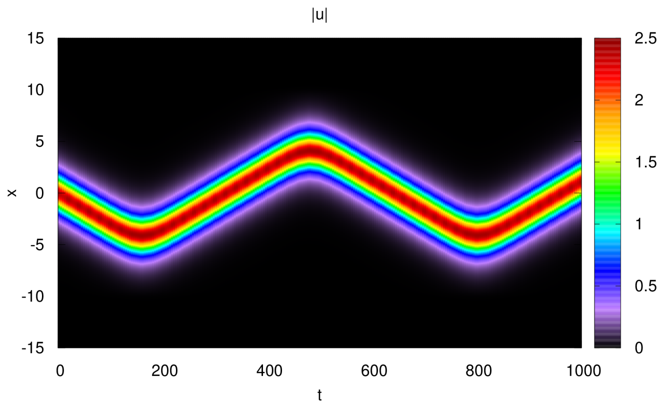


FIGURE 4. Time evolution of the travelling soliton for the first case.

3.2.2. Second example: Crossing of solitons. In this second example, we will simulate a cross between two solitons. For this we consider the exact solution of Hirota described in [9]. Then we consider our numerical scheme to approximate the Hirota equation:

$$iu_t + u_{xx} + |u|^2 u + i\frac{1}{10}u_{xxx} + i\frac{3}{10}|u|^2 u_x = 0.$$

for $(x,t) \in [-50,50]\times(0,15]$. At that time, we consider as initial condition, the solution of Hirota [9] for $t = 0$, that approximately corresponds to the sum of two hyperbolic secants of different amplitudes and centered in distant points. In this way, we calculate the numerical solution described in section 2 of this paper and we compare the result with the exact solution described in Hirota [9].

Given the absence of a damping function and because $3a_2a_3 = a_1a_4$, we conclude that there should not exist energy decay at L^2 and H^1 levels. For our calculations, we have made $\Delta t = 0.0001$, and $\Delta x = \frac{100}{2^{15}} \approx 0.00305$.

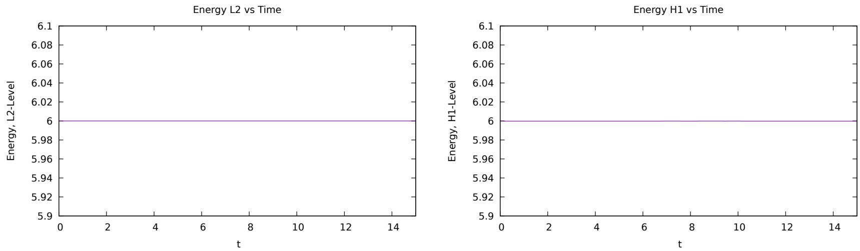


FIGURE 5. Preserved quantities for the 2 soliton experiment (second case). Left: L^2 level energy. Right: H^1 level energy.

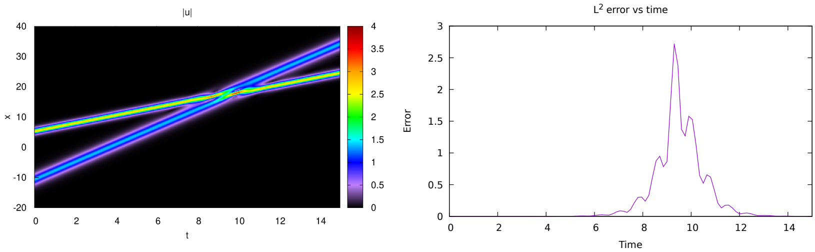


FIGURE 6. Left: the 2 soliton solution over time. Right: numerical error.

Regarding the error, we observed that the shape of the numerical solution moves away from the exact solution just at the moment of crossing between both solitons (see Figure 6 right). However, surprisingly we can notice that past the crossing, the numerical solution returns to reasonable levels of error ($t > 13$). The time evolution of the preserved quantities can be seen in Figure 5, where $\Delta E_{L^2} = 1.216 \times 10^{-8}$ and $\Delta E_{H^1} = 4.088 \times 10^{-5}$.

3.3. Computational performance

The computational performance will be now discussed. Using the same solution (3.1) from the previous subsection, Tables 1 and 2 illustrate how our scheme performs for two different timesteps; $\Delta t = 10^{-3}$ and $\Delta t = 10^{-4}$. $\Delta t = 10^{-2}$ was not included because it fails to guarantee the contraction condition of the operator involved in the nonlinear problem (2.4). For our computations, we have used a home computer equipped with a Linux operative system, an Intel Core i5-2400 chip with 4 processors at 3.10GHz each, and 9.7 GB of RAM memory. Parallelization was not implemented. From both tables, we see that the decrease of the numerical error is evident. We can also see a real improvement on the preservation of the H^1 energy.

Δx	$\lvert\lvert e\rvert\rvert^2_{L^2(0,T;L^2(\Omega))}$	t_{comp} [s]	ΔE_{L^2}	ΔE_{H^1}
$190/2^{10} \approx$ 1.885E-1	4.948753E-1	25.951	3.519406E-12	6.137745E-3
$190/2^{11} \approx$ 9.277E-2	3.024600E-2	52.292	4.636734E-12	3.440563E-4
$190/2^{12} \approx$ 4.638E-2	1.874793E-3	118.843	2.009226E-11	2.076023E-5
$190/2^{13} \approx$ 2.319E-2	1.163071E-4	230.372	6.786449E-11	1.250147E-6
$190/2^{14} \approx$ 1.159E-2	7.168988E-6	450.836	6.883855E-10	7.176423E-8
$190/2^{15} \approx$ 5.798E-3	4.869383E-7	1162.440	1.334082E-9	4.808387E-9

Table 1. Computational performance using (3.1) as initial condition and reference solution for $\Delta t = 0.001$.

Δx	$\lvert\lvert e\rvert\rvert^2_{L^2(0,T;L^2(\Omega))}$	t_{comp} [s]	ΔE_{L^2}	ΔE_{H^1}
$190/2^{10} \approx$ 1.885E-1	4.949154E-1	221.080	2.506295E-11	6.141681E-3
$190/2^{11} \approx$ 9.277E-2	3.025991E-2	442.384	2.235878E-12	3.449260E-4
$190/2^{12} \approx$ 4.638E-2	1.878424E-3	1099.167	9.875344E-11	2.097018E-5
$190/2^{13} \approx$ 2.319E-2	1.171612E-4	2168.396	2.507041E-6	2.301061E-6
$190/2^{14} \approx$ 1.159E-2	7.316259E-6	3558.420	5.695187E-10	8.371821E-8
$190/2^{15} \approx$ 5.798E-3	4.567125E-7	7768.200	1.664837E-9	2.460758E-9

Table 2. Computational performance using (3.1) as initial condition and reference solution for $\Delta t = 0.0001$.

4. Conclusion

In this work, we have proposed a new way to solve equation (1.1) using a finite difference method. The procedure involved the re-writing of a particular nonlinearity as a convex combination in order to get the conservation of the numerical energy at L^2 level when no damping term is present. The energy at H^1 level can also be controlled for sufficiently small values of Δt and Δx. When the damping term is present, the L^2 energy decays exponentially with time. We have also compared our proposal with the one from Smadi and Bahloul, observing an evident difference between both outputs. We deduce that our numerical method adapts better and more efficiently to the numerical resolution of the HNSL equation with respect to the known methods in the literature (see Smadi and Bahloul [19]), for various examples, with or without damping. Our method can also compute reasonable results using a small computer time at a home PC. Nevertheless, there is still room for improvement; in particular, about the contrast between the numerical and the exact solution when a collision between solitons happens. Also, the Picard iteration can be modified in order to be able to perform more calculations for smaller timesteps. Further studies are needed in both regards.

References

[1] Ablowitz, M.J., Segur, H., *Solitons and the Inverse Scattering Transform*, SIAM, 1981.

[2] Alves, M., Sepúlveda, M., Vera, O., *Smoothing properties for the higher-order nonlinear Schrödinger equation with constant coefficients*, Nonlinear Anal., **71**(2009), 948-966.

[3] Anderson, D., Lisak, M., *Nonlinear asymmetric self-phase modulation and self-steepening of pulses in long optical waveguides*, Phys. Rev. A, **27**(1983), no. 3, 1393-1398.

[4] Carvajal, X., *Sharp global well-posedness for a higher order Schrödinger equation*, J. Fourier Anal. Appl., **12**(2006), no. 1, 53-70.

[5] Chen, H.H., Lee, Y.C., Liu, C.S., *Integrability of nonlinear Hamiltonian systems by inverse scattering method*, Phys. Scr., **20**(1979), 490-492.

[6] Delfour, M., Fortin, M., Payre, G., *Finite-difference solutions of a non-linear Schrödinger equation*, J. Comput. Phys., **44**(1981), 277-288.

[7] Fibich, G., *Adiabatic law for self-focusing of optical beams*, Opt. Lett., **21**(1996), no. 21, 1735-1737.

[8] Furihata, D., Matsuo, T., *Discrete Variational Derivative Method: A Structure-Preserving Numerical Method for Partial Differential Equations*, Chapman & Hall/CRC, 2011.

[9] Hirota, R., *Exact envelope-soliton solutions of a nonlinear wave equation*, J. Math. Phys., **14**(1973), no. 7, 805-809.

[10] Kivshar, Y., Malomed, B., *Dynamics of solitons in nearly integrable systems*, Rev. Modern Phys., **61**(1989), no. 4, 763-915.

[11] Kodama, Y., Hasegawa, A., *Nonlinear pulse propagation in a monomode dielectric guide*, IEEE J. Quantum Electron., **QE-23**(1989), 510-524.

[12] Korteweg, D.J., De Vries, G., *On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves*, Philos. Mag., **539**(1895), 422-443.

[13] Lele, S.K., *Compact finite difference schemes with spectral-like resolution*, J. Comput. Phys., **103**(1992), 16-42.

[14] Li, B., *Exact soliton solutions for the higher-order nonlinear Schrödinger equation*, Int. J. Modern Phys. C, **16**(2005), no. 8, 1225-1237.

[15] Pazoto, A.F., Sepúlveda, M., Vera, O., *Uniform stabilization of numerical schemes for the critical generalized Korteweg-de Vries equation with damping*, Numer. Math., **116**(2010), 317-356.

[16] Perel'man, T.L., Fridman, A.Kh., El'yashevich, M.M., *A modified Korteweg-de Vries equation in electrohydrodynamics*, Sov. Phys. JETP, **39**(1974), no. 4, 643-646.

[17] Potasek, M.J., Tabor, M., *Exact solutions for an extended nonlinear Schrödinger equation*, Phys. Lett. A, **154**(1991), no. 9, 449-452.

[18] Sasa, N., Satsuma, J., *New-type of soliton solutions for a higher-order nonlinear Schrödinger equation*, J. Phys. Soc. Jpn., **60**(1991), no. 2, 409-417.

[19] Smadi, M., Bahloul, D., *A compact split step Padé scheme for higher-order nonlinear Schrödinger equation (HNLS) with power law nonlinearity and fourth order dispersion*, Comput. Phys. Commun., **182**(2011), 366-371.

[20] Smadi, M., Bahloul, D., *Dynamic of HNLS solitons using compact split step Padé scheme*, J. Phys. Conf. Ser., **574**(2015).

[21] Takaoka, H., *Well-posedness for the one-dimensional nonlinear Schrödinger equation with the derivative nonlinearity*, Adv. Differential Equations, **4**(1999), no. 4, 561-580.

[22] Takaoka, H., *Well-posedness for the higher order nonlinear Schrödinger equation*, Adv. Math. Sci. Appl., **10**(2000), no. 1, 149-171.

[23] Zakharov, V.E., Shabat, A.B., *Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media*, Sov. Phys. JETP, **34**(1972), no. 1, 62-69.

Marcelo M. Cavalcanti
Department of Mathematics, State University of Maringá
87020-900, Maringá, PR, Brazil
e-mail: mmcavalcanti@uem.br

Wellington J. Corrêa
Academic Department of Mathematics, Federal Technological University of Paraná
Campuses Campo Mourão, 87301-899, Campo Mourão, PR, Brazil
e-mail: wcorrea@utfpr.edu.br

Mauricio A. Sepúlveda C.
CI^2 MA and Departamento de Ingeniería Matemática
Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción
Casilla 160-C, Concepción, Chile
e-mail: mauricio@ing-mat.udec.cl

Rodrigo Véjar Asem
CI^2 MA and Departamento de Ingeniería Matemática
Facultad de Ciencias Físicas y Matemáticas, Universidad de Concepción
Casilla 160-C, Concepción, Chile
e-mail: rodrigovejar@ing-mat.udec.cl

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 173–182
DOI: 10.24193/subbmath.2019.2.04

Quantitative results for the convergence of the iterates of some King type operators

Marius Mihai Birou

Abstract. In this article we construct three q-King type operators which fix the functions e_0 and $e_2+\alpha e_1$, $\alpha > 0$. We study the rates of convergence for the iterates of these operators using the first and the second order modulus of continuity. We show that the convergence is faster in the case of q operators ($q < 1$) than in the classical case ($q = 1$).

Mathematics Subject Classification (2010): 41A17, 41A25, 41A36.

Keywords: King type operators, q-operators, convergence, modulus of smoothness.

1. Introduction

In [4] the authors introduced the operators $B_{n,\alpha} : C[0,1] \to C[0,1]$, $n > 1$, given by

$$B_{n,\alpha}f(x) = \sum_{k=0}^{n} \binom{n}{k} (u_{n,\alpha}(x))^k (1 - u_{n,\alpha}(x))^{n-k} f\left(\frac{k}{n}\right), \tag{1.1}$$

where $\alpha \in [0,\infty)$ and

$$u_{n,\alpha}(x) = -\frac{n\alpha + 1}{2(n-1)} + \sqrt{\frac{(n\alpha+1)^2}{4(n-1)^2} + \frac{n(\alpha x + x^2)}{n-1}}.$$

The operators $B_{n,\alpha}$ preserve the functions e_0 and $e_2+\alpha e_1$, where $e_i(x) = x^i$, $i = 0, 1, 2$. For $\alpha = 0$ the operator $B_{n,\alpha}$ reduces to the King operator (see [7]).

In this article we consider three q-operators of King type which fix the functions e_0 and $e_2 + \alpha e_1$. We study the convergence of the iterates of these operators.

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

Rates of convergence are obtained by using the first and the second order modulus of smoothness, i.e.

$$\omega_1(f,\delta) = \sup\left\{|f(x+h)-f(x)| : x, x+h \in [0,1],\ 0 \le h \le \delta\right\},$$

$$\omega_2(f,\delta) = \sup\left\{|f(x+h)-2f(x)+f(x-h)| : x, x\pm h \in [0,1],\ 0 \le h \le \delta\right\},$$

where $f \in C[0,1]$ and $\delta \ge 0$. We get better results in the case of the considered q-operators ($q<1$) than in the classical case ($q=1$).

Other quantitative results related to the convergence of the iterates of some positive linear operators may be found in [1], [2], [3], [6], [9], [8], [12].

We remind some notations from q-calculus which we use in the construction of the operators. For $q \in (0,1)$ we have:

- q-integer:

$$[n]_q = \frac{1-q^n}{1-q},\ n \in \mathbb{N},$$

- q-factorial:

$$[n]_q! = [n]_q[n-1]_q...[1]_q,\ n = 1,2,...,\ [0]_q! = 1,$$

- q-binomial coefficients:

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{[n]_q!}{[k]_q![n-k]_q!},$$

- q-integral

$$\int_0^1 f(x)d_qx = (1-q)\sum_{n=0}^{\infty} f(q^n)q^n.$$

2. Convergence of the iterates of the positive linear operators which preserve some functions

Let $\tau : [0,1] \to [0,1]$ a continuous strictly increasing functions satisfying the conditions $\tau(0) = 0$ and $\tau(1) = 1$.

Let $P : C[0,1] \to C[0,1]$ the operator given by

$$Pf(x) = (1-\tau(x))\, f(0) + \tau(x)f(1). \tag{2.1}$$

The following theorem is a direct consequence of Theorem 3.1 from [5].

Theorem 2.1. *Let $L : C[0,1] \to C[0,1]$ a positive linear operator which preserves the functions e_0, τ and has the set of interpolation points $\{0,1\}$. If there exists $\varphi \in C[0,1]$ such that*

$$L\varphi \gtrless \varphi \ \text{on}\ (0,1),$$

then

$$\lim_{m\to\infty} L^m f = Pf,$$

uniformly on $[0,1]$.

Theorem 2.2. [3] *Let $\tau : [0,1] \to [0,1]$, $\tau \in C^1[0,1]$, strictly increasing, $\tau(0) = 0$, $\tau(1) = 1$, $\tau'(0) \neq 1$, $\tau'(1) \neq 1$ and $\tau(x) \neq x$, $x \in (0,1)$. Let $L : C[0,1] \to C[0,1]$ be a linear positive operator which preserves e_0 and τ and let*

$$c = \sup_{0 \leq x \leq 1} \frac{\tau(x) - 2x\tau(x) + x^2}{|x - \tau(x)|} \tag{2.2}$$

and

$$\delta_m(x) = |L^m e_1(x) - \tau(x)|\,, \ x \in [0,1]. \tag{2.3}$$

If

$$0 < \delta_m(x) < 1/4, \ x \in (0,1),$$

then we have, for every $x \in [0,1]$,

$$|L^m f(x) - Pf(x)| \leq \sqrt{\delta_m(x)}\omega_1(f, \sqrt{\delta_m(x)}) + \left(1 + \frac{c}{2}\right)\omega_2(f, \sqrt{\delta_m(x)})$$

and

$$|L^m f(x) - Pf(x)| \leq 2\delta_m(x)\,\|f\| + \left(\frac{3}{2}\sqrt{\delta_m(x)} + \frac{7+c}{2}\right)\omega_2(f, \sqrt{\delta_m(x)}).$$

If we take

$$\tau = \frac{e_2 + \alpha e_1}{1 + \alpha},$$

with $\alpha \geq 0$, then the operator P from (2.1) becomes

$$Pf(x) = \frac{(1-\alpha)(1+x+\alpha)}{1+\alpha} f(0) + \frac{x(x+\alpha)}{1+\alpha} f(1), \ x \in [0,1], \ f \in C[0,1] \tag{2.4}$$

and for the constant c from (2.2) we get

$$c = \alpha + 2.$$

In the next sections we obtain estimations for $\delta_m(x)$, $x \in [0,1]$ given by (2.3) for three new operators which preserve e_0 and $e_2 + \alpha e_1$, $\alpha \geq 0$. Using Theorem 2.2 we get quantitative results for the convergence of the iterates of these operators.

3. The King modified q-Bernstein operator

The classical Berstein operator is given by

$$B_n f(x) = \sum_{k=0}^{n} p_{n,k}(x) f\left(\frac{k}{n}\right), \ f \in C[0,1], \ x \in [0,1],$$

where

$$p_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}.$$

In [11] Phillips constructed the q-Bernstein operator:

$$B_{n,q} f(x) = \sum_{k=0}^{n} p_{n,k}(q;x) f\left(\frac{[k]_q}{[n]_q}\right), \ q \in (0,1], \ f \in C[0,1], \ x \in [0,1],$$

where

$$p_{n,k}(q;x) = \left[\begin{array}{c} n \\ k \end{array} \right]_q x^k \prod_{s=0}^{n-k-1} (1 - q^s x).$$

The King modified q-Bernstein operator is given by

$$K^1_{n,q,\alpha} f(x) = B_n f(u_{n,q,\alpha}(x)),\ f \in C[0,1],\ x \in [0,1],$$

where $u_{n,q,\alpha} : [0,1] \to [0,1]$, $n > 1$ are continuous and strictly increasing functions having the properties $u_{n,q,\alpha}(0) = 0$, $u_{n,q,\alpha}(1) = 1$.
From properties of the the q-Bernstein operator we have

$$\begin{aligned} K^1_{n,q,\alpha} e_0(x) &= 1, \\ K^1_{n,q,\alpha} e_1(x) &= u_{n,q,\alpha}(x), \\ K^1_{n,q,\alpha} e_2(x) &= (u_{n,q,\alpha}(x))^2 + \frac{u_{n,q,\alpha}(x)(1 - u_{n,q,\alpha}(x))}{[n]_q}. \end{aligned}$$

Imposing that the operator $K^1_{n,q,\alpha}$ preserves the function $e_2 + \alpha e_1$, $\alpha \geq 0$ we get

$$u_{n,q,\alpha}(x) = -\frac{[n]_q \alpha + 1}{2([n]_q - 1)} + \sqrt{\frac{([n]_q \alpha + 1)^2}{4([n]_q - 1)^2} + \frac{[n]_q (\alpha x + x^2)}{[n]_q - 1}}.$$

Particular cases:

- $q = 1$ – the operator constructed in [4]
- $q = 1$, $\alpha = 0$ – the King operator (see [7])
- $\alpha = 0$ – the q-Bernstein King operator (see [8])

Theorem 3.1. *The sequence of the iterates of the operator $K^1_{n,q,\alpha}$ converges uniformly to the operator P given by (2.4).*

Proof. For every $x \in [0,1]$ we have

$$\begin{aligned} K^1_{n,q,\alpha} e_1(x) - x &= u_{n,q,\alpha}(x) - x \qquad (3.1) \\ &= \frac{x(x-1)}{[n]_q - 1} \cdot \frac{1}{\frac{[n]_q \alpha + 1}{2([n]_q - 1)} + \sqrt{\frac{([n]_q \alpha + 1)^2}{4([n]_q - 1)^2} + \frac{[n]_q (\alpha x + x^2)}{[n]_q - 1}} + x} \end{aligned}$$

It follows that

$$K^1_{n,q,\alpha} e_1(x) - x \leq 0,\ x \in [0,1],$$

with equality if and only if $x \in \{0,1\}$. The conclusion follows from Theorem 2.1 by taking $\varphi = e_1$. □

Theorem 3.2. *If*

$$\delta^1_{m,n,q,\alpha}(x) = \left(K^1_{n,q,\alpha}\right)^m e_1(x) - \frac{x^2 + \alpha x}{1 + \alpha},\ x \in [0,1],$$

then we have the estimation

$$\delta^1_{m,n,q,\alpha}(x) \leq \left(\frac{(\alpha + 2)([n]_q - 1)}{(\alpha + 2)[n]_q - 1} \right)^m \frac{x(1-x)}{1+\alpha} = \lambda^1_{m,n,q,\alpha}(x),\ x \in [0,1]. \qquad (3.2)$$

Proof. For $x \in \{0,1\}$ we get $\delta^1_{m,n,q,\alpha}(x) = 0$ and (3.2) holds.
For $x \in (0,1)$, using (3.1) we get

$$\frac{K^1_{n,q,\alpha}e_1(x) - x}{x(1-x)} \le \frac{-1}{[n]_q\alpha + 2[n]_q - 1}.$$

We observe that

$$\frac{K^1_{n,q,\alpha}e_1(x) - x^2}{x(1-x)} = \frac{K^1_{n,q,\alpha}e_1(x) - x}{x(1-x)} + 1 \le \frac{[n]_q\alpha + 2[n]_q - 2}{[n]_q\alpha + 2[n]_q - 1}.$$

It follows that

$$\begin{aligned} K^1_{n,q,\alpha}e_1(x) - \frac{x^2+\alpha x}{1+\alpha} &= K^1_{n,q,\alpha}e_1(x) - x^2 + x^2 - \frac{x^2+\alpha x}{1+\alpha} \\ &= K^1_{n,q,\alpha}e_1(x) - x^2 - \frac{\alpha}{\alpha+1}x(1-x) \\ &\le \left(\frac{[n]_q\alpha + 2[n]_q - 2}{[n]_q\alpha + 2[n]_q - 1} - \frac{\alpha}{\alpha+1}\right)x(1-x) \\ &= \frac{(\alpha+2)\left([n]_q - 1\right)}{(\alpha+2)[n]_q - 1}\cdot\left(x - \frac{x^2+\alpha x}{1+\alpha}\right) \end{aligned}$$

Taking into account that the operator $K^1_{n,q,\alpha}$ preserves $\left(e_2 + \alpha e_1\right)/(1+\alpha)$ we obtain

$$\left(K^1_{n,q,\alpha}\right)^m e_1(x) - \frac{x^2+\alpha x}{1+\alpha} \le \left(\frac{(\alpha+2)\left([n]_q - 1\right)}{(\alpha+2)[n]_q - 1}\right)^m\cdot\left(x - \frac{x^2+\alpha x}{1+\alpha}\right)$$

and the conclusion follows. □

Theorem 3.3. *We have the following estimations:*

$$\left|(K^1_{n,q,\alpha})^m f(x) - Pf(x)\right| \tag{3.3}$$

$$\le \sqrt{\lambda^1_{m,n,q,\alpha}(x)}\omega_1\left(f, \sqrt{\lambda^1_{m,n,q,\alpha}(x)}\right) + \left(2 + \frac{\alpha}{2}\right)\omega_2\left(f, \sqrt{\lambda^1_{m,n,q,\alpha}(x)}\right),\ x \in [0,1]$$

and

$$\left|(K^1_{n,q,\alpha})^m f(x) - Pf(x)\right| \tag{3.4}$$

$$\le 2\lambda^1_{m,n,q,\alpha}(x)\,\|f\| + \left(\frac{3}{2}\sqrt{\lambda^1_{m,n,q,\alpha}(x)} + \frac{9+\alpha}{2}\right)\omega_2(f, \sqrt{\lambda^1_{m,n,q,\alpha}(x)}),\ x \in [0,1].$$

Proof. The conclusion follows using Theorem 2.2 and Theorem 3.2. □

For $\alpha = 0$ the estimations (3.3) and (3.4) were obtained in [2].
The function $h^1_{m,\alpha} : [1,\infty) \to \mathbb{R}, m \ge 1$ defined by

$$h^1_{m,\alpha}(t) = \left(\frac{(2+\alpha)(t-1)}{t(2+\alpha) - 1}\right)^m$$

is strictly increasing. If $0 < q_1 < q_2 \le 1$ then $[n]_{q_1} < [n]_{q_2}$ and therefore

$$h^1_{m,\alpha}([n]_{q_1}) < h^1_{m,\alpha}([n]_{q_2}),\ m \ge 1.$$

From Theorem 3.3 it follows that the estimation $\left|(K^1_{n,q_1,\alpha})^m f(x) - Pf(x)\right|$, $x \in [0,1]$ is smaller than the estimation $\left|(K^1_{n,q_2,\alpha})^m f(x) - Pf(x)\right|$, $x \in [0,1]$. Taking $q_1 = q \in$

$(0,1)$ and $q_2 = 1$ we get that the q-operator $K^1_{n,q,\alpha}$ provides better convergence of the iterates than the classical operator $K^1_{n,1,\alpha}$.

4. The King modified q-Stancu operator

The q-Stancu operator constructed by Nowak in [10] is given by

$$P_n^{q,a} f(x) = \sum_{k=0}^{n} w_{n,k}^{q,a}(x) f\left(\frac{[k]_q}{[n]_q}\right), \ a \geq 0, \ q \in (0,1], \ f \in C[0,1], \ x \in [0,1],$$

where

$$w_{n,k}^{q,a}(x) = \begin{bmatrix} n \\ k \end{bmatrix}_q \frac{\prod_{i=0}^{k-1}(x + a[i]_q) \prod_{i=0}^{n-1-k}(1 - q^i x + a[i]_q)}{\prod_{i=0}^{n-1}(1 + a[i]_q)}.$$

Particular cases:

- $q = 1$ – the Stancu operator (see [13])
- $a = 0$ – the q-Bernstein operator (see [11])

We consider the King modified q-Stancu operator

$$K^2_{n,q,a,\alpha} f(x) = P_n^{q,a} f(u_{n,q,a,\alpha}(x)),$$

where $u_{n,q,a,\alpha} : [0,1] \to [0,1]$, $n > 1$ are continuous and strictly increasing functions having the properties $u_{n,q,a,\alpha}(0) = 0$, $u_{n,q,a,\alpha}(1) = 1$.
From [10] we have

$$K^2_{n,q,a,\alpha} e_0(x) = 1,$$
$$K^2_{n,q,a,\alpha} e_1(x) = u_{n,q,a,\alpha}(x),$$
$$K^2_{n,q,a,\alpha} e_2(x) = \tfrac{1}{a+1}\left(\tfrac{u_{n,q,a,\alpha}(x)(1-u_{n,q,a,\alpha}(x))}{[n]_q} + u_{n,q,a,\alpha}(x)(x + u_{n,q,a,\alpha}(x))\right).$$

If

$$u_{n,q,a,\alpha}(x) = -\beta_{n,q,a,\alpha} + \sqrt{\beta^2_{n,q,a,\alpha} + \frac{[n]_q(1+a)(\alpha x + x^2)}{[n]_q - 1}},$$

where

$$\beta_{n,q,a,\alpha} = \frac{1 + [n]_q(a + \alpha + a\alpha)}{2([n]_q - 1)},$$

then the operator $K^2_{n,q,a,\alpha}$ fixes the functions e_0 and $e_2 + \alpha e_1$, $\alpha \geq 0$.

Theorem 4.1. *The sequence of the iterates of the operator $K^2_{n,q,a,\alpha}$ converges uniformly to the operator P given by* (2.4).

Proof. We use Theorem 2.1 with $\varphi = e_1$. Indeed, we have

$$K^2_{n,q,a,\alpha} e_1(x) - x = u_{n,q,a,\alpha}(x) - x$$

and therefore

$$K^2_{n,q,a,\alpha} e_1(x) - x = \frac{x(x-1)}{[n]_q - 1} \cdot \frac{1 + [n]_q a}{\beta_{n,q,a,\alpha} + \sqrt{\beta^2_{n,q,a,\alpha} + \frac{[n]_q(1+a)(\alpha x + x^2)}{[n]_q - 1}} + x} \leq 0, \quad (4.1)$$

for all $x \in [0,1]$, with equality only for $x \in \{0,1\}$. □

Theorem 4.2. *If*

$$\delta^2_{m,n,q,a,\alpha}(x) = \left(K^2_{n,q,a,\alpha}\right)^m e_1(x) - \frac{x^2+\alpha x}{1+\alpha}, \ x \in [0,1],$$

then we have the estimation

$$\delta^2_{m,n,q,a,\alpha}(x) \le \left(\frac{(\alpha+2)\left([n]_q - 1\right)}{(a\alpha+\alpha+a+2)\,[n]_q - 1}\right)^m \frac{x(1-x)}{1+\alpha} = \lambda^2_{m,n,q,a,\alpha}(x),$$

for all $x \in [0,1]$.

Proof. From (4.1) it follows that

$$\frac{\left(K^2_{n,q,a,\alpha}\right)^m e_1(x) - x}{x(1-x)} \le -\frac{1+[n]_q a}{2\left([n]_q - 1\right)\left(1+\beta_{n,q,a,\alpha}\right)}, \ x \in (0,1),$$

Using the same steps as in Theorem 3.2 we get the conclusion. □

Theorem 4.3. *We have the following estimations:*

$$\left|(K^2_{n,q,a,\alpha})^m f(x) - Pf(x)\right| \le \sqrt{\lambda^2_{m,n,q,a,\alpha}(x)}\,\omega_1\left(f, \sqrt{\lambda^2_{m,n,q,a,\alpha}(x)}\right)$$

$$+\left(2+\frac{\alpha}{2}\right)\omega_2\left(f, \sqrt{\lambda^2_{m,n,q,a,\alpha}(x)}\right), \ x \in [0,1]$$

and

$$\left|(K^2_{n,q,a,\alpha})^m f(x) - Pf(x)\right| \le 2\lambda^2_{m,n,q,a,\alpha}(x)\,\|f\|$$

$$+\left(\frac{3}{2}\sqrt{\lambda^2_{m,n,q,a,\alpha}(x)} + \frac{9+\alpha}{2}\right)\omega_2(f, \sqrt{\lambda^2_{m,n,q,a,\alpha}(x)}), \ x \in [0,1].$$

Proof. The conclusion follows from Theorem 2.2 and Theorem 4.2. □

The function $h^2_{m,a,\alpha} : [1,\infty) \to \mathbb{R}$, $m \ge 1$ defined by

$$h^2_{m,a,\alpha}(t) = \left(\frac{(2+\alpha)(t-1)}{t(2+\alpha+a+a\alpha)-1}\right)^m$$

is strictly increasing. If $0 < q_1 < q_2 \le 1$ then

$$h^2_{m,a,\alpha}([n]_{q_1}) < h^2_{m,a,\alpha}([n]_{q_2}), \ m \ge 1.$$

From Theorem 4.3 it follows that the estimation $\left|(K^2_{n,q_1,a,\alpha})^m f(x) - Pf(x)\right|$, $x \in [0,1]$ is smaller than the estimation $\left|(K^2_{n,q_2,a,\alpha})^m f(x) - Pf(x)\right|$, $x \in [0,1]$. In particular, taking $q_1 = q \in (0,1)$ and $q_2 = 1$ we get that the q-operator $K^2_{n,q,a,\alpha}$ has a better rate of convergence for the iterates than the operator $K^2_{n,1,a,\alpha}$.

5. The King modified q-genuine Bernstein-Durrmeyer operator

The q-genuine Bernstein-Durrmeyer operator introduced in [9] is given by

$$U_{n,q}f(x) = p_{n,0}(q;x)f(0) + p_{n,n}(q;x)f(1) \tag{5.1}$$

$$+[n-1]_q \sum_{k=1}^{n-1} q^{1-k} p_{n,k}(q;x) \int_0^1 p_{n-2,k-1}(q;qt)f(t)d_qt,$$

for every $f \in C[0,1]$ and every $x \in [0,1]$. For $q=1$ we get the classical genuine Bernstein-Durrmeyer operator.
We consider the King modification of the q-genuine Bernstein-Durrmeyer operator

$$K^3_{n,q,\alpha}f(x) = U_{n,q}f(u_{n,q,\alpha}(x)),$$

where $u_{n,q,\alpha} : [0,1] \to [0,1]$, $n > 1$ are continuous and strictly increasing functions having the properties $u_{n,q,\alpha}(0) = 0$, $u_{n,q,\alpha}(1) = 1$.
From [9] we have

$$K^3_{n,q,\alpha}e_0(x) = 1,$$
$$K^3_{n,q,\alpha}e_1(x) = u_{n,q,\alpha}(x),$$
$$K^3_{n,q,\alpha}e_2(x) = (u_{n,q,\alpha}(x))^2 + \frac{[2]_q u_{n,q,\alpha}(x)(1-u_{n,q,\alpha}(x))}{[n+1]_q}.$$

If

$$u_{n,q,\alpha}(x) = -\frac{[n+1]_q\alpha + [2]_q}{2([n+1]_q - [2]_q)} + \sqrt{\frac{([n+1]_q\alpha + [2]_q)^2}{4([n+1]_q - [2]_q)^2} + \frac{[n+1]_q(\alpha x + x^2)}{[n+1]_q - [2]_q}}, \tag{5.2}$$

then the operator $K^3_{n,q,\alpha}$ preserves the functions e_0 and $e_2 + \alpha e_1$, $\alpha > 0$.

Theorem 5.1. *The sequence of the iterates of the operator $K^3_{n,q,\alpha}$ converges uniformly to the operator P given by (2.4).*

Proof. We have

$$\begin{aligned} K^3_{n,q,\alpha}e_1(x) - x &= u_{n,q,\alpha}(x) - x \\ &= \frac{x(x-1)}{[n+1]_q - [2]_q} \cdot \frac{[2]_q}{\gamma_{n,q,\alpha} + \sqrt{\gamma^2_{n,q,\alpha} + \frac{[n+1]_q(\alpha x + x^2)}{[n+1]_q - [2]_q}} + x}, \end{aligned}$$

where

$$\gamma_{n,q,\alpha} = \frac{[n+1]_q\alpha + [2]_q}{2([n+1]_q - [2]_q)}.$$

It follows that

$$K^3_{n,q,\alpha}e_1(x) - x \le 0, \; x \in [0,1],$$

with equality only for $x \in \{0,1\}$. Using Theorem 2.1 with $\varphi = e_1$ we get the conclusion. □

Theorem 5.2. *If*

$$\delta^3_{m,n,q,\alpha}(x) = (K^3_{n,q,\alpha})^m e_1(x) - \frac{x^2+\alpha x}{1+\alpha},\ x \in [0,1],$$

then we get

$$\delta^3_{m,n,q,\alpha}(x) \le \left(\frac{(\alpha+2)\left([n+1]_q - [2]_q\right)}{(\alpha+2)\,[n+1]_q - [2]_q}\right)^m \frac{x(1-x)}{1+\alpha} = \lambda^3_{m,n,q,\alpha}(x),$$

for all $x \in [0,1]$.

Proof. We get the conclusion using the same steps as in Theorem 3.2 and taking into account the inequality

$$\frac{\left(K^3_{n,q,\alpha}\right)^m e_1(x) - x}{x(1-x)} \le -\frac{[2]_q}{(\alpha+1)[n+1]_q - [2]_q},\ x \in (0,1). \qquad \square$$

Theorem 5.3. *We have the following estimations:*

$$\left|(K^3_{n,q,\alpha})^m f(x) - Pf(x)\right| \le \tag{5.3}$$

$$\sqrt{\lambda^3_{m,n,q,\alpha}(x)}\omega_1\left(f, \sqrt{\lambda^2_{m,n,q}(x)}\right) + \left(2+\frac{\alpha}{2}\right)\omega_2\left(f, \sqrt{\lambda^3_{m,n,q,\alpha}(x)}\right),\ x \in [0,1]$$

and

$$\left|(K^3_{n,q,\alpha})^m f(x) - Pf(x)\right| \le \tag{5.4}$$

$$2\lambda^2_{m,n,q,\alpha}(x)\,\|f\| + \left(\frac{3}{2}\sqrt{\lambda^3_{m,n,q,\alpha}(x)} + \frac{9+\alpha}{2}\right)\omega_2(f, \sqrt{\lambda^3_{m,n,q,\alpha}(x)}),\ x \in [0,1].$$

Proof. The conclusion follows from Theorem 2.2 and Theorem 5.2. $\square$

For $\alpha = 0$ the estimations (5.3) and (5.4) were obtained in [3].

The function $h^3_{m,n,\alpha} : (0,1] \to \mathbb{R}$, $m \ge 1$ defined by

$$h^3_{m,n,\alpha}(q) = \left(\frac{(\alpha+2)(q^2-q^{n+1})}{(\alpha+2)(1-q^{n+1}) - 1 + q^2}\right)^m$$

is strictly increasing. From Theorem 5.3 it follows that if $0 < q_1 < q_2 \le 1$ then the estimation $\left|(K^3_{n,q_1,\alpha})^m f(x) - Pf(x)\right|$, $x \in [0,1]$ is smaller than the estimation $\left|(K^3_{n,q_2,\alpha})^m f(x) - Pf(x)\right|$, $x \in [0,1]$. Taking $q_1 = q \in (0,1)$ and $q_2 = 1$ we get that the convergence of the iterates of the q-operator $K^3_{n,q,\alpha}$ is better than that of the operator $K^3_{n,1,\alpha}$.

References

[1] Altomare, F., Cappelletti Montano, M., Leonessa, V., Rasa, I., *Markov Operators, Positive Semigroups and Approximation Process*, De Gruyter Stud. Math., **61**(2015).

[2] Birou, M.M., *New rates of convergence for the iterates of some positive linear operators*, Mediterr. J. Math., **14**(2017), no. 3, Art. no. UNSP 129.

[3] Birou, M.M., *New quantitative results for the convergence of the iterates of some positive linear operators*, Positivity, **32**(2019), no. 2, 315-326.

[4] Cardenas-Morales, D., Garrancho, P., Muñoz-Delgado, F.J., *Shape preserving approximation by Bernstein-type operators which fix polynomials*, Appl. Math. Comput., **182**(2006), no. 2, 1615-1622.

[5] Gavrea, I., Ivan, M., *Asymptotic behaviour of the iterates of positive linear operators*, Abstr. Appl. Anal., 2011, Art. ID 670509, 11 pp.

[6] Gonska, H., Raşa, I., *On infinite products of positive linear operators reproducing linear functions*, Positivity, **17**(2013), 67-79.

[7] King, J.P, *Positive linear operators which preserve* x^2, Acta Math. Hungar., **99**(2003), no. 3, 203-208.

[8] Mahmudov, N.I., *Asymptotic properties of powers of linear positive operators which preserve* e_2, Comput. Math. Appl., **62**(2011), 4568-4575.

[9] Mahmudov, N.I., Sabancigil, P., *On genuine q-Bernstein-Durrmeyer operators*, Publ. Math. Debrecen, **76**(2010), no. 4, 221-229.

[10] Nowak, G., *Approximation properties for generalized q-Bernstein polynomials*, J. Math. Anal. Appl., **350**(2009), 50-55.

[11] Philips, G.M., *Bernstein polynomial based on q-integers*, Ann. Numer. Math., **4**(1997), 511-518.

[12] Raşa, I., C_0*-semigroups and iterates of positive linear operators: asymptotic behaviour*, Rend. Circ. Mat. Palermo (2), **82**(2010), 1-20.

[13] Stancu, D.D., *Approximation of functions by a new class of linear polynomial operators*, Rev. Roumaine Math. Pures Appl., **13**(1968), no. 8, 1173-1194.

Marius Mihai Birou
Technical University of Cluj Napoca
Faculty of Automatics and Computer Sciences
28, Memorandumului Street,
400114 Cluj-Napoca, Romania
e-mail: Marius.Birou@math.utcluj.ro

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 183–196
DOI: 10.24193/subbmath.2019.2.05

Some approximation properties of Urysohn type nonlinear operators

Harun Karsli

Abstract. The central issue of this paper is to continue the investigation of convergence properties of Urysohn type operators. By using Urysohn type operators we will extend the theory of interpolation to functionals and operators. In details, the present paper centers around Urysohn type nonlinear counterpart of the two dimensional Stancu operators defined on a triangle. We construct our nonlinear operators by defining a nonlinear forms of the kernel functions. Afterwards, we investigate the convergence problem for these operators.

Mathematics Subject Classification (2010): 41A25, 41A35, 47G10, 47H30.

Keywords: Urysohn integral operators, Stancu operator, two dimensional nonlinear Stancu operators, Urysohn type nonlinear Stancu operators.

1. Introduction

In functional analysis, the superposition problem is known as the problem of representing a function f as the composition of "simpler and more easily calculated" functions. In 1885, Weierstrass gave a positive answer to this problem with his famous theorem, which states that every continuous function defined on a closed interval $[a, b]$ can be uniformly approximated by a sequence of polynomials. Since that time many researchers try to find an explicit form of such polynomials to give a simple proof of this theorem. A well-known and most celebrated proof of the Weierstrass approximation theorem for $f \in C[0, 1]$ is due to Bernstein, in which he defined the following polynomials

$$(B_n f)(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) p_{n,k}(x), \quad n \geq 1, \tag{1.1}$$

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

where $p_{n,k}(x) = \left(\begin{array}{c} n \\ k \end{array} \right) x^k(1-x)^{n-k}$ is the Binomial distribution, and proved that $B_n f$ converges uniformly to any $f \in C[0,1]$ (see [7]). Further investigations are obtained by Lorentz in [19]. Since Bernstein operators are the prototype of many positive linear operators used in the theory of approximation, a great number of generalizations of these operators are given.

For the same functions, Stancu defined another positive linear operator as follows

$$\left(P_n^{\alpha} f\right)(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) p_{n,k}^{\alpha}(x), \quad n \geq 1,$$

where α is a non-negative parameter, which may depend only on the natural number n and $p_{n,k}^{\alpha}(x)$ called Markov-Polya distribution (see [23]).

The special case $\alpha = 0$ yields the Bernstein operator, while the Szasz-Mirakyan operator is shown to be a limiting case of P_n^{α}. When $\alpha = 1/n$ we obtain the Lupaş and Lupaş [20] operators corresponding to the equally spaced points k/n $(k = 0, 1, ..., n)$.

Up to the work of the famous polish mathematician Julian Musielak in 1981, see [22], the theory of approximation was strongly related with the linearity of the considered operators. Based on the idea developed in [22] and afterwards the works of C. Bardaro, G. Vinti and their research group on nonlinear operators, the approximation problem was proved by using nonlinear operators in some function spaces (see the fundamental book due to Bardaro, Musielak and Vinti [5]). For the approximation by linear and nonlinear operators, please see also the papers [3]-[2] and the monographs [10] and [26].

In view of the approach due to Musielak [22] and the techniques introduced by Bardaro-Mantellini in [4], Karsli-Tiryaki and Altin [18] considered the following nonlinear Bernstein operators;

$$(NB_n f)(x) = \sum_{k=0}^{n} P_{n,k}\left(x, f\left(\frac{k}{n}\right)\right), \ \ 0 \leq x \leq 1, \quad n \in \mathbb{N}, \tag{1.2}$$

acting on bounded functions f on an interval $[0,1]$, where $P_{n,k}$ satisfy some suitable assumptions. For further results we refer the papers [17], [16] and [18].

To generalize and extend the superposition or approximation problem for the functionals and operators, very recently in [13] and [14] Karsli defined and investigated the Urysohn type nonlinear Bernstein operators as;

$$(NB_n F)x\left(t\right) = \int_0^1 \left[\sum_{k=0}^{n} P_{k,n}\left(x(s), f\left(t, s, \frac{k}{n}\right)\right)\right] ds, \ \ 0 \leq x(s) \leq 1,$$

where $P_{k,n}$ satisfy some suitable assumptions.

As a continuation of the above studies, in [14] the author also obtained Voronovskaya-type theorems for these operators.

For the linear forms of the Urysohn Bernstein and Urysohn Stancu operators we refer to the reader [11] and [21].

Moreover, in [15], Karsli considered a sequence $NBF = (NB_n F)$ of operators, which represents the Urysohn type nonlinear form of the two dimensional Bernstein operators defined by P.L. Butzer on the square $S=[0,1]\times[0,1]$ (see [8], [9]), having the form:

$$(NB_nF)(x(t),y(t))=\int_0^1\int_0^1\left[\sum_{k=0}^{n}\sum_{i=0}^{n}P_{k,i,n}\left(x(s),y(z),f\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right)\right]dsdz,$$

$$0\leq x(s),y(z)\leq 1,\ n\in\mathbb{N},$$

acting on bounded functions f on $[0,1]^5$, where $P_{k,i,n}$ satisfy some suitable assumptions.

The central issue of this paper is to give a positive answer to the superposition problem for functionals and operators by introducing the Urysohn nonlinear operators of the two dimensional Stancu operators $(P_n^{\alpha}f)(x,y)$ defined on the triangle

$$\triangle:=\{(s,z):s,z\geq 0,s+z\leq 1\}.$$

Afterwards, we investigate the convergence problem for these nonlinear operators.

This paper is organized as follows: in Section 2, we construct the operators and further we present a basic lemma together with some definitions, which will be used in the sequel. Section 3 deals with the main convergence results for these operators.

2. Preliminaries and auxiliary results

This section is devoted to collecting some definitions and results which will be needed further on.

Now, we consider the following two dimensional Urysohn integral operator over the triangle $\triangle:=\{(s,z):s,z\geq 0,s+z\leq 1\}$,

$$F(x(t),y(t))=\iint_{\triangle}f(t,s,z,x(s),y(z))dsdz,\qquad t,s,z\in[0,1]$$

with unknown kernel f. If this representation exists, then $f(t,s,z,x(.),y(.))$ is called the two dimensional Green's function, which is strongly related to the functions x and y (see [25] and [26]).

In view of the above relations, we assume that the two dimensional continuous interpolation conditions hold:

$$F(x_i(t),y_j(t))=\iint_{\triangle}f(t,s,z,x_i(s),y_j(z))dsdz,\qquad t\in[0,1],\tag{2.1}$$

where

$$\begin{aligned}x_i(s)&=\frac{i}{n}H(s-\xi);\xi\in[0;1]\\ y_j(z)&=\frac{j}{n}H(z-\varsigma);\varsigma\in[0;1]\end{aligned}$$

and $i,j=0,1,2,...n$. By a straightforward calculation we have

$$\begin{aligned}\frac{\partial^2 F\left(\frac{i}{n}H(s-\xi),\frac{j}{n}H(z-\varsigma)\right)}{\partial\xi\partial\varsigma} &= f(t,\xi,\varsigma,\frac{i}{n},\frac{j}{n})-f(t,\xi,\varsigma,\frac{i}{n},0)\\ &\quad +f(t,\xi,\varsigma,0,0)-f(t,\xi,\varsigma,0,\frac{j}{n}).\end{aligned}$$

Say

$$F_1\left(t,\xi,\varsigma,\frac{i}{n},\frac{j}{n}\right):=\frac{\partial^2 F\left(\frac{i}{n}H(s-\xi),\frac{j}{n}H(z-\varsigma)\right)}{\partial\xi\partial\varsigma}.$$

According to the above definition, it is possible to construct an approximation operator in order to generalize and extend of the theory of interpolation of functions to operators.

For a bounded function defined on the triangle $\triangle:=\{(x,y):x,y\geq 0,x+y\leq 1\}$, two dimensional Stancu polynomials is given by:

$$\left(P_n^\alpha f\right)(x,y)=\sum_{k=0}^{n}\sum_{j=0}^{n-k}p_{n,k,j}^\alpha(x,y)f\left(\frac{k}{n},\frac{j}{n}\right),$$

where α is a non-negative parameter, which may depend only on the natural number n and

$$\begin{aligned}p_{n,k,j}^\alpha(x,y) &= \binom{n}{k}\binom{n-k}{j}\\ &\quad * \frac{\prod_{l_1=0}^{k-1}(x+l_1\alpha)\prod_{l_2=0}^{j-1}(y+l_2\alpha)\prod_{l_3=0}^{n-k-j-1}(1-x-y+l_3\alpha)}{\prod_{l_4=0}^{n-1}(1+l_4\alpha)}\end{aligned}$$

is the two dimensional Markov-Polya distribution ([24]).

Finally, let us now consider a sequence $NP^\alpha F=(NP_n^\alpha F)$ of operators, which represents Urysohn type nonlinear counterpart of the two dimensional Stancu operators defined on the triangle $\triangle:=\{(s,z):s,z\geq 0,s+z\leq 1\}$, having the form:

$$\begin{aligned}&\left(NP_n^\alpha F\right)(x(t),y(t))\\ &=\iint\limits_{\triangle}\left[\sum_{k=0}^{n}\sum_{i=0}^{n-k}P_{k,i,n}^\alpha\left(x(s),y(z),f\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right)\right]dsdz, \qquad (2.2)\\ &\qquad 0\leq x(s),y(z) \text{ and } x(s)+y(z)\leq 1,\ n\in\mathbb{N},\end{aligned}$$

acting on bounded functions f on $[0,1]^5$, where $P_{k,i,n}^\alpha$ satisfy some suitable assumptions. In particular, we will put $Dom\ NP^\alpha F=\bigcap\limits_{n\in\mathbb{N}}Dom\ NP_n^\alpha F$, where $Dom\ NP_n^\alpha F$ is the set of all functions $f:[0,1]^5\to\mathbb{R}$ for which the operator is well defined.

Let X be the set of all bounded Lebesgue measurable functions

$$f:[0,1]^5\to\mathbb{R}_0^+=[0,\infty).$$

Let Ψ be the class of all functions $\psi : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ such that the function ψ is continuous and concave with $\psi(0) = 0$, $\psi(u) > 0$ for $u > 0$.

We now introduce a sequence of functions. Let $\left\{P_{k,i,n}^{\alpha}\right\}_{n\in\mathbb{N}}$ be a sequence of functions $P_{k,i,n}^{\alpha} : [0,1] \,\mathrm{x}\, [0,1] \,\mathrm{x}\, \mathbb{R} \to \mathbb{R}$ defined by

$$P_{k,i,n}^{\alpha}(t,l,u) = p_{k,n}^{\alpha}(t) p_{i,n}^{\alpha}(l) H_n(u) \tag{2.3}$$

for every $t, l \in [0,1], u \in \mathbb{R}$, where $H_n : \mathbb{R} \to \mathbb{R}$ is such that $H_n(0) = 0$ and $p_{k,n}^{\alpha}(\bullet)$ is the Markov-Polya basis.

Throughout the paper we assume that $\mu : \mathbb{N} \to \mathbb{R}^+$ is an increasing and continuous function such that $\lim\limits_{n\to\infty} \mu(n) = \infty$.

Assume that the following conditions hold:

a) $H_n : \mathbb{R} \to \mathbb{R}$ *is such that*

$$|H_n(u) - H_n(v)| \leq \psi(|u - v|), \tag{2.4}$$

holds for every $u, v \in \mathbb{R}$, *for every* $n \in \mathbb{N}$. *That is,* H_n *satisfies a* $(L - \Psi)$ *Lipschitz condition.*

b) *Denoting by* $r_n(u) := H_n(u) - u$, $u \in \mathbb{R}$ *and* $n \in \mathbb{N}$, *such that for* n *sufficiently large*

$$\sup_u |r_n(u)| = \sup_u |H_n(u) - u| \leq \frac{1}{\mu(n)}, \tag{2.5}$$

holds.

The symbol $[a]$ will denote the greatest integer not greater than a.

Following our announced aim, in this part we recall results regarding the univariate and linear case of the celebrated Stancu operators.

Lemma 2.1. [23] *For* $(P_n^{\alpha} t^s)(x)$, $s = 0, 1, 2$, *one has*

$$\begin{aligned} (P_n^{\alpha} 1)(x) &= 1 \\ (P_n^{\alpha} t)(x) &= x \\ (P_n^{\alpha} t^2)(x) &= x^2 + \frac{(1 + \alpha n)\, x(1 - x)}{n(1 + \alpha)}. \end{aligned}$$

By direct calculation, we find the following equalities:

$$(P_n^{\alpha}(t - x)^2)(x) = \frac{x(1 - x)(1 + \alpha n)}{n(1 + \alpha)} \quad , \quad (P_n^{\alpha}(t - x))(x) = 0\,.$$

Moreover, for the second order central moment one has

$$(P_n^{\alpha}(t - x)^2)(x) \leq \frac{1 + \alpha n}{4n(1 + \alpha)}.$$

Definition 2.2. Let $f \in C\left([a,b]^5\right)$ and $\delta > 0$ be given. Then the complete modulus of continuity is given by:

$$\omega(f;\delta) = \sup_{\sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2} \leq \delta} |f(t,s,z,u_1,v_1) - f(t,s,z,u_2,v_2)|\,. \tag{2.6}$$

Further on, the partial modulus of continuity with respect to forth and fifth variables are defined by

$$\omega_1(f;\delta) = \sup_{t,s,z,v_1}\left(\sup_{|u_1-u_2|\le\delta}|f(t,s,z,u_1,v_1)-f(t,s,z,u_2,v_1)|\right),$$

and

$$\omega_2(f;\delta) = \sup_{t,s,z,u_1}\left(\sup_{|v_1-v_2|\le\delta}|f(t,s,z,u_1,v_1)-f(t,s,z,u_1,v_2)|\right),$$

respectively. Note that $\omega(f;\delta)$ has the following properties;

(i) Let $\lambda \in \mathbb{R}^+$, then

$$\omega(f;\lambda\delta) \le (\lambda+1)\,\omega(f;\delta),$$

(ii) $\lim_{\delta\to 0^+}\omega(f;\delta) = 0$,

(iii) $|f(t,s,z,u_1,v_1) - f(t,s,z,u_2,v_2)|$

$$\le \omega(f;\delta)\left(1+\frac{\sqrt{(u_1-u_2)^2+(v_1-v_2)^2}}{\delta}\right).$$

The same properties also hold for partial moduli of continuity.

Now, we are ready to state some convergence results of the operators defined on the triangle.

3. Main theorems

Theorem 3.1. *Let F be the Urysohn integral operator with $0 \le x(s), y(z)$ and*

$$x(s)+y(z)\le 1.$$

Then $(NP_n^\alpha F)$ converges to F uniformly in $x, y \in C[0,1]$. That is

$$\lim_{n\to\infty}\|(NP_n^\alpha F)(x(t),y(t)) - F(x(t),y(t))\|_{C(\triangle)} = 0.$$

Proof. Owing to the definition of the operator given by (2.2), by considering (2.1), (2.3), (2.4) and (2.5), we have

$$|(NP_n^\alpha F)(x(t),y(t)) - F(x(t),y(t))|$$

$$= \left|\iint_{\triangle}\left[\sum_{k=0}^{n}\sum_{i=0}^{n-k}P_{k,i,n}^\alpha\left(x(s),y(z),f\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right)\right]dsdz - F(x(t),y(t))\right|$$

$$
\begin{aligned}
\leq\ & \iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \cdot \\
& \cdot \left| H_n \left(f \left(t, s, z, \frac{k}{n}, \frac{i}{n} \right) \right) - H_n \left(f \left(t, s, z, x(s), y(z) \right) \right) \right| dsdz \\
& + \iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \cdot \\
& \cdot \left| H_n \left(f \left(t, s, z, x(s), y(z) \right) \right) - f(t, s, z, x(s), y(z)) \right| dsdz \\
& := I_1 + I_2.
\end{aligned}
$$

Owing to the assumption b), one has

$$
\begin{aligned}
I_2 &= \iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \cdot \\
& \quad \cdot \left| H_n \left(f \left(t, s, z, x(s), y(z) \right) \right) - f(t, s, z, x(s), y(z)) \right| dsdz \\
&\leq \iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \frac{1}{\mu(n)} dsdz \\
&= \frac{1}{\mu(n)},
\end{aligned}
$$

which tends to zero as $n \to \infty$.
Using the definition of the function $F_1(t, s, z, x(s), y(z))$, by concavity of the function ψ, and using Jensen inequality, we obtain

$$
I_1 \leq \iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z))
$$

$$
\times \psi \left(\left| f \left(t, s, z, \frac{k}{n}, \frac{i}{n} \right) - f \left(t, s, z, x(s), y(z) \right) \right| \right) dsdz
$$

$$
\leq \psi \left(\iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \left| f \left(t, s, z, \frac{k}{n}, \frac{i}{n} \right) - f \left(t, s, z, x(s), y(z) \right) \right| dsdz \right)
$$

$$
\begin{aligned}
\leq \psi \Bigg\{ & \iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \times \left| F_1 \left(t, s, z, x(s), y(z) \right) - F_1 \left(t, s, z, \tfrac{k}{n}, \tfrac{i}{n} \right) \right| dsdz \\
& + \iint\limits_{\triangle} \left| f \left(t, s, z, x(s), 0 \right) - \sum_{k=0}^{n} p_{k,n}^{\alpha}\left(x(s) \right) f \left(t, s, z, \frac{k}{n}, 0 \right) \right| dsdz \\
& + \iint\limits_{\triangle} \left| f \left(t, s, z, 0, y(z) \right) - \sum_{i=0}^{n} p_{i,n}^{\alpha}\left(y(z) \right) f \left(t, s, z, 0, \frac{i}{n} \right) \right| dsdz \Bigg\} \leq I_{1,1} + I_{1,2} + I_{1,3}.
\end{aligned}
$$

Let us divide the first term into four parts as;

$$\begin{aligned} I_{1,1} &= \psi\left(\begin{array}{c} \iint\limits_{\triangle} \sum\limits_{k=0}^{n} \sum\limits_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\ \cdot \left|F_1\left(t,s,z,x(s),y(z)\right) - F_1\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right| dsdz \end{array}\right) \\ &: \leq I_{1,1,1} + I_{1,1,2} + I_{1,1,3} + I_{1,1,4}, \end{aligned}$$

where

$$I_{1,1,1} = \psi\left(\begin{array}{c} \iint\limits_{\triangle} \sum\limits_{\left|\frac{k}{n}-x(s)\right|<\delta_1} \sum\limits_{\left|\frac{i}{n}-y(z)\right|<\delta_2} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\ \cdot \left|F_1\left(t,s,z,x(s),y(z)\right) - F_1\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right| dsdz \end{array}\right),$$

$$I_{1,1,2} = \psi\left(\begin{array}{c} \iint\limits_{\triangle} \sum\limits_{\left|\frac{k}{n}-x(s)\right|<\delta_1} \sum\limits_{\left|\frac{i}{n}-y(z)\right|\geq\delta_2} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\ \cdot \left|F_1\left(t,s,z,x(s),y(z)\right) - F_1\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right| dsdz \end{array}\right),$$

$$I_{1,1,3} = \psi\left(\begin{array}{c} \iint\limits_{\triangle} \sum\limits_{\left|\frac{k}{n}-x(s)\right|\geq\delta_1} \sum\limits_{\left|\frac{i}{n}-y(z)\right|<\delta_2} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\ \cdot \left|F_1\left(t,s,z,x(s),y(z)\right) - F_1\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right| dsdz \end{array}\right),$$

and

$$I_{1,1,4} = \psi\left(\begin{array}{c} \iint\limits_{\triangle} \sum\limits_{\left|\frac{k}{n}-x(s)\right|\geq\delta_1} \sum\limits_{\left|\frac{i}{n}-y(z)\right|\geq\delta_2} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\ \cdot \left|F_1\left(t,s,z,x(s),y(z)\right) - F_1\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right| dsdz \end{array}\right).$$

Since $x, y \in C[0,1]$, then there exist $\delta_1, \delta_2 > 0$ such that

$$\left|F_1\left(t,s,z,x(s),y(z)\right) - F_1\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right| < \epsilon$$

holds true when $\left|\frac{k}{n} - x(s)\right| < \delta_1$ and $\left|\frac{i}{n} - y(z)\right| < \delta_2$. So one can easily obtain

$$I_{1,1,1} < \psi\left(\epsilon\right).$$

As to the other terms

$$\left|F_1\left(t,s,z,x(s),y(z)\right) - F_1\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right| \leq 2M$$

holds true for some $M > 0$, when $\left|\frac{k}{n} - x(s)\right| \geq \delta_1$ or $\left|\frac{i}{n} - y(z)\right| \geq \delta_2$.

In view of Lemma 2.1, we obtain

$$I_{1,1,2} = \psi\left(\begin{array}{l} \iint\limits_{\triangle} \sum\limits_{\left|\frac{k}{n}-x(s)\right|<\delta_1} \sum\limits_{\left|\frac{i}{n}-y(z)\right|\geq\delta_2} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\ \cdot \left|F_1\left(t,s,z,x(s),y(z)\right) - F_1\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right| dsdz \end{array} \right)$$

$$\leq \psi\left(2M \iint\limits_{\triangle} \sum_{\left|\frac{k}{n}-x(s)\right|<\delta_1} \sum_{\left|\frac{i}{n}-y(z)\right|\geq\delta_2} \left(\frac{i-ny(z)}{\delta_2}\right)^2 p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z))dsdz\right)$$

$$\leq \psi\left(2M \iint\limits_{\triangle} \sum_{\left|\frac{k}{n}-x(s)\right|<\delta_1} \sum_{\left|\frac{i}{n}-y(z)\right|\geq\delta_2} \left(\frac{i-ny(z)}{\delta_2}\right)^2 p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z))dsdz\right)$$

$$\leq \psi\left(\frac{2M}{\delta_2^2}\frac{1+\alpha n}{4n(1+\alpha)}\right).$$

Similarly one has

$$I_{1,1,3} \leq \psi\left(\frac{2M}{\delta^2}\frac{1+\alpha n}{4n(1+\alpha)}\right),$$

and

$$I_{1,1,4} \leq \psi\left(\frac{2M}{\delta_1^2\delta_2^2}\left(\frac{1+\alpha n}{4n(1+\alpha)}\right)^2\right).$$

Collecting these estimates we have

$$\begin{aligned} & \left|\left(NP_n^{\alpha}F\right)\left(x\left(t\right),y(t)\right) - F\left(x\left(t\right),y(t)\right)\right| \\ \leq\ & \psi\left(\epsilon\right) + \psi\left(\frac{2M}{\delta_1^2}\frac{1+\alpha n}{4n(1+\alpha)}\right) + \psi\left(\frac{2M}{\delta_2^2}\frac{1+\alpha n}{4n(1+\alpha)}\right) \\ & +\psi\left(\frac{2M}{\delta_1^2\delta_2^2}\left(\frac{1+\alpha n}{4n(1+\alpha)}\right)^2\right) + \frac{1}{\mu\left(n\right)}. \end{aligned}$$

That is

$$\lim_{n\to\infty} \left\|\left(NP_n^{\alpha}F\right)\left(x\left(t\right),y(t)\right) - F\left(x\left(t\right),y(t)\right)\right\|_{C([0,1]^2)} = 0.$$

This completes the proof.

Theorem 3.2. *Let F be the Urysohn integral operator with $x, y \in C[0,1]$ and $0 \leq x(s)$, $y(z) \leq 1$. Then*

$$\left|\left(NP_n^{\alpha}F\right)\left(x\left(t\right),y(t)\right) - F\left(x\left(t\right),y(t)\right)\right| \leq 2\psi\left(\omega\left(f;\delta\right)\right) + \frac{1}{\mu\left(n\right)}$$

holds true, where $\delta = \sqrt{\frac{1+\alpha n}{2n(1+\alpha)}}$.

Proof. Clearly one has

$$
\begin{aligned}
& \left|\left(NP_n^{\alpha}F\right)(x(t),y(t)) - F(x(t),y(t))\right| \\
\leq\ & \iint\limits_{\triangle} \sum_{k=0}^{n}\sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\
& \cdot\left|H_n\left(f\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right) - H_n\left(f\left(t,s,z,x(s),y(z)\right)\right)\right| dsdz \\
& +\frac{1}{\mu(n)} \\
& := I_{n,1}(x) + \frac{1}{\mu(n)}, \qquad (3.1)
\end{aligned}
$$

say. Since $x, y \in C[0,1]$ we can rewrite (3.1) as follows

$$
\begin{aligned}
I_{n,1}(x) &\leq \iint\limits_{\triangle} \sum_{k=0}^{n}\sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\
&\quad \cdot \psi\left(\left|f\left(t,s,z,\frac{k}{n},\frac{i}{n}\right) - f\left(t,s,z,x(s),y(z)\right)\right|\right) dsdz \\
&\leq \iint\limits_{\triangle} \sum_{k=0}^{n}\sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z))\psi\left(\omega\left(f;\delta\right)\right) dsdz \\
&\leq \psi\left(\iint\limits_{\triangle} \sum_{k=0}^{n}\sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z))\omega\left(f;\delta\right) dsdz\right) \\
&\leq \psi\left(\begin{array}{c} \iint\limits_{\triangle} \sum\limits_{k=0}^{n}\sum\limits_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\ \cdot\left(\frac{\sqrt{\left(\frac{k}{n}-x(s)\right)^2+\left(\frac{i}{n}-y(z)\right)^2}}{\delta}+1\right)\omega\left(f;\delta\right) dsdz \end{array}\right) \\
&= \psi\left(\begin{array}{c} \omega\left(f;\delta\right) \iint\limits_{\triangle} \sum\limits_{k=0}^{n}\sum\limits_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\ \cdot\frac{\sqrt{\left(\frac{k}{n}-x(s)\right)^2+\left(\frac{i}{n}-y(z)\right)^2}}{\delta} dsdz \end{array}\right) \\
&\quad + \psi\left(\omega\left(f;\delta\right) \iint\limits_{\triangle} \sum_{k=0}^{n}\sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) dsdz\right) \\
&\leq \psi\left(\frac{\omega\left(f;\delta\right)}{\delta} \iint\limits_{\triangle} \left(\begin{array}{c} \sum\limits_{k=0}^{n}\sum\limits_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z)) \cdot \\ \cdot\left[\left(\frac{k}{n}-x(s)\right)^2+\left(\frac{i}{n}-y(z)\right)^2\right] \end{array}\right)^{1/2} dsdz\right) + \psi\left(\omega\left(f;\delta\right)\right)
\end{aligned}
$$

$$\leq \psi\left(\frac{\omega\left(f;\delta\right)}{\delta}\left[\frac{1+\alpha n}{2n(1+\alpha)}\right]^{1/2}\right)+\psi\left(\omega\left(f;\delta\right)\right).$$

Taking into account that $\omega\left(f;\delta\right)$ is the modulus of continuity defined as (2.6). If we choose

$$\delta=\sqrt{\frac{1+\alpha n}{2n(1+\alpha)}},$$

then one can obtain the desired estimate, namely,

$$\left|\left(NP_n^{\alpha}F\right)\left(x\left(t\right),y(t)\right)-F\left(x\left(t\right),y(t)\right)\right|\leq 2\psi\left(\omega\left(f;\delta\right)\right)+\frac{1}{\mu\left(n\right)}.$$

Thus the proof is now complete.

Theorem 3.3. *Let F be the Urysohn integral operator with $x,y\in C[0,1]$, and $0\leq x(s)$, $y(z)\leq 1$. Then*

$$\left|\left(NP_n^{\alpha}F\right)\left(x\left(t\right),y(t)\right)-F\left(x\left(t\right),y(t)\right)\right|$$

$$\leq 2\left[\psi\left(\omega_1\left(f;\left[\frac{1+\alpha n}{4n(1+\alpha)}\right]^{1/2}\right)\right)+\psi\left(\omega_2\left(f;\left[\frac{1+\alpha n}{4n(1+\alpha)}\right]^{1/2}\right)\right)\right]+\frac{1}{\mu\left(n\right)}$$

holds true.

Proof. In view of the definition of the considered operator, one has

$$\begin{aligned}
&\left|\left(NP_n^{\alpha}F\right)\left(x\left(t\right),y(t)\right)-F\left(x\left(t\right),y(t)\right)\right|\\
\leq\ &\iint\limits_{\triangle}\sum_{k=0}^{n}\sum_{i=0}^{n-k}p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z))\cdot\\
&\cdot\left|H_n\left(f\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right)-H_n\left(f\left(t,s,z,x(s),y(z)\right)\right)\right|dsdz\\
&+\frac{1}{\mu\left(n\right)}\\
=\ &\iint\limits_{\triangle}\sum_{k=0}^{n}\sum_{i=0}^{n-k}p_{k,n}^{\alpha}(x(s))p_{i,n}^{\alpha}(y(z))\cdot\\
&\cdot\left|\begin{array}{c}H_n\left(f\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right)-H_n\left(f\left(t,s,z,x(s),\frac{i}{n}\right)\right)\\+H_n\left(f\left(t,s,z,x(s),\frac{i}{n}\right)\right)-H_n\left(f\left(t,s,z,x(s),y(z)\right)\right)\end{array}\right|dsdz\\
&+\frac{1}{\mu\left(n\right)}
\end{aligned}$$

$$
\begin{aligned}
&\leq && \iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \cdot \\
& && \cdot \left| H_n\left(f\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right) - H_n\left(f\left(t,s,z,x(s),\frac{i}{n}\right)\right)\right| dsdz \\
& && + \iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \\
& && \left| H_n\left(f\left(t,s,z,x(s),\frac{i}{n}\right)\right) - H_n\left(f\left(t,s,z,x(s),y(z)\right)\right)\right| dsdz \\
& && + \frac{1}{\mu(n)} \\
&: && = I_{n,1}(x) + I_{n,2}(x) + \frac{1}{\mu(n)},
\end{aligned}
$$

say. Since $x, y \in C[0,1]$ we can rewrite (3.1) as follows: By concavity of the function ψ, and using Jensen inequality, we obtain

$$
\begin{aligned}
I_{n,1}(x) &= \iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \cdot \\
& \quad \cdot \left| H_n\left(f\left(t,s,z,\frac{k}{n},\frac{i}{n}\right)\right) - H_n\left(f\left(t,s,z,x(s),\frac{i}{n}\right)\right)\right| dsdz \\
&\leq \iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \psi\left(\omega_1\left(f;\left|\frac{k}{n} - x(s)\right|\right)\right) dsdz \\
&\leq \psi\left(\iint\limits_{\triangle} \sum_{k=0}^{n} \sum_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \omega_1\left(f;\left|\frac{k}{n} - x(s)\right|\right) dsdz\right)
\end{aligned}
$$

Since ψ is non decreasing, then one has

$$
\begin{aligned}
I_{n,1}(x) &\leq \psi\left(\begin{array}{c} \iint\limits_{\triangle} \sum\limits_{k=0}^{n} \sum\limits_{i=0}^{n-k} p_{k,n}^{\alpha}(x(s)) p_{i,n}^{\alpha}(y(z)) \cdot \\ \cdot \left(\frac{\sqrt{\left(\frac{k}{n} - x(s)\right)^2}}{\delta_1} + 1\right) \omega_1(f;\delta)\, dsdz \end{array}\right) \\
&\leq \psi\left(\frac{\omega_1(f;\delta)}{\delta}\left[\frac{1+\alpha n}{4n(1+\alpha)}\right]^{1/2}\right) + \psi\left(\omega_1(f;\delta)\right).
\end{aligned}
$$

Similarly

$$
I_{n,1}(x) \leq \psi\left(\frac{\omega_2(f;\delta)}{\delta}\left[\frac{1+\alpha n}{4n(1+\alpha)}\right]^{1/2}\right) + \psi\left(\omega_2(f;\delta)\right).
$$

If we choose $\delta = \left[\frac{1+\alpha n}{4n(1+\alpha)}\right]^{1/2}$, so we get the desired estimate.

Acknowledgments. The author is thankful to the referee for his/her valuable remarks and suggestions leading to a better presentation of this paper.

References

[1] Altomare, F., Campiti, M., *Korovkin-Type Approximation Theory and its Applications*, De Gruyter Studies in Mathematics, 17, Walter de Gruyter and Co., Berlin, 1994.

[2] Bardaro, C., Karsli, H., Vinti, G., *Nonlinear integral operators with homogeneous kernels: pointwise approximation theorems*, Appl. Anal., **90**(2011), no 3-4, 463–474.

[3] Bardaro, C., Mantellini, I., *On the reconstruction of functions by means of nonlinear discrete operators*, J. Concr. Appl. Math., **1**(2003), no. 4, 273–285.

[4] Bardaro, C., Mantellini, I., *Approximation properties in abstract modular spaces for a class of general sampling-type operators*, Appl. Anal., **85**(2006), no. 4, 383–413.

[5] Bardaro, C., Musielak, J., Vinti, G., *Nonlinear Integral Operators and Applications*, De Gruyter Series in Nonlinear Analysis and Applications, Vol. 9, xii + 201 pp., 2003.

[6] Bardaro, C., Vinti, G., *Urysohn integral operators with homogeneous kernel: approximation properties in modular spaces*, Comment. Math. (Prace Mat.), **42**(2002), no. 2, 145–182.

[7] Bernstein S.N., *Demonstration du Théoreme de Weierstrass fondée sur le calcul des probabilités*, Comm. Soc. Math. Kharkow, **13**(1912/13), 1-2.

[8] Butzer, P.L., *On Bernstein Polynomials*, Ph.D. Thesis, University of Toronto, 1951.

[9] Butzer, P.L., *On two dimensional Bernstein polynomials*, Canad. J. Math., **5**(1953), 107–113.

[10] Butzer, P.L., Nessel, R.J., *Fourier Analysis and Approximation*, Vol. 1, Academic Press, New York, London, 1971.

[11] Demkiv, I.I., *On Approximation of the Urysohn operator by Bernstein type operator polynomials*, Visn. L'viv. Univ., Ser. Prykl. Mat. Inform., (2000), no. 2, 26-30.

[12] Karsli, H., *Approximation by Urysohn type Meyer-König and Zeller operators to Urysohn integral operators*, Results Math., **72**(2017), no. 3, 1571–1583.

[13] Karsli, H., *Approximation Results for Urysohn Type Nonlinear Bernstein Operators*, Advances in Summability and Approximation Theory, Book Chapter, Springer Nature Singapore Pte Ltd., 223-241, 2018.

[14] Karsli, H., *Voronovskaya-type theorems for Urysohn type nonlinear Bernstein operators*, Mathematical Methods in the Applied Sciences, (2018), accepted.

[15] Karsli, H., *Approximation results for Urysohn type two dimensional nonlinear Bernstein operators*, Const. Math. Anal., **1**(2018), no. 1, 45-57.

[16] Karsli, H., Altin, H.E., *A Voronovskaya-type theorem for a certain nonlinear Bernstein operators*, Stud. Univ. Babeş-Bolyai Math., **60**(2015), no. 2, 249–258.

[17] Karsli, H., Tiryaki, I.U., Altin, H.E., *On convergence of certain nonlinear Bernstein operators*, Filomat, **30**(2016), no. 1, 141–155.

[18] Karsli, H., Tiryaki I.U., Altin H.E., *Some approximation properties of a certain nonlinear Bernstein operators*, Filomat, **28**(6)(2014), 1295-1305.

[19] Lorentz, G.G., *Bernstein Polynomials*, University of Toronto Press, Toronto, 1953.

[20] Lupaş, L., Lupaş, A., *Polynomials of binomial type and approximation operators*, Studia Univ. Babeş-Bolyai, Mathematica, **32**(1987), 61-69.

[21] Makarov, V.L., Demkiv, I.I., *Approximation of the Urysohn operator by operator polynomials of Stancu type*, Ukrainian Math Journal, **64**(2012), no. 3, 356-386.

[22] Musielak, J., *On some approximation problems in modular spaces*, In: Constructive Function Theory 1981, (Proc. Int. Conf., Varna, June 1-5, 1981), 455-461, Publ. House Bulgarian Acad. Sci., Sofia 1983.

[23] Stancu, D.D., *Approximation of functions by a new class of linear polynomial operators*, Rev. Roumaine Math. Pures Appl., **13**(1968), 1173–1194.

[24] Stancu, D.D., *A new class of uniform approximating polynomial operators in two and several variables*, Proceedings of the Conference on the Constructive Theory of Functions (Approximation Theory) (Budapest, 1969), 443-455, Akademiai Kiado, Budapest, 1972.

[25] Urysohn, P., *On a type of nonlinear integral equation*, Mat. Sb., **31**(1924), 236–255.

[26] Zabreiko, P.P., Koshelev, A.I., Krasnosel'skii, M.A., Mikhlin, S.G., Rakovscik, L.S., Stetsenko, V.Ja., *Integral Equations: A Reference Text*, Noordhoff Int. Publ., Leyden, 1975.

Harun Karsli
Bolu Abant Izzet Baysal University Faculty of Science and Arts
Department of Mathematics
14030 Golkoy Bolu, Turkey
e-mail: `karsli_h@ibu.edu.tr`

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 197–205
DOI: 10.24193/subbmath.2019.2.06

Modified Kantorovich-Stancu operators (II)

Ioan Gavrea and Adonia-Augustina Opriş

Abstract. In this paper, we introduce a new kind of Bernstein-Kantorovich-Stancu operators. These operators generalize the operators introduced in the paper [2] by V. Gupta, G. Tachev and A.M. Acu.

Mathematics Subject Classification (2010): 41A25, 41A36.

Keywords: Approximation by linear operators, Kantorovich-Stancu operators.

1. Introduction

For $f \in C([0,1])$, the Bernstein operator of degree n is defined by

$$B_n(f;x) = \sum_{k=0}^{n} p_{n,k}(x) f\left(\frac{k}{n}\right),\ x \in [0,1]$$

where

$$p_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k},\ k = 0, 1, \ldots, n$$

and $p_{n,k}(x) = 0$ if $k < 0$ or $k > n$.
In [3], H. Khosravian-Arab, M. Delghan and M.R. Eslahchi, starting from well-known equalities

$$p_{n,k}(x) = (1-x)p_{n-1,k}(x) + x p_{n-1,k-1}(x)$$

and

$$p_{n,k}(x) = (1-x)^2 p_{n-2,k}(x) + 2x(1-x)p_{n-2,k-1}(x) + x^2 p_{n-2,k-2}(x),\ 0 < k < n$$

have introduced modified Bernstein operators:

(i) $B_n^{M,1}$ defined by

$$B_n^{M,1}(f;x) = \sum_{k=0}^{n} p_{n,k}^{M,1}(x) f\left(\frac{k}{n}\right),\ x \in [0,1] \tag{1.1}$$

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

where

$$p_{n,k}^{M,1} = a(x,n)p_{n-1,k}(x) + a(1-x,n)p_{n-1,k-1}(x)$$

and

$$a(x,n) = a_1(n)x + a_0(n),\ n = 0,1,\ldots$$

(ii) $B_n^{M,2}$ defined by

$$B_n^{M,2}(f;x) = \sum_{k=0}^{n} p_{n,k}^{M,2}(x) f\left(\frac{k}{n}\right),\ x \in [0,1] \tag{1.2}$$

where

$$p_{n,k}^{M,2}(x) = b(x,n)p_{n-2,k}(x) + d(x,n)p_{n-2,k-1}(x) + b(1-x,n)p_{n-2,k-2}(x)$$

and

$$b(x,n) = b_2(n)x^2 + b_1(n)x + b_0(n),$$
$$d(x,n) = d_0(n)x(1-x),\ n = 0,1,\ldots$$

$a_0(n)$, $a_1(n)$, $b_0(n)$, $b_1(n)$, $b_2(n)$ and $d_0(n)$ are the unknown sequences which are determined in appropriate way for each forms.

V. Gupta, G. Tachev and A.M. Acu ([2]) have considered the operators:

$$K_n^{M,1}(f;x) = (n+1)\sum_{k=0}^{n} p_{n,k}^{M,1}(x)\int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}} f(s)ds \tag{1.3}$$

and

$$K_n^{M,2}(f;x) = (n+1)\sum_{k=0}^{n} p_{n,k}^{M,1}(x)\int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}} f(s)ds. \tag{1.4}$$

Here, they have discussed a uniform convergence estimate for these modified operators. In 1968, D.D. Stancu ([5]) has introduced the linear positive operators

$$P_n^{(\alpha,\beta)} : C([0,1]) \to C([0,1])$$

defined by

$$P_n^{(\alpha,\beta)}(f;x) = \sum_{k=0}^{n} p_{n,k}(x) f\left(\frac{k+\alpha}{n+\beta}\right)$$

where α, β are two fixed real numbers such that $0 \le \alpha \le \beta$.

In 2004, D. Bărbosu ([1]) has introduced Kantorovich-Stancu operators

$$K_n^{(\alpha,\beta)} : L_1([0,1]) \to C([0,1])$$

defined by

$$K_n^{(\alpha,\beta)}(f;x) = (n+\beta+1)\sum_{k=0}^{n} p_{n,k}(x)\int_{\frac{k+\alpha}{n+\beta+1}}^{\frac{k+\alpha+1}{n+\beta+1}} f(s)ds.$$

Regarding the previously modified operators, we note the following:

(a) The operators $B_n^{M,1}$ and $B_n^{M,2}$ are linear combinations of the operators $P_{n-1}^{(0,1)}$ and $P_{n-1}^{(1,1)}$, respectively of the operators $P_{n-2}^{(0,2)}$, $P_{n-2}^{(1,2)}$ and $P_{n-2}^{(2,2)}$, more precisely

$$B_n^{M,1}(f;x) = a(x,n)P_{n-1}^{(0,1)}(f;x) + a(1-x,n)P_{n-1}^{(1,1)}(f;x)$$

and

$$B_n^{M,2}(f;x) = b(x,n)P_{n-2}^{(0,2)}(f;x) + d(x,n)P_{n-2}^{(1,2)}(f;x) + b(1-x,n)P_{n-2}^{(2,2)}(f;x);$$

(b) The operators $K_n^{M,1}$ and $K_n^{M,2}$ are linear combinations of the operators $K_{n-1}^{(0,1)}$ and $K_{n-1}^{(1,1)}$, respectively of the operators $K_{n-2}^{(0,2)}$, $K_{n-2}^{(1,2)}$ and $K_{n-2}^{(2,2)}$, therefore

$$K_n^{M,1}(f;x) = a(x,n)K_{n-1}^{(0,1)}(f;x) + a(1-x,n)K_{n-1}^{(1,1)}(f;x)$$

and

$$K_n^{M,2}(f;x) = b(x,n)K_{n-2}^{(0,2)}(f;x) + d(x,n)K_{n-2}^{(1,2)}(f;x) + b(1-x,n)K_{n-2}^{(2,2)}(f;x).$$

From the above reasons, in this paper we introduce, for any $\alpha, \beta \in \mathbb{R}$, $0 \le \alpha \le \beta$ the operators

$$\overline{K}_n^{(\alpha,\beta)}(f;x) = (n+\beta+1)\sum_{k=0}^{n} p_{m,k}^{M,1}(x)\int_{\frac{k+\alpha}{n+\beta+1}}^{\frac{k+\alpha+1}{n+\beta+1}} f(s)ds \tag{1.5}$$

and

$$\overline{\overline{K}}_n^{(\alpha,\beta)}(f;x) = (n+\beta+1)\sum_{k=0}^{n} p_{m,k}^{M,2}(x)\int_{\frac{k+\alpha}{n+\beta+1}}^{\frac{k+\alpha+1}{n+\beta+1}} f(s)ds. \tag{1.6}$$

We mention that the Kantorovich-Stancu type operators $\overline{K}_n^{(\alpha,\beta)}$ was studied in a recent paper submitted for publication ([4]).

2. Auxiliary results

Lemma 2.1. *The central moments of $K_n^{(\alpha,\beta)}$ are given by:*

$$K_n^{(\alpha,\beta)}((t-x)^s;x) = \frac{1}{s+1}\sum_{k=0}^{n} p_{n,k}(x)\Bigg\{\sum_{i=0}^{s+1}\binom{s+1}{i}\left(\frac{k+1}{n+1}-x\right)^i \frac{1}{(n+\beta+1)^{s-i}}$$
$$\times\left[\sum_{j=1}^{s+1-i}(-1)^{j+1}\binom{s+1-i}{j}\left(\alpha-\beta\frac{k+1}{n+1}\right)^{s+1-i-j}\right]\Bigg\}. \tag{2.1}$$

Proof.

$$K_n^{(\alpha,\beta)}((t-x)^s;x) = (n+\beta+1)\sum_{k=0}^{n} p_{n,k}(x)\int_{\frac{k+\alpha}{n+\beta+1}}^{\frac{k+\alpha+1}{n+\beta+1}} (t-x)^s dt$$

$$= (n+\beta+1)\frac{1}{s+1}\sum_{k=0}^{n} p_{n,k}(x)\left[\left(\frac{k+\alpha+1}{n+\beta+1}-x\right)^{s+1} - \left(\frac{k+\alpha}{n+\beta+1}-x\right)^{s+1}\right] \tag{2.2}$$

Because

$$\frac{k+\alpha+1}{n+\beta+1}-x = \frac{k+1}{n+1}-x+\frac{k+\alpha+1}{n+\beta+1}-\frac{k+1}{n+1}$$
$$= \frac{k+1}{n+1}-x+\left(\alpha-\beta\frac{k+1}{n+1}\right)\frac{1}{n+\beta+1},$$

we have

$$\left(\frac{k+\alpha+1}{n+\beta+1}-x\right)^{s+1}$$
$$=\sum_{i=0}^{s+1}\binom{s+1}{i}\left(\frac{k+1}{n+1}-x\right)^{i}\left(\alpha-\beta\frac{k+1}{n+1}\right)^{s+1-i}\frac{1}{(n+\beta+1)^{s+1-i}}$$

and

$$\left(\frac{k+\alpha}{n+\beta+1}-x\right)^{s+1}$$
$$=\sum_{i=0}^{s+1}\binom{s+1}{i}\left(\frac{k+1}{n+1}-x\right)^{i}\left(\alpha-1-\beta\frac{k+1}{n+1}\right)^{s+1-i}\frac{1}{(n+\beta+1)^{s+1-i}}.$$

So, (2.2) becomes

$$K_n^{(\alpha,\beta)}((t-x)^s;x)=\frac{1}{s+1}\sum_{k=0}^{n}p_{n,k}(x)\left\{\sum_{i=0}^{s+1}\binom{s+1}{i}\left(\frac{k+1}{n+1}-x\right)^{i}\frac{1}{(n+\beta+1)^{s-i}}\right.$$
$$\left.\times\left[\left(\alpha-\beta\frac{k+1}{n+1}\right)^{s+1-i}-\left(\alpha-1-\beta\frac{k+1}{n+1}\right)^{s+1-i}\right]\right\}$$
$$=\frac{1}{s+1}\sum_{k=0}^{n}p_{n,k}(x)\left\{\sum_{i=0}^{s+1}\binom{s+1}{i}\left(\frac{k+1}{n+1}-x\right)^{i}\frac{1}{(n+\beta+1)^{s-i}}\right.$$
$$\left.\times\left[\sum_{j=1}^{s+1-i}(-1)^{j+1}\binom{s+1-i}{j}\left(\alpha-\beta\frac{k+1}{n+1}\right)^{s+1-i-j}\right]\right\}.$$

Remark 2.2. For $s=\overline{1,6}$ we have

$$K_n^{(\alpha,\beta)}(t-x;x)=-\frac{\beta+1}{n+\beta+1}x+\frac{2\alpha+1}{2(n+\beta+1)},$$

$$K_n^{(\alpha,\beta)}((t-x)^2;x)=\frac{n-(2\alpha+1)(\beta+1)}{(n+\beta+1)^2}x(1-x)$$
$$+\frac{(\beta-2\alpha)(\beta+1)}{(n+\beta+1)^2}x^2+\frac{3\alpha^2+3\alpha+1}{3(n+\beta+1)^2},$$

$$K_n^{(\alpha,\beta)}((t-x)^3;x)=-\frac{(3\beta+5)n-(\beta+1)^3}{(n+\beta+1)^3}x^2(1-x)$$
$$+\frac{(12\alpha+10)n-6(2\alpha+1)(\beta+1)^2+4(\beta+1)^3}{4(n+\beta+1)^3}x(1-x)$$
$$-\frac{4(3\alpha^2+3\alpha+1)(\beta+1)-6(2\alpha+1)(\beta+1)^2+4(\beta+1)^3}{4(n+\beta+1)^3}x$$
$$+\frac{4\alpha^3+6\alpha^2+4\alpha+1}{4(n+\beta+1)^3},$$

$$K_n^{(\alpha,\beta)}((t-x)^4;x)=\frac{3n^2-2(3+4(\beta+1)+3(\beta+1)^2)n}{(n+\beta+1)^4}(x(1-x))^2$$

$$-\frac{[4(2\alpha+1)+2(6\alpha+1)(\beta+1)-6(\beta+1)^2]n-2(2\alpha+1)(\beta+1)^3}{(n+\beta+1)^4}x^2(1-x)$$

$$+\frac{(6\alpha^2+10\alpha+5)n+2(2\alpha+1)(\beta+1)^3-2(3\alpha^2+3\alpha+1)(\beta+1)^2}{(n+\beta+1)^4}x(1-x)$$

$$-\frac{2(2\alpha+1)(\beta+1)^3-2(3\alpha^2+3\alpha+1)(\beta+1)^2+(4\alpha^3+6\alpha^2+4\alpha+1)(\beta+1)}{(n+\beta+1)^4}x$$

$$+\frac{5\alpha^4+10\alpha^3+10\alpha^2+5\alpha+1}{5(n+\beta+1)^4},$$

$$K_n^{(\alpha,\beta)}((t-x)^5;x)=\frac{[(30\beta+70)x^2(1-x)^3+(30\alpha-30\beta-35)x^2(1-x)^2]}{2(n+\beta+1)^5}n^2$$

$$+\Bigg[-\frac{30(\beta+1)^3+60(\beta+1)^2+90(\beta+1)+72}{3(n+\beta+1)^5}x^2(1-x)^3$$

$$+\frac{60(\beta+1)^3-45(2\alpha-1)(\beta+1)^2-30(4\alpha-1)(\beta+1)-9(10\alpha+1)}{3(n+\beta+1)^5}x^2(1-x)^2$$

$$-\frac{30(\beta+1)^3-15(6\alpha+1)(\beta+1)^2+15(6\alpha^2+2\alpha+1)(\beta+1)+2(30\alpha^2+30\alpha+13)}{3(n+\beta+1)^5}x^2(1-x)$$

$$+\frac{30\alpha^3+75\alpha^2+75\alpha+28}{3(n+\beta+1)^5}x(1-x)\Bigg]n+O\left(\frac{1}{n^5}\right),$$

$$K_n^{(\alpha,\beta)}((t-x)^6;x)=\frac{15x^3(1-x)^3}{(n+\beta+1)^6}n^3+\Bigg[\frac{45(\beta+1)^2x^4(1-x)^2}{(n+\beta+1)^6}$$

$$-\frac{(120x^3(1-x)^3+15(6\alpha-1)x^3(1-x)^2)(\beta+1)}{(n+\beta+1)^6}$$

$$+\frac{130x^2(1-x)^4+10(12\alpha-7)x^2(1-x)^3+5(9\alpha^2-3\alpha+2)x^2(1-x)^2}{(n+\beta+1)^6}\Bigg]n^2+O\left(\frac{1}{n^5}\right).$$

3. Main results

Here, we will extend the results from [4] for modified operators $\overline{\overline{K}}_n^{(\alpha,\beta)}$ defined by (1.6).

It is easy to see that

$$\overline{\overline{K}}_n^{(\alpha,\beta)}(f;x)=b(x;n)K_{n-2}^{(\alpha,\beta+2)}(f;x)+d(x;n)K_{n-2}^{(\alpha+1,\beta+2)}(f;x)$$
$$+b(1-x;n)K_{n-2}^{(\alpha+2,\beta+2)}(f;x). \tag{3.1}$$

Lemma 3.1. *For $i = 0, 1, 2$, the moments of $\overline{\overline{K}}_n(t^i; x)$ are given by:*

$$\overline{\overline{K}}_n^{(\alpha,\beta)}(1;x) = (2b_2(n) - d_0(n))x^2 - (2b_2(n) - d_0(n))x + b_2(n) + b_1(n) + 2b_0(n),$$

$$\begin{aligned}\overline{\overline{K}}_n^{(\alpha,\beta)}(t;x) &= \frac{(2(n-2)x + 2\alpha + 3)(2b_2(n) - d_0(n))}{2(n+\beta+1)}x(x-1)\\ &+ \frac{(n-4)(b_2(n) + b_1(n)) + 2(n-2)b_0(n)}{n+\beta+1}x\\ &+ \frac{(2\alpha+5)(b_2(n) + b_1(n)) + 2(2\alpha+3)b_0(n)}{2(n+\beta+1)},\end{aligned}$$

$$\begin{aligned}\overline{\overline{K}}_n^{(\alpha,\beta)}(t^2;x) &= [(2b_2(n) - d_0(n))x(x-1) + b_2(n) + b_1(n) + 2b_0(n)]\\ &\times \left[\frac{(n-2)(n-3)}{(n+\beta+1)^2}x^2 + \frac{2(\alpha+1)(n-2)}{(n+\beta+1)^2}x + \frac{3\alpha^2 + 3\alpha + 1}{3(n+\beta+1)^2}\right]\\ &+ \left[\frac{2(2b_2(n) - d_0(n))x^3}{(n+\beta+1)^2} - \frac{(2(2b_2(n) - d_0(n)) + 4(b_2(n) + b_1(n)))x^2}{(n+\beta+1)^2}\right.\\ &\left. + \frac{4(b_2(n) + b_1(n) + b_0(n))}{(n+\beta+1)^2}x\right](n-2)\\ &+ \frac{2(\alpha+1)(2b_2(n) - d_0(n)) + 2b_2(n)}{(n+\beta+1)^2}x^2\\ &- \left[\frac{2(\alpha+1)(2b_2(n) - d_0(n))}{(n+\beta+1)^2} + \frac{4(\alpha+2)b_2(n) + 2(2\alpha+3)b_1(n)}{(n+\beta+1)^2}\right]x\\ &+ \frac{2(2\alpha+3)(b_2(n) + b_1(n) + b_0(n))}{(n+\beta+1)^2}.\end{aligned}$$

We want to demonstrate the uniform convergence of the sequence $(\overline{\overline{K}}_n^{(\alpha,\beta)} f)_{n\geq 2}$. For this purpose, we will consider that

$$\overline{\overline{K}}_n^{(\alpha,\beta)}(1;x) = 1$$

$$\Leftrightarrow 2b_2(n) - d_0(n) = 0 \text{ and } b_2(n) + b_1(n) + 2b_0(n) = 1. \tag{3.2}$$

Using these, we obtain that

$$\overline{\overline{K}}_n^{(\alpha,\beta)}(t;x) = x + \frac{4b_0(n) - \beta - 5}{n+\beta+1}x + \frac{2\alpha + 5 - 4b_0(n)}{2(n+\beta+1)}$$

$$\begin{aligned}\overline{\overline{K}}_n^{(\alpha,\beta)}(t^2;x) &= \frac{n^2 - (9 - 8b_0(n))n + 16 - 2b_1(n) - 20b_0(n)}{(n+\beta+1)^2}x^2\\ &+ \frac{2(\alpha + 3 - 2b_0(n))n + 2b_1(n) + 8(\alpha+3)b_0(n) - 4(2\alpha+5)}{(n+\beta+1)^2}x\\ &+ \frac{3\alpha^2 + 15\alpha + 19 - 6(2\alpha+3)b_0(n)}{3(n+\beta+1)^2}.\end{aligned}$$

Assume that $\beta = 2\alpha$, for $b_0(n) = \dfrac{\beta+5}{4}$ the above expressions become

$$\widetilde{K}_n^{(\alpha,\beta)}(t;x) = x$$

$$\widetilde{K}_n^{(\alpha,\beta)}(t^2;x) = x^2 + \frac{n+2b_1(n)+(\beta+2)(\beta+5)}{(n+\beta+1)^2}x(1-x) - \frac{3\beta^2+18\beta+14}{12(n+\beta+1)^2}.$$

Taking $b_1(n) = -\dfrac{n+(\beta+2)(\beta+5)}{2}$ we have that

$$\widetilde{K}_n^{(\alpha,\beta)}(t^2;x) = x^2 - \frac{3\beta^2+18\beta+14}{12(n+\beta+1)^2}.$$

By (3.2) we obtain

$$b_2(n) = \frac{n+2+(\beta+1)(\beta+5)}{2}$$

and

$$d_0(n) = n+2+(\beta+1)(\beta+5).$$

In this situation, we can give other expressions for the first six central moments.

Lemma 3.2.

$$\widetilde{K}_n^{(\alpha,\beta)}(t-x;x) = 0,$$
$$\widetilde{K}_n^{(\alpha,\beta)}((t-x)^2;x) = -\frac{3\beta^2+18\beta+14}{12(n+\beta+1)^2},$$
$$\widetilde{K}_n^{(\alpha,\beta)}((t-x)^3;x) = -\frac{1-2x}{4(n+\beta+1)^3}[2(3\beta+7)nx(1-x)$$
$$+2(\beta^3+9\beta^2+21\beta+13)x(1-x)+\beta^3+9\beta^2+23\beta+15],$$
$$\widetilde{K}_n^{(\alpha,\beta)}((t-x)^4;x) = \frac{-3n^2}{(n+\beta+1)^4}x^2(1-x)^2 + O\left(\frac{1}{n^3}\right),$$
$$\widetilde{K}_n^{(\alpha,\beta)}((t-x)^5;x) = \frac{15(\beta+3)n^2}{(n+\beta+1)^5}x^2(1-x)^2(2x-1) + O\left(\frac{1}{n^4}\right),$$
$$\widetilde{K}_n^{(\alpha,\beta)}((t-x)^6;x) = \frac{-30n^3}{(n+\beta+1)^6}x^3(1-x)^3 + O\left(\frac{1}{n^4}\right).$$

Using this, we will prove the following result:

Theorem 3.3. *For $x \in [0,1]$, if $f \in C^{(6)}([0,1])$, we have*

$$\widetilde{K}_n^{(\alpha,\beta)}(f;x) - f(x) = O\left(\frac{1}{n^2}\right), \tag{3.3}$$

for sufficient large n.

Proof. Applying the Taylor's formula to the operators $\widetilde{K}_n^{(\alpha,\beta)}$ we have

$$\widetilde{K}_n^{(\alpha,\beta)}(f;x) = f(x) + \sum_{k=1}^{6}\frac{1}{k!}\widetilde{K}_n^{(\alpha,\beta)}((t-x)^k;x)f^{(k)}(x)$$
$$+\widetilde{K}_n^{(\alpha,\beta)}(\rho(t;x)(t-x)^6;x),$$

where ρ is a continuous function.

It is sufficient to prove that

$$|\widetilde{K}_n^{(\alpha,\beta)}(\rho(t;x)(t-x)^6;x)| = O\left(\frac{1}{n^2}\right). \tag{3.4}$$

We know that operators are not positive, so we rewrite them like this

$$\widetilde{K}_n^{(\alpha,\beta)}(f;x) = \widetilde{K}_{n,1}^{(\alpha,\beta)}(f;x) - \widetilde{K}_{n,2}^{(\alpha,\beta)}(f;x)$$

where

$$\widetilde{K}_{n,1}^{(\alpha,\beta)}(f;x) = (b_2(n)x^2 + b_0(n)) \cdot K_{n-2}^{(\alpha,\beta+2)}(f;x) + d_0(n)x \cdot K_{n-2}^{(\alpha+1,\beta+2)}(f;x) + b_2(n)x^2 \cdot K_{n-2}^{(\alpha+2,\beta+2)}(f;x)$$

and

$$\widetilde{K}_{n,2}^{(\alpha,\beta)}(f;x) = -b_1(n)x \cdot K_{n-2}^{(\alpha,\beta+2)}(f;x) + d_0(n)x^2 \cdot K_{n-2}^{(\alpha+1,\beta+2)}(f;x) +((2b_2(n) + b_1(n))x - (b_2(n) + b_1(n) + b_0(n))) \cdot K_{n-2}^{(\alpha+2,\beta+2)}(f;x).$$

We note that $\widetilde{K}_{n,1}^{(\alpha,\beta)}$ and $\widetilde{K}_{n,2}^{(\alpha,\beta)}$ are linear and positive operators.

$$|\widetilde{K}_n^{(\alpha,\beta)}(\rho(t;x)(t-x)^6;x)| \le |\widetilde{K}_{n,1}^{(\alpha,\beta)}(\rho(t;x)(t-x)^6;x)| + |\widetilde{K}_{n,2}^{(\alpha,\beta)}(\rho(t;x)(t-x)^6;x)|. \tag{3.5}$$

Computing $\widetilde{K}_{n,i}^{(\alpha,\beta)}((t-x)^6;x)$, $i=1,2$, we obtain the following expressions

$$\widetilde{K}_{n,i}^{(\alpha,\beta)}((t-x)^6;x) = \frac{15x^4(1-x)^3(1+x)n^4}{(n+\beta+1)^6} + A_i(\alpha,\beta,x)\frac{n^3}{(n+\beta+1)^6}, \ i=1,2$$

where

$$A_1(\alpha,\beta,x) = 15(\beta^2+6\beta+7)x^4(1-x)^3(1+x) -[120x^4(1-x)^3(1+x) + 15(6\alpha+5)x^4(1-x)^2(1+x)](\beta+3) +45x^5(1-x)^2(1+x)(\beta+3)^2 + 130x^3(1-x)^4(1+x) +10(12\alpha+5)x^3(1-x)^3(1+x) + 5(9\alpha^2+15\alpha+8)x^3(1-x)^2(1+x) +45x^4(1-x)^2 + \frac{15(\beta+5)x^3(1-x)^3}{4}$$

and

$$A_2(\alpha,\beta,x) = 15(\beta^2+6\beta+7)x^4(1-x)^3(1+x) -[120x^4(1-x)^3(1+x) + 15(6\alpha+5)x^4(1-x)^2(1+x)](\beta+3) +45x^5(1-x)^2(1+x)(\beta+3)^2 + 130x^3(1-x)^4(1+x) +10(12\alpha+5)x^3(1-x)^3(1+x) + 5(9\alpha^2+15\alpha+8)x^3(1-x)^2(1+x) +45x^3(1-x)^2 - \frac{15(\beta+1)x^3(1-x)^3}{4}.$$

Because ρ is a continuous function, there exists an $M>0$ such that $|\rho(t;x)|<M$, $\forall\, x,t\in[0,1]$. Using the above results for $\widetilde{K}_{n,i}^{(\alpha,\beta)}((t-x)^6;x)$, we obtain

$$|\widetilde{K}_{n,i}^{(\alpha,\beta)}(\rho(t;x)(t-x)^6;x)| \le M\left|\frac{15x^4(1-x)^3(1+x)n^4}{(n+\beta+1)^6} + O\left(\frac{1}{n^3}\right)\right| = O\left(\frac{1}{n^2}\right), \ i=1,2.$$

So, (3.4) is proved.

Combining this with Lemma 3.2, we complete the proof of theorem.

References

[1] Bărbosu, D., *Kantorovich-Stancu type operators*, Journal of Inequalities in Pure and Applied Mathematics, **5**(2004), no. 3, Article 53.

[2] Gupta, V., Tachev, G., Acu, A.M., *Modified Kantorovich operators with better approximation properties*, Numer. Algo. (2018). https://doi.org/10.1007/s11075-018-0538-7

[3] Khosravian-Arab, H., Dehghan, M., Eslahchi, M.R., *A new approach to improve the order of approximation of the Bernstein operators: theory and applications*, Numer. Algo., **77**(2018), no. 1, 111-150.

[4] Opriş, A.A., *Approximation by modified Kantorovich-Stancu operators*, JIAP-D-18-00423, april, 2018 (summited for publication).

[5] Stancu, D.D., *Asupra unei generalizări a polinoamelor lui Bernstein*, Studia Univ. Babeş-Bolyai, Ser. Math.-Phys., **14**(1969), no. 2, 31-45.

Ioan Gavrea
Department of Mathematics
Technical University of Cluj-Napoca
str. Memorandumului, nr. 28, 400144
Cluj-Napoca, Romania
e-mail: ioan.gavrea@math.utcluj.ro

Adonia-Augustina Opriş
Department of Mathematics
Technical University of Cluj-Napoca
str. Memorandumului, nr. 28, 400144
Cluj-Napoca, Romania
e-mail: mate.salaj@yahoo.com

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 207–223
DOI: 10.24193/subbmath.2019.2.07

Approximation by max-product operators of Kantorovich type

Lucian Coroianu and Sorin G. Gal

Dedicated to Professor Heiner Gonska on the occasion of his 70th anniversary.

Abstract. We associate to various linear Kantorovich type approximation operators, nonlinear max-product operators for which we obtain quantitative approximation results in the uniform norm, shape preserving properties and localization results.

Mathematics Subject Classification (2010): 41A35, 41A25, 41A20.

Keywords: Max-product operators, max-product operators of Kantorovich kind, uniform approximation, shape preserving properties, localization results, max-product Kantorovich-Choquet operators.

1. Introduction

The general form of a linear and positive discrete operator attached to $f : I \to [0, +\infty)$ can be defined by

$$D_n(f)(x) = \sum_{k \in I_n} p_{n,k}(x) f(x_{n,k}), x \in I, n \in \mathbb{N},$$

where $p_{n,k}(x)$ are various kinds of function basis on I with $\sum_{k \in I_n} p_{n,k}(x) = 1$, I_n are finite or infinite families of indices and $\{x_{n,k}; k \in I_n\}$ represents a division of I.

Based on the Open Problem 5.5.4, pp. 324-326 in [7], to each $D_n(f)(x)$, can be attached the max-product type operator defined by

$$L_n^{(M)}(f)(x) = \frac{\bigvee_{k \in I_n} p_{n,k}(x) \cdot f(x_{n,k})}{\bigvee_{k \in I_n} p_{n,k}(x)}, x \in I, n \in \mathbb{N}. \tag{1.1}$$

Here $\bigvee_{k \in A} a_k = \sup_{k \in A} a_k$.

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

Thus, in a series of papers we have introduced and studied the so-called max-product operators attached to the Bernstein polynomials and to other linear Bernstein-type operators, like those of Favard-Szász-Mirakjan operators (truncated and nontruncated case), Baskakov operators (truncated and nontruncated case), Meyer-König and Zeller operators and Bleimann-Butzer-Hahn operators. All these results were collected in the very recent research monograph [2].

Remark 1.1. The max-product operators can also be naturally called as **possibilistic operators**, since they can be obtained by analogy with the Feller probabilistic scheme used to generate positive and linear operators, by replacing the probability (σ-additive), with a maxitive set function and the classical integral with the possibilistic integral (see, e.g. [2], Chapter 10, Section 10.2). If, for example, $p_{n,k}(x)$, $n \in \mathbb{N}$, $k = 0, \ldots, n$ is a polynomial basis, then the operators $L_n^{(M)}(f)(x)$ become piecewise rational functions.

Now, to each max-product operator $L_n^{(M)}$, we can formally attach its Kantorovich variant, defined by

$$LK_n^{(M)}(f)(x) = \frac{\bigvee_{k \in I_n} p_{n,k}(x) \cdot (1/(x_{n,k+1} - x_{n,k})) \cdot \int_{x_{n,k}}^{x_{n,k+1}} f(t)dt}{\bigvee_{k \in I_n} p_{n,k}(x)}, \tag{1.2}$$

with $\{x_{n,k}; k \in I_n\}$ a division of the finite or infinite interval I.

The goal of this paper is to study these Kantorovich-type versions for various max-product operators. Firstly, we prove that these operators are subadditive, positively homogeneous and monotone. For continuous functions we prove quantitative estimates, in most of the cases very good Jackson type estimates, shape preserving properties and localization results.

2. Uniform and pointwise approximation

Keeping the notations in the formulas (1.1) and (1.2), let us denote

$$C_+(I) = \{f : I \to \mathbb{R}_+; f \text{ is continuous on } I\},$$

where I is a bounded or unbounded interval and suppose that all $p_{n,k}(x)$ are continuous functions on I, satisfying $p_{n,k}(x) \geq 0$, for all $x \in I, n \in \mathbb{N}, k \in I_n$ and $\sum_{k \in I_n} p_{n,k}(x) = 1$, for all $x \in I, n \in \mathbb{N}$.

In many cases, for the Kantorovich max-product operator $K_n^{(M)}$ we could deduce quantitative estimates in approximation, by using the elaborated methods we used for the Bernstein-type max-product in the book [2]. However, here we will use a more simple method, which will be based on the already obtained estimates for the original type max-product operators denoted by $L_n^{(M)}$.

Firstly, we present the following result.

Lemma 2.1. *(i) For any $f \in C_+(I)$, $LK_n^{(M)}(f)$ is continuous on I.*

(ii) If $f \leq g$ then $LK_n^{(M)}(f) \leq LK_n^{(M)}(g)$.

(iii) $LK_n^{(M)}(f + g) \leq LK_n^{(M)}(f) + LK_n^{(M)}(g)$.

(iv) If $f \in C_+(I)$ *and* $\lambda \geq 0$ *then* $LK_n^{(M)}(\lambda f) = \lambda LK_n^{(M)}(f)$.

(v) If $LK_n^{(M)}(e_0) = e_0$, *where* $e_0(x) = 1$, *for all* $x \in I$, *then for any* $f \in C_+(I)$, *we have*

$$\left| LK_n^{(M)}(f)(x) - f(x) \right| \leq \left[1 + \frac{1}{\delta} LK_n^{(M)}(\varphi_x)(x) \right] \omega_1(f;\delta),$$

for any $x \in I$ *and* $\delta > 0$. *Here,* $\varphi_x(t) = |t - x|$, $t \in I$ *and*

$$\omega_1(f;\delta) = \sup\{|f(x) - f(y)|;\ x, y \in I,\ |x - y| \leq \delta\}.$$

(vi) $\left| LK_n^{(M)}(f) - LK_n^{(M)}(g) \right| \leq LK_n^{(M)}\left(|f - g|\right)$.

Proof. The proofs of (i)-(iv) are immediate from the definition of $K_n^{(M)}$. As for the proof of (v) and (vi), we exactly follow the proof of e.g., Theorem 1.1.2, pp. 16-17 in [2]. □

Lemma 2.2. *With the notations in (*1.1*) and (*1.2*), suppose that, in addition,*

$$|x_{n,k+1} - x_{n,k}| \leq \frac{C}{n+1}$$

for all $k \in I_n$, *with* $C > 0$ *an absolute constant. Then, for all* $x \in I$ *and* $n \in \mathbb{N}$, *we have*

$$LK_n^{(M)}(\varphi_x)(x) \leq L_n^{(M)}(\varphi_x)(x) + \frac{C}{n+1}.$$

Proof. If $f \in C_+(I)$, then by the integral mean value theorem, there exists $\xi_{n,k} \in (x_{n,k}, x_{n,k+1})$, such that

$$\int_{x_{n,k}}^{x_{n,k+1}} f(t)dt = (x_{n,k+1} - x_{n,k}) \cdot f(\xi_{n,k}),$$

which immediately leads to

$$LK_n^{(M)}(f)(x) = \frac{\bigvee_{k \in I_n} p_{n,k}(x) \cdot f(\xi_{n,k})}{\bigvee_{k \in I_n} p_{n,k}(x)}. \tag{2.1}$$

Applying this form for $f(t) = \varphi_x(t)$, we get

$$LK_n^{(M)}(\varphi_x)(x) = \frac{\bigvee_{k \in I_n} p_{n,k}(x) \cdot |\xi_{n,k} - x|}{\bigvee_{k \in I_n} p_{n,k}(x)}$$

$$\leq \frac{\bigvee_{k \in I_n} p_{n,k}(x) \cdot |\xi_{n,k} - x_{n,k}|}{\bigvee_{k \in I_n} p_{n,k}(x)} + L_n^{(M)}(\varphi_x)(x) \leq \frac{C}{n+1} + L_n^{(M)}(\varphi_x)(x),$$

which proves the lemma. □

Corollary 2.3. *With the notations in (*1.1*) and (*1.2*) and supposing that, in addition,*

$$|x_{n,k+1} - x_{n,k}| \leq \frac{C}{n+1}$$

for all $k \in I_n$, *for any* $f \in C_+(I)$, *we have*

$$\left| LK_n^{(M)}(f)(x) - f(x) \right| \leq 2\left[\omega_1(f; L_n^{(M)}(\varphi_x)(x)) + \omega_1(f; C/(n+1)) \right] \tag{2.2}$$

for any $x \in I$ *and* $n \in \mathbb{N}$.

Proof. By using Lemma 2.2, from the estimate in Lemma 2.1, (v), we immediately get

$$\left|LK_n^{(M)}(f)(x)-f(x)\right| \le 2\omega_1(f;L_n^{(M)}(\varphi_x)(x)+C/(n+1))$$
$$\le 2\left[\omega_1(f;L_n^{(M)}(\varphi_x)(x))+\omega_1(f;C/(n+1))\right],$$

which proves the corollary. □

This corollary shows that knowing quantitative estimates in approximation by a given max-product operator, we can deduce a quantitative estimate for its Kantorovich variant. Also, this method does not worsen the orders of approximation of the original operators. Let us exemplify below for several known max-product operators.

Firstly, let us choose $p_{n,k}(x)=\binom{n}{k}x^k(1-x)^{n-k}$, $I=[0,1]$, $I_n=\{0,\dots,n-1\}$ and $x_{n,k}=\frac{k}{n+1}$. In this case, $L_n^{(M)}$ in (1.1) become the Bernstein max-product operators. Let us denote by $BK_n^{(M)}$ their Kantorovich variant, given by the formula

$$BK_n^{(M)}(f)(x)=\frac{\bigvee_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}\cdot(n+1)\int_{k/(n+1)}^{(k+1)/(n+1)}f(t)dt}{\bigvee_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}}. \quad (2.3)$$

We can state the following result.

Theorem 2.4. *(i) If $f\in C_+([0,1])$, then we have*

$$|BK_n^{(M)}(f)(x)-f(x)|\le 24\omega_1(f;1/\sqrt{n+1})+2\omega_1(f;1/(n+1)), x\in[0,1], n\in\mathbb{N}.$$

(ii) If $f\in C_+([0,1])$ is concave on $[0,1]$, then we have

$$|BK_n^{(M)}(f)(x)-f(x)|\le 6\omega_1(f;1/n), x\in[0,1], n\in\mathbb{N}.$$

(iii) If $f\in C_+([0,1])$ is strictly positive on $[0,1]$, then we have

$$|BK_n^{(M)}(f)(x)-f(x)|\le 2\omega_1(f;1/n)\cdot\left(\frac{n\omega_1(f;1/n)}{m_f}+4\right)+2\omega_1(f;1/n),$$

for all $x\in[0,1], n\in\mathbb{N}$, where $m_f=\min\{f(x);x\in[0,1]\}$.

Proof. (i) is immediate from Corollary 2.3 (with $C=1$) and from Theorem 2.1.5, p. 30, in [2].

(ii) is immediate from Corollary 2.3 (with $C=1$) and from Corollary 2.1.10, p. 36 in [2].

(iii) is immediate from Corollary 2.3 (with $C=1$) and from Theorem 2.2.18, p. 63 in [2]. □

Now, let us choose $p_{n,k}(x)=\frac{(nx)^k}{k!}$, $I=[0,+\infty)$, $I_n=\{0,\dots,n,\dots,\}$ and $x_{n,k}=\frac{k}{n+1}$. In this case, $L_n^{(M)}$ in (1.1) become the non-truncated Favard-Szász-Mirakjan max-product operators. Let us denote by $FK_n^{(M)}$ their Kantorovich variant defined by

$$FK_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!} \cdot (n+1)\int_{k/(n+1)}^{(k+1)/(n+1)} f(t)dt}{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!}}. \tag{2.4}$$

We can state the following result.

Theorem 2.5. *(i) If* $f : [0, +\infty) \to [0, +\infty)$ *is bounded and continuous on* $[0, +\infty)$*, then we have*

$$|FK_n^{(M)}(f)(x) - f(x)| \leq 16\omega_1(f; \sqrt{x}/\sqrt{n}) + 2\omega_1(f; 1/n), x \in [0, +\infty), n \in \mathbb{N}.$$

(ii) If $f : [0, +\infty) \to [0, +\infty)$ *is continuous, bounded, non-decreasing, concave function on* $[0, +\infty)$*, then we have*

$$|FK_n^{(M)}(f)(x) - f(x)| \leq 4\omega_1(f; 1/n), x \in [0, +\infty), n \in \mathbb{N}.$$

Proof. (i) is immediate from Corollary 2.3 (with $C = 1$) and from Theorem 3.1.4, p. 162, in [2].

(ii) is immediate from Corollary 2.3 (with $C = 1$) and from Corollary 3.1.8, p. 168 in [2]. □

If we choose $p_{n,k}(x) = \frac{(nx)^k}{k!}$, $I = [0, 1]$, $I_n = \{0, \ldots, n\}$ and $x_{n,k} = \frac{k}{n+1}$. In this case, $L_n^{(M)}$ in (1.1) become the truncated Favard-Szász-Mirakjan max-product operators. Let us denote by $TK_n^{(M)}$ their Kantorovich variant given by the formula

$$TK_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} \frac{(nx)^k}{k!} \cdot (n+1)\int_{k/(n+1)}^{(k+1)/(n+1)} f(t)dt}{\bigvee_{k=0}^{n} \frac{(nx)^k}{k!}}. \tag{2.5}$$

We can state the following result.

Theorem 2.6. *(i) If* $f \in C_+([0, 1])$*, then we have*

$$|TK_n^{(M)}(f)(x) - f(x)| \leq 12\omega_1(f; 1/\sqrt{n}) + 2\omega_1(f; 1/n), x \in [0, 1], n \in \mathbb{N}.$$

(ii) If $f \in C_+([0, 1])$ *is non-decreasing, concave function on* $[0, 1]$*, then we have*

$$|TK_n^{(M)}(f)(x) - f(x)| \leq 4\omega_1(f; 1/n), x \in [0, 1], n \in \mathbb{N}.$$

Proof. (i) is immediate from Corollary 2.3 (with $C = 1$) and from Theorem 3.2.5, p. 178, in [2].

(ii) is immediate from Corollary 2.3 (with $C = 1$) and from Corollary 3.2.7, p. 182 in [2]. □

Now, let us choose $p_{n,k}(x) = \binom{n+k-1}{k} x^k/(1 + x)^{n+k}$, $I = [0, +\infty)$, $I_n = \{0, \ldots, n, \ldots, \}$ and $x_{n,k} = \frac{k}{n+1}$. In this case, $L_n^{(M)}$ in (1.1) become the non-truncated Baskakov max-product operators. Let us denote by $VK_n^{(M)}$ their Kantorovich variant defined by

$$VK_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} \binom{n+k-1}{k} \frac{x^k}{(1+x)^{n+k}} \cdot (n+1)\int_{k/(n+1)}^{(k+1)/(n+1)} f(t)dt}{\bigvee_{k=0}^{\infty} \binom{n+k-1}{k} \frac{x^k}{(1+x)^{n+k}}}. \tag{2.6}$$

We can state the following result.

Theorem 2.7. *(i) If $f : [0, +\infty) \to [0, +\infty)$ is bounded and continuous on $[0, +\infty)$, then for all $x \in [0, +\infty)$ and $n \geq 3$, we have*

$$|VK_n^{(M)}(f)(x) - f(x)| \leq 24\omega_1(f; \sqrt{x(x+1)}/\sqrt{n-1}) + 2\omega_1(f; 1/(n+1)).$$

(ii) If $f : [0, +\infty) \to [0, +\infty)$ is continuous, bounded, non-decreasing, concave function on $[0, +\infty)$, then for $x \in [0, +\infty)$ and $n \geq 3$ we have

$$|VK_n^{(M)}(f)(x) - f(x)| \leq 4\omega_1(f; 1/n).$$

Proof. (i) is immediate from Corollary 2.3 (with $C = 1$) and from Theorem 4.1.6, p. 196, in [2].

(ii) is immediate from Corollary 2.3 (with $C = 1$) and from Corollary 4.1.9, p. 206 in [2]. □

If we choose $p_{n,k}(x) = \binom{n+k-1}{k} x^k/(1+x)^{n+k}$, $I = [0, 1]$, $I_n = \{0, \ldots, n\}$ and $x_{n,k} = \frac{k}{n+1}$, then in this case, $L_n^{(M)}$ in (1.1) become the truncated Baskakov max-product operators. Let us denote by $UK_n^{(M)}$ their Kantorovich variant defined by

$$UK_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} \binom{n+k-1}{k} \frac{x^k}{(1+x)^{n+k}} \cdot (n+1) \int_{k/(n+1)}^{(k+1)/(n+1)} f(t)dt}{\bigvee_{k=0}^{\infty} \binom{n+k-1}{k} \frac{x^k}{(1+x)^{n+k}}}. \qquad (2.7)$$

We can state the following result.

Theorem 2.8. *(i) If $f \in C_+([0, 1])$, then we have,*

$$|UK_n^{(M)}(f)(x) - f(x)| \leq 48\omega_1(f; 1/\sqrt{n+1}) + 2\omega_1(f; 1/(n+1)), x \in [0, 1], n \geq 2.$$

(ii) If $f \in C_+([0, 1])$ is non-decreasing, concave function on $[0, 1]$, then we have

$$|UK_n^{(M)}(f)(x) - f(x)| \leq 6\omega_1(f; 1/n), x \in [0, 1], n \in \mathbb{N}.$$

Proof. (i) is immediate from Corollary 2.3 (with $C = 1$) and from Theorem 4.2.6, p. 217, in [2].

(ii) is immediate from Corollary 2.3 (with $C = 1$) and from Corollary 4.2.9, p. 223 in [2]. □

Now, let us choose $p_{n,k}(x) = \binom{n+k}{k} x^k$, $I = [0, 1]$, $I_n = \{0, \ldots, n, \ldots\}$ and $x_{n,k} = \frac{k}{n+1+k}$. In this case, $L_n^{(M)}$ in (1.1) become the Meyer-König and Zeller max-product operators. Also, it is easy to see that $|x_{n,k+1} - x_{n,k}| \leq \frac{1}{n+1}$, for all $k \in I_n$. Let us denote by $ZK_n^{(M)}$ their Kantorovich variant defined by

$$ZK_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} \binom{n+k}{k} x^k \cdot \frac{(n+k+1)(n+k+2)}{n+1} \int_{k/(n+1+k)}^{(k+1)/(n+k+2)} f(t)dt}{\bigvee_{k=0}^{\infty} \binom{n+k}{k} x^k}. \qquad (2.8)$$

The following result holds.

Theorem 2.9. *(i) If $f \in C_+([0, 1])$, then for $n \geq 4$, $x \in [0, 1]$, we have*

$$|ZK_n^{(M)}(f)(x) - f(x)| \leq 36\omega_1(f; \sqrt{x}(1-x)/\sqrt{n}) + 2\omega_1(f; 1/n).$$

(ii) If $f \in C_{+}([0,1])$ is non-decreasing concave function on $[0,1]$, then for $x \in [0,1]$ and $n \geq 2x$ we have

$$|ZK_n^{(M)}(f)(x) - f(x)| \leq 4\omega_1(f;1/n).$$

Proof. (i) is immediate from Corollary 2.3 (with $C = 1$) and from Theorem 6.1.4, p. 248, in [2].

(ii) is immediate from Corollary 2.3 (with $C = 1$) and from Corollary 6.1.7, p. 256 in [2]. □

In what follows, let us choose $p_{n,k}(x) = h_{n,k}(x)$-the fundamental Hermite-Fejér interpolation polynomials based on the Chebyshev knots of first kind

$$x_{n,k} = \cos\left(\frac{2(n-k)+1}{2(n+1)}\pi\right),$$

$I = [-1,1]$, and $I_n = \{0,\dots,n\}$. In this case, $L_n^{(M)}$ in (1.1) become the Hermite-Fejér max-product operators. Also, applying he mean value theorem to cos, it is easy to see that $|x_{n,k+1} - x_{n,k}| \leq \frac{4}{n+1}$, for all $k \in I_n$. Let us denote by $HK_n^{(M)}$ their Kantorovich variant defined by

$$HK_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} h_{n,k}(x) \cdot \frac{1}{x_{n,k} - x_{n,k+1}} \cdot \int_{x_{n,k}}^{x_{n,k+1}} f(t)dt}{\bigvee_{k=0}^{\infty} h_{n,k}(x)}, \tag{2.9}$$

where $x_{n,k} = \cos\left(\frac{2(n-k)+1}{2(n+1)}\pi\right)$.

The following result holds.

Theorem 2.10. *If $f \in C_{+}([-1,1])$, then for $n \in \mathbb{N}$, $x \in [-1,1]$, we have*

$$|HK_n^{(M)}(f)(x) - f(x)| \leq 30\omega_1(f;1/n).$$

Proof. It is immediate from Corollary 2.3 (with $C = 4$) and from Theorem 7.1.5, p. 286, in [2]. □

Now, let us consider choose $p_{n,k}(x) = e^{-|x-k/(n+1)|}$, $I = (-\infty,+\infty)$, $I_n = \mathbb{Z}$-the set of integers and $x_{n,k} = \frac{k}{n+1}$. In this case, $L_n^{(M)}$ in (1.1) become the Picard max-product operators. Let us denote by $\mathcal{P}K_n^{(M)}$ their Kantorovich variant defined by

$$\mathcal{P}K_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} e^{-|x-k/(n+1)|} \cdot (n+1) \int_{k/(n+1)}^{(k+1)/(n+1)} f(t)dt}{\bigvee_{k=0}^{\infty} e^{-|x-k/(n+1)|}}. \tag{2.10}$$

We can state the following result.

Theorem 2.11. *If $f : \mathbb{R} \to [0,+\infty)$ is bounded and uniformly continuous on $\mathbb{R}$, then we have*

$$|\mathcal{P}K_n^{(M)}(f)(x) - f(x)| \leq 6\omega_1(f;1/n), x \in \mathbb{R}, n \in \mathbb{N}.$$

Proof. It is immediate from Corollary 2.3 (with $C = 1$) and from Theorem 10.3.1, p. 423, in [2]. □

In what follows, let us choose $p_{n,k}(x) = e^{-(x-k/(n+1))^2}$, $I = (-\infty, +\infty)$, $I_n = \mathbb{Z}$-the set of integers and $x_{n,k} = \frac{k}{n+1}$. In this case, $L_n^{(M)}$ in (1.1) become the Weierstrass max-product operators. Let us denote by $\mathcal{W}K_n^{(M)}$ their Kantorovich variant defined by

$$\mathcal{W}K_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} e^{-(x-k/(n+1))^2} \cdot (n+1) \int_{k/(n+1)}^{(k+1)/(n+1)} f(t)dt}{\bigvee_{k=0}^{\infty} e^{-(x-k/(n+1))^2}}. \tag{2.11}$$

We can state the following result.

Theorem 2.12. *If $f : \mathbb{R} \to [0, +\infty)$ is bounded and uniformly continuous on $\mathbb{R}$, then we have*

$$|\mathcal{W}K_n^{(M)}(f)(x) - f(x)| \le 4\omega_1(f; 1/\sqrt{n}) + 2\omega_1(f; 1/n), x \in \mathbb{R}, n \in \mathbb{N}.$$

Proof. It is immediate from Corollary 2.3 (with $C = 1$) and from Theorem 10.3.3, p. 425, in [2]. □

At the end of this section, let us choose $p_{n,k}(x) = \frac{1}{n^2(x-k/n)^2+1}$, $I = (-\infty, +\infty)$, $I_n = \mathbb{Z}$-the set of integers and $x_{n,k} = \frac{k}{n+1}$. In this case, $L_n^{(M)}$ in (1.1) become the Poisson-Cauchy max-product operators. Let us denote by $\mathcal{C}K_n^{(M)}$ their Kantorovich variant

$$\mathcal{C}K_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} \frac{1}{n^2(x-k/(n+1))^2+1} \cdot (n+1) \int_{k/(n+1)}^{(k+1)/(n+1)} f(t)dt}{\bigvee_{k=0}^{\infty} \frac{1}{n^2(x-k/(n+1))^2+1}}. \tag{2.12}$$

We can state the following result.

Theorem 2.13. *If $f : \mathbb{R} \to [0, +\infty)$ is bounded and uniformly continuous on $\mathbb{R}$, then we have*

$$|\mathcal{C}K_n^{(M)}(f)(x) - f(x)| \le 6\omega_1(f; 1/n), x \in \mathbb{R}, n \in \mathbb{N}.$$

Proof. It is immediate from Corollary 2.3 (with $C = 1$) and from Theorem 10.3.5, p. 426, in [2]. □

Remark 2.14. All the Kantorovich kind max-product operators $LK_n^{(M)}$ given by (1.2) are defined and used for approximation of positive valued functions. But, they can be used for approximation of lower bounded functions of variable sign too, by introducing the new operators

$$N_n^{(M)}(f)(x) = LK_n^{(M)}(f + c)(x) - c,$$

where $c > 0$ is such that $f(x) + c > 0$, for all x in the domain of definition of f.

It is easy to see that the operators $N_n^{(M)}$ give the same approximation orders as $LK_n^{(M)}$.

3. Shape preserving properties for the Bernstein-Kantorovich max-product operators

In this section we deal with the shape preserving properties of the Bernstein-Kantorovich max-product operators $BK_n^{(M)}$ given by (2.3).

We can prove the following.

Theorem 3.1. *Let* $f \in C_+([0,1])$.

(i) If f *is non-decreasing (non-increasing) on* $[0,1]$, *then for all* $n \in \mathbb{N}$, $BK_n^{(M)}(f)$ *is non-decreasing (non-increasing, respectively) on* $[0,1]$.

(ii) If f *is quasi-convex on* $[0,1]$ *then for all* $n \in \mathbb{N}$, $BK_n^{(M)}(f)$ *is quasi-convex on* $[0,1]$. *Here quasi-convexity on* $[0,1]$ *means that* $f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$, *for all* $x, y, \lambda \in [0,1]$.

Proof. (i) By using the formula (2.1) for $LK_n^{(M)}$, we can write $BK_n^{(M)}(f)$ under the form

$$BK_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \cdot f(\xi_{n,k})}{\bigvee_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}},$$

where $\xi_{n,k} \in (x_{n,k}, x_{n,k+1})$, for all $k = 0, \ldots, n$.

Then, by analogy with the proofs for the Bernstein max-product operators (see [2], pp. 39-41, the proofs for the Bernstein-Kantorovich max-product operators, will be based on the properties of the functions

$$f_{k,n,j}(x) = \frac{\binom{n}{k}}{\binom{n}{j}} \cdot \left(\frac{x}{1-x}\right)^{k-j} \cdot f(\xi_{n,k}).$$

Now, analyzing the proofs of Lemma 2.1.13, Corollary 2.1.14, Theorem 2.1.15 and Corollary 2.1.16 in [2], pp. 39-41, it is easy to see that they work identically for the above $f_{k,n,j}$ too and we immediately obtain the required conclusions.

(ii) Since as in the case of the max-product Bernstein operators in Corollary 2.1.18, p. 41 in [2], this point is based on the properties from the above point (i) and on the properties in the above Lemma 2.1, (i)-(iv), we easily get the required conclusion for this point too. □

In what follows, we will prove that $BK_n^{(M)}$ preserves quasi-concavity too. This property holds in the case of the operator $B_n^{(M)}$ (By Theorem 5.1 in [5]). However, it is difficult to adapt the proof to our case. Instead, we can prove this property by finding a direct correspondence between the operators $B_n^{(M)}$ and $BK_n^{(M)}$.

Let us notice that the operator $BK_n^{(M)}$ can be obtained from the operator $B_n^{(M)}$. Suppose that f is arbitrary in $C_+([0,1])$. Let us consider

$$f_n(x) = (n+1) \int_{nx/(n+1)}^{(nx+1)/(n+1)} f(t)dt \tag{3.1}$$

It is readily seen that $B_n^{(M)}(f_n)(x) = BK_n^{(M)}(f)(x)$, for all $x \in [0,1]$. We also notice that $f_n \in C_+([0,1])$. What is more, if f is strictly positive then so is f_n.

A function $f : [a, b] \to \mathbb{R}$ is quasi-concave if $-f$ is quasi-convex. If f is continuous, quasi-concavity equivalently means that there exists $c \in [a, b]$ such that f is nondecreasing on $[a, c]$ and nonincreasing on $[c, b]$.

We are now in position to prove that $BK_n^{(M)}$ preserves quasi-concavity too.

Theorem 3.2. *Let $f \in C_+([0, 1])$. If fis quasi-concave on $[0, 1]$ then $BK_n^{(M)}(f)$ is quasi-concave on $[0, 1]$.*

Proof. For some arbitrary $n \geq 1$ let us consider the function f_n given by (3.1). Moreover, let $c \in [0, 1]$ such that f is nondecreasing on $[0, c]$ and nonincreasing on $[c, 1]$. Then, let $j(c) \in \{0, \ldots, n\}$ such that

$$\frac{j(c)}{n+1} \leq c \leq \frac{j(c)+1}{n+1}.$$

Next, we consider the function g_n which interpolates f_n at all the knots $\frac{k}{n}$, $k = 0, 1, \ldots, n$, and which is continuous on $[0, 1]$ and affine on any interval $\left[\frac{k}{n}, \frac{k+1}{n}\right]$, $k = 0, 1, \ldots, n-1$. It means that g_n is the continuous polygonal line which interpolates f_n at all the knots $\frac{k}{n}$, $k = 0, 1, \ldots, n$. This easily implies that

$$B_n^{(M)}(f_n)(x) = B_n^{(M)}(g_n)(x),\ x \in [0, 1],$$

hence,

$$BK_n^{(M)}(f)(x) = B_n^{(M)}(g_n)(x),\ x \in [0, 1].$$

Let us now choose arbitrary $0 \leq k_1 < k_2 \leq j(c) - 1$. We have

$$g_n\left(\frac{k_1}{n}\right) = (n+1)\int_{k_1/(n+1)}^{(k_1+1)/(n+1)} f(t)dt$$

and

$$g_n\left(\frac{k_2}{n}\right) = (n+1)\int_{k_2/(n+1)}^{(k_2+1)/(n+1)} f(t)dt.$$

As $\frac{k_1+1}{n+1} \leq \frac{k_2}{n+1}$ and f is nondecreasing on $[0, \frac{k_2+1}{n+1}]$, we easily obtain (after applying the mean value theorem) that $g_n\left(\frac{k_1}{n}\right) \leq g_n\left(\frac{k_2}{n}\right)$. The construction of g_n easily implies that g_n is nondecreasing on $\left[0, \frac{j(c)-1}{n}\right]$. By similar reasoning we get that g_n is nonincreasing on $\left[\frac{j(c)+1}{n}, 1\right]$. Now, suppose that $f\left(\frac{j(c)}{n+1}\right) \geq f\left(\frac{j(c)+1}{n+1}\right)$. The quasi-concavity of f implies that $f(x) \geq f\left(\frac{j(c)+1}{n+1}\right)$ for any $x \in \left[\frac{j(c)}{n+1}, \frac{j(c)+1}{n+1}\right]$. Since there exists $x_0 \in \left[\frac{j(c)}{n+1}, \frac{j(c)+1}{n+1}\right]$ such that

$$(n+1)\int_{j(c)/(n+1)}^{(j(c)+1)/(n+1)} f(t)dt = f(x_0) = g_n\left(\frac{j(c)}{n}\right),$$

and since $f\left(\frac{j(c)+1}{n+1}\right) \geq g_n\left(\frac{j(c)+1}{n}\right)$ (this is true indeed as f is nondecreasing on $\left[\frac{j(c)+1}{n+1}, 1\right]$), we get that $g_n\left(\frac{j(c)}{n}\right) \geq g_n\left(\frac{j(c)+1}{n}\right)$. Therefore, g_n is nonincreasing on $\left[\frac{j(c)}{n}, \frac{j(c)+1}{n}\right]$. This implies that g_n is nondecreasing on $\left[0, \frac{j(c)-1}{n}\right]$ and nonincreasing on $\left[\frac{j(c)}{n}, 1\right]$. But f is affine on $\left[\frac{j(c)-1}{n}, \frac{j(c)}{n}\right]$ which means that it is monotone on this interval. Clearly this implies that g_n is either nondecreasing on $\left[0, \frac{j(c)-1}{n}\right]$ and nonincreasing on $\left[\frac{j(c)-1}{n}, 1\right]$ or, it is nondecreasing on $\left[0, \frac{j(c)}{n}\right]$ and nonincreasing on $\left[\frac{j(c)}{n}, 1\right]$. It means that g_n is quasi-concave on $[0, 1]$. By similar reasonings we get to the same conclusion if $f\left(\frac{j(c)}{n+1}\right) \leq f\left(\frac{j(c)+1}{n+1}\right)$. The only difference is that now g_n is either nondecreasing on $\left[0, \frac{j(c)}{n}\right]$ and nonincreasing on $\left[\frac{j(c)}{n}, 1\right]$ or, it is nondecreasing on $\left[0, \frac{j(c)+1}{n}\right]$ and nonincreasing on $\left[\frac{j(c)+1}{n}, 1\right]$. Thus, we just proved that g_n is quasi-concave on $[0, 1]$. By Theorem 5.1 in [5] (see also Theorem 2.2.22 in the book, it follows that $B_n^{(M)}(g_n)$ is quasi-concave on $[0, 1]$. As $B_n^{(M)}(g_n) = BK_n^{(M)}(f)$, it follows that $BK_n^{(M)}(f)$ is quasi-concave on $[0, 1]$. □

As an important side remark, let us note that in Theorem 5.1 of paper [5](see also the book [2]), it is proved that if f is quasi-concave and c is a maximum point of f then there exists a maximum point of $B_n^{(M)}(f)$ such that $|c - c\prime| \leq \frac{1}{n+1}$. By the construction of g_n it follows that one maximum point of g_n is between the values $\frac{j(c)-1}{n}$, $\frac{j(c)}{n}$ or $\frac{j(c)+1}{n}$. If we denote this value with c_n then one can easily check that $|c_n - c| \leq \frac{2}{n}$. Now, applying the afore mentioned property obtained in [5], let $c\prime$ be a maximum point of $B_n^{(M)}(g_n) = BK_n^{(M)}(f)$, such that $|c\prime - c_n| \leq \frac{1}{n+1}$. This easily implies that $|c\prime - c| \leq \frac{3}{n}$. So, we obtained a quite similar result for the operator $BK_n^{(M)}$ in comparison with the operator $B_n^{(M)}$.

4. Approximation of Lipschitz functions by Bernstein-Kantorovich max-product operators

Let us return to the functions f_n given in (3.1) and let us find now an upper bound for the approximation of f by f_n in terms of the uniform norm. For some $x \in [0, 1]$, using the mean value theorem, there exists $\xi_x \in \left[\frac{nx}{n+1}, \frac{nx+1}{n+1}\right]$ such that $f_n(x) = f(\xi_x)$. We also easily notice that $|\xi_x - x| \leq \frac{1}{n+1}$. It means that

$$|f(x) - f_n(x)| \leq \omega_1(f; 1/(n+1)), x \in \mathbb{R}, n \in \mathbb{N}. \tag{4.1}$$

In particular, if f is Lipschitz with constant C then f_n is Lipschitz continuous with constant $3C$. These estimation are useful to prove some inverse results in the case of the operator $BK_n^{(M)}$ by using analogue results already obtained for the operator $B_n^{(M)}$.

Below we present a result which gives for the class of Lipschitz function the order of approximation $1/n$ in the approximation by the operator $BK_n^{(M)}$, hence an analogue result which holds in the case of the operator $B_n^{(M)}$.

Theorem 4.1. *Suppose that f is Lipschitz on $[0,1]$ with Lipschitz constant C and suppose that the lower bound of f is $m_f > 0$. Then we have*

$$\left\| BK_n^{(M)}(f) - f \right\| \le 2C\left(\frac{C}{m_f} + 5\right)\cdot\frac{1}{n},\ n \ge 1.$$

Proof. The estimation is immediate using the estimation from Corollary 2.4, (iii), taking into account that $\omega_1(f;1/n) \le C/n$. □

5. Localization results for Bernstein-Kantorovich max-product operators

We firstly prove a very strong localization property of the operator $BK_n^{(M)}$.

Theorem 5.1. *Let $f, g : [0,1] \to [0,\infty)$be both bounded on $[0,1]$ with strictly positive lower bounds and suppose that there exist $a, b \in [0,1]$, $0 < a < b < 1$ such that $f(x) = g(x)$ for all $x \in [a,b]$. Then for all $c, d \in [a,b]$ satisfying $a < c < d < b$ there exists $\widetilde{n} \in \mathbb{N}$ depending only on f, g, a, b, c, d such that $BK_n^{(M)}(f)(x) = BK_n^{(M)}(g)(x)$ for all $x \in [c,d]$and $n \in \mathbb{N}$ with $n \ge \widetilde{n}$.*

Proof. Let us choose arbitrary $x \in [c,d]$ and for each $n \in \mathbb{N}$ let $j_x \in \{0,1,\dots,n\}$ be such that $x \in [j_x/(n+1), (j_x+1)/(n+1)]$. Then by relation (4.17) in [1] we have

$$BK_n^{(M)}(f)(x) = B_n^{(M)}(f_n)(x) = \bigvee_{k=0}^{n} (f_n)_{k,n,j_x}(x), \tag{5.1}$$

where for $k \in \{0,1,\dots,n\}$ we have

$$(f_n)_{k,n,j_x} = \frac{\binom{n}{k}}{\binom{n}{j_x}}\left(\frac{x}{1-x}\right)^{k-j_x} f_n\left(\frac{k}{n}\right). \tag{5.2}$$

and each f_n is given by (3.1). Let us denote with m_f, M_f and m_{f_n}, M_{f_n} respectively, the minimums and maximum values of the functions f and f_n, respectively. By the mean value theorem, one can easily notice that for any $x \in [0,1]$ there exists $\xi_{n,x} \in [0,1]$ such that $f_n(x) = f(\xi_{n,x})$. It means that $0 < m_f \le m_{f_n} \le M_{f_n} \le M_f$. In what follows, the proof is very similar to the proof of Theorem 2.1 in [6] (see also Theorem 2.4.1 in [2]). However, as often we will use f_n instead of f, especially since the constants obtained in the proof of Theorem 2.1 in [6] depend on f, in our setting these constants would depend on f_n, hence, they would depend on n, if we would apply directly the results in [6]. Therefore, there are some differences in the two proofs as our intention is to obtain constants that do not depend on f_n.

We need the set $I_{n,x} = \{k \in \{0,1,\dots,n\} : j_x - a_n \le k \le j_x + a_n\}$, where $a_n = \left[\sqrt[3]{n^2}\right]$ (here $[a]$ denotes the integer part of a). Now, suppose that $k \notin I_{n,x}$, and let us discuss first the case when $k < j_x - a_n$. If we look over the proof of Theorem 2.1 in [6], we observe that this proof is split in cases i) and ii). Case i) corresponds to the case when $k < j_x - a_n$. Furthermore this case is divided in two subcases i_a) and i_b). In subcase i_a) the inequality

$$\frac{f_{j_x,n,,j_x}(x)}{f_{k,n,j_x}(x)} \geq \left(1 + \frac{a_n}{nb - a_n}\right)^{a_n} \cdot \frac{f(j_x/n)}{f(k/n)}$$

is obtained which then gives

$$\frac{f_{j_x,n,j_x}(x)}{f_{k,n,j_x}(x)} \geq \left(1 + \frac{a_n}{nb - a_n}\right)^{a_n} \cdot \frac{m_f}{M_f}.$$

Applying this reasoning but considering f_n instead of f, we get

$$\frac{(f_n)_{j_x,n,j_x}(x)}{(f_n)_{k,n,j_x}(x)} \geq \left(1 + \frac{a_n}{nb - a_n}\right)^{a_n} \cdot \frac{f_n(j_x/n)}{f_n(k/n)}.$$

But since $m_f \leq m_{f_n} \leq M_{f_n} \leq M_f$, we get

$$\frac{(f_n)_{j_x,n,j_x}(x)}{(f_n)_{k,n,j_x}(x)} \geq \left(1 + \frac{a_n}{nb - a_n}\right)^{a_n} \cdot \frac{m_f}{M_f}.$$

We get the same conclusion all cases and subcases, that is, any lower bound for $\frac{f_{j_x,n,,j_x}(x)}{f_{k,n,j_x}(x)}$ is also a lower bound for $\frac{(f_n)_{j_x,n,j_x}(x)}{(f_n)_{k,n,j_x}(x)}$, for any k outside of $I_{n,x}$. Since in..., we proved that there exists $N_0 \in \mathbb{N}$ which may depend only on f, a, b, c, d, such that for any $n \geq N_0$, $k \in \{0, 1, \ldots, n\}$, with $k < j_x - a_n$ or $k > j_x + a_n$, we have $\frac{f_{j_x,n,,j_x}(x)}{f_{k,n,j_x}(x)} \geq 1$, it follows that $\frac{(f_n)_{j_x,n,j_x}(x)}{(f_n)_{k,n,j_x}(x)} \geq 1$, for any $n \geq N_0$, $k \in \{0, 1, \ldots, n\}$, with $k < j_x - a_n$ or $k > j_x + a_n$. Combining this fact with relations (5.1)-(5.2), we get that

$$BK_n^{(M)}(f)(x) = \bigvee_{k \in I_{n,x}} (f_n)_{k,n,j_x}(x), x \in [c, d],\ n \geq N_0.$$

Using a similar reasoning as in the proof of Theorem 2.1 in [6], in what follows, we will prove that N_0 can be replaced if necessary with a larger value $\widetilde{N}_1$ such that $[\frac{k}{n+1}, \frac{k+1}{n+1}] \subseteq [a, b]$ for any $k \in I_{n,x}$. Let us choose arbitrary $x \in [c, d]$ and $n \in \mathbb{N}$ so that $n \geq N_0$. If there exists $k \in I_{n,x}$ such that $k/(n+1) \notin [c, d]$ then we distinguish two cases. Either $\frac{k}{n+1} < c$ or $\frac{k}{n+1} > d$. In the first case we observe that

$$0 < c - \frac{k}{n+1} \leq x - \frac{k}{n+1} \leq \frac{j_x + 1}{n+1} - \frac{k}{n+1} \leq \frac{j_x + 1}{n+1} - \frac{k}{n+1} \leq \frac{a_n + 1}{n+1}.$$

Since $\lim_{n \to \infty} \frac{a_n+1}{n+1} = 0$, it results that for sufficiently large n we necessarily have $\frac{a_n+1}{n+1} < c - a$ which clearly implies that $\frac{k}{n+1} \in [a, c]$. In the same manner, when $\frac{k}{n+1} > d$, for sufficiently large n we necessarily have $\frac{k}{n+1} \in [d, b]$. By similar reasoning it results that for sufficiently large n we necessarily have $\frac{k}{n+1} \in [a, b]$.Summarizing, there exists a constant $\widetilde{N}_1 \in \mathbb{N}$ independent of any $x \in [c, d]$ such that

$$BK_n^{(M)}(f)(x) = \bigvee_{k \in I_{n,x}} (f_n)_{k,n,j_x}(x),\ x \in [c, d],\ n \geq \widetilde{N}_1$$

and in addition for any $x \in [c,d]$, $n \geq \widetilde{N}_1$ and $k \in I_{n,x}$, we have $[\frac{k}{n+1}, \frac{k+1}{n+1}] \subseteq [a,b]$. Also, it is easy to check that $\widetilde{N}_1$ depends only on a,b,c,d,f.

Now, for $k \in \{0,1,\dots,n\}$ taking

$$(g_n)_{k,n,j_x} = \frac{\binom{n}{k}}{\binom{n}{j_x}} \left(\frac{x}{1-x}\right)^{k-j_x} g_n\left(\frac{k}{n}\right),$$

applying the same reasoning, there exists $\widetilde{N}_2 \in \mathbb{N}$ which may depend only on a,b,c,d,g, such that

$$BK_n^{(M)}(g)(x) = \bigvee_{k\in I_{n,x}} (g_n)_{k,n,j_x}(x),\ x \in [c,d],\ n \geq \widetilde{N}_2$$

and in addition for any $x \in [c,d]$, $n \geq \widetilde{N}_2$ and $k \in I_{n,x}$, we have $[\frac{k}{n+1}, \frac{k+1}{n+1}] \subseteq [a,b]$. Since $f(x) = g(x)$, $x \in [a,b]$, we get that for any $n \geq \widetilde{n} = \max\{\widetilde{N}_1, \widetilde{N}_2\}$, $k \in I_{n.x}$ and $x \in [c,d]$, it holds that $(f_n)_{k,n,j_x}(x) = (g_n)_{k,n,j_x}(x)$. Thus, for any $n \geq \widetilde{n}$ and $x \in [c,d]$, we have $BK_n^{(M)}(f)(x) = BK_n^{(M)}(g)(x)$. The proof is complete now. □

As in the case of the Bernstein max-product operator, we can present a local direct approximation result as an immediate consequence of the localization result in Theorem 5.1.

Corollary 5.2. *Let $f : [0,1] \to [0,\infty)$ be bounded on $[0,1]$ with the lower bound strictly positive and $0 < a < b < 1$ be such that $f|_{[a,b]} \in Lip\,[a,b]$ with Lipschitz constant $\overline{C}$. Then, for any $c,d \in [0,1]$ satisfying $a < c < d < b$, we have*

$$\left|BK_n^{(M)}(f)(x) - f(x)\right| \leq \frac{C}{n} \text{ for all } n \in \mathbb{N} \text{ and } x \in [c,d],$$

where the constant C depends only on f and a,b,c,d.

Proof. Let us define the function $F : [0,1] \to \mathbb{R}$,

$$F(x) = \begin{cases} f(a) & \text{if} \quad x \in [0,a], \\ f(x) & \text{if} \quad x \in [a,b], \\ f(b) & \text{if} \quad x \in [b,1]. \end{cases}$$

The hypothesis immediately imply that F is a strictly positive Lipschitz function on $[0,1]$. Then, according to Theorem 4.1 and noting that the minimum of F is above the minimum of f, m_f, it results that

$$\left|BK_n^{(M)}(F)(x) - F(x)\right| \leq 2\overline{C}\left(\frac{\overline{C}}{m_f} + 5\right) \cdot \frac{1}{n}, \text{for all } x \in [0,1], n \in \mathbb{N}.$$

Now, let us choose arbitrary $c,d \in [a,b]$ such that $a < c < d < b$. Then, by Theorem 5.1 it results the existence of $\widetilde{n} \in \mathbb{N}$ which depends only on a,b,c,d,f,F such that $BK_n^{(M)}(F)(x) = BK_n^{(M)}(f)(x)$ for all $x \in [c,d]$. But since actually the function F depends on the function f, by simple reasonings we get that in fact $\widetilde{n}$ depends only on a,b,c,d and f.

Therefore, for arbitrary $x \in [c,d]$ and $n \in \mathbb{N}$ with $n \geq \widetilde{n}$ we obtain

$$\left|BK_n^{(M)}(f)(x) - f(x)\right| = \left|BK_n^{(M)}(F)(x) - F(x)\right| \leq 2\overline{C}\left(\frac{\overline{C}}{m_f} + 5\right) \cdot \frac{1}{n},$$

where C_1 and $\widetilde{n}$ depend only on a, b, c, d and f.

Now, denoting

$$C_2 = \max_{1 \leq n < \widetilde{n}} \{n \cdot \|BK_n^{(M)}(f) - f\|_{[c,d]}\},$$

we finally obtain

$$|BK_n^{(M)}(f)(x) - f(x)| \leq \frac{C}{n}, \text{for all } n \in \mathbb{N},\, x \in [c,d],$$

with $C = \max\{2\overline{C}\left(\frac{\overline{C}}{m_f} + 5\right), C_2\}$ depending only on a, b, c, d and f. □

In a previous section we proved that $BK_n^{(M)}$ preserves monotonicity and more generally quasi-convexity. By the localization result in Theorem 5.1 and then applying a very similar reasoning to the one used in the proof of Corollary 5.2, we obtain local versions for these shape preserving properties. Indeed, in all cases it will suffice to consider the same F as in the proof of Corollary 5.2 as this function will be monotone or quasi-convex/quasi-concave, respectively, whenever f will be monotone or quasi-convex/quasi-concave, respectively. For this reason we omit the proofs of the following corollaries (see also the corresponding local shape preserving properties proved for the operator $B_n^{(M)}$ in [6]).

Corollary 5.3. *Let $f : [0,1] \to [0,\infty)$be bounded on $[0,1]$ with strictly positive lower bound and suppose that there exists $a, b \in [0,1]$, $0 < a < b < 1$, such that f is nondecreasing (nonincreasing) on $[a,b]$. Then for any $c, d \in [a,b]$ with $a < c < d < b$, there exists $\widetilde{n} \in \mathbb{N}$ depending only on a, b, c, d and f, such that $B_n^{(M)}(f)$ is nondecreasing (nonincreasing) on $[c,d]$ for all $n \in \mathbb{N}$ with $n \geq \widetilde{n}$.*

Corollary 5.4. *Let $f : [0,1] \to [0,\infty)$be a continuous and strictly positive function and suppose that there exists $a, b \in [0,1]$, $0 < a < b < 1$, such that f is quasi-convex on $[a,b]$. Then for any $c, d \in [a,b]$ with $a < c < d < b$, there exists $\widetilde{n} \in \mathbb{N}$ depending only on a, b, c, d and $f$such that $B_n^{(M)}(f)$ is quasi-convex on $[c,d]$ for all $n \in \mathbb{N}$ with $n \geq \widetilde{n}$.*

Corollary 5.5. *Let $f : [0,1] \to [0,\infty)$be a continuous and strictly positive function and suppose that there exists $a, b \in [0,1]$, $0 < a < b < 1$, such that f is quasi-concave on $[a,b]$. Then for any $c, d \in [a,b]$ with $a < c < d < b$, there exists $\widetilde{n} \in \mathbb{N}$ depending only on a, b, c, d and f, such that $B_n^{(M)}(f)$ is quasi-concave on $[c,d]$ for all $n \in \mathbb{N}$ with $n \geq \widetilde{n}$.*

Remark 5.6. As in the cases of Bernstein-type max-product operators studied in the research monograph [2], for the the max-product Kantorovich type operators we can find natural interpretation as possibilistic operators, which can be deduced from the Feller scheme written in terms of the possibilistic integral. These approaches also offer new proofs for the uniform convergence, based on a Chebyshev type inequality in the theory of possibility.

Remark 5.7. In the recently submitted paper [3], we have introduced the more general Kantorovich max-product operators based on a generalized (φ, ψ)-kernel, by the formula

$$K_n^{(M)}(f; \varphi, \psi)(x) = \frac{1}{b} \cdot \frac{\bigvee_{k=0}^{n} \frac{\varphi(nx-kb)}{\psi(nx-kb)} \cdot \left[(n+1)\int_{kb/(n+1)}^{(k+1)b/(n+1)} f(v)\, dv\right]}{\bigvee_{k=0}^{n} \frac{\varphi(nx-kb)}{\psi(nx-kb)}}, \tag{5.3}$$

where $b > 0$, $f : [0, b] \to \mathbb{R}_+$, $f \in L^p[0, b]$, $1 \leq p \leq \infty$ and φ and ψ satisfy some properties specific to max-product operators and proved pointwise, uniform or L^p convergence quantitative approximation results. For particular choices of (φ, ψ), we have obtained approximation results for many other max-product Kantorovich operators, including for example the sampling operators based on sinc-type kernels.

Remark 5.8. In another recently in preparation paper [4], we have generalized the max-product Kantorovich operators from the above Remark 2), by replacing the classical linear integral $\int dv$, by the nonlinear Choquet integral $(C)\int d\mu(v)$ with respect to a monotone and submodular set function μ obtaining and studying the max-product Kantorovich-Choquet operators given by the formula

$$K_n^{(M)}(f; \varphi, \psi)(x)$$
$$= \frac{1}{b} \cdot \frac{\bigvee_{k=0}^{n} \frac{\varphi(nx-kb)}{\psi(nx-kb)} \cdot \left[(C)\int_{kb/(n+1)}^{(k+1)b/(n+1)} f(v)\, d\mu(v)/\mu\left(\left[\frac{kb}{n+1}, \frac{(k+1)b}{n+1}\right]\right)\right]}{\bigvee_{k=0}^{n} \frac{\varphi(nx-kb)}{\psi(nx-kb)}}, \tag{5.4}$$

It is worth noting that the max-product Kantorovich-Choquet operators are **doubly nonlinear operators**: firstly due to max and secondly, due to the Choquet integral.

Acknowledgement. The work of both authors was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PN-III-P1-1.1-PD-2016-1416.

References

[1] Bede, B., Coroianu, L., Gal, S.G., *Approximation and shape preserving properties of the Bernstein operator of max-product kind*, Intern. J. Math. Math. Sci., **2009**, Art. ID 590589, 2009.

[2] Bede, B., Coroianu, L., Gal, S.G., *Approximation by Max-Product Type Operators*, Springer, 2016, xv+458 pp.

[3] Coroianu, L., Gal, S.G., *Approximation by truncated max-product operators of Kantorovich-type based on generalized (φ, ψ)-kernels*, Math. Meth. Appl. Sci., online access, 2018: 1-14, DOI: 10.1002/mma.5262.

[4] Coroianu, L., Gal, S.G., *Approximation by max-product operators of Kantorovich-Choquet type based on generalized (φ, ψ)-kernels*, in preparation.

[5] Coroianu, L., Gal, S.G., *Classes of functions with improved estimates in approximation by the max-product Bernstein operator,* Anal. Appl., **9**(2011), 249–274.

[6] Coroianu, L., Gal, S.G., *Localization results for the Bernstein max-product operator,* Appl. Math. Comp., **231**(2014), 73–78.

[7] Gal, S.G., *Shape-Preserving Approximation by Real and Complex Polynomials*, XIV+352 pp., Birkhäuser, Boston, Basel, Berlin, 2008.

Lucian Coroianu
University of Oradea
Department of Mathematics and Computer Sciences
Universităţii Street, No. 1
410087 Oradea, Romania
e-mail: lcoroianu@uoradea.ro

Sorin G. Gal
University of Oradea
Department of Mathematics and Computer Sciences
Universităţii Street, No. 1
410087 Oradea, Romania
e-mail: galso@uoradea.ro

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 225–237
DOI: 10.24193/subbmath.2019.2.08

Operator norms of Gauß-Weierstraß operators and their left quasi interpolants

Ulrich Abel

Abstract. The paper deals with the Gauß–Weierstraß operators W_n and their left quasi interpolants $W_n^{[r]}$. The quasi interpolants were defined by Paul Sablonnière in 2014. Recently, their asymptotic behaviour was studied by Octavian Agratini, Radu Păltănea and the author by presenting complete asymptotic expansions. In this paper we derive estimates for the operator norms of W_n and $W_n^{[r]}$ when acting on various function spaces.

Mathematics Subject Classification (2010): 41A36, 41A45, 47A30.

Keywords: Approximation by positive operators, operator norm.

1. Introduction

For $1 \le p \le +\infty$ and $c > 0$, let $L_c^p(\mathbb{R})$ denote the space of all locally integrable functions $f : \mathbb{R} \to \mathbb{R}$, such that the weighted norm $\|fw_c\|_{L^p(\mathbb{R})}$

$$\|f\|_{L_c^p(\mathbb{R})} := \left(\int_{-\infty}^{\infty} |f(t)|^p \, w_c(t) \, dt\right)^{1/p} \qquad (1 \le p < +\infty),$$

$$\|f\|_{L_c^\infty(\mathbb{R})} := \operatorname*{ess\,sup}_{t\in\mathbb{R}} |f(t)| \, w_c(t) \qquad (p = +\infty)$$

is finite, where the weight function w_c is given by

$$w_c(t) := e^{-ct^2}.$$

In the particular case $c = 0$, we obtain the ordinary spaces $L_0^p(\mathbb{R}) = L^p(\mathbb{R})$ and $L_0^\infty(\mathbb{R}) = L^\infty(\mathbb{R})$, respectively.

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

The nth Gauß–Weierstraß convolution operator W_n (see, e.g., [7, Section 5.2.9]) is defined by

$$(W_n f)(x) = \sqrt{\frac{n}{\pi}} \int_{-\infty}^{\infty} e^{-n(t-x)^2} f(t)\, dt. \tag{1.1}$$

Note that the integral on the right-hand side exists, for $f \in L_c^p(\mathbb{R})$, provided that $n > c$. We have convergence $\lim_{n\to\infty} (W_n f)(x) = f(x)$ in each continuity point $x \in \mathbb{R}$ of $f \in L_c^p(\mathbb{R})$. The operator W_n played a key role in the original proof of the Weierstraß approximation theorem. The properties of W_n have been studied by many authors (we refer to [8] for details). What regards the local rate of convergence as n tends to infinity the sequence (W_n) satisfies the Voronovskaja-type formula

$$\lim_{n\to\infty} n\left((W_n f)(x) - f(x)\right) = \frac{1}{4} f''(x),$$

provided that the derivative $f''(x)$ exists. For more smooth functions the operators W_n possess the complete asymptotic expansion

$$(W_n f)(x) \sim f(x) + \sum_{k=1}^{\infty} \frac{1}{4^k k! n^k} f^{(2k)}(x) \qquad (n \to \infty). \tag{1.2}$$

This formula follows from [4, Theorem 5.1] where it was proved for a more general sequence of operators introduced by Altomare and Milella [6, Eq. (2.5)]. Eq. (1.2) is valid also with respect to simultaneous approximation [2, Proposition 3.4.] where it turns out that the asymptotic expansion can be differentiated term-by-term. In particular, for $m = 0, 1, 2, \ldots$, we have

$$\lim_{n\to\infty} n\left((W_n f)^{(m)}(x) - f^{(m)}(x)\right) = \frac{1}{4} f^{(m+2)}(x).$$

Eq. (1.2) was rediscovered, for polynomial functions, by Sablonnière [11, Theorem 1] in 2014. With this recent paper he renewed the interest in Gauß–Weierstraß operators. Sablonnière defined left and right quasi-interpolants $W_n^{[r]}$ and $W_n^{(r)}$, resp., of W_n, presented their explicit integral representations and derived a plenty of nice properties. In particular, Sablonnière [11, Theorem 5] expressed the operator norm of $W_n^{[r]}$ with respect to the uniform norm in terms of a certain integral which cannot be exactly evaluated. He proved that $r + \sqrt{2}$ is an upper bound on this operator norm [11, Theorem 6].

In this paper we considerably improve the upper bound. Furthermore, we study the operator norms of W_n and $W_n^{[r]}$ when acting on various function spaces.

2. The left quasi interpolants

The Gauß–Weierstraß operators possess the representation

$$W_n = \sum_{k\geq 0} \frac{1}{4^k k! n^k} D^{2k}$$

as a differential operator on the space of algebraic polynomials [11, Theorem 1]. Here D denotes the differentiation operator. The inverse operator [11, Theorem 1] is given by

$$V_n = \sum_{k\geq 0} \frac{(-1)^k}{4^k k! n^k} D^{2k}.$$

Composition of the the partial sums $V_n^{[r]}$ of order r and W_n defines the left quasi interpolants

$$W_n^{[r]} := V_n^{[r]} \circ W_n = \sum_{k=0}^{r} \frac{(-1)^k}{4^k k! n^k} D^{2k} W_n.$$

[11, Subsection 4.1].

By [11, Theorem 3] (where $\tilde{H}_{2r}(x-t)$ correctly reads $\tilde{H}_{2r}(\sqrt{n}(x-t))$ in the first representation), the left quasi-interpolants $W_n^{[r]}$ of the Gauß–Weierstraß operators possess the integral representation

$$\left(W_n^{[r]} f\right)(x) = \sqrt{\frac{n}{\pi}} \int_{-\infty}^{\infty} \tilde{H}_{2r}\left(\sqrt{n}(x-t)\right) e^{-n(t-x)^2} f(t)\, dt. \tag{2.1}$$

The polynomials $\tilde{H}_{2r}$ are defined by

$$\tilde{H}_{2r}(x) = \sum_{k=0}^{r} \frac{(-1)^k}{4^k k!} H_{2k}(x), \tag{2.2}$$

where H_k denote the Hermite polynomials [11, p. 38]. Sablonnière proved the explicit representation [11, Theorem 4]

$$\tilde{H}_{2r}(x) = \frac{(2r+1)!}{r!} \sum_{k=0}^{r} \frac{(-1)^{r-k}}{4^k k! (2r-2k+1)!} x^{2(r-k)}. \tag{2.3}$$

In the next section we frequently make use of the following Lemma.

Lemma 2.1. *For $r = 0, 1, 2, \ldots$, the polynomials $\tilde{H}_{2r}$ satisfy the relation*

$$\int_{-\infty}^{\infty} \left(\tilde{H}_{2r}(t)\right)^2 e^{-t^2} dt = \sqrt{\pi}\binom{r+1/2}{r}. \tag{2.4}$$

Furthermore, we have the estimate

$$2\sqrt{r+3/4} < \sqrt{\pi}\binom{r+1/2}{r} \leq 2\sqrt{r+1} \qquad (r \geq 0) \tag{2.5}$$

and the asymptotic relation

$$\sqrt{\pi}\binom{r+1/2}{r} \sim 2\sqrt{r} \qquad (r \to \infty). \tag{2.6}$$

Remark 2.2. In other words, we have

$$\left\|\tilde{H}_{2r}\right\|_{L_1^2(\mathbb{R})} = \sqrt{\sqrt{\pi}\binom{r+1/2}{r}}.$$

Proof of Lemma 2.1. Taking advantage of the orthogonality of the Hermite polynomials (see, e.g., [5, formula 22.2.14]) we obtain

$$\begin{aligned}
\int_{-\infty}^{\infty} \tilde{H}_{2r}^{2}(t) \exp\left(-t^{2}\right) dt &= \sum_{k=0}^{r} \frac{(-1)^{k}}{4^{k} k!} \sum_{j=0}^{r} \frac{(-1)^{j}}{4^{j} j!} \int_{-\infty}^{\infty} H_{2k}(x) H_{2j}(x) e^{-t^{2}} dt \\
&= \sum_{k=0}^{r} \frac{1}{4^{2k} (k!)^{2}} \int_{-\infty}^{\infty} H_{2k}^{2}(x) e^{-t^{2}} dt \\
&= \sum_{k=0}^{r} \frac{1}{4^{2k} (k!)^{2}} \sqrt{\pi} 2^{2k} (2k)! \\
&= \sqrt{\pi} \sum_{k=0}^{r} \frac{1}{4^{k}} \binom{2k}{k}.
\end{aligned}$$

Application of the well-known identity [10, formulas (1.108) and (1.109)]

$$\sum_{k=0}^{r} \frac{1}{4^{k}} \binom{2k}{k} = \binom{r+1/2}{r}$$

proves Eq. (2.4). The bounds from below and above are a consequence of the estimate [12]

$$\sqrt{y+1/4} < \frac{\Gamma(y+1)}{\Gamma(y+1/2)} \leq \sqrt{y+1/\pi} \qquad (y \geq 0).$$

Using $\Gamma(3/2) = \sqrt{\pi}/2$, this implies

$$\begin{aligned}
\frac{2}{\sqrt{\pi}} \sqrt{r+1/2+1/4} &< \binom{r+1/2}{r} = \frac{\Gamma(r+3/2)}{\Gamma(3/2)\Gamma(r+1)} \\
&\leq \frac{2}{\sqrt{\pi}} \sqrt{r+1/2+1/\pi}.
\end{aligned}$$

Application of the well-known formula [5, formula 6.1.47]) yields the asymptotic relation

$$\sqrt{\pi} \binom{r+1/2}{r} = \Gamma(1/2) \frac{\Gamma(r+3/2)}{\Gamma(3/2)\Gamma(r+1)} \sim 2\sqrt{r} \left(1+\frac{2}{r}\right) \sim 2\sqrt{r}$$

as $r \to \infty$. □

3. The operator norms of W_n and $W_n^{[r]}$ in the space $L^{\infty}(\mathbb{R})$

We consider the operator norm of

$$W_n^{[r]} : L^{\infty}(\mathbb{R}) \to L^{\infty}(\mathbb{R})$$

with respect to the sup-norm on $L^{\infty}(\mathbb{R})$. Sablonnière [11, Theorem 5] gave the following result.

Proposition 3.1. *The operator norm with respect to the sup-norm on $L^\infty(\mathbb{R})$ is given by*

$$\left\|W_n^{[r]}\right\|_{(L^\infty(\mathbb{R}),L^\infty(\mathbb{R}))} = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right| e^{-t^2}dt.$$

Note that the value of $\left\|W_n^{[r]}\right\|_{(L^\infty(\mathbb{R}),L^\infty(\mathbb{R}))}$ is independent of n. As in [11, Theorem 5] we put, for the sake of brevity,

$$N_r := \left\|W_n^{[r]}\right\|_{(L^\infty(\mathbb{R}),L^\infty(\mathbb{R}))}.$$

Remark 3.2. In the special case $r = 0$ we obtain the well-known operator norm $\|W_n\|_{(L^\infty(\mathbb{R}),L^\infty(\mathbb{R}))} = 1$ of the Gauß–Weierstraß operator W_n, since $\tilde{H}_0(x) = 1$.

Since the proof given in [11] is not completely correct we present a proof.

Proof of Prop. 3.1. Let $f \in L^\infty(\mathbb{R})$. By Eq. (2.1), we have

$$\left(W_n^{[r]}f\right)(x) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\tilde{H}_{2r}(t)\, e^{-t^2} f\left(x - \frac{t}{\sqrt{n}}\right)dt.$$

Hence, for all $x \in \mathbb{R}$,

$$\left|\left(W_n^{[r]}f\right)(x)\right| \le N_r \cdot \|f\|_{L^\infty(\mathbb{R})}$$

which implies

$$\left\|W_n^{[r]}f\right\|_{L^\infty(\mathbb{R})} \le N_r.$$

The function $f_0(t) = \operatorname{sgn}\left(\tilde{H}_{2r}(-\sqrt{n}t)\right)$ satisfies

$$\left\|W_n^{[r]}f_0\right\|_{L^\infty(\mathbb{R})} \ge \left(W_n^{[r]}f_0\right)(0) = N_r = N_r \cdot \|f_0\|_{L^\infty(\mathbb{R})}$$

which completes the proof. □

Using the well-known estimate $|H_{2r}(x)| \le 2^r\sqrt{(2r)!}e^{x^2/2}$ (see, e.g., [9, Subsection 1.5.1, p. 31]), Sablonnière [11, Theorem 6] proves, for $r = 0, 1, 2, \ldots$, the both estimates

$$N_r \le \sqrt{2}\sigma_r := \sqrt{2}\left(1 + \sum_{p=1}^{r}\sqrt{\prod_{j=1}^{p}\frac{2j-1}{2j}}\right) \le C_r := r + \sqrt{2}.$$

We improve these upper bounds as follows.

Theorem 3.3. *For $r = 0, 1, 2, \ldots$, the operator norm $\left\|W_n^{[r]}\right\|_{(L^\infty(\mathbb{R}),L^\infty(\mathbb{R}))} = N_r$ satisfies the estimate*

$$N_r \le \binom{r+1/2}{r}^{1/2} =: D_r.$$

Lemma 2.3 implies the asymptotic relation

$$D_r \sim \sqrt[4]{\frac{4r}{\pi}} \qquad (r \to \infty)$$

and the following estimate.

Corollary 3.4. *For $r = 0, 1, 2, \ldots$, the operator norm $\left\| W_n^{[r]} \right\|_{(L^\infty(\mathbb{R}), L^\infty(\mathbb{R}))} = N_r$ possesses the upper bound*

$$N_r \leq \sqrt[4]{4 \frac{r+1}{\pi}}.$$

The next table shows some numerical values of N_r up to $r = 10$, its estimates $\sqrt{2}\sigma_r$ and C_r by Sablonnière, followed by the new estimate D_r from Theorem 3.3:

r	N_r	$\sqrt{2}\sigma_r$	C_r	D_r
1	1.14	2.41	2.41	1.22
2	1.22	3.28	3.41	1.37
3	1.28	4.07	4.41	1.48
4	1.33	4.81	5.41	1.57
5	1.37	5.51	6.41	1.65
6	1.40	6.18	7.41	1.71
7	1.43	6.83	8.41	1.77
8	1.45	7.46	9.41	1.83
9	1.47	8.07	10.41	1.88
10	1.49	8.66	11.41	1.92

Proof of Theorem 3.3. By Prop. 3.1, we have

$$\left\| W_n^{[r]} \right\|_{(L^\infty(\mathbb{R}), L^\infty(\mathbb{R}))} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left| \tilde{H}_{2r}(t) \right| \sqrt{e^{-t^2}} \sqrt{e^{-t^2}} dt.$$

Application of the Schwarz inequality implies that

$$N_r \leq \frac{1}{\sqrt{\pi}} \sqrt{\int_{-\infty}^{\infty} \left| \tilde{H}_{2r}(t) \right|^2 e^{-t^2} dt} \sqrt{\int_{-\infty}^{\infty} e^{-t^2} dt} = \binom{r+1/2}{r}^{1/2},$$

where the last equality follows from Eq. (2.4) of Lemma 2.1. □

4. The operator norms of W_n and $W_n^{[r]}$ in weighted spaces

4.1. Weighted spaces

In the following we suppose that $c > 0$. Put $f_c = w_{-c}$, i.e.,

$$f_c(t) = e^{ct^2}.$$

Then, for all $n > c$,

$$W_n f_c = \sqrt{\frac{n}{n-c}} f_{nc/(n-c)}.$$

Note that, for $c > 0$, it holds $c < nc/(n-c)$. This means that the space $L_c^\infty(\mathbb{R})$ is not invariant under the mapping W_n. The function $f_c \in L_c^\infty(\mathbb{R})$ satifies $\|f_c\|_{L_c^\infty(\mathbb{R})} = 1$. However, $W_n f_c \notin L_c^\infty(\mathbb{R})$. Therefore, we consider the mapping $W_n : L_c^\infty(\mathbb{R}) \to L_\gamma^\infty(\mathbb{R})$, for some $\gamma > c$. Note that $\lim_{n\to\infty} nc/(n-c) = c$ implies that $nc/(n-c) < \gamma$, for sufficiently large integers n. More precisely, we have $W_n f_c \in L_\gamma^\infty(\mathbb{R})$, for all integers n satisfying

$$n > \gamma c/(\gamma - c).$$

In the following we consider only such values of n.

4.2. The space $L_c^\infty(\mathbb{R})$ equipped with the weighted sup-Norm

Let $c > 0$. Suppose that $f \in L_c^\infty(\mathbb{R})$ and $W_n^{[r]} f \in L_\gamma^\infty(\mathbb{R})$, for some $\gamma > c$. As we know, this is the case if $n > \gamma c/(\gamma - c)$. For these n, the operator norm $\left\|W_n^{[r]}\right\| = \left\|W_n^{[r]}\right\|_{\left(L_c^\infty(\mathbb{R}), L_\gamma^\infty(\mathbb{R})\right)}$ is defined by

$$\left\|W_n^{[r]}\right\| = \sup_{\substack{f\in L_c^\infty(\mathbb{R})\\ f\neq 0}} \frac{\left\|W_n^{[r]} f\right\|_{L_\gamma^\infty(\mathbb{R})}}{\|f\|_{L_c^\infty(\mathbb{R})}} = \sup_{\substack{f\in L_c^\infty(\mathbb{R})\\ \|f\|_{L_c^\infty(\mathbb{R})}=1}} \left\|W_n^{[r]} f\right\|_{L_\gamma^\infty(\mathbb{R})}.$$

Theorem 4.1. *Let $r \in \mathbb{N}$. If $0 < c < \gamma$, for all integers $n > 2\gamma c/(\gamma - c)$, the operator norm $\left\|W_n^{[r]}\right\|_{\left(L_c^\infty(\mathbb{R}), L_\gamma^\infty(\mathbb{R})\right)}$ satisfies the estimate*

$$\left\|W_n^{[r]}\right\|_{\left(L_c^\infty(\mathbb{R}), L_\gamma^\infty(\mathbb{R})\right)} \le \binom{r+1/2}{r}^{1/2} \left(\frac{n(\gamma-c)}{n(\gamma-c)-2c\gamma}\right)^{1/4}.$$

Remark 4.2. By Eq. (2.5) of Lemma 2.1, we have the upper bound

$$\|W_n\|_{\left(L_c^\infty(\mathbb{R}), L_\gamma^\infty(\mathbb{R})\right)} \le \sqrt[4]{4\frac{r+1}{\pi}\left(\frac{n(\gamma-c)}{n(\gamma-c)-2c\gamma}\right)}.$$

Remark 4.3. In the special case $r = 0$, we obtain the estimate

$$\|W_n\|_{\left(L_c^\infty(\mathbb{R}), L_\gamma^\infty(\mathbb{R})\right)} \le \left(\frac{n(\gamma-c)}{n(\gamma-c)-2c\gamma}\right)^{1/4}$$

for the classical Gauß–Weierstraß operators $W_n = W_n^{[0]}$.

Remark 4.4. The limit $n \to \infty$ leads for $\left\|W_n^{[r]}\right\|_{\left(L_c^\infty(\mathbb{R}), L_\gamma^\infty(\mathbb{R})\right)}$ to the upper bound

$$D_r = \sqrt{\binom{r+1/2}{r}} \sim \sqrt[4]{\frac{4r}{\pi}} \qquad (r \to \infty).$$

Proof of Theorem 4.1. Let $f \in L_c^\infty(\mathbb{R})$. By Eq. (2.1), we have

$$\left(W_n^{[r]} f\right)(x) = \sqrt{\frac{n}{\pi}} \int_{-\infty}^{\infty} \tilde{H}_{2r}\left(\sqrt{n}(x-t)\right) e^{-n(t-x)^2} w_{-c}(t) \cdot w_c(t) f(t)\, dt.$$

which can be estimated by

$$\begin{aligned}\left|\left(W_n^{[r]}f\right)(x)\right| &\leq \sqrt{\frac{n}{\pi}}\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}\left(\sqrt{n}\,(x-t)\right)\right|\exp\left(-n\,(t-x)^2\right)w_{-c}(t)\ dt\\ &\quad\times\sup_{t\in\mathbb{R}}|f(t)|\,w_c(t)\,.\end{aligned}$$

A change of variable replacing t with $x-t/\sqrt{n}$ yields

$$\left|\left(W_n^{[r]}f\right)(x)\right| \leq \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|e^{-t^2+c\left(x-t/\sqrt{n}\right)^2}\ dt\cdot\|f\|_{L_c^\infty(\mathbb{R})}\,.$$

Hence,

$$\begin{aligned}\left\|W_n^{[r]}f\right\|_{L_\gamma^\infty(\mathbb{R})} &= \sup_{x\in\mathbb{R}}\left|\left(W_n^{[r]}f\right)(x)\right|e^{-\gamma x^2}\\ &\leq \sup_{x\in\mathbb{R}}\left(\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|e^{-t^2+c\left(x-t/\sqrt{n}\right)^2}\ dt\cdot e^{-\gamma x^2}\right)\cdot\|f\|_{L_c^\infty(\mathbb{R})}\,.\end{aligned}$$

For fixed t, the expression $e^{c\left(x-t/\sqrt{n}\right)^2}e^{-\gamma x^2}$ attains (as a function of x) its maximum at $x=-ct/\left(\sqrt{n}\,(\gamma-c)\right)$, such that

$$\sup_{x\in\mathbb{R}}e^{c\left(x-t/\sqrt{n}\right)^2}e^{-\gamma x^2}=\exp\left(\frac{c\gamma}{n\,(\gamma-c)}t^2\right).$$

Consequently,

$$\left\|W_n^{[r]}f\right\|_{L_\gamma^\infty(\mathbb{R})} \leq \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|\exp\left(-\left(1-\frac{c\gamma}{n\,(\gamma-c)}\right)t^2\right)\ dt\cdot\|f\|_{L_c^\infty(\mathbb{R})}\,.$$

Hence,

$$\left\|W_n^{[r]}\right\|_{\left(L_c^\infty(\mathbb{R}),L_\gamma^\infty(\mathbb{R})\right)} \leq \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|\exp\left(-\left(1-\frac{c\gamma}{n\,(\gamma-c)}\right)t^2\right)\ dt.$$

Application of the Schwarz inequality yields

$$\begin{aligned}&\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|\exp\left(-\left(1-\frac{c\gamma}{n\,(\gamma-c)}\right)t^2\right)\ dt\\ &= \int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|\sqrt{e^{-t^2}}\cdot\sqrt{e^{t^2}}\exp\left(-\left(1-\frac{c\gamma}{n\,(\gamma-c)}\right)t^2\right)\ dt\\ &\leq \sqrt{\int_{-\infty}^{\infty}\left(\tilde{H}_{2r}(t)\right)^2e^{-t^2}dt}\sqrt{\int_{-\infty}^{\infty}\exp\left(-\left(1-\frac{2c\gamma}{n\,(\gamma-c)}\right)t^2\right)\ dt}\\ &= \sqrt{\sqrt{\pi}\binom{r+1/2}{r}}\sqrt{\sqrt{\pi}\left(1-\frac{2c\gamma}{n\,(\gamma-c)}\right)^{-1/2}}.\end{aligned}$$

By Lemma 2.1, and using the well-known identity $\int_{-\infty}^{\infty}\exp\left(-at^2\right)\ dt=\sqrt{\pi/a}$, for $a>0$, we obtain

$$\left\|W_n^{[r]}\right\|_{\left(L_c^\infty(\mathbb{R}),L_\gamma^\infty(\mathbb{R})\right)} \leq \sqrt{\binom{r+1/2}{r}\left(1-\frac{2c\gamma}{n\,(\gamma-c)}\right)^{-1/2}}$$

which completes the proof. □

4.3. The space $L_c^1(\mathbb{R})$

Now we consider the case $p = 1$.

Theorem 4.5. *Let $f \in L_c^1(\mathbb{R})$. Then, for all integers $n > 2c$ and each real $\gamma > c(1-2c/n)^{-1}$, it holds*

$$\left\| W_n^{[r]} \right\|_{\left(L_c^1(\mathbb{R}), L_\gamma^\infty(\mathbb{R})\right)} \le \tau_r \sqrt{\frac{n}{\pi}}$$

and

$$\left\| W_n^{[r]} \right\|_{\left(L_c^1(\mathbb{R}), L_\gamma^1(\mathbb{R})\right)} \le \tau_r \sqrt{\frac{n-2c}{(1-2c/n)\gamma - c}},$$

where

$$\tau_r := \sum_{j=0}^{r} \frac{1}{2^j} \binom{2j}{j}^{1/2}.$$

Proof. Let $n > 2c$. For $f \in L_c^1(\mathbb{R})$, we obtain

$$\left| \left(W_n^{[r]} f \right)(x) \right| \le K_{n,c}(x) \left\| f \right\|_{L_c^1(\mathbb{R})}$$

with

$$\begin{aligned} K_{n,c}(x) &= \sqrt{\frac{n}{\pi}} \sup_{t\in\mathbb{R}} \left| \tilde{H}_{2r}\left(\sqrt{n}(x-t)\right) \right| \exp\left(-n(t-x)^2\right) w_{-c}(t) \\ &\le \sqrt{\frac{n}{\pi}} \sup_{u\in\mathbb{R}} \left| \tilde{H}_{2r}(u) \right| e^{-u^2/2} \cdot \exp\left(\sup_{t\in\mathbb{R}} \left(ct^2 - \frac{n}{2}(t-x)^2 \right) \right). \end{aligned}$$

The well-known estimate $|H_{2r}(x)| \le 2^r \sqrt{(2r)!} e^{x^2/2}$ (see, e.g., [9, Subsection 1.5.1, p. 31]) implies the estimate

$$\left| \tilde{H}_{2r}(u) \right| e^{-u^2/2} \le \sum_{j=0}^{r} \frac{1}{2^j} \binom{2j}{j}^{1/2} = \tau_r$$

which was already remarked by Sablonnière [11, Page 42]. Elementary calculus shows that $\left(ct^2 - \frac{n}{2}(t-x)^2\right)$ attains its maximum at $t = x(1-2c/n)^{-1}$. Therefore,

$$K_{n,c}(x) \le \sqrt{\frac{n}{\pi}} \tau_r \exp\left(\frac{cx^2}{1-2c/n} \right).$$

Hence, for $n > 2c$,

$$\left\| W_n^{[r]} \right\|_{\left(L_c^1(\mathbb{R}), L_\gamma^\infty(\mathbb{R})\right)} \le \sqrt{\frac{n}{\pi}} \tau_r \sup_{x\in\mathbb{R}} \exp\left(-\left(\gamma - \frac{c}{1-2c/n} \right) x^2 \right) = \sqrt{\frac{n}{\pi}} \tau_r$$

and

$$\left\| W_n^{[r]} \right\|_{\left(L_c^1(\mathbb{R}), L_\gamma^1(\mathbb{R})\right)} \le \sqrt{\frac{n}{\pi}} \tau_r \int_{-\infty}^{\infty} \exp\left(-\left(\gamma - \frac{c}{1-2c/n} \right) x^2 \right) dx$$

which implies the assertions. □

4.4. The spaces $L_c^p(\mathbb{R})$ with $p>1$

Now we turn to the case $p>1$. As usual let q denote the conjugate number of p satisfying $1/p+1/q=1$.

Theorem 4.6. *Let $p>1$, $q=p/(p-1)$, and $f\in L_c^p(\mathbb{R})$. Then, for $\gamma>c/(p-1)$ and sufficiently large integers n, it holds*

$$\begin{aligned}\left\|W_n^{[r]}\right\|_{(L_c^p(\mathbb{R}),L_\gamma^q(\mathbb{R}))} \leq\ & \left(\frac{n}{\pi}\right)^{\frac{1}{2p}}\left(\frac{p}{p\gamma-qc}\right)^{\frac{1}{2q}}\\ &\times\left(\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|^q\exp\left(-qt^2\right)\exp\left(\frac{cq\gamma}{(p\gamma-qc)\,n}t^2\right)dt\right)^{1/q}.\end{aligned}$$

Remark 4.7. In the special case $r=0$ (note that $\tilde{H}_0(t)=1$) one can explicitly calculate the integral

$$\int_{-\infty}^{\infty}\exp\left(-qt^2\right)\exp\left(\frac{c\gamma q}{(p\gamma-cq)\,n}t^2\right)dt=\sqrt{\frac{\pi}{q-\frac{c\gamma q}{(p\gamma-cq)n}}}$$

which tends to $\sqrt{\pi/q}$ as $n\to\infty$. Hence, for the Gauß–Weierstrass operators W_n holds

$$\|W_n\|_{(L_c^p(\mathbb{R}),L_\gamma^q(\mathbb{R}))}\leq n^{\frac{1}{2p}}\pi^{\frac{1}{2}-\frac{1}{p}}\left(\frac{p-1}{p\gamma-cq-\frac{c\gamma}{n}}\right)^{\frac{1}{2q}}.$$

Proof. We estimate

$$\left(W_n^{[r]}f\right)(x)=\sqrt{\frac{n}{\pi}}\int_{-\infty}^{\infty}\tilde{H}_{2r}\left(\sqrt{n}\,(x-t)\right)e^{-n(t-x)^2}w_{-c/p}(t)\cdot w_{c/p}(t)\,f(t)\,dt$$

by Hölder's inequality (note that $\frac{1}{2}-\frac{1}{2q}=\frac{1}{2p}$):

$$\begin{aligned}\left|\left(W_n^{[r]}f\right)(x)\right|&\leq\sqrt{\frac{n}{\pi}}\left(\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}\left(\sqrt{n}\,(x-t)\right)\right|^q e^{-nq(t-x)^2}w_{-c/p}^q(t)\,dt\right)^{1/q}\\ &\quad\times\left(\int_{-\infty}^{\infty}w_{c/p}^p(t)\,|f(t)|^p\,dt\right)^{1/p}\\ &=\frac{n^{\frac{1}{2p}}}{\sqrt{\pi}}\left(\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|^q\exp\left(-qt^2\right)w_{-cq/p}\left(x-\frac{t}{\sqrt{n}}\right)dt\right)^{1/q}\|f\|_{L_c^p(\mathbb{R})}\\ &=:C(n,r,p,x)\cdot\|f\|_{L_c^p(\mathbb{R})}\end{aligned}$$

Then, for $\gamma>c/(p-1)$, we have

$$\left\|W_n^{[r]}f\right\|_{L_\gamma^q(\mathbb{R})}\leq\|C(n,r,p,\cdot)\|_{L_\gamma^q(\mathbb{R})}\|f\|_{L_c^p(\mathbb{R})}$$

with

$$\begin{aligned}
&\|C(n,r,p,\cdot)\|_{L^q_\gamma(\mathbb{R})}\\
&= \frac{n^{\frac{1}{2p}}}{\sqrt{\pi}}\left(\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|^q \exp\left(-qt^2\right) w_{-cq/p}\left(x-\frac{t}{\sqrt{n}}\right) dt\right) w_\gamma(x)\,dx\right)^{1/q}\\
&= \frac{n^{\frac{1}{2p}}}{\sqrt{\pi}}\left(\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} w_{-cq/p}\left(x-\frac{t}{\sqrt{n}}\right) w_\gamma(x)\ dx\right)\left|\tilde{H}_{2r}(t)\right|^q \exp\left(-qt^2\right) dt\right)^{1/q}.
\end{aligned}$$

A short calculation yields

$$\begin{aligned}
&\int_{-\infty}^{\infty} w_{-cq/p}\left(x-\frac{t}{\sqrt{n}}\right)\cdot w_\gamma(x)\ dx\\
&= \int_{-\infty}^{\infty}\exp\left(c\frac{q}{p}\left(x-\frac{t}{\sqrt{n}}\right)^2-\gamma x^2\right)\ dx\\
&= \sqrt{\frac{\pi p}{p\gamma - cq}}\exp\left(\frac{cq\gamma}{(p\gamma-qc)\,n}t^2\right).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\|C(n,r,p,\cdot)\|_{L^q_\gamma(\mathbb{R})}\\
&= \frac{n^{\frac{1}{2p}}}{\sqrt{\pi}}\left(\sqrt{\frac{\pi p}{p\gamma-qc}}\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|^q \exp\left(-qt^2\right)\exp\left(\frac{cq\gamma}{(p\gamma-qc)\,n}t^2\right) dt\right)^{1/q}.
\end{aligned}$$

The estimate now follows by noting that $\left(\sqrt{\pi}\right)^{-1+\frac{1}{q}} = \pi^{-\frac{1}{2p}}$. □

4.5. The special case $p = q = 2$

In the special case $p = q = 2$, we obtain, for $\gamma > c$ and sufficiently large integers n,

$$\|C(n,r,2,\cdot)\|_{L^2_\gamma(\mathbb{R})} = \left(\frac{n}{\pi(\gamma-c)}\right)^{\frac{1}{4}}\left(\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|^2 \exp\left(-\left(2-\frac{c\gamma}{(\gamma-c)\,n}\right)t^2\right) dt\right)^{1/2}.$$

Note that $\left|\tilde{H}_{2r}(t)\right|^2 = \tilde{H}^2_{2r}(t)$. Therefore, one can explicitly evaluate the integrals for each integer value r.

In particular, for $r = 0$, we obtain the explicit expression

$$\begin{aligned}
\|C(n,r,2,\cdot)\|_{L^2_\gamma(\mathbb{R})} &= \left(\frac{n}{\pi(\gamma-c)}\right)^{\frac{1}{4}}\left(\sqrt{\frac{\pi}{2-\frac{c\gamma}{(\gamma-c)n}}}\right)^{1/2}\\
&= \sqrt[4]{\frac{n}{2(\gamma-c)-c\gamma/n}}.
\end{aligned}$$

Hence, Theorem 4.6 reduces to the next corollary.

Corollary 4.8. *Let $f \in L_c^2(\mathbb{R})$. Then, for $\gamma > c$ and sufficiently large integers n, it holds*

$$\left\|W_n^{[r]}\right\|_{\left(L_c^2(\mathbb{R}),L_\gamma^2(\mathbb{R})\right)} \leq \left(\frac{n}{(\gamma-c)\pi}\right)^{\frac{1}{4}}\left(\int_{-\infty}^{\infty}\left|\tilde{H}_{2r}(t)\right|^2\exp\left(-\left(2-\frac{c\gamma}{(\gamma-c)n}\right)t^2\right)dt\right)^{1/2}.$$

In particular, for $r=0$, the operator norm of the classical Gauß–Weierstraß operators satisfies the estimate

$$\|W_n\|_{\left(L_c^2(\mathbb{R}),L_\gamma^2(\mathbb{R})\right)} \leq \sqrt[4]{\frac{n}{2(\gamma-c)-c\gamma/n}}.$$

Concluding remark. If we allow γ to depend on n, we can choose $\gamma = \frac{n}{n-c}c \equiv \gamma(n)$, such that $\exp\left(-\left(2-\frac{c\gamma}{n(\gamma-c)}\right)t^2\right) = \exp\left(-t^2\right)$, then

$$\left\|W_n^{[r]}\right\|_{\left(L_c^2(\mathbb{R}),L_{\gamma(n)}^2(\mathbb{R})\right)} \leq \left(\frac{n(n-c)}{c^2\pi}\right)^{\frac{1}{4}}\left\|\tilde{H}_{2r}\right\|_{L_1^2(\mathbb{R})}.$$

Finally, by Eq. (2.4) of Lemma 2.3,

$$\left\|W_n^{[r]}\right\|_{\left(L_c^2(\mathbb{R}),L_{\gamma(n)}^2(\mathbb{R})\right)} \leq \sqrt[4]{\frac{n(n-c)}{c^2}}\binom{r+1/2}{r}^{1/2}.$$

By estimate (2.5), it follows that

$$\left\|W_n^{[r]}\right\|_{\left(L_c^2(\mathbb{R}),L_{\gamma(n)}^2(\mathbb{R})\right)} \leq \sqrt[4]{4\frac{n(n-c)}{c^2\pi}(r+1)}.$$

Note that $\gamma(n)$ tends to c (from above) as $n \to \infty$. For large values of n both norms $L_c^2(\mathbb{R})$ and $L_{\gamma(n)}^2(\mathbb{R})$ are close together.

References

[1] Abel, U., *Asymptotic expansions for Favard operators and their left quasi-interpolants*, Stud. Univ. Babeş-Bolyai Math., **56**(2011), 199–206.

[2] Abel, U., Agratini, O., Păltănea, R., *A complete asymptotic expansion for the quasi-interpolants of Gauß–Weierstraß operators*, Mediterr. J. Math., **15**(2018), Online First.

[3] Abel, U., Butzer, P.L., *Complete asymptotic expansion for generalized Favard operators*, Constr. Approx., **35**(2012), 73–88.

[4] Abel, U., Ivan, M., *Simultaneous approximation by Altomare operators*, Proceedings of the 6th international conference on functional analysis and approximation theory, Acquafredda di Maratea (Potenza), Italy, Sept. 24–30, 2009, Palermo: Circolo Matemàtico di Palermo, Rend. Circ. Mat. Palermo, Serie II, Suppl., **82**(2010), 177–193.

[5] Abramowitz, M., Stegun, A., *Handbook of Mathematical Functions*, Appl. Math. Ser. 55, National Bureau of Standards, Washington D.C., 1972.

[6] Altomare, F., Milella, S., *Integral-type operators on continuous function spaces on the real line*, J. Approx. Theory **152**(2008), 107–124.

[7] Altomare, F., Campiti, M., *Korovkin-type Approximation Theory and its Applications*, de Gruyter Studies in Mathematics 17, W. de Gruyter, Berlin, New York, 1994.

[8] Butzer, P.L., Nessel, R.J., *Fourier Analysis and Approximation*, Birkhäuser Verlag, Basel, 1971.

[9] Gautschi, W., *Orthogonal Polynomials*, Oxford University Press, 2004.

[10] Gould, H.W., *Combinatorial Identities*, Morgantown Print & Bind., Morgantown, WV, 1972.

[11] Sablonnière, P., *Weierstrass quasi-interpolants*, J. Approx. Theory, **180**(2014), 32–48.

[12] Watson, G.N., *A note on gamma functions*, Proc. Edinb. Math. Soc., **2**(11)(1958/59); Edinburgh Math. Notes, **42**(1959), 7–9.

Ulrich Abel
Technische Hochschule Mittelhessen
Department Mathematik, Naturwissenschaften und Datenverarbeitung
Wilhelm-Leuschner-Straße 13, 61169 Friedberg, Germany
e-mail: `Ulrich.Abel@mnd.thm.de`

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 239–252
DOI: 10.24193/subbmath.2019.2.09

A generalization of Bernstein-Durrmeyer operators on hypercubes by means of an arbitrary measure

Mirella Cappelletti Montano and Vita Leonessa

Abstract. In this paper we introduce and study a sequence of Bernstein-Durrmeyer type operators $(M_{n,\mu})_{n\geq 1}$, acting on spaces of continuous or integrable functions on the multi-dimensional hypercube Q_d of $\mathbf{R}^d$ $(d \geq 1)$, defined by means of an arbitrary measure μ. We investigate their approximation properties both in the space of all continuous functions and in L^p-spaces with respect to μ, also furnishing some estimates of the rate of convergence. Further, we prove an asymptotic formula for the $M_{n,\mu}$'s. The paper ends with a concrete example.

Mathematics Subject Classification (2010): 41A36, 41A63, 41A10.

Keywords: Bernstein operator, Bernstein-Durrmeyer operator, approximation process, asymptotic formula.

1. Introduction

Bernstein-Durrmeyer operators were introduced, independently, by Durrmeyer ([15]) and Lupaş ([18]) in their respective dissertations, as a modification of the classical Bernstein operators acting on spaces of integrable functions. More precisely, they are defined by setting

$$M_n(f)(x) := (n+1)\sum_{k=0}^{n}\left(\int_0^1 f(u)p_{n,k}(u)\,du\right)p_{n,k}(x),$$

for every $n \geq 1$, $x \in [0,1]$ and $f : [0,1] \to \mathbf{R}$ such that $fp_{n,k} \in L^1([0,1])$ for every $k = 0,\dots,n$, where $p_{n,k}(x) := \binom{n}{k}x^k(1-x)^{n-k}$.

These operators were intensively studied by Derriennic ([12]), and during the years they have been subject to many generalizations. The most renowned one is due

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

to Paltanea ([19], see also [11, 21, 22]), who replaced the weighted measure $p_{n,k}\lambda_1$ (λ_1 being the Lebesgue-Borel measure on [0, 1]) with the absolutely continuous Borel measure with respect to λ_1 with density the normalized Jacobi weights

$$w_{a,b}(x) := \frac{x^a(1-x)^b}{\int_0^1 y^a(1-y)^b\,dy} \quad a > -1, b > -1, x \in (0,1).$$

The Bernstein-Durrmeyer operators with Jacobi weights have been matter of many investigations, in the context of the interval $[0,1]$, in the multi-dimensional framework of simplices ([13, 1]) and, more recently, for hypercubes ([3]). In particular, in [1, 3] the connection between such operators and the study of the so-called Fleming-Viot differential operators is investigated.

A further step in the possible generalizations of the Bernstein-Durrmeyer operators, which has significant applications in learning theory, consists in replacing the Lebesgue measure with an arbitrary regular Borel measure. This generalization was briefly mentioned by Berens and Xu in [11] in the context of the interval $[0,1]$, and then intensively studied in [8, 6, 7, 17, 9] for the multi-dimensional simplex.

Inspired by these last works, in this paper we introduce and study a sequence of Bernstein-Durrmeyer type operators $(M_{n,\mu})_{n\geq 1}$ with respect to an arbitrary measure μ (see (3.1)-(3.3)), acting both on spaces of continuous and integrable functions on the hypercube $Q_d := [0,1]^d$ of $\mathbf{R}^d$ $(d \geq 1)$.

First, we prove a necessary and sufficient condition, which involves only properties of the measure μ, in order that the sequence $(M_{n,\mu})_{n\geq 1}$ is an approximation process with respect to the uniform norm.

Moreover, following the reasoning in [17] (see also [9]), we show that $(M_{n,\mu})_{n\geq 1}$ is an approximation process in $L^p(Q_d,\mu)$ $(1 \leq p < +\infty)$ for any Borel measure μ for which the $M_{n,\mu}$'s are well defined; this entails that the space $L^p(Q_d,\mu)$ is the most natural environment in which studying these operators.

Further, we produce, under suitable conditions, an asymptotic formula for the sequence $(M_{n,\mu})_{n\geq 1}$, that involves a second order differential operator.

Finally, a concrete example of Bernstein-Durrmeyer operators on Q_d is illustrated.

2. Notation and preliminary results

Let us start by fixing some notation. Let X be a compact Hausdorff space. As usual, we denote by $C(X)$ the space of all continuous real-valued functions on X. $C(X)$ will be endowed with the uniform norm $\|\cdot\|_\infty$, with respect to which it is a Banach space.

A linear operator T on $C(X)$ is called a Markov operator on $C(X)$ if it is positive and $T(\mathbf{1}) = \mathbf{1}$, where $\mathbf{1}$ indicates the constant function of constant value 1 on X.

If B_X is the σ-algebra of all Borel subsets of X, the symbol $M^+(X)$ (resp., $M_b^+(X)$) stands for the cone of all regular Borel measures on X (resp., the cone of all bounded Borel measures on X).

For a measure $\mu \in M^+(X)$, we denote by $\mathrm{supp}(\mu)$ the support of μ, i.e., the complement of the largest open subset of X on which μ is zero. We recall that a measure μ on X is said to be strictly positive on X if $\mu(X \cap A) > 0$ for every open subset A such that $A \cap X \neq \emptyset$. We remark that, on the account of [20, Prop. 13, p. 408], μ is a strictly positive measure if and only if $\mathrm{supp}(\mu) = X$.

Let $\mu \in M^+(X)$ and $1 \leq p < +\infty$. As usual, $L^p(X,\mu)$ is the space of all (the equivalence classes of) Borel measurable, real-valued functions on X which are μ-integrable in the p^{th} power. The space $L^p(X,\mu)$ is endowed with the norm

$$\|f\|_p := \left(\int_X |f|^p \, d\mu\right)^{1/p} \qquad (f \in L^p(X,\mu)).$$

Now, let $\gamma = (\gamma_1, \dots, \gamma_d) \in \mathbf{R}^d$, $d \geq 1$. If $x = (x_1, \dots, x_d) \in \mathbf{R}^d$, $x_i > 0$ for every $i = 1, \dots, d$, we set

$$x^\gamma := \prod_{i=1}^d x_i^{\gamma_i}.$$

If $\gamma_i \geq 0$ and $x_i \geq 0$ for each $i = 1, \dots, d$, then x^γ is similarly defined as above assuming $0^0 := 1$.

If $x = (x_1, \dots, x_d)$, $y = (y_1, \dots, y_d) \in \mathbf{R}^d$, then we write $x \leq y$ whenever $x_i \leq y_i$ for every $i = 1, \dots, d$.

Let $j = (j_1, \dots, j_d)$, $k = (k_1, \dots, k_d) \in \mathbf{N}^d$ be two multi-indices such that $k \leq j$; we define

$$\binom{j}{k} := \prod_{i=1}^d \binom{j_i}{k_i}.$$

We also set $0_d := (0, \dots, 0)$ and, for every $n \geq 1$, $n_d := (n, \dots, n)$.

All the results of this paper concern the case where X is the d-dimensional hypercube $Q_d := [0,1]^d$, $d \geq 1$.

First, for each $i = 1, \dots, d$, the symbol pr_i stands for the i-th coordinate function on Q_d, which is defined by setting $pr_i(x) = x_i$ for every $x \in Q_d$.

Moreover, let us consider the partition $P = \{\xi_0, \xi_1, \xi_2\}$ of the unit interval $[0,1]$ such that $\xi_0 = 0, \xi_1 = \frac{1}{2}$ and $\xi_2 = 1$. Then, P yields a partition of Q_d composed by closed sub-hypercubes of the form

$$Q_{d,j} := [\xi_{j_1}, \xi_{j_1+1}] \times \cdots \times [\xi_{j_d}, \xi_{j_d+1}] \tag{2.1}$$

where $j = (j_1, \dots, j_d)$ is such that $j_i = 0$ or $j_i = 1$ for every $i = 1, \dots, d$ (briefly, $j \in \{0,1\}^d$).

For $a \in Q_d$ and $r > 0$, we define the open d-dimensional hypercube

$$I_d(a;r) := (a_1 - r, a_1 + r) \times \cdots \times (a_d - r, a_d + r), \tag{2.2}$$

and the closed d-dimensional hyperrectangle

$$J_d(a;r) := [a_1, a_1 + r] \times \cdots \times [a_d, a_d + r]. \tag{2.3}$$

Remark 2.1. Observe that, if a belongs to some $Q_{d,j}$ and if $r \leq \frac{1}{2}$, then $J_d(a;r) \subset Q_d$. Indeed, $a \in Q_{d,j}$ means that $\xi_{j_i} \leq a_i \leq \xi_{j_i+1}$ for every $i = 1, \dots, d$. Hence, if $x \in J_d(a;r)$, for any $i = 1, \dots, d$, $x_i \geq 0$, since $j_i \in \{0, \frac{1}{2}, 1\}$. Now, fix $i = 1, \dots, d$ and suppose that $j_i = 0$; in this case $0 \leq a_i \leq \frac{1}{2}$ from which it easily follows $x_i \leq 1$.

Otherwise, if $j_i = 1$, $\frac{1}{2} \leq a_i \leq 1$, that is $0 \leq 1 - a_i \leq \frac{1}{2}$, therefore $a_i + r \leq 1$ is equivalent to the true inequality $r \leq 1 - a_i \leq \frac{1}{2}$, and again $x_i \leq 1$.

The following lemma will play an important role in the next section. It deals with the case of the unit interval $[0, 1]$.

Lemma 2.2. *Let $0 < r \leq \frac{1}{2}$. Then, there exists a positive constant $K = K(r)$, depending only on r, such that*

$$\frac{\max_{x\in[0,1]\setminus(a-r,a+r)} x^a(1-x)^{1-a}}{\min_{x\in[a,a+r^2]} x^a(1-x)^{1-a}} < K < 1\,.$$

Proof. The proof can be found in [6, Lemma 3 and Lemma 4] (see also [7, p. 738]). In particular it is worth noticing that

$$\max_{x\in[0,1]\setminus(a-r,a+r)} x^a(1-x)^{1-a} = \begin{cases} (a+r)^a(1-a-r)^{1-a} & \text{if } 0 \leq a \leq \frac{1}{2}\,, \\ (a-r)^a(1-a+r)^{1-a} & \text{if } \frac{1}{2} \leq a \leq 1\,, \end{cases}$$

and

$$\min_{x\in[a,a+r^2]} x^a(1-x)^{1-a} = (a+r^2)^a(1-a-r^2)^{1-a}\,.$$

□

Finally, coming back to the case $d \geq 1$, we state the following result.

Lemma 2.3. *Consider $\mu \in M_b^+(Q_d)$ such that $supp(\mu) = Q_d$. Then, for every $r > 0$ and $j \in \{0,1\}^d$,*

$$\inf_{a\in Q_{d,j}} \mu(J_d(a;r)) > 0,$$

where $Q_{d,j}$ and $J_d(a;r)$ are defined, respectively, by (2.1) and (2.3).

Proof. We begin with supposing $r \leq 1/2$. In this case, from Remark 2.1 it follows that the interior $\overset{\circ}{J}_d(a;r)$ of $J_d(a;r)$ is contained into Q_d, so that $\overset{\circ}{J}_d(a;r) = \overset{\circ}{J}_d(a;r) \cap Q_d$. Since μ is strictly positive on Q_d, we have that

$$0 < \mu(\overset{\circ}{J}_d(a;r) \cap Q_d) = \mu(\overset{\circ}{J}_d(a;r)) \leq \mu(J_d(a;r)).$$

The case $r > 1/2$ is an easy consequence of the fact that $\mu(J_d(a;r)) \geq \mu(J_d(a;\frac{1}{2})) > 0$, where the last inequality is true for the first part of the proof. □

On account of Lemma 2.3, we set

$$C(\mu,r) := \min_{j\in\{0,1\}^d} \inf_{a\in Q_{d,j}} \mu(J_d(a;r)) > 0\,. \tag{2.4}$$

3. Bernstein-Durrmeyer operators on Q_d with respect to arbitrary measures

Let $\mu \in M_b^+(Q_d)$ be a nonnegative Borel measure on Q_d satisfying the assumption

$$\operatorname{supp}(\mu) \setminus \partial Q_d \neq \emptyset. \tag{3.1}$$

For every $n \geq 1$, let us consider the operator $M_{n,\mu} : L^1(Q_d, \mu) \to C(Q_d)$ defined by setting, for every $f \in L^1(Q_d, \mu)$ and $x \in Q_d$,

$$M_{n,\mu}(f)(x) := \sum_{\substack{h \in \mathbf{N}^d \\ 0_d \leq h \leq n_d}} \omega_{n_d,h}(f, \mu) \binom{n_d}{h} x^h (1_d - x)^{n_d - h}, \tag{3.2}$$

where, for every $n \geq 1$ and $h = (h_1, \dots h_d) \in \mathbf{N}^d$, $0_d \leq h \leq n_d$,

$$\omega_{n_d,h}(f, \mu) := \frac{1}{\int_{Q_d} y^h (1_d - y)^{n_d - h} \, d\mu(y)} \int_{Q_d} y^h (1_d - y)^{n_d - h} f(y) \, d\mu(y). \tag{3.3}$$

We remark that assumption (3.1) guarantees that, for every $n \geq 1$ and $h \in \mathbf{N}^d$, $0_d \leq h \leq n_d$,

$$\int_{Q_d} y^h (1_d - y)^{n_d - h} \, d\mu(y) > 0.$$

Clearly, the operators $M_{n,\mu}$ are linear, positive, and $M_{n,\mu}(\mathbf{1}) = \mathbf{1}$, so that the restriction of each $M_{n,\mu}$ to $C(Q_d)$ is a Markov operator on $C(Q_d)$ with unitary norm. Moreover, for any $f \in L^1(Q_d, \mu)$ and $n \geq 1$, $M_{n,\mu}(f)$ is a polynomial of total degree at most n.

In order to discuss the convergence of the sequence $(M_{n,\mu})_{n \geq 1}$ both on $C(Q_d)$ and $L^p(Q_d, \mu)$ ($p \geq 1$), first of all, we recall the definition of the classical Bernstein operators on Q_d (see [16] and the references therein). They are defined by setting, for any $n \geq 1$, $f \in C(Q_d)$, and $x \in Q_d$,

$$B_n(f)(x) = \sum_{\substack{h \in \mathbf{N}^d \\ 0_d \leq h \leq n_d}} f\left(\frac{h}{n}\right) \binom{n_d}{h} x^h (1_d - x)^{n_d - h}. \tag{3.4}$$

The sequence $(B_n)_{n \geq 1}$ is an approximation process on $C(Q_d)$, i.e., for any $f \in C(Q_d)$

$$\lim_{n \to \infty} B_n(f) = f \quad \text{uniformly on } Q_d. \tag{3.5}$$

Observe in particular that $B_n(\mathbf{1}) = \mathbf{1}$, or equivalently, that

$$\sum_{\substack{h \in \mathbf{N}^d \\ 0_d \leq h \leq n_d}} \binom{n_d}{h} x^h (1_d - x)^{n_d - h} = 1 \quad \text{for every} \quad x \in Q_d. \tag{3.6}$$

Finally, we have that, for every $n \geq 1$ and $i = 1, \dots, d$,

$$B_n(pr_i) = pr_i \quad \text{and} \quad B_n(pr_i^2) = \frac{1}{n} pr_i + \frac{n-1}{n} pr_i^2. \tag{3.7}$$

3.1. Approximation properties on $C(Q_d)$

In what follows we study the convergence properties of the sequence $(M_{n,\mu}(f))_{n\geq 1}$ on the space $C(Q_d)$.

Theorem 3.1. *The following statements are equivalent:*

(i) *For every $f \in C(Q_d)$,*

$$\lim_{n\to\infty} M_{n,\mu}(f) = f \quad \text{uniformly on } \; Q_d\,. \tag{3.8}$$

(ii) *$supp(\mu) = Q_d$.*

Proof. $(i) \Longrightarrow (ii)$. Suppose that there exists a nonempty open set $A \subset Q_d$ such that $\mu(A) = 0$. Then, for every $f \in C(Q_d)$, $f = 0$ on $Q_d \setminus A$, on account of (3.5) for operators B_n, we have

$$\int_{Q_d} y^h(1_d - y)^{n_d-h} f(y)\, d\mu(y) = \left\{ \int_A + \int_{Q_d\setminus A} \right\} y^h(1_d - y)^{n_d-h} f(y)\, d\mu(y) = 0\,.$$

Therefore, $M_{n,\mu}(f) = 0$ and this leads to a contradiction because of (i).
$(ii) \Longrightarrow (i)$. For $n \geq 1$, $f \in C(Q_d)$ and $x \in Q_d$, we have

$$\begin{aligned} |M_{n,\mu}(f)(x) - f(x)| &\leq |M_{n,\mu}(f)(x) - B_n(f)(x)| + |B_n(f)(x) - f(x)| \\ &\leq \sum_{\substack{h\in\mathbf{N}^d \\ 0_d\leq h\leq n}} \left|\omega_{n_d,h}(f,\mu) - f\left(\frac{h}{n_d}\right)\right| \binom{n_d}{h} x^h(1_d - x)^{n_d-h} + \|B_n(f) - f\|_\infty \\ &\leq \max_{\substack{h\in\mathbf{N}^d \\ 0_d\leq h\leq n}} \left|\omega_{n_d,h}(f,\mu) - f\left(\frac{h}{n_d}\right)\right| + \|B_n(f) - f\|_\infty, \end{aligned}$$

as (3.6) holds true. Thus, keeping (3.5) in mind, in order to get the claim it is sufficient to prove that

$$\lim_{n\to\infty} \max_{\substack{h\in\mathbf{N}^d \\ 0_d\leq h\leq n_d}} \left|\omega_{n_d,h}(f,\mu) - f\left(\frac{h}{n}\right)\right| = 0. \tag{3.9}$$

We begin to observing that, since f is uniformly continuous on Q_d,

$$\begin{aligned} &\text{for a fixed } \varepsilon > 0, \text{ there exists } \tilde{\delta} > 0 \text{ such that, for every} \\ &x, y \in Q_d \text{ with } \|x - y\| < \tilde{\delta}, \text{ then } |f(x) - f(y)| < \varepsilon, \end{aligned} \tag{3.10}$$

where by $\|\cdot\|$ we indicate the l^∞-norm on $\mathbf{R}^d$ defined by setting $\|x\| := \max_{i=1,\dots,d} |x_i|$ $(x \in \mathbf{R}^d)$.

Fix $\delta = \min\{\tilde{\delta}, \frac{1}{2}\}$, $n \geq 1$, $h \in \mathbf{N}^d$, $0_d \leq h \leq n_d$, and $j \in \{0,1\}^d$ such that, the point $\frac{h}{n}$ belongs to the piece $Q_{d,j}$ of Q_d (see (2.1)).

Then

$$\left|\omega_{n_d,h}(f,\mu) - f\left(\frac{h}{n}\right)\right| \leq \frac{\int_{Q_d} \left|f(y) - f\left(\frac{h}{n}\right)\right| y^h(1_d - y)^{n_d-h}\, d\mu(y)}{\int_{Q_d} y^h(1_d - y)^{n_d-h}\, d\mu(y)}\,.$$

Now, rewrite $Q_d = \big(Q_d \cap I_d(\frac{h}{n};\delta)\big) \cup \big(Q_d \setminus I_d(\frac{h}{n};\delta)\big)$ (see (2.2)), and observe that, from (3.10) it follows that $\left|f(y) - f\left(\frac{h}{n}\right)\right| < \varepsilon$ when $y \in Q_d \cap I_d(\frac{h}{n};\delta)$. Hence, setting $M := \|f\|_\infty$, from Remark 2.1, (2.3) and (2.4), we get

$$\left|\omega_{n_d,h}(f,\mu) - f\left(\frac{h}{n}\right)\right| \le \varepsilon + 2M \frac{\int_{Q_d \setminus I_d(\frac{h}{n};\delta)} y^h (1_d - y)^{n_d - h}\, d\mu(y)}{\int_{Q_d} y^h (1_d - y)^{n_d - h}\, d\mu(y)}$$

$$< \varepsilon + 2M \frac{\int_{Q_d \setminus I_d(\frac{h}{n};\delta)} y^h (1_d - y)^{n_d - h}\, d\mu(y)}{\int_{J_d(\frac{h}{n};\delta^2)} y^h (1_d - y)^{n_d - h}\, d\mu(y)}$$

$$< \varepsilon + 2M \frac{\mu(Q_d)}{\mu(J_d(\frac{h}{n};\delta^2))} \frac{\max_{y \in Q_d \setminus I_d(\frac{h}{n};\delta)} y^h (1_d - y)^{n_d - h}}{\min_{y \in J_d(\frac{h}{n};\delta^2)} y^h (1_d - y)^{n_d - h}}$$

$$< \varepsilon + 2M \frac{\mu(Q_d)}{C(\mu,\delta^2)} \frac{\big(\max_{y \in Q_d \setminus I_d(a;\delta)} y^a (1_d - y)^{1_d - a}\big)^n}{\big(\min_{y \in J_d(a;\delta^2)} y^a (1_d - y)^{1_d - a}\big)^n},$$

where in the last inequality $a := \frac{h}{n}$.
Since

$$\max_{y \in Q_d \setminus I_d(a;\delta)} y^a (1_d - y)^{1_d - a} = \prod_{i=1}^d \max_{y_i \in [0,1] \setminus (a_i - \delta, a_i + \delta)} y_i^{a_i} (1 - y_i)^{1 - a_i},$$

and

$$\min_{y \in J_d(a;\delta^2)} y^a (1_d - y)^{1_d - a} = \prod_{i=1}^d \min_{y_i \in [a_i, a_i + \delta^2]} y_i^{a_i} (1 - y_i)^{1 - a_i},$$

by applying Lemma 2.2 we get that

$$\frac{\big(\max_{y \in Q_d \setminus I_d(a;\delta)} y^a (1_d - y)^{1_d - a}\big)^n}{\big(\min_{y \in J_d(a;\delta^2)} y^a (1_d - y)^{1_d - a}\big)^n} < K(r)^n \to 0 \quad \text{as } n \to \infty\,.$$

Therefore, there exists $n_\varepsilon \in \mathbf{N}$ such that, for every $n \ge n_\varepsilon$, $(K(r))^n < \varepsilon \frac{C(\mu,\delta^2)}{2M\mu(Q_d)}$. Accordingly, for every $n \ge n_\epsilon$ and $h \in \mathbf{N}^d$, $0_d \le h \le n_d$,

$$\left|\omega_{n_d,h}(f,\mu) - f\left(\frac{h}{n}\right)\right| < 2\varepsilon$$

and this completes the proof of (3.9). □

At the end of Section 4 we present, under suitable assumptions, an estimate of the convergence in (3.8).

3.2. Approximation properties on $L^p(Q_d,\mu)$

In this section we are interested in the convergence properties of the Bernstein-Durrmeyer operators $M_{n,\mu}$ defined by (3.1)-(3.3) in the space $L^p(Q_d,\mu)$ $(1 \le p < +\infty)$.

First note that, if $n \ge 1$ and $f \in L^1(Q_d,\mu)$, we get

$$\int_{Q_d} M_{n,\mu}(f)\, d\mu = \int_{Q_d} f\, d\mu\,. \tag{3.11}$$

Indeed, keeping (3.6) in mind,

$$\int_{Q_d} M_{n,\mu}(f)\,d\mu = \sum_{\substack{h\in\mathbf{N}^d\\ 0_d\le h\le n_d}} \omega_{n_d,h}(f,\mu)\binom{n_d}{h}\int_{Q_d} x^h(1_d-x)^{n_d-h}\,d\mu(x)$$
$$= \sum_{\substack{h\in\mathbf{N}^d\\ 0_d\le h\le n_d}} \binom{n_d}{h}\int_{Q_d} y^h(1_d-y)^{n_d-h}f(y)\,d\mu(y)$$
$$= \int_{Q_d}\sum_{\substack{h\in\mathbf{N}^d\\ 0_d\le h\le n_d}} \left[\binom{n_d}{h}y^h(1_d-y)^{n_d-h}\right]f(y)\,d\mu(y) = \int_{Q_d} f(y)\,d\mu(y)\,.$$

Moreover, by using the convexity of the function $|t|^p (t\in\mathbf{R})$ and (3.6), for any $n\ge 1$ and $f\in L^p(Q_d,\mu)$, we obtain

$$|M_{n,\mu}(f)(x)|^p \le \sum_{\substack{h\in\mathbf{N}^d\\ 0_d\le h\le n_d}} (\omega_{n_d,h}(|f|,\mu))^p\binom{n_d}{h}x^h(1_d-x)^{n_d-h}.$$

Now, by applying the integral Jensen inequality (see, e.g. [5]) to the probability measure ρ on Q_d which is absolutely continuous with respect to μ with density the weight function $w_{n_d,h}$ defined on the interior $\overset{\circ}{Q}_d$ of Q_d as

$$w_{n_d,h}(x) = \frac{x^h(1_d-x)^{n_d-h}}{\int_{Q_d} y^h(1_d-y)^{n_d-h}\,d\mu(y)} \quad (x\in\overset{\circ}{Q}_d),$$

we get (see (3.3))

$$(\omega_{n_d,h}(|f|,\mu))^p = \left(\int_{Q_d}|f(y)|\,d\rho(y)\right)^p \le \int_{Q_d}|f(y)|^p\,d\rho(y) = \omega_{n_d,h}(|f|^p,\mu)\,,$$

and hence, $|M_{n,\mu}(f)|^p \le M_{n,\mu}(|f|^p)$. Therefore, by integrating with respect to μ over Q_d, we gain

$$\int_{Q_d}|M_{n,\mu}(f)|^p\,d\mu \le \int_{Q_d} M_{n,\mu}(|f|^p)\,d\mu = \int_{Q_d}|f|^p\,d\mu\,, \tag{3.12}$$

where in the last equality we have used (3.11). Inequality (3.12) means that $M_{n,\mu}$ maps $L^p(Q_d,\mu)$ into itself and, in particular that each restriction $M_n|_{L^p(Q_d,\mu)}$ coincides with the extension of $M_n|_{C(Q_d)}$ to $L^p(Q_d,\mu)$.

Thanks to these considerations, we are able to get the following result.

Proposition 3.2. *Consider $\mu\in M_b^+(Q_d)$ satisfying (3.1). Then, for every $n\ge 1$ and $1\le p<+\infty$, for the operator $M_{n,\mu}: L^p(Q_d,\mu)\to L^p(Q_d,\mu)$ the following inequality holds:*

$$\|M_{n,\mu}(f)\|_{L^p} \le \|f\|_{L^p} \quad (f\in L^p(Q_d,\mu))\,. \tag{3.13}$$

Property (ii) on μ for the convergence in $C(Q_d)$ seems to be too strong for spaces of integrable functions, and in fact, following the idea used by Li in the case of simplices (see [17]; see also [9]), we prove that $(M_{n,\mu})_{n\ge 1}$ constitutes a positive approximation process in $L^p(Q_d,\mu)$ requiring only that μ satisfies (3.1).

We need some additional notation. Consider the space $C^1(Q_d)$ of all real-valued continuous functions on Q_d which are continuously differentiable on $\overset{\circ}{Q}_d$ and whose partial derivatives can be continuously extended to Q_d. We shall continue to denote by $\frac{\partial}{\partial x_i}$ the continuous extensions to Q_d of $\frac{\partial}{\partial x_i}$. Moreover, the space $C^1(Q_d)$ will be equipped with the seminorm $|||\nabla g||| := \max_{i=1,\dots,d} \|\frac{\partial}{\partial x_i} g\|_\infty$.

Further, let $\mathcal{K}(f,t)_p$ be the $\mathcal{K}$-functional (see, e.g. [14]) defined by

$$\mathcal{K}(f,t)_p = \inf_{g\in C^1(Q_d)} \{\|f-g\|_{L^p} + t|||\nabla g|||\} \quad \text{for } p \geq 1\,, t > 0.$$

In particular, for every $f \in L^p(Q_d, \mu)$, one has

$$\mathcal{K}(f,t)_p \to 0 \quad \text{as } t \to 0\,. \tag{3.14}$$

Theorem 3.3. *Let* $\mu \in M_b^+(Q_d)$ *satisfying* (3.1) *and consider the operators* $M_{n,\mu}$ *defined by* (3.2). *For every* $f \in L^p(Q_d,\mu)$, $1 \leq p < +\infty$ *and* $n \geq 1$,

$$\|M_{n,\mu}(f) - f\|_p \leq 2\mathcal{K}(f, C_p(\mu(Q_d))^{\frac{1}{p}} n^{-\frac{1}{2}})_p \tag{3.15}$$

where C_p *is a constant depending only on* p *and* d. *In particular,*

$$\lim_{n\to\infty} M_{n,\mu}(f) = f \quad \textit{in } L^p(Q_d,\mu)\,. \tag{3.16}$$

Proof. By an inspection of Theorems 2.1 and 2.2 in [17], one notes that the arguments used there work also for hypercubes. In fact, first we have that, for any $n \geq 1$, $f \in L^p(Q_d,\mu)$ and $p \geq 1$, the following estimate holds:

$$\|M_{n,\mu}(f) - f\|_p \leq 2\mathcal{K}(f, \Delta_{n,p}/2)_p\,, \tag{3.17}$$

where

$$\begin{aligned}\Delta_{n,p} &:= \sum_{i=1}^{d} \left(\int_{Q_d} |M_{n,\mu}(|pr_i - pr_i(x)\mathbf{1}|)(x)|^p \, d\mu(x) \right)^{1/p} \\ &= \sum_{i=1}^{d} \|M_{n,\mu}(|pr_i - pr_i(x)\mathbf{1}|)\|_{L^p}.\end{aligned}$$

The proof of (3.17) runs as in [17, Theorem 2.1] (see also [8, Theorem 4.5]) on account of (3.13) and the well-known equivalence between l^∞-norm and l^1-norm in $\mathbf{R}^d$.

Subsequently, estimates of $\Delta_{n,p}$ similar to those in [9, Theorem 1.1, Lemma 1.1] can be obtained, thanks to (3.6) and, moreover, to the fact that the expressions of $B_n(pr_i)$ and $B_n(pr_i^2)$ in the case of hypercubes (see (3.7)) are the same as those for simplices (see, e.g [17, Lemma 2.1]). In particular, for every $i = 1, \dots, d$,

$$\|M_{n,\mu}(|pr_i - pr_i(x)\mathbf{1}|)\|_{L^p} \leq c_p(\mu(Q_d))^{\frac{1}{p}} n^{-\frac{1}{2}},$$

where c_p is a constant depending only on p. Hence,

$$\|M_{n,\mu}(f) - f\|_p \leq 2\mathcal{K}(f, dc_p(\mu(Q_d))^{\frac{1}{p}} n^{-\frac{1}{2}}/2)_p,$$

which leads to (3.15) and, letting $n \to \infty$, we get (3.16) by virtue of (3.14). $\square$

4. Asymptotic formula for the operators $M_{n,\mu}$

In order to present an asymptotic formula for the operators $M_{n,\mu}$, we need some further notation.

In particular, we denote by $C^2(Q_d)$ the space of all real-valued continuous functions on Q_d which are twice-continuously differentiable on $\overset{\circ}{Q}_d$ and whose partial derivatives up to the order 2 can be continuously extended to Q_d. We shall continue to indicate by $\frac{\partial}{\partial x_i}$ and $\frac{\partial^2}{\partial x_i \partial x_j}$ the continuous extensions to Q_d of $\frac{\partial}{\partial x_i}$ and $\frac{\partial^2}{\partial x_i \partial x_j}$. Moreover, for every $x \in Q_d$, we denote by $\Psi_x, d_x \in C(Q_d)$ the functions defined by

$$\Psi_x(y) := y - x \quad \text{and} \quad d_x(y) := \left(\sum_{i=1}^{d} |y_i - x_i|^2 \right)^{1/2} \quad (y \in Q_d).$$

We notice that, for every $y = (y_1, \dots, y_d) \in Q_d$, and $i = 1, \dots, d$,

$$(pr_i \circ \Psi_x)(y) = pr_i(y - x) = y_i - x_i = pr_i(y) - x_i;$$

accordingly,

$$d_x^2 = \sum_{i=1}^{d} (pr_i \circ \Psi_x)^2 \quad \text{and} \quad d_x^4 = \sum_{i,j=1}^{d} (pr_i \circ \Psi_x)^2 (pr_j \circ \Psi_x)^2. \tag{4.1}$$

Theorem 4.1. *Under the hypothesis $supp(\mu) = Q_d$, assume also that:*

(i) *For every $i = 1, \dots, d$, there exists $\beta_i \in C(Q_d)$ such that*

$$\lim_{n \to \infty} n(M_{n,\mu}(pr_i) - pr_i) = \beta_i \quad \textit{uniformly on } Q_d. \tag{4.2}$$

(ii) *For every $i, j = 1, \dots, d$, there exists $\gamma_{ij} \in C(Q_d)$ such that*

$$\lim_{n \to \infty} n(M_{n,\mu}(pr_i pr_j) - pr_i pr_j) = \gamma_{ij} \quad \textit{uniformly on } Q_d. \tag{4.3}$$

(iii) *For every $x \in Q_d$, $n \geq 1$ and for every $h \in \mathbf{N}^d$, $0_d \leq h \leq n_d$, one has*

$$\lim_{n \to \infty} n \max_{\substack{h \in \mathbf{N}^d \\ 0_d \leq h \leq n_d}} \left(\omega_{n_d,h}(d_x^4, \mu) - \sum_{i,j=1}^{d} \left(\frac{h_i}{n} - x_i \right)^2 \left(\frac{h_j}{n} - x_j \right)^2 \right) = 0 \tag{4.4}$$

uniformly w.r.t. $x \in Q_d$.

Then, for every $u \in C^2(Q_d)$,

$$\lim_{n \to \infty} n(M_{n,\mu}(u) - u) = \sum_{i,j=1}^{d} \alpha_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{d} \beta_i \frac{\partial u}{\partial x_i} \quad \textit{uniformly on } Q_d, \tag{4.5}$$

where, for every $x \in Q_d$,

$$\alpha_{ij}(x) := \gamma_{ij}(x) - x_i \beta_j(x) - x_j \beta_i(x).$$

Proof. On account of Theorem 1.5.2 in [4], (4.5) will be proved once we show that, for every $i, j = 1, \dots, d$,

(a) $\lim_{n \to \infty} n M_{n,\mu}(pr_i \circ \Psi_x) - \beta_i = 0$ uniformly on Q_d;

(b) $\lim_{n \to \infty} n M_{n,\mu}((pr_i \circ \Psi_x)(pr_j \circ \Psi_x)) - 2\alpha_{ij} = 0$ uniformly on Q_d;

(c) $\sup_{n\geq 1, x\in Q_d} nM_{n,\mu}(d_x^2)(x) < +\infty$;

(d) $\lim_{n\to\infty} nM_{n,\mu}(d_x^4)(x) = 0$ uniformly w.r.t. $x \in Q_d$.

In order to prove statement (a), we first observe that, for any $n \geq 1$, $h \in \mathbf{N}^d$, $0 \leq h \leq n_d$ and $x \in Q_d$,

$$\omega_{n_d,h}((pr_i \circ \Psi_x), \mu) = \omega_{n_d,h}(pr_i, \mu) - x_i;$$

this, together with assumption (4.2), completes the proof.

We pass now to prove statement (b). It is, indeed, a consequence of (4.2) and (4.3), once one notices that

$$\omega_{n_d,h}((pr_i \circ \Psi_x)(pr_j \circ \Psi_x, \mu) = (\omega_{n_d,h}(pr_i pr_j, \mu) - x_i x_j)$$
$$- x_i(\omega_{n_d,h}(pr_j, \mu) - x_j) - x_j(\omega_{n_d,h}(pr_i, \mu) - x_i).$$

Statement (c) follows directly from (4.1).

Finally, we have to prove statement (d). We recall that for the sequence $(B_n)_{n\geq 1}$ of the Bernstein operators (3.4), one has that

$$\lim_{n\to\infty} nB_n(d_x^4)(x) = 0$$

uniformly w.r.t. $x \in Q_d$ (see [2, Formula (5), p. 434 and Proposition 6.2.3]).

Under assumption (4.4), as for every $x \in Q_d$,

$$nM_{n,\mu}(d_x^4)(x) = nB_n(d_x^4)(x) + n\left(M_{n,\mu}(d_x^4)(x) - B_n(d_x^4)(x)\right)$$
$$\leq nB_n(d_x^4)(x) + n|M_{n,\mu}(d_x^4)(x) - B_n(d_x^4)(x)| \leq nB_n(d_x^4)(x)$$
$$+ n \sum_{\substack{h\in\mathbf{N}^d \\ 0_d\leq h\leq n_d}} \left| \omega_{n_d,h}(d_x^4, \mu) - \sum_{i,j=1}^{d} \left(\frac{h_i}{n} - x_i\right)^2 \left(\frac{h_j}{n} - x_j\right)^2 \right| \binom{n_d}{h} x^h (1_d - x)^{n_d - h},$$

we easily get that $\lim_{n\to\infty} nM_{n,\mu}(d_x^4)(x) = 0$ uniformly w.r.t. $x \in Q_d$, and this finishes the proof. □

Remark 4.2. As we have shown in (3.15), some estimates for the rate of convergence in (3.16), in terms of the K-functionals for L^p-spaces, are available. A more difficult question is to establish the rate of convergence of $M_{n,\mu}$ with respect to the uniform norm. Under the assumptions (i)-(ii) of Theorem 4.1, we can give a partial answer to this question.

From [10, Theorem 2], we infer the general estimate in terms of the second modulus of continuity $\omega_2(f, \delta)$: for every $n \geq 1$ and $f \in C(Q_d)$,

$$\|M_{n,\mu}(f) - f\|_\infty \leq C\left(\lambda_{n,\infty}\|f\|_\infty + \omega_2(f, \lambda_{n,\infty}^{1/2})\right),$$

where C is an absolute constant depending only on d, $\lambda_{n,\infty}$ is defined by

$$\lambda_{n,\infty} := \max\{\|M_{n,\mu}(\mathbf{1}) - \mathbf{1}\|_\infty, \|M_{n,\mu}(pr_1) - pr_1\|_\infty,$$
$$\ldots, \|M_{n,\mu}(pr_d) - pr_d\|_\infty, \|M_{n,\mu}(e_2) - e_2\|_\infty\},$$

and $e_2 := \sum_{i=1}^d pr_i^2$.

Assumption (i)-(ii) in Theorem 4.1 yield that there exists $M > 0$ such that, for every $i = 1, \dots, d$

$$\|M_{n,\mu}(pr_i) - pr_i\|_\infty \le \frac{M}{n} \quad \text{and} \quad \|M_{n,\mu}(e_2) - e_2\|_\infty \le \frac{M}{n}.$$

Hence, since $M_{n,\mu}(\mathbf{1}) = \mathbf{1}$, we have that

$$\|M_{n,\mu}(f) - f\|_\infty \le C\left(\frac{M}{n} + \omega_2(f, \sqrt{M/n})\right).$$

5. An example

In this last section, we discuss a concrete example where the previous results apply. We begin to choose the measure μ. Let $a = (a_1, \dots, a_d)$, $b = (b_1, \dots, b_d) \in \mathbf{R}^d$ such that $a_i > -1$ and $b_i > -1$ for any $i = 1, \dots, d$. As measure μ we consider the absolutely continuous measure $\mu_{a,b} \in M_1^+(Q_d)$ with respect to the Borel-Lebesgue measure λ_d on Q_d with density the normalized Jacobi weight

$$w_{a,b}(x) := \frac{x^a(1-x)^b}{\int_{Q_d} y^a(1-y)^b\,dy} \qquad (x \in \overset{\circ}{Q}_d).$$

Note that $\mu_{a,b}$ satisfies property (ii) in Theorem 3.1.

In such a case, the operators $M_{n,\mu_{a,b}}$ turn into the so-called Bernstein-Durrmeyer operators on Q_d with Jacobi weights, which were introduced and studied in [3]. More precisely, they are defined, for every $n \ge 1$, $f \in L^1(Q_d, \mu_{a,b})$ and $x \in Q_d$, by setting

$$M_n(f)(x) := \sum_{\substack{h \in \mathbf{N}^d \\ 0_d \le h \le n_d}} \omega_{n_d,h}(f) \binom{n_d}{h} x^h (1_d - x)^{n_d - h},$$

where, for every $n \ge 1$ and $h = (h_1, \dots h_d) \in \mathbf{N}^d$, $0_d \le h \le n_d$,

$$\begin{aligned}\omega_{n_d,h}(f) &:= \frac{1}{\int_{Q_d} y^{h+a}(1_d - y)^{n_d-h+b}\,dy} \int_{Q_d} y^{h+a}(1_d - y)^{n_d-h+b} f(y)\,dy \\ &= \prod_{i=1}^d \frac{\Gamma(n + a_i + b_i + 2)}{\Gamma(h_i + a_i + 1)\Gamma(n - h_i + b_i + 1)} \int_{Q_d} y^{h+a}(1_d - y)^{n_d-h+b} f(y)\,dy\end{aligned}$$

and $\Gamma(u)$ $(u \ge 0)$ denotes the classical Euler Gamma function.

The operators $M_{n,\mu_{a,b}}$ satisfy assumptions (4.2)-(4.4); in particular, they verify the following asymptotic formula: for every $u \in C^2(Q_d)$,

$$\lim_{n\to\infty} n(M_{n,\mu_{a,b}}(u) - u) = A(u) \quad \text{uniformly on } Q_d,$$

where the differential operator A is defined by

$$A(u)(x) = \sum_{i=1}^d x_i(1-x_i)\frac{\partial^2 u}{\partial x_i^2}(x) + (a_i + 1 - (a_i + b_i + 2)x_i)\frac{\partial u}{\partial x_i}(x)$$

for every $u \in C^2(Q_d)$ and $x = (x_1, \dots, x_d) \in Q_d$.

Note that such a differential operator falls into the category of so-called Fleming-Viot operators. In [3], the authors proved that the operator A is closable and its closure (pre)-generates a Markov semigroup $(T(t))_{t\geq 0}$ on $C(Q_d)$ such that, if $f \in C(Q_d)$ and $t \geq 0$, then

$$T(t)(f) = \lim_{n\to\infty} M_n^{[nt]}(f) \quad \text{uniformly on } Q_d\,,$$

$[nt]$ denoting the integer part of nt $(n \geq 1)$.

A similar relation holds true also in $L^p(Q_d, \mu_{a,b})$. An open problem should be to understand under which conditions this holds true in the more general context of the Bernstein-Durrmeyer operators $M_{n,\mu}$ with respect to arbitrary measures.

References

[1] Albanese, A.A., Campiti, M., Mangino, E.M., *Regularity properties of semigroups generated by some Fleming-Viot type operators*, J. Math. Anal. Appl., **335**(2007), 1259-1273.

[2] Altomare, F., Campiti, M., *Korovkin-Type Approximation Theory and its Applications*, de Gruyter Studies in Mathematics, **17**, Walter de Gruyter, Berlin-New York, 1994.

[3] Altomare, F., Cappelletti Montano, M., Leonessa, V., *On the positive semigroups generated by Fleming-Viot type differential operators*, Comm. Pure Appl. Anal., **18**(1)(2019), 323-340.

[4] Altomare, F., Cappelletti Montano, M., Leonessa, V., Raşa, I., *Markov Operators, Positive Semigroups and Approximation Processes*, de Gruyter Studies in Mathematics, **61**, Walter de Gruyter GmbH, Berlin/Boston, 2014.

[5] Bauer, H., *Probability Theory*, de Gruyter Studies in Mathematics, **23**, Walter de Gruyter & Co., Berlin, 1996.

[6] Berdysheva, E.E., *Uniform convergence of Bernstein-Durrmeyer operators with respect to arbitrary measure*, J. Math. Anal. Appl., **394**(2012), 324-336.

[7] Berdysheva, E.E., *Bernstein-Durrmeyer operators with respect to arbitrary measure, II: Pointwise convergence*, J. Math. Anal. Appl., **418**(2014), 734-752.

[8] Berdysheva, E.E., Jetter, K., *Multivariate Bernstein Durrmeyer operators with arbitrary weight functions*, J. Approx. Theory, **162**(2010), 576-598.

[9] Berdysheva, E.E., Li, B.Z., *On L^p-convergence of Bernstein Durrmeyer operators with respect to arbitrary measure*, Publ. Inst. Math., Beograd (N.S.), **96**(110)(2014), 23-29.

[10] Berens, H., De Vore, R., *Quantitative Korovkin Theorems for positive linear operators on L_p-spaces*, Trans. Amer. Math. Soc., **245**(1978), 349-361.

[11] Berens, H., Xu, Y., *On Bernstein-Durrmeyer polynomials with Jacobi weights*, in: C.K. Chui (ed.), Approximation Theory and Functional Analysis, Academic Press, Boston, 1991, 25-46.

[12] Derriennic, M.M., *Sur l'approximation de fonctions intégrables sur* $[0,1]$ *par des polynômes de Bernstein modifies*, J. Approx. Theory, **31**(1981), 325-343.

[13] Ditzian, Z., *Multidimensional Jabobi-type Bernstein-Durrmeyer operators*, Acta Sci. Math., Szeged, **60**(1995), 225-243.

[14] Ditzian, Z., Totik, V., *Moduli of Smoothness*, Springer Series in Computational Mathematics, **9**, Springer-Verlag, New York, 1987.

[15] Durrmeyer, J.L., *Une formule d'inversion de la transformée de Laplace: applications à la théorie des moments*, Thèse de 3e cycle, Faculté des Sciences de l'Université de Paris, 1967.

[16] Lorentz, G.G., *Bernstein Polynomials*, University of Toronto Press, Toronto, 1953.

[17] Li, B.Z., *Approximation by multivariate Bernstein-Durrmeyer operators and learning rates of least-squares regularized regression with multivariate polynomial kernels*, J. Approx. Theory, **173**(2013), 33-55.

[18] Lupaş, A., *Die Folge der Betaoperatoren*, Dissertation, Universität Stuttgart, 1972.

[19] Păltănea, R., *Sur un opérateur polynomial défini sur l'ensemble des fonctions intégrables*, Univ. Babeş-Bolyai, Cluj-Napoca, **83**(1983), no. 2, 101-106.

[20] Royden, H.L., *Real Analysis*, third ed., Macmillan Publishing Company, New York, 1988.

[21] Vladislav, T., Raşa, I., *Analiză Numerică, Aproximare, Problema lui Cauchy Abstractă, Proiectorii Altomare*, Ed. Tehnică, Bucureşti, 1999.

[22] Waldron, Sh., *A generalization of beta integral and the limit of Bernstein-Durrmeyer operator with Jacobi weights*, J. Approx. Theory, **122**(2003), 141-150.

Mirella Cappelletti Montano
Università degli Sudi di Bari Aldo Moro
Dipartimento di Matematica
Via E. Orabona 4
70125 Bari, Italy
e-mail: `mirella.cappellettimontano@uniba.it`

Vita Leonessa
Università degli Studi della Basilicata
Dipartimento di Matematica, Informatica ed Economia
Viale dell'Ateneo Lucano 10
85100 Potenza, Italy
e-mail: `vita.leonessa@unibas.it`

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 253–262
DOI: 10.24193/subbmath.2019.2.10

Stone-Weierstrass theorems for random functions

Hans-Jörg Starkloff

Abstract. We present several generalizations of the *Stone-Weierstrass* theorem concerning the approximation of continuous functions on a compact set by using functions from a subalgebra to the case of random functions and random variables in the space of continuous functions. The continuity of the random functions is allowed to be only with respect to a metric, hence including the case of stochastically continuous random functions. These results could be cornerstones for the general theory of approximation for random functions.

Mathematics Subject Classification (2010): 41A65, 60G07, 60B11.

Keywords: Stone-Weierstrass theorem, approximation of random functions, stochastic convergence, random polynomial.

1. Introduction

It is well known that approximation theory plays an important role in the mathematical investigation of deterministic equations and for other problems. Therefore it seems to be natural that approximation theory should also play a similar important role in the investigation of different types of random equations. Moreover in recent time an increased use of stochastic models and a very intensive investigation of random and stochastic differential equations can be observed. So the systematic investigation of approximation procedures and possibilities for random mathematical objects seems to be useful, leading possibly to the development of an approximation theory for random functions and random variables in function spaces.

One fundamental result in deterministic approximation theory is the *Stone-Weierstrass* theorem about the uniform closure of a subalgebra of continuous functions in the space of all continuous functions on a compact set. This result is a generalization of the classical theorem of *Weierstrass* about the denseness (with respect to the norm of uniform convergence) of polynomials in the space of continuous functions on a closed interval of the real line.

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

The presented work deals with analogous questions for random functions. Hereby different generalizations can be considered, due for example to the use of different norms or metrics in spaces of random variables.

In the remaining part of this section a short remainder on deterministic *Stone-Weierstrass* theorems together with some needed definitions is given. The following section is devoted to different theorems of *Stone-Weierstrass* type for random functions or random variables with values in function spaces.

The following notations and concepts are used throughout the article.

- T denotes a compact *Hausdorff* topological space.
- $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$ denotes the scalar field (of real or complex numbers), $\mathbb{N}^* = \{1, 2, 3, \ldots\}$ is the set of positive natural numbers.
- $\mathrm{C}(T, \mathbb{K})$ denotes the linear space of continuous $\mathbb{K}$-valued functions on T, endowed with the maximum norm (this way it becomes a *Banach* space and moreover a *Banach* algebra).
- $\mathbf{1}_S$ denotes the indicator function of the set S (also called characteristic function) with values 1 for elements from S and 0 otherwise.
- A set A of $\mathbb{K}$-valued functions on T is called an algebra if it holds

$$f, g \in A, \alpha \in \mathbb{K} \quad \Rightarrow \quad f + g \in A,\ \alpha f \in A,\ f \cdot g \in A\,.$$

The classical theorem of *Weierstrass* about the approximation of real continuous functions by algebraic polynomials can be formulated as follows.

Theorem 1.1. *The set (algebra) of polynomials with real coefficients is dense in the space of real-valued continuous functions on a finite interval $[a, b] \subset \mathbb{R}$, endowed with the maximum norm.*

Analogous results hold for complex-valued functions and for functions defined on $[a_1, b_1] \times \ldots \times [a_d, b_d] \subset \mathbb{R}^d$ $(d \in \mathbb{N}^*)$.

The approximability of continuous functions by algebraic polynomials is strongly related to the approximability of periodic functions by trigonometric polynomials.

Theorem 1.2. *The set (algebra) of trigonometric polynomials with real coefficients is dense in the space of real-valued $2\pi-$periodic continuous functions on the interval $[0, 2\pi] \subset \mathbb{R}$, endowed with the maximum norm.*

Analogous results hold for complex-valued functions, for periodic functions defined on $[0, 2\pi]^d \subset \mathbb{R}^d$ $(d \in \mathbb{N}^*)$ and for periodic functions with other periods.

One generalization of this result, namely a version of the theorem of *Stone-Weierstrass* can be given as follows.

Theorem 1.3. *A subalgebra $A \subset \mathrm{C}(T, \mathbb{K})$ is dense in $\mathrm{C}(T, \mathbb{K})$, if and only if*

(i) *A separates the points of T (A is point-separating on T), i.e.,*

$$\forall\, t_1 \neq t_2 \in T\ \exists\, f \in A\, :\, f(t_1) \neq f(t_2);$$

(ii) *A is not point-vanishing on T, i.e., $\forall\, t \in T\ \exists\, f \in A\, :\, f(t) \neq 0$;*
(iii) *A is self-adjoint in the case $K = \mathbb{C}$, i.e., $\forall\, f \in A\ \Rightarrow \overline{f} \in A$.*

These theorems can be found in many textbooks for analysis or approximation theory. Other versions of the *Stone-Weierstrass* theorem, different proofs, etc. can be found for example in *Prolla* [6].

2. Stone-Weierstrass theorems for random functions and random variables with values in a function space

In the following all random objects are assumed to be defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. The corresponding expectation operator is denoted by E. $\mathrm{L}^0(\mathbb{K})$ denotes the linear space of equivalence classes of random variables with values in $\mathbb{K}$ (and analogously for other ranges). For $\mathbb{K}$-valued random variables the stochastic convergence is connected to the concept of metric, one possible metric is

$$d_p(\eta_1, \eta_2) = \mathrm{E}\left[\frac{|\eta_1 - \eta_2|}{1 + |\eta_1 - \eta_2|}\right], \quad \eta_1, \eta_2 \in \mathrm{L}^0(\mathbb{K}),$$

with analogous expression in the case of a (semi-)normed space as range instead of $\mathbb{K}$. As usual $\mathrm{L}^p(\mathbb{K})$ is the linear space of equivalence classes of $\mathbb{K}$-valued random variables with finite moment of order p, endowed with the corresponding norm ($1 \le p < \infty$) or metric ($0 < p < 1$).

Stochastic generalizations of the *Stone-Weierstrass* theorem can be derived

- for random variables (random elements) with values in $\mathrm{C}(T, \mathbb{K})$ or
- for $\mathbb{K}$-valued random functions $(\xi_t)_{t \in T}$, continuous in some sense.

Further generalizations can be given for example for random functions with values in separable *Banach* spaces.

The approximating set is often the linear span (which is not necessarily an algebra)

$$\mathrm{span}\{\eta f \,:\, \eta \in V \subseteq \mathrm{L}^0(\mathbb{K}),\ f \in A \subseteq \mathrm{C}(T, \mathbb{K})\}$$

with suitable sets V and A.

The notion or type of convergence for (generalized) sequences of random variables or random functions can be generated by norms or metrics or otherwise. In case of norms many theorems can be generalized straightforwardly. More difficult is the study of these results and definitions in the context of a metric (which is for example valid for the case of stochastic convergence).

As a first generalization of the deterministic *Stone-Weierstrass* theorem a result for random variables in the space of continuous functions and stochastic convergence is given.

Theorem 2.1. *Assume that $\mathbb{X}$ is a separable Banach space with norm $\|\cdot\|$ or more generally a complete separable metrizable locally convex space with metric $\|\cdot\|_{\mathbb{X}}$ over the field $\mathbb{K}$, T is a compact metric space with metric r and $A \subseteq \mathrm{C}(T, \mathbb{K})$ is a self-adjoint algebra, which is point-separating and not point-vanishing on T. Then*

$$S := \mathrm{span}\{\eta f \,:\, \eta \in \mathrm{L}^0(\mathbb{X}),\ f \in A\}$$

is dense with respect to the stochastic convergence in $\mathrm{L}^0(\mathrm{C}(T,\mathbb{X}))$*, i.e.,*

$$\forall\,\varepsilon>0\ \forall\,\xi\in\mathrm{L}^0(\mathrm{C}(T,\mathbb{X}))\ \exists\,\zeta\in S\,:\,\mathrm{P}(\max_{t\in T}\|\xi_t-\zeta_t\|_{\mathbb{X}}>\varepsilon)<\varepsilon.$$

Proof. Convergence in a complete separable metrizable locally convex space is equivalent to the convergence for each of an at most countable defining system of half-norms (see, e.g. *Rudin* [9], Theorem 1.24, Remarks 1.38). So the proof is given for the case of a separable *Banach* space $\mathbb{X}$ with norm $\|\cdot\|$ and it is identical for each of the half-norms in the locally convex spaces.

In this case the asserted property is equivalent to

$$\forall\,\varepsilon>0\ \forall\,\xi\in\mathrm{L}^0(\mathrm{C}(T,\mathbb{X}))\ \exists\,\zeta\in S\,:\,\mathrm{E}\left[\frac{\max_{t\in T}\|\xi_t-\zeta_t\|}{1+\max_{t\in T}\|\xi_t-\zeta_t\|}\right]<\varepsilon.$$

$\mathrm{C}(T,\mathbb{X})$ is separable if T is a compact metric space. Hence for each $\varepsilon>0$ there exists a compact set $K\subset\mathrm{C}(T,\mathbb{X})$ with $\mathrm{P}(\xi\in K)>1-\varepsilon/3$ (see, e.g. *Billingsley* [1], Theorem 1.4).

Define $\xi':=\xi\cdot\mathbf{1}_{\{\xi\in K\}}$, so that $\mathrm{P}(\max_{t\in T}\|\xi_t-\xi'_t\|>0)<\varepsilon/3$ and hence

$$\mathrm{E}\left[\frac{\max_{t\in T}\|\xi_t-\xi'_t\|}{1+\max_{t\in T}\|\xi_t-\xi'_t\|}\right]<\varepsilon/3.$$

The functions in K and hence the realizations of ξ' are bounded and equicontinuous (see, e.g., *Rudin* [9], Theorems A 4, A 5, *Dieudonné* [3], Theorem 7.5.7). Hence there exists $\delta>0$ with $\|\xi'_t(\omega)-\xi'_s(\omega)\|<\varepsilon/3$ for all $\omega\in\Omega$ if $r(t,s)<\delta$. The compact set T can be covered by a finite number of $\delta-$neighbourhoods of points $t_i\in T, i=1,\dots,n$. Assume $(z_i, i=1,\dots,n)$ is a partition of unity subordinate to this open cover (see, e.g., *Rudin* [8], Theorem 2.13). Defining

$$\hat{\xi}_t:=\sum_{i=1}^{n}\xi'_{t_i}z_i(t),\qquad t\in T,$$

it holds for arbitrary $t\in T$ and $\omega\in\Omega$

$$\|\hat{\xi}_t(\omega)-\xi'_t(\omega)\|\le\sum_{i=1}^{n}\|\xi'_{t_i}-\xi'_t\|z_i(t)\le\sum_{i=1}^{n}\frac{\varepsilon}{3}z_i(t)=\frac{\varepsilon}{3}.$$

This means that the set $\operatorname{span}\{\eta f\,:\,\eta\in\mathrm{L}^0(\mathbb{X})\,,\ f\in\mathrm{C}(T,\mathbb{K})\}$ is dense with respect to the stochastic convergence in $\mathrm{L}^0(\mathrm{C}(T,\mathbb{X}))$. It remains to approximate the functions $z_i\in\mathrm{C}(T,\mathbb{K})$ uniformly by functions from $A\subseteq\mathrm{C}(T,\mathbb{K})$, which is possible using the deterministic *Stone-Weierstrass* Theorem 1.3. □

Corollary 2.2. *Assume that* T *is a compact metric space with metric* r *and* $A\subseteq\mathrm{C}(T,\mathbb{K})$ *is a self-adjoint algebra, which is point-separating and not point-vanishing on* T*. Then* $S:=\operatorname{span}\{\eta f\,:\,\eta\in\mathrm{L}^0(\mathbb{K})\,,\ f\in A\}$ *is dense with respect to the stochastic convergence in* $\mathrm{L}^0(\mathrm{C}(T,\mathbb{K}))$*, i.e.,*

$$\forall\,\varepsilon>0\ \forall\,\xi\in\mathrm{L}^0(\mathrm{C}(T,\mathbb{K}))\ \exists\,\zeta\in S\,:\,\mathrm{P}(\max_{t\in T}|\xi_t-\zeta_t|>\varepsilon)<\varepsilon.$$

Remark 2.3. In relation to the above theorem one can remark the following.

- One can find for every $\xi \in \mathrm{L}^0(\mathrm{C}(T,\mathbb{X}))$ a sequence $(\zeta_n)_{n\in\mathbb{N}^*} \subset S$, which converges almost surely to ξ.
- Analogous results hold for $\mathrm{L}^p(\mathrm{C}(T,\mathbb{X}))$ $(1 \le p < \infty)$ with convergence in $p-$th mean.
- $\mathrm{C}(T,\mathbb{K})$ is separable iff T is a compact metric space (see, e.g., *Rolewicz* [7], Proposition 1.6.6). The investigation of random variables with values in non-separable normed spaces is much more complicated, so we restrict ourselves to the case of a separable *Banach* space $\mathrm{C}(T,\mathbb{K})$.

Now we consider the case of $\mathbb{K}$-valued random functions $(\xi_t)_{t\in T}$ which are continuous in some sense. Correspondig results can be found in the literature for example in *Dugué* [4] ($T \subset \mathbb{R}$ finite interval, approximation of a random function $(\xi_t)_{t\in T}$, which is continuous in probability, by random polynomials with respect to the uniform stochastic convergence), *Onicescu, Istrăţescu* [5] ($T \subset \mathbb{R}^d$ $(d \in \mathbb{N}^*)$ convex compact set, uniform stochastic approximation by multivariate random polynomials) and *Ryabykh, Tokmakova, Yablonskiĭ* [10] ($T \subset \mathbb{R}^d$ $(d \in \mathbb{N}^*)$ compact set, uniform stochastic approximation by elements of a subalgebra of random functions). It can be remarked, that, in general, spaces of random variables endowed with the topology of stochastic convergence are not locally convex spaces, so that generalizations of *Stone-Weierstrass* theorems for locally convex spaces (see, e.g. *Timofte* [11]) are not applicable in this situation.

A stochastic generalization of *Stone-Weierstrass* theorem is easy to obtain if a topology in the space of random variables induced by a norm is used.

Theorem 2.4. *Assume $(V, \|\cdot\|)$ is a normed subspace of $\mathrm{L}^0(\mathbb{K})$ and $A \subseteq \mathrm{C}(T,\mathbb{K})$ is a self-adjoint algebra, which is point-separating and not point-vanishing on the compact Hausdorff topological space T. Then*

$$S := \mathrm{span}\{\eta f : \eta \in V\,,\ f \in A\}$$

is dense in $\mathrm{C}(T,V)$ with respect to the uniform $\|\cdot\|-$convergence on T.

Proof. For $\xi \in \mathrm{C}(T,V)$ and $\varepsilon > 0$ consider the open cover of T defined by sets $U_t := \{s \in T : \|\xi_s - \xi_t\| < \varepsilon/2\}$. Due to the compactness one can choose a finite open subcover of T with sets U_{t_i}, where $t_i \in T$, $i = 1,\dots,n \in \mathbb{N}^*$, are points which are pairwise disjoint. Assume $(z_i, i = 1,\dots,n)$ is a partition of unity subordinate to this open subcover and define

$$\hat{\xi}_t := \sum_{i=1}^{n} \xi_{t_i} z_i(t), \qquad t \in T.$$

Then it holds $\hat{\xi} \in \mathrm{span}\{\eta f : \eta \in V\,,\ f \in \mathrm{C}(T,\mathbb{K})\}$ and for all $t \in T$

$$\|\hat{\xi}_t - \xi_t\| \le \sum_{i=1}^{n} \|\xi_{t_i} - \xi_t\| z_i(t) < \varepsilon/2,$$

because it holds $z_i(t) > 0$ only for points $t \in U_{t_i}$. It remains to approximate the functions $z_i \in \mathrm{C}(T,\mathbb{K})$ uniformly by functions $\tilde{z}_i \in A \subseteq \mathrm{C}(T,\mathbb{K})$, which is possible by the deterministic *Stone-Weierstrass* Theorem 1.3, in such a way that

$$\max_{t\in T} |z_i(t) - \tilde{z}_i(t)| < \varepsilon/(2nM), \quad i = 1, \ldots, n,$$

where $M := \max_{t\in T} \|\xi_t\| < \infty$. Then for

$$\tilde{\xi}_t := \sum_{i=1}^{n} \xi_{t_i} \tilde{z}_i(t), \qquad t \in T,$$

it holds $\tilde{\xi} \in \operatorname{span}\{\eta f : \eta \in V\,,\ f \in A\}$ and for all $t \in T$

$$\|\tilde{\xi}_t - \hat{\xi}_t\| \leq \sum_{i=1}^{n} \|\xi_{t_i}\| \cdot |\tilde{z}_i(t) - z_i(t)| < \varepsilon/2.$$

Now from the triangle inequality the assertion follows. □

This result is valid for arbitrary functions with values in a (half-)normed space, defined on a compact set and can be deduced easily also for example from Theorem 1 in Chapter 2 of *Prolla* [6].

Measuring the nearness of random variables with the help of a metric it is desirable that $V \subseteq \mathrm{L}^0(\mathbb{K})$ is a metric linear space, i.e., the linear operations are continuous. Then from general theory it follows (see, e.g., *Rolewicz* [7], Theorems 1.1.1 and 1.2.2) that there exists a translation-invariant non-decreasing metric (an "F-norm" denoted by $\|\cdot\|$), which is equivalent to the given metric. Basic properties of an F-norm on $\mathrm{L}^0(\mathbb{K})$ are

(F 1) $\|\mathbf{1}_\Omega\| < \infty\,.$
(F 2) $\|\xi\| = 0 \Leftrightarrow \xi = 0$ a. s.
(F 3) $\|\alpha\xi\| \leq \|\xi\|$ for all $\alpha \in \mathbb{K}$ with $|\alpha| \leq 1$.
(F 4) $\|\xi + \eta\| \leq \|\xi\| + \|\eta\|$.
(F 5) $\|a_n\xi\| \to 0$, if $a_n \to 0 \ \ (n \to \infty)$.

In certain cases further properties of the F-norm are imposed.

(F 6) $|\xi| \leq |\eta|$ a.s. and $\|\eta\| < \infty \ \Rightarrow \|\xi\| \leq \|\eta\|$.
(F 7) For each sequence $(F_n)_{n\in\mathbb{N}^*}$, $F_n \in \mathcal{F}$ with $\mathrm{P}(F_n) \to 0$ and every $\xi \in \mathrm{L}^0(\mathbb{K})$ with $\|\xi\| < \infty$ it holds $\|\xi \mathbf{1}_{F_n}\| \to 0 \ \ (n \to \infty)$.
(F 8) There exists a constant $\kappa = \kappa(c) > 0$ such that for arbitrary random variables ξ with $\mathrm{P}(|\xi| \leq c) = 1$ for a real number $c > 0$ and for arbitrary $a \in \mathbb{R}$ it holds

$$\|a\xi\| \leq |a|\kappa(c)\|\xi\|.$$

The F-norm generating the stochastic convergence is $\|\xi\|_p = d_p(\xi, 0)$ and fulfills all the properties (F 1)-(F 7) from above and also (F 8) as is proven below in Lemma 2.9.

Some further properties related to F-norms are stated now.

Lemma 2.5. *Let $(V, \|\cdot\|)$ be an $F-$normed subspace of $\mathrm{L}^0(\mathbb{K})$, such that the basic and the additional properties* (F 1)-(F 7) *of F-norms are fulfilled and let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of elements from V which converges stochastically to $\xi \in V$ and such that for a random variable $\eta \in V$ it holds* $\mathrm{P}(|\xi_n| \leq \eta) = 1$ *for all $n \in \mathbb{N}$. Then it holds also $\|\xi_n - \xi\| \to 0$ for $n \to \infty$.*

Proof. One shows that from each subsequence $(\xi_{n'})$ a subsubsequence $(\xi_{n''})$, $\{n''\} \subseteq \{n'\}$, can be chosen such that $\|\xi_{n''} - \xi\| \to 0$ for $n'' \to \infty$. Then the assertion follows.

Due to the stochastic convergence of $(\xi_{n'})$ to ξ one can choose a subsubsequence $(\xi_{n''})$ which converges almost surely to ξ. By the Theorem of *Egorov* (see, e.g. *Bogachev* [2], Theorem 2.2.1) it follows that for every $\varepsilon > 0$ there exists $B_\varepsilon \in \mathcal{F}$ with $\mathrm{P}(\Omega \backslash B_\varepsilon) < \varepsilon$ and uniform convergence of $(\xi_{n''})$ to ξ on B_ε. In

$$\|\xi_{n''} - \xi\| \leq \|(\xi_{n''} - \xi)\mathbf{1}_{B_\varepsilon}\| + +\|(\xi_{n''} - \xi)\mathbf{1}_{\Omega\backslash B_\varepsilon}\|$$

the first summand on the right hand side converges for each $\varepsilon > 0$ for $n'' \to \infty$ to zero due to the uniform convergence and properties (F 5) and (F 6). From property (F 7) it follows hat the second summand on the right hand side converges for $\varepsilon \to 0$ to zero. □

Corollary 2.6. *Let* $(V, \|\cdot\|)$ *be an* $F-$*normed subspace of* $\mathrm{L}^0(\mathbb{K})$*, such that the basic and the additional properties* (F 1)-(F 7) *of F-norms are fulfilled and let* T *be a set. Assume* $(\xi_t^n, t \in T)_{n\in\mathbb{N}}$ *is a sequence of functions with values in* V *such that there exists* $\eta \in V$ *with* $\mathrm{P}(|\xi_t^n| \leq \eta) = 1$ *for all* $t \in T$ *and* $n \in \mathbb{N}$*. Then it holds for a random function* $(\xi_t)_{t\in T}$ *with values in* V

$$\lim_{n\to\infty} \sup_{t\in T} \|\xi_t^n - \xi_t\|_p = 0 \quad \Rightarrow \quad \lim_{n\to\infty} \sup_{t\in T} \|\xi_t^n - \xi_t\| = 0.$$

Proof. Denote $V_\eta = \{\xi \in V : |\xi| \leq \eta\}$. The assertion follows from the fact that the identity operator from $(V_\eta, \|\cdot\|_p)$ to $(V_\eta, \|\cdot\|)$ is continuous by Lemma 2.5. □

Lemma 2.7. *Let* $(V, \|\cdot\|)$ *be an* $F-$*normed subspace of* $\mathrm{L}^0(\mathbb{K})$*, such that the basic properties* (F 1)-(F 5) *of F-norms are fulfilled,* $\xi \in V$ *and let be given* $f_n \in \mathrm{C}(T, \mathbb{K})$*,* $n \in \mathbb{N}^*$*, such that* $f_n \to f$ $(n \to \infty)$ *in* $\mathrm{C}(T, \mathbb{K})$*. Then it holds also*

$$\lim_{n\to\infty} \sup_{t\in T} \|\xi\,(f_n(t) - f(t))\,\| = 0.$$

Proof. This follows directly from property (F 5) and the assumptions. □

For a real-valued random variable ξ and a real number $c > 0$ its truncation is defined by $\xi^{(c)} := \sup\{-c, \inf\{\xi, c\}\}$ (and analogously for the real and imaginary part in the complex-valued case). So it follows that $\mathrm{P}(|\xi^{(c)}| \leq c) = 1$.

Lemma 2.8. *Assume* T *is a compact Hausdorff topological space,* $(V, \|\cdot\|)$ *is an* $F-$*normed subspace of* $\mathrm{L}^0(\mathbb{K})$*, such that the basic and the additional properties* (F 1)-(F 7) *of F-norms are fulfilled and such that* $\forall c > 0\ \xi \in V \Rightarrow \xi^{(c)} \in V$*. Furthermore let* $\xi = (\xi_t)_{t\in T}$ *be a continuous function with values in* V*.*

(i) $(\xi_t^{(c)})_{t\in T}$ *is continuous in* $(V, \|\cdot\|)$ *for arbitrary* $c > 0$*.*
(ii) $\lim_{c\to\infty} \sup_{t\in T} \|\xi_t^{(c)} - \xi_t\| = 0$*.*

Proof. (i) The assertion follows from property (F 6).

(ii) For $\xi \in \mathrm{C}(T, V)$ and $\varepsilon > 0$ consider the open cover of T defined by sets

$$U_t := \{s \in T : \|\xi_s - \xi_t\| < \varepsilon/3\}.$$

Due to the compactness one can choose a finite open subcover of T with sets U_{t_i}, where $t_i \in T$, $i = 1, \ldots, n \in \mathbb{N}^*$, are points which are pairwise disjoint. Then for each

$t \in T$ one has $\|\xi_{t_i} - \xi_t\| < \varepsilon/3$ for one t_i and consequently by property (F 6) also $\|\xi_{t_i}^{(c)} - \xi_t^{(c)}\| < \varepsilon/3$ for arbitrary $c > 0$. Property (F 7) allows us to find $c > 0$ such that for all $i = 1, \ldots, n$ it holds $\|\xi_{t_i}^{(c)} - \xi_{t_i}\| < \varepsilon/3$. Then it follows

$$\|\xi_t^{(c)} - \xi_t\| \le \|\xi_t^{(c)} - \xi_{t_i}^{(c)}\| + \|\xi_{t_i}^{(c)} - \xi_{t_i}\| + \|\xi_{t_i} - \xi_t\| < \varepsilon. \qquad \square$$

Lemma 2.9. *Let ξ be a random variable with* $\mathrm{P}(|\xi| \le c) = 1$ *for a real number $c > 0$. Then $\|\cdot\|_p$ fulfills* (F 8)*, in particular it holds for arbitrary $a \in \mathbb{R}$*

$$\|a\xi\|_p \le |a|(c+1)\|\xi\|_p.$$

Proof. It holds

$$\|a\xi\|_p = \mathrm{E}\left[\frac{|a\xi|}{1+|a\xi|}\right] \le |a|\,\mathrm{E}\left[\frac{1+|\xi|}{1+|a\xi|}\frac{|\xi|}{1+|\xi|}\right]$$

and

$$\mathrm{P}\left(\frac{1+|\xi|}{1+|a\xi|} \le 1+c\right) = 1. \qquad \square$$

Theorem 2.10. *Assume T is a compact Hausdorff topological space, $(V, \|\cdot\|)$ is an $F-$normed subspace of* $\mathrm{L}^0(\mathbb{K})$*, such that the basic and the additional properties* (F 1)-(F 8) *of F-norms are fulfilled and $A \subseteq \mathrm{C}(T,\mathbb{K})$ is a self-adjoint algebra, which is point-separating and not point-vanishing on T. Then the set $S := \mathrm{span}\{\eta f \,:\, \eta \in V\,,\ f \in A\}$ is dense with respect to the uniform $\|\cdot\|-$convergence on T, i.e., in $\mathrm{C}(T,V)$.*

Proof. First it is proved that $\mathrm{span}\{\eta f \,:\, \eta \in V\,,\ f \in \mathrm{C}(T,\mathbb{K})\}$ is dense in $\mathrm{C}(T,V)$.

One can use the truncation procedure. From Lemma 2.8 (ii) it follows

$$\lim_{c\to\infty} \sup_{t\in T} \|\xi_t^{(c)} - \xi_t\| = 0,$$

hence for some $c > 0$ it holds $\sup_{t\in T} \|\xi_t^{(c)} - \xi_t\| < \varepsilon/3$. One constructs as in the proof of Theorem 2.4 using (F 8)

$$\hat{\xi}_t^{(c)} := \sum_{i=1}^n \xi_{t_i}^{(c)} z_i(t)$$

with $\hat{\xi}^c \in \mathrm{span}\{\eta f \,:\, \eta \in V, \mathrm{P}(|\eta| \le c) = 1, f \in \mathrm{C}(T,\mathbb{K})\}$ and

$$\sup_{t\in T} \|\hat{\xi}_t^{(c)} - \xi_t^{(c)}\| < \varepsilon/3.$$

For finishing the proof it again remains to approximate the functions $z_i \in \mathrm{C}(T,\mathbb{K})$ by functions $\tilde{z}_i(t) \in A \subseteq \mathrm{C}(T,\mathbb{K})$ which is possible by the deterministic *Stone-Weierstrass* Theorem 1.3. Based on (F 3) and (F 5) on finds $\delta > 0$ such that $\|\xi_{t_i}^{(c)}\Delta\| < \varepsilon/(3n)$ for $i = 1, \ldots, n$ and $\Delta \le \delta$. Then for suitable $\tilde{z}_i(t) \in A \subseteq \mathrm{C}(T,\mathbb{K})$ with $\sup_{t\in T} |z_i(t) - \tilde{z}_i(t)| < \delta$ one gets

$$\sup_{t\in T} \|\xi_t - \sum_{i=1}^n \xi_{t_i}^{(c)} \tilde{z}_i(t)\|$$

$$\le \sup_{t\in T} \|\xi_t - \xi_t^{(c)}\| + \sup_{t\in T} \|\xi_t^{(c)} - \hat{\xi}_t^{(c)}\| + \sum_{i=1}^n \sup_{t\in T} \|\hat{\xi}_{t_i}^c (z_i(t) - \tilde{z}_i(t))\| < \varepsilon. \qquad \square$$

Due to Lemma 2.9 this theorem includes the case of uniform stochastic convergence on T.

Theorem 2.11. *If $T \subset \mathbb{R}^d$ ($d \in \mathbb{N}^*$) is a compact set, the additional properties* (F 6)-(F 8) *of the F-norm in the previous theorem are not needed.*

This is due to the fact that there exist open covers of T, such that each point of T is an element of at most 2^d open sets from this cover. Analogous results can also be stated for periodic random functions.

Theorem 2.12. *Theorem 2.5 and Theorem 2.6 remain true if one considers instead of scalar-valued random variables random variables with values in a separable Hilbert space $\mathbb{X}$ and corresponding $F-$normed subspaces of $\mathrm{L}^0(\mathbb{X})$.*

This can be shown again by using norms instead of absolute values in the proofs. The truncation procedure can also be adapted. To see this remark that the set of distributions of the random variables of $(\xi_t\,;\, t \in T)$ is a relatively compact set. Then by the Theorem of *Prokhorov* (see, e.g., *Billingsley* [1], Theorem 6.2) there exist for each $\varepsilon > 0$ a compact set $K_\varepsilon \subset \mathbb{X}$ with $\mathrm{P}(\xi_t \in K_\varepsilon) > 1 - \varepsilon$ for all $t \in T$. This set can also be assumed to be convex. Then the orthogonal projection on this set can be used for the truncation procedure.

3. Conclusions

Several generalizations of the *Stone-Weierstrass* theorem about the possibility of approximation of certain continuous random functions on compact sets by some random functions from some subset are presented. Also the case of random variables in the space of continuous functions is considered.

References

[1] Billingsley, P., *Convergence of Probability Measures*, John Wiley & Sons, New York, 1968.

[2] Bogachev, V.I., *Measure Theory*, Springer, Berlin, 2007.

[3] Dieudonné, D., *Foundations of Modern Analysis*, Academic Press, New York, 1969.

[4] Dugué, D., *Traité de Statistique Théoretique et Appliquée*, Masson, Paris, 1958.

[5] Onicescu, O., Istrăţescu, V.I., *Approximation theorems for random functions*, Rend. Mat. Serie VI, **8**(1975), 65-81.

[6] Prolla, J.B., *Weierstrass-Stone, the Theorem*, Lang, Frankfurt am Main, 1993.

[7] Rolewicz, S., *Metric Linear Spaces*, D. Reidel Publishing Company, Dordrecht, 1984.

[8] Rudin, W., *Real and Complex Analysis*, McGraw-Hill Book Company, New York, 1970.

[9] Rudin, W., *Functional Analysis*, McGraw-Hill Book Company, New York, 1973.

[10] Ryabykh, V.G., Tokmakova,N.N., Yablonskiĭ, A.Ya., *Approksimatsiya sluchaĭnykh funktsiĭ (Approximation of random functions)*, In: Matematicheskiĭ analiz i ego prilozheniya, Rostov na Donu, (1985), 131–135.

[11] Timofte, V., *Stone-Weierstrass theorems revisited*, J. Approx. Theory, **136**(2005), 45-59.

Hans-Jörg Starkloff
Technische Universität Bergakademie Freiberg
Faculty of Mathematics and Computer Sciences
Prüferstraße 9, D-09599 Freiberg, Germany
e-mail: `Hans-Joerg.Starkloff@math.tu-freiberg.de`

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 263–277
DOI: 10.24193/subbmath.2019.2.11

Numerical optimal control for satellite attitude profiles

Ralf Rigger

Abstract. Many modern science satellites are 3-axis stabilized. The construction of attitude profiles therefore play a central role in satellite control. Besides the dynamical properties numerous constraints need to be fulfilled. In [6] a generic way for calculating such attitudes is given. Other options to design slews connecting two attitudes have been published in various papers (e.g. [3, 11]) including approaches using optimal control techniques (e.g. [4, 8, 11]).
In this paper we will present a new approach for optimal control of slews and attitude profiles. After the description of a set of the considered Hamiltonian functions and the respective slew maneuvers some analytical consequences of the choices are given. A comparison with the actual operational Euler angle slew in [6] is given and shows a close match. The performed numerical investigations of direct solutions help to gain a clearer picture on the underlaying analytical problem. By applying the Pontryagin maximum principle to the Hamiltonian equation, a family of closed dynamics ordinary differential equation for the direct optimal control problem is presented and their solutions and properties are investigated.

Mathematics Subject Classification (2010): 49J15, 70Q05, 93C10, 93C15.

Keywords: Attitude, slew, numerical optimal control, Hamiltonian function, Pontryagin maximum principle, system of ordinary differential equations.

1. Introduction

1.1. Dynamic Optimization

For the numerical solution of optimal control problems there are two fundamentally different approaches. Formulating the solution of the optimization problem and then using a discretization method to approximate the solution is called indirect approach [2]. In the so called direct approach the problem is first discretized and

---

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

then optimization methods are used to find an approximate solution [9]. The well known indirect methods are the Hamilton Jacobi Bellmann equation and the Hamilton equations together with the Pontryagin maximum principle. Direct methods have been popular in the recent past. There are several reasons that support the direct approach: To a limited extend realtime applications are possible and it is rather easy and straight forward to incorporate constraints into the procedure. We will consider the second indirect approach in this paper. The well known result from the calculus of variations is given by:

Theorem 1.1. [5, 2] *Let $F\colon I\times\mathbb{R}^n\times\mathbb{R}^m \to \mathbb{R}^n$ and $L\colon I\times\mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}$ be differentiable. Then the variation of the Hamiltonian $H(t,x,u,\lambda) = L(t,x,\dot{x}) + \lambda\cdot F(t,x,u)$ with respect to the independent variables $x\in\mathbb{R}^n$, $\lambda\in\mathbb{R}^n$ and $u\in\mathbb{R}^m$, results in the equations*

$$\begin{aligned}\dot{\lambda}(t) &= -\tfrac{\partial}{\partial x}H(t,x(t),u(t),\lambda(t))\,,\\ \dot{x}(t) &= \tfrac{\partial}{\partial \lambda}H(t,x(t),u(t),\lambda(t))\,,\\ 0 &= \tfrac{\partial}{\partial u}H(t,x(t),u(t),\lambda(t))\,.\end{aligned}$$

The first two equations are differential equations for $x(t)$ and $\lambda(t)$, the so called **Hamiltonian equations**. The last one is the optimality condition, an algebraic equation for $u(t)$, which is valid for all t. The generalization of the optimality condition for the optimal trajectory $\lambda^*(t), x^*(t)$ is:

$$H(t,x^*(t),u^*(t),\lambda^*(t)) = \max_u H(t,x^*(t),u(t),\lambda^*(t))\,.$$

This equation is often referred to as the **Pontryagin maximum principle**. Since we want to prescribe the initial and final values x_{ini} and x_{fin} of our state variables, we will end up with a two-point boundary value problem of the following kind, where u^* is the optimal control to be determined:

$$\begin{aligned}\dot{x} &= \tfrac{\partial}{\partial\lambda}H(t,x,u^*,\lambda)\,, \quad x(t_{\text{ini}}) = x_{\text{ini}}\,,\\ \dot{\lambda} &= -\tfrac{\partial}{\partial x}H(t,x,u^*,\lambda)\,, \quad \lambda(t_{\text{fin}}) = \lambda_{\text{fin}}\,.\end{aligned}$$

Remark 1.2. The exact list of state variables will depend on the exact statement of the problem, e.g. we will have to add an integral constraint to the state variables in order to be able to enforce further constraints on the solution trajectory.

1.2. Numerical Dynamic Optimization

There are numerous ways in order to solve two-point boundary value problems numerically. There are many standard schemes, but with the desire to be able to solve real-time problems time critical approaches have surfaced in the recent years. The numerical simulations for this paper where undertaken by three different schemes.

- A single shooting method using symbolic differentiation, symbolic solvers and standard ordinary differential equation integrators (of Runge-Kutta and Adams type) and the derivative free optimization method of Nelder-Mead.

- A fast direct approach using the CasADi tool with algorithmic differentiation, symbolic ordinary differential equation solver and nonlinear optimization techniques [1].
- A commercial software package with built in boundary value problem solvers. Here the exact solution approach is undisclosed.

Besides the obvious difference in time consumption of the different approaches, there have been no inconsistencies in the respective results. Further the analytical results presented in this paper match the characteristic of the numerical solutions.

2. Optimal Slews

Unit quaternions provide a mathematical way for representing orientations in 3-space. We will denote the field of quaternions by $\mathbb{H}$ and the quaternions themselves by q. The quaternion multiplication is written as $*$. In the following sections vectors x of the $\mathbb{R}^3$ are embedded in $\mathbb{H} \approx \mathbb{R}^4$ in the canonical way by setting the scalar part to 0. With $\overline{q}$ we denote the complex conjugate quaternion of q and for all q_1, q_2 and q_3 we have $(q_1 * q_2) * q_3 = q_1 * (q_2 * q_2)$ and $\overline{q_1 * q_2} = \overline{q_2} * \overline{q_1}$. Further we can explicitly express $*$ by

$$q_1 * q_2 = \begin{pmatrix} \mathrm{Re}(q_1)\,\mathrm{Re}(q_2) - \mathrm{Im}\,m(q_1) \cdot \mathrm{Im}(q_2) \\ \mathrm{Im}(q_1) \times \mathrm{Im}(q_2) + \mathrm{Re}(q_1)\,\mathrm{Im}(q_2) + \mathrm{Re}(q_2)\,\mathrm{Im}(q_1) \end{pmatrix},$$

where $\mathrm{Re}(q)$ and $\mathrm{Im}(q)$ denote the real- and imaginary part of q.

2.1. Eigenaxis Slews

An attitude slew is a time profile $q(t)\colon [t_0, t_1] \to \mathbb{H}$ connecting two orientations in 3-space. The rotation slew of a rigid body has therefore the state vector $x = q = (q_s, q_x, q_y, q_z) \in \mathbb{H}$. q_s is the scalar part of the quaternion and q_x, q_y and q_z indicate the vector parts. As control variable u the angular velocity $\omega = (\omega_x, \omega_y, \omega_z)$ is chosen. The kinematic equation of the rotational movement can be written as

$$\dot{x} = \dot{q} = \tfrac{1}{2}\omega * q = F(\omega\,,\, q)\,.$$

A constant of integration is given namely by the length of the quaternion q:

Lemma 2.1. [4] *Let $\omega \in \mathcal{C}(\mathbb{R}, \mathbb{R}^3 \subset \mathbb{H})$ be given. Then for the solution $q \in \mathcal{C}^1(\mathbb{R}, \mathbb{H})$ of the differential equation $\dot{q} = \frac{1}{2}\,\omega * q$ we get $\|q(t)\| = \|q(t_0)\|$ $\forall t$.*

This does exclude $\|q\|^2 = 1$ from the design as a cost term for reducing the duration of the slew – it is automatically built in. The cost function we choose is therefore $L = \|\omega\|^2 = \omega_x^2 + \omega_y^2 + \omega_z^2$. This results in the Hamiltonian

$$H \;=\; L + \lambda^\top F = \|\omega\|^2 + \tfrac{1}{2}\,\lambda^\top \omega * q$$

and the Hamilton equations are (see also [4])

$$\begin{aligned} \dot{q} &= +\tfrac{\partial H}{\partial \lambda} = +\tfrac{1}{2} \cdot \omega * q \\ \dot{\lambda} &= -\tfrac{\partial H}{\partial q} = -\tfrac{1}{2} \cdot \overline{\omega} * \lambda\,. \end{aligned}$$

From the Pontyagin maximum principle follows

$$
\begin{aligned}
& 0 = \tfrac{\partial H}{\partial u} = \tfrac{\partial H}{\partial \omega} = 2\omega + \tfrac{1}{2} \cdot \lambda * \overline{q} \\
\Rightarrow \quad & \omega = -\tfrac{1}{4} \cdot \lambda * \overline{q} \\
\Rightarrow \quad & \dot{\omega} = -\tfrac{1}{4} \left(\dot{\lambda} * \overline{q} + \cdot \lambda * \dot{\overline{q}} \right) \\
& \quad = +\tfrac{1}{8} \left([\overline{\omega} * \lambda] * \overline{q} - \lambda * \overline{[\omega * q]} \right) = \tfrac{1}{8} \left(\overline{\omega} * [\lambda * \overline{q}] - [\lambda * \overline{q}] * \overline{\omega} \right) \\
& \quad = -\tfrac{1}{2} \left(\overline{\omega} * \omega - \omega * \overline{\omega} \right) = -\tfrac{1}{2} \left(\|\omega\|^2 - \|\omega\|^2 \right) = 0 \quad \text{i.e.} \quad \boxed{\dot{\omega} = 0} .
\end{aligned}
$$

Lemma 2.2. *The unconstraint optimal control slew connecting two attitudes* q_1 *and* q_2 *is an eigenaxis slew with constant angular velocity.*

In [7] the same result can be found, formulated in the language of Lie theory. For now i.e. in this paper we will not make use of this formalism, since we are in the end interested in numerical solution schemes and do not see the benefit at this point. Nevertheless with respect to the the Euler-Poincaré equations it could be beneficial to consider this in the future. Although the analytic solution can be explicitly stated, it is interesting to note that the numerical integrators do preserve the constant of integration $\|q\|$ flawlessly.

2.2. Geometric Optimal Slews

Geometric and dynamic constraints often lead to cost terms that contradict each other. This can be easily demonstrated by the means of examples. Therefore they shall not be mixed as optimization terms. A rather stepwise approach by first constructing a geometrically optimal path and then use e.g. weight functions like in [8] for optimizing the dynamics and speed is suggested. This idea is related to engineering solutions where the relative slow motion of the celestial bodies is completely neglected. So we consider the rotational motion of a rigid body with the state vector as

$$x = (q, \omega) = (q_s, q_x, q_y, q_z, \omega_x, \omega_y, \omega_z) \in \mathbb{R}^7 \ .$$

As a control u a torque term $T = (t_x, t_y, t_z)^\top$ is used and the kinematic equation is:

$$\dot{x} = (\dot{q}, \dot{\omega}) = F\,(\omega\ ,\ q) = (\tfrac{1}{2} \cdot \omega * q, T)$$

The cost function chosen is $L = \|T\|^2 = t_x^2 + t_y^2 + t_z^2$. Then

$$H = L + \lambda^\top F = \|T\|^2 + \lambda_1^\top \cdot \tfrac{1}{2} \cdot \omega * q + \lambda_2^\top \cdot T$$

and the respective Hamilton equations are:

$$
\begin{aligned}
\begin{pmatrix} \dot{q} \\ \dot{\omega} \end{pmatrix} &= +\frac{\partial H}{\partial \lambda} = \begin{pmatrix} \frac{1}{2} \cdot \omega * q \\ T \end{pmatrix} \\
\begin{pmatrix} \dot{\lambda}_1 \\ \dot{\lambda}_2 \end{pmatrix} &= -\frac{\partial H}{\partial x} = \begin{pmatrix} -\frac{1}{2} \cdot \overline{\omega} * \lambda_1 \\ -\frac{1}{2} \cdot \lambda_1 * \overline{q} \end{pmatrix}
\end{aligned}
$$

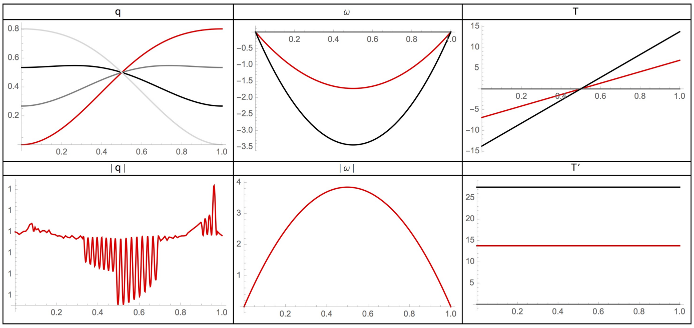
From the Pontyagin maximum principle follows

$$
\begin{aligned}
& 0 = \tfrac{\partial H}{\partial u} = \tfrac{\partial H}{\partial T} = 2T + \lambda_2 \\
\Rightarrow\quad & T = -\tfrac{1}{2} \cdot \lambda_2 = \dot{\omega} \\
\Rightarrow\quad & \ddot{\omega} = -\tfrac{1}{2} \cdot \dot{\lambda}_2 = \tfrac{1}{4} \lambda_1 * \overline{q} \\
\Rightarrow\quad & \dddot{\omega} = +\tfrac{1}{4} \left(\dot{\lambda}_1 * \overline{q} + \lambda_1 * \dot{\overline{q}} \right) \\
& \quad = -\tfrac{1}{8} \left([\overline{\omega} * \lambda_1] * \overline{q} - \lambda_1 * \overline{[\omega * q]} \right) = -\tfrac{1}{8} \left(\overline{\omega} * [\lambda_1 * \overline{q}] - [\lambda_1 * \overline{q}] * \overline{\omega} \right) \\
& \quad = +\tfrac{1}{4} \left(\overline{\omega} * \dot{\lambda}_2 - \dot{\lambda}_2 * \overline{\omega} \right) = -\tfrac{1}{2} \left(\overline{\omega} * \ddot{\omega} - \ddot{\omega} * \overline{\omega} \right) .
\end{aligned}
$$

Theorem 2.3 **(ω - ode).** *The unconstraint optimal control slew connecting two attitudes q_1, ω_1 and q_2, ω_2 is governed by the an angular velocity ω for which*

$$\boxed{\dddot{\omega} = -\tfrac{1}{2} \left(\overline{\omega} * \ddot{\omega} - \ddot{\omega} * \overline{\omega} \right)} \quad \text{or equivalent} \quad \boxed{\dddot{\omega} = \omega \times \ddot{\omega}} \quad \text{holds.}$$

Example 2.4. Shown is a geometric slew connecting the initial and final state $(q(0), \omega(0)) = (\frac{1}{\sqrt{14}}(0, 1, 2, 3)^\top, 0)$ and $(q(1), \omega(1)) = (\frac{1}{\sqrt{14}}(3, 2, 1, 0)^\top, 0)$:



2.3. Constraint Optimal Slews

If we add integral terms to the dynamics of the slew, additional constraints can be considered. For the motion of a rigid body, the state vector then becomes

$$x = (q, \omega, c) = (q_s, q_x, q_y, q_z, \omega_s, \omega_x, \omega_y, \omega_z, c) \in \mathbb{R}^{8+m}$$

$m = 1$ or 2. As control we again consider a torque $T = (t_x, t_y, t_z)^\top$ and the kinematic equation is:

$$\dot{x} = (\dot{q}, \dot{\omega}, \dot{c}) = F\,(\omega\;, \,q) = (\tfrac{1}{2} \cdot \omega * q, T, C(q, \omega))$$

As cost function we choose $L = c^2 + \|T\|^2 = c^2 + t_s^2 + t_x^2 + t_y^2 + t_z^2$. Then

$$H = L + \lambda^\top F = c^2 + \|T\|^2 + \lambda_1^\top \cdot \tfrac{1}{2} \cdot \omega * q + \lambda_2^\top \cdot T + \lambda_3^\top \cdot C(q, \omega)$$

and the Hamilton equations are:

$$\begin{pmatrix} \dot{q} \\ \dot{\omega} \\ \dot{c} \end{pmatrix} = +\frac{\partial H}{\partial \lambda} = \begin{pmatrix} \frac{1}{2}\cdot \omega * q \\ T \\ C(q,\omega) \end{pmatrix}$$

$$\begin{pmatrix} \dot{\lambda}_1 \\ \dot{\lambda}_2 \\ \dot{\lambda}_3 \end{pmatrix} = -\frac{\partial H}{\partial x} = \begin{pmatrix} -\frac{1}{2}\cdot \overline{\omega} * \lambda_1 - \lambda_3^\top \cdot \frac{\partial}{\partial q} C(q,\omega) \\ -\frac{1}{2}\cdot \lambda_1 * \overline{q} - \lambda_3^\top \cdot \frac{\partial}{\partial \omega} C(q,\omega) \\ 2c \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}\cdot \overline{\omega} * \lambda_1 - C_q \\ -\frac{1}{2}\cdot \lambda_1 * \overline{q} - C_\omega \\ 2c \end{pmatrix}$$

From the Pontyagin maximum principle follows with $\dot{\lambda}_2 = -\frac{1}{2}\lambda_1 * \overline{q} - C_\omega$ and $\dot{\omega} = -\frac{1}{2}\lambda_2$:

$$\begin{aligned} 0 &= \tfrac{\partial H}{\partial u} = \tfrac{\partial H}{\partial T} = 2T + \lambda_2 \\ &\Rightarrow \; T = -\tfrac{1}{2}\cdot \lambda_2 = \dot{\omega} \\ &\Rightarrow \; \ddot{\omega} = -\tfrac{1}{2}\cdot \dot{\lambda}_2 = \tfrac{1}{4}\lambda_1 * \overline{q} + \tfrac{1}{2} C_\omega \\ &\Rightarrow \; \dddot{\omega} = +\tfrac{1}{4}\left(\dot{\lambda}_1 * \overline{q} + \lambda_1 * \dot{\overline{q}}\right) + \tfrac{1}{2}\dot{C}_\omega \\ &\qquad = -\tfrac{1}{8}\left(\overline{\omega} * [\lambda_1 * \overline{q}] + 2C_q * \overline{q} - [\lambda_1 * \overline{q}] * \overline{\omega}\right) + \tfrac{1}{2}\dot{C}_\omega \\ &\qquad = +\tfrac{1}{8}\left(\overline{\omega} * [2C_\omega + 2\dot{\lambda}_2] - 2C_q * \overline{q} - [2C_\omega + 2\dot{\lambda}_2] * \overline{\omega}\right) + \tfrac{1}{2}\dot{C}_\omega \\ &\qquad = +\tfrac{1}{4}\left(\overline{\omega} * [C_\omega + \dot{\lambda}_2] - C_q * \overline{q} - [C_\omega + \dot{\lambda}_2] * \overline{\omega}\right) + \tfrac{1}{2}\dot{C}_\omega \\ &\qquad = -\tfrac{1}{2}\left(\overline{\omega} * \ddot{\omega} - \ddot{\omega} * \overline{\omega}\right) + \tfrac{1}{4}\left(\overline{\omega} * C_\omega - C_q * \overline{q} - C_\omega * \overline{\omega}\right) + \tfrac{1}{2}\dot{C}_\omega \end{aligned}$$

Theorem 2.5. *With the above definitions we have*

$$\boxed{\dddot{\omega} = \tfrac{1}{2}\left(\overline{\omega} * \ddot{\omega} - \ddot{\omega} * \overline{\omega}\right) + \tfrac{1}{4}\left(\overline{\omega} * C_\omega - C_q * \overline{q} - C_\omega * \overline{\omega}\right) + \tfrac{1}{2}\dot{C}_\omega}$$

and for $C(q,\omega) = C(q)$ *we have* $\boxed{\dddot{\omega} = -\tfrac{1}{2}\left(\overline{\omega} * \ddot{\omega} - \ddot{\omega} * \overline{\omega}\right) - \tfrac{1}{4} C_q * \overline{q}}$.

We want to derive a geometric form of the disturbance term $C(q)$ and to this end we need the following well known fact:

Lemma 2.6. [10] *For a orthogonal matrix* $A \in \mathbb{R}^{3\times 3}$ *and the respective quaternion* $q_A \in \mathbb{H}$ *and a vector* $x \in \mathbb{R}^3$ *we have for the coordinate change from inertial coordinates* x_{in} *to satellite coordinates* x_{sc} *:*

$$x_{sc} = A \cdot x_{in} = \overline{q}_A * x_{in} * q_A.$$

Theorem 2.7. *Let* $a_{sc}(t) = a_{sc}$ *be fixed in spacecraft frame, and* $b_{in}(t) = b_{in}$ *be fixed in inertial frame. For* $C(q) = \langle a_{in}, b_{in}\rangle - c_0 = \langle a_{sc}, b_{sc}\rangle - c_0$ *we have:*

$$\boxed{C_q * \overline{q} = -2\, b_{in} * a_{in} = 2\begin{pmatrix} \langle a_{in}\,,\, b_{in}\rangle \\ a_{in} \times b_{in} \end{pmatrix} = 2\begin{pmatrix} \langle a_{sc}\,,\, b_{sc}\rangle \\ a_{in} \times b_{in} \end{pmatrix}}\;.$$

Proof.

$$\begin{aligned} C_{q_i} &= \tfrac{\partial}{\partial q_i}[\langle a_{\rm sc}, b_{\rm sc}\rangle - c_0] = \tfrac{\partial}{\partial q_i}[a_{\rm sc}\cdot(\overline{q} * b_{\rm in} * q)] = a_{\rm sc}\cdot \tfrac{\partial}{\partial q_i}[\overline{q} * b_{\rm in} * q] \\ &= [\tfrac{\partial}{\partial q_i}\overline{q} * b_{\rm in} * q + \overline{q} * b_{\rm in} * \tfrac{\partial}{\partial q_i} q]\cdot a_{\rm sc} \\ &= [\overline{e}_i * b_{\rm in} * q]\cdot a_{\rm sc} + [\overline{q} * b_{\rm in} * e_i]\cdot a_{\rm sc} \end{aligned}$$

With the notation

$$\begin{aligned} R(q) &:= \big(\overline{e}_1 * q, \overline{e}_2 * q, \overline{e}_3 * q, \overline{e}_4 * q\big)^\top \quad \text{and} \\ L(q) &:= \big(\overline{q} * e_1, \overline{q} * e_2, \overline{q} * e_3, \overline{q} * e_4\big)^\top \end{aligned}$$

we can write $R(q_1)\cdot q_2 = q_1 * \overline{q}_2$ and $L(\overline{q_1})\cdot q_2 = q_1 * q_2$. Since the complex conjugate of a vector in 3-space is $\overline{a}_{\rm sc} = -a_{\rm sc}$ and $\overline{b}_{\rm in} = -b_{\rm in}$ we finally have

$$\begin{aligned} C_q * \overline{q} &= [\begin{pmatrix} \overline{e}_1 * b_{\rm in} * q \\ \overline{e}_2 * b_{\rm in} * q \\ \overline{e}_3 * b_{\rm in} * q \\ \overline{e}_4 * b_{\rm in} * q \end{pmatrix}\cdot a_{\rm sc} + \begin{pmatrix} \overline{q} * b_{\rm in} * e_1 \\ \overline{q} * b_{\rm in} * e_2 \\ \overline{q} * b_{\rm in} * e_3 \\ \overline{q} * b_{\rm in} * e_4 \end{pmatrix}\cdot a_{\rm sc}\ ] * \overline{q} \\ &= [R(b_{\rm in} * q)\cdot a_{\rm sc} + L(\overline{\overline{q} * b_{\rm in}})\cdot a_{\rm sc}\ ] * \overline{q} \\ &= [b_{\rm in} * q * \overline{a}_{\rm sc} + \overline{\overline{q} * b_{\rm in}} * a_{\rm sc}\ ] * \overline{q} \\ &= [b_{\rm in} * q * \overline{a}_{\rm sc} + \overline{b}_{\rm in} * q * a_{\rm sc}\ ] * \overline{q} = -2\ b_{\rm in} * q * a_{\rm sc} * \overline{q} \\ &= -2\ b_{\rm in} * a_{\rm in}. \end{aligned}$$

□

Example 2.8. A slew with the prescribed constraint $\langle a_{\rm sc}, b_{\rm sc}\rangle = 0$ or $a_{sc} \perp b_{sc}$ and ω, $a_{\rm in}$, and $b_{\rm in} \in \mathbb{R}^3$ will have the following dynamics:

$$\begin{aligned} \dddot{\omega} &= -\tfrac{1}{2}\left(\overline{\omega} * \ddot{\omega} - \ddot{\omega} * \overline{\omega}\right) - \tfrac{1}{4} C_q * \overline{q} = -\tfrac{1}{2}\left(\overline{\omega} * \ddot{\omega} - \ddot{\omega} * \overline{\omega}\right) + \tfrac{1}{2}\ b_{\rm in} * a_{\rm in} \\ &= \tfrac{1}{2}[\begin{pmatrix} \langle \ddot{\omega}\,,\,\overline{\omega}\rangle \\ \ddot{\omega}\times\overline{\omega} \end{pmatrix} - \begin{pmatrix} \langle \overline{\omega}\,,\,\ddot{\omega}\rangle \\ \overline{\omega}\times\ddot{\omega} \end{pmatrix}] + \tfrac{1}{2}\begin{pmatrix} \langle a_{\rm sc}\,,\,b_{\rm sc}\rangle \\ a_{\rm in}\times b_{\rm in} \end{pmatrix} = \omega\times\ddot{\omega} + \tfrac{1}{2}\begin{pmatrix} \langle a_{\rm sc}\,,\,b_{\rm sc}\rangle \\ a_{\rm in}\times b_{\rm in} \end{pmatrix}. \end{aligned}$$

And if $\langle a_{\rm sc}, b_{\rm sc}\rangle = 0$ for $t \geq t_0$, then the dynamic simplifies to

$$\boxed{\dddot{\omega} = \omega\times\ddot{\omega} + \tfrac{1}{2} a_{\rm in}\times b_{\rm in}}\ .$$

This is a new ω-ode $\dddot{\omega} = \omega\times\ddot{\omega} + c(t)$ which is similar to the second constant of integration $\ddot{\omega} = \omega\times\dot{\omega} + c$.

2.4. Comparison with a Euler Angle Slew

In [6] a description of a slew maneuver using an appropriate reference frame and then perform three successive rotations is given:

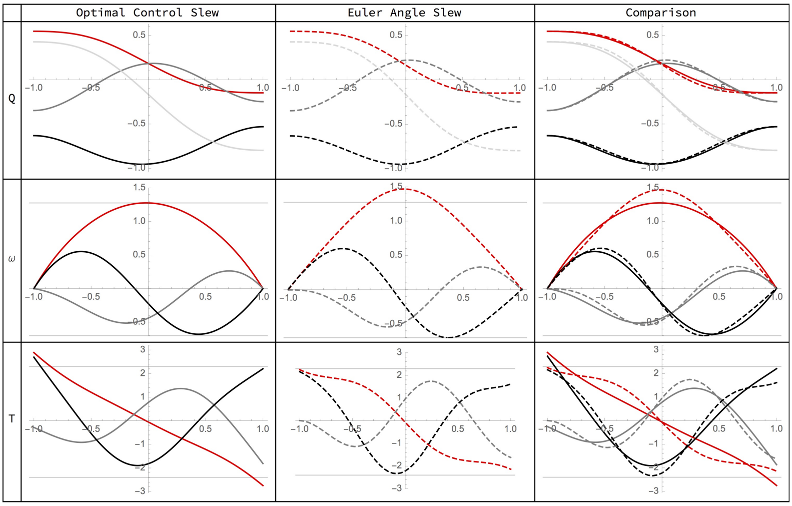
1. Rotation around the reference e_1-axis, i.e. the sun direction $s_{\rm in}$. Here the $y_{\rm sc}$-axis stays orthogonal to the sun line.
2. Rotation around the new reference e_2-axis, i.e. around the axis of the solar arrays $y_{\rm sc}$. Here the $y_{\rm sc}$-axis stays orthogonal to the sun line.

3. Rotation around the new reference e_3-axis. Here inconsistencies of the boundary values may compromise the orthogonality of the y_{sc}-axis with respect to the sun line.

The boundary values then determine the Euler angles and their derivative the slew is given by three cubic splines for these Euler angles and the respective rotation $R_3(\eta_3(t)) \cdot R_2(\eta_2(t)) \cdot R_1(\eta_1(t))$.

Example 2.9. The following graphs show a comparison of the constraint optimal slew ($s_{\mathrm{in}} \perp y_{\mathrm{sc}}$) and the Euler angle slew. The values that have been used are the fixed inertial sun direction $s_{\mathrm{in}} = (-0.930975, -0.344742, -0.120159)$ and

$$q_{\mathrm{ini}} = (0.546232, -0.34778, -0.631268, 0.426827), \quad \omega_{\mathrm{ini}} = (0, -0.000012, 0),$$
$$q_{\mathrm{fin}} = (0.148181, \quad 0.24793, \quad 0.530584, 0.796904), \quad \omega_{\mathrm{fin}} = (0, \quad 0.000012, 0).$$



The constraint optimal slew has overall lower rates, but higher torques at both ends of the interval. Note for the constraint optimal slew additional constraints could still be added.

3. Dynamics of the Angular Velocity

In this section we perform further investigations of dynamics of angular velocity. Two types of solution families are described. The relation of these solutions to the two following constants of integration is described.

3.1. Analytic Solutions

Theorem 3.1. *Let the following initial value problem (IVP)*

$$\boxed{\dddot{\omega}(t) = \omega(t) \times \ddot{\omega}(t)}, \quad \omega(t_0) = \omega_1, \quad \dot{\omega}(t_0) = \omega_2, \quad \ddot{\omega}(t_0) = \omega_3$$

with $\omega(t) \in \mathbb{R}^3$ and $t \in [t_0, t_1] \subset \mathbb{R}$ be given. Then there exist the following two constants of integration:

1. $\forall\, t \in [t_0, t_1]\colon\ \ \|\ddot\omega(t)\| = \|\ddot\omega(t_0)\|$ *and*
2. $\forall\, t \in [t_0, t_1]\colon\ \ \ddot\omega(t) - \omega(t) \times \dot\omega(t) = \ddot\omega(t_0) - \omega(t_0) \times \dot\omega(t_0)$

Proof. 1. $\frac{d}{dt}\|\ddot\omega(t)\|^2 = 2\ddot\omega(t) \cdot \dddot\omega(t) = 2\ddot\omega(t) \cdot (\omega(t) \times \ddot\omega(t)) = 0.$

2. $\frac{d}{dt}[\ddot\omega(t) - \omega(t) \times \dot\omega(t)] = \dddot\omega(t) - \underbrace{\dot\omega(t) \times \dot\omega(t)}_{=0} - \omega(t) \times \ddot\omega(t) = 0.$ □

Theorem 3.2 (Quadratic Solutions). *For the IVP $\dddot\omega(t) = \omega(t) \times \ddot\omega(t)$ with $\omega(t_0) = \omega_1$, $\dot\omega(t_0) = \omega_2$, $\ddot\omega(t_0) = \omega_3$ and $\omega(t) \in \mathbb{R}^3$, $t \in [t_0, t_1] \subset \mathbb{R}$ we have:*

1. *If the initial values ω_1, ω_2 and ω_3 are colinear, i.e. $c_1 \cdot \omega_1 = c_2 \cdot \omega_2 = c_3 \cdot \omega_3$ for c_1, c_2 and $c_3 \in \mathbb{R}$, then $\dddot\omega(t) \equiv 0$ and the $\omega_i(t)$ stay for $t \in [t_0, t_1] \subset \mathbb{R}$ colinear. Further the solution of the ordinary differential equation in this case is given by*
$$\omega(t) := \omega_1 + \omega_2(t - t_0) + \frac{\omega_3}{2}(t - t_0)^2\ .$$
2. *If two components of a solution of the ordinary differential equation are linear dependent, so is the third component. And therefore this is a quadratic solution.*

Proof. 1. Since the differential equation is Lipschitz continuous and apparently $\omega(t)$ as given above is a solution of the IVP with $\omega(t_0) = \omega_1$, $\dot\omega(t_0) = \omega_2$, $\ddot\omega(t_0) = \omega_3$ and $\dddot\omega(t) \equiv 0$ the claim follows.

2. Let $d \in \mathbb{R}$ be given, without loss of generality we assume
$$\begin{pmatrix} \omega_x(t) \\ d\,\omega_x(t) \\ \omega_z(t) \end{pmatrix} \times \begin{pmatrix} \ddot\omega_x(t) \\ d\,\ddot\omega_x(t) \\ \ddot\omega_z(t) \end{pmatrix} = \begin{pmatrix} d\,[\omega_x(t)\ddot\omega_z(t) - \omega_z(t)\ddot\omega_x(t)] \\ -[\omega_x(t)\ddot\omega_z(t) - \omega_z(t)\ddot\omega_x(t)] \\ 0 \end{pmatrix} = \begin{pmatrix} \dddot\omega_x(t) \\ d\,\dddot\omega_x(t) \\ \dddot\omega_z(t) \end{pmatrix}$$
and get from $\dddot\omega_z(t) = 0$ and $\dddot\omega_x(t) = -d^2\dddot\omega_x(t)$ solutions of the form
$$\omega_x(t) = a_0 + a_1\,t + a_2\,t^2 \text{ and } \omega_z(t) = b_0 + b_1\,t + b_2\,t^2.$$ □

Remark 3.3. Observe that the quadratic functions for ω become cubic when the integration to a quaternion profile is considered.

Example 3.4. A non zero solution of the ω - ode with boundary values zero at $t = 1$ and $t = 2$ is given by $\omega(t) = (0, 0, 0)^\top + (1, 2, 3)^\top[t - 1] - (2, 4, 6)^\top \frac{[t-1]^2}{2}$.

Theorem 3.5 (Periodic Solutions). 1. *For a_0, a_1, $a_2 \in \mathbb{R}$ the differential equation $\dddot\omega(t) = \omega(t) \times \ddot\omega(t)$ has the following periodic solutions on $[t_0, t_1] \subset \mathbb{R}$:*
$$\omega_1(t) = \begin{pmatrix} a_1 \\ a_0\cos(a_1 t + a_2) \\ a_0\sin(a_1 t + a_2) \end{pmatrix}, \quad \omega_{-1}(t) = \begin{pmatrix} -a_1 \\ a_0\sin(a_1 t + a_2) \\ a_0\cos(a_1 t + a_2) \end{pmatrix},$$
$$\omega_2(t) = \begin{pmatrix} a_0\sin(a_1 t + a_2) \\ a_1 \\ a_0\cos(a_1 t + a_2) \end{pmatrix}, \quad \omega_{-2}(t) = \begin{pmatrix} a_0\cos(a_1 t + a_2) \\ -a_1 \\ a_0\sin(a_1 t + a_2) \end{pmatrix},$$
$$\omega_3(t) = \begin{pmatrix} a_0\cos(a_1 t + a_2) \\ a_0\sin(a_1 t + a_2) \\ a_1 \end{pmatrix}, \quad \omega_{-3}(t) = \begin{pmatrix} a_0\sin(a_1 t + a_2) \\ a_0\cos(a_1 t + a_2) \\ -a_1 \end{pmatrix}.$$

2. *For all these solutions the second constant of integration has the form*

$$\ddot{\omega} = \omega \times \ddot{\omega} \pm a_0^2 a_1 e_i.$$

3. *There are no other solutions of the form*

$$\omega(t) = \begin{pmatrix} \cos(f(t)) \\ g(t) \\ \sin(f(t)) \end{pmatrix}.$$

Proof. A straight forward calculation shows 1. and 2. by inspection. To show 3. note first that

$$\dddot{\omega} - \omega \times \ddot{\omega} = \begin{pmatrix} * \\ \dddot{g}(t) + \ddot{f}(t) \\ * \end{pmatrix},$$

so that $f(t) = -\dot{g}(t) + c_0 + c_1 t$. With

$$\omega(t) = \begin{pmatrix} \cos(-\dot{g}(t) + c_0 + c_1 t) \\ g(t) \\ \sin(-\dot{g}(t) + c_0 + c_1 t) \end{pmatrix}$$

and

$$\begin{aligned} f_1(t) &:= \dddot{g}(t)\,[g(t) + 3(c_1 - \ddot{g}(t))] \\ f_2(t) &:= g(t)\,(c_1 - \ddot{g}(t))^2 + (c_1 - \ddot{g}(t))^3 + \ddot{g}(t) + \ddddot{g}(t) \end{aligned}$$

we get

$$\begin{aligned} \dddot{\omega} - \omega \times \ddot{\omega} &= \begin{pmatrix} f_1(t)\cos(-\dot{g}(t) + c_0 + c_1 t) + f_2(t)\sin(-\dot{g}(t) + c_0 + c_1 t) \\ 0 \\ f_1(t)\sin(-\dot{g}(t) + c_0 + c_1 t) - f_2(t)\cos(-\dot{g}(t) + c_0 + c_1 t) \end{pmatrix} \\ &= \begin{pmatrix} f_1(t)\cos(\ldots) + f_2(t)\sin(\ldots) \\ 0 \\ f_1(t)\sin(\ldots) - f_2(t)\cos(\ldots) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

From the last equation it can be seen, that $f_2(t) = f_1(t)\frac{\sin(\ldots)}{\cos(\ldots)}$ and therefore:

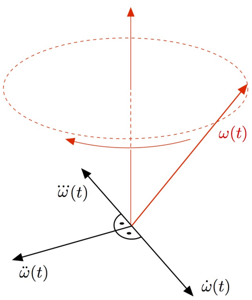
$$0 = f_1(t)\cos(\ldots)^2 + f_2(t)\sin(\ldots)\cos(\ldots) = f_1(t)[\cos(\ldots)^2 + \sin(\ldots)^2] = f_1(t)\ .$$

The differential equation $f_1(1) = \dddot{g}(t)\,[g(t) + 3(c_1 - \ddot{g}(t))] = 0$ has the two solutions $g_1(t) = b_0 + b_1 t + b_2 t^2$ and $g_2(t) = -3c_1 + b_0 e^{\frac{t}{\sqrt{3}}} + b_1 e^{-\frac{t}{\sqrt{3}}}$. The solution g_1 implies $f_2(t) = 0$ for $c_1 = 0$ and $b_2 = 0$ which results in quadratic solutions. g_2 is not a solution of the second differential equation. □

Example 3.6. With the initial values

$$\omega(t_0) = (0,1,1)^\top,\ \dot{\omega}(t_0) = (1,0,0)^\top,\ \ddot{\omega}(t_0) = -(0,0,1)^\top$$

a periodic motion is performed by the solution of the ω - ode. The vectors $\dot{\omega}(t)$, $\ddot{\omega}(t)$ and $\dddot{\omega}(t)$ lie in a plane and are rotated 90° each and $\dot{\omega}(t)$ points towards $-\dddot{\omega}(t)$.



$\omega(t)$ moves on a 45° cone and all three derivatives rotate in the common plane.

3.2. Properties of the Solutions

Now we look at a more general type of solutions, where yet no analytic representation is known to the author.

Theorem 3.7 (Step Response Solution). *If for the first constant of integration holds*

$$\ddot{\omega}(t_0) - \omega(t_0) \times \dot{\omega}(t_0) = 0,$$

then we get with $\hat{\omega} = \frac{\omega}{\|\omega\|}$*:*

1. *We have* $\dot{\omega}(t) \cdot \ddot{\omega}(t) = 0$ *and the vectors* $\omega(t)$, $\ddot{\omega}(t)$ *and* $\dddot{\omega}(t)$ *form an orthogonal basis i.e.* $\forall\ t \in [t_0, t_1]:\ \omega(t) \cdot \ddot{\omega}(t) = \dddot{\omega}(t) \cdot \ddot{\omega}(t) = \dddot{\omega}(t) \cdot \omega(t) = 0.$
2. *We have* $\|\dot{\omega}(t)\| \equiv \|\dot{\omega}(t_0)\| = c_1$, $\|\ddot{\omega}(t)\| \equiv \|\ddot{\omega}(t_0)\| = c_2$ *and*
$$\|\dddot{\omega}(t)\| = c_2 \cdot \|\omega(t)\|\ \forall\ t \in [t_0, t_1].$$
3. $\|\omega(t)\| = \sqrt{(c_1 \cdot (t \pm c))^2 + \left(\frac{c_2}{c_1}\right)^2}, \quad c = c(\|\omega(t_0)\|) \in \mathbb{R}.$
4. $|1 - (\hat{\omega} \cdot \hat{\dot{\omega}})^2| \le O(\frac{1}{\|\omega(t)\|^2})$ *for* $t \to \infty$ *i.e.* $\measuredangle(\hat{\omega}, \hat{\dot{\omega}}) \to 0$ *for* $t \to \infty$.

Proof. First we get

1. $\ddot{\omega} = \omega \times \dot{\omega} \quad \Rightarrow \quad \omega \cdot \ddot{\omega} = \dot{\omega} \cdot \ddot{\omega} = \frac{1}{2} \cdot \frac{d}{dt}\|\dot{\omega}(t)\|^2 = 0$ and $\|\dot{\omega}(t)\| = c_1$.
2. $\dddot{\omega} = \omega \times \ddot{\omega} \quad \Rightarrow \quad \omega \cdot \dddot{\omega} = \ddot{\omega} \cdot \dddot{\omega} = \frac{1}{2} \cdot \frac{d}{dt}\|\ddot{\omega}(t)\|^2 = 0$ and $\|\ddot{\omega}(t)\| = c_2$.

For the Norms of the derivatives of ω then holds

1. $\|\omega(t)\|^2 =: f^2(t)$.
2. $\|\dot{\omega}(t)\|^2 = \|\dot{f}\hat{\omega} + f\dot{\hat{\omega}}\|^2 = \dot{f}^2\|\hat{\omega}\|^2 + 2f\dot{f} \cdot \hat{\omega} \cdot \dot{\hat{\omega}} + f^2\|\dot{\hat{\omega}}\|^2 = \dot{f}^2 + f^2\|\dot{\hat{\omega}}\|^2$.
3. $\|\ddot{\omega}(t)\|^2 = \|\omega \times \dot{\omega}\|^2 = f^4\|\hat{\omega} \times \dot{\hat{\omega}}\|^2 = f^4\|\hat{\omega}\|^2\|\dot{\hat{\omega}}\|^2 = f^4\|\dot{\hat{\omega}}\|^2$.
4. $\|\dddot{\omega}(t)\|^2 = \|\omega \times \ddot{\omega}\|^2 = \|\omega\|^2\|\ddot{\omega}\|^2 = f^6 \cdot \|\dot{\hat{\omega}}\|^2$, da $\omega \cdot \ddot{\omega} = 0$.

For the unit vector $\hat{\omega}(t)$ we get further

$$\hat{\omega}(t) = \begin{pmatrix} \sin(\vartheta(t))\cos(\varphi(t)) \\ \sin(\vartheta(t))\sin(\varphi(t)) \\ \cos(\vartheta(t)) \end{pmatrix},$$

$$\partial_\vartheta \hat{\omega}(t) = \begin{pmatrix} \cos(\vartheta(t))\cos(\varphi(t)) \\ \cos(\vartheta(t))\sin(\varphi(t)) \\ -\sin(\vartheta(t)) \end{pmatrix},$$

$$\partial_\varphi \hat{\omega}(t) = \begin{pmatrix} -\sin(\vartheta(t))\sin(\varphi(t)) \\ \sin(\vartheta(t))\cos(\varphi(t)) \\ 0 \end{pmatrix}$$

with

$$\hat{\omega}(t) \perp \partial_\vartheta \hat{\omega}(t) \perp \partial_\varphi \hat{\omega}(t) \perp \hat{\omega}(t)$$

and

$$\|\partial_\vartheta \hat{\omega}(t)\| = 1,\ \|\partial_\varphi \hat{\omega}(t)\| = \sin(\vartheta(t))^2.$$

And finally

$$\begin{aligned} \Phi(t) &:= \|\dot{\hat{\omega}}\|^2 = \|\partial_\varphi \hat{\omega} \cdot \dot{\varphi} + \partial_\vartheta \hat{\omega} \cdot \dot{\vartheta}\|^2 \\ &= \|\partial_\varphi \hat{\omega}\|^2 \cdot \dot{\varphi}^2 + \underbrace{\|\partial_\vartheta \hat{\omega}\|^2}_{=1} \cdot \dot{\vartheta}^2 + \underbrace{\partial_\varphi \hat{\omega} \cdot \partial_\vartheta \hat{\omega}}_{=0} \cdot \dot{\varphi}\,\dot{\vartheta} \\ &= \sin(\vartheta(t))^2 \cdot \dot{\varphi}^2 + \dot{\vartheta}^2\,. \end{aligned}$$

From the equations

$$\|\dot{\omega}(t)\|^2 = \dot{f}^2 + f^2 \|\dot{\hat{\omega}}\|^2 = c_1^2$$

and

$$\|\ddot{\omega}(t)\|^2 = f^4 \cdot \|\dot{\hat{\omega}}\|^2 = c_2^2$$

we construct the differential equation

$$\dot{f}(t)^2 + \frac{c_2^2}{f(t)^2} = c_1^2$$

and using standard techniques we get

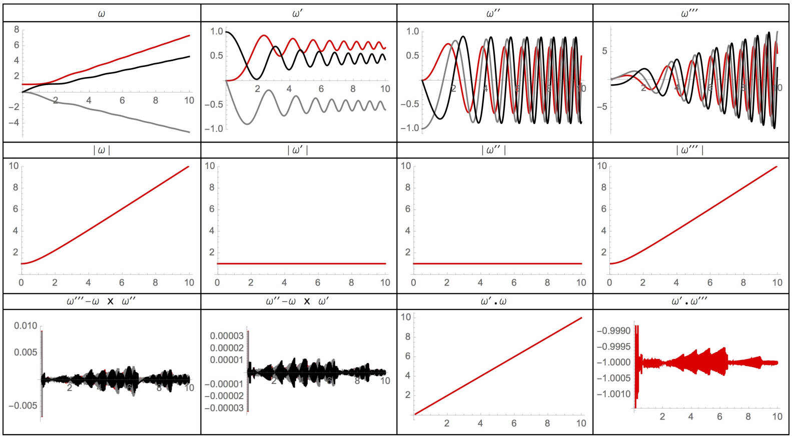
$$f(t) = \sqrt{c_1^2 (t-t_0)^2 + \left(\tfrac{c_2}{c_1}\right)^2}.$$

Finally

$$\begin{aligned} \|\hat{\omega} \times \hat{\dot{\omega}}\|^2 &= \|\omega \times \dot{\omega}\|^2 \frac{1}{\|\omega\|^2} \frac{1}{\|\dot{\omega}\|^2} = \frac{1}{\|\omega\|^2} \frac{\|\ddot{\omega}\|^2}{\|\dot{\omega}\|^2} = \frac{1}{\|\omega\|^2} \cdot \frac{c_2^2}{c_1^2} \\ &= \left[\|\omega\|^2 \|\dot{\omega}\|^2 - (\omega \cdot \dot{\omega})^2\right] \frac{1}{\|\omega\|^2} \frac{1}{\|\dot{\omega}\|^2} = 1 - (\hat{\omega} \cdot \hat{\dot{\omega}})^2 \\ &\Rightarrow\ 1 - (\hat{\omega} \cdot \hat{\dot{\omega}})^2 = \frac{1}{\|\omega\|^2} \cdot \frac{c_2^2}{c_1^2} = \sin^2[\measuredangle(\hat{\omega}, \hat{\dot{\omega}})] \end{aligned}$$

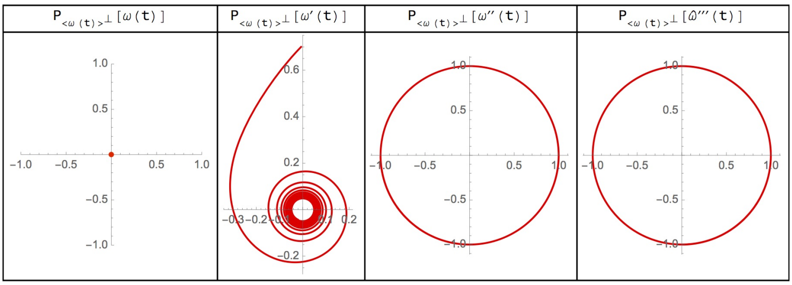
□

Example 3.8 (Step Response Solution). In this example the statements of the above theorem are confirmed. And it is clearly recognizable, that the norms of ω and its derivatives exhibit the expected and tranquil behavior.



$$t_0 = 0\ ,\ t_1 = 10\ ,\ \omega(t_0) = \begin{pmatrix}1\\0\\0\end{pmatrix},\ \dot\omega(t_0) = \begin{pmatrix}0\\0\\1\end{pmatrix},\ \ddot\omega(t_0) = -\begin{pmatrix}0\\1\\0\end{pmatrix}$$

Example 3.9 (Step Response Solution). A better representation of this type of solutions can be given by projecting all vectors to $\omega(t)^\perp$. But one has to beware that this is a moving frame and all calculations in this frame need the respective corrections.



$$t_0 = 0\ ,\ t_1 = 10\ ,\ \omega(t_0) = \begin{pmatrix}0\\1\\0\end{pmatrix},\ \dot\omega(t_0) = \begin{pmatrix}0\\1\\1\end{pmatrix},\ \ddot\omega(t_0) = -\begin{pmatrix}1\\0\\0\end{pmatrix}$$

Remark 3.10. The linearization of the ordinary differential equation has the characteristic polynomial

$$\boxed{\chi(x) = x^3\ (x^3 + \|\omega\, x - \ddot\omega\|^2)^2}$$

and thus the eigenvalues are $(0, 0, 0, \pm\sqrt{\xi_1}, \pm\sqrt{\xi_2}, \pm\sqrt{\xi_3})$, where ξ_1,ξ_2 and ξ_3 are the solutions of the equation

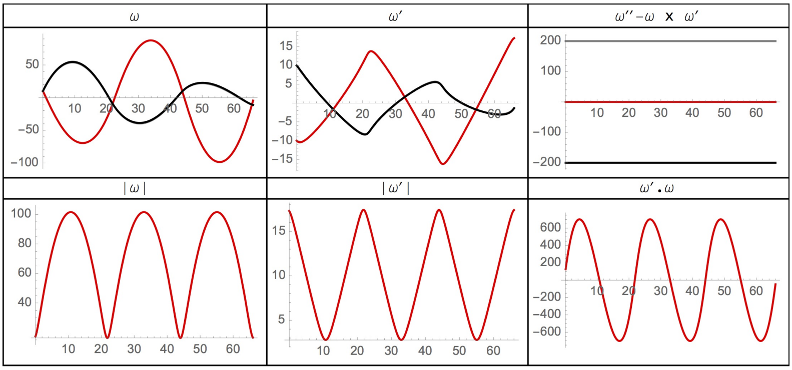
$$x^3 + \|\omega\, x - \ddot\omega\|^2 = x^3 + \|\omega\|^2\ x^2 - 2\,\omega\cdot\ddot\omega\ x + \|\ddot\omega\|^2 = 0$$

and are therefore determined by the terms $\|\omega\|$, $\|\ddot\omega\|$ and $\omega\cdot\ddot\omega$. Since

$$\frac{\|\dddot\omega\|^2}{\|\omega\|^2} = \|\ddot\omega\|^2 - \frac{\frac{1}{2}\frac{d}{dt}\|\dot\omega\|^2 - \frac{1}{2}\frac{d}{dt}\|\dot\omega(t_0)\|^2}{\|\omega\|^2}$$

holds, we have for $\frac{d}{dt}\|\dot\omega\| \approx 0$ that $\|\omega(t)\|\|\ddot\omega(t)\| \approx \|\dddot\omega(t)\|$, if $\|\omega(t)\| \to \infty$.

Example 3.11 (Quasi Periodic Solution). A particular type of solution is given when the second constant of integration is e.g. set to $(\delta\,,+\Delta\,,-\Delta)$, where $0 \leq \delta \ll \Delta$. Then the second and third component are close to each other (appearing almost identical) and stay this way over time, as can be seen in the figure. Since by Theorem 3.2 all three components must be pairwise different over time, these solutions appear therefore somewhat paradox. At the same time the values of $|\omega|$, $|\dot{\omega}|$ and $\dot{\omega}\cdot\omega$ exhibit a periodic behavior.



$$t_0 = 0\ ,\ t_1 = 66\ ,\ \omega(t_0) = \begin{pmatrix} 10 \\ 10 \\ 10 \end{pmatrix}\ ,\ \dot{\omega}(t_0) = \begin{pmatrix} -10 \\ 10 \\ 10 \end{pmatrix}\ ,\ \ddot{\omega}(t_0) = -\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

4. Summary

There are still several open questions. It has not yet been investigated, if it is possible and practical to build the gyroscopic term into the slew by $C(q,\omega)$. Unbalanced boundary conditions may prevent a corresponding solution. This is a very important topic for the practical usability of the whole approach. The details of the numerical methods especially the direct method using [1] have not been described. All this are topics for future work.

References

[1] Andersson, J.A.E., Gillis, J., Horn, G., Rawlings, J.B., Diehl, M., *CasADi – A software framework for nonlinear optimization and optimal control*, Mathematical Programming Computation, 2018.

[2] Bryson, A.E., Ho, Y.C., *Applied Optimal Control: Optimization, Estimation and Control*, 1981.

[3] Frazzoli, E., Dahleh, M.A., Feron, E., *A randomized attitude slew planning algorithm for autonomous spacecraft*, AIAA Conference, Montreal, 2001.

[4] Junkins, J.L., Turner, J.D. *Optimal Continuous Torque Attitude Maneuvers*, Virginia Polytechnic Institute and State University, Blacksburg, Va. 1980.

[5] Kirk, D.E., *Optimal Control Theory*, Dover, 2016.

[6] Rigger, R., Yde, J., Müller, M., Companys, V., *The optimization of attitude profiles for SMART-1: A highly constrained problem*, Proceedings of the 18th International Symposium on Space Flight Dynamics, 2004.

[7] Spindler, K., *Optimal control on Lie groups with applications to attitude control*, Math. Control Signals and Systems, **11**(1998), no. 3, 197-219.

[8] Spindler, K., *Attitude Maneuvers which avoid a forbidden direction*, Journal of Dynamical and Control Systems, **8**(2002), no. 1.

[9] Wang, L., *Model Predictive Control System Design and Implementation*, Springer, Advances in Industrial Control, 2009.

[10] Wertz, J., *Spacecraft Attitude Determination and Control*, Kluwer Academic Publishers, 1990.

[11] Wisniewski, R., Kulczycki, P., *Slew Maneuver Control for Spacecraft Equipped with Star Camera and Reaction Wheels*, Preprint submitted to Elsevier Science, 2003.

Ralf Rigger
Technische Hochschule Mittelhessen
Department Mathematik, Naturwissenschaften und Datenverarbeitung
Wilhelm-Leuschner-Straße 13
61169 Friedberg, Germany
e-mail: `ralf.rigger@mnd.thm.de`

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 279–289
DOI: 10.24193/subbmath.2019.2.12

Parameter estimations for linear parabolic fractional SPDEs with jumps

Wilfried Grecksch, Hannelore Lisei and Jens Lueddeckens

Abstract. We give an unbiased and consistent estimator for the drift coefficient of a linear parabolic stochastic partial differential equation driven by a multiplicative cylindrical fractional Brownian motion with Hurst index $1/2 < h < 1$ and a cylindrical centered Poisson process, if the observations of the solution process are given in discrete time points. The presented method is based on mean square estimations.

Mathematics Subject Classification (2010): 60H15, 62F12, 60G22.

Keywords: Parameter estimation, SPDE, cylindrical fractional Brownian motion, cylindrical Poisson process.

1. Introduction

There are lots of papers concerning parameter estimations for stochastic differential equations (SDEs). Results for SDEs driven by a fractional Brownian motion (fBm) were given, for example, by Y. Kozachenko, A. Melnikov, Y. Mishura [4] and W.L. Xiao, W.G. Zhang, X. Zhang [11] (see also the references therein). If the driving process is a Lévy process, then see, for example, H. Long [5] and the references therein. If the equations are driven by a fBm and a fractional Poisson measure one can find interesting results and applications in the PhD thesis of J. Lueddeckens [7].

Many papers are devoted to the parameter estimation of fractional stochastic partial differential equations (SPDEs). As a representative result we quote here the paper [8] of B. Maslowski and C.A. Tudor.

The following estimation criteria are mainly used in constructing estimators for the parameters of SPDEs:

- maximum likelihood type methods by considering fundamental martingales and theorems of Girsanov type;

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

- time continuous and time discrete least square criteria;
- Kalman-Bucy filters;
- L^1-norm estimations and contrast estimations.

Often SPDEs are considered as stochastic evolution equations in Hilbert spaces. For example, a parameter estimation problem for linear diagonalized stochastic partial differential equations driven by a multiplicative fBm is considered by I. Cialenco in [2]. The stochastic processes defined by the random Fourier coefficients of the solution process describe one dimensional geometric fractional Brownian motions. Based on these processes, consistent parameter estimates for the SPDEs are determined using a maximum likelihood type method. So we see, that one dimensional results of parameter estimations are useful for parameter estimations of SPDEs.

Parameter estimations for diagonal SPDEs are also considered in Chapter 6 in [6] by S.V. Lototsky, B.L. Rozovsky.

The aim of the present paper is to give new contributions in the estimation theory of the coefficients of linear homogeneous SPDEs, which are driven by cylindrical fractional Brownian motions and cylindrical Poisson processes. The applied estimation criterion uses covariances (as a generalization of the mean squared method), such that the long range dependence property of the fractional Brownian motion with Hurst index $h \in]1/2, 1[$ is taken into account. Moreover, in this paper weakly, respectively strongly consistent estimators are constructed by using only information about the underlying process in discrete time points.

The paper starts with a preliminary section, containing the assumptions needed throughout the paper. A linear SPDE driven by a multiplicative cylindrical fractional Brownian motion and a cylindrical Poisson process is introduced in Section 3. In Section 4 the one dimensional stochastic differential equations for the Fourier coefficients of the solution process of the SPDE is considered and similar to the results from [7] an estimation criterion of least squares type in discrete time points of the observations for the drift term is formulated. The estimator of the drift coefficient is unbiased. Conditions for choosing the time points are given such that the constructed estimator is unbiased and weakly consistent (Theorem 4.2), respectively strongly consistent (see Theorem 4.4).

2. Preliminaries

Definition 2.1. A real-valued Gaussian process $(B^h(t))_{t\geq 0}$ with $E(B^h(t)) = 0$, for all $t \geq 0$, $B^h(0) = 0$ and Hurst index $h \in]0, 1[$ is called fractional Brownian motion (fBm) if

$$E(B^h(t)B^h(s)) = \frac{1}{2}(t^{2h} + s^{2h} - |t-s|^{2h}) \text{ for all } s, t \geq 0.$$

The fBm is not a semimartingale and it is not a Markovian process for $h \neq 1/2$. The fBm is a Wiener process for $h = 1/2$. In this paper we consider $h \in]1/2, 1[$. Then, the fBm has the so-called long range dependence property.

Assumptions:

(A1) Let (V, H, V^*) be a triplet of rigged Hilbert spaces, where V is compactly embedded into H and $A : V \to V^*$ is linear and $\langle Av, v\rangle + \alpha_1 \|v\|_V^2 \leq \alpha_2 \|v\|_H^2$ for all $v \in V$ and $\alpha_1 > 0,\ \alpha_2 \in \mathbb{R}$ are constants.

Observe, that $A : D(A) \to H$ is linear and unbounded with $D(A) = \{v \in V : Av \in H\}$, which is dense in H. The eigenvalues $(\lambda_k)_{k\geq 1}$ of this operator are negative and satisfy $\lim\limits_{k\to\infty} \lambda_k = -\infty$.

(A2) Let $(h_k)_{k\geq 1} \subset H$ be the complete orthonormal system constructed by the eigenfunctions of A.

(A3) The C_0 semigroup $(\mathcal{T}_t)_{t\geq 0}$ defined by

$$\mathcal{T}_t(x) = \sum_{k=1}^{\infty} \exp\{\lambda_k t\}(x, h_k)h_k,\ x \in H$$

is generated by $-A$, where $(\cdot, \cdot)$ denotes the scalar product in H.

(A4) $\Phi_1, \Phi_2 :\ H \to H$ are Hilbert-Schmidt operators of the type

$$\Phi_i(x) = \sum_{k=1}^{\infty} \mu_{ik}(x, h_k)h_k, \quad x \in H,$$

where $\sum\limits_{k=1}^{\infty} \mu_{ik}^2 < \infty$ for $i \in \{1, 2\}$.

(A5) Let $(B_k^h(t))_{t\geq 0}$, $k \in \{1, 2, \ldots\}$, be independent fractional Brownian motions with Hurst index $h \in]1/2, 1[$ and let

$$B^h(t) = \sum_{k=1}^{\infty} B_k^h(t)h_k, \quad t \geq 0,$$

denote the cylindrical fBm.

(A6) Let $(\pi_j(t))_{t\geq 0}$, $j \in \{1, 2, \ldots\}$, be independent homogeneous Poisson processes with parameter ν.

(A7) Consider $\tilde{\pi}_j(t) = \pi_j(t) - \nu t,\ j \in \{1, 2, \ldots\}$ and we denote by

$$\tilde{\pi}(t) = \sum_{j=1}^{\infty} \tilde{\pi}_j(t)h_j, \quad t \geq 0,$$

the cylindrical centered Poisson process.

(A8) The processes B_k^h and $\tilde{\pi}_j$ are independent stochastic processes for all $j, k \in \{1, 2, ...\}$.

(A9) Let $X_0 \in H$ be a deterministic initial value.

(A10) All stochastic processes are defined on the same complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, where $\mathcal{F}_t = \sigma(\mathcal{F}_t^{B^h} \vee \mathcal{F}_t^{\tilde{\pi}})$ and $\mathcal{F}_t^{B^h}$ and $\mathcal{F}_t^{\tilde{\pi}}$ denote the σ-algebras generated by $(B^h(s))_{s\in[0,t[}$ and $(\tilde{\pi}(s))_{s\in[0,t[}$.

3. A linear fractional parabolic SPDE with jumps

At first we introduce for $k \in \{1, 2, \ldots\}$ the one dimensional linear stochastic differential equations

$$dY_k(t) = a\lambda_k Y_k(t)dt + \sigma\mu_{1k} Y_k(t)dB_k^h(t) + \eta\mu_{2k} Y_k(t-)d\tilde{\pi}_k(t), \tag{3.1}$$

where a, σ, η are positive constants and $Y_k(0) = (X_0, h_k)$.
The stochastic equation (3.1) is defined by

$$Y_k(t) = Y_k(0) + a\lambda_k \int_0^t Y_k(s)ds + \sigma\mu_{1k} \int_0^t Y_k(s)dB_k^h(s) + \eta\mu_{2k} \int_0^t Y_k(s-)d\tilde{\pi}_k(s), \tag{3.2}$$

for all $t \geq 0$, where the stochastic integral with respect to B_k^h is defined by a divergence integral as in [9] and the stochastic integral with respect to the compensated Poisson process is defined as in [10], Chapter II (or [3], page 246).

Theorem 3.1. *The process*

$$Y_k(t) = Y_k(0)(1 + \eta\mu_{2k})^{\pi_k(t)} \cdot \exp\left\{\sigma\mu_{1k} B_k^h(t) - \frac{1}{2}\sigma^2\mu_{1k}^2 t^{2h} + a\lambda_k t - \nu\eta\mu_{2k} t\right\} \tag{3.3}$$

solves equation (3.1) *for all* $t \geq 0$ *with probability 1.*

Proof. We prove that the process $Y_k(t) = Y_{1k}(t)Y_{2k}(t)$ with

$$Y_{1k}(t) = \exp\left\{\sigma\mu_{1k} B_k^h(t) - \frac{1}{2}\sigma^2\mu_{1k}^2 t^{2h}\right\}$$

and

$$Y_{2k}(t) = Y_k(0)\,(1 + \eta\mu_{2k})^{\pi_k(t)} \cdot \exp\{a\lambda_k t - \nu\eta\mu_{2k} t\},$$

is the solution of (3.1). Since the fBms and the Poisson processes are independent, we get

$$dY_k(t) = Y_{1k}(t)dY_{2k}(t) + Y_{2k}(t)dY_{1k}(t). \tag{3.4}$$

Obviously it holds

$$Y_{2k}(t) = Y_k(0)\exp\{a\lambda_k t - \nu\eta\mu_{2k} t + \pi_k(t)\ln(1 + \eta\mu_{2k})\}. \tag{3.5}$$

It follows by a result from [3] (see formula (15) on page 261) that (3.5) solves

$$dY_{2k}(t) = Y_{2k}(t)[a\lambda_k - \nu\eta\mu_{2k}]dt + \eta\mu_{2k} Y_{2k}(t-)d\pi_k(t)$$

with $Y_{2k}(0) = Y_k(0)$. By the definition of $\tilde{\pi}(t)$ it follows

$$dY_{2k}(t) = a\lambda_k Y_{2k}(t)dt + \eta\mu_{2k} Y_{2k}(t-)d\tilde{\pi}_k(t)$$

with $Y_{2k}(0) = Y_k(0)$.
The fractional Itô formula in [9] (see formula (2.18)) gives that the process $Y_{1k}(t)$ solves

$$dY_{1k}(t) = \mu_{1k}\sigma Y_{1k}(t)dB_k^h(t)$$

with $Y_{1k}(0) = 1$.

If we substitute these results in (3.4), then we get that the process (3.3) solves equation (3.1). □

We will prove the following a priori estimates:

Lemma 3.2. *There are constants* $C_0 > 0,\ C_1 > 0$ *and* $C_2 > 1$ *such that for all* $t \geq 0$ *and* $k \in \{1, 2, ...\}$ *it holds*

$$E|Y_k(t)|^2 \leq F(t)|Y_k(0)|^2 \tag{3.6}$$

with

$$F(t) = C_0 \exp\{2\sigma^2 C_1 t^{2h} + \nu t(C_2 - 1)\}.$$

Proof. Since the fBms and the Poisson processes are independent, we get

$$E|Y_k(t)|^2 = |Y_k(0)|^2 \left(E \exp\left\{2\sigma\mu_{1k} B_k^h(t)\right\}\right) \left(E\left((1+\eta\mu_{2k})^{2\pi_k(t)}\right)\right)$$

$$\times \exp\left\{-\sigma^2\mu_{1k}^2 t^{2h} + 2a\lambda_k t - 2\nu\eta\mu_{2k} t\right\}. \tag{3.7}$$

Since $\lambda_k < 0$ for each $k \in \{1, 2, ..\}$ and $(\mu_{2k})_{k\geq 1}$ is a bounded sequence, we have

$$\exp\left\{-\sigma^2\mu_{1k}^2 t^{2h} + 2a\lambda_k t - 2\nu\eta\mu_{2k} t\right\} \leq C_0 \text{ for each } t \geq 0,$$

where $C_0 > 0$ is a constant.

The random variable $Z_1 := \exp\left\{2\sigma\mu_{1k} B_k^h(t)\right\}$ is log-normally distributed, so we get for its expectation

$$E(Z_1) = \exp\{2\sigma^2\mu_{1k}^2 t^{2h}\}.$$

From the boundedness of $(\mu_{1k})_{k\geq 1}$ it follows the existence of a positive constant C_1 with

$$E(Z_1) \leq \exp\{2\sigma^2 C_1 t^{2h}\}.$$

For $Z_2 := (1+\eta\mu_{2k})^{2\pi_k(t)}$ we compute

$$E(Z_2) = E\left((1+\eta\mu_{2k})^{2\pi_k(t)}\right) = \sum_{j=0}^{\infty} (1+\eta\mu_{2k})^{2j} \exp\{-\nu t\} \frac{(\nu t)^j}{j!}.$$

But the sequence $(\mu_{2k}^2)_{k\geq 1}$ is bounded, hence there is a constant $C_2 > 1$ such that

$$E(Z_2) \leq \exp\{-\nu t\} \sum_{j=0}^{\infty} \frac{(\nu t C_2)^j}{j!} = \exp\{\nu t(C_2 - 1)\}.$$

Then we get with (3.7) the inequality (3.6). □

We now consider a solution definition of mild solution type.

Theorem 3.3. *Let* $k \in \{1, 2, \ldots\}$. *The process* $(Y_k(t))_{t\geq 0}$ *defined by* (3.3) *solves* (3.2) *if and only if the equation*

$$Y_k(t) = Y_k(0)\exp\{\lambda_k a t\} + \int_0^t \exp\{\lambda_k a(t-s)\}\sigma\mu_{1k} Y_k(s) dB_k^h(s)$$

$$+ \int_0^t \exp\{\lambda_k a(t-s)\}\eta\mu_{2k} Y_k(s-) d\tilde{\pi}_k(s), \text{ for all } t \geq 0 \tag{3.8}$$

holds.

Proof. If the process $(Y_k(t))_{t\geq 0}$ is the solution of (3.2) defined by (3.3), then this process solves (3.8) too.

Let $(\tilde{Y}_k(t))_{t\geq 0}$ be a solution process of (3.8). Then we get

$$\tilde{Y}_k(t) = Y_k(0)\exp\{a\lambda_k t\} + \exp\{a\lambda_k t\}\int_0^t \exp\{-a\lambda_k s\}\sigma\mu_{1k}\tilde{Y}_k(s)dB_k^h(s)$$
$$+\exp\{a\lambda_k t\}\int_0^t \exp\{-a\lambda_k s\}\eta\mu_{2k}\tilde{Y}_k(s-)d\tilde{\pi}_k(s). \tag{3.9}$$

Obviously $\exp\{a\lambda_k t\}$ is deterministic and differentiable and the stochastic differentials of the stochastic integrals in formula (3.9) exist. If we use the stochastic product formula to the two last terms of the sum in formula (3.9), then we get

$$d\tilde{Y}_k(t) = a\lambda_k\tilde{Y}_k(t)dt + \sigma\mu_{1k}\tilde{Y}_k(t)dB_k^h(t) + \eta\mu_{2k}\tilde{Y}_k(t-)d\tilde{\pi}_k(t).$$

It follows from formula (3.9) that $\tilde{Y}_k(0) = Y_k(0)$. That is, $\tilde{Y}_k(t) = Y_k(t)$ for all $t \geq 0$ with probability 1. □

We introduce for $n \geq 1$ and $t \geq 0$

$$X_n(t) := \sum_{k=1}^{n} Y_k(t)h_k$$

By (3.6) and $Y_k(0) = (X_0, h_k)$ we get

$$E\|X_n(t)\|_H^2 = \sum_{k=1}^{n} E|Y_k(t)|^2 \leq \|X_0\|_H^2 F(t) \tag{3.10}$$

for every $t > 0$. It follows also from this inequality and the definition of $F(t)$ that there is for all $T > 0$ a positive constant C_T such that

$$E\|X_n(t)\|_H^2 \leq C_T\|X_0\|_H^2 \tag{3.11}$$

and

$$E\int_0^t \|X_n(s)\|_H^2 ds \leq TC_T\|X_0\|_H^2 \tag{3.12}$$

for all $t \in [0,T]$ and all $n \geq 1$.
Consequently for $t > 0$ there exists in $L^2(\Omega;\, H)$ and in $L^2(\Omega\times[0,T]\,;H)$ the process

$$X(t) := \sum_{k=1}^{\infty} Y_k(t)h_k \tag{3.13}$$

and the a priori estimates (3.11) and (3.12) hold also for $X(t)$.

Since $Y_k(0) = (X_0, h_k)$ holds, we obtain $X_0 = \sum_{k=1}^{\infty}(Y_k(0), h_k)h_k$.

It holds for $X_n(t)$

$$X_n(t) = \sum_{k=1}^{n} \exp\{\lambda_k at\}(X_0, h_k)h_k$$
$$+\sigma\int_0^t \sum_{k=1}^{n} \exp\{\lambda_k a(t-s)\}\mu_{1k}(X_n(s), h_k)h_k dB_k^h(s)$$

$$+\eta\int_0^t\sum_{k=1}^n \exp\{\lambda_k a(t-s)\}\mu_{2k}(X_n(s-),h_k)h_k d\tilde{\pi}_k(s).$$

Consequently, we get by the definition of the semigroup $\mathcal{T}_t$ and the operators Φ_1, Φ_2

$$dX_n(t)=\mathcal{T}_tX_n(0)+\sigma\int_0^t\mathcal{T}_{t-s}\Phi_1X_n(s)dB^h(s)+\eta\int_0^t\mathcal{T}_{t-s}\Phi_2X_n(s-)d\tilde{\pi}(s), \quad (3.14)$$

where B^h and $\tilde{\pi}$ are the cylindrical fractional Brownian motion and the cylindrical centered Poisson process defined by the sequences $(B_k^h(t))_{t\geq 0},\ k\in\{1,2,\ldots\}$, and $(\tilde{\pi}_j(t))_{t\geq 0},\ j\in\{1,2,\ldots\}$.

With the definition of $X(t)$ from formula (3.13) and the a priori estimates (3.11), (3.12), with (3.14), by the definition of $Y_k(t)$ and the definitions of the stochastic integrals it is easy to prove, that the following result holds:

Theorem 3.4. *The process $(X(t))_{t\in[0,T]}$ with $X(t)=\sum_{k=1}^{\infty}Y_k(t)h_k$, $t\in[0,T]$, solves*

$$X(t)=\mathcal{T}_tX_0+\sigma\int_0^t\mathcal{T}_{t-s}\Phi_1X(s)dB^h(s)+\eta\int_0^t\mathcal{T}_{t-s}\Phi_2X(s-)d\tilde{\pi}(s) \quad (3.15)$$

for all $t\in[0,T]$.

Remark 3.5. The last theorem shows, that $(X(t))_{t\in[0,T]}$ is the *mild solution* of

$$dX(t)=aAX(t)dt+\sigma\Phi_1X(t)dB^h(t)+\eta\Phi_2X(t-)d\tilde{\pi}(t)\ ,X(0)=X_0.$$

4. Parameter estimation

In what follows we assume $Y_k(0)>0$ and $1+\eta\mu_{2k}>0$ for all $k\in\{1,2,\ldots\}$. Let $k\in\{1,2,\ldots\}$ be arbitrary. We introduce a method to estimate the parameter a in equation (3.2) for the process $(Y_k(t))_{t\in[0,T]}$. By construction we can interpret this process as the process corresponding to the k-th Fourier coefficient of $(X(t))_{t\in[0,T]}$ with respect to h_k. We introduce the process

$$\xi_k(t):=\ln(Y_k(t))=\ln(Y_k(0))+\sigma\mu_{1k}B_k^h(t)-\frac{1}{2}\sigma^2\mu_{1k}^2t^{2h}$$
$$+a\lambda_kt-\nu\eta\mu_{2k}t+\ln(1+\eta\mu_{2k})\pi_k(t). \quad (4.1)$$

Then,

$$E(\xi_k(t))=E(\ln(Y_k(t)))=\ln(Y_k(0))-\frac{1}{2}\sigma^2\mu_{1k}^2t^{2h}$$
$$+a\lambda_kt-\nu\eta\mu_{2k}t+(1+\eta\mu_{2k})\nu t. \quad (4.2)$$

Remark 4.1. Parameter estimation problems involving maximum likelihood methods for equations of type (4.1) without Poisson processes were considered in [1].

We consider for $n \in \{1, 2, \dots\}$ and for $\beta > 1$, such that $n^\beta \in \mathbb{N}$, the partitions

$$t_1 = \frac{1}{n} < t_2 = \frac{2}{n} < \dots < t_{n^\beta} = \frac{n^\beta}{n} =: T(n^\beta)\,.$$

Having statistical observations for $X(\cdot)$ in this time points, we can calculate $\xi_k(t_1), \dots, \xi_k(T(n^\beta))$.
We introduce the following estimation criterion analogous to the one given in [7]:

$$\min \left\{ \sum_{i,j=1}^{n^\beta} \operatorname{cov}\left(\xi_k(t_i), \xi_k(t_j)\right) \ : \ a > 0 \right\}. \tag{4.3}$$

Equation (4.3) is a quadratic function with respect to a. The factor in front of a^2 is given by the positive term $\lambda_k^2 \Big(\sum_{i=1}^{n^\beta} t_i \Big)^2$. Consequently, there is a unique estimator $\hat{a}(n^\beta)$.

Theorem 4.2. *For every $k \in \{1, 2, \dots\}$ the estimator $\hat{a}(n^\beta) =$*

$$-\frac{\sum_{i=1}^{n^\beta} \left(\ln(Y_k(0)) - \eta\mu_{2k}\nu t_i + \ln(1 + \eta\mu_{2k})\nu t_i - \frac{1}{2}\sigma^2\mu_{1k}^2 t_i^{2h} - \ln(Y_k(t_i)) \right)}{\lambda_k \sum_{i=1}^{n^\beta} t_i} \tag{4.4}$$

is unbiased and weakly consistent for the parameter a.

Proof. If we substitute $\ln(Y_k(t_i))$ in the right hand side of $\hat{a}(n^\beta)$, then we get

$$\hat{a}(n^\beta) = a + \frac{\sum_{i=1}^{n^\beta} (\sigma B_k^h(t_i) + \ln(1 + \eta\mu_{2k})\tilde{\pi}_k(t_i))}{\lambda_k \sum_{i=1}^{n^\beta} t_i}. \tag{4.5}$$

Consequently, the estimator is unbiased. We get

$$E[\hat{a}(n^\beta) - a]^2 = E \left[\frac{\sum_{i=1}^{n^\beta} \left(\sigma\mu_{1k} B_k^h(t_i) + \ln(1 + \eta\mu_{2k})\tilde{\pi}_k(t_i)\right)}{\lambda_k \sum_{i=1}^{n^\beta} t_i} \right]^2$$

$$= \frac{\sum_{i,j=1}^{n^\beta} \left(\frac{\sigma^2\mu_{1k}^2}{2} \left(t_i^{2h} + t_j^{2h} - |t_i - t_j|^{2h}\right) + (\ln(1 + \eta\mu_{2k}))^2 \nu \min\{t_i, t_j\} \right)}{\lambda_k^2 \sum_{i,j=1}^{n^\beta} t_i t_j}.$$

Further we have

$$\sum_{i,j=1}^{n^\beta} \frac{\sigma^2\mu_{1k}^2}{2}\left(t_i^{2h}+t_j^{2h}-|t_i-t_j|^{2h}\right) \le \sum_{i,j=1}^{n^\beta} \frac{\sigma^2\mu_{1k}^2}{2}\left(t_i^{2h}+t_j^{2H}\right)$$

$$\le \sigma^2\mu_{1k}^2 \sum_{i,j=1}^{n^\beta}\left(\frac{n^\beta}{n}\right)^{2h} = \sigma^2\mu_{1k}^2\frac{n^{\beta(2+2h)}}{n^{2h}},$$

$$\sum_{i,j=1}^{n^\beta}(\ln(1+\eta\mu_{2k}))^2\nu\min\{t_i,t_j\} \le (\ln(1+\eta\mu 2k))^2\nu\sum_{i,j=1}^{n^\beta}\left(\frac{n^\beta}{n}\right)$$

$$= (\ln(1+\eta\mu_{2k}))^2\nu\frac{n^{3\beta}}{n}$$

and

$$\sum_{i,j=1}^{n^\beta} t_it_j = \frac{1}{n^2}\left(\sum_{i=1}^{n^\beta} i\right)^2 = \frac{n^{2\beta}(n^\beta+1)^2}{4n^2} \ge \frac{n^{4\beta}}{4n^2}.$$

Consequently,

$$\lim_{n\to\infty} E[\hat{a}(n^\beta)-a]^2$$

$$\le \lim_{n\to\infty}\frac{4}{\lambda_k^2}\left(\sigma^2\mu_{1k}^2\cdot n^{(2-2h)+\beta(2+2h-4)} + (\ln(1+\eta\mu_{2k}))^2\nu\cdot n^{(2-1)+\beta(3-4)}\right)$$

$$\le \lim_{n\to\infty}\frac{4}{\lambda_k^2}\left(\sigma^2\mu_{1k}^2\cdot n^{(2-2h)-\beta(2-2h)} + (\ln(1+\eta\mu_{2k}))^2\nu\cdot n^{1-\beta}\right) = 0$$

and the weak consistency follows for $\beta > 1$. □

Remark 4.3. 1. In a similar manner we can calculate estimates for σ^2 and ν. Then, we need the condition $\beta > 4h-1$ to prove the weak consistency.

2. The estimation of η is difficult. A possibility consists in the application of an approximation of $\tilde{\pi}_k$ by Brownian motions $(B_k)_{k\ge 1}$ with $E(B_k^2(t)) = \nu t$ by using the Central Limit Theorem.
3. If $\eta = 0$, then the random variable

$$\frac{\lambda_k^2}{4}\cdot\frac{\hat{a}(n^\beta)-a}{\sigma^2\mu_{1k}^2\cdot n^{(2-2h)-\beta(2-2h)}}$$

is asymptotically $N(0,1)$ distributed.

Moreover, we prove the following result:

Theorem 4.4. *Consider $\beta > \frac{3-2h}{2-2h}$ such that $n^\beta \in \mathbb{N}$. Then the estimate (4.4) for the parameter a is strongly consistent for all $k \in \{1,2,\ldots\}$.*

Proof. It is well known that, if for all $\varepsilon > 0$ the relation

$$\sum_{n=1}^{\infty} P(|\hat{a}(n^\beta)-a| > \varepsilon) < \infty$$

holds, then $\lim_{n\to\infty} \hat{a}(n^\beta) = a$ with probability 1.

Let $k \in \{1, 2, \ldots\}$ be arbitrary. We know from the end of the proof of the last theorem that for the variance of $\hat{a}(n^\beta)$ we can write

$$V(\hat{a}(n^\beta)) \leq \frac{4}{\lambda_k^2}\left(\sigma^2\mu_{1k}^2 \cdot n^{(2-2h)-\beta(2-2h)} + (\ln(1+\eta\mu_{2k}))^2\nu \cdot n^{1-\beta}\right).$$

Then, by using Chebyshev's inequality we obtain

$$\sum_{n=1}^{\infty} P(|\hat{a}(n^\beta) - a| > \varepsilon) \leq \sum_{n=1}^{\infty} \frac{1}{\varepsilon^2} V(\hat{a}(n^\beta))$$
$$\leq \frac{4}{\lambda_k^2}\sum_{n=1}^{\infty}\sigma^2\mu_{1k}^2 \cdot n^{(2-2h)-\beta(2-2h)} + \frac{4}{\lambda_k^2}\sum_{n=1}^{\infty}(\ln(1+\eta\mu_{2k}))^2\nu \cdot n^{1-\beta}.$$

Obviously, the first, respectively, the second sum on the right hand side of the last inequality are convergent, if

$$\beta > \frac{3-2h}{2-2h}, \text{ respectively, } \beta > 2.$$

Hence, we get the statement for $\beta > \frac{3-2h}{2-2h} > 2$ (since $1/2 < h < 1$). □

References

[1] Bertin, K., Torres, S., Tudor, C. A., *Maximum-likelihood estimators and random walks in long memory models*, Statistics, **45**(2011), no. 4, 361-374.

[2] Cialenco, I., *Parameter estimation for SPDEs with multiplicative fractional noise*, Stoch. Dyn., **10**(2010), no. 4, 561-576.

[3] Gichman, I.I., Skorochod, A.W., *Stochastische Differentialgleichungen*, Akademie-Verlag, 1971.

[4] Kozachenko, Y., Melnikov, A., Mishura, Y., *On drift parameter estimation in models with fractional Brownian motion*, Statistics, **49**(2015), no. 1, 35-62.

[5] Long, H., *Parameter estimation for a class of stochastic differential equations driven by small stable noises from discrete observations*, Acta Math. Sci. Ser. B (Engl. Ed.), **30**(2010), no. 3, 645-663.

[6] Lototsky, S.V., Rozovsky, B.L., *Stochastic partial differential equations*, Springer, 2017.

[7] Lueddeckens, J., *Fraktale stochastische Integralgleichungen im White-Noise-Kalkül*, Dissertation Martin-Luther-Universität Halle-Wittenberg, April 2017.

[8] Maslowski, B., Tudor, C.A., *Drift parameter estimation for infinite-dimensional fractional Ornstein–Uhlenbeck process*, Bull. Sci. Math., **137**(2013), no. 7, 880-901.

[9] Nualart, D., *Stochastic calculus with respect to fractional Brownian motion*, Ann. Fac. Sci. Toulouse Math. (6), **15**(2006), no. 1, 63-78.

[10] Protter, Ph., *Stochastic Integration and Differential Equations*, Springer, 1990.

[11] Xiao, W., Zhang, W., Zhang, X., *Parameter identification for discretely observed geometric fractional Brownian motion*, J. Stat. Comput. Simul., **85**(2015), no. 2, 269-283.

Wilfried Grecksch
Martin-Luther-University Halle-Wittenberg
Faculty of Natural Sciences II, Institute of Mathematics
06099 Halle (Saale), Germany
e-mail: `wilfried.grecksch@mathematik.uni-halle.de`

Hannelore Lisei
Babeş-Bolyai University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street
400084 Cluj-Napoca, Romania
e-mail: `hanne@math.ubbcluj.ro`

Jens Lueddeckens
Martin-Luther-University Halle-Wittenberg
Faculty of Natural Sciences II, Institute of Mathematics
06099 Halle (Saale), Germany
e-mail: `jlueddeckens@web.de`

Stud. Univ. Babeş-Bolyai Math. 64(2019), No. 2, 291–296
DOI: 10.24193/subbmath.2019.2.13

An application of inverse Padé interpolation

Radu T. Trîmbiţaş

Abstract. We use inverse Padé interpolation to find a fourth order method for the solution of scalar nonlinear equations. Our approach is based on Computer Algebra. Maple Computer Algebra system assisted us to find the method and to establish its order.

Mathematics Subject Classification (2010): 65H05, 65Y99.

Keywords: Nonlinear equations, inverse interpolation, Computer Algebra.

1. Introduction

Suppose we wish to approximate the solution of the nonlinear equation

$$f(x) = 0, \tag{1.1}$$

where $f : \Omega \subseteq \mathbb{R} \to \mathbb{R}$. Let α be a solution of (1.1).

Suppose there exists $g = f^{-1}$ on a neighbourhood V of α. The inverse interpolation consists of approximating

$$\alpha = g(0),$$

by the value of an interpolant $\widehat{g}$ for g at 0

$$\alpha \approx \widehat{g}(0).$$

In this paper we will use inverse Padé interpolation. Let $\mathcal{R}_{m,n}$ be the set of rational functions with numerator degree m and denominator degree n. Suppose f has a formal Taylor series

$$f(z) = c_0 + c_1 z + c_2 z^2 + \cdots$$

For any pair of natural numbers (m, n), $r_{mn} \in \mathcal{R}_{m,n}$ is the type (m, n) Padé approximant to f if their Taylor series at $z = 0$ agree as far as possible:

$$\left(f - r_{mn}\right)(z) = O\left(z^{\max}\right). \tag{1.2}$$

This paper has been presented at the fourth edition of the International Conference on Numerical Analysis and Approximation Theory (NAAT 2018), Cluj-Napoca, Romania, September 6-9, 2018.

The formula we look for will have the form

$$x_{k+1} = r_{mn}(x_k), \ k = 0, 1, \ldots \tag{1.3}$$

For details on inverse interpolation see [1, 5, 7]. The paper [7] uses rational interpolation to derive methods for the solution of scalar nonlinear equations.

The paper is structured as follows. The second section establishes the formula, its order and the efficiency index. The third section studies the convergence by using a fixed point approach. The next section gives a MATLAB implementation. Finally, the last section gives two numerical examples and compares the new method to Newton and Halley methods respectively.

2. The formula and its order

We use the Maple package `numapprox` (see [4, 6]). Let us start with a $(1,1)$-degree inverse Padé interpolation.

```
> restart;
> with(numapprox):
> eval(pade((f@@(-1))(y),y=f(x),[1,1]),y=0);
```

$$1/2\,\frac{4\,x\,(\mathrm{D}(f)(x))^2 - \left(2\,\left(D^{(2)}\right)(f)(x)\,x + 4\,\mathrm{D}(f)(x)\right)f(x)}{2\,(\mathrm{D}(f)(x))^2 - \left(D^{(2)}\right)(f)(x)\,f(x)}$$

```
> collect(%,x,simplify);
```

$$x - 2\,\frac{\mathrm{D}(f)(x)\,f(x)}{2\,(\mathrm{D}(f)(x))^2 - \left(D^{(2)}\right)(f)(x)\,f(x)}$$

We rewrite the formula as

$$\Phi(x) = x - \frac{f(x)}{f'(x) - \frac{f''(x)f(x)}{2f'(x)}}.$$

This is the well-known Halley's formula. This formula was obtained using direct Padé approximation in [2, 3].

The next step is to try a $(2,1)$-degree Padé formula.

```
> eval(pade((f@@(-1))(y),y=f(x),[2,1]),y=0):
> F1:=collect(%,x,simplify);
```

$$F1 := x - 1/2\,\frac{f(x)\left(6\,(\mathrm{D}(f)(x))^2\left(D^{(2)}\right)(f)(x) + 2\,\mathrm{D}(f)(x)f(x)\left(D^{(3)}\right)(f)(x) - 3\left(\left(D^{(2)}\right)(f)(x)\right)^2 f(x)\right)}{\mathrm{D}(f)(x)\left(3\,(\mathrm{D}(f)(x))^2\left(D^{(2)}\right)(f)(x) + \mathrm{D}(f)(x)f(x)\left(D^{(3)}\right)(f)(x) - 3\left(\left(D^{(2)}\right)(f)(x)\right)^2 f(x)\right)}$$

Rewrite the previous formula as

$$\Phi(x) = x - \frac{f(x)}{f'(x)}\left\{1 + \frac{\frac{1}{2}}{\dfrac{f'(x)}{f''(x)}\left[\dfrac{f'(x)}{f(x)} + \dfrac{f'''(x)}{3f''(x)}\right] - 1}\right\}. \tag{2.1}$$

The formula (2.1) is equivalent to `F1` from previous Maple code.

```
> FF:=x-f(x)/D(f)(x)*(1+(1/2)/((D(f)(x)/(D@@2)(f)(x))*(D(f)(x)/f(x)+
> (D@@3)(f)(x)/(3*(D@@2)(f)(x)))-1)):
> simplify(F1-FF);
```

$$0$$

We compute the order of (2.1) as follows:

```
> Phi:=unapply(F1,x);
```

$$\Phi := x \mapsto x - 1/6\,\frac{\left(x^3-a\right)\left(324\,x^5-72\,x^2\left(x^3-a\right)\right)}{x^2\left(162\,x^5-90\,x^2\left(x^3-a\right)\right)}$$

```
> simplify(Phi(alpha),[f(alpha)=0]);
```

$$\alpha$$

```
> simplify(D(Phi)(alpha),[f(alpha)=0]);
```

$$0$$

```
> simplify((D@@2)(Phi)(alpha),[f(alpha)=0]);
```

$$0$$

```
> simplify((D@@3)(Phi)(alpha),[f(alpha)=0]);
```

$$0$$

```
> simplify((D@@4)(Phi)(alpha),[f(alpha)=0]);
```

$$1/3\,\frac{\left(3\left(D^{(2)}\right)(f)(\alpha)\left(D^{(4)}\right)(f)(\alpha)-4\left(\left(D^{(3)}\right)(f)(\alpha)\right)^2\right)(\mathrm{D}(f)(\alpha))^2-6\mathrm{D}(f)(\alpha)\left(\left(D^{(2)}\right)(f)(\alpha)\right)^2\left(D^{(3)}\right)(f)(\alpha)+9\left(\left(D^{(2)}\right)(f)(\alpha)\right)^4}{(\mathrm{D}(f)(\alpha))^3\left(D^{(2)}\right)(f)(\alpha)}$$

The last expression is the asymptotic error constant. We write it as

$$C_\alpha = \frac{\left(3f''(\alpha)f^{(4)}(\alpha)-4\left[f'''(\alpha)\right]^2\right)f'^2(\alpha)-6f'(\alpha)\left[f''(\alpha)\right]^2f'''(\alpha)+9\left[f''(\alpha)\right]^4}{3\left[f'(\alpha)\right]^3 f''(\alpha)}. \tag{2.2}$$

The order of (2.1) is $d = 4$, and the efficiency index is $4^{\frac{1}{4}} = \sqrt{2}$. See [5, Section 3.2] for a definition and properties of the efficiency index.

3. The convergence

Let $I_\varepsilon = \{x \in \Omega : |x-\alpha| < \varepsilon\}$ and

$$M(\varepsilon) = \max_{x\in I_\varepsilon}\left|\frac{\Phi^{(4)}(x)}{4!}\right|. \tag{3.1}$$

Theorem 3.1. *If $\Phi \in C^4\left(I_\varepsilon\right)$ and*

$$\varepsilon^3 M(\varepsilon) < 1, \tag{3.2}$$

then

(a) $x_n \in I_\varepsilon, n = 1, 2, 3\ldots, \forall x_0 \in I_\varepsilon$;
(b) $\lim\limits_{n\to\infty} x_n = \alpha$.

Proof. (a) Since the method has order $d = 4$ we have

$$e_{n+1} \leq Ce_n^4 \tag{3.3}$$

where $e_n = |x_n - \alpha|$ and M is given by (3.1). If $x_0 \in I_\varepsilon$, the conclusion follows by using complete induction.

(b) From (3.3) using the Maple code

```
> rsolve({e(n+1)=C*e(n)^4,e(0)=e0},e(n));
```

$$\frac{e0^{4^n} C^{1/3 \cdot 4^n}}{\sqrt[3]{C}}$$

it follows

$$e_{n+1} \leq \frac{\left[\left(Ce_0^3\right)^{4^n}\right]^{1/3}}{C^{1/3}}.$$

The right hand side tends to 0 if $Ce_0^3 < 1$ which is equivalent to (3.2). □

[1, Theorem 26.1.4, pag. 317] leads us to the same conclusion.

4. Implementation

The function invPade gives a MATLAB implementation for the formula (2.1). The meaning of the input and output parameters are explained in function header.

```
function [y,ni]=invPade(f, fp1, fp2, fp3, x0, ea, er, nmax)
%INVPADE - solution of f(x) = 0 by inverse Pade interpolation
%f, fp1, fp2, fp3 - f and its derivatives
%x0 - starting value
%ea, er - absolute and relative error
%nmax - maximum number of iterations
%y - result
%ni - #iterations

if nargin < 8, nmax=50; end;
if nargin < 7, er=0; end
if nargin < 6, ea=1e-4; end
for k=1:nmax
  f0=f(x0); f1=fp1(x0); f2=fp2(x0); f3=fp3(x0);
  ffp=f0/f1; ifp=f1/f0;
  x1=x0-ffp*(1+0.5/(f1/f2*(ifp+f3/3/f2)-1));
  if abs(x1-x0)<ea+er*abs(x1) %success
    y=x1; ni=k;
    return
  end
  x0=x1;
end
error('max #iterations exceeded')
```

5. Numerical examples

We tested our implementation at the computation of $\sqrt[3]{a}$. We compared our method to Newton and Halley method, respectively. See the source below. We took $a = 201$.

```
a = input('a=');
f = @(x) x^3-a;
fd1 = @(x) 3*x^2;
fd2 = @(x) 6*x;
fd3 = @(x) 6;

[z0,ni0]=Newton(f,fd1,(a+2)/3, 0, eps,100)
[z1,ni1]=Halley(f,fd1,fd2,(a+2)/3, 0, eps, 100)
[z2,ni2]=invPade(f,fd1,fd2,fd3,(a+2)/3, 0, eps, 100)
```

```
z0 =
5.857766002650652
ni0 =
12
z1 =
5.857766002650653
ni1 =
8
z2 =
5.857766002650652
ni2 =
6
```

In order to compute the result with a relative error of machine epsilon and the starting value $x_0 = (a+2)/3$, inverse Padé method requires 6 iteration, while Newton and Halley method require 12 and 8 iterations respectively.

The second example solves numerically the equation

$$xe^x + x^2 - 6 = 0.$$

```
g = @(x) x*exp(x)+x^2-6;
gd1 = @(x) (x+1)*exp(x)+2*x;
gd2 = @(x) 2+(x+2)*exp(x);
gd3 = @(x) (x+3)*exp(x);
tic
[z0,ni0]=Newton(g,gd1,5, 0, eps,100);
t0=toc;
tic
[z1,ni1]=Halley(g,gd1,gd2, 5, 0, eps,100);
t1=toc;
tic
```

```
[z2,ni2]=invPade(g,gd1,gd2,gd3, 5, 0, eps,100);
t2=toc;
fprintf('Newton,    z=%17.15f, ni=%2d, elapsed time=%f\n',z0,ni0,t0)
fprintf('Halley,    z=%17.15f, ni=%2d, elapsed time=%f\n',z1,ni1,t1)
fprintf('Inv. Pade, z=%17.15f, ni=%2d, elapsed time=%f\n',z2,ni2,t2)
```

```
Newton,    z=1.257169468081542, ni=11, elapsed time=0.000570
Halley,    z=1.257169468081542, ni= 6, elapsed time=0.000560
Inv. Pade, z=1.257169468081542, ni= 5, elapsed time=0.000549
```

For a relative tolerance equal to machine epsilon and a starting value $x_0 = 5$, the inverse Padé method requires 5 iterations, while Newton and Halley method require 11 and 6 iterations, respectively.

References

[1] Agratini, O., Blaga, P., Chiorean, I., Coman, Gh., Stancu, D.D., Trîmbiţaş, R.T., *Numerical Analysis and Approximation Theory* (vol. III), Cluj University Press, Cluj-Napoca, 2002 (in Romanian).

[2] Gander, W., Gander, M.J., Kwok, F., *Scientific Computing. An Introduction Using Maple and MATLAB*, Springer, 2014.

[3] Gander, W., Gruntz, D., *Derivation of numerical methods using Computer Algebra*, SIAM Rev., **41**(1999), no. 3, 577-593.

[4] Garvan, F., *The Maple Book*, 1st Edition, Chapman & Hall/CRC, 2001.

[5] Gautschi, W., *Numerical Analysis*, Second Edition, Springer Science+Business Media, 2012.

[6] Heck, A., *Introduction to Maple*, Third Edition, Springer-Verlag, New York, 2003.

[7] Păvăloiu, I., *Equations Solution through Interpolation*, Dacia Publishers, 1981 (in Romanian).

Radu T. Trîmbiţaş
Babeş-Bolyai University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street,
400084 Cluj-Napoca, Romania
e-mail: tradu@math.ubbcluj.ro