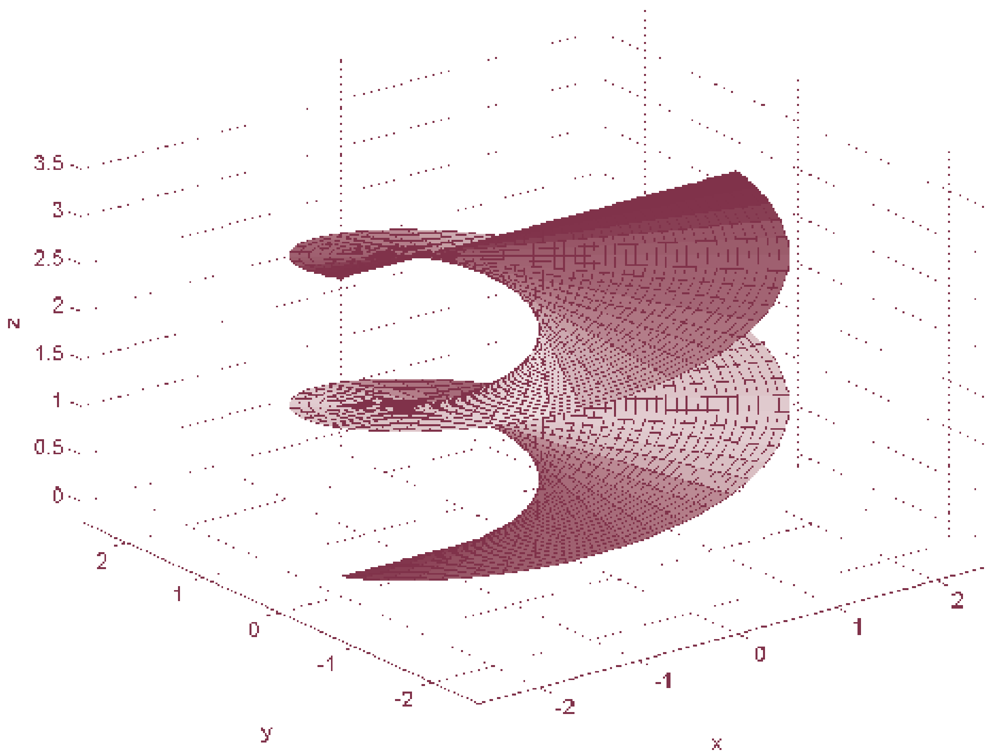
STUDIA UNIVERSITATIS
BABEŞ-BOLYAI



MATHEMATICA

2/2011

Volume 56, Number 2, June, 2011

STUDIA
UNIVERSITATIS BABEŞ-BOLYAI
MATHEMATICA

CONTENTS

NUMBER 2A

NUMBER 2B

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 197–197

Preface

Between September 23-26, 2010, The Second International Conference on Numerical Analysis and Approximation Theory (NAAT 2010) was held in Cluj-Napoca, Romania. It was organized by the Applied Mathematics Department of Faculty of Mathematics and Computer Science, Babeş-Bolyai University. The first edition took place in 2006.

Promoting interactions between specialists, young researchers and PhD students, the meeting was devoted to some significant aspects of mathematical areas on functions approximation, integral and differential operators, numerical analysis and stability methods, positive operators, splines, wavelets, stochastic processes, approximation of linear functionals.

A special word of thanks goes to the invited speakers: Francesco Altomare (Bari, Italy), Carsten Carstensen (Berlin, Germany), Sorin Gal (Oradea, Romania), Heiner Gonska (Duisburg, Germany), Kurt Helmes (Berlin, Germany), Willi Jäger (Heidelberg, Germany), Gradimir Milovanović (Beograd, Serbia), Paul Sablonniere (INSA Rennes, France).

In the frame of the Conference, a symposium was dedicated to Professor Dr. h.c. Willi Jäger on the occasion of his 70^{th} birthday.

The Conference was attended by over 80 mathematicians coming from 13 countries: Czech Republic, France, Georgia, Germany, Hungary, Israel, Italy, Norway, Romania, Serbia, Spain, Syria and Turkey. Besides the plenary lecturers the programme included 71 research talks. The participants appreciated the diversity of topics covered and the quality level of the talks. They showed enjoyment over the opportunity of sharing their ideas with each other.

In this volume we collect a selection of 37 refereed papers corresponding to some presented talks.

Finally, we would like to express our gratitude to all participants who transformed NAAT 2010 into a successful event by creating a warm and cordial atmosphere impregnated with a high degree of professionalism.

The courtain fell on this meeting. Since every end is a new beginning, see you in 2014.

Octavian Agratini,
On behalf of the Organizing Committee

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 199–206

Asymptotic expansions for Favard operators and their left quasi-interpolants

Ulrich Abel

Abstract. In 1944 Favard [5, pp. 229, 239] introduced a discretely defined operator which is a discrete analogue of the familiar Gauss-Weierstrass singular convolution integral. In the present paper we consider a slight generalization F_{n,σ_n} of the Favard operator and its Durrmeyer variant $\tilde{F}_{n,\sigma_n}$ and study the local rate of convergence when applied to locally smooth functions. The main result consists of the complete asymptotic expansions for the sequences $(F_{n,\sigma_n} f)(x)$ and $\left(\tilde{F}_{n,\sigma_n} f\right)(x)$ as n tends to infinity. Furthermore, these asymptotic expansions are valid also with respect to simultaneous approximation. Finally, we define left quasi-interpolants for the Favard operator and its Durrmeyer variant in the sense of Sablonniere.

Mathematics Subject Classification (2010): 41A36, 41A60, 41A28.

Keywords: Approximation by positive operators, asymptotic expansions, simultaneous approximation.

1. Introduction

In 1944 J. Favard [5, pp. 229, 239] introduced the operator

$$(F_n f)(x) = \frac{1}{\sqrt{\pi n}} \sum_{\nu=-\infty}^{\infty} f\left(\frac{\nu}{n}\right) \exp\left(-n\left(\frac{\nu}{n} - x\right)^2\right) \tag{1.1}$$

which is a discrete analogue of the familiar Gauss-Weierstrass singular convolution integral

$$(W_n f)(x) = \sqrt{\frac{n}{\pi}} \int_{-\infty}^{\infty} f(t) \exp\left(-n(t-x)^2\right) dt.$$

Basic properties such as saturation in weighted spaces can be found in [3] and [2]. For a sequence of positive reals σ_n, the generalization

$$\left(F_{n,\sigma_n} f\right)(x) = \sum_{\nu=-\infty}^{\infty} p_{n,\nu,\sigma_n}(x)\, f\left(\frac{\nu}{n}\right), \tag{1.2}$$

where

$$p_{n,\nu,\sigma_n}(x) = \frac{1}{\sqrt{2\pi} n \sigma_n} \exp\left(-\frac{1}{2\sigma_n^2}\left(\frac{\nu}{n} - x\right)^2\right),$$

was introduced and studied by Gawronski and Stadtmüller [7]. The particular case $\sigma_n^2 = \gamma/(2n)$ with a constant $\gamma > 0$ reduces to Favard's classical operators (1.1). The operators can be applied to functions f defined on $\mathbb{R}$ satisfying the growth condition

$$f(t) = O\left(e^{Kt^2}\right) \qquad \text{as } |t| \to \infty, \tag{1.3}$$

for a constant $K > 0$.

In 2007 Nowak and Sikorska-Nowak [11] considered a Kantorovich variant [11, Eq. (1.5)]

$$\left(\hat{F}_{n,\sigma_n} f\right)(x) = n \sum_{\nu=-\infty}^{\infty} p_{n,\nu,\sigma_n}(x) \int_{\nu/n}^{(\nu+1)/n} p_{n,\nu,\sigma_n}(t)\, f(t)\, dt$$

and a Durrmeyer variant [11, Eq. (1.6)]

$$\left(\tilde{F}_{n,\sigma_n} f\right)(x) = n \sum_{\nu=-\infty}^{\infty} p_{n,\nu,\sigma_n}(x) \int_{-\infty}^{\infty} p_{n,\nu,\sigma_n}(t)\, f(t)\, dt \tag{1.4}$$

of Favard operators. Further related papers are [12] and [13].

The main result of this paper consists of the complete asymptotic expansions

$$F_{n,\sigma_n} f \sim f + \sum_{k=0}^{\infty} c_k(f)\, \sigma_n^k \quad \text{and} \quad \tilde{F}_{n,\sigma_n} f \sim f + \sum_{k=0}^{\infty} \tilde{c}_k(f)\, \sigma_n^k \qquad (n \to \infty),$$

for f sufficiently smooth. The coefficients c_k and $\tilde{c}_k$, which depend on f but are independent of n, are explicitly determined. It turns out that $c_k(f) = 0$, for all odd integers $k > 0$. Moreover, we deal with simultaneous approximation by the operators (1.2).

Finally, we define left quasi-interpolants for the Favard operator and its Durrmeyer variant in the sense of Sablonniere.

2. Complete asymptotic expansions

Throughout the paper, we assume that

$$\sigma_n > 0, \qquad \sigma_n \to 0, \qquad \sigma_n^{-1} = O\left(n^{1-\eta}\right) \qquad (n \to \infty) \tag{2.1}$$

with (an arbitrarily small) constant $\eta > 0$. Note that the latter condition implies that $n\sigma_n \to \infty$ as $n \to \infty$.

Under these conditions, the operators possess the basic property that $(F_n f)(x)$ converges to $f(x)$ in each continuity point x of f. Among other results, Gawronski and Stadtmüller [7, Eq. (0.6)] established the Voronovskaja-type theorem

$$\lim_{n\to\infty} \frac{1}{\sigma_n^2} \left[(F_{n,\sigma_n} f)(x) - f(x) \right] = \frac{1}{2} f''(x) \tag{2.2}$$

uniformly on proper compact subsets of $[a, b]$, for $f \in C^2[a, b]$ $(a, b \in \mathbb{R})$ and $\sigma_n \to 0$ as $n \to \infty$, provided that certain conditions on the first three moments of F_{n,σ_n} are satisfied. Actually, Eq. (2.2) was proved for a truncated variant of (1.2) which possesses the same asymptotic properties as (1.2) [7, cf. Theorem 1 (iii) and Remark (i), p. 393]. For a Voronovskaja-type theorem in the particular case $\sigma_n^2 = \gamma/(2n)$ cf. [3, Theorem 4.3]. Abel and Butzer extended Formula (2.2) by deriving a complete asymptotic expansion of the form

$$F_{n,\sigma_n} f \sim f + \sum_{k=0}^{\infty} c_k(f)\, \sigma_n^k \qquad (n \to \infty),$$

for f sufficiently smooth. The latter formula means that, for all positive integers q, there holds pointwise on $\mathbb{R}$

$$F_{n,\sigma_n} f = f + \sum_{k=1}^{q} c_k(f)\, \sigma_n^k + o(\sigma_n^q) \qquad (n \to \infty).$$

The following theorem presents the main result of this paper, the complete asymptotic expansion for the sequence $\left(\tilde{F}_{n,\sigma_n}\right)(x)$ as $n \to \infty$. For $r \in \mathbb{N}$ and $x \in \mathbb{R}$ let $W[r; x]$ be the class of functions on $\mathbb{R}$ satisfying growth condition (1.3), which admit a derivative of order r at the point x.

Theorem 2.1. *Let $q \in \mathbb{N}$ and $x \in \mathbb{R}$. Suppose that the real sequence (σ_n) satisfies the conditions* (2.1)*. For each function $f \in W[2q; x]$, the Favard-Durrmeyer operators* (1.4) *possess the complete asymptotic expansions*

$$(F_{n,\sigma_n} f)(x) = f(x) + \sum_{k=1}^{q} \frac{f^{(2k)}(x)}{(2k)!!} \sigma_n^{2k} + o\left(\sigma_n^{2q}\right) \tag{2.3}$$

and

$$\left(\tilde{F}_{n,\sigma_n} f\right)(x) = f(x) + \sum_{k=1}^{q} \frac{f^{(2k)}(x)}{k!} \sigma_n^{2k} + o\left(\sigma_n^{2q}\right) \tag{2.4}$$

as $n \to \infty$.

Here $m!!$ denote the double factorial numbers defined by $0!! = 1!! = 1$ and $m!! = m \times (m-2)!!$ for integers $m \geq 2$. It turns out that the asymptotic expansions contain only terms with even order derivatives of the function f.

As an immediate consequence we obtain the following Voronosvkaja-type theorems.

Corollary 2.2. *Let $x \in \mathbb{R}$. Suppose that the real sequence (σ_n) satisfies the conditions* (2.1). *For each function $f \in W[2;x]$, there hold the asymptotic relations*

$$\lim_{n\to\infty} \sigma_n^{-2} \left(\left(F_{n,\sigma_n} f\right)(x) - f(x)\right) = \frac{1}{2} f''(x)$$

and

$$\lim_{n\to\infty} \sigma_n^{-2} \left(\left(\tilde{F}_{n,\sigma_n} f\right)(x) - f(x)\right) = f''(x).$$

Concerning simultaneous approximation, it turns out that the complete asymptotic expansion (2.3) can be differentiated term-by-term. Indeed, there holds

Theorem 2.3. *Let $\ell \in \mathbb{N}_0$, $q \in \mathbb{N}$ and $x \in \mathbb{R}$. Suppose that the real sequence (σ_n) satisfies condition* (2.1). *For each function $f \in W[2(\ell+q);x]$, the following complete asymptotic expansions are valid as $n \to \infty$:*

$$\left(F_{n,\sigma_n} f\right)^{(\ell)}(x) = f(x) + \sum_{k=1}^{q} \frac{f^{(2k+\ell)}(x)}{(2k)!!} \sigma_n^{2k} + o\left(\sigma_n^{2q}\right) \tag{2.5}$$

and

$$\left(\tilde{F}_{n,\sigma_n} f\right)^{(\ell)}(x) = f^{(\ell)}(x) + \sum_{k=1}^{q} \frac{f^{(2k+\ell)}(x)}{k!} \sigma_n^{2k} + o\left(\sigma_n^{2q}\right). \tag{2.6}$$

Remark 2.4. The latter formulas can be written in the equivalent form

$$\lim_{n\to\infty} \sigma_n^{-2q} \left(\left(F_{n,\sigma_n} f\right)^{(\ell)}(x) - f^{(\ell)}(x) - \sum_{k=1}^{q} \frac{f^{(2k+\ell)}(x)}{(2k)!!} \sigma_n^{2k} \right) = 0,$$

$$\lim_{n\to\infty} \sigma_n^{-2q} \left(\left(\tilde{F}_{n,\sigma_n} f\right)^{(\ell)}(x) - f^{(\ell)}(x) - \sum_{k=1}^{q} \frac{f^{(2k+\ell)}(x)}{k!} \sigma_n^{2k} \right) = 0.$$

Assuming smoothness of f on intervals $I = (a,b)$, $a,b \in \mathbb{R}$, it can be shown that the above expansions hold uniformly on compact subsets of I.

The proofs are based on localization theorems which are interesting in themselves. We quote only the result for the ordinary Favard operator (1.2).

Proposition 2.5. *Fix $x \in \mathbb{R}$ and let $\delta > 0$. Assume that the function $f : \mathbb{R} \to \mathbb{R}$ vanishes in $(x-\delta, x+\delta)$ and satisfies, for positive constants M_x, K_x, the growth condition*

$$|f(t)| \le M_x e^{K_x (t-x)^2} \qquad (t \in \mathbb{R}). \tag{2.7}$$

Then, for positive $\sigma < 1/\sqrt{2K_x}$, there holds the estimate

$$\left|\left(F_{n,\sigma} f\right)(x)\right| \le \sqrt{\frac{2}{\pi}} \frac{M_x \sigma/\delta}{1 - 2K_x \sigma^2} \exp\left(-\frac{1 - 2K_x \sigma^2}{2} \left(\frac{\delta}{\sigma}\right)^2 \right).$$

Consequently, under the general assumption (2.1) a positive constant A (independent of δ) exists such that the sequence $((F_{n,\sigma_n} f)(x))$ can be estimated by

$$(F_{n,\sigma_n} f)(x) = o\left(\exp\left(-A\frac{\delta^2}{\sigma_n^2}\right)\right) \qquad (n \to \infty).$$

Remark 2.6. Note that a function $f : \mathbb{R} \to \mathbb{R}$ satisfies condition (2.1) if and only if condition (2.7) is valid. The elementary inequality $(t-x)^2 \leq 2(t^2 + x^2)$ implies that

$$M_x e^{K_x (t-x)^2} \leq M e^{Kt^2} \qquad (t, x \in \mathbb{R})$$

with constants $M = M_x e^{2K x^2}$ and $K = 2K_x$.

3. Quasi-interpolants

The results of the preceding section show that the optimal degree of approximation cannot be improved in general by higher smoothness properties of the function f. In order to obtain much faster convergence quasi-interpolants were considered. Let us shortly recall the definition of the quasi-interpolants in the sense of Sablonniere [14]. For another method to construct quasi-interpolants see [8] and [9].

If the operators $\mathcal{B}_n$ let invariant the space of algebraic polynomials Π_j of each order $j = 0, 1, 2, \ldots$ (the most approximation operators possess this property), i.e.,

$$\mathcal{B}_n(\Pi_j) \subseteq \Pi_j \qquad (0 \leq j \leq n),$$

$B_n : \Pi_n \to \Pi_n$ is an isomorphism which can be represented by linear differential operators

$$\mathcal{B}_n = \sum_{k=0}^{n} \beta_{n,k} D^k$$

with polynomial coefficients $\beta_{n,k}$ and $Df = f'$, $D^0 =$id. The inverse operator $\mathcal{B}_n^{-1} \equiv \mathcal{A} : \Pi_n \to \Pi_n$ satisfies

$$\mathcal{A} = \sum_{k=0}^{n} \alpha_{n,k} D^k$$

with polynomial coefficients $\alpha_{n,k}$. Sablonniere defined new families of intermediate operators obtained by composition of B_n and its truncated inverses

$$\mathcal{A}_n^{(r)} = \sum_{k=0}^{r} \alpha_{n,k} D^k.$$

In this way he obtained a family of left quasi-interpolants (LQI) defined by

$$\mathcal{B}_n^{(r)} = \mathcal{A}_n^{(r)} \circ \mathcal{B}_n, \quad 0 \leq r \leq n,$$

and a family of right quasi-interpolants (RQI) defined by

$$\mathcal{B}_n^{[r]} = \mathcal{B}_n \circ \mathcal{A}_n^{(r)}, \quad 0 \leq r \leq n.$$

Obviously, there holds $\mathcal{B}_n^{(0)} = \mathcal{B}_n^{[0]} = \mathcal{B}_n$, and $\mathcal{B}_n^{(n)} = \mathcal{B}_n^{[n]} = I$ when acting on Π_n. In the following we consider only the family of LQI. The definition reveals that $\mathcal{B}_n^{(r)} f$ is a linear combination of derivatives of $\mathcal{B}_n f$. Furthermore, $\mathcal{B}_n^{(r)}$ $(0 \le r \le n)$ has the nice property to preserve polynomials of degree up to r, because, for $p \in \Pi_r$, we have

$$\begin{aligned}\mathcal{B}_n^{(r)} p &= \left(\mathcal{A}_n^{(r)} \circ \mathcal{B}_n\right) p = \sum_{k=0}^{r} \alpha_{n,k} D^k \underbrace{(\mathcal{B}_n p)}_{\in \Pi_r} = \sum_{k=0}^{n} \alpha_{n,k} D^k \left(\mathcal{B}_n p\right) \\ &= \left(\mathcal{A}_n^{-1} \circ \mathcal{B}_n\right) p = p.\end{aligned}$$

In many instances there holds $L_n^{(r)} f - f = O\left(n^{-\lfloor r/2+1 \rfloor}\right)$ as $n \to \infty$.

Unfortunately, the Favard operator as well as its Durrmeyer variant doesn't let invariant the spaces Π_j, for $0 \le j \le n$. However, under appropriate assumptions on the sequence (σ_n) they do it asymptotically up to a remainder which decays exponentially fast as n tends to infinity. Writing $\simeq$ for this "asymptotic equality" we obtain, for fixed $n \in \mathbb{N}$,

$$\begin{aligned} F_{n,\sigma_n} p_k &\simeq e_k \\ \text{with } p_k &= k! \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j \frac{\sigma_n^{2j}}{2^j j! \, (k-2j)!} e_{k-2j},\end{aligned}$$

where e_m denote the monomials $e_m(t) = t^m$ $(m = 0, 1, 2, \ldots)$. Hence, for the inverse,

$$\left(F_{n,\sigma_n}\right)^{-1} e_k \simeq p_k = \sum_{j=0}^{\lfloor k/2 \rfloor} \underbrace{(-1)^j \frac{\sigma_n^{2j}}{2^j j!}}_{=\alpha_{n,2j}} D^{2j} e_k$$

Note that $\beta_{n,2k+1} = \alpha_{n,2k+1} = 0$ $(k = 0, 1, 2, \ldots)$ and that neither $\beta_{n,k}$ nor $\alpha_{n,k}$ depend on the variable x. The analogous results for the Favard-Durrmeyer operators are similar. Proceeding in this way we define the following operators:

Definition 3.1 (Favard quasi-interpolants). *The left quasi-interpolants* $F_{n,\sigma_n}^{(r)}$ *and* $\tilde{F}_{n,\sigma_n}^{(r)}$ $(r = 0, 1, 2, \ldots)$ *of the Favard and Favard-Durrmeyer operators, respectively, are given by*

$$F_{n,\sigma_n}^{(r)} = \sum_{k=0}^{r} \alpha_{n,k} D^k F_{n,\sigma_n} := \sum_{k=0}^{\lfloor r/2 \rfloor} (-1)^k \frac{\sigma_n^{2k}}{2^k k!} D^{2k} F_{n,\sigma_n}$$

and

$$\tilde{F}_{n,\sigma_n}^{(r)} = \sum_{k=0}^{r} \tilde{\alpha}_{n,k} D_{n,\sigma_n}^k \tilde{F}_{n,\sigma_n} := \sum_{k=0}^{\lfloor r/2 \rfloor} (-1)^k \frac{\sigma_n^{2k}}{k!} D^{2k} \tilde{F}_{n,\sigma_n}.$$

Remark 3.2. Note that $F_{n,\sigma_n}^{(2r)} = F_{n,\sigma_n}^{(2r+1)}$ and $\tilde{F}_{n,\sigma_n}^{(2r)} = \tilde{F}_{n,\sigma_n}^{(2r+1)}$ $(r = 0, 1, 2, \ldots)$.

The local rate of convergence is given by the next theorem.

Theorem 3.3. *Let $\ell \in \mathbb{N}_0$, $q \in \mathbb{N}$ and $x \in \mathbb{R}$. Suppose that the real sequence (σ_n) satisfies condition* (2.1)*. For each function $f \in W[2(\ell+q);x]$, the following complete asymptotic expansions are valid as $n \to \infty$:*

$$\left(F_{n,\sigma_n}^{(2r)} f\right)^{(\ell)}(x) \sim f^{(\ell)}(x) + (-1)^r \sum_{k=r+1}^{\infty} \binom{k-1}{r} \frac{f^{(2k+\ell)}(x)}{(2k)!!} \sigma_n^{2k}$$

and

$$\left(\tilde{F}_{n,\sigma_n}^{(2r)} f\right)^{(\ell)}(x) = f^{(\ell)}(x) + (-1)^r \sum_{k=1}^{q} \binom{k-1}{r} \frac{f^{(2k+\ell)}(x)}{k!} \sigma_n^{2k} + o\left(\sigma_n^{2q}\right).$$

Remark 3.4. An immediate consequence are the asymptotic relations

$$\left(F_{n,\sigma_n}^{(2r)} f\right)(x) - f(x) = O\left(\sigma_n^{2(r+1)}\right)$$

and

$$\left(\tilde{F}_{n,\sigma_n}^{(2r)} f\right)(x) - f(x) = O\left(\sigma_n^{2(r+1)}\right)$$

as $n \to \infty$.

Acknowledgment. The author would like to thank the anonymous referee for the very careful reading of the manuscript. The valuable comments and suggestions led to several improvements in the paper.

References

[1] Abel, U., Butzer, P.L., *Complete asymptotic expansion for generalized Favard operators*, to appear in Constr. Approx.

[2] Becker, M., *Inverse theorems for Favard operators in polynomial weight spaces*, Ann. Soc. Math. Pol., Ser. I, Commentat. Math., **22**(1981), 165-173.

[3] Becker, M., Butzer, P.L., Nessel, R.J., *Saturation for Favard operators in weighted function spaces*, Studia Math., **59**(1976), 139-153.

[4] Butzer, P.L., and Nessel, R.J., *Fourier analysis and approximation. Vol. 1: One-dimensional theory*, Birkhäuser, Basel and Stuttgart, 1971.

[5] Favard, J., *Sur les multiplicateurs d'interpolation*, J. Math. Pures Appl., IX, **23**(1944), 219-247.

[6] Favard, J., *Sur les fonctions harmoniques presque périodiques*, Journ. de Math., **9**(1927), no. 6, 229-336.

[7] Gawronski, W., Stadtmüller, U., *Approximation of continuous functions by generalized Favard operators*, J. Approx. Theory, **34**(1982), 384-396.

[8] Kageyama, Y., *Generalization of the left Bernstein quasi-interpolants*, J. Approx. Theory, **94**(1998), 306-329.

[9] Kageyama, Y., *A new class of modified Bernstein operators*, J. Approx. Theory, **101**(1999), 121-147.

[10] Müller, M.W., *The central approximation theorems for the method of left gamma quasi-interpolants in L_p spaces*, J. Comput. Anal. Appl., **3**(2001), 207-222.

[11] Nowak, G., Sikorska-Nowak, A., *On the generalized Favard-Kantorovich and Favard-Durrmeyer operators in exponential function spaces*, J. Inequal. Appl., 2007, Article ID 75142, 12 p.

[12] Pych-Taberska, P., *On the generalized Favard operators*, Funct. Approximatio, Comment. Math., **26**(1998), 265-273.

[13] Pych-Taberska, P., Nowak, G., *Approximation properties of the generalized Favard-Kantorovich operators*, Ann. Soc. Math. Pol., Ser. I, Commentat. Math., **39**(1999), 139-152.

[14] Sablonnière, P., *Representation of quasi-interpolants as differential operators and applications*, In: "New Developments in Approximation Theory" (International Series of Numerical Mathematics, Vol. 132), (M. W. Müller, M. Buhmann, D. H. Mache, and M. Felten, eds.) Birkhäuser-Verlag, Basel 1998, pp. 233-253.

[15] Voronovskaja, E., *Détermination de la forme asymptotique de l'approximation des fonctions par les polynômes de M. Bernstein*, C. R. Acad. Sc. URSS, 1932, 79-85.

[16] Sikkema, P. C., *On some linear positive operators*, Indag. Math., **32**(1970), 327-337.

Ulrich Abel
Technische Hochschule Mittelhessen
University of Applied Sciences
Fachbereich MND
Wilhelm-Leuschner-Straße 13
61169 Friedberg
Germany
e-mail: `ulrich.abel@mnd.th-mittelhessen.de`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 207–218

Applying the Backus-Gilbert theory to function approximation

Flavian Abramovici

Abstract. In this paper are given new results within the project I started some years ago, of using inverse problems methods for recovering the values at points $\mathbf{x}_0$ of a continuous function f with compact support $\mathbb{E} \subseteq \mathbb{R}^m$, when N of its values are given at the nodes $\mathbf{x}_i$. After showing in [1] how to obtain Shepard's formula with two different versions of the well known Backus-Gilbert process, building averaging kernels that resemble δ - "functions" centered at the nodes and consist in linear combinations of the data representers. In the present paper I am showing how to attach a spread to the Shepard formula itself, leading to a convergence theorem concerning the recovery of the considered function.

Mathematics Subject Classification (2010): 41A30.

Keywords: Backus-Gilbert theory, Shepard's formula, deltaness, inverse problems, moving least-squares.

1. Introduction

In order to use the classical Backus-Gilbert process, the data should consist in a set of bounded functionals. Not having such functionals, I tried to use the internal products

$$\int_{\mathbb{R}^m} f(\mathbf{x}) G_i^{(\lambda)}(\mathbf{x}) dV \tag{1.1}$$

between the given function f and the elements of a Dirac sequence ([9]), which for high enough λ could approximate such functionals, e.g. the elements of the Dirac sequence used in [1]

$$G^{(\lambda)}(\mathbf{x}) = \begin{cases} \lambda^m & \text{for } \mathbf{x} \in \mathbb{S} \\ \\ 0 & \text{otherwise} \end{cases}, \tag{1.2}$$

where $\mathbb{S}$ is the reqular hypercube having the center at the origin and edges of length $\frac{1}{\lambda}$, with λ a positive real parameter. The Backus-Gilbert classical theory is looking for the optimal linear combination of the form

$\sum_{i=1}^{N} a_i^{(\lambda)}(\mathbf{x}_0) G_i^{(\lambda)}(\mathbf{x})$ with $G_i^{(\lambda)}(\mathbf{x}) = G^{(\lambda)}(\mathbf{x} - \mathbf{x}_i)$, that gives the best approximation of the value of the function f at $\mathbf{x}_0$

$$\begin{aligned} \tilde{f}^{(\lambda)}(\mathbf{x}_0) &= \sum_{i=1}^{N} \int_{\mathbb{R}^m} a_i^{(\lambda)}(\mathbf{x}_0) f(\mathbf{x}) G_i^{(\lambda)}(\mathbf{x}) dV \\ &= \sum_{i=1}^{N} \int_{\mathbb{R}^m} a_i^{(\lambda)}(\mathbf{x}_0) f(\mathbf{x}) G^{(\lambda)}(\mathbf{x} - \mathbf{x}_i) dV \ , \end{aligned} \tag{1.3}$$

with dV the volume element $dx_1, \ldots, dx_m$. Taking the limit of the result for $\lambda \to \infty$ in order to compensate for the errors involved in using only finite values of λ, I obtained the well-known Shepard's formula [7],[12],[13]:

$$\lim_{\lambda \to \infty} \tilde{f}^{(\lambda)}(\mathbf{x}_0) = \sum_{k=1}^{N} \frac{1}{||\mathbf{x}_k - \mathbf{x}_0||^2 \sum_{i=1}^{N} \frac{1}{||\mathbf{x}_i - \mathbf{x}_0||^2}} \ . \tag{1.4}$$

This result surprised some people working in seismology and others working in numerical analysis, as the Backus-Gilbert theory [2],[3],[4] was known mainly to geophysicists, while Shepard's formula was familar to mathematicians working in numerical analysis. In order to make sure that my results were correct, I looked for another approximation of $f(\mathbf{x}_0)$ having all the ingredients of the Backus-Gilbert theory including a spread. What I actually did was to discretize the integrals involved, obtaining the discrete version of the Backus-Gilbert theory. Following closely the way the classical theory was built, I applied the Backus-Gilbert linear representation theorem marked below as Theorem 2.1, finding that necessarily the average given by my discrete version had to be of the form obtained by discretization. After finishing the preliminary report, I passed copies to people who showed a special interest in my results and with whom I had many discussions, among them Prof. Kes Salkauskas from the University of Calgary in Canada and Prof. David Levin from the Hebrew University in Jerusalem. They extended my findings obtaining new results, e.g. making the same steps I did including taking a limit, Bos and Salkoskas obtained in [5] the moving least-squares approximation [8], a generalisation of Shepard's formula. For this purpose they defined a special type of Dirac sequences called regular, in order to handle the quadratic integrals needed for the optimal solutions. On the other hand David Levin using not only my results but also those of Bos and Salkauskas, presented an elegant way to obtain the moving least squares process using block matrices and applied with great success in [10] my discrete Backus-Gilbert process to scattered interpolation, smoothing and numerical differentition and, in [11], to numerical integration.

Once I obtained the Shepard formula using two independent Backus-Gilbert theories and being further stimulated by the results obtained by Bos, Salkauskas and Levin, I decided to try to settle an important question that was still open: is it possible to attach to the Shepard approximation also a sort of spread, which would lead to further properties? Although both Backus-Gilbert theories led to Shepard's formula, only within the discrete one I could define a spread, as for $\lambda \to \infty$ the process was divergent, the spread tending to infinity. The solution came to me while reading the paper by Bos and Salkauskas, as I realized that the simple Dirac sequence I used was regular and therefore what I had to do was "only" to modify a little the classical Backus-Gilbert spread by normalizing its integrand. As a result, the integral representing the spread for every λ became convergent, the limit having all the characteristic properties of a spread. It was therefore justified to define this limit as being the Backus-Gilbert spread of Shepard's formula. Moreover, it turned out that the limit obtained as described coincides with the spread attached to the Shepard formula within the discrete theory directly, not in combination with taking a limit and with its own justification. For the benefit of those not familiar with the classical Backus-Gilbert theory, a short description is given in Section 2. In Section 3 is shown how the Shepard formula is obtained and in the last Section is described the discrete Backus-Gilbert version.

2. The classical Backus-Gilbert theory

Clearly, if we have only a finite number of data, it is not possible to determine exactly the properties of the Earth at every location, but it may be possible to get averages of the so called "Earth models", functions f belonging to the Hilbert space $\mathcal{H} = L_2(\mathbb{E})$ with $\mathbb{E}$ the closed, connected and bounded support of f in $\mathbb{R}^m$ representing the properties of the Earth. Hence, the most we may hope to achieve using the data described in Section 1 is to find significant quantities that characterize the entire family of models, e.g. the average f_{av} at every point $\mathbf{x}_0$ of $\mathbb{E}$, corresponding to an averaging kernel $\mathcal{A}(\mathbf{x}_0, \mathbf{x})$:

$$f_{av}(\mathbf{x}_0) = \int_{\mathbb{E}} \mathcal{A}(\mathbf{x}_0, \mathbf{x}) f(\mathbf{x}) dV \simeq f(\mathbf{x}_0) \ . \tag{2.1}$$

Backus and Gilbert proved in [3] the following general result.

Theorem 2.1. *Let $f \in \mathcal{H}$ be a function for which N linearly independent bounded linear functionals γ_i on $\mathcal{H}$ are known. If it is possible to obtain a linear average $\mathcal{L}_{av}(f)$ of f at a point $\mathbf{x}_0$ using only the given N functionals, then the average $\mathcal{L}_{av}(f)$ is necessarily a linear combination of these functionals:*

$$\mathcal{L}_{av}(f) = \sum_{i=1}^{N} a_i \mathcal{L}_i(f) \tag{2.2}$$

with coefficients a_i that depend upon $\mathbf{x}_0$.

Using this result taking as $\mathcal{L}_i(f)$ the values $G_i(\mathbf{x})$, we find that

$$\mathcal{L}_{av}(f) = \sum_{i=1}^{N} a_i \int_{\mathbb{E}} G_i(\mathbf{x}) f(\mathbf{x}) dV = \int_{\mathbb{E}} \left[\sum_{i=1}^{N} a_i G_i(\mathbf{x}) \right] f(\mathbf{x}) dV \ , \tag{2.3}$$

i.e. the averaging kernels are indeed linear combination of the representers $G_i(\mathbf{x})$:

$$\mathcal{A}(\mathbf{x}_0, \mathbf{x}) = \sum_{i=1}^{N} a_i(\mathbf{x}_0), G_i(\mathbf{x}) \ . \tag{2.4}$$

Consequently, Backus and Gilbert looked for an optimal unimodular averaging kernel $\mathcal{A}(\mathbf{x}_0, \mathbf{x})$ i.e satisfying the condition

$$\int_{\mathbb{E}} \mathcal{A}(\mathbf{x}_0, \mathbf{x}) dV = 1 \tag{2.5}$$

and having the highest deltaness, that is the highest likeness to the Dirac δ-function centered at $\mathbf{x}_0$, condition checked by using the "spread" of the average kernel defined, by Backus and Gilbert as follows.

Definition 2.2. For every averaging kernel $\mathcal{A}(\mathbf{x}_0, \mathbf{x})$ on a compact set $\mathbb{E} \subseteq \mathbb{R}^m$, the function

$$s_0 = s(\mathbf{x}_0) = \frac{12}{m} \int_{\mathbb{E}} J(\mathbf{x}_0, \mathbf{x}) \mathcal{A}^2(\mathbf{x}_0, \mathbf{x}) dV \tag{2.6}$$

with $J(\mathbf{x}_0, \mathbf{x})$ a "sink" function i.e. a non-negative function that vanishes for $\mathbf{x} = \mathbf{x}_0$ and grows rapidly away from this point, is called a spread of $\mathcal{A}$ at $\mathbf{x}_0$. A typical "sink" function is $J(\mathbf{x}_0, \mathbf{x}) = ||\mathbf{x} - \mathbf{x}_0||^2$, with $||\mathbf{x} - \mathbf{x}_0||$ the Euclidean norm.

Thus, the Backus-Gilbert process solves the following variational problem: *find the coefficients $a_i(\mathbf{x}_0)$ for which the averaging kernel has the highest δ-ness, i.e. the smallest spread.*

Using a Lagrange multiplier, Backus and Gilbert solved this classical variational problem, obtaining the following relation giving the coefficients of the averaging kernel:

$$\mathbf{a}(\mathbf{x_0}) = \frac{1}{\mathbf{u}^T \left[\mathbf{Z}(\mathbf{x_0})\right]^{-1} \mathbf{u}} \left[\mathbf{Z}(\mathbf{x_0})\right]^{-1} \mathbf{u} \tag{2.7}$$

with $\mathbf{Z}(\mathbf{x}_0)$ a Gram matrix [6] of components

$$Z_{ik}(\mathbf{x}_0) = \frac{12}{m} \int_{\mathbf{E}} J(\mathbf{x}_0, \mathbf{x}) G_i(\mathbf{x}) G_k(\mathbf{x}) dV \ , \tag{2.8}$$

the corresponding spread $s(\mathbf{x}_0)$ being given by

$$s_0 = \mathbf{a}(\mathbf{x_0})^T \mathbf{Z}(\mathbf{x_0}) \mathbf{a}(\mathbf{x_0}) \ , \tag{2.9}$$

Consider now the functions $f \in \mathcal{H}$ for which the following integrals

$$\int_{\mathbf{E}} \frac{|f(\mathbf{x}) - f(\mathbf{x}_0)|}{||\mathbf{x} - \mathbf{x}_0||} dV \quad \text{and} \quad \int_{\mathbf{E}} \frac{|f(\mathbf{x}) - f(\mathbf{x}_0)|^2}{||\mathbf{x} - \mathbf{x}_0||^2} dV \tag{2.10}$$

are finite. Defining the square root of the second integral as being the "$\mathbf{x}_0$-norm" $||f(\mathbf{x})||_{\mathbf{x}_0}$ of a function f that belongs to $\mathcal{H}$ and is either identically zero or not constant, Backus and Gilbert proved the following result:

Theorem 2.3. *Under the conditions described above. if the averaging kernel* $\mathcal{A}(\mathbf{x}_0, \mathbf{x})$ *given by (2.4) is unimodular, the error functional*

$$\mathcal{E}(f(\mathbf{x}_0)) = \int_{\mathbf{E}} \mathcal{A}(\mathbf{x}_0, \mathbf{x}) \left[f(\mathbf{x}) - f(\mathbf{x}_0)\right] dV \tag{2.11}$$

is bounded, its x_0*-norm being given by*

$$||\mathcal{E}||_{\mathbf{x}_0} = \sqrt{\frac{m}{12} s(\mathbf{x}_0)} \ , \tag{2.12}$$

with $s(\mathbf{x}_0)$ *the considered spread.*

Corollary 2.4. *(The boundedness inequality for a given* λ*) (For every* $\lambda > 0$ *the following inequality takes place:*

$$|\mathcal{E}(f(\mathbf{x}_0))| \leq ||\mathcal{E}||_{\mathbf{x}_0} ||f||_{\mathbf{x}_0} \tag{2.13}$$

Corollary 2.5. *(A convergence theorem) Let* $\mathcal{A}^{(\lambda)}(\mathbf{x}_0, \mathbf{x}) = \sum_{i=1}^{N} a_i^{(\lambda)}(\mathbf{x}_0) G_i^{(\lambda)}(\mathbf{x})$ *be a family of unimodular averaging kernels. The average* $f^{(\lambda)}(\mathbf{x}_0)$ *tends to* $f(\mathbf{x}_0)$ *when* $\lambda \to \mu$*, if and only if* $\lim_{\lambda \to \mu} s^{(\lambda)}(\mathbf{x}_0) = 0$*, in particular* $f^{(\lambda)}(\mathbf{x}_0)$ *tends to* $f(\mathbf{x}_0)$ *when* $\lambda \to \infty$ *if and only if* $s^{(\lambda)}(\mathbf{x}_0) \to 0$ *when* $\lambda \to \infty$*.*

Remark 2.6. This theorem is important as a general result but in many cases the computer time needed to reach the wanted precision is very large. This is why it is important to have an efficient method, giving an effective growth of accuracy at every step, which is precisely why the Backus-Gilbert process is preferable to other methods, as one may see on the examples given by David Levin in the articles mentioned above.

3. Approximating a function with given values using the classical Backus-Gilbert theory

Using the general relations (2.7) - (2.8), we prove the following result:

Theorem 3.1. *Let* f *be a unimodular Earth model satisfying the conditions of Theorem 2.1. For every set of data of the form*

$$\tilde{f}^{(\lambda)}(\mathbf{x}_i) = \int_{\mathbb{E}_i} \tilde{G}_i^{(\lambda)}(\mathbf{x}) f(\mathbf{x}) dV \ , \tag{3.1}$$

the coefficients that minimize the spread of the averaging kernel are

$$\tilde{\mathbf{a}}^{(\lambda)}(\mathbf{x}_0) = \frac{1}{\left[\tilde{\mathbf{u}}^{(\lambda)}\right]^T \left[\tilde{\mathbf{Z}}^{(\lambda)}(\mathbf{x}_0)\right]^{-1} \tilde{\mathbf{u}}^{(\lambda)}} \left[\tilde{\mathbf{Z}}^{(\lambda)}(\mathbf{x}_0)\right]^{-1} \tilde{\mathbf{u}}^{(\lambda)} \ . \tag{3.2}$$

with $\tilde{u}^{(\lambda)} = 1$ *for every* λ *and*

$$\tilde{\mathbf{Z}}_{ik}^{(\lambda)}(\mathbf{x}_0) = \int_{\mathbf{R}^m} J(\mathbf{x}_0, \mathbf{x}) \tilde{G}_i^{(\lambda)}(\mathbf{x}) \tilde{G}_k^{(\lambda)}(\mathbf{x}) dV \ . \tag{3.3}$$

As for $\lambda > 0$ large enough, the matrix $\tilde{\mathbf{Z}}^{(\lambda)}(\mathbf{x}_0)$ is diagonal with diagonal elements $\lambda^2 \tilde{X}_k^{(\lambda)}(\mathbf{x}_0)$ where

$$\tilde{X}_k^{(\lambda)}(\mathbf{x}_0) = ||\mathbf{x}_k - \mathbf{x}_0||^2 + \frac{1}{\lambda^2} \ . \tag{3.4}$$

Making λ tend to infinity, we find that

$$\check{\mathbf{a}}(\mathbf{x}_0) = \lim_{\lambda\to\infty} \tilde{a}^{(\lambda)}(\mathbf{x}_0) = \frac{1}{\sum\limits_{i=1}^{N} \frac{1}{||\mathbf{x}_i - \mathbf{x}_0||^2}} \begin{pmatrix} \frac{1}{||\mathbf{x}_1 - \mathbf{x}_0||^2} \\ \vdots \\ \frac{1}{||\mathbf{x}_N - \mathbf{x}_0||^2} \end{pmatrix} , \tag{3.5}$$

i.e. we arrive to the following result:

Corollary 3.2. *The considerate Earth model is approximated by Shepard's formula*

$$\tilde{f}(x_0) = \frac{1}{\sum\limits_{i=1}^{N} \frac{1}{||\mathbf{x}_i - \mathbf{x}_0||^2}} \sum_{k=1}^{N} \frac{f(\mathbf{x}_k)}{||\mathbf{x}_k - \mathbf{x}_0||^2} \quad . \tag{3.6}$$

Having obtained the Shepard formula using the Backus-Gilbert theory, it is only natural to try to attach to it a Backus-Gilbert spread for characterizing the way Shepard's formula approximates the function f. In order to do so, we calculate the spread of the optimal averaging kernel for a given λ:

$$\begin{aligned} \tilde{s}^{(\lambda)}(\mathbf{x}_0) = {} & \frac{12}{m} \sum_{k=1}^{N} \left[\tilde{a}_k^{(\lambda)}(\mathbf{x}_0)\right]^2 \int_{\mathbb{E}_k} J(\mathbf{x}_0, \mathbf{x}) \left[G_k^{(\lambda)}(\mathbf{x})\right]^2 dV \\ & + \frac{12}{m} \sum_{k,l=1}^{N} \tilde{a}_k^{(\lambda)}(\mathbf{x}_0)\tilde{a}_\ell^{(\lambda)}(\mathbf{x}_0) \int_{\mathbb{E}_k \bigcap \mathbb{E}_\ell} J(\mathbf{x}_0, \mathbf{x})\tilde{G}_k^{(\lambda)}(\mathbf{x}_0)\tilde{G}_\ell^{(\lambda)}(\mathbf{x}_0) dV \ . \end{aligned} \tag{3.7}$$

and see if it tends to a finite limit when λ tends to infinity. For λ large enough the intersection $\mathbb{E}_k \bigcap \mathbb{E}_\ell$ is empty, so that the double sum in the second term of the right hand side is equal to zero as it is easy to see and therefore we are left with

$$\tilde{s}^{(\lambda)}(\mathbf{x}_0) \quad = \quad \frac{12}{m} \sum_{k=1}^{N} \left[\tilde{a}_k^{(\lambda)}(\mathbf{x}_0)\right]^2 \int_{\mathbb{E}_k} J(\mathbf{x}_0, \mathbf{x}) \left[G_k^{(\lambda)}(\mathbf{x})\right]^2 dV \ , \tag{3.8}$$

a divergent integral ! However, as already explained, the simple Dirac sequence $G^{(\lambda)}(\mathbf{x})$ is regular, according to the following definition.

Definition 3.3. A Dirac sequence $G^{(\lambda)}(\mathbf{x})$ is called regular if the following conditions hold:

1. $G^{(\lambda)} \in L_2(\mathbb{R}^m)$.
2. For every bounded and continuous function $f \in L_2(\mathbb{R}^m)$

$$\lim_{\lambda\to\infty} \int_{\mathbb{R}^m} \frac{G^{(\lambda)}(\mathbf{x}-\mathbf{a})G^{(\lambda)}(\mathbf{x}-\mathbf{b})}{\kappa_m^{(\lambda)}} f(\mathbf{x}) dV = \begin{cases} \mathbf{0} & \text{if} \quad \mathbf{b} \neq \mathbf{a} \\ f(\mathbf{a}) & \text{if} \quad \mathbf{b} = \mathbf{a} \end{cases} \tag{3.9}$$

with

$$\kappa_m^{(\lambda)} = \int_{\mathbb{R}^m} \left[G^{(\lambda)}(\mathbf{x})\right]^2 dV \ . \tag{3.10}$$

Dividing both sides of relation (38) by $\tilde{\kappa}_m^{(\lambda)}$ and knowing that the ratio $\frac{\tilde{s}^{(\lambda)}(\mathbf{x}_0)}{\tilde{\kappa}_m^{(\lambda)}}$ is convergent for $\lambda \to \infty$, we find that the right hand side is convergent and therefore the left hand side is also convergent to $\check{s}(\mathbf{x}_0)$ that may be called the normalized spread of Shepard's formula:

$$\begin{aligned} \check{s}(\mathbf{x}_0) &= \frac{12}{m} \sum_{k=1}^{N} \lim_{\lambda\to\infty} \left[\tilde{a}_k^{(\lambda)}(\mathbf{x}_0)\right]^2 \lim_{\lambda\to\infty} \int_{\mathbf{E}_k} J(\mathbf{x}_0, \mathbf{x}) \frac{\left[\tilde{G}_k^{(\lambda)}(\mathbf{x})\right]^2}{\tilde{\kappa}_m^{(\lambda)}} dV \ . \\ &= \frac{12}{m} \sum_{k=1}^{N} \check{a}_k^2(\mathbf{x}_0) J(\mathbf{x}_0, \mathbf{x}_k) \ . \end{aligned} \tag{3.11}$$

Moreover, we may attach to Shepard's formula a boudedness inequality leading to a convergence property. Indeed, consider the boundedness inequality (2.13) for a the minimal solution for any λ

$$|\tilde{\mathcal{E}}^{(\lambda)}(\tilde{f}^{(\lambda)})| \leq ||\tilde{\mathcal{E}}^{(\lambda)}||_{\mathbf{x}_0} \ ||\tilde{f}^{(\lambda)}||_{\mathbf{x}_0} \ , \tag{3.12}$$

the error functional $\tilde{\mathcal{E}}^{(\lambda)}(f^{(\lambda)})$ being defined by

$$\tilde{\mathcal{E}}^{(\lambda)}(f^{(\lambda)}) = \tilde{f}^{(\lambda)}(\mathbf{x}_0) - f(\mathbf{x}_0) = \sum_{i=1}^{N} \tilde{a}_i^{(\lambda)}(\mathbf{x}_0) \tilde{f}_i^{(\lambda)} - f(\mathbf{x}_0) \ , \tag{3.13}$$

with $\tilde{f}_i^{(\lambda)} = \int_{\mathbf{E}} \tilde{G}_i^{(\lambda)}(\mathbf{x}) f(\mathbf{x}) dV$. Using the expression of the averaging kernel (2.4) and its unimodularity, we find that

$$\lim_{\lambda\to\infty} \tilde{\mathcal{E}}^{(\lambda)}(f) = \lim_{\lambda\to\infty} \int_{\mathbf{E}} \tilde{\mathcal{A}}^{(\lambda)}(\mathbf{x}_0, \mathbf{x})[f(\mathbf{x}) - f(\mathbf{x}_0)] dV = \check{\mathcal{E}}(f) \ . \tag{3.14}$$

As to the right hand side, the first factor is equal to $\sqrt{\frac{m}{12} \tilde{s}^{(\lambda)}(\mathbf{x}_0)}$ and therefore tends to infinity when λ does, due to the presence of the spread according to (2.12). Dividing by κ_m the first factor of the right hand side becomes convergent but in this case we have to multiply the second factor also by κ_m, bringing the inequality to the following form:

$$|\tilde{\mathcal{E}}^{(\lambda)}(\tilde{f}^{(\lambda)})| \leq \sqrt{\frac{m}{12} \frac{\tilde{s}^{(\lambda)}(\mathbf{x}_0)}{\tilde{\kappa}_m^{(\lambda)}}} \ \sqrt{\tilde{\kappa}_m^{(\lambda)}} ||\tilde{f}^{(\lambda)}||_{\mathbf{x}_0} \ . \tag{3.15}$$

In order to prove that the right hand side is convergent and that its limit is also a "discrete" quantity, consider the "shrinking" function

$$F(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{x} = \mathbf{x}_0 \text{ or } \mathbf{x} \in \mathbb{E}_i \\ & \text{for } i = 1, \ldots, N \text{ and} \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

and its approximation $\tilde{F}^{(\lambda)}(\mathbf{x}_0) = \sum\limits_{i=1}^{N} \tilde{a}_i^{(\lambda)}(\mathbf{x}_0) \int_{\mathbb{E}_i} \tilde{G}_i^{(\lambda)}(\mathbf{y}) F(\mathbf{y}) dV$. For λ large enough this approximation also satisfies the boundedness inequality, in fact $\tilde{F}^{(\lambda)}(\mathbf{x}_0)$ is equal to $\tilde{f}^{(\lambda)}(\mathbf{x}_0)$ for every $\mathbf{x}_0$ different from any node $\mathbf{x}_i$ for $i = 1, \dots N$, so that the error functionals and the x_0-norms coincide. Taking into account that $\tilde{f}^{(\lambda)}$ is sectionally continuous we find, using the mean value theorem, that there exists at least one point $\xi_i^{(\lambda)} \in \mathbb{E}_i$ that depends upon λ, such that

$$\int_{\mathbb{E}_i} \frac{\left|\tilde{f}^{(\lambda)}(\mathbf{x}) - \tilde{f}^{(\lambda)}(\mathbf{x}_0)\right|^2}{(\mathbf{x} - \mathbf{x}_0)^2} dx = \frac{\left|\tilde{f}^{(\lambda)}(\xi_i^{(\lambda)}) - \tilde{f}^{(\lambda)}(\mathbf{x}_0)\right|^2}{(\xi_i^{(\lambda)} - \mathbf{x}_0)^2} \epsilon \tag{3.17}$$

and therefore

$$\sqrt{\tilde{\kappa}_1^{(\lambda)}} ||\tilde{F}^{(\lambda)}||_{\mathbf{x}_0} = \sqrt{\sum_{i=1}^{N} \frac{\left|\tilde{f}^{(\lambda)}(\xi|_i^{(\lambda)}) - \tilde{f}^{(\lambda)}(\mathbf{x}_0)\right|^2}{(\xi_i^{(\lambda)} - \mathbf{x}_0)^2}}\,, \tag{3.18}$$

as $\lambda\epsilon = 1$. Using this relation we get from (3.15) the inequality

$$|\tilde{\mathcal{E}}^{(\lambda)}(\tilde{f}^{(\lambda)})| \leq \sqrt{\frac{m\tilde{s}^{(\lambda)}(\mathbf{x}_0)}{12\tilde{\kappa}_1^{(\lambda)}}} \sqrt{\sum_{i=1}^{N} \frac{\left|\tilde{f}^{(\lambda)}(\xi_i^{(\lambda)}) - \tilde{f}^{(\lambda)}(\mathbf{x}_0)\right|^2}{(\xi_i^{(\lambda)} - \mathbf{x}_0)^2}}\,, \tag{3.19}$$

with the first factor of the right hand side tending to the "discrete" quantity $\sqrt{\frac{m}{12}\check{s}(\mathbf{x}_0)}$. On the other hand when λ tends to infinity, ϵ tends to zero so that $\xi_i^{(\lambda)}$ tends to the center $\mathbf{x}_i$ of $\mathbb{E}_i$ and $\tilde{f}^{(\lambda)}$ is continuous at $\mathbf{x}_i$. Therefore, $\lim\limits_{\lambda\to\infty} \tilde{f}^{(\lambda)}(\xi) = \tilde{f}^{(\lambda)}(\mathbf{x}_i)$ implying that

$$\lim_{\lambda\to\infty} \sqrt{\sum_{i=1}^{N} \frac{\left|\tilde{f}^{(\lambda)}(\xi_i^{(\lambda)}) - \tilde{f}^{(\lambda)}(\mathbf{x}_0)\right|^2}{(\xi_i^{(\lambda)} - \mathbf{x}_0)^2}} = \sqrt{\sum_{i=1}^{N} \frac{|f(\mathbf{x}_i) - f(\mathbf{x}_0)|^2}{(\mathbf{x}_i - \mathbf{x}_0)^2}}\,. \tag{3.20}$$

As a result, we get using (3.13) the following boundedness inequality for the discrete error functional:

$$|\check{\mathcal{E}}(f)| \leq \sqrt{\frac{\check{s}(\mathbf{x}_0)}{12}} \sqrt{\sum_{i=1}^{N} \frac{|f(\mathbf{x}_i) - f(\mathbf{x}_0)|^2}{(\mathbf{x}_i - x_0)^2}}\,. \tag{3.21}$$

making possible to define the following discrete quantity.

Definition 3.4. *The discrete* $\mathbf{x}_0$*-norm of a function* $f(\mathbf{x})$ *corresponding to the points* $x_1, \dots, x_N \in \mathbb{E}$ *is*

$$\|f\|_{\mathbf{x}_0} = \sqrt{\sum_{k=1}^{N} \frac{|f(\mathbf{x}_k) - f(\mathbf{x}_0)|^2}{(\mathbf{x}_k - \mathbf{x}_0)^2}}\,. \tag{3.22}$$

As a result, the discrete boundedness inequality (3.21) becomes

$$|\check{\mathcal{E}}(f)| \leq \sqrt{\frac{\check{s}(\mathbf{x}_0)}{12}}\|f\|_{\mathbf{x}_0} \tag{3.23}$$

leading to the following conclusion:

Corollary 3.5. *The continuous boundedness inequality* (3.12) *written in the form*

$$|\tilde{\mathcal{E}}^{(\lambda)}(\tilde{f}^{(\lambda)})| \leq \sqrt{\frac{\check{s}(\mathbf{x}_0)}{12}}||\tilde{f}^{(\lambda)}||_{\mathbf{x}_0} \tag{3.24}$$

tends to the discrete boundedness inequality (3.23) *when $\lambda \to \infty$, i.e. the left hand side of* (3.24) *tends to the left hand side of* (3.23) *and similarly for the right hand sides.*

Corollary 3.6. *The discrete error functional tends to zero if the discrete spread tends to zero.*

4. A discrete Backus-Gilbert theory

In this Section are presented formally without going into all the proofs, the main definitions and properties of my version of the Backus-Gilbert discrete process, built by similarity with the classical process: similar discrete spread, discrete $\mathbf{x}_0$-norm and discrete boundedness inequality with similar properties. As already explained, my initial intention was just to check the results obtained applying the classical theory combined with taking the limit for λ tending to infinity, but it turned out from my own results as well as from those of other people, that the discrete theory is very effective and gives very good numerical results in many cases. Moreover, I found that the way to write the corresponding approximation is dictated by Theorem 1.2 like in the continuous case, the difference being, of course the different data set. As a result, one finds that the form one writes usually a discrete average

$$f_{av}(\mathbf{x}_0) = \sum_{i=1}^{N} a_i(\mathbf{x}_0) f(\mathbf{x}_i) \ . \tag{4.1}$$

is the only possible one under the adopted assumptions.

Definition 4.1. *Let $\mathcal{A} = \mathcal{A}(\mathbf{x}_0)$ be a set of real non-negative numbers a_i*

$$\mathcal{A} = \{a_i(\mathbf{x}_0)\}_{i=1}^{N} \ , \tag{4.2}$$

used as coefficients of the average value of a function $f \in \mathcal{H}$ at a point $\mathbf{x}_0 \in \mathbb{E} \subset \mathbb{R}^m$. The set $\mathcal{A}$ is called a discrete averaging kernel of f at $\mathbf{x}_0$ and the Euclidean norm of the vector $\mathbf{a}$ *having as components the coefficients a_i is called the norm $||\mathcal{A}|| = ||\mathcal{A}(\mathbf{x}_0)||$ of the averaging kernel $\mathcal{A}$.*

It is not difficult to prove the following property.

Theorem 4.2. *The $\mathbf{x}_0$-norm* (3.22) *of any sectionally continuous function f which is either zero identically or non-constant on $\mathbb{E}$, is indeed a norm.*

Definition 4.3. *For every discrete averaging kernel $\mathcal{A}$ at $\mathbf{x}_0$, one may define its discrete spread at $\mathbf{x}_0$*

$$s_0 = s(\mathbf{x}_0, \mathcal{A}) = \frac{12}{m} \sum_{i=1}^{N} J_i a_i^2(\mathbf{x}_0) \ , \tag{4.3}$$

the factors $J_i = J(\mathbf{x}_i, \mathbf{x}_0) = ||\mathbf{x}_i - \mathbf{x}_0||^2$ called the spread's coefficients, measuring the "location separation" between the nodes $\mathbf{x}_i$ and the target point $\mathbf{x}_0$.

Remark 4.4. Like with the "deltaness" property in the classical Backus-Gilbert theory, these coefficients depend upon the distances between the nodes $\mathbf{x}_i$ and the current point $\mathbf{x}_0$ through a "sink" function.

Definition 4.5. *A discrete averaging kernel $\mathcal{A} = \{a_i(\mathbf{x}_0)\}_{i=1}^{N}$ is called unimodular if $\sum\limits_{i=1}^{N} a_i(\mathbf{x}_0) = 1$.*

Theorem 4.6. *For every unimodular averaging kernel and every function f belonging to $\mathcal{C}_{\mathbb{E}}$ or to L^2, the error functional*

$$\mathcal{E}(f, \mathbf{x}_0) = f_{av}(\mathbf{x}_0) - \ f(\mathbf{x}_0) \tag{4.4}$$

with f_{av} given by (4.1) is bounded, satisfies the inequality

$$|\mathcal{E}(f, \mathbf{x}_0)|_{\mathbf{x}_0} \leq ||\mathcal{E}||_{\mathbf{x}_0} ||f||_{\mathbf{x}_0} \ , \tag{4.5}$$

and its $\mathbf{x}_0$-norm is given by

$$||\mathcal{E}||_{\mathbf{x}_0} = \sqrt{\frac{m}{12} s(\mathbf{x}_0, \mathcal{A})} \ . \tag{4.6}$$

Proof. Using (4.1) and the unimodularity of the averaging kernel, we may write

$$\mathcal{E}(f, \mathbf{x}_0) = \sum_{i=1}^{N} a_i(\mathbf{x}_0)[f(\mathbf{x}_i) - f(\mathbf{x}_0)] = \sum_{i=1}^{N} u_i v_i \tag{4.7}$$

with

$$u_i = \sqrt{J_i} a_i(\mathbf{x}_0) \quad \text{and} \quad v_i = \frac{f(\mathbf{x}_i) - f(\mathbf{x}_0)}{\sqrt{J_i}} \ , \tag{4.8}$$

so that applying the Cauchy-Schwarz inequality we get

$$|\mathcal{E}(f, \mathbf{x}_0)| \leq \sqrt{\sum_{i=1}^{N} J_i a_i^2(\mathbf{x}_0)} \sqrt{\sum_{i=1}^{N} \frac{|f(\mathbf{x}_i) - f(\mathbf{x}_0)|^2}{J_i}} \ . \tag{4.9}$$

However, the first factor on the right is the discrete spread and the second one is the $\mathbf{x}_0$-norm of f so that

$$|\mathcal{E}(f, \mathbf{x}_0)| \leq \sqrt{\frac{m}{12} s\left(\mathbf{x}_0, \mathcal{A}\right)} \ ||f||_{\mathbf{x}_0} \ , \tag{4.10}$$

Hence, the linear functional $\mathcal{E}(f, \mathbf{x}_0)$ is bounded, its norm $||\mathcal{E}||_{\mathbf{x}_0}$ being not larger than any of its upper bounds, in particular not larger than

$\sqrt{\frac{m}{12} s\left(\mathbf{x}_0, \mathcal{A}\right)}$, which proves inequality (4.5). As to (4.6), consider the function $p(\mathbf{x}) = ||\mathbf{x} - \mathbf{x}_0||^2 q(\mathbf{x})$ with $q(\mathbf{x})$ continuous in $\mathbb{E}$ and satisfying the condition $q(\mathbf{x}_i) = a_i(\mathbf{x}_0)$ for $i = 1, \ldots, N$. It turns out that the inequality (4.10) is actually an equality for $f(\mathbf{x}) = p(\mathbf{x})$:

$$|\mathcal{E}(p, \mathbf{x}_0)| = \sqrt{\frac{m}{12} s\left(\mathbf{x}_0, \mathcal{A}\right)}\ ||p||_{\mathbf{x}_0} \ . \tag{4.11}$$

Indeed, the error functional (4.4) corresponding to $f(\mathbf{x}) = p(\mathbf{x})$ is equal just to $f_{av}(\mathbf{x}_0)$ as $p(\mathbf{x}_0) = 0$, so that in this case we get using (4.1) $\mathcal{E}(p, \mathbf{x}_0) = \sum_{i=1}^{N} a_i p(\mathbf{x}_i)$. However, $p(\mathbf{x}_i) = ||\mathbf{x}_i - \mathbf{x}_0||^2 q(\mathbf{x}_i)$ and $||\mathbf{x}_i - \mathbf{x}_0||^2 = J_i$ while $q(\mathbf{x}_i) = a_i(\mathbf{x}_0)$, so that $p(\mathbf{x}_i) = J_i a_i(\mathbf{x}_0)$ and therefore

$$\mathcal{E}(p, \mathbf{x}_0) = \sum_{i=1}^{N} a_i^2(\mathbf{x}_0) J_i = \sqrt{\sum_{i=1}^{N} a_i^2(\mathbf{x}_0) J_i} \sqrt{\sum_{i=1}^{N} a_i^2(\mathbf{x}_0) J_i} \ . \tag{4.12}$$

On the other hand, using the definition of the $\mathbf{x}_0$-norm we find that

$$||p||_{\mathbf{x}_0} = \sqrt{\sum_{i=1}^{N} \frac{|p(\mathbf{x}_i) - p(\mathbf{x}_0)|^2}{J_i}} = \sqrt{\sum_{i=1}^{N} \frac{|p(\mathbf{x}_i)|^2}{J_i}} = \sqrt{\sum_{i=1}^{N} a_i^2(\mathbf{x}_0) J_1} \ , \tag{4.13}$$

enabling us to replace one of the square roots in (4.12) by the $\mathbf{x}_0$-norm $||p||_{\mathbf{x}_0}$, whereas using the definition (4.3) of the discrete spread,, we find that we may replace the second square root in the right hand side of (4.12) by $\sqrt{\frac{m}{12} s(\mathbf{x}_0, \mathcal{A})}$, obtaining precisely (4.11). □

Based on this result, we arrive to the following easy to prove pointwise convergence theorem.

Theorem 4.7. *For any sequence $\mathcal{A}^{(\nu)}$ $(\nu = 1, 2, \ldots)$ of $\mathbf{x_0}-$ restricted unimodular discrete averaging kernels, the sequence $f_{av}^{(\nu)}(\mathbf{x}_0)$ tends to the exact value $f(\mathbf{x}_0)$ for every $\mathbf{x}_0$, if and only if $\lim_{n\to\infty} s_0^{(\nu)} = \lim_{\nu\to\infty} s\left(\mathbf{x}_0, \mathcal{A}^{(\nu)}\right) = 0$.*

In order to obtain the optimal average one solve here also a variational problem using also a Lagrange multiplier, the coefficients and the multiplier satisfying the same system of equations

$$\frac{\partial \tau}{\partial a_k} = 0 \ \text{ for } \ k = 1, \ldots, N \text{ and } \frac{\partial \tau}{\partial \eta} = 0 \ . \tag{4.14}$$

Hence, the solution of this equation is

$$a_k = \frac{m\eta}{24||\mathbf{x}_0 - \mathbf{x}_k||2} \quad (k = 1, \ldots, N) \ . \tag{4.15}$$

with

$$\eta = \frac{24}{m} \left(\sum_{i=1}^{N} \frac{1}{||\mathbf{x}_0 - \mathbf{x}_i||^2} \right)^{-1} \ . \tag{4.16}$$

Substituting the obtained value of η into the expression of a_k, we obtain precisely the components of $\check{\mathbf{a}}(x_0)$ given by (3.5), where the components of this vector are denoted using i as index instead of k. Hence we get again Shepard formula.

References

[1] Abramovici, F., *The Shepard interpolation as the best average of a set of data*, School of Mathematical Sciences, Tel Aviv University, Israel, 1984.

[2] Backus, G.E., Gilbert, J.F., *Numerical application of a formalism for geophysical inverse problems*, Geophys. J. R. Astr. Soc., **13**(1967), 247-276.

[3] Backus, G.E., Gilbert, J.F., *The resolving power of gross earth data*, Geophys. J. R. Astr. Soc., **16**(1968), 169-205.

[4] Backus, G.E., Gilbert, J.F., *Uniqueness in the inversion of inaccurate gross earth data*, Phil. Trans. R. Soc. London ser. A, **266**(1970), 123-192.

[5] Bos, L.P., Salkauskas, K., *Moving least-squares are Backus-Gilbert optimals*, J. Approx. Theory, **59**(1989), 267-275.

[6] Courant, R., Hilbert, D., *Methods of mathematical physics*, Vol. I, Interscience Publishers, New York, 1959.

[7] Franke, R., *Scattered data interpolation: tests of some methods*, Math. Comp., **38**(1982), 181-200.

[8] Lancaster, P., Salkaukas, *Surfaces generated by moving least squares methods*, Math. Comp., **37**(1981), 141-158.

[9] Lang, S., *Undergraduate Analysis*, 1998.

[10] Levin, D., *The approximation power of moving least - squares*, Math. Comp., **67**(1998), 1517-1531.

[11] Levin, D., *Stable integration rules with scattered integration points*, J. Comp. Appl. Math. **112**(1999), 181-187.

[12] Shepard, D., *A two-dimensional interpolation function for irregularly spaced data*, Proc. 23-rd Nat. Conf. A.C.M., 1968, 517-523.

[13] Schumaker, L.L., *Fitting surfaces to scattered data*, In: Approximation Theory ed. by G.G. Lorentz, C. K. Chui and L. L. Schumaker, **II**(1976), 203-268.

Flavian Abramovici
University of Tel-Aviv
School of Mathematical Sciences
Ramat-Aviv, Israel
e-mail: `c35@post.tau.ac.il`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 219–235

Iterates of multidimensional Kantorovich-type operators and their associated positive C_0-semigroups

Francesco Altomare, Mirella Cappelletti Montano and Vita Leonessa

Abstract. In this paper we deepen the study of a sequence of positive linear operators acting on $L^1([0,1]^N)$, $N \geq 1$, that have been introduced in [3] and that generalize the multidimensional Kantorovich operators (see [15]). We show that particular iterates of these operators converge on $\mathscr{C}([0,1]^N)$ to a Markov semigroup and on $L^p([0,1]^N)$, $1 \leq p < +\infty$, to a positive contractive C_0-semigroup (that is an extension of the previous one). The generators of these C_0-semigroups are the closures of some partial differential operators that belong to the class of Fleming-Viot operators arising in population genetics.

Mathematics Subject Classification (2010): 41A36, 47D06, 47F05.

Keywords: Multidimensional Kantorovich operator, positive approximation process, iterate of operators, positive C_0-semigroup, approximation of semigroups, Fleming-Viot operator.

1. Introduction

In the paper [3] we introduced and studied a sequence $(C_n)_{n\geq 1}$ of positive linear operators on $L^1([0,1]^N)$, $N \geq 1$, that are a generalization of the multidimensional Kantorovich operators, first introduced in [15], and that also extend to a multidimensional setting another sequence of positive linear operators on $L^1([0,1])$ studied in [5] and [6].

The operators C_n, $n \geq 1$, offer the advantage to reconstruct any Lebesgue-integrable function on $[0,1]^N$ by means of its mean values on a finite numbers of sub-cells of $[0,1]^N$ that do not constitute a subdivision of $[0,1]^N$.

Both in [6] and in [11] particular iterates of the (generalized) Kantorovich operators have been also investigated in connection with the existence of related C_0-semigroups of operators on $\mathscr{C}([0,1])$ and on $L^1([0,1])$.

Then, it seemed quite natural to tackle similar problems in a multidimensional setting and for the operators C_n, $n \geq 1$.

By using different methods from those employed in [6] and [11], in fact we first show that there exists a Markov semigroup $(T(t))_{t\geq 0}$ on $\mathscr{C}([0,1]^N)$ such that

$$T(t)(f) = \lim_{n\to\infty} C_n^{\rho_n}(f) \qquad \text{in } \mathscr{C}([0,1]^N) \tag{1.1}$$

for any $f \in \mathscr{C}([0,1]^N)$, $t \geq 0$ and for any sequence $(\rho_n)_{n\geq 1}$ of positive integers such that $\rho_n/n \to t$ as $n \to \infty$.

The generator $(A, D(A))$ of the Markov semigroup is determined on a core of $D(A)$, namely on $\mathscr{C}^2([0,1]^N)$, where it coincides with the second-order elliptic differential operator

$$V_l(u)(x) := \frac{1}{2}\sum_{i=1}^{N} x_i(1-x_i)\frac{\partial^2 u}{\partial x_i^2}(x) + \sum_{i=1}^{N}\left(\frac{l}{2} - x_i\right)\frac{\partial u}{\partial x_i}(x)$$

$(u \in \mathscr{C}^2([0,1]^N), x = (x_i)_{1\leq i\leq N} \in [0,1]^N)$, where $l \in [0,2]$.

Accordingly, formula (1.1) provides a constructive approximation of the solutions to the abstract Cauchy problem

$$\begin{cases} \dfrac{\partial u}{\partial t}(x,t) = A(u(\cdot,t))(x) & x \in [0,1]^N, \quad t \geq 0, \\[2ex] u(x,0) = u_0(x) & u_0 \in D(A), \quad x \in [0,1]^N, \end{cases}$$

that, as it is well-known, are given by $u(x,t) = T(t)(u_0)(x)$ $(x \in [0,1]^N, t \geq 0)$.

The differential operator V_l falls in a class of Fleming-Viot operators arising in population genetics (see [2], [7], [10] for some additional references).

In addition, we also show that the subspace of all polynomials with a given degree and the subspace of all Hölder continuous functions on $[0,1]^N$ are invariant under $(T(t))_{t\geq 0}$. In some particular cases we finally show that the semigroup $(T(t))_{t\geq 0}$ can be extended to a positive contractive C_0-semigroup on $L^p([0,1]^N)$ for every $1 \leq p < +\infty$ and this semigroup can be equally approximated in the L^p-norm by iterates of the operators C_n, as in formula (1.1).

2. Preliminary results

Throughout this paper $[0,1]^N$ denotes the canonical hypercube in $\mathbf{R}^N$, $N \geq 1$, i.e.,

$$[0,1]^N := \{(x_i)_{1\leq i\leq N} \in \mathbf{R}^N \mid 0 \leq x_i \leq 1 \text{ for every } i = 1,\ldots,N\}.$$

As usual we denote by $\mathscr{C}([0,1]^N)$ the space of all real valued continuous functions on $[0,1]^N$ and by $\mathscr{C}^2([0,1]^N)$ the space of all real valued continuous functions on $[0,1]^N$ which are twice continuously differentiable in the interior of $[0,1]^N$ and whose partial derivatives up to the order two can be continuously extended on $[0,1]^N$. The space $\mathscr{C}([0,1]^N)$, endowed with the natural (pointwise) ordering and the sup-norm $\|\cdot\|_\infty$, is a Banach lattice.

We also denote by $\mathbf{1}$ the constant function of constant value 1 on $[0,1]^N$. For a given $i \in \{1,\dots,N\}$, the symbol pr_i stands for the i^{th} coordinate function on $[0,1]^N$, i.e., $pr_i(x) := x_i$ $(x = (x_i)_{1\le i\le N} \in [0,1]^N)$. Moreover, fixed $x \in [0,1]^N$, we denote by Ψ_x the function defined as $\Psi_x(y) = y - x$ for every $y \in [0,1]^N$ (whenever $N = 1$ we use the symbol ψ_x) and by d_x the function defined by

$$d_x(y) := \|y - x\|_2 \quad (y \in [0,1]^N), \tag{2.1}$$

where $\|\cdot\|_2$ stands for the Euclidean norm on $\mathbf{R}^N$, i.e., $\|x\|_2 := \left(\sum\limits_{i=1}^{N} x_i^2\right)^{1/2}$ $(x = (x_i)_{1\le i\le N} \in \mathbf{R}^N)$.

We note that, given $x = (x_i)_{1\le i\le N} \in [0,1]^N$ and $i \in \{1,\dots,N\}$,

$$pr_i \circ \Psi_x = pr_i - x_i\mathbf{1}, \tag{2.2}$$

and hence

$$(pr_i \circ \Psi_x)^2 = pr_i^2 - 2x_i pr_i + x_i^2\mathbf{1}. \tag{2.3}$$

Moreover,

$$d_x^2 = \sum_{i=1}^{N} (pr_i \circ \Psi_x)^2 \tag{2.4}$$

and

$$d_x^4 = \sum_{i=1}^{N} (pr_i \circ \Psi_x)^4 + 2 \sum_{1\le i<j\le N} (pr_i \circ \Psi_x)^2 (pr_j \circ \Psi_x)^2. \tag{2.5}$$

Given $1 \le p < +\infty$, the symbol $L^p([0,1]^N)$ stands for the spaces of all (equivalence classes of) Borel measurable functions f defined on $[0,1]^N$ such that

$$\|f\|_p := \left(\int_{[0,1]^N} |f|^p \, dx\right)^{1/p} < +\infty.$$

In [3] we introduced and studied a new sequence of positive linear operators acting on $L^1([0,1]^N)$, that will be also the object of interest of this paper.

More precisely, let $(a_n)_{n\ge1}$ and $(b_n)_{n\ge1}$ be two sequences of real numbers such that $0 \le a_n < b_n \le 1$ for every $n \ge 1$.

If $n \ge 1$ and $h = (h_i)_{1\le i\le N} \in \{0,\dots,n\}^N$, set

$$Q_{n,h}^{a_n,b_n} := \prod_{i=1}^{N} \left[\frac{h_i + a_n}{n+1}, \frac{h_i + b_n}{n+1}\right]$$

and consider the positive linear operator $C_n : L^1([0,1]^N) \longrightarrow \mathscr{C}([0,1]^N)$ defined by setting, for any $f \in L^1([0,1]^N)$ and $x = (x_i)_{1\le i\le N} \in [0,1]^N$,

$$\begin{aligned} C_n(f)(x) &= \sum_{h\in\{0,\ldots,n\}^N} P_{n,h}(x)\left(\frac{n+1}{b_n - a_n}\right)^N \int_{Q_{n,h}^{a_n,b_n}} f(t)\, dt \\ &= \sum_{\substack{h=(h_i)_{1\le i\le N}\\ h_i\in\{0,\ldots,n\}}} P_{n,h}(x)\left(\frac{n+1}{b_n - a_n}\right)^N \int_{\frac{h_1+a_n}{n+1}}^{\frac{h_1+b_n}{n+1}} \cdots \int_{\frac{h_N+a_n}{n+1}}^{\frac{h_N+b_n}{n+1}} f(t_1,\ldots,t_N)\, dt_1\cdots dt_N, \end{aligned} \tag{2.6}$$

where

$$P_{n,h}(x) := \prod_{i=1}^N p_{n,h_i}(x_i) = \prod_{i=1}^N \binom{n}{h_i} x_i^{h_i}(1-x_i)^{n-h_i} \tag{2.7}$$

for every $x = (x_i)_{1\le i\le N} \in [0,1]^N$ and $h = (h_i)_{1\le i\le N} \in \{0,\ldots,n\}^N$.

Note that C_n is positive and continuous and that, as an operator from $\mathscr{C}([0,1]^N)$ into itself, its norm is $||C_n|| = 1$, since $C_n(\mathbf{1}) = \mathbf{1}$ for any $n \ge 1$.

We point out that the sequence $(C_n)_{n\ge1}$ represents a generalization of Kantorovich operators on $[0,1]^N$, that were introduced and studied by Zhou in [15] and that can be obtained from (2.6) by setting, for any $n \ge 1$, $a_n = 0$ and $b_n = 1$.

On the other hand, the C_n's generalize to the multidimensional case a class of operators first studied in [5, Examples 1.2, 1] and defined by

$$K_n(f)(x) = \sum_{h=0}^n p_{n,h}(x)\frac{n+1}{b_n - a_n}\int_{\frac{h+a_n}{n+1}}^{\frac{h+b_n}{n+1}} f(t)\, dt \tag{2.8}$$

for every $n \ge 1$, $f \in L^1([0,1])$ and $x \in [0,1]$, where, as above, $p_{n,h}(x) := \binom{n}{h}x^h(1-x)^{n-h}$.

A possible interest in the study of the sequence $(C_n)_{n\ge1}$ lies in the fact that it allows to reconstruct a Lebesgue-integrable function by means of its mean values on the sets $Q_{n,h}^{a_n,b_n}$ which are smaller than the corresponding ones considered in [15]. In fact, the following result holds (see [3, Theorems 2.2 and 2.5]).

Proposition 2.1. *For every $f \in \mathscr{C}([0,1]^N)$,*

$$\lim_{n\to\infty} C_n(f) = f \quad \textit{uniformly on } \; [0,1]^N. \tag{2.9}$$

Moreover, for every $n \ge 1$ and $p \in [1,+\infty[$, the operator C_n is continuous from $L^p([0,1]^N)$ into itself and

$$\|C_n\|_{L^p,L^p} \le \frac{1}{(b_n - a_n)^{N/p}}. \tag{2.10}$$

Finally, if $\sup_{n\ge1} 1/(b_n - a_n) < +\infty$, then, for every $f \in L^p([0,1]^N)$,

$$\lim_{n\to\infty} C_n(f) = f \quad \textit{in } \; L^p([0,1]^N). \tag{2.11}$$

In [3, Propositions 2.4, 2.6 and 2.7] estimates of the rate of convergence in the previous approximation formulae are also given.

The main aim of this paper is to show that suitable iterates of the operators C_n converge to a positive C_0-semigroup of operators both in $\mathscr{C}([0,1]^N)$ and in $L^p([0,1]^N)$, $p \geq 1$.

To this end, first of all we recall some properties of the operators K_n defined in (2.8), that will be useful in the sequel (for a proof see [6, Section 2]).

Lemma 2.2. *For every $n \geq 1$, let K_n be the positive linear operator defined by (2.8) and, for every $0 \leq x \leq 1$, consider the functions $\psi_x(y) = y - x$ ($y \in [0,1]$). Then*

(i) $\lim_{n\to\infty} K_n(\psi_x^2)(x) = 0$ *uniformly on* $[0,1]$*;*

(ii) $\lim_{n\to\infty} nK_n(\psi_x^2)(x) = x(1-x)$ *uniformly on* $[0,1]$*;*

(iii) $\lim_{n\to\infty} nK_n(\psi_x^4)(x) = 0$ *uniformly on* $[0,1]$.

As regards the operators C_n, we have the following result (see [3, Lemma 2.1]).

Lemma 2.3. *Given $n \geq 1$ and $i \in \{1, \ldots, N\}$, then*

$$C_n(\mathbf{1}) = \mathbf{1}, \tag{2.12}$$

$$C_n(pr_i) = \frac{n}{n+1} pr_i + \frac{a_n + b_n}{2(n+1)} \mathbf{1} \tag{2.13}$$

and

$$\begin{aligned} C_n(pr_i^2) = \frac{1}{(n+1)^2} \Big\{ & n^2 pr_i^2 + n pr_i(1 - pr_i) + n(a_n + b_n) pr_i \\ & + \frac{1}{3}(a_n^2 + a_n b_n + b_n^2)\mathbf{1} \Big\}. \end{aligned} \tag{2.14}$$

Further, the following equalities will be useful (see [3, Lemma 2.1]).

Proposition 2.4. *For every $x = (x_i)_{1\leq i\leq N} \in [0,1]^N$ and $n \geq 1$,*

$$C_n(pr_i \circ \Psi_x)(x) = -\frac{1}{n+1} x_i + \frac{a_n + b_n}{2(n+1)}, \tag{2.15}$$

$$\begin{aligned} C_n((pr_i \circ \Psi_x)^2)(x) = \frac{1}{(n+1)^2} \Big\{ & x_i^2 + n x_i(1 - x_i) - (a_n + b_n) x_i \\ & + \frac{a_n^2 + a_n b_n + b_n^2}{3} \Big\}, \end{aligned} \tag{2.16}$$

$$\begin{aligned} C_n(d_x^2)(x) = \frac{1}{(n+1)^2} \Bigg\{ & (1-n)\|x\|_2^2 + (n - a_n - b_n) \sum_{i=1}^N x_i \\ & + N \frac{a_n^2 + a_n b_n + b_n^2}{3} \Bigg\} \end{aligned} \tag{2.17}$$

and

$$C_n(d_x^4)(x) = \sum_{i=1}^{N} K_n(\psi_{x_i}^4)(x_i) + 2 \sum_{1\le i<j\le N} K_n(\psi_{x_i}^2)(x_i)K_n(\psi_{x_j}^2)(x_j), \quad (2.18)$$

where, for any $n \geq 1$, *the operator* K_n *is defined by* (2.8) *and, for a given* $i \in \{1, \ldots, N\}$, $\psi_{x_i}(t_i) = t_i - x_i$ $(t = (t_i)_{1\le i\le N} \in [0,1]^N)$.

Proof. Formulae (2.15)-(2.17) are a direct consequence of Lemma 2.3 and formulas (2.2)-(2.4). Taking both definition (2.6) of C_n's and formulae (2.2) and (2.5) into account, we obtain

$$\begin{aligned}
C_n(d_x^4)(x) &= \sum_{h\in\{0,\ldots,n\}^N} P_{n,h}(x)\left(\frac{n+1}{b_n-a_n}\right)^N \int_{Q_{n,h}^{a_n,b_n}} d_x^4(t)\,dt \\
&= \sum_{h\in\{0,\ldots,n\}^N} P_{n,h}(x)\left(\frac{n+1}{b_n-a_n}\right)^N \int_{Q_{n,h}^{a_n,b_n}} \sum_{i=1}^{N}(t_i-x_i)^4(t)\,dt \\
&+ \sum_{h\in\{0,\ldots,n\}^N} P_{n,h}(x)\left(\frac{n+1}{b_n-a_n}\right)^N \int_{Q_{n,h}^{a_n,b_n}} 2\sum_{1\le i<j\le N}(t_i-x_i)^2(t_j-x_j)^2\,dt \\
&= \sum_{\substack{h=(h_i)_{1\le i\le N}\\ h_i\in\{0,\ldots,n\}}} P_{n,h}(x)\left(\frac{n+1}{b_n-a_n}\right)^N \int_{\frac{h_1+a_n}{n+1}}^{\frac{h_1+b_n}{n+1}}\cdots\int_{\frac{h_N+a_n}{n+1}}^{\frac{h_N+b_n}{n+1}} \sum_{i=1}^{N}\psi_{x_i}^4(t_i)\,dt_1\cdots dt_N \\
&+2\sum_{\substack{h=(h_i)_{1\le i\le N}\\ h_i\in\{0,\ldots,n\}}} P_{n,h}(x)\left(\frac{n+1}{b_n-a_n}\right)^N \int_{\frac{h_1+a_n}{n+1}}^{\frac{h_1+b_n}{n+1}}\cdots\int_{\frac{h_N+a_n}{n+1}}^{\frac{h_N+b_n}{n+1}} \sum_{1\le i<j\le N}\psi_{x_i}^2(t_i)\psi_{x_j}^2(t_j)dt_1\cdots dt_N \\
&= \sum_{\substack{h=(h_i)_{1\le i\le N}\\ h_i\in\{0,\ldots,n\}}} P_{n,h}(x)\left(\frac{n+1}{b_n-a_n}\right)\sum_{i=1}^{N}\int_{\frac{h_i+a_n}{n+1}}^{\frac{h_i+b_n}{n+1}} \psi_{x_i}^4(t_i)\,dt_i \\
&+ 2\sum_{\substack{h=(h_i)_{1\le i\le N}\\ h_i\in\{0,\ldots,n\}}} P_{n,h}(x)\left(\frac{n+1}{b_n-a_n}\right)^2 \sum_{1\le i<j\le N}\int_{\frac{h_i+a_n}{n+1}}^{\frac{h_i+b_n}{n+1}}\int_{\frac{h_j+a_n}{n+1}}^{\frac{h_j+b_n}{n+1}} \psi_{x_i}^2(t_i)\psi_{x_j}^2(t_j)\,dt_i\,dt_j.
\end{aligned}$$

Now keeping (2.7) in mind and using the identities

$$\sum_{h_k=0}^{n} p_{n,h_k}(x_k) = 1 \quad \text{for every } k \in \{1,\ldots,N\},$$

we have

$$C_n(d_x^4)(x) = \sum_{i=1}^{N} \sum_{h_i=0}^{n} p_{n,h_i}(x_i) \left(\frac{n+1}{b_n - a_n}\right) \int_{\frac{h_i+a_n}{n+1}}^{\frac{h_i+b_n}{n+1}} \psi_{x_i}^4(t_i)\, dt_i$$

$$+ 2 \sum_{1 \le i < j \le N} \sum_{h_i=0}^{n} p_{n,h_i}(x_i) \left(\frac{n+1}{b_n - a_n}\right) \int_{\frac{h_i+a_n}{n+1}}^{\frac{h_i+b_n}{n+1}} \psi_{x_i}^2(t_i)\, dt_i$$

$$\times \sum_{h_j=0}^{n} p_{n,h_j}(x_j) \left(\frac{n+1}{b_n - a_n}\right) \int_{\frac{h_j+a_n}{n+1}}^{\frac{h_j+b_n}{n+1}} \psi_{x_j}^2(t_j)\, dt_j,$$

and hence formula (2.18) follows. □

Remark 2.5. A more explicit expression of (2.18) can be obtained using some computations contained in the proof of [6, Theorem 2.2].

Another useful result is shown below.

Proposition 2.6. *Under each of the following sets of conditions:*

(a) $a_n = 0$ *and* $b_n = 1$ *for every* $n \ge 1$,

or

(b) (i) $0 < b_n - a_n < 1$ *for every* $n \ge 1$*;*
(ii) *there exist* $\lim_{n\to\infty} a_n = 0$ *and* $\lim_{n\to\infty} b_n = 1$*;*
(iii) $M_1 := \sup_{n\ge 1} n(1 - (b_n - a_n)) < +\infty$,

for every $p \ge 1$ *there exists* $\omega_p \ge 0$ *such that, for every* $k \ge 1$ *and* $n \ge 1$,

$$\|C_n^k\|_{L^p, L^p} \le e^{\frac{k}{n}\omega_p}, \tag{2.19}$$

where C_n^k *denotes the iterate of* C_n *of order* k.

Proof. Fix $p \ge 1$. Under assumption (a), on account of (2.10), the result obviously follows with $\omega_p = 0$.

Assume that conditions (i), (ii) and (iii) of (b) hold true; since

$$\lim_{n\to\infty} \frac{\log(b_n - a_n)}{1 - (b_n - a_n)} = -1,$$

there exists

$$M_2 := \sup_{n \ge 1} \frac{-\log(b_n - a_n)}{1 - (b_n - a_n)} > 0. \tag{2.20}$$

By means of (2.10), we then get

$$\|C_n^k\|_{L^p, L^p} \le \frac{1}{(b_n - a_n)^{kN/p}} = e^{-\frac{kN}{p}\log(b_n - a_n)}$$

$$= e^{\frac{k}{n}\left(-\frac{N}{p} n(1-(b_n-a_n)) \frac{\log(b_n - a_n)}{1-(b_n - a_n)}\right)} \le e^{\frac{k}{n}\omega_p},$$

where $\omega_p := NM_1M_2/p$, and this completes the proof of (2.19). □

We also point out that, as in the one-dimensional case (see [5, formula (4.2)]), the operators C_n are closely related to the Bernstein operators on $[0,1]^N$ that are defined by

$$B_n(f)(x) := \sum_{\substack{h=(h_i)_{1\le i\le N}\\ h_i\in\{0,\dots,N\}}} P_{n,h}(x) f\left(\frac{h_1}{n},\dots,\frac{h_N}{n}\right) \tag{2.21}$$

($f \in \mathscr{C}([0,1]^N)$, $x = (x_i)_{1\le i\le N} \in [0,1]^N, n \ge 1$), $P_{n,h}(x)$ being defined by (2.7).

More precisely, for every $f \in L^1([0,1]^N)$, considering the function

$$\begin{aligned} F_n(f)(x) &:= \left(\frac{n+1}{b_n-a_n}\right)^N \int_{\frac{nx_1+a_n}{n+1}}^{\frac{nx_1+b_n}{n+1}} dt_1 \cdots \int_{\frac{nx_N+a_n}{n+1}}^{\frac{nx_N+b_n}{n+1}} f(t_1,\dots,t_N)\, dt_N \\ &= \int_0^1 dt_1 \cdots \int_0^1 f\left(\frac{(b_n-a_n)t_1+a_n+nx_1}{n+1},\dots,\frac{(b_n-a_n)t_N+a_n+nx_N}{n+1}\right) dt_N \end{aligned} \tag{2.22}$$

($x = (x_i)_{1\le i\le N} \in [0,1]^N), n \ge 1$), it turns out that

$$C_n(f)(x) = B_n(F_n(f))(x) \tag{2.23}$$

($f \in L^1([0,1]^N)$, $x = (x_i)_{1\le i\le N} \in [0,1]^N, n \ge 1$).

Formula (2.23) allows us to easily determine some subsets of $\mathscr{C}([0,1]^N)$ that are invariant under the operators C_n, $n \ge 1$.

Given any $m \in \mathbf{N}$, we shall denote by $\mathbb{P}_m$ the linear subspace of the (restrictions to $[0,1]^N$ of the) polynomials of degree no greater than m.

Finally, given $M \ge 0$ and $0 < \alpha \le 1$, the symbol $Lip^1_M\alpha$ stands for the subset of all functions $f \in \mathscr{C}([0,1]^N)$ such that, for every $x, y \in [0,1]^N$,

$$|f(x) - f(y)| \le M\|x-y\|_1^\alpha,$$

where $\|\cdot\|_1$ denotes the l_1-norm on $\mathbf{R}^N$, i.e., $\|z\|_1 := \sum_{i=1}^N |z_i|$ for every $z = (z_i)_{1\le i\le N} \in \mathbf{R}^N$.

Proposition 2.7. *The subsets $\mathbb{P}_m$, $m \ge 1$, and $Lip^1_M\alpha$ are invariant under the operators C_n, $n \ge 1$, i.e.,*

$$C_n(\mathbb{P}_m) \subset \mathbb{P}_m \tag{2.24}$$

and

$$C_n(Lip^1_M\alpha) \subset Lip^1_M\alpha. \tag{2.25}$$

Proof. Both the subsets $\mathbb{P}_m$ and $Lip^1_M\alpha$ are invariant under the operators B_n, $n \ge 1$ (see, respectively, [1, Section 6.3.12, condition (6.2.18) and the proof of Theorem 6.2.6, p. 441] and [1, Corollary 6.1.22 and Section 6.3.12, p. 476]).

Therefore, on account of (2.23), it suffices to show that $F_n(f) \in \mathbb{P}_m$ (resp., $F_n(f) \in Lip^1_M\alpha$) provided that $f \in \mathbb{P}_m$ or $f \in Lip^1_M\alpha$, respectively, and this can be easily verified by virtue of (2.22). $\square$

3. The C_0-semigroups associated with the operators C_n

In this section we shall prove that suitable iterates of the operators C_n converge on $\mathscr{C}([0,1]^N)$ to a Markov semigroup and on $L^p([0,1]^N)$, $1 \le p < +\infty$, to a positive contractive C_0-semigroup (that is an extension of the previous one).

From now on we assume that there exists

$$l := \lim_{n\to\infty} (a_n + b_n) \in \mathbf{R}. \tag{3.1}$$

Clearly, $0 \le l \le 2$.

Under this assumption we shall prove that the sequence $(C_n)_{n\ge 1}$ satisfies an asymptotic formula with respect to the elliptic second order differential operator $V_l : \mathscr{C}^2([0,1]^N) \longrightarrow \mathscr{C}([0,1]^N)$ defined by setting

$$V_l(u)(x) := \frac{1}{2}\sum_{i=1}^{N} x_i(1-x_i)\frac{\partial^2 u}{\partial x_i^2}(x) + \sum_{i=1}^{N}\left(\frac{l}{2} - x_i\right)\frac{\partial u}{\partial x_i}(x), \tag{3.2}$$

for every $u \in \mathscr{C}^2([0,1]^N)$ and $x = (x_i)_{1\le i\le N} \in [0,1]^N$.

Theorem 3.1. *Under assumption* (3.1), *for every* $u \in \mathscr{C}^2([0,1]^N)$,

$$\lim_{n\to\infty} n(C_n(u) - u) = V_l(u) \tag{3.3}$$

uniformly on $[0,1]^N$ *and hence in* $L^p([0,1]^N)$.

Proof. According to [4, Theorem 3.5], the claim will be proved after showing that, for every $i \in \{1,\dots,N\}$,

(a) $\lim\limits_{n\to\infty} [nC_n(pr_i \circ \Psi_x)(x) - (l/2 - x_i)] = 0$ uniformly on $[0,1]^N$,

(b) $\lim\limits_{n\to\infty} [nC_n((pr_i \circ \Psi_x)^2)(x) - x_i(1-x_i)] = 0$ uniformly on $[0,1]^N$,

(c) $\sup\limits_{\substack{n\ge 1\\ x\in[0,1]^N}} nC_n(d_x^2)(x) < +\infty$

and

(d) $\lim\limits_{n\to\infty} nC_n(d_x^4)(x) = 0$ uniformly on $[0,1]^N$,

where d_x is defined by (2.1).

We proceed to verify (a). According to formula (2.15) we get that, for every $i = 1,\dots,N$,

$$\begin{aligned}&\left|nC_n(pr_i \circ \Psi_x)(x) - \left(\frac{l}{2} - x_i\right)\right| \le \frac{1}{n+1}|x_i| + \left|\frac{n}{n+1}\frac{a_n+b_n}{2} - \frac{l}{2}\right|\\ &\le \frac{1}{n+1} + \left|\frac{n}{n+1}\frac{a_n+b_n}{2} - \frac{l}{2}\right|;\end{aligned}$$

hence the required assertion follows from (3.1).

To prove statement (b) we preliminary notice that, by virtue of formula (2.16), for every $i = 1, \dots, N$,

$$nC_n((pr_i \circ \Psi_x)^2)(x) - x_i(1-x_i)$$
$$= \left[\frac{n^2}{(n+1)^2} - 1\right] x_i(1-x_i) + \frac{n}{(n+1)^2}\left\{x_i^2 - (a_n+b_n)x_i + \frac{a_n^2+a_nb_n+b_n^2}{3}\right\};$$

therefore

$$|nC_n((pr_i \circ \Psi_x)^2)(x) - x_i(1-x_i)|$$
$$\leq \left|\frac{n^2}{(n+1)^2} - 1\right| x_i(1-x_i) + \frac{n}{(n+1)^2}\left(x_i^2 + (a_n+b_n)x_i + \frac{a_n^2+a_nb_n+b_n^2}{3}\right)$$
$$\leq \frac{2n+1}{4}\frac{1}{(n+1)^2} + \frac{4n}{(n+1)^2}$$

and this completes the proof of (b).

As regards conditions (c) and (d), from (2.17) we achieve that, for every $x \in [0,1]^N$,

$$C_n(d_x^2)(x) \leq \frac{N}{n+1},$$

and hence condition (c) follows. Finally, condition (d) is a consequence of (2.18) and Lemma 2.2. $\square$

We recall that a Markov semigroup on $\mathscr{C}([0,1]^N)$ is a C_0-semigroup $(T(t))_{t\geq 0}$ of positive linear operators on $\mathscr{C}([0,1]^N)$ such that $T(t)(\mathbf{1}) = \mathbf{1}$ for every $t \geq 0$ (for more details on the theory of C_0-semigroups of operators we refer, e.g., to [8], [9] and [12]). In particular, we refer to [8, Section 13.6] for some remarkable aspects concerning Markov semigroups (see also [1, Section 1.6]).

We also recall that, given a Banach space $(E, \|\cdot\|)$, a *core* for a linear operator $A : D(A) \longrightarrow E$, defined on a linear subspace $D(A)$ of E, is a linear subspace D_0 of E that is dense in $D(A)$ with respect to the graph norm $\|u\|_A := \|u\| + \|A(u)\|$ $(u \in D(A))$.

If $(A, D(A))$ is the generator of a C_0-semigroup $(T(t))_{t\geq 0}$ of operators on E, then a dense (in E) linear subspace D_0 of $D(A)$ that is invariant under $(T(t))_{t\geq 0}$, i.e., $T(t)(D_0) \subset D_0$ for every $t \geq 0$, is a core for $(A, D(A))$ (see, e.g., [9, Chapter II, Proposition 1.7]). Moreover, if D_0 is a core for $(A, D(A))$, then $(A, D(A))$ is the closure of (A, D_0) as well.

As in Section 2, given any $m \in \mathbf{N}$, we denote by $\mathbb{P}_m$ the linear subspace of the (restrictions to $[0,1]^N$ of the) polynomials on $\mathbf{R}^N$ of degree no greater than m. Thus $\mathbb{P} := \bigcup_{m=0}^{+\infty} \mathbb{P}_m$ is the subalgebra of all the (restrictions to $[0,1]^N$ of the) polynomials on $\mathbf{R}^N$ and it is dense in $\mathscr{C}([0,1]^N)$ by the Stone-Weierstrass theorem.

Fix $0 \leq l \leq 2$ and consider the differential operator $V_l : \mathscr{C}^2([0,1]^N) \longrightarrow \mathscr{C}([0,1]^N)$ defined by (3.2). This operator falls in the class of Fleming-Viot operators arising in population genetics, that are usually studied in the setting of the multidimensional simplex. However, in the framework of hypercubes they have been investigated in [2], [7], [10].

Theorem 3.2. *There exists a Markov semigroup $(T_l(t))_{t\geq 0}$ on $\mathscr{C}([0,1]^N)$ satisfying the following properties:*

(1) *If $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ are two sequences of real numbers satisfying $0 \leq a_n < b_n \leq 1$ for every $n \geq 1$ and $\lim\limits_{n\to\infty}(a_n+b_n) = l$, then for every $t \geq 0$ and for every sequence $(\rho_n)_{n\geq 1}$ of positive integers such that $\lim\limits_{n\to\infty} \rho_n/n = t$*

$$\lim_{n\to\infty} C_n^{\rho_n}(f) = T_l(t)(f) \qquad \textit{uniformly on } [0,1]^N \tag{3.4}$$

for every $f \in \mathscr{C}([0,1]^N)$, where each $C_n^{\rho_n}$ denotes the iterate of C_n of order ρ_n. In particular,

$$\lim_{n\to\infty} C_n^{[nt]}(f) = T_l(t)(f) \qquad \textit{uniformly on } [0,1]^N \tag{3.5}$$

for every $f \in \mathscr{C}([0,1]^N)$, where $[nt]$ stands for the integer part of nt.

(2) *Denoted by $(A_l, D(A_l))$ the generator of the semigroup $(T_l(t))_{t\geq 0}$, then $\mathscr{C}^2([0,1]^N)$ is a core for $(A_l, D(A_l))$, so that $(A_l, D(A_l))$ is the closure of $(V_l, \mathscr{C}^2([0,1]^N))$.*

(3) *The subalgebra $\mathbb{P}$ is a core for $(A_l, D(A_l))$ and $T_l(t)(\mathbb{P}_m) \subset \mathbb{P}_m$ for every $t \geq 0$ and $m \geq 0$.*

(4) *$T_l(t)(Lip_M^1\alpha) \subset Lip_M^1\alpha$ for every $t \geq 0$, $M \geq 0$ and $0 < \alpha \leq 1$.*

Proof. The proof is similar in spirit to the one of Theorem 4.1 of [2]. Consider two sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ of real numbers satisfying $0 \leq a_n < b_n \leq 1$ for every $n \geq 1$ and $\lim\limits_{n\to\infty}(a_n+b_n) = l$, and denote by $(C_n)_{n\geq 1}$ the relevant operators defined by (2.6).

Moreover, consider the linear operator $B : D(B) \longrightarrow \mathscr{C}([0,1]^N)$ defined by

$$B(u) := \lim_{n\to\infty} n(C_n(u) - u) \qquad (u \in D(B)),$$

where

$$D(B) := \left\{ u \in \mathscr{C}([0,1]^N) \;\middle|\; \text{there exists } \lim_{n\to\infty} n(C_n(u)-u) \text{ in } \mathscr{C}([0,1]^N) \right\}.$$

By Theorem 3.1, $\mathscr{C}^2([0,1]^N) \subset D(B)$ and $B = V_l$ on $\mathscr{C}^2([0,1]^N)$. In particular, each $\mathbb{P}_m$ is contained in $D(B)$, it is finite dimensional and invariant under the operators C_n by virtue of Proposition 2.7. By a result of Schnabl ([14]; see also [13] or [1, Theorem 1.6.8]) the operator $(B, D(B))$ is then closable in $\mathscr{C}([0,1]^N)$ and its closure, that we denote by $(A_l, D(A_l))$, is the generator of a positive C_0-semigroup $(T_l(t))_{t\geq 0}$ of linear contractions of $\mathscr{C}([0,1]^N)$, satisfying (3.4) and (3.5).

Since $C_n(\mathbf{1}) = \mathbf{1}$ for any $n \geq 1$, from (3.5) it follows that $T_l(t)(\mathbf{1}) = \mathbf{1}$ for every $t \geq 0$. Moreover, each $\mathbb{P}_m$ is closed in $\mathscr{C}([0,1]^N)$ and it is invariant under the C_n's. Therefore, iterating and passing to the limit, we obtain that $T_l(t)(\mathbb{P}_m) \subset \mathbb{P}_m$ for every $t \geq 0$.

Accordingly, we get that $T_l(t)(\mathbb{P}) \subset \mathbb{P}$ for any $t \geq 0$ and hence $\mathbb{P}$ is a core for $(A_l, D(A_l))$. In particular, $\mathscr{C}^2([0,1]^N)$ is a core for $(A_l, D(A_l))$ as well and $A_l = B = V_l$ on $\mathscr{C}^2([0,1]^N)$, which implies that $(A_l, D(A_l))$ is the closure of $(V_l, \mathscr{C}^2([0,1]^N))$, too.

This last statement shows, indeed, that the generator $(A_l, D(A_l))$ is independent on the sequence $(C_n)_{n\geq 1}$ and hence on the sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$. On the other hand, the generator $(A_l, D(A_l))$ determines the generated semigroup uniquely (see [9, Chapter II, Theorem 1.4]) and so the semigroup $(T_l(t))_{t\geq 0}$ does not depend on the particular sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$, as well.

Finally, statement (4) follows from formula (2.25) of Proposition 2.7 and from the fact that $Lip_M^1\alpha$ is closed under the pointwise (and hence under the uniform) convergence on $[0,1]^N$. □

Remarks 3.3.

1. Let us now consider the abstract Cauchy problem associated with $(A_l, D(A_l))$, i.e.,

$$\begin{cases} \dfrac{\partial u}{\partial t}(x,t) = A_l(u(\cdot,t))(x) & x \in [0,1]^N, \quad t \geq 0, \\ \\ u(x,0) = u_0(x) & u_0 \in D(A_l), \quad x \in [0,1]^N. \end{cases}$$

Since $(A_l, D(A_l))$ generates a C_0-semigroup, the above Cauchy problem admits a unique solution $u : [0,1]^N \times [0,+\infty[\to \mathbf{R}$ given by $u(x,t) = T_l(t)(u_0)(x)$ for every $x \in [0,1]^N$ and $t \geq 0$ (see, e.g., [12, Chapter A-II]). Hence, by Theorem 3.2, it is possible to approximate such solutions by means of iterates of the C_n's, i.e.,

$$u(x,t) = T_l(t)(u_0)(x) = \lim_{n\to\infty} C_n^{[nt]}(u_0)(x),$$

the limit being uniform with respect to $x \in [0,1]^N$.

Moreover, since A_l coincides with the elliptic second-order differential operator V_l defined by (3.2) on $\mathbb{P}_m$, $m \geq 1$, if $u_0 \in \mathbb{P}_m$, then $u(x,t)$ is the unique solution to the Cauchy problem

$$\begin{cases} \dfrac{\partial u}{\partial t}(x,t) = \dfrac{1}{2}\sum\limits_{i=1}^{N} x_i(1-x_i)\dfrac{\partial^2 u(x,t)}{\partial x_i^2} + \sum\limits_{i=1}^{N}\left(\dfrac{l}{2} - x_i\right)\dfrac{\partial u(x,t)}{\partial x_i} & x \in [0,1]^N, \\ & t \geq 0, \\ \\ u(x,0) = u_0(x) & x \in [0,1]^N \end{cases}$$

and $u(\cdot,t) \in \mathbb{P}_m$ for every $t \geq 0$ (see statement (3) of Theorem 3.2).

Finally, according to statement (4) of Theorem 3.2, if $u_0 \in D(A_l) \cap Lip_M^1\alpha$ $(M \geq 0, 0 < \alpha \leq 1)$, then $u(\cdot,t) \in Lip_M^1\alpha$ for every $t \geq 0$.

2. Theorem 3.2 extends Theorem 3.3 of [6] from the one-dimensional case to a multidimensional context. However, there an explicit description of

the generator $(A_l, D(A_l))$ is given, namely

$$D(A_l) := \left\{ u \in \mathscr{C}([0,1]) \;\; | \; u \in \mathscr{C}^2(]0,1[) \text{ and } \lim_{\substack{x \to 0^+ \\ x \to 1^-}} A_l(u)(x) \in \mathbf{R} \right\} \tag{3.6}$$

and

$$A_l(u)(x) := \begin{cases} \dfrac{x(1-x)}{2} u''(x) + \left(\dfrac{l}{2} - x \right) u'(x) & \text{if } 0 < x < 1, \\ \\ \lim\limits_{t \to x} A_l(u)(t) & \text{if } x = 0, 1 \end{cases} \tag{3.7}$$

$(u \in D(A_l), 0 \le x \le 1)$.

An analogous description of $(A_l, D(A_l))$ in multidimensional setting seems to be a difficult but very interesting problem.

3. Statement (2) of Theorem 3.2 has been also obtained in [7, Theorem 2.1] with a different approach.

Next, we shall show that, in some particular cases, the Markov semigroup considered in Theorem 3.2 extends to a positive contractive C_0-semigroup on $L^p([0,1]^N)$, $1 \le p < +\infty$.

In fact, in these cases the limit (3.1) is $l = 1$, that leads to consider the differential operator

$$\begin{aligned} V(u)(x) := V_1(u)(x) &= \frac{1}{2} \sum_{i=1}^N x_i(1-x_i) \frac{\partial^2 u}{\partial x_i^2}(x) + \sum_{i=1}^N \left(\frac{1}{2} - x_i \right) \frac{\partial u}{\partial x_i}(x) \\ &= \sum_{i=1}^N \frac{\partial}{\partial x_i} \left(\frac{x_i(1-x_i)}{2} \frac{\partial u}{\partial x_i} \right)(x) \end{aligned} \tag{3.8}$$

$(u \in \mathscr{C}^2([0,1]^N)$ and $x = (x_i)_{1 \le i \le N} \in [0,1]^N)$.

Similarly, we shall simply denote by $(T(t))_{t \ge 0}$ and by $(A, D(A))$ the semigroup $(T_1(t))_{t \ge 0}$ and its generator $(A_1, D(A_1))$.

Theorem 3.4. *The Markov semigroup $(T(t))_{t \ge 0}$ extends to a positive contractive C_0-semigroup $(\widetilde{T}(t))_{t \ge 0}$ on $L^p([0,1]^N)$ for each $p \in [1, +\infty[$.*

Moreover, $\mathscr{C}^2([0,1]^N)$ is a core for the generator $(\widetilde{A}, D(\widetilde{A}))$ of $(\widetilde{T}(t))_{t \ge 0}$, so that $(\widetilde{A}, D(\widetilde{A}))$ is the closure of $(V, \mathscr{C}^2([0,1]^N))$ in $L^p([0,1]^N)$.

Finally, if $(a_n)_{n \ge 1}$ and $(b_n)_{n \ge 1}$ are two sequences of real numbers such that $0 \le a_n < b_n \le 1$ and if, in addition, they satisfy one of the following sets of conditions:

(a) $a_n = 0$ *and* $b_n = 1$ *for every* $n \ge 1$,

or

(b) (i) $0 < b_n - a_n < 1$ *for every* $n \ge 1$;
(ii) *there exist* $\lim\limits_{n \to \infty} a_n = 0$ *and* $\lim\limits_{n \to \infty} b_n = 1$;
(iii) $M_1 := \sup\limits_{n \ge 1} n(1 - (b_n - a_n)) < +\infty$,

then for every $t \geq 0$, for every sequence $(\rho_n)_{n\geq 1}$ of positive integers such that $\lim_{n\to\infty} \rho_n/n = t$ and for every $f \in L^p([0,1]^N)$,

$$\lim_{n\to\infty} C_n^{\rho_n}(f) = \widetilde{T}(t)(f) \qquad \text{in } L^p([0,1]^N). \tag{3.9}$$

In particular, for every $f \in L^p([0,1]^N)$,

$$\lim_{n\to\infty} C_n^{[nt]}(f) = \widetilde{T}(t)(f) \qquad \text{in } L^p([0,1]^N). \tag{3.10}$$

Here, again, the operators C_n, $n \geq 1$, are defined by (2.6).

Proof. Fix $t \geq 0$ and consider an arbitrary sequence $(\rho_n)_{n\geq 1}$ of positive integers such that $\rho_n/n \to t$. Furthermore, consider the sequence $(C_n)_{n\geq 1}$ associated with $a_n = 0$ and $b_n = 1$, $n \geq 1$. From (2.10) it follows that $\|C_n\|_{L^p,L^p} \leq 1$ and hence, on account of (3.4)

$$\|T(t)f\|_p = \lim_{n\to\infty} \|C_n^{\rho_n}(f)\|_p \leq \|f\|_p$$

for every $f \in \mathscr{C}([0,1]^N)$.

Therefore, there exists a unique linear continuous extension $\widetilde{T}(t) : L^p([0,1]^N) \longrightarrow L^p([0,1]^N)$ of $T(t)$. Moreover, $\|\widetilde{T}(t)\|_{L^p,L^p} \leq 1$ for every $t \geq 0$.

It is not difficult to show that $\widetilde{T}(t)$ is positive because if $f \in L^p([0,1]^N)$, $f \geq 0$, then there exists a sequence $(f_n)_{n\geq 1}$ in $\mathscr{C}([0,1]^N)$ such that $\lim_{n\to\infty} f_n = f$ in $L^p([0,1]^N)$. We may assume that $f_n \geq 0$ for every $n \geq 1$ (if not, we replace f_n with its positive part f_n^+). Therefore,

$$\widetilde{T}(t)(f) = \lim_{n\to\infty} \widetilde{T}(t)(f_n) = \lim_{n\to\infty} T(t)(f_n) \geq 0.$$

The family $(\widetilde{T}(t))_{t\geq 0}$ is obviously a semigroup and, in addition, it is strongly continuous; this easily follows, for instance, from ([9, Chapter I, Proposition 5.3]) thanks to the fact that, for every $t \in [0,1]$ and for every $f \in \mathscr{C}([0,1]^N)$,

$$\lim_{t\to 0^+} \widetilde{T}(t)(f) = \lim_{t\to 0^+} T(t)(f) = f$$

in $\mathscr{C}([0,1]^N)$ and hence in $L^p([0,1]^N)$, because $(T(t))_{t\geq 0}$ is a C_0-semigroup on $\mathscr{C}([0,1]^N)$.

Let $(\widetilde{A}, D(\widetilde{A}))$ be the generator of $(\widetilde{T}(t))_{t\geq 0}$. Then, from the definition of domain of generators, it follows that $D(A) \subset D(\widetilde{A})$ and $\widetilde{A} = A$ on $D(A)$. Moreover, $D(A)$ is a core for $(\widetilde{A}, D(\widetilde{A}))$, since $\widetilde{T}(t)(D(A)) = T(t)(D(A)) \subset D(A)$ for every $t \geq 0$.

In order to show that $\mathscr{C}^2([0,1]^N)$ is a core for $(\widetilde{A}, D(\widetilde{A}))$, fix $u \in D(\widetilde{A})$ and $\varepsilon > 0$; then there exists $v \in D(A)$ such that

$$\|u - v\|_p \leq \frac{\varepsilon}{2} \quad \text{and} \quad \|\widetilde{A}(u) - A(v)\|_p \leq \frac{\varepsilon}{2}. \tag{3.11}$$

On the other hand, by Theorem 3.2, $\mathscr{C}^2([0,1]^N)$ is a core for $(A, D(A))$ and hence there exists $w \in \mathscr{C}^2([0,1]^N)$ such that

$$\|v-w\|_\infty \leq \frac{\varepsilon}{2} \quad \text{and} \quad \|A(v)-A(w)\|_\infty \leq \frac{\varepsilon}{2}. \tag{3.12}$$

From (3.11) and (3.12) it follows that

$$\|u-w\|_p \leq \|u-v\|_p + \|v-w\|_p \leq \|u-v\|_p + \|v-w\|_\infty \leq \varepsilon$$

and, analogously,

$$\|\widetilde{A}(u) - A(w)\|_p \leq \varepsilon.$$

In order to prove (3.9), fix $t \geq 0$ and consider a sequence $(\rho_n)_{n\geq 1}$ of positive integers such that $\lim\limits_{n\to\infty} \rho_n/n = t$; formula (3.4) implies that, for every $f \in \mathscr{C}([0,1]^N)$,

$$\lim_{n\to\infty} C_n^{\rho_n}(f) = \widetilde{T}(t)(f)$$

in $L^p([0,1]^N)$. Since $\|C_n^{\rho_n}\|_{L^p,L^p} \leq 1$ for every $n \geq 1$, then (3.9) and (3.10) follow.

Finally, consider two sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ satisfying assumption (b) and denote by $(C_n)_{n\geq 1}$ the relevant operators. Given $t \geq 0$ and a sequence $(\rho_n)_{n\geq 1}$ of positive integers such that $\rho_n/n \to t$, from (3.4) it follows that

$$\widetilde{T}(t)(f) = \lim_{n\to\infty} C_n^{\rho_n}(f) \qquad \text{in } L^p([0,1]^N)$$

for every $f \in \mathscr{C}([0,1]^N)$. Moreover, (2.19) implies that

$$\|C_n^{\rho_n}\|_{L^p,L^p} \leq \exp\left(\omega_p \frac{\rho_n}{n}\right) \leq \exp(\rho\, \omega_p),$$

where $\rho := \sup\limits_{n\geq 1} \rho_n/n$ and $\omega_p = NM_1M_2/p$, M_2 being defined by formula (2.20) in the proof of Proposition 2.6. Consequently, $(C_n^{\rho_n})_{n\geq 1}$ is equibounded in $L^p([0,1]^N)$ and hence the above limit relationship extends from $\mathscr{C}([0,1]^N)$ to $L^p([0,1]^N)$. $\square$

Remarks 3.5.

1. Examples of sequences satisfying assumptions (b) in Theorem 3.4 can be easily furnished. For instance, fix $\alpha \geq 1$ and, for every $n \geq 1$, set $a_n := \dfrac{1}{2}\left(1 + \dfrac{1}{2n^\alpha} - \dfrac{n^\alpha}{n^\alpha+1}\right)$ and $b_n := \dfrac{1}{2}\left(1 + \dfrac{1}{2n^\alpha} + \dfrac{n^\alpha}{n^\alpha+1}\right)$.
2. Theorem 3.4 seems to be new even in the one-dimensional case where, according to Remark 3.3, 2, the generator $(A, D(A))$ is described by (3.6) and (3.7). However, for $N = 1$ and for $a_n = 0$ and $b_n = 1$, $n \geq 1$, a similar result has been already proved in [11, Theorem 1] with a completely different method. Moreover, in the same paper a representation of the semigroup in terms of the Legendre polynomials is also given.
3. The differential operator $(V_l, \mathscr{C}^2([0,1]^N))$ falls within a more general class of second order differential operators that have been investigated in [2] (see, in particular, Section 4, formula (4.1) and Examples 2.2, 2). From Theorem 4.1 of that paper it already follows that $(V_l, \mathscr{C}^2([0,1]^N))$ is closable

and its closure is the generator of a Markov semigroup on $\mathscr{C}([0,1]^N)$ that can be approximated, as in (3.4), by iterates of modified Bernstein-Schnabl operators. However, in general, these approximating operators are not defined on $L^p([0,1]^N)$, so that formulae (3.9) and (3.10) cannot be available for them.

4. The generation property of the operator $(V, \mathscr{C}^2([0,1]^N))$ in the space $L^p([0,1]^N)$ has been also investigated in [10, Theorem 2.5]. Moreover, in this paper it is shown that the semigroup $(\widetilde{T}(t))_{t\geq 0}$ is analytic and a description of the domain $D(\widetilde{A})$ in terms of weighted Sobolev spaces is given.

References

[1] Altomare, F., Campiti, M., *Korovkin-type approximation theory and its applications*, de Gruyter Studies in Mathematics, **17**, Walter de Gruyter & Co., Berlin, 1994.

[2] Altomare, F., Cappelletti Montano, M., Diomede, S., *Degenerate elliptic operators, Feller semigroups and modified Bernstein-Schnabl operators*, Math. Nachr., 2011, DOI:10.1002/mana.200810196.

[3] Altomare, F., Cappelletti Montano, M., Leonessa, V., *On a generalization of Kantorovich operators on simplices and hypercubes*, Adv. Pure Appl. Math., **1**(2010), no. 3, 359-385.

[4] Altomare, F., Diomede, S., *Asymptotic formulae for positive linear operators on convex subsets of Banach spaces*, J. Concr. Appl. Math., **3**(2005), no. 4, 413-435.

[5] Altomare, F., Leonessa, V., *On a sequence of positive linear operators associated with a continuous selection of Borel measures*, Mediterr. J. Math., **3**(2006), no. 3-4, 363-382.

[6] Altomare, F., Leonessa, V., *Continuous selections of Borel measures, positive operators and degenerate evolution problems*, Rev. Anal. Numér. Théor. Approx., **36**(2007), no. 1, 9-23.

[7] Cerrai, S., Clément, Ph., *Schauder estimates for a degenerate second order elliptic operator on a cube*, J. Differential Equations, **242**(2007), 287-321.

[8] Davies, E.B., *Linear operators and their spectra*, Cambridge Studies in Advanced Mathematics **106**, Cambridge University Press, Cambridge, 2007.

[9] Engel, K.J., Nagel, R., *One-parameter semigroups for linear evolution equations*, Graduate Texts in Mathematics, **194**, Springer, New York-Berlin, 2000.

[10] Mugnolo, D., Rhandi, A., *On the domain of a Fleming-Viot type operator on an L^p-space with invariant measure*, preprint n. 63, 2009, Dipartimento di Ingegneria dell'Informazione e Matematica Applicata, Università degli Studi di Salerno.

[11] Nagel, J., *Asymptotic properties of powers of Kantorovič operators*, J. Approx. Theory, **36**(1982), no. 3, 268-275.

[12] Nagel, R., (ed.), *One-parameter semigroups of positive operators*, Lecture Notes in Math., **1184**, Springer-Verlag, Berlin, 1986.

[13] Nishishiraho, T., *Saturation of bounded linear operators*, Tôhoku Math. J., **30**(1978), 69-81.

[14] Schnabl, R., *Über gleichmäßige Approximation durch positive lineare Operatoren*, in Constructive Theory of Functions (Proc. Internat. Conf. Varna, 1970) 287-296; Izdat. Bolgar. Akad. Nauk, Sofia, 1972.

[15] Zhou, D.X., *Converse theorems for multidimensional Kantorovich operators*, Anal. Math., **19**(1993), 85-100.

Francesco Altomare
Dipartimento di Matematica
Università degli Studi di Bari "A. Moro"
Campus Universitario, Via E. Orabona, 4
70125 Bari-Italia
e-mail: `altomare@dm.uniba.it`

Mirella Cappelletti Montano
Dipartimento di Matematica
Università degli Studi di Bari "A. Moro"
Campus Universitario, Via E. Orabona, 4
70125 Bari-Italia
e-mail: `montano@dm.uniba.it`

Vita Leonessa
Dipartimento di Matematica e Informatica
Università degli Studi della Basilicata
Viale dell'Ateneo Lucano no. 10, Macchia Romana
85100 Potenza-Italia
e-mail: `vita.leonessa@unibas.it`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 237–245

Blending surfaces on ellipse generated using the Bernstein operators

Marius Birou

Abstract. In this paper we present some blending surfaces using univariate Bernstein operators. The surfaces stay on a ellipse which is the border of the domain and they have a fixed height in the point $(0,0)$. Some results about monotonicity, concavity and type of surfaces are given.

Mathematics Subject Classification (2010): 65D17, 41A36, 41A63.

Keywords: Bernstein polynomial, surfaces, monotonicity, concavity.

1. Introduction

The blending surfaces were introduced by Coons in [4] and they have the property of fitting some given curves. In papers [2],[3] and references therein, some blending surfaces on rectangular and triangular domain were constructed. These surfaces can be used in civil engineering as roof surfaces for building. In [1], using the univariate Bernstein operators we constructed blending surfaces which stay on the border of a domain bounded by a simple and closed curve and having fixed weight in the point $(0,0)$ from the domain. In this paper we present some examples in the case in which the curve is a ellipse. Some results about monotonicity and concavity of the surfaces are given. We study the type of the obtained surfaces.

2. Preliminaries

The univariate Bernstein polynomial of a function $f:[0,1]\to\mathbb{R}$ is given by

$$(B_nf)(t)=\sum_{j=0}^{n}b_{jn}(t)f\left(\frac{j}{n}\right), \tag{2.1}$$

where the functions b_{jn} are given by formula

$$b_{jn}(t) = \binom{n}{j} t^j (1-t)^{n-j}$$

for $j = 0, ..., n$. It has the interpolation properties

$$(B_n f)(0) = f(0),\ (B_n f)(1) = f(1)$$

More about the Bernstein polynomials can be found in the book [9].

Next we give some definitions and remarks about the monotonicity and convexity of the bivariate functions (see [7], [8], [5]).

Definition 2.1. *The bivariate function $G : A \to \mathbb{R}$, $A \subseteq \mathbb{R}^2$ is increasing (decreasing) in the direction $d = (d_1, d_2) \in \mathbb{R}^2$ if and only if*

$$G(x + \lambda d_1, y + \lambda d_2) \geq (\leq) G(x, y),$$

for every $(x, y) \in A$ and $\lambda > 0$ such that $(x + \lambda d_1, y + \lambda d_2) \in D$.

Remark 2.2. If G is a C^1 function on the set A we have that the function G is increasing (decreasing) in the direction $d = (d_1, d_2)$ if

$$D_d G \geq (\geq 0)$$

on A, where $D_d G$ is the first order directional derivative in the direction $d = (d_1, d_2)$ of the function G, i.e.

$$D_d G = d_1 G_x + d_2 G_y.$$

Definition 2.3. *The bivariate function $G : A \to \mathbb{R}$, $A \subseteq \mathbb{R}^2$ is convex (concave) on the convex set A if and only if*

$$G(\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \leq (\geq) \lambda G(x_1, y_1) + (1-\lambda) G(x_2, y_2)$$

for every $(x_1, y_1), (x_2, y_2) \in A$ and every $\lambda \in [0, 1]$.

Remark 2.4. If G is a C^2 function on the convex set A we have that the function G is convex (concave) if

$$D_d^2 G \geq (\leq) 0 \tag{2.2}$$

on A for every $(d_1, d_2) \in \mathbb{R}^2$, where $D_d^2 G$ is the second order directional derivative in the direction $d = (d_1, d_2)$ of the function G

$$D_d^2 G = d_1^2 G_{xx} + 2 d_1 d_2 G_{xy} + d_2^2 G_{yy}.$$

The conditions (2.2) hold if and only if

$$G_{xx} \geq (\leq) 0,\ G_{yy} \geq (\leq) 0,\ G_{xx} G_{yy} - G_{xy}^2 \geq 0.$$

Definition 2.5. *Let G a C^2 function on the set $A \subseteq R^2$. The point $(x, y) \in A$ of the surface $z = G(x, y)$ is parabolic point if*

$$PG(x, y) = 0,$$

where

$$PG(x, y) = G_{xx}(x, y) G_{yy}(x, y) - (G_{xy}(x, y))^2. \tag{2.3}$$

If we have $PG(x,y) < 0$ (> 0) then the point (x,y) is called hyperbolic point (elliptic point). The surface G is called of parabolic (hyperbolic, elliptic) type if all the points of the surface are parabolic (hyperbolic, elliptic).

We note

$$\Delta_1 h_j = h_{j+1} - h_j,\ j = 0, ..., n-1,$$
$$\Delta_2 h_j = h_{j+2} - 2h_{j+1} + h_j,\ j = 0, ..., n-2.$$

3. The first family of surfaces

Let $n \in \mathbb{N}$, $n \geq 2$ and $h_i, h \in \mathbb{R}$, $i = 1, ..., n-1$ such that

$$0 = h_n < ... < h_1 < h_0 = h \tag{3.1}$$

and let $f : [0,1] \to \mathbb{R}$ be a function with the properties

$$\begin{array}{l} f(0) = h, \\ f(\frac{j}{n}) = h_j,\ j = 1, ..., n-1, \\ f(1) = 0. \end{array} \tag{3.2}$$

From (2.1) and (3.2), we obtain the Bernstein function

$$(B_n f)(t) = b_{0n}(t)h + \sum_{j=1}^{n-1} b_{jn}(t)h_j. \tag{3.3}$$

The function in (3.3) has the properties

$$(B_n f)(0) = h,\ (B_n f)(1) = 0.$$

Let

$$D = \{(x,y) \in \mathbb{R}^2 : 0 \leq \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1\}.$$

Let u a bivariate positive function such that the curve $C : u(x,y) = 1$ is the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. We assume that the curve $u(x,y) = 0$ is reduced to the point $(0,0)$.

If we make the substitution

$$y = u(x,y)$$

in (3.3), we obtain the bivariate function

$$F(x,y) = (B_n f)(u(x,y)) = \tag{3.4}$$

$$= b_{0n}(u(x,y))h + \sum_{j=1}^{n-1} b_{jn}(u(x,y))h_j,\ (x,y) \in D.$$

The function F from (3.4) has the properties

$$F|_{\partial D} = 0,$$
$$F(0,0) = h.$$

It follows that the surfaces $z = F(x,y)$ match the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, $z = 0$ (the surfaces stay on the border of domain D), and the height of the surfaces in the point $(0,0)$ of domain is h.

From [1] we have the following results about the monotonicity and concavity of the function F.

Theorem 3.1. *If the function u is increasing (decreasing) in the direction (d_1, d_2) then the function F is decreasing (increasing) in the same direction.*

Theorem 3.2. *If $\Delta_2 h_j \leq 0$, $j = 0, ..., n-2$ and the function u is convex then the function F is concave.*

The following theorem gives the expression of the function PF from (2.3) corresponding the surfaces F.

Theorem 3.3. *We have*

$$PF(x, y) = A(x, y)B(x, y)P(x, y) + (A(x, y))^2 Q(x, y)$$

with

$$A(x, y) = n \sum_{i=0}^{n-1} b_{i,n-1}(u(x, y))\Delta_1 h_j,$$

$$B(x, y) = n(n-1) \sum_{i=0}^{n-2} b_{i,n-2}(u(x, y))\Delta_2 h_j,$$

$$P(x, y) = u_x^2(x, y)u_{yy}(x, y) + u_y^2(x, y)u_{xx}(x, y) - 2u_x(x, y)u_y(x, y)u_{xy}(x, y),$$

$$Q(x, y) = u_{xx}(x, y)u_{yy}(x, y) - u_{xy}^2(x, y).$$

Proof. Using some relations from [6], the second order partial derivative of the function F are given by

$$F_{xx}(x, y) = A(x, y)u_x^2(x, y) + B(x, y)u_{xx}(x, y),$$

$$F_{xy}(x, y) = A(x, y)u_x(x, y)u_y(x, y) + B(x, y)u_{xy}(x, y),$$

$$F_{yy}(x, y) = A(x, y)u_y^2(x, y) + B(x, y)u_{yy}(x, y).$$

Taking into account (2.3) we get the expression of $PF(x, y)$. □

We take the function u in the form

$$u(x, y) = \varphi\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right) \tag{3.5}$$

with $\varphi \in C^2(0, 1)$ and $\varphi(0) = 0$, $\varphi(1) = 1$.

Let

$$D_1 = \{(x, y) \in D : x \geq 0, y \geq 0\},$$
$$D_2 = \{(x, y) \in D : x \leq 0, y \geq 0\},$$
$$D_3 = \{(x, y) \in D : x \leq 0, y \leq 0\},$$
$$D_4 = \{(x, y) \in D : x \geq 0, y \leq 0\}.$$

Theorem 3.4. *We have*

i) If $\varphi' \geq 0$ on $(0, 1)$ and $(d_1, d_2) \geq 0 (\leq 0)$ then the function F is decreasing (increasing) on D_1.

ii) If $\varphi' \geq 0$ on $(0, 1)$ and $(-d_1, d_2) \geq 0 (\leq 0)$ then the function F is decreasing (increasing) on D_2.

iii) *If* $\varphi' \geq 0$ *on* $(0,1)$ *and* $(d_1, d_2) \leq 0 (\geq 0)$ *then the function* F *is decreasing (increasing) on* D_3.

iv) *If* $\varphi' \geq 0$ *on* $(0,1)$ *and* $(d_1, -d_2) \geq 0 (\leq 0)$ *then the function* F *is decreasing (increasing) on* D_4.

Proof. The first order directional derivative of the function u is

$$D_d u = 2\left(\frac{xd_1}{a^2} + \frac{yd_2}{b^2}\right)\varphi'\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)$$

Using Theorem 3.1 and Remark 2.2 it follows the conclusion. □

Theorem 3.5. *If* $\Delta_2 h_j \leq 0,\ j = 0, ..., n-2$ *and* $\varphi', \varphi'' \geq 0$ *on* $(0,1)$ *the surfaces* F *are concave and of elliptic type.*

Proof. The second order partial derivatives of the function u are given by

$$u_{xx}(x,y) = \frac{2}{a^2}\varphi'\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right) + \frac{4x^2}{a^4}\varphi''\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right),$$

$$u_{yy}(x,y) = \frac{2}{b^2}\varphi'\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right) + \frac{4y^2}{b^4}\varphi''\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right),$$

$$u_{xy}(x,y) = \frac{4xy}{a^2b^2}\varphi''\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right).$$

It follows that

$$u_{xx}(x,y) \geq 0,\ u_{yy}(x,y) \geq 0$$

and

$$P(x,y) = \frac{8}{a^2b^2}\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)\varphi'^3\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right),$$

$$Q(x,y) = \frac{4}{a^2b^2}\varphi'\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right) \times$$

$$\times \left(\varphi'\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right) + 2\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)\varphi''\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)\right) \geq 0.$$

Using Theorems 3.2, Theorem 3.3 and Remark 2.4 it follows the conclusion. □

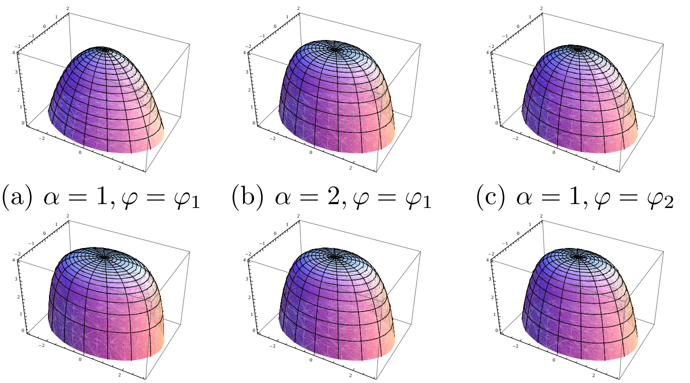
We give some examples of functions which satisfy the conditions of Theorem 3.4 and Theorem 3.5:

$$\varphi_1(t) = t^\alpha,\ \alpha \in \{1\} \cup [2, \infty),$$

$$\varphi_2(t) = \frac{e^{\alpha t} - 1}{e^\alpha - 1},\ \alpha > 0,$$

$$\varphi_3(t) = \frac{cos\alpha t - 1}{\cos\alpha - 1}, \alpha \in \left[0, \frac{\pi}{2}\right).$$

In Figure 1 we plot the surfaces F using the functions φ_1, φ_2, φ_3 and different values of α. We take $n = 3$ $a = 2$, $b = 3$ and $h_0 = 4$, $h_1 = 3$, $h_2 = 1.7$, $h_3 = 0$ (i.e. $\Delta_2 h_j < 0$, $i = 0, 1$).



(a) $\alpha = 1, \varphi = \varphi_1$ (b) $\alpha = 2, \varphi = \varphi_1$ (c) $\alpha = 1, \varphi = \varphi_2$

(d) $\alpha = 3, \varphi = \varphi_2$ (e) $\alpha = 1, \varphi = \varphi_3$ (f) $\alpha = \pi/2, \varphi = \varphi_3$

FIGURE 1. The surface $z = F(x, y)$

If we take take the function

$$\varphi_4(t) = t^\alpha,\ \alpha \in (0,1) \cup (1,2)$$

we get the surfaces

$$F(x,y) = b_{0n}\left(\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)^\alpha\right)h + \sum_{j=1}^{n-1} b_{jn}\left(\left(\frac{x^2}{a^2}+\frac{y^2}{b^2}\right)^\alpha\right)h_j,\ (x,y) \in D.$$

If $\alpha < 1$ the point $(0,0)$ is singular point for the first order partial derivatives of the function F. If $\alpha > 1$ the point $(0,0)$ is singular point for the second order partial derivatives of the function F.

Let $D_0 = \{(x,y) \in D :\ (x,y) \neq (0,0)\}$.

If $\alpha > 1/2$ and $\Delta_2 h_j \leq 0,\ j = 0, ..., n-2$ the points of the surface F from D_0 are of elliptic type.

If $\alpha = 1/2$ and $\Delta_2 h_j = 0,\ j = 0, ..., n-2$ the points of the surface F from D_0 are of parabolic type.

If $\alpha < 1/2$ and $\Delta_2 h_j \geq 0,\ j = 0, ..., n-2$ the points of the surface F from D_0 are of hyperbolic type.

In Figure 2.a we plot the surface F for $n = 3$, $a = 2$, $b = 3$, $\alpha = 1/3$ and $h_0 = 4$, $h_1 = 3$, $h_2 = 1.7$, $h_3 = 0$ (i.e. $\Delta_2 h_j < 0$, $i = 0, 1$).

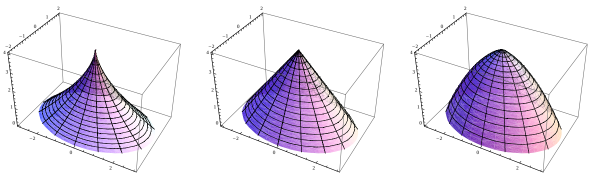
In Figure 2.b we plot the surface F for $n = 3$, $a = 2$, $b = 3$, $\alpha = 1/2$ and $h_0 = 4$, $h_1 = 8/3$, $h_2 = 4/3$, $h_3 = 0$ (i.e. $\Delta_2 h_j = 0$, $i = 0, 1$).

In Figure 2.c we plot the surface F for $a = 2$, $b = 3$, $n = 3$, $\alpha = 2/3$ and $h_0 = 4$, $h_1 = 2.5$, $h_2 = 1.2$, $h_3 = 0$ (i.e. $\Delta_2 h_j > 0$, $i = 0, 1$).

4. The second family of surfaces

Let $n \in \mathbb{N}$, $n \geq 2$ and $\widetilde{h}_i, \widetilde{h} \in \mathbb{R}$, $i = 1, ..., n-1$ such that

$$0 = \widetilde{h}_0 < \widetilde{h}_1 < ...\widetilde{h}_{n-1} < \widetilde{h}_n = h \tag{4.1}$$



(a) $\alpha = 1/3$ (b) $\alpha = 1/2$ (c) $\alpha = 2/3$

FIGURE 2. The surface $z = F(x, y)$

and let $\widetilde{f} : [0,1] \to \mathbb{R}$ be a function with the properties

$$\begin{aligned}&\widetilde{f}(0) = 0,\\&\widetilde{f}(\tfrac{j}{n}) = \widetilde{h}_j,\ j = 1, ..., n-1,\\&\widetilde{f}(1) = h.\end{aligned} \tag{4.2}$$

We obtain

$$(B_n\widetilde{f})(y) = \sum_{j=1}^{n-1} b_{jn}(y)\widetilde{h}_j + b_{n,n}(y)h. \tag{4.3}$$

The function in (4.3) has the properties

$$(B_n\widetilde{f})(0) = 0,\ (B_n\widetilde{f})(1) = h.$$

Let $\widetilde{u}$ a bivariate positive function such that the curve $\widetilde{C} : \widetilde{u}(x, y) = 0$ is the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. We assume that the curve $\widetilde{u}(x, y) = 1$ is reduced to the point $(0, 0)$.

We obtain the bivariate function

$$\begin{aligned}\widetilde{F}(x, y) &= (B_n\widetilde{f})(\widetilde{u}(x, y)) = \qquad (4.4)\\&= \sum_{j=1}^{n-1} b_{jn}(\widetilde{u}(x, y))\widetilde{h}_j + b_{nn}(\widetilde{u}(x, y))h,\ (x, y) \in D.\end{aligned}$$

The function $\widetilde{F}$ from (4.4) has the properties

$$\widetilde{F}|_{\partial D} = 0,$$

$$\widetilde{F}(0, 0) = h.$$

It follows that the surfaces $z = \widetilde{F}(x, y)$ match the curve $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, $z = 0$ (the surfaces stay on the border of domain D), and the height of the surfaces in the point $(0, 0)$ of domain is h.

From [1] we have the following results about the monotonicity and concavity of the function $\widetilde{F}$.

Theorem 4.1. *If the function $\widetilde{u}$ is increasing (decreasing) in direction (d_1, d_2) then the function $\widetilde{F}$ is increasing (decreasing) in the same direction.*

Theorem 4.2. *If $\Delta_2\widetilde{h}_j \le 0$, $j = 0, ..., n-2$ and the function $\widetilde{u}$ is concave then the function $\widetilde{F}$ is concave.*

The expression of function $P\widetilde{F}$ is given in Theorem 3.3, with $\widetilde{u}$ instead of the function u.

Next we assume that the function $\widetilde{u}$ is in the form

$$\widetilde{u}(x,y) = \widetilde{\varphi}\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right) \tag{4.5}$$

with $\widetilde{\varphi} \in C^2(0,1)$ and $\widetilde{\varphi}(0) = 1$, $\widetilde{\varphi}(1) = 0$.

Theorem 4.3. *We have*

i) If $\widetilde{\varphi}' \le 0$ on $(0,1)$ and $(d_1, d_2) \ge 0 (\le 0)$ then the function F is decreasing (increasing) on D_1

ii) If $\widetilde{\varphi}' \le 0$ on $(0,1)$ and $(-d_1, d_2) \ge 0 (\le 0)$ then the function F is decreasing (increasing) on D_2

iii) If $\widetilde{\varphi}' \le 0$ on $(0,1)$ and $(d_1, d_2) \le 0 (\ge 0)$ then the function F is decreasing (increasing) on D_3

iv) If $\widetilde{\varphi}' \le 0$ on $(0,1)$ and $(d_1, -d_2) \ge 0 (\le 0)$ then the function F is decreasing (increasing) on D_4

Theorem 4.4. *If $\Delta_2 h_j \le 0$, $j = 0, ..., n-2$ and $\widetilde{\varphi}', \widetilde{\varphi}'' \le 0$ on $(0,1)$ the surface is concave and of elliptic type.*

The proofs of Theorem 4.3 and Theorem 4.4 are analogous with the proofs of Theorem 3.4 and Theorem 3.5 respectively.

We give some examples of functions which satisfy the conditions of Theorem 4.3 and Theorem 4.4:

$$\widetilde{\varphi}_1(t) = (1-t)^\alpha, \ \alpha \in (0,1],$$

$$\widetilde{\varphi}_2(t) = \frac{e^{\alpha(1-t)} - 1}{e^\alpha - 1}, \ \alpha < 0,$$

$$\widetilde{\varphi}_3(t) = \frac{\sin\alpha(1-t)}{\sin\alpha}, \alpha \in \left(0, \frac{\pi}{2}\right].$$

The surfaces $\widetilde{F}$ obtained using the function $\widetilde{\varphi}_1$ have singular points on the border of the domain for the first order partial derivatives of the function $\widetilde{F}$.

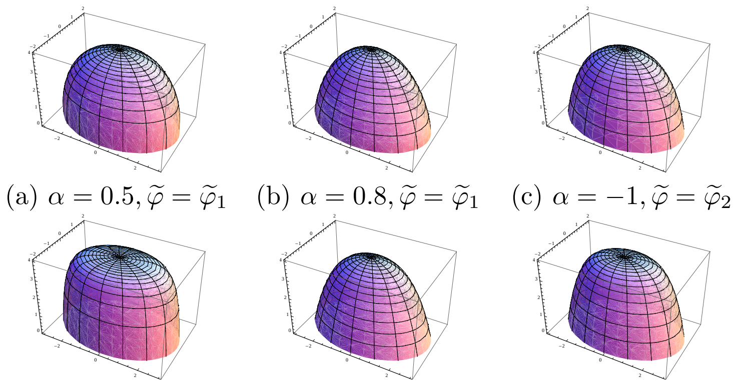
In Figure 3 we plot the surfaces $\widetilde{F}$ using the functions $\widetilde{\varphi}_1$, $\widetilde{\varphi}_2$, $\widetilde{\varphi}_3$ and different values of α. We take $a = 2$, $b = 3$, $n = 3$ and $h_0 = 4$, $h_1 = 3$, $h_2 = 1.7$, $h_3 = 0$ (i.e. $\Delta_2 h_j < 0$, $i = 0, 1$).

References

[1] Birou, M., *Blending surfaces generated using the Bernstein operator*, International Conference on Applied Mathematics, Baia Mare, 2010.

[2] Coman, Gh., Gânscă, I., Ţâmbulea, L., *Some new roof-surfaces generated by blending interpolation technique*, Stud. Univ. Babeş-Bolyai Math., **36**(1991), no. 1, 119-130.

[3] Coman, Gh., Gânscă, I., Ţâmbulea, L., *Surfaces generated by blending interpolation*, Stud. Univ. Babeş-Bolyai Math., **38**(1993), no. 3, 39-48.



(a) $\alpha = 0.5, \widetilde{\varphi} = \widetilde{\varphi}_1$ (b) $\alpha = 0.8, \widetilde{\varphi} = \widetilde{\varphi}_1$ (c) $\alpha = -1, \widetilde{\varphi} = \widetilde{\varphi}_2$

(d) $\alpha = -2, \widetilde{\varphi} = \widetilde{\varphi}_2$ (e) $\alpha = 1, \widetilde{\varphi} = \widetilde{\varphi}_3$ (f) $\alpha = \pi/2, \widetilde{\varphi} = \widetilde{\varphi}_3$

FIGURE 3. The surface $z = \widetilde{F}(x, y)$.

[4] Coons, S.A., *Surface for computer aided design of space forms*, Project MAC, Design Div., Dep. of Mech. Engineering, MIT, 1964.

[5] Ekeland, I., Temam, R., *Convex analysis and variational problems,* Classics in Applied Mathematics, 28. (SIAM), Philadelphia, PA, 1999.

[6] Farin, G., *Curves and surfaces for computer aided geometric*, Design. Academic Press, Inc Harcourt Brace Jovanovich, Publishers, Boston, San Diego, New York, London, Sidney, Tokyo, Toronto, 1990.

[7] Floater, M.S., *Shape properties of tensor-product Bernstein polynomials*, In: Shape preserving representations in computer-aided geometric design, J. M. Pena (ed.), Nova Science Publications, New York, 1999, 149-162.

[8] Floater, M.S., Pena, J.M., *Tensor-product monotonicity preservation*, Advances in Comp. Math., **9**(1998), 353-362.

[9] Lorentz, G.G., *Bernstein polynomials*, Chelsea, New York, 1986.

Marius Birou
Technical University of Cluj Napoca,
Faculty of Mathematics and Computer Sciences
28-30, Memorandumului Street,
400084 Cluj-Napoca,
Romania
e-mail: marius.birou@math.utcluj.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 247–259

Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the core and the expected value

Adriana Brândaş

Abstract. In this paper, we have suggested a new trapezoidal approximation of a fuzzy number, preserving the core and the expected value of fuzzy numbers. We have proved that the trapezoidal approximation of fuzzy numbers preserving the core and the expected value is always a fuzzy number. We have discussed the properties of this approximation.

Mathematics Subject Classification (2010): 03E72, 47S40.

Keywords: Fuzzy numbers, trapezoidal fuzzy numbers, trapezoidal approximation.

1. Introduction

Many recent papers process and transform imprecise information using the fuzzy theory. For a more efficient handling of information there is a natural need to approximate the fuzzy numbers by trapezoidal fuzzy numbers with or without additional conditions [2], [10], [8], [11], [14]. In [7] the trapezoidal approximation is a reasonable compromise between two opposite tendencies: to lose to much information and to introduce too sophisticated form of approximation from the point of view of computation.

In this paper we have proposed the trapezoidal approximation preserving the core and the expected value of a fuzzy number. Important properties (translation invariance, scale invariance, etc.) of this new trapezoidal approximation are studied in Section 4.

2. Preliminaries

We consider the following well-known description of a fuzzy number A:

$$A(x)=\begin{cases} l_A(x), & \text{if} \quad a_1 \le x \le a_2, \\ 1 & \text{if} \quad a_2 \le x \le a_3, \\ r_A(x), & \text{if} \quad a_3 \le x \le a_4, \\ 0, & \quad\text{otherwise,} \end{cases}$$

where $a_1,\ a_2,\ a_3,\ a_4,\ \in \mathbb{R}$, $l_A : [a_1, a_2] \longrightarrow [0,1]$ is a nondecreasing upper semicontinuous function, $l_A(a_1) = 0$, $l_A(a_2) = 1$, called the left side of the fuzzy number and $r_A : [a_3, a_4] \longrightarrow [0,1]$ is a nonincreasing upper semicontinuous function, $r_A(a_3) = 1$, $r_A(a_4) = 0$, called the right side of the fuzzy number. The $\alpha-$cut, $\alpha \in (0,1]$, of a fuzzy number A is a crisp set defined as $A_\alpha = \{x \in \mathbb{R} : A(x) \ge \alpha\}$.

Every $\alpha-$cut $\alpha \in [0,1]$, of a fuzzy number is a closed interval $A_\alpha = [A_L(\alpha), A_U(\alpha)]$ where

$$A_L(\alpha) = \inf\{x \in \mathbb{R} : A(x) \ge \alpha\},\ A_U(\alpha) = \sup\{x \in \mathbb{R} : A(x) \ge \alpha\}$$

for any $\alpha \in (0,1]$. If the sides of the fuzzy number A are strictly monotone then one can easily see that A_L and A_U are inverse functions of l_A and r_A, respectively.

The core or $1-$cut of a fuzzy number is defined as $core(A) = [a_2, a_3]$.

We denote by $F(\ \mathbb{R})$ the set of all fuzzy numbers.

Let $A, B \in F(\mathbb{R})\,, A_\alpha = [A_L(\alpha)\,, A_U(\alpha)]\,,\ B_\alpha = [B_L(\alpha)\,, B_U(\alpha)]\,, \alpha \in [0,1]$ and $\lambda \in \mathbb{R}$. We consider the sum $A + B$ and the scalar multiplication $\lambda \cdot A$ by (see [5]) $(A+B)_\alpha = A_\alpha + B_\alpha = [A_L(\alpha) + B_L(\alpha)\,, A_U(\alpha) + B_U(\alpha)]$ and $(\lambda \cdot A)_\alpha = \lambda A_\alpha = \begin{cases} [\lambda A_L(\alpha)\,, \lambda A_U(\alpha)]\,, & \text{if } \lambda \ge 0, \\ [\lambda A_U(\alpha)\,, \lambda A_L(\alpha)]\,, & \text{if } \lambda < 0, \end{cases}$ respectively, for every $\alpha \in [0,1]$.

A metric on the set of fuzzy numbers, which is and extension of the Euclidean distance, is defined by (see [9]) $D^2(A,B) = \int_0^1 (A_L(\alpha) - B_L(\alpha))^2 d\alpha + \int_0^1 (A_U(\alpha) - B_U(\alpha))^2 d\alpha$.

Fuzzy numbers with simple membership functions are preferred in practice. The most used such fuzzy numbers are so-called trapezoidal fuzzy numbers, given by

$$T(x)=\begin{cases} \frac{x-t_1}{t_2-t_1}, & \text{if} \quad t_1 \le x \le t_2, \\ 1, & \text{if} \quad t_2 \le x \le t_3, \\ \frac{t_4-x}{t_4-t_3}, & \text{if} \quad t_3 \le x \le t_4, \\ 0, & \quad\text{otherwise.} \end{cases}$$

We denote $T = (t_1, t_2, t_3, t_4)$ a trapezoidal fuzzy number as above. It is easy to prove that $T_L(\alpha) = t_1 + (t_2 - t_1)\alpha$ and $T_U(\alpha) = t_4 - (t_4 - t_3)\alpha$ for every $\alpha \in [0,1]$.

We denote by $F^T(\mathbb{R})$ the set of all trapezoidal fuzzy numbers.

The ambiguity Amb of $A \in F(\mathbb{R})$ is defined by (see [4])

$$Amb(A) = \int_0^1 \alpha\left(A_U(\alpha) - A_L(\alpha)\right) d\alpha$$

and the value Val of $A \in F(\mathbb{R})$ is defined by (see [4])

$$Val(A) = \int_0^1 \alpha \left(A_U(\alpha) + A_L(\alpha)\right) d\alpha.$$

The expected interval $EI(A)$ of a fuzzy number A,

$$A_\alpha = [A_L(\alpha), A_U(\alpha)]$$

is defined by (see [6], [12])

$$EI(A) = [E_*(A), E^*(A)] = \left[\int_0^1 A_L(\alpha)\, d\alpha, \int_0^1 A_U(\alpha)\, d\alpha\right],$$

expected value is given by (see [12]): $EV(A) = \frac{E_*(A)+E^*(A)}{2}$, core of A is given by (see [1]): $core\,(A) = [A_L(1), A_U(1)]$. The expected value for a trapezoidal fuzzy number $T = (t_1, t_2, t_3, t_4)$ is $EV(T) = \frac{t_1+t_2+t_3+t_4}{4}$.

Another kind of fuzzy number was introduced in [3] as follows:

$$A(x) = \begin{cases} \left(\frac{x-a}{b-a}\right)^n, x \in [a,b] \\ 1, x \in [b,c] \\ \left(\frac{d-x}{d-c}\right)^n, x \in [c,d] \\ 0, \text{ otherwise,} \end{cases}$$

where $n > 0$, $A = (a,b,c,d)_n$ with the parametric representation:

$$A_L(\alpha) = (b-a)\sqrt[n]{\alpha} + a,\ A_U(\alpha) = d - (d-c)\sqrt[n]{\alpha}, \alpha \in [0,1].$$

3. Trapezoidal approximation of fuzzy numbers

The below version of the well-known Karush-Kuhn-Tucker theorem is useful in the solving of the proposed problem.

Theorem 3.1. (*Rockafellar,* [13]) *Let $f, g_1, g_2, ..., g_m : \mathbb{R}^n \to \mathbb{R}$ be convex and differentiable function. Then $\overline{x}$ solves the convex programming problem*

$\min f(x)$

s.t. $g_i(x) \le b_i\ i \in \{1,2,3,...,m\}$

if and only if exists μ_i, $i \in \{1,2,3,...,m\}$, *such that*

(*i*) $\nabla f\left(\overline{x}\right) + \sum_{i=1}^{m} \mu_i \nabla g_i\left(\overline{x}\right) = 0;$

(*ii*) $g_i\left(\overline{x}\right) - b_i \le 0;$

(*iii*) $\mu_i \ge 0;$

(*iv*) $\mu_i\left(b_i - g_i\left(\overline{x}\right)\right) = 0.$

Given a fuzzy number A, $A_\alpha = [A_L(\alpha), A_U(\alpha)]$, $\alpha \in [0,1]$, the problem is to find a trapezoidal fuzzy number $T(A) = (t_1, t_2, t_3, t_4)$ which is the nearest to A with respect to metric D and preserves the expected value and the core of A, that is:

$$EV(A) = EV(T(A)),\, core\,(A) = core\,(T(A)).$$

The problem is reduced to minimize the distance between the fuzzy number A and the trapezoidal fuzzy number $T(A)$

$$\begin{aligned} F(t_1,t_2,t_3,t_4) &= \int_0^1 \left[A_L(\alpha) - (t_1 + (t_2 - t_1)\alpha)\right]^2 d\alpha + \\ &\quad + \int_0^1 \left[A_U(\alpha) - (t_4 + (t_3 - t_4)\alpha)\right]^2 d\alpha \end{aligned}$$

s.t.

$$t_2 = A_L(1), \tag{3.1}$$

$$t_3 = A_U(1), \tag{3.2}$$

so

$$2\int_0^1 \left[A_L(\alpha) + A_U(\alpha)\right] d\alpha = t_1 + t_2 + t_3 + t_4. \tag{3.3}$$

The conditions for $T(A) = (t_1, t_2, t_3, t_4)$ to be a trapezoidal fuzzy number are

$$t_1 \le t_2 \tag{3.4}$$

$$t_2 \le t_3 \tag{3.5}$$

and

$$t_3 \le t_4. \tag{3.6}$$

Taking into account the relations $(3.1) - (3.3)$, t_4 becomes:

$$t_4 = 2\int_0^1 (A_L(\alpha) + A_U(\alpha))\, d\alpha - t_1 - A_L(1) - A_U(1),$$

so $F(t_1,t_2,t_3,t_4)$ becomes $g(t_1)$

$$\begin{aligned} g(t_1) = \tfrac{2t_1^2}{3} + 2t_1 \textstyle\int_0^1 (A_L(\alpha) - E(\alpha))(\alpha - 1)\, d\alpha + \tfrac{t_1}{3} A_L(1) + \\ + \textstyle\int_0^1 (E(\alpha))^2 d\alpha + \int_0^1 (A_L(\alpha) - A_L(1)\alpha)^2 d\alpha, \end{aligned} \tag{3.7}$$

where

$$E(\alpha) = A_U(\alpha)$$

$$+2(\alpha - 1)\int_0^1 (A_L(\alpha) + A_U(\alpha))\, d\alpha - A_L(1)(\alpha - 1) - A_U(1)(2\alpha - 1),$$

so conditions $(3.4) - (3.6)$ are:

$$t_1 \le A_L(1) \tag{3.8}$$

$$A_L(1) \le A_U(1) \tag{3.9}$$

and

$$t_1 \le 2\int_0^1 A_L(\alpha)\, d\alpha + 2\int_0^1 A_U(\alpha)\, d\alpha - A_L(1) - 2A_U(1). \tag{3.10}$$

For any fuzzy number the relation (3.9) is always true.

Theorem 3.2. *If A, $A_\alpha = [A_L(\alpha), A_U(\alpha)]$ is a fuzzy number and $T(A) = (t_1, t_2, t_3, t_4)$ denotes the nearest (with respect to metric D) trapezoidal fuzzy number A preserving the expected value and the core, then*

(i) If

$$\int_0^1 [(2-6\alpha) A_U(\alpha) + (6\alpha - 10) A_L(\alpha)]\, d\alpha + 7A_L(1) + A_U(1) < 0 \quad (3.11)$$

and

$$\int_0^1 [A_L(\alpha) + A_U(\alpha)]\, d\alpha \geq A_L(1) + A_U(1), \quad (3.12)$$

then

$$t_1 = t_2 = A_L(1);\ t_3 = A_U(1);$$

$$t_4 = 2\int_0^1 A_L(\alpha)\, d\alpha + 2\int_0^1 A_U(\alpha)\, d\alpha - 2A_L(1) - A_U(1).$$

(ii) If

$$\int_0^1 [(2-6\alpha) A_L(\alpha) + (6\alpha - 10) A_U(\alpha)]\, d\alpha + A_L(1) + 7A_U(1) > 0 \quad (3.13)$$

and

$$\int_0^1 [A_L(\alpha) + A_U(\alpha)]\, d\alpha \leq A_L(1) + A_U(1), \quad (3.14)$$

then

$$t_1 = 2\int_0^1 A_L(\alpha)\, d\alpha + 2\int_0^1 A_U(\alpha)\, d\alpha - A_L(1) - 2A_U(1);$$

$$t_2 = A_L(1);\ t_3 = t_4 = A_U(1).$$

(iii) If

$$\int_0^1 [(2-6\alpha) A_U(\alpha) + (6\alpha - 10) A_L(\alpha)]\, d\alpha + 7A_L(1) + A_U(1) \geq 0 \quad (3.15)$$

and

$$\int_0^1 [(2-6\alpha) A_L(\alpha) + (6\alpha - 10) A_U(\alpha)]\, d\alpha + A_L(1) + 7A_U(1) \leq 0 \quad (3.16)$$

then

$$t_1 = -\frac{3}{2}\int_0^1 \alpha (A_L(\alpha) - A_U(\alpha))\, d\alpha - \frac{3}{4}A_L(1)$$
$$+\frac{1}{2}\int_0^1 (5A_L(\alpha) - A_U(\alpha))\, d\alpha - \frac{1}{4}A_U(1);$$

$$t_2 = A_L(1);\ t_3 = A_U(1);$$

$$t_4 = \frac{3}{2}\int_0^1 \alpha (A_L(\alpha) - A_U(\alpha))\, d\alpha - \frac{1}{4}A_L(1)$$
$$-\frac{1}{2}\int_0^1 (A_L(\alpha) - 5A_U(\alpha))\, d\alpha - \frac{3}{4}A_U(1).$$

Proof. Let us remark that the hypothesis of convexity and differentiability in the Karush-Kuhn-Tucker theorem are satisfied for the function g given by (3.7) under conditions (3.8) – (3.10). After some calculations we can write the conditions of Karush-Kuhn-Tucker theorem to minimize the function g, in the following way:

$$\frac{4t_1}{3} + 2\int_0^1 \alpha \left[A_L(\alpha) - A_U(\alpha)\right] d\alpha + \frac{2}{3}\int_0^1 \left(A_U(\alpha) - 5A_L(\alpha)\right) d\alpha +$$

$$+A_L(1) + \frac{A_U(1)}{3} + \mu_1 + \mu_2 = 0 \tag{3.17}$$

$$\mu_1\left(t_1 - A_L(1)\right) = 0 \tag{3.18}$$

$$\mu_2\left(t_1 - 2\int_0^1 A_L(\alpha)\, d\alpha - 2\int_0^1 A_U(\alpha)\, d\alpha + A_L(1) + 2A_U(1)\right) = 0 \tag{3.19}$$

$$\mu_1 \geq 0 \tag{3.20}$$

$$\mu_2 \geq 0 \tag{3.21}$$

$$t_1 - A_L(1) \leq 0 \tag{3.22}$$

$$t_1 - 2\int_0^1 A_L(\alpha)\, d\alpha - 2\int_0^1 A_U(\alpha)\, d\alpha + A_L(1) + 2A_U(1) \leq 0 \tag{3.23}$$

If $\mu_1 \neq 0$ and $\mu_2 \neq 0$, then the solution is: $t_1 = A_L(1)$ and

$$t_1 = 2\int_0^1 A_L(\alpha)\, d\alpha + 2\int_0^1 A_U(\alpha)\, d\alpha - A_L(1) - 2A_U(1),$$

so

$$t_4 = 2\int_0^1 A_L(\alpha)\, d\alpha + 2\int_0^1 A_U(\alpha)\, d\alpha - 2A_L(1) - A_U(1)$$

or $t_4 = A_U(1)$, from (3.17) we obtain that

$$\mu_1 + \mu_2 = -2\int_0^1 \alpha\left[A_L(\alpha) - A_U(\alpha)\right] d\alpha$$

$$-\frac{2}{3}\int_0^1 \left(A_U(\alpha) - 5A_L(\alpha)\right) d\alpha - A_L(1) - \frac{A_U(1)}{3} \tag{3.24}$$

but $\mu_1 \neq 0$ and $\mu_2 \neq 0$, so

$$\int_0^1 A_L(\alpha)\, d\alpha + \int_0^1 A_U(\alpha)\, d\alpha = A_L(1) + A_U(1),$$

by (3.24), we have

$$\mu_1 + \mu_2 = \int_0^1 (1 - 2\alpha) A_L(\alpha)\, d\alpha + \int_0^1 (2\alpha - 3) A_U(\alpha)\, d\alpha + 2A_U(1)$$

so

$$\mu_1 + \mu_2 = \int_0^1 (1 - 2\alpha) A_L(\alpha)\, d\alpha + \int_0^1 (2\alpha - 1) A_U(\alpha)\, d\alpha$$

$$+2\left(A_U(1) - \int_0^1 A_U(\alpha)\, d\alpha\right)$$

taking into account Lemma 1 from [2] results that $\mu_1 + \mu_2 \leq 0$. In fact, because $\mu_1 \neq 0$ and $\mu_2 \neq 0$ we obtain that $\mu_1 + \mu_2 < 0$, so Karush-Kuhn-Tucker conditions can not be verified, which means that we have no solution in this case.

(i) If $\mu_1 \neq 0$ and $\mu_2 = 0$, then from (3.17) and (3.18) we obtain that:

$$t_1 = A_L(1)$$

and

$$\mu_1 = 2\int_0^1 \alpha \left[A_U(\alpha) - A_L(\alpha)\right] d\alpha$$

$$-\frac{2}{3}\int_0^1 \left(A_U(\alpha) - 5A_L(\alpha)\right) d\alpha - \frac{A_U(1)}{3} - \frac{7A_L(1)}{3}, \mu_2 = 0$$

and from (3.1), (3.2) and (3.3) we obtain that

$$t_4 = 2\int_0^1 A_L(\alpha)\, d\alpha + 2\int_0^1 A_U(\alpha)\, d\alpha - 2A_L(1) - A_U(1).$$

(ii) If $\mu_1 = 0$ and $\mu_2 \neq 0$, then from (3.17) and (3.19) we obtain that

$$t_1 = 2\int_0^1 A_L(\alpha)\, d\alpha + 2\int_0^1 A_U(\alpha)\, d\alpha - A_L(1) - 2A_U(1) \text{ and } \mu_1 = 0,$$

$$\mu_2 = \frac{1}{3}\int_0^1 \left[(2-6\alpha) A_L(\alpha) + (6\alpha - 10) A_U(\alpha)\right] d\alpha + \frac{1}{3}\left(A_L(1) + 7A_U(1)\right)$$

and $t_4 = A_U(1)$.

(iii) If $\mu_1 = 0$ and $\mu_2 = 0$, then from (3.17), (3.22) and (3.23) we obtain that

$$t_1 = -\frac{3}{2}\int_0^1 \alpha \left(A_L(\alpha) - A_U(\alpha)\right) d\alpha + \frac{1}{2}\int_0^1 \left(5A_L(\alpha) - A_U(\alpha)\right) d\alpha$$

$$-\frac{3}{4}A_L(1) - \frac{1}{4}A_U(1)$$

and

$$t_4 = \frac{3}{2}\int_0^1 \alpha \left(A_L(\alpha) - A_U(\alpha)\right) d\alpha - \frac{1}{2}\int_0^1 \left(A_L(\alpha) - 5A_U(\alpha)\right) d\alpha$$

$$-\frac{1}{4}A_L(1) - \frac{3}{4}A_U(1).$$

□

Remark 3.3. Any fuzzy number can apply one and only case of the Theorem 3.2.

Proof. Let us denote

$$\Gamma_1 = \{A : A \in F(\mathbb{R}) \text{ and the case } (i) \text{ is applicable to } A\},$$

$$\Gamma_2 = \{A : A \in F(\mathbb{R}) \text{ and the case } (ii) \text{ is applicable to } A\}$$

and

$$\Gamma_3 = \{A : A \in F(\mathbb{R}) \text{ and the case } (iii) \text{ is applicable to } A\}.$$

It is obvious that $\Gamma_3 = (\Gamma_1 \cup \Gamma_2)^c$, so, the three cases of Theorem 3.2 cover the set of all fuzzy numbers. On the other hand $\Gamma_1 \cap \Gamma_3 = \varnothing$ because the relation (3.11) is complementary with the relation (3.15). $\Gamma_2 \cap \Gamma_3 = \varnothing$ because the relation (3.13) is complementary with the relation (3.16). □

Example 3.4. Let A be a fuzzy number $A_\alpha = [1 + 99\sqrt{\alpha}, 200 - 95\sqrt{\alpha}]$ then the trapezoidal approximation preserving the core and the expected value is $T(A) = (t_1, t_2, t_3, t_4)$ and can be calculated with case (iii) of Theorem 3.2 as follows: $t_1 = \frac{923}{30}$; $t_2 = 100$; $t_3 = 105$; $t_4 = \frac{5147}{30}$.

Theorem 3.5. *Let $A = (a, b, c, d)_n$ be a fuzzy number and $T(A) = (t_1, t_2, t_3, t_4)$ the nearest (with respect to metric D) trapezoidal fuzzy number A preserving the expected value and the core*

(i) If $(n-1)(d-c) + (17n+7)(b-a) < 0$ and $a - b - c + d \geq 0$, then

$$t_1 = t_2 = b;\ t_3 = c;\ t_4 = \frac{2a - 2b - c + 2d + cn}{n+1}.$$

(ii) If $(b-a)(1-n) - (17n+7)(d-c) > 0$ and $a - b - c + d \leq 0$, then

$$t_1 = \frac{2a - b - 2c + 2d + bn}{n+1};\ t_2 = b;\ t_3 = t_4 = c.$$

(iii) If $(b-a)(1-n) - (17n+7)(d-c) \leq 0$ and

$$(n-1)(d-c) + (17n+7)(b-a) \geq 0$$

then

$$t_1 = \frac{7a - 3b - c + d + 8bn^2 + 17an - 5bn + cn - dn}{4(n+1)(2n+1)};\ t_2 = b;\ t_3 = c;$$

$$t_4 = \frac{a - b - 3c + 7d + 8cn^2 - an + bn - 5cn + 17dn}{4(n+1)(2n+1)}.$$

Proof. Let $A = (a, b, c, d)_n$ be a fuzzy number, then $A_L(1) = b$, $A_U(1) = c$ and

$$\int_0^1 A_L(\alpha)\, d\alpha = \frac{a + bn}{n+1};\ \int_0^1 A_U(\alpha)\, d\alpha = \frac{cn + d}{n+1};$$

$$\int_0^1 \alpha A_L(\alpha)\, d\alpha = \frac{a + 2bn}{4n+2};\ \int_0^1 \alpha A_U(\alpha)\, d\alpha = \frac{2cn + d}{4n+2}.$$

Applying Theorem 3.2 the result is immediately. □

Example 3.6. For a fuzzy number $A = (2, 3, 4, 40)_{\frac{1}{2}}$ applying the case (i) of Theorem 3.5 we obtain the trapezoidal approximation which preserves the expected value and the core of A: $T(A) = \left(3, 3, 4, \frac{152}{3}\right)$.

Example 3.7. For a fuzzy number $A = (-200, 0, 10, 20)_{\frac{1}{5}}$ applying the case (ii) of Theorem 3.5 we obtain the trapezoidal approximation which preserves the expected value and the core of A: $T(A) = \left(-\frac{950}{3}, 0, 10, 10\right)$.

Example 3.8. For a fuzzy number $A = (1, 2, 3, 100)_2$ applying the case (iii) of Theorem 3.5 we obtain the trapezoidal approximation which preserves the expected value and the core of A: $T(A) = \left(-\frac{3}{10}, 2, 3, \frac{693}{10}\right)$.

4. Properties

In fuzzy theory are many approximate methods of fuzzy numbers and an infinite number of approximation operators. For the present approximation operator we study some properties proposed by Grzgorzewski and Mrowka in paper [7].

Theorem 4.1. *The trapezoidal approximation preserving the core and the expected value given in Theorem* 3.2*:*

(i) is invariant to translation;
(ii) is scale invariant;
(iii) fulfills the nearness criterion;
(iv) fulfills the identity criterion.

Proof. (i) If $A \in \Gamma_1$ then the conditions $(3.11 - 3.12)$ are verified, so

$$\int_0^1 \left[(2-6\alpha)\left(A+z\right)_U(\alpha) + (6\alpha - 10)\left(A+z\right)_L(\alpha)\right] d\alpha$$
$$+7\left(A+z\right)_L(1) + \left(A+z\right)_U(1)$$
$$= \int_0^1 \left[(2-6\alpha) A_U(\alpha) + 2z + (6\alpha-10) A_L(\alpha) - 10z\right] d\alpha$$
$$+7A_L(1) + A_U(1) + 8z$$
$$= \int_0^1 \left[(2-6\alpha) A_U(\alpha) + (6\alpha-10) A_L(\alpha)\right] d\alpha + 7A_L(1) + A_U(1) < 0,$$

and

$$\int_0^1 \left[\left(A+z\right)_L(\alpha) + \left(A+z\right)_U(\alpha)\right] d\alpha - \left(A+z\right)_L(1) - \left(A+z\right)_U(1)$$
$$= \int_0^1 \left[A_L(\alpha) + A_U(\alpha)\right] d\alpha - A_L(1) - A_U(1) \geq 0,$$

so $A+z \in \Gamma_1$.

We obtain that

$$t_1(A+z) = \left(A+z\right)_L(1) = A_L(1) + z = t_1(A) + z,$$
$$t_2(A+z) = \left(A+z\right)_L(1) = A_L(1) + z = t_2(A) + z,$$
$$t_3(A+z) = \left(A+z\right)_U(1) = A_U(1) + z = t_3(A) + z,$$

and

$$t_4(A+z) = 2\int_0^1 \left[\left(A+z\right)_L(\alpha) + \left(A+z\right)_U(\alpha)\right] d\alpha$$
$$-2\left(A+z\right)_L(1) - \left(A+z\right)_U(1)$$
$$= 2\int_0^1 A_L(\alpha)\, d\alpha + 2\int_0^1 A_U(\alpha)\, d\alpha - 2A_L(1) - A_U(1) + 4z - 2z - z = t_4(A) + z,$$

so $T(A+z) = T(A) + z$, $A \in \Gamma_1$.

It is obviously that for $A \in \Gamma_2$ we obtain that $A+z \in \Gamma_2$ and for $A \in \Gamma_3$ we obtain that $A+z \in \Gamma_3$ so $T(A+z) = T(A) + z$, for every $A \in \Gamma_i$, $i \in \{1,2,3\}$.

(ii) For $\lambda > 0$ the proof is immediate because $(\lambda A)_L(\alpha) = \lambda A_L(\alpha)$ and $(\lambda A)_U(\alpha) = \lambda A_U(\alpha)$, $\alpha \in [0,1]$ so $T(\lambda \cdot A) = \lambda \cdot T(A)$, $A \in \Gamma_i$, $i \in \{1,2,3\}$ and from $A \in \Gamma_i$ results that $\lambda \cdot A \in \Gamma_i$, $i \in \{1,2,3\}$.

In case $\lambda < 0$ we have $(\lambda A)_L(\alpha) = \lambda A_U(\alpha)$ and $(\lambda A)_U(\alpha) = \lambda A_L(\alpha)$, for every $\alpha \in [0,1]$.

If $A \in \Gamma_1$ then the conditions $(3.11 - 3.12)$ are verified

$$\lambda \int_0^1 \left[(2-6\alpha) A_U(\alpha) + (6\alpha - 10) A_L(\alpha)\right] d\alpha + 7\lambda A_L(1) + \lambda A_U(1) > 0,$$

so

$$\int_0^1 \left[(2-6\alpha) \lambda A_U(\alpha) + (6\alpha - 10) \lambda A_L(\alpha)\right] d\alpha + 7\lambda A_L(1) + \lambda A_U(1) > 0$$

and is equivalent to

$$\int_0^1 \left[(2-6\alpha)(\lambda A)_L(\alpha) + (6\alpha - 10)(\lambda A)_U(\alpha)\right] d\alpha$$
$$+7(\lambda A)_U(1) + (\lambda A)_L(1) > 0$$

and

$$\lambda \int_0^1 \left[A_L(\alpha) + A_U(\alpha)\right] d\alpha \le \lambda A_L(1) + \lambda A_U(1)$$

so

$$\int_0^1 \left[\lambda A_L(\alpha) + \lambda A_U(\alpha)\right] d\alpha \le \lambda A_L(1) + \lambda A_U(1)$$

and is equivalent to

$$\int_0^1 \left[(\lambda A)_U(\alpha) + (\lambda A)_L(\alpha)\right] d\alpha \le (\lambda A)_U(1) + (\lambda A)_L(1),$$

we obtain that $\lambda \cdot A \in \Gamma_2$, so

$$T(\lambda \cdot A) = \Bigg(2\int_0^1 (\lambda \cdot A)_L(\alpha)\, d\alpha + 2\int_0^1 (\lambda \cdot A)_U(\alpha)\, d\alpha - 2(\lambda \cdot A)_L(1) -$$
$$-(\lambda \cdot A)_U(1), (\lambda \cdot A)_L(1), (\lambda \cdot A)_U(1), (\lambda \cdot A)_U(1))$$
$$= (\lambda \cdot A_L(1), \lambda \cdot A_L(1), \lambda \cdot A_U(1),$$
$$2\int_0^1 \lambda \cdot A_L(\alpha)\, d\alpha + 2\int_0^1 \lambda \cdot A_U(\alpha)\, d\alpha - 2\lambda \cdot A_U(1) - \lambda \cdot A_L(1)\Bigg)$$
$$= \lambda \cdot T(A)$$

then $T(\lambda \cdot A) \in \Gamma_1$. $\lambda \cdot A$ is in case (ii) of Theorem 3.2 if and only if A is in case (i) of Theorem 3.2.

For $A \in \Gamma_2$ we obtain that $\lambda \cdot A \in \Gamma_2$, so $\lambda \cdot A$ is in case (i) of Theorem 3.2 if and only if A is in case (ii) of Theorem 3.2.

Similarly it can be shown that $\lambda \cdot A$ is in case (iii) of Theorem 3.2 if and only if A is in case (iii) of Theorem 3.2.

(iii) By the construction of the operator under study.

(iv) If $A = (a_1, a_2, a_3, a_4)$, $A_\alpha = [a_1 + \alpha (a_2 - a_1), a_4 - \alpha (a_4 - a_3)]$, then

$$\int_0^1 [(2 - 6\alpha) A_U(\alpha) + (6\alpha - 10) A_L(\alpha)] \, d\alpha + 7A_L(1) + A_U(1)$$
$$= 4(a_2 - a_1) \geq 0$$

and

$$\int_0^1 [(2 - 6\alpha) A_L(\alpha) + (6\alpha - 10) A_U(\alpha)] \, d\alpha + A_L(1) + 7A_U(1)$$
$$= 4(a_3 - a_4) \leq 0$$

so (3.15) and (3.16) are verified and the case (iii) of Theorem 3.2 is applicable to A. We obtain that

$$t_1 = -\frac{3}{2} \int_0^1 \alpha \left((a_1 + \alpha (a_2 - a_1)) - (a_4 - \alpha (a_4 - a_3)) \right) d\alpha$$
$$-\frac{1}{4}(a_4 - (a_4 - a_3)) + \frac{1}{2} \int_0^1 \left(5(a_1 + \alpha (a_2 - a_1)) - (a_4 - \alpha (a_4 - a_3)) \right) d\alpha$$
$$-\frac{3}{4}(a_1 + (a_2 - a_1)) = a_1,$$
$$t_2 = A_L(1) = a_2; t_3 = A_U(1) = a_3$$

and

$$t_4 = \frac{3}{2} \int_0^1 \alpha \left((a_1 + \alpha (a_2 - a_1)) - (a_4 - \alpha (a_4 - a_3)) \right) d\alpha$$
$$-\frac{3}{4}(a_4 - (a_4 - a_3)) - \frac{1}{2} \int_0^1 \left((a_1 + \alpha (a_2 - a_1)) - 5(a_4 - \alpha (a_4 - a_3)) \right) d\alpha$$
$$-\frac{1}{4}(a_1 + (a_2 - a_1)) = a_4,$$

so $T(A) = A, \forall A \in F^T(\mathbb{R})$. □

Remark 4.2. Let A be a trapezoidal fuzzy number, the trapezoidal approximation preserving the core and the expected value does not preserve the ambiguity and the value of a fuzzy number: $Amb(A) \neq Amb(T(A))$, $Val(A) \neq Val(T(A))$, for any $A \in F(\mathbb{R})$.

Example 4.3. Let A be a fuzzy number $A = (0, 1, 2, 3)_2$, $A_L(\alpha) = \sqrt{\alpha}$, $A_U(\alpha) = 3 - \sqrt{\alpha}$. The trapezoidal approximation which preserves the expected value and core is $T(A) = \left(\frac{3}{10}, 1, 2, \frac{27}{10}\right)$ .The ambiguity of the fuzzy number A is: $Amb(A) = \frac{7}{10}$ and the ambiguity of the trapezoidal approximation is: $Amb(T(A)) = \frac{11}{15}$, so $Amb(A) \neq Amb(T(A))$.

Example 4.4. For a fuzzy number $A = [\sqrt{\alpha} + 1, 30 - 27\sqrt{\alpha}]$ the value of the fuzzy number is $Val(A) = \frac{51}{10}$, the trapezoidal approximation which preserves the expected value and the core is $T(A) = \left(\frac{13}{15}, 2, 3, \frac{322}{15}\right)$ so it's value is: $Val(T(A)) = \frac{97}{18}$, then $Val(A) \neq Val(T(A))$.

The following example studies the continuity of the trapezoidal operator which preserves the core and the expected value.

Example 4.5. The trapezoidal operator introduced by Theorem 3.2 is not continuous, with respect to the distance D. Indeed, if $A \in F(\mathbb{R}), A_n \in F(\mathbb{R}), n \in \mathbb{N}$ such that $A_L(\alpha) = \sqrt{\alpha}$, $A_U(\alpha) = 3 - \sqrt{\alpha}$ and $(A_n)_L = \sqrt{\alpha} + \alpha^n$, $(A_n)_U = 3 - \sqrt{\alpha}, \alpha \in [0,1]$, then the trapezoidal fuzzy numbers which preserve the expected value and the core of A and A_n are:

$$T(A) = \left(\frac{3}{10}, 1, 2, \frac{27}{10}\right),$$

$$T(A_n) = \left(-\frac{7n + 9n^2 - 52}{20(n+2)(n+1)}, 2, 2, \frac{27}{10}\right), n \in \mathbb{N}$$

and

$$\lim_{n\to\infty} D(T(A_n), T(A)) = \frac{\sqrt{39}}{12}$$

and

$$\lim_{n\to\infty} D(A_n, A) = \lim_{n\to\infty} \sqrt{\frac{1}{2n+1}} = 0,$$

so $\lim\limits_{n\to\infty} D(T(A_n), T(A)) = \frac{\sqrt{39}}{12} \neq 0 = \lim\limits_{n\to\infty} D(A_n, A)$.

5. Conclusion

In the present paper a new trapezoidal approximation of a fuzzy number was added to trapezoidal approximations already introduced in [1], [2], [7], [8], [14]. It has multiple advantages: can be easy calculated, has some important properties: scale invariance, identity, translation invariance, nearness criterion.

References

[1] Allahviranloo, T., Adabitabar Firozja, M., *Note on "Trapezoidal approximation of fuzzy numbers"*, Fuzzy Sets and Systems, **158**(2007), 755-756.

[2] Ban, A., *Approximation of fuzzy numbers by trapezoidal fuzzy numbers preserving the expected interval*, Fuzzy Sets and Systems, **159**(2008), 1327-1344.

[3] Bodjanova, S., *Median value interval of a fuzzy number*, Inform. Sci., **172**(2005), 73-89.

[4] Delgado, M., Vila, M.A., Voxman, W., *On a canonical representation of fuzzy numbers*, Fuzzy Sets and Systems, **93**(1998), 125-135.

[5] Diamond, P., Kloeden, P., *Metric spaces of fuzzy sets*, Theory and Applications, World Scientific, Singapore, 1994.

[6] Dubois, D., Prade, H., *The mean value of a fuzzy number*, Fuzzy Sets and Systems, **24**(1987), 279-300.

[7] Grzegorzewski, P., Mrówka, E., *Trapezoidal approximations of fuzzy numbers*, Fuzzy Sets and Systems, **153**(2005), 115-135.

[8] Grzegorzewski, P., Mrówka, E., *Trapezoidal approximations of fuzzy numbers - revisited*, Fuzzy Sets and Systems, **158**(2007), 757-768.

[9] Grzegorzewski, P., *Metrics and orders in space of fuzzy numbers*, Fuzzy Sets and Systems, **97**(1998), 83-94.

[10] Grzegorzewski, P., *Trapezoidal approximations of fuzzy numbers preserving the expected interval - Algorithms and properties*, Fuzzy Sets and Systems, **47**(2008), 1354-1353.

[11] Grzegorzewski, P., *New algorithms for trapezoidal approximation of fuzzy numbers preserving the expected interval*, IPMU'08, 2008, 117-123.

[12] Heilpern, S., *The expected value of a fuzzy number*, Fuzzy Sets and Systems, **47**(1992), 81-86.

[13] Rockafellar, R.T., *Convex analysis*, Princeton University Press, NJ, 1970.

[14] Yeh, C.T., *Trapezoidal and triangular approximations preserving the expected interval*, Fuzzy Sets and Systems, **159**(2008), 1345-1353.

Adriana Brândaş
Babeş-Bolyai University
Cluj-Napoca, Romania
e-mail: adrianabrandas@yahoo.com

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 261–272

Approximation of cosine functions and Rogosinski type operators

Michele Campiti, Giusy Mazzone and Cristian Tacelli

Abstract. We study some quantitative estimates of the convergence of the iterates of some Rogosinski type operators to their associated cosine functions. We also consider a general cosine counterpart of the quantitative version of Trotter's theorem on the approximation of C_0-semigroups.

Mathematics Subject Classification (2010): 41A25, 42A10, 47D09.

Keywords: Convergence of iterates of trigonometric polynomials, Rogosinski operators, representation of cosine functions.

1. Introduction and preliminary results

The convergence of iterates of trigonometric polynomials to suitable cosine functions in the setting of spaces of continuous periodical functions has been considered in [7, 8] from a qualitative point of view; these results extend to cosine function the possibility of using iterates of positive operators in the approximation of C_0-semigroups (see [2, Chapter 6] for more details). Recently, some quantitative versions of the classical Trotter's theorem [17] on the approximation of C_0-semigroups have been obtained in [14, 15] and [9, 10]. Here, we consider the possibility of obtaining quantitative estimates of the convergence to suitable cosine functions. We study in particular the Rogosinski type operators introduced in [7, 8] and establish some quantitative estimates of the convergence of their iterates to a cosine function generated by the square of a first order differential operator. Our discussion is based on the following general quantitative cosine version of Trotter's approximation theorem [17, Theorem 5.3], which provides a quantitative estimate of the convergence and, besides the Rogosinski type operators, can be applied also to other sequences of operators, such as Fejér operators and the general sequences of averages of trigonometric interpolating operators considered in [6]. A partial result on the generation of cosine functions has been obtained in [8, Theorem 1.2] without quantitative estimates.

Theorem 1.1. *Let E be a Banach space, let $(L_n)_{n\geq 1}$ and $(M_n)_{n\geq 1}$ be two sequences of bounded linear operators from E into itself and assume that there exist $M \geq 1$ and $\omega \geq 0$ such that*

$$\|L_n^k\| \leq M\, e^{\omega\, k/n}\,, \qquad \|M_n^k\| \leq M e^{\omega\, k/n}\,, \quad n\,, k \geq 1. \tag{1.1}$$

Suppose also that D is a dense subspace of E and $A : D \to E$ is a linear operator such that

$$\lim_{n\to+\infty} n(L_n u - u) = Au\,, \qquad \lim_{n\to+\infty} n(M_n u - u) = -Au$$

and $(\lambda - A)(D)$ is dense in E for some $\lambda > \omega$.

Then the closure of (A, D) generates a C_0-group $(G(t))_{t\in\mathbb{R}}$ in E and then the square A^2 of the closure of (A, D) generates a cosine function $(C(t))_{t\in\mathbb{R}}$ in E and, for every $t \geq 0$,

$$C(t) = \frac{1}{2} \lim_{n\to\infty} \left(L_n^{k(n)} + M_n^{k(n)}\right)\,, \tag{1.2}$$

where $(k(n)_n)_{n\in\mathbb{N}}$ is a sequence of positive integers such that

$$\lim_{n\to+\infty} k(n)/n = t$$

(in particular, we can take $k(n) = [n\,t]$). Consequently, for every $t \in \mathbb{R}$, we have $\|C(t)\| \leq M\, e^{\omega\,|t|}$.

Moreover, for every $t \geq 0$ and for every increasing sequence $(k(n))_{n\geq 1}$ of positive integers and $u \in \{v \in D|\, G(s)v,\ G(-s)v \in D$ for every $0 \leq s \leq t\}$, we have

$$\begin{aligned}
&\left\|C(t)u - \frac{1}{2}\left(L_n^{k(n)}u + M_n^{k(n)}u\right)\right\| \qquad\qquad (1.3)\\
&\quad\leq \frac{M}{2}\,\exp(\omega\, e^{\omega/n}\, t) \int_0^t \exp(-\omega e^{\omega/n} s)\,(\|(n(L_n - I) - A)G(s)u\|\\
&\qquad\qquad + \|(n(M_n - I) + A)G(-s)u\|)\, ds\\
&\qquad + \frac{M}{2}\left(\exp(\omega\, e^{\omega/n}\, t_n)\ |k(n) - nt| + \sqrt{\frac{2k(n)}{\pi}}\, e^{\omega\, k(n)/n}\right.\\
&\qquad\qquad \left. + \frac{\omega k(n)}{n}\,\exp\left(\omega\, e^{\omega/n}\,\frac{k(n)}{n}\right)\right)(\|L_n u - u\| + \|M_n u - u\|)
\end{aligned}$$

where $t_n := \sup\{t, k(n)/n\}$.

Proof. From the classical Trotter's theorem [17, Theorem 5.3] it follows that the closure of the operators A and $-A$ generate a C_0-semigroup $(T_+(t))_{t\geq 0}$ and respectively $(T_-(t))_{t\geq 0}$ in E. Consequently, the closure of A generates a C_0-group $(G(t))_{t\in\mathbb{R}}$ in E and, for every $t \geq 0$,

$$G(t) = T_+(t)\,, \qquad G(-t) = T_-(t)\,.$$

Moreover, again from [17, Theorem 5.3], we obtain the representation of the group $(G(t))_{t\in\mathbb{R}}$ in terms of iterates of the operators L_n and M_n; indeed, for every $t \geq 0$ and for every sequence $(k(n)_n)_{n\in\mathbb{N}}$ of positive integers such that $\lim_{n\to+\infty} k(n)/n = t$, we have

$$G(t) = \lim_{n\to+\infty} L_n^{k(n)}\,, \qquad G(-t) = \lim_{n\to+\infty} M_n^{k(n)}\,.$$

Consequently, it follows that the square of the closure of (A, D) generates a cosine function $(C(t))_{t\in\mathbb{R}}$ in E (see [4, Example 3.14.15, p. 217]) and, for every $t \in \mathbb{R}$, $C(t) = (G(t) + G(-t))/2$. Hence the representation of the cosine function is a consequence of the representation of $(G(t))_{t\in\mathbb{R}}$ and the estimate $\|C(|t|)\| \leq M\, e^{\omega\, t}$ follows from (1.1) and (1.2).

Finally, we prove (1.3).

Let $t \geq 0$, $(k(n))_{n\geq 1}$ an increasing sequence of positive integers and $u \in \{v \in D|\, G(s)v,\ G(-s)v \in D$ for every $0 \leq s \leq t\}$.

From [9, Theorem 1.2] and our assumptions we get

$$\begin{aligned}
&\left\|T_+(t)u - L_n^{k(n)}u\right\| \\
&\quad\leq M\,\exp(\omega\, e^{\omega/n}\, t)\int_0^t \exp(-\omega e^{\omega/n} s)\|(n(L_n - I) - A)T_+(s)u\|ds \\
&\qquad + M\left(\exp(\omega\, e^{\omega/n}\, t_n)\ |k(n) - nt| + \sqrt{\frac{2k(n)}{\pi}}\, e^{\omega\, k(n)/n}\right. \\
&\qquad \left. + \frac{\omega k(n)}{n}\,\exp\left(\omega\, e^{\omega/n}\,\frac{k(n)}{n}\right)\right)\ \|L_n u - u\|
\end{aligned}$$

and

$$\begin{aligned}
&\left\|T_-(t)u - M_n^{k(n)}u\right\| \\
&\quad\leq M\,\exp(\omega\, e^{\omega/n}\, t)\int_0^t \exp(-\omega e^{\omega/n} s)\|(n(M_n - I) + A)T_-(s)u\|ds \\
&\qquad + M\left(\exp(\omega\, e^{\omega/n}\, t_n)\ |k(n) - nt| + \sqrt{\frac{2k(n)}{\pi}}\, e^{\omega\, k(n)/n}\right. \\
&\qquad \left. + \frac{\omega k(n)}{n}\,\exp\left(\omega\, e^{\omega/n}\,\frac{k(n)}{n}\right)\right)\ \|M_n u - u\|\,.
\end{aligned}$$

Taking into account that

$$\begin{aligned}
&\left\|C(t)u - \frac{1}{2}\left(L_n^{k(n)}u + M_n^{k(n)}u\right)\right\| \\
&\qquad = \frac{1}{2}\left\|T_+(t)u + T_-(t)u - L_n^{k(n)}u - M_n^{k(n)}u\right\| \\
&\qquad \leq \frac{1}{2}\left(\left\|T_+(t)u - L_n^{k(n)}u\right\| + \left\|T_-(t)u - M_n^{k(n)}u\right\|\right)
\end{aligned}$$

the proof follows from the preceding inequalities. □

Remark 1.2. In many applications it is natural to consider the sequence $k(n) = [nt]$ for which $t_n = t$ and $|[nt]/n - t| = nt/n - [nt]/n \leq 1/n$. Hence estimate (1.3) yields

$$\begin{aligned}
&\left\| C(t)u - \frac{1}{2}\left(L_n^{[nt]}u + M_n^{[nt]}u\right)\right\| \qquad (1.4)\\
&\quad \le \frac{M}{2}\,\exp(\omega\, e^{\omega/n}\,t)\int_0^t \exp(-\omega e^{\omega/n}s)\times\\
&\quad \times\left(\|(n(L_n - I) - A)G(s)u\| + \|(n(M_n - I) + A)G(-s)u\|\right)ds\\
&\quad + \frac{M}{2}\left(\exp(\omega\, e^{\omega/n}\,t) + \sqrt{\frac{2nt}{\pi}}\,e^{\omega\,t} + \omega\,t\,\exp\left(\omega\, e^{\omega/n}\,t\right)\right)\times\\
&\qquad \times\left(\|L_n u - u\| + \|M_n u - u\|\right).
\end{aligned}$$

From the classical theory of the cosine functions (see [16] and [13, Chapter II] for more details) we have that the unique solution of the following second-order Cauchy problem

$$\begin{cases} \dfrac{\partial^2}{\partial t^2}u(t,x) = A^2u(t,x)\,, & t\in\mathbb{R}\,;\\ u(0,x) = u_0(x)\,, & x\in\mathbb{R}\,;\\ \dfrac{\partial}{\partial t}u(t,x)|_{t=0} = u_1(x)\,, & x\in\mathbb{R}\,, \end{cases} \qquad (1.5)$$

with $u_0,\ u_1\in D$, is given by

$$\begin{aligned}
u(t,x) &= C(t)u_0(x) + \int_0^t C(v)u_1(x)\,dv \qquad (1.6)\\
&= \frac{1}{2}\lim_{n\to\infty}\left(L_n^{[n\,t]}u_0 + M_n^{[n\,t]}u_0 + \int_0^t\left(L_n^{[n\,v]}u_1 + M_n^{[n\,v]}u_1\right)dv\right),
\end{aligned}$$

for every $t\in\mathbb{R}$ and $x\in\mathbb{R}$. We explicitly observe that the sequences $(L_n^{[n\,v]}u_1)_{n\ge1}$ and $(M_n^{[n\,v]}u_1)_{n\ge1}$ are equibounded with respect to $v\in[0,t]$ and this allows us to apply the Lebesgue dominated convergence theorem.

2. Rogosinski type operators

Denote by $C_{2\pi}$ the space of all 2π-periodic continuous real functions on $\mathbb{R}$ and put $\Pi := \{\pi + 2k\pi \mid k\in\mathbb{Z}\}$. Moreover, let $a\in C_{2\pi}\cap C^1(\mathbb{R}\setminus\Pi)$ be such that $a\ne0$ in $]-\pi,\pi[$ and consider the first-order differential operator $(A, D(A))$ defined by

$$Au := au', \qquad u\in D(A) := \left\{u\in C_{2\pi}\cap C^1(\mathbb{R}\setminus\Pi)\mid Au\in C_{2\pi}\right\}.$$

In order to consider the generation of cosine functions, we also consider the operator A^2 on the following domain

$$D(A^2) := \left\{u\in C_{2\pi}\cap C^2(]-\pi,\pi[)\mid a(au')'\in C_{2\pi}\right\}.$$

It is well-known (see e.g. [8, Theorem 1.1]) that $(A^2, D(A^2))$ generates a cosine functions $(C(t))_{t\in\mathbb{R}}$ in $C_{2\pi}$ if and only if

$$\frac{1}{a}\in L^1(-\pi,0)\,,\ \frac{1}{a}\in L^1(0,\pi)\,. \qquad (2.1)$$

Now, we consider the Rogosinski kernel defined by setting, for every $n \in \mathbb{N}$ and $x \in \mathbb{R}$,

$$r_n(x) := 1 + 2\sum_{k=1}^{n} \cos\left(\frac{k\pi}{2n+1}\right)\cos(kx)\ ,$$

and the corresponding n-th *Rogosinski operator* $R_n : C_{2\pi} \to C_{2\pi}$ given by

$$R_n f(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-v) r_n(v)\, dv\ , \qquad f \in C_{2\pi}\ ,\ x \in \mathbb{R}\ .$$

The n-th generalized Rogosinski operator $R_{a,n} : C_{2\pi} \to C_{2\pi}$ introduced in [8] is defined by putting

$$R_{a,n} f(x) = R_n f\left(x + \frac{2\pi}{2n+1}\, a(x)\right)\ , \qquad f \in C_{2\pi}\ , \quad x \in \mathbb{R}\ .$$

From [8, Theorem 2.1] the sequence $(\|R_{a,n}\|)_{n\in\mathbb{N}}$ is equibounded and moreover $\|R_{a,n}^k\| \leq 2\pi$ for every $n, k \geq 1$. Further, there exists a positive constant $C > 0$ such that

$$\|R_{a,n} f - f\| \leq C\ \omega\left(f; \frac{1}{n}\right)\ , \qquad f \in C_{2\pi}\ . \tag{2.2}$$

In order to apply Theorem 1.1, our next aim is to establish a quantitative estimate of the Voronovskaja-type formula associated with these operators.

Lemma 2.1. *Let $0 < \alpha \leq 1$. Then, for every $f \in C_{2\pi}^{1,\alpha}$,*

$$\left\|\frac{2n+1}{2\pi}\,(R_{a,n} f - f) - Af\right\|_{\infty} \leq 49(\|a\|_{\infty} + 1)\left(\frac{2\pi}{2n+1}\right)^{\alpha} L_{f'}\ ,$$

where $L_{f'}$ is the constant of α-hölderianity of f'.

Proof. For every $f \in C_{2\pi}^{1,\alpha}$ we have

$$\begin{aligned}\left\|\frac{2n+1}{2\pi}\,(R_{a,n} f - f) - Af\right\|_{\infty} &\leq \left\|\frac{2n+1}{2\pi}\,(R_{a,n} f - R_n f) - Af\right\|_{\infty} \\ &\quad + \left\|\frac{2n+1}{2\pi}\,(R_n f - f)\right\|_{\infty}\ .\end{aligned} \tag{2.3}$$

As regards to the first term at the right-hand side of (2.3), from Lagrange's theorem we can write

$$f(y+t) - f(y) = f'(y)t + (f'(\xi) - f'(y))\, t\ , \qquad y, t \in \mathbb{R}$$

where ξ lies between y and $y+t$. For every $x \in \mathbb{R}$ and $n \in \mathbb{N}$ we have

$$\begin{aligned}
&\frac{2n+1}{2\pi}(R_{a,n}f(x) - R_n f(x)) - a(x)f'(x) \\
&\quad = \frac{2n+1}{2\pi}\,\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(f\left(x - v + \frac{2\pi}{2n+1}\,a(x)\right) - f(x-v)\right)\,r_n(v)\,dv \\
&\qquad - a(x)f'(x) \\
&\quad = \frac{2n+1}{2\pi}\,\frac{1}{2\pi}\int_{-\pi}^{\pi} f'(x-v)\frac{2\pi}{2n+1}\,a(x)\,r_n(v)dv - a(x)f'(x) \\
&\qquad + \frac{2n+1}{2\pi}\,\frac{1}{2\pi}\int_{-\pi}^{\pi}(f'(\xi) - f'(x-v))\,\frac{2\pi}{2n+1}a(x)\,r_n(v)\,dv \\
&\quad = a(x)(R_n f'(x) - f'(x)) + a(x)\frac{1}{2\pi}\int_{-\pi}^{\pi}(f'(\xi) - f'(x-v))\;r_n(v)\,dv\;,
\end{aligned}$$

where $\xi \in]x - v, x - v + 2\pi a(x)/(2n+1)[$.

We recall that (see e.g. [5, Theorem 2.4.8, p. 106])

$$\|R_n g - g\|_\infty \le (2\pi+1)E_n(g) + 4\,\omega\left(g; \frac{1}{n}\right)\;, \qquad g \in C_{2\pi}\;,$$

where $E_n(g)$ is the best approximation of the function g by trigonometric polynomials of degree n and hence, from the classical Jackson's theorem,

$$\|R_n g - g\|_\infty \le 6(2\pi+1)\,\omega\left(g; \frac{1}{n}\right) + 4\,\omega\left(g; \frac{1}{n}\right) \le (12\pi+10)\omega\left(g; \frac{1}{n}\right)\;.$$

Applying the above inequality to f' and f we get

$$\begin{aligned}
&\left|\frac{2n+1}{2\pi}(R_{a,n}f(x) - R_n f(x)) - a(x)f'(x)\right| \\
&\qquad \le \|a\|_\infty\left((12\pi+10)\omega\left(f'; \frac{1}{n}\right) + \omega\left(f'; \frac{2\pi}{2n+1}\right)\right) \\
&\qquad \le \|a\|_\infty(12\pi+11)\,\omega\left(f'; \frac{2\pi}{2n+1}\right)\;,
\end{aligned}$$

and consequently

$$\begin{aligned}
\left\|\frac{2n+1}{2\pi}\,(R_{a,n}f - f) - Af\right\|_\infty \le{}& \|a\|_\infty(12\pi+11)\omega\left(f'; \frac{2\pi}{2n+1}\right) \\
&+ \frac{2n+1}{2\pi}(12\pi+10)\omega\left(f; \frac{1}{n}\right)\;.
\end{aligned}$$

Since $f \in C^{1,\alpha}_{2\pi}$ we have $\omega(f,\delta) \le \frac{L_{f'}}{2}\delta^{\alpha+1}$ and $\omega(f',\delta) \le L_{f'}\delta^{\alpha}$. Thus we conclude that

$$\begin{aligned}
&\left\| \frac{2n+1}{2\pi}\left(R_{a,n}f - f\right) - Af \right\|_\infty \\
&\qquad \le \|a\|_\infty (12\pi+11)\left(\frac{2\pi}{2n+1}\right)^\alpha L_{f'} + (12\pi+10)\frac{2n+1}{4n\pi}\frac{1}{n^\alpha}L_{f'} \\
&\qquad \le (12\pi+11)(\|a\|_\infty+1)\left(\frac{2\pi}{2n+1}\right)^\alpha L_{f'}.
\end{aligned}$$

□

From [8, Theorem 1.1] we already know that if $a \in C^{0,1}_{2\pi} \cap C^1(\mathbb{R}\setminus\Pi)$ satisfies condition (2.1), then the operator $(A, D(A))$ generates a C_0-semigroup $(T(t))_{t\ge 0}$ of positive contractions on $C_{2\pi}$. In the next lemma we state a more precise quantitative estimate of Voronovskaja's formula.

Lemma 2.2. *Let $a \in C^3_{2\pi}$ satisfy condition* (2.1) *and let $(T(t))_{t\ge0}$ be the C_0-semigroup on $C_{2\pi}$ generated by $(A, D(A))$.*

Then, for every $t \ge 0$ and $f \in C^{1,\alpha}_{2\pi}$,

$$\begin{aligned}
&\left\| \left(\frac{2n+1}{2\pi}\left(R_{a,n} - I\right) - A\right) T(t)f \right\|_\infty \qquad (2.4)\\
&\qquad \le 49(\|a\|_\infty+1)\left(\frac{2\pi}{2n+1}\right)^\alpha (C_1+t)\|a''\|_\infty e^{C_2\|a\|_{C^2}t}\|f\|_{C^{1,\alpha}},
\end{aligned}$$

where $K = C_2\|a\|_{C^2}$ and the constants C, C_1 and C_2 are independent of t and n.

Proof. Let us consider the flow $\phi : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ as the unique solution of this problem

$$\begin{cases} \dfrac{\partial \phi(t,x)}{\partial t} = a(\phi(t,x)), & \text{for every } t,\ x \in \mathbb{R} \\ \phi(0,x) = x, & \text{for every } x \in \mathbb{R}. \end{cases}$$

Now consider the C_0-semigroup $(T(t))_{t\ge0}$ defined by

$$T(t)f(x) := f(\phi(t,x)) \qquad \text{for all } t \ge 0,\ x \in \mathbb{R},\ f \in C_{2\pi}. \tag{2.5}$$

Notice that the operator $(A, D(A))$ is the generator of the semigroup defined in (2.5).

Since $a \in C^3_{2\pi}$, then $\phi \in C^3_{2\pi}$ (see [3, Theorem 10.3]), and hence for all $f \in C^m_{2\pi}$, we have that $T(t)f = f(\phi(t,\cdot)) \in C^m_{2\pi}$, $m = 0,1,2,3$.

Let us consider the following Cauchy problem

$$\begin{cases} \dfrac{\partial u}{\partial t} = Au & \text{in } (0,\infty)\times\mathbb{R}, \\ u(0,x) = f(x) & \text{in } \mathbb{R}, \end{cases} \tag{2.6}$$

with $f \in C^3_{2\pi}$. Let $u(t,x) = T(t)f(x)$ be the solution of the previous problem, we have $\|u(t,\cdot)\|_\infty = \|T(t)f\|_\infty \le \|f\|_\infty$ for every $t \ge 0$; moreover $u \in C^3_{2\pi}$ since $f \in C^3_{2\pi}$.

Consider also the problem

$$\begin{cases} \dfrac{\partial v}{\partial t} = Av + Bv & \text{in } (0,\infty)\times\mathbb{R}\,, \\ v(0,\cdot) = f' & \text{in } \mathbb{R}\,, \end{cases} \tag{2.7}$$

where $Bv := a'v$ for every $v \in D(A)$. Since $\|B\| = \|a'\|_\infty$, then the operator $A+B$ on $D(A)$ is a bounded perturbation of the operator $(A, D(A))$ and it generates a semigroup $(S(t))_{t\geq 0}$ on $C_{2\pi}$ such that

$$\|S(t)\| \leq e^{\|a'\|_\infty t}\,, \qquad t \geq 0$$

see [12, Chapter 3].

Let us notice that u' solves (2.7) on $[0,\infty)\times\mathbb{R}$, indeed

$$(Au)' = a'u' + au'' = Bu' + Au'\;;$$

then $u'(t,x) = S(t)f'(x)$ for every $t \geq 0$, $x \in \mathbb{R}$ and

$$\|T(t)\|_{\mathcal{L}(C^1;C^1)} \leq 1 + e^{\|a'\|_\infty t}\,, \qquad t \geq 0\,.$$

Now let us consider the problem

$$\begin{cases} \dfrac{\partial w(t,x)}{\partial t} = Aw + Cw + a''u'(t,0) & \text{in } (0,\infty)\times\mathbb{R}\,, \\ w(0,\cdot) = f'' & \text{in } \mathbb{R}\,, \end{cases} \tag{2.8}$$

where $Cw := a''\int_0^x \left(w(y) - \frac{1}{2\pi}\int_0^{2\pi} w\right) dy + 2a'w$ for every $w \in D(A)$. Then $A+C$ is a bounded perturbation of $(A, D(A))$ and hence $(A+C, D(A))$ generates the C_0-semigroup $\left(\tilde{S}(t)\right)_{t\geq 0}$ on $C_{2\pi}$. Since $\|C\| \leq 2\pi\|a''\|_\infty + 2\|a'\|_\infty \leq 2\pi\|a\|_{C^2}$, we have (see [12, Chapter 3])

$$\|\tilde{S}(t)\| \leq e^{2\pi\|a\|_{C^2}\,t}\,, \qquad t \geq 0\,.$$

Therefore

$$w(t,x) = \tilde{S}(t)f'' + \int_0^t \tilde{S}(t-s)\left(a''u'(t,0)\right) ds$$

is a mild solution of (2.8) in $C_{2\pi}$; moreover

$$\begin{aligned} \|w(t,\cdot)\|_\infty &\leq \|\tilde{S}(t)f''\|_\infty + \int_0^t \|\tilde{S}(t-s)\left(a''u'(s,0)\right)\| ds \\ &\leq e^{2\pi\|a\|_{C^2}\,t}\|f''\|_\infty + \int_0^t \|\tilde{S}(t-s)\left(a''(T(s)f)'(0)\right)\| ds \\ &\leq e^{2\pi\|a\|_{C^2}\,t}\|f''\|_\infty + \int_0^t e^{2\pi\|a\|_{C^2}\,(t-s)}\|a''(T(s)f)'(0)\| ds \\ &\leq e^{2\pi\|a\|_{C^2}\,t}\|f''\|_\infty + \int_0^t e^{2\pi\|a\|_{C^2}\,(t-s)}\|a''\|_\infty\|f'\|_\infty e^{\|a'\|_\infty s} ds \\ &\leq e^{2\pi\|a\|_{C^2}\,t}\|f''\|_\infty + t\|a''\|_\infty\|f'\|_\infty e^{2\pi\|a\|_{C^2}\,t}\,. \end{aligned}$$

Since $(Au)'' = a''u' + 2a'u'' + Au''$, we have

$$\begin{aligned}\frac{\partial}{\partial t}u''(t,x) &= Au''(t,x) + a''(x)u''(t,x) + 2a'(x)u''(x)\\ &= Au''(t,x) + a''(x)\left(\int_0^x u'(t,y)dy + u'(t,0)\right) + 2a'(x)u''(x)\\ &= Au''(t,x) + Cu''(t,x) + a''(x)u'(t,0)\ ,\end{aligned}$$

then $u''(t,x)$ is a solution of (2.8) and $u(t,\cdot) = w(t,\cdot)$. So we can apply the previous estimate to $u''(t,\cdot)$ and we get, for every $t \geq 0$,

$$\|(T(t)f)''\|_\infty = \|u''(t,\cdot)\|_\infty \leq e^{2\pi\|a\|_{C^2}t}\|f''\|_\infty + t\|a''\|_\infty\|f'\|_\infty e^{2\pi\|a\|_{C^2}t}\ .$$

Therefore

$$\|T(t)\|_{\mathcal{L}(C^2;C^2)} \leq 1 + e^{\|a'\|_\infty t} + e^{2\pi\|a\|_{C^2}t} + t\|a''\|_\infty e^{2\pi\|a\|_{C^2}t}\ , \qquad t \geq 0\ .$$

Finally, since $C^{1,\alpha}_{2\pi}$ is an intermediate space between $C^1_{2\pi}$ and $C^2_{2\pi}$, then we get

$$\begin{aligned}\|T(t)\|_{\mathcal{L}(C^{1,\alpha};C^{1,\alpha})} &\leq C\|T(t)\|^{1-\alpha}_{\mathcal{L}(C^1;C^1)}\|T(t)\|^{\alpha}_{\mathcal{L}(C^2;C^2)}\\ &\leq (C_1+t)\|a''\|_\infty e^{C_2\|a\|_{C^2}\,t} \qquad \text{for all } t \geq 0\ ,\end{aligned} \tag{2.9}$$

where C_1 and C_2 are positive constant independent of t.

Finally from Lemma 2.1 and taking into account (2.9), we get

$$\begin{aligned}&\left\|\left(\frac{2n+1}{2\pi}(R_{a,n}-I) - A\right)T(t)f\right\|_\infty \\ &\quad\leq 49(\|a\|_\infty+1)\left(\frac{2\pi}{2n+1}\right)^\alpha L_{(T(t)f)'}\\ &\quad\leq 49(\|a\|_\infty+1)\left(\frac{2\pi}{2n+1}\right)^\alpha \|T(t)f\|_{C^{1,\alpha}}\\ &\quad\leq 49(\|a\|_\infty+1)\left(\frac{2\pi}{2n+1}\right)^\alpha \|T(t)\|_{\mathcal{L}(C^{1,\alpha};C^{1,\alpha})}\|f\|_{C^{1,\alpha}}\\ &\quad\leq 49(\|a\|_\infty+1)\left(\frac{2\pi}{2n+1}\right)^\alpha (C_1+t)\|a''\|_\infty e^{C_2\|a\|_{C^2}t}\|f\|_{C^{1,\alpha}}\ ,\end{aligned} \tag{2.10}$$

for all $t \geq 0$. □

In [8, Theorem 2.7] Campiti and Ruggeri established that besides the generation of the cosine function $(C(t))_{t\in\mathbb{R}}$, condition (2.1) also ensures that $C^1_{2\pi} \cap D(A^2)$ is a core for $(A^2, D(A^2))$ and further, for every $t > 0$,

$$C(t) = \frac{1}{2}\lim_{n\to\infty}\left(R^{k(n)}_{a,n}u + R^{k(n)}_{-a,n}u\right)\ , \tag{2.11}$$

where $(k(n))_{n\geq 1}$ is a sequence of positive integers such that

$$\lim_{n\to+\infty}\frac{2\pi k(n)}{2n+1} = t.$$

From (2.2) it follows that there exits a constant $C > 0$ such that

$$\|R_{a,n}f - f\|_\infty \le \frac{C}{n^{\alpha+1}} L_{f'}$$

for all $f \in C^{1,\alpha}_{2\pi}$ and the same estimate also holds for $R_{-a,n}$. We obtain the following quantitative version of (2.11).

Theorem 2.3. *Let $a \in C^3_{2\pi}$ satisfy* (2.1)*. Then for every $t \ge 0$ and $u \in C^{1,\alpha} \cap D(A^2)$*

$$\left\| C(t)u - \frac{1}{2}\left(R^{k(n)}_{a,n}u + R^{k(n)}_{-a,n}u\right)\right\|_\infty \tag{2.12}$$

$$\le 2\pi\, C \left(\frac{2\pi}{2n+1}\right)^\alpha \left(\|a''\|_\infty \left(\|a\|_\infty + 1\right)\left[\frac{e^{Kt}-1}{K}\left(C_1 + \frac{1}{K}\right) + t\frac{e^{Kt}}{K}\right]\right. \tag{2.13}$$

$$\left. + \left|\frac{2\pi k(n)}{2n+1} - t\right| + \sqrt{\frac{2}{n}}\sqrt{\frac{2k(n)}{2n+1}}\right)\|u\|_{C^{1,\alpha}}\ ,$$

where $K = C_2\|a\|_{C^2}$, $(k(n))_{n\ge1}$ is a sequence of positive integers such that

$$\lim_{n\to+\infty}\frac{2\pi k(n)}{2n+1} = t$$

and C, C_1 and C_2 are positive constants independent of $n \in \mathbb{N}$ and $t \ge 0$.

Proof. Consider $u \in C^{1,\alpha} \cap D(A^2)$, taking into account (2.4) we have

$$\int_0^t \left\|\left(\frac{2n+1}{2\pi}\left(R_{a,n} - I\right) - A\right)T(s)u\right\|_\infty ds \tag{2.14}$$

$$\le 49\|a''\|_\infty(\|a\|_\infty + 1)\left(\frac{2\pi}{2n+1}\right)^\alpha \|u\|_{C^{1,\alpha}}\left[\frac{e^{Kt}-1}{K}\left(C_1 + \frac{1}{K}\right) + t\frac{e^{Kt}}{K}\right]\ ,$$

where $K = C_2\|a\|_{C^2}$. The same estimate also holds for the sequences of operators $(R_{-a,n})_{n\ge1}$ and the differential operator $-A$. Then from (1.3) we have

$$\left\| C(t)u - \frac{1}{2}\left(R^{k(n)}_{a,n}u + R^{k(n)}_{-a,n}u\right)\right\|_\infty \tag{2.15}$$

$$\le 2\pi C\|a''\|_\infty \left(\|a\|_\infty + 1\right)\left(\frac{2\pi}{2n+1}\right)^\alpha \left[\frac{e^{Kt}-1}{K}\left(C_1 + \frac{1}{K}\right) + t\frac{e^{Kt}}{K}\right]\|u\|_{C^{1,\alpha}}$$

$$+ 2\pi\left(\left|k(n) - \frac{2n+1}{2\pi}t\right| + \sqrt{\frac{2k(n)}{2\pi}}\right) C\left(\frac{1}{n}\right)^{\alpha+1}\|u\|_{C^{1,\alpha}}\ .$$

□

Finally, we observe that arguing as in [11] we can also establish a quantitative estimate of the resolvent operators.

Exactly the same procedure can be also applied to other sequences of trigonometric polynomials such as Fejér operators and more general averages of trigonometric interpolating operator considered in [8, 6]. Since in these cases the cosine function is the same, we limit ourselves to observe that (2.12) remains still valid when considering these other sequences of trigonometric interpolating operators too.

References

[1] Albanese, A., Campiti, M., Mangino, E., *Approximation formulas for C_0-semigroups and their resolvent operators,* J. Applied Funct. Anal., **1**(2006), no. 3, 343-358.

[2] Altomare, F., Campiti, M., *Korovkin-type approximation theory and its applications,* de Gruyer Studies in Mathematics **17**, W. de Gruyer, Berlin, New York, 1994.

[3] Amann, H., *Ordinary differential equations - An introduction to nonlinear analysis,* de Gruyer Studies in Mathematics **13** W. de Gruyer, Berlin, New York, 1990.

[4] Arendt, W., Batty, C., Hieder, M., Neubrander, F., *Vector-valued Laplace transforms and Cauchy problems,* Monographs in Mathematics **96** Birkhäuser-Verlag, Basel, 2001.

[5] Butzer, P.L., Nessel, R.J., *Fourier analysis and approximation I,* Academic Press, New York-London, 1971.

[6] Campiti, M., Metafune, G., Pallara, D. *On some averages of trigonometric interpolating operators,* in "Approximation Theory, Wavelets and Applications", (S. P. Singh ed.), Acquafredda di Maratea, 16-26 maggio 1994, Kluwer Academic Publishers, 1995, 303-313.

[7] Campiti, M., Ruggeri, S.P., *C_0-semigroups in spaces of 2π-periodic continuous functions generated by a second-order differential operator,* Internat. J. Pure Appl. Math., **7**(2003), no. 1, 27-48.

[8] Campiti, M., Ruggeri, S.P., *Approximation of semigroups and cosine functions in spaces of periodic functions,* Applicable Analysis, **86**(2007), no. 2, 167-186.

[9] Campiti, M., Tacelli, C., *Rate of convergence in the Trotter's approximation theorem,* Constr. Approx., **28**(2008), no. 3, 333-341. *Erratum to: Rate of convergence in Trotter's approximation theorem,* Constr. Approx., **31**(2010), no. 3, 459-462.

[10] Campiti, M., Tacelli, C., *Trotter's approximation of semigroups and order of convergence in $C^{2,\alpha}$-spaces,* J. Approx. Theory, **162**(2010), 2303-2316.

[11] Campiti, M., Tacelli, C., *Approximation processes for resolvent operators,* Calcolo (Springer), **45**(2008), no. 4, 235-245.

[12] Engel, K.-J., Nagel, R., *One-parameter semigroups for linear evolution equations,* Graduate Texts in Mathematics, Springer, 1999.

[13] Fattorini, H.O., *Second order linear differential equations in Banach spaces,* North-Holland **108** North-Holland, Amsterdam, New York, Oxford, 1985.

[14] Gonska, H., Raşa, I., *The limiting semigroup of the Bernstein iterates: degree of convergence,* Acta Math. Hungar., **111**(2006), 111-122.

[15] Mangino, E., Raşa, I., *A quantitative version of Trotter's theorem,* J. Approx. Theory, **146**(2007), no. 2, 149-156.

[16] Sova, M., *Cosine operator functions,* Rozprawy Matematyczne, **49**(1966), 1-47.

[17] Trotter, H. F., *Approximation of semigroups of operators,* Pacific J. Math., **8**(1958), 887-919.

Michele Campiti
University of Salento
Department of Mathematics
P.O.Box 193
73100 Lecce
Italy
e-mail: `michele.campiti@unisalento.it`

Giusy Mazzone
University of Salento
Department of Mathematics
P.O.Box 193
73100 Lecce
Italy
e-mail: `giusy.mazzone@unisalento.it`

Cristian Tacelli
University of Salento
Department of Mathematics
P.O.Box 193
73100 Lecce
Italy
e-mail: `cristian.tacelli@unisalento.it`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 273–293

Remarks on the state of the art of a posteriori error control of elliptic PDEs in energy norms in practise

Carsten Carstensen and Christian Merdon

Abstract. Five classes of up to 9 a posteriori error estimators compete in three second-order model problems, namely the conforming and nonconforming first-order approximation of the Poisson-Problem plus some conforming obstacle problem. Our numerical results provide sufficient evidence that guaranteed error control in the energy norm is indeed possible with efficiency indices between one and three. The five classes of error estimator consist of the standard residual-based error estimators, averaging error estimators, equilibration error estimators, e.g. the ones of Braess or Luce and Wohlmuth, least-square error estimators and the localisation error estimator of Carstensen and Funken. For the error control for obstacle problems, Braess considers Lagrange multipliers and some resulting auxiliary equation to view the a posteriori error control of the error in the obstacle problem as computable terms plus errors and residuals in the auxiliary equation. Hence all the former a posteriori error estimators apply to this benchmark as well and lead to surprisingly accurate guaranteed upper error bounds. This approach allows an extension to more general boundary conditions and a discussion of efficiency for the affine benchmark examples. The Luce-Wohlmuth and the least-square error estimators win the competition in several computational benchmark problems. Novel equilibration of nonconsistency residuals and novel conforming averaging error estimators win the competition for Crouzeix-Raviart nonconforming finite element methods. Furthermore, accurate error control is slightly more expensive but pays off in all applications under consideration while adaptive mesh-refinement is sufficiently pleasant as accurate when based on explicit residual-based error estimates.

Mathematics Subject Classification (2010): 65N30, 65R20, 73C50.

Keywords: A posteriori error estimators, finite element methods.

This work was supported by DFG Research Center MATHEON and by the World Class University (WCU) program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology R31-2008-000-10049-0. The second author was also supported by the German Academic Exchange Service (DAAD, D/10/44641) during his stay at the Yonsei University in 2010.

1. Introduction

A posteriori finite element error control for second-order elliptic boundary value problems involves the computation of guaranteed upper bounds of some residual Res in the dual $H^{-1}(\Omega)$ of $H^1_0(\Omega)$ with respect to the dual norm. The majority of applications in computational PDEs [19, 20] applies to the residual

$$\mathrm{Res}(v) = \int_\Omega (fv - \sigma_h \cdot Dv)\, dx$$

with some given Lebesgue integrable functions f and σ_h. Traditional equilibration techniques compute some $q \in H(\mathrm{div}, \Omega)$ such that (via an integration by parts) the residual becomes

$$\mathrm{Res}(v) = \int_\Omega (f + \mathrm{div}\, q)v\, dx + \int_\Omega (q - \sigma_h) \cdot Dv\, dx$$

and leads to the error estimate

$$|||\mathrm{Res}|||_\star := \sup_{v\in H^1_0(\Omega)} \mathrm{Res}(v)/|||v||| \leq \eta(q) := |||f + \mathrm{div}\, q|||_\star + \|q - \sigma_h\|_{L^2(\Omega)}.$$

This paper concentrates on three model problems to support the observation of published and ongoing error estimator competitions [11, 22, 24, 23] that accurate error control is possible with efficiency between 1 and 2. Section 2 introduces the setting for the Poisson model problem and Section 3 recalls the five classes of error estimators from Table 1 to control $|||\mathrm{Res}|||_\star$.

Table 1. Classes of a posteriori error estimators used in this paper.

No	Classes of error estimators	Class representatives
1	explicit residual-based	η_{R}
2	averaging	$\eta_{\mathrm{MP1}}, \eta_{\mathrm{A1}}$
3	equilibration	η_{B}, η_{MFEM}, η_{LW}, η_{EQL}
4	least-square	η_{LS}
5	localisation	η_{CF}

Subsection 4.1 explains our adaptive mesh-refinement algorithm. In this paper the adaptive mesh-refinement is driven by local error estimator contributions from any estimator from Table 1 to observe that mesh refinement with the standard residual-based error estimator η_{R} is suitable and does not need to be replaced by any other marking strategy.

Section 5 deals with nonconforming Crouzeix-Raviart approximations u_{CR} for the Poisson model problem. The Helmholtz decomposition allows a split of the error in the broken energy norm into

$$|||e|||^2_{\mathrm{NC}} \leq \eta^2 + |||\mathrm{Res}_{\mathrm{NC}}|||^2_\star.$$

The first term η on the right-hand side involves contributions of the right-hand side f and is directly computable (up to quadrature errors). The second term in the upper error bound is the dual norm of some residual $\mathrm{Res}_{\mathrm{NC}}$ that enjoys Galerkin orthogonality properties,

$$|||\mathrm{Res}_{\mathrm{NC}}|||_\star = \min_{\substack{v \in H^1(\Omega) \\ v = u_D \text{ on } \partial\Omega}} \|\nabla_{\mathrm{NC}} u_{\mathrm{CR}} - \nabla v\|_{L^2(\Omega)}.$$

Upper bounds of $|||\mathrm{Res}_{\mathrm{NC}}|||_\star$ are computed by the error estimators of Table 1 or by the design of some $v \in H^1(\Omega)$ with Dirichlet data $v = u_D$ along $\partial\Omega$.

Section 6 extends applications to obstacle problems with affine obstacles by introduction of some auxiliary Poisson problem after [12].

2. Model Poisson problem

This section specifies the setting in the Poisson model problem.

2.1. Discrete problem

Given a bounded Lipschitz domain Ω and right-hand side $f \in L^2(\Omega)$, the Poisson model problem seeks the exact solution $u \in H^1(\Omega)$ with $u = 0$ along $\partial\Omega$ and

$$-\Delta u = f \text{ in } \Omega.$$

Given a regular triangulation $\mathcal{T}$ of $\Omega \subseteq \mathbb{R}^2$ into triangles with edges $\mathcal{E}$, nodes $\mathcal{N}$, and free nodes $\mathcal{K}$, let $P_k(T)$ denote the polynomials of degree $\leq k$ on $T \in \mathcal{T}$ and

$$P_k(\mathcal{T}) := \{v_h \in L^2(\Omega) \mid \forall T \in \mathcal{T},\ v_h|_T \in P_k(T)\}.$$

The first-order Courant finite element method computes the discrete solution $u_h \in V(\mathcal{T}) := P_1(\mathcal{T}) \cap C_0(\Omega)$ with gradient $\sigma_h := \nabla u_h$ as

$$\int_\Omega \nabla u_h \cdot \nabla v_h \, dx = \int_\Omega f v_h \, dx \quad \text{for all } v_h \in V(\mathcal{T}). \tag{2.1}$$

2.2. Residual

The related residual $\mathrm{Res} \in V^\star$ is a linear and bounded functional

$$\mathrm{Res}(v) := \int_\Omega f v \, dx - \int_\Omega \sigma_h \cdot \nabla v \, dx$$

for the Sobolev functions v in the Hilbert space $V := H^1_0(\Omega)$ endowed with the (semi-) norm $|||\cdot||| := \|\nabla \cdot\|_{L^2(\Omega)}$. It is clear from the Riesz representation theorem that the energy norm $|||e|||$ of the error $e := u - u_h$ equals the norm of $|||\mathrm{Res}|||_\star$ of the residual Res (cf., e.g., [15, Section 5.1.2, p. 86] and [20, Section 3.3]). A posteriori equilibration error estimators derive computable upper bounds of $|||\mathrm{Res}|||_\star$ through the introduction of some equilibrated $q \in H(\mathrm{div}, \Omega)$. An integration by parts shows

$$\mathrm{Res}(v) = \int_\Omega (f + \mathrm{div}\, q) v \, dx + \int_\Omega (q - \sigma_h) \cdot \nabla v \, dx$$

and therefore leads to

$$|||\mathrm{Res}|||_\star \leq |||f + \mathrm{div}\, q|||_\star + \|q - \sigma_h\|_{L^2(\Omega)}.$$

The equilibration error estimator of Braess [13, 26] is one modern example for a proper choice of q in $RT_0(\mathcal{T}) \subseteq H(\mathrm{div}, \Omega)$,

$$RT_0(\mathcal{T}) := \Big\{ q(x) = a_{\mathcal{T}} x + \big(b_{\mathcal{T}}, c_{\mathcal{T}}\big) \in H(\mathrm{div}, \Omega) \mid a_{\mathcal{T}}, b_{\mathcal{T}}, c_{\mathcal{T}} \in P_0(\mathcal{T}) \Big\}.$$

Earlier examples of Ladeveze [29, 3] and [21] also provide a source of a posteriori error estimators compared in [11, 22]. If the local problems therein are solved exactly, they also yield guaranteed upper bounds. It is unrealistic to assume an exact solve of those local problems in practise and so the displayed numbers in [21, 11, 22] are only lower bounds for the guaranteed upper bounds. This fundamental difficulty is circumvented by modern equilibration error estimators, like the ones of Braess and Luce-Wohlmuth.

2.3. Inhomogenous Dirichlet boundary conditions

In case of inhomogenous boundary conditions $u = u_D$ along the boundary edges $\mathcal{E}(\partial\Omega) := \{E \in \mathcal{E} \mid E \subset \partial\Omega\}$, the discrete solution u_h satisfies $u_h = \mathcal{I}u_D := \sum_{z \in \mathcal{N}} u_D(z)\varphi_z$. Since $e = u - u_h = u_D - \mathcal{I}u_D \notin H^1_0(\Omega)$, the equation $|||e||| = |||\mathrm{Res}|||_\star$ does not hold.

Theorem 2.1. *Assume that $u_D \in H^1(\Omega) \cap C(\Omega)$ satisfies $u_D \in H^2(E)$ for all $E \in \mathcal{E}(\partial\Omega)$. Let $\partial^2_{\mathcal{E}} u_D/\partial s^2$ denote the edgewise second partial derivative of u_D along $\partial\Omega$. Then there exists $w_D \in H^1(\Omega)$ and some constant $C_\gamma \lesssim 1$ (which depends only on the interior angles of $\mathcal{T}$) with*

$$\begin{aligned} w_D|_{\partial\Omega} &= u_D|_{\partial\Omega} - \mathcal{I}u_D|_{\partial\Omega}, \\ \mathrm{supp}(w_D) &\subset \bigcup \{T \in \mathcal{T} \mid T \cap \partial\Omega \neq \varnothing\}, \\ \|w_D\|_{L^\infty(\Omega)} &= \|u_D - \mathcal{I}u_D\|_{L^\infty(\partial\Omega)}, \\ |||w_D||| &\leq C_\gamma \|h_{\mathcal{E}}^{3/2} \partial^2_{\mathcal{E}} u_D/\partial s^2\|_{L^2(\partial\Omega)}. \end{aligned}$$

Furthermore it holds

$$|||e|||^2 \leq |||\mathrm{Res}|||_\star^2 + |||w_D|||^2.$$

Proof. For the proof of the existence of w_D see [9]. For the proof of the last equation, assume the optimal $w \in H^1(\Omega)$ with $w|_{\partial\Omega} = u|_{\partial\Omega} - \mathcal{I}u|_{\partial\Omega}$ and $\mathrm{div}\,\nabla w \equiv 0$. Then, it holds the orthogonality from [9],

$$|||e|||^2 = |||e - w|||^2 + |||w|||^2 \leq |||\mathrm{Res}|||_\star^2 + |||w|||^2 \leq |||\mathrm{Res}|||_\star^2 + |||w_D|||^2.$$

This concludes the proof. □

Remark 2.2. More explicit calculations in [24] show $C_\gamma \leq 0.7043$ for triangulations with right isosceles triangles. However, for the numerical examples in this paper, we use $C_\gamma = 1$.

3. Five types of a posteriori error estimators

This section recalls some representatives of the five classes of error estimators from Table 1.

3.1. Notation

Consider a regular triangulation $\mathcal{T}$ of $\Omega \subseteq \mathbb{R}^2$ into triangles with nodes $\mathcal{N}$, free nodes $\mathcal{K} := \mathcal{N} \setminus \partial\Omega$, edges $\mathcal{E}$, Dirichlet boundary edges $\mathcal{E}(\partial\Omega) := \{E \in \mathcal{E} \mid E \subseteq \partial\Omega\}$. Each node z in $\mathcal{N}$ is associated with its nodal basis functions φ_z and node patch $\omega_z := \{\varphi_z > 0\}$ with diameter $h_z := \operatorname{diam}(\omega_z)$. Each triangle $T \in \mathcal{T}$ is the closed convex hull of the set $\mathcal{N}(T)$ of its vertices and associated to its element patch $\omega_T := \bigcup_{z \in \mathcal{N}(T)} \omega_z$. The set $\mathcal{E}(T)$ denotes the edges of T in $\mathcal{T}$ and the set $\mathcal{E}(z)$ denotes all edges connected to $z \in \mathcal{N}$.

3.2. Standard residual error estimator

The standard residual error estimator

$$\eta_R := \|h_{\mathcal{T}} f\|_{L^2(\Omega)} + \left(\sum_{E \in \mathcal{E}} h_E \| [\sigma_h \cdot \nu_E]_E \|^2_{L^2(E)} \right)^{1/2}$$

is a guaranteed upper bound of $|||u - u_h|||$. In all our examples, $\mathcal{T}$ consists of right isosceles triangles and then the generic reliability constant is even 1, i.e. $|||u - u_h||| \le \eta_R$ [21]. Here, $[\sigma_h \cdot \nu_E]_E$ denotes the jump of $[\sigma_h \cdot \nu_E]_E$ across $E \in \mathcal{E}$, which is set to zero along any Dirichlet edge $E \in \mathcal{E}(\partial\Omega)$.

3.3. Minimal $P_1(\mathcal{T}; \mathbb{R}^2)$ averaging

The error estimator

$$\eta_{\mathrm{MP1}} := \min_{q \in P_1(\mathcal{T};\mathbb{R}^2) \cap C(\Omega;\mathbb{R}^2)} \|\sigma_h - q\|_{L^2(\Omega)}$$

shows very accurate results for the Laplace equation, but only yields an upper bound for $|||u - u_h|||$ up to some reliability constant C_{rel} [18], which is *not* displayed and expected to be too large to be competitive. Simple averagings $q_A \in P_1(\mathcal{T}; \mathbb{R}^2)$ compute approximations of η_{MP1}, e.g.

$$\eta_{\mathrm{A1}} := \|\sigma_h - q_{\mathrm{A1}}\|_{L^2(\Omega)} \quad \text{with} \quad q_{\mathrm{A1}}(z) = \int_{\omega_z} \sigma_h \, dx / |\omega_z| \quad \text{for all } z \in \mathcal{N}.$$

3.4. Least-square error estimator

An integration by parts yields, for any $q \in H(\operatorname{div}, \Omega)$ and with elementwise integral mean $f_{\mathcal{T}} \in P_0(\mathcal{T})$, that

$$\begin{aligned}\int_\Omega (\nabla u - \sigma_h) \cdot \nabla v \, dx \\ = \int_\Omega (f - f_{\mathcal{T}}) v \, dx + \int_\Omega (f_{\mathcal{T}} + \operatorname{div} q) v \, dx + \int_\Omega (\sigma_h - q) \cdot \nabla v \, dx.\end{aligned}$$

After [33, 35, 22], this results in the error estimator

$$\eta_{\mathrm{LS}} := \min_{q \in RT_0(\mathcal{T})} C_F \|f_{\mathcal{T}} + \operatorname{div} q\|_{L^2(\Omega)} + \|\sigma_h - q\|_{L^2(\Omega)} + \operatorname{osc}(f, \mathcal{T})/\pi$$

with Friedrichs' constant $C_F := \sup_{v \in V \setminus \{0\}} \|v\|_{L^2(\Omega)} / |||v|||$, and oscillations

$$\operatorname{osc}(f, \mathcal{T}) := \|h_{\mathcal{T}}(f - f_{\mathcal{T}})\|_{L^2(\Omega)}.$$

Our interpretation of Repin's variant (without the oscillation split) reads

$$\eta_{\mathrm{REPIN}} := \min_{q \in RT_0(\mathcal{T})} C_F \|f + \operatorname{div} q\|_{L^2(\Omega)} + \|\sigma_h - q\|_{L^2(\Omega)}.$$

This paper studies the least-square variant η_{LS} rather than Repin's majorant η_{REPIN} for reasons discussed in [22, Subsection 4.2]. Supercloseness results from [14] show that asymptotically the minimiser q_{LS} equals the gradient q_{MFEM} of the mixed finite element method with lowest-order Raviart-Thomas finite elements $RT_0(\mathcal{T})$. In practise η_{LS} is approximated by a series of least-square problems as in [37].

3.5. Luce-Wohlmuth error estimator

Luce and Wohlmuth [30] suggest to solve local problems around each node on the dual triangulation $\mathcal{T}^\star$ of $\mathcal{T}$ and compute some equilibrated quantity q_{LW}. The dual triangulation $\mathcal{T}^\star$ connects each triangle center $\mathrm{mid}(T)$, $T \in \mathcal{T}$, with the edge midpoints $\mathrm{mid}(\mathcal{E}(T))$ and nodes $\mathcal{N}(T)$ and so divides each triangle $T \in \mathcal{T}$ into 6 subtriangles of area $|T|/6$.

Consider some node $z \in \mathcal{N}(\mathcal{T})$ and its nodal basis function $\varphi_z^\star$ with the fine patch $\omega_z^\star := \{\varphi_z^\star > 0\}$ of the dual triangulation $\mathcal{T}^\star$ and its neighbouring triangles $\mathcal{T}^\star(z) := \{T^\star \in \mathcal{T}^\star \mid z \in \mathcal{N}^\star(T)\}$. Since $\sigma_h \in P_0(\mathcal{T})$ is continuous along $\partial\omega_z^\star \cap T$ for any $T \in \mathcal{T}$, $q \cdot \nu = \sigma_h \cdot \nu \in P_0(\mathcal{E}^\star(\partial\omega_z^\star))$ is well-defined on the boundary edges $\mathcal{E}^\star(\partial\omega_z^\star)$ of $\omega_z^\star$. With $f_{T,z} := -\int_T f\varphi_z\, dx/|T^\star|$ and the local spaces

$$\begin{aligned} Q(\mathcal{T}^\star(z)) := \big\{\tau_h \in RT_0(\mathcal{T}^\star(z)) \mid \operatorname{div}\tau_h|_{T^\star} + f_{T,z} = 0 \text{ on } T^\star \in \mathcal{T}^\star \text{ with}\\ \mathcal{N}^\star(T^\star) \cap \mathcal{N}(T) = \{z\} \text{ and } q \cdot \nu = \sigma_h \cdot \nu \text{ along } \partial\omega_z^\star \setminus \partial\Omega\big\}, \end{aligned}$$

the mixed finite element method solves

$$q|_{\omega_z^\star} := \operatorname*{argmin}_{\tau_h \in Q(\mathcal{T}^\star(z))} \|q_h - \tau_h\|_{L^2(\omega_z^\star)}.$$

This choice of the divergence [25] differs from the original one of [30] for an improved bound for $|||f + \operatorname{div} q_{\mathrm{LW}}|||_\star$ with explicitly known constants, namely

$$|||f + \operatorname{div} q_{\mathrm{LW}}|||_\star \le \|h_{\mathcal{T}}(f + \operatorname{div} q_{\mathrm{LW}})\|_{L^2(\Omega)}/\pi.$$

For details cf. [25]. The remaining degrees of freedom permit proper boundary fluxes and

$$\int_\Omega q_{\mathrm{LW}} \cdot \mathrm{Curl}\varphi_z^\star\, dx = \int_\Omega \sigma_h \cdot \mathrm{Curl}\varphi_z^\star\, dx \quad \text{for all } z \in \mathcal{N}.$$

Here, Curl denotes the rotated gradient $\mathrm{Curl}v := (-\partial v/\partial x_2, \partial v/\partial x_1)$. Then, the Luce-Wohlmuth error estimator reads

$$\eta_{\mathrm{LW}} := \|\sigma_h - q_{\mathrm{LW}}\|_{L^2(\Omega)} + \|h_{\mathcal{T}}(f + \operatorname{div} q_{\mathrm{LW}})\|_{L^2(\Omega)}/\pi.$$

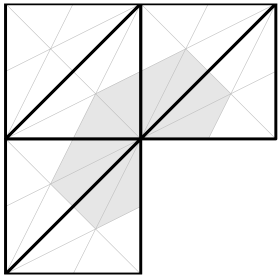


FIGURE 1. Triangulation $\mathcal{T}$ (thick lines), fine triangulation $\mathcal{T}^\star$ (thin lines) and $\omega_z^\star$ (lightgray) around the reentering corner of the L-shaped domain for the Luce-Wohlmuth error estimator.

3.6. Equilibration error estimator by Braess

Braess [13, 26] designs patchwise broken Raviart-Thomas functions $r_z \in RT_{-1}(\mathcal{T}(z))$ that satisfy

$$\begin{aligned} \operatorname{div} r_z|_T &= -\int_T f\varphi_z\,dx/|T| && \text{for } T \in \mathcal{T}(z)\\ [r_z\cdot\nu_E]_E &= -[\sigma_h\cdot\nu_E]_E/2 && \text{on } E \in \mathcal{E}(z)\cap\mathcal{E}(\partial\Omega)\\ r_z\cdot\nu &= 0 && \text{along } \partial\omega_z\setminus\mathcal{E}(\partial\Omega). \end{aligned}$$

The solution r_z of these problems is unique up to multiplicatives of $\operatorname{Curl}\varphi_z$ and may be chosen such that $\|r_z\|_{L^2(\omega_z)}$ is minimal. Eventually, the quantity $q_{\mathrm{B}} := \sigma_h + \sum_{z\in\mathcal{N}} r_z \in RT_0(\mathcal{T})$ satisfies

$$\operatorname{div} q_{\mathrm{B}}|_T = -\int_T f\,dx/|T|.$$

and allows the dual norm estimate

$$|||f + \operatorname{div} q_{\mathrm{B}}|||_\star \le \operatorname{osc}(f,\mathcal{T})/\pi.$$

The estimator reads

$$\eta_{\mathrm{B}} := \|\sigma_h - q_{\mathrm{B}}\|_{L^2(\Omega)} + \operatorname{osc}(f,\mathcal{T})/\pi.$$

3.7. Equilibration error estimator by Ladeveze

The fluxes q_L designed by Ladeveze-Leguillon [29] act as Neumann boundary conditions for local problems on each triangle, cf. also [3] for details. Given the local function space

$$H^1_D(T) = \begin{cases} H^1(T)/\mathbb{R} & \text{if } |T\cap\Gamma_D| = 0 \text{ and else}\\ \{v\in H^1(T) \mid v = 0 \text{ on } \partial T\cap\Gamma_D\}, \end{cases}$$

seek $\phi_T \in H^1_D(T)$ such that, for all $v \in H^1_D(T)$,

$$\int_T \phi_T \cdot \nabla v\, dx = \int_T fv\, dx - \int_T \sigma_h \cdot \nabla v\, dx + \int_{\partial T} q_{\mathrm{L}} \cdot \nu_T\, v\, ds\,.$$

Then the error estimate reads

$$|||u - u_h||| \leq \eta_{\mathrm{EQL}} := \left(\sum_{T \in \mathcal{T}} \|\nabla \phi_T\|^2_{L^2(T)} \right)^{1/2}.$$

3.8. Carstensen-Funken error estimator

The partition of unity property of the nodal basis functions $(\varphi_z \mid z \in \mathcal{N})$ leads in [21] to the solution of local problems on node patches ω_z: Seek $w_z \in W_z := \{v \in H^1_{\mathrm{loc}}(\omega_z) \mid \|\varphi_z^{1/2} \nabla v\|_{L^2(\omega_z)} < \infty, v = 0 \text{ on } \partial\Omega \cap \partial\omega_z\}$ if $z \in \mathcal{N}(\partial\Omega)$, or $w_z \in W_z := \{v \in H^1_{\mathrm{loc}}(\omega_z) \mid \|\varphi_z^{1/2} \nabla v\|_{L^2(\omega_z)} < \infty\}/\mathbb{R}$ if $z \in \mathcal{N}(\Omega)$, such that

$$\int_{\omega_z} \varphi_z \nabla w_z \cdot \nabla v\, dx = \int_{\omega_z} \varphi_z f v\, dx - \int_{\omega_z} \sigma_h \cdot \nabla(\varphi_z v)\, dx \quad \text{for all } v \in W_z.$$

Then the error estimator reads

$$|||u - u_h||| \leq \eta_{\mathrm{CF}} := \left(\sum_{z \in \mathcal{N}} \|\varphi_z^{1/2} \nabla w_z\|^2_{L^2(\omega_z)} \right)^{1/2}.$$

In the computations for η_{CF} and η_{EQL}, all the local problems are solved with fourth-order polynomials for simplicity. The computed values are regarded as very good approximations. However, strictly speaking the values displayed for η_{EQL} or η_{CF} are lower bounds of the guaranteed upper bounds.

4. Conforming finite element method

4.1. Uniform and adaptive mesh refinement

Automatic mesh refinement generates a sequence of meshes $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2$... by successive mesh refining using local refinement indicators derived from some η_{xyz} from Section 3.

Algorithm 4.1. *INPUT* coarse mesh $\mathcal{T}_0$. For any level $\ell = 0, 1, 2, \ldots$ do *COMPUTE* discrete solution u_ℓ on $\mathcal{T}_\ell$ with ndof $:= |\mathcal{N}_\ell(\Omega)|$ degrees of freedom, error estimator η_{xyz}, efficiency indices $EI := \eta_{\mathrm{xyz}}(k)/|||e|||$, and refinement indicators

$$\eta_\ell(T)^2 = \eta_{\mathrm{xyz}}(T)^2 + \|h_{\mathcal{E}}^{3/2} \partial^2_{\mathcal{E}} u_D / \partial s^2\|^2_{L^2(\partial T \cap \partial\Omega)}.$$

MARK minimal set (for adaptive mesh-refinement) $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ of elements such that

$$1/2 \sum_{T \in \mathcal{T}_\ell} \eta_\ell(T)^2 \leq \sum_{T \in \mathcal{M}_\ell} \eta_\ell(T)^2.$$

(For uniform mesh-refinement set $\mathcal{M}_\ell = \mathcal{T}_\ell$.)
REFINE $\mathcal{T}_\ell$ by *red*-refinement of elements in $\mathcal{M}_\ell$ and *red-green-blue*-refinement of further elements to avoid hanging nodes and compute $\mathcal{T}_{\ell+1}$.
od

4.2. Numerical example on L-shaped domain

The first benchmark problem employs $f \equiv 0$ and inhomogenous Dirichlet data u_D of the exact solution

$$u(r, \varphi) = r^{2/3} \sin(2\varphi/3)$$

on the L-shaped domain $\Omega = (-1,1)^2 \backslash([0,1] \times [-1,0])$. The problem involves a typical corner singularity and shows an empirical convergence rate of 1/3 for uniform mesh refinement. This can be improved by adaptive refinement as shown in Figure 4. All error estimators induce meshes with the optimal empirical convergence rate 0.5.

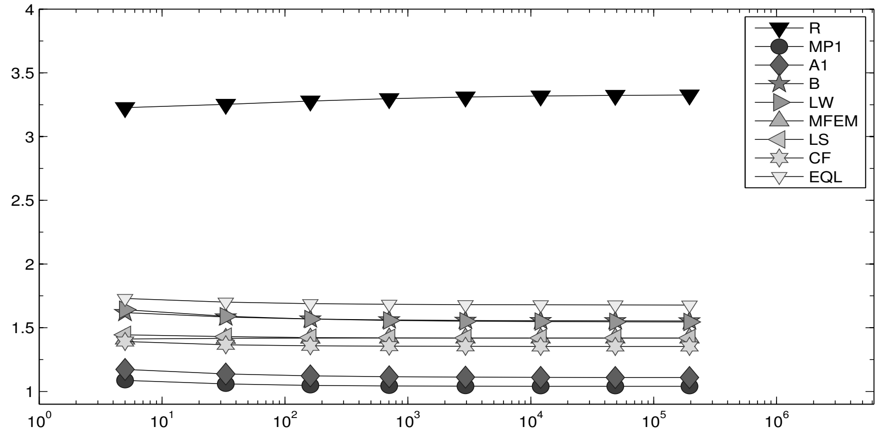


FIGURE 2. History of efficiency indices $\eta_{\mathrm{xyz}}/|||e|||$ of various a posteriori error estimators η_{xyz} labelled xyz in the figure as functions of the number of unknowns on uniform meshes in Subsection 4.2.

Figures 2 and 3 display the efficiency indices for uniform and adaptive mesh refinement. The optimal averaging η_{MP1} turns out to be asymptotic exact, but η_{MP1} as well as η_{A1} yield no guaranteed upper bound as the other estimators. While η_{R} takes efficiency indices of almost 4, all other error estimators arrive at efficiency indices below 1.7. The localisation error estimator η_{CF} is very accurate with values about 1.35 and is only beaten by η_{LW} for adaptive mesh refinement.

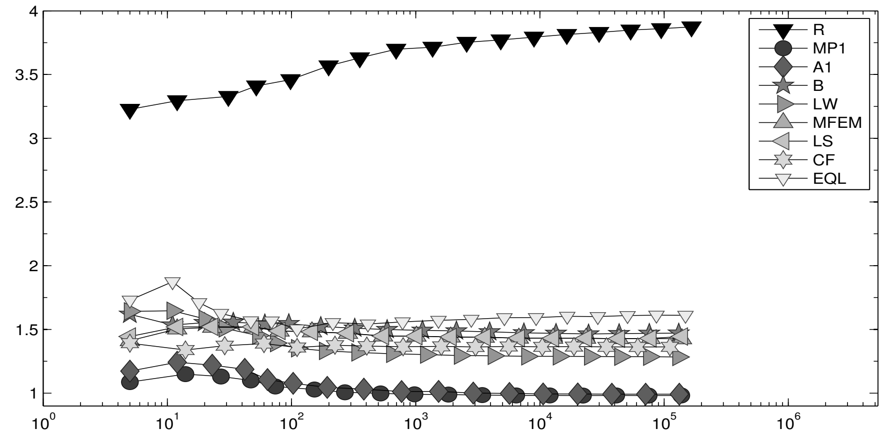


FIGURE 3. History of efficiency indices $\eta_{\mathrm{xyz}}/|||e|||$ of various a posteriori error estimators η_{xyz} labelled xyz in the figure as functions of the number of unknowns on adaptive meshes in Subsection 4.2.

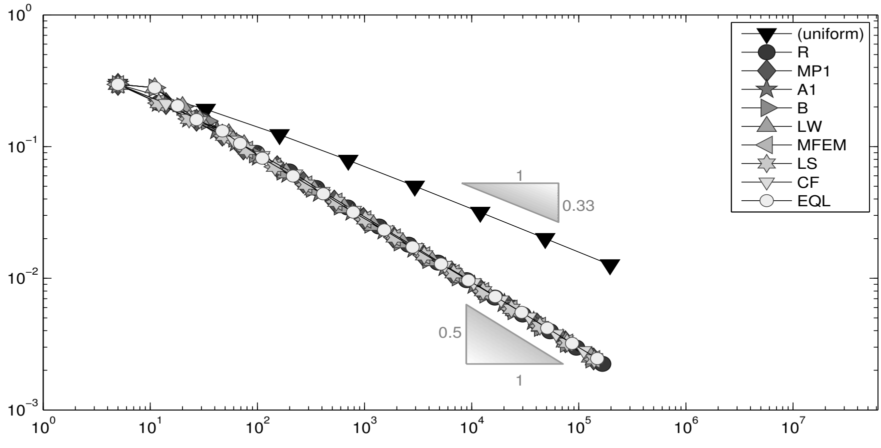


FIGURE 4. Convergence history of the energy errpr $|||e|||(\eta_{\mathrm{xyz}})$ for uniform and adaptive mesh refinement driven by various a posteriori error estimators η_{xyz} as functions of the number of unknowns in Subsection 4.2.

5. Nonconforming finite element method

This section deals with error control for noncoforming approximation for the Poisson model problem.

5.1. Discrete problem and notation

With the elementwise first-order polynomials $P_1(\mathcal{T})$, the nonconforming Crouzeix-Raviart finite element spaces read

$$\mathrm{CR}^1(\mathcal{T}) := \{v \in P_1(\mathcal{T}) \mid v \text{ is continuous at } \mathrm{mid}(\mathcal{E})\},$$
$$\mathrm{CR}^1_0(\mathcal{T}) := \{v \in CR^1(\mathcal{T}) \mid \forall E \in \mathcal{E}(\partial\Omega),\ v(\mathrm{mid}(E)) = 0.\}$$

The Crouzeix-Raviart finite elements form a subspaces of the broken Sobolev functions $H^1(\mathcal{T}) := \{v \in L^2(\Omega) \mid \forall T \in \mathcal{T}, v|_T \in H^1(T)\}$ with piecewise gradient $(\nabla_{\mathrm{NC}} v)|_T = \nabla v|_T$ for $v \in H^1(\mathcal{T})$ and $T \in \mathcal{T}$.

5.2. Error control via nonconforming residual

The error control dervied in [24] consists of two contributions. The first component contains the right-hand side f and its elementwise oscillations,

$$\mathrm{osc}(f, \mathcal{T}) := \|h_{\mathcal{T}}(f - f_{\mathcal{T}})\|_{L^2(\Omega)},$$

with the piecewise integral mean $f_{\mathcal{T}}$ and the piecewise constant mesh-size $h_{\mathcal{T}}$, $h_{\mathcal{T}}|_T := h_T$ for $T \in \mathcal{T}$. It reads

$$\eta := \|f_{\mathcal{T}}/2\ (\bullet - \mathrm{mid}(\mathcal{T}))\|_{L^2(\Omega)} + 1/\pi\ \mathrm{osc}(f, \mathcal{T}). \tag{5.1}$$

The second component derives from the residual defined, for any test function $v \in H^1(\Omega)$, by

$$\mathrm{Res}_{\mathrm{NC}}(v) := \int_{\partial\Omega} v\, \partial u_D/\partial s\, ds - \int_\Omega \nabla_{\mathrm{NC}} u_{\mathrm{CR}} \cdot \mathrm{Curl}\, v\, dx.$$

Its dual norm reads

$$|||\mathrm{Res}_{\mathrm{NC}}|||_\star := \sup_{\substack{v \in H^1(\Omega)\\ \mathrm{Curl}\, v \not\equiv 0}} \mathrm{Res}_{\mathrm{NC}}(v)/\|\mathrm{Curl}\, v\|_{L^2(\Omega)}.$$

The Helmholtz decomposition allows a split of the error in the broken energy norm

$$|||e|||^2_{\mathrm{NC}} \le \eta^2 + |||\mathrm{Res}_{\mathrm{NC}}|||^2_\star.$$

The dual norm $|||\mathrm{Res}_{\mathrm{NC}}|||_\star$ can be estimated with the error estimators from Section 3 with the data $f := 0$ and $\sigma_h := \mathrm{Curl}\, u_{\mathrm{CR}}$ and Neumann boundary data $g := \partial u_D/\partial s$. On the other hand, there exists an alternative characterisation of $|||\mathrm{Res}_{\mathrm{NC}}|||_\star$,

$$|||\mathrm{Res}_{\mathrm{NC}}|||_\star = \min_{\substack{v \in H^1(\Omega)\\ v = u_D \text{ on } \partial\Omega}} |||u_{\mathrm{CR}} - v|||_{\mathrm{NC}}.$$

Any conforming interpolation $v \in H^1(\Omega)$ with $v = u_D$ on $\partial\Omega$ gives an upper bound for $|||\mathrm{Res}_{\mathrm{NC}}|||_\star$.

5.3. Interpolation after Ainsworth

This subsection introduces the interpolation operator after Ainsworth [1] that designs some piecewise linear $I_A u_{\mathrm{CR}} \in H_0^1(\Omega)$ with respect to the original triangulation $\mathcal{T}$.

$$(I_{\mathrm{A}} v)(z) := \begin{cases} u_D(z) & \text{if } z \in \mathcal{N} \setminus \mathcal{K}, \\ \sum_{T \in \mathcal{T}(z)} u_{\mathrm{CR}}|_T(z)/|\mathcal{T}(z)| & \text{if } z \in \mathcal{K}. \end{cases}$$

The error estimator reads

$$\mu_{\mathrm{A}} := \|\nabla_{\mathrm{NC}} u_{\mathrm{CR}} - \nabla(I_{\mathrm{A}} u_{\mathrm{CR}})\|_{L^2(\Omega)}.$$

5.4. Modified interpolation operator

This subsection introduces an improved interpolation operator that designs some piecewise linear $I_{\mathrm{RED}} u_{\mathrm{CR}} \in H_0^1(\Omega)$ with respect to the red refined triangulation $\mathrm{red}(\mathcal{T})$. The nodes of $\mathrm{red}(\mathcal{T})$ consists of the nodes $\mathcal{N}$ and the edge midpoints $\mathrm{mid}(\mathcal{E})$ of $\mathcal{T}$. At the boundary the interpolation equals the nodal interpolation of u_D and on all edge midpoints it equals u_{CR}.

$$(I_{\mathrm{RED}} v)(z) := \begin{cases} u_{\mathrm{CR}}(z) & \text{for } z \in \mathrm{mid}(\mathcal{E}) \setminus \mathrm{mid}(\mathcal{E}(\partial\Omega)), \\ u_D(z) & \text{for } z \in (\mathcal{N} \cup \mathrm{mid}(\mathcal{E})) \cap \partial\Omega, \\ v_z & \text{for } z \in \mathcal{K}. \end{cases}$$

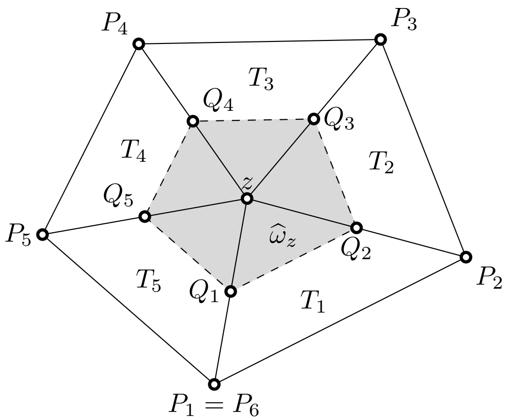


Figure 5. Interior patch

In this way, the interpolation equals u_{CR} on all central subtriangles like T_4 in Figure 6 and it remains to determine the values v_z at free nodes $z \in \mathcal{K}$. They may be chosen as in the design of I_{A}, but we suggest to choose them locally optimal as follows. Consider the node patch $\widehat{\omega}_z$ with respect to the red-refined triangulation as in Figure 5. Then minimise the contribution $\|\nabla_{\mathrm{NC}} u_{\mathrm{CR}} - \nabla v)\|_{L^2(\widehat{\omega}_z)}$ under the side condition of the fixed values at the edge midpoints Q_j of the adjacent edges. The value v_z at z remains the only degree of freedom in this local problem. The complete error estimator reads

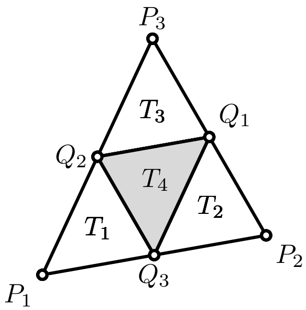


FIGURE 6. Central subtriangle $T_4 = \operatorname{conv}\{\operatorname{mid}(\mathcal{E}(T))\}$ in $\operatorname{red}(T)$ for $T \in \mathcal{T}$.

$$\mu_{\mathrm{RED}} := \|\nabla_{\mathrm{NC}} u_{\mathrm{CR}} - \nabla(I_{\mathrm{RED}} u_{\mathrm{CR}})\|_{L^2(\Omega)}.$$

We distinguish between the optimal version μ_{PMRED}, where v_z is chosen patchwise minimal (PM) as described above, and μ_{MARED} with the suboptimal choice v_z as in Subsection 5.3. This can be seen as a modification of I_{A} at the edge midpoints.

5.5. Optimal choices

The optimal $v \in P_1(\mathcal{T}) \cap C(\Omega)$ is attained at the solution u_C of the conforming formulation of the Poisson problem, since the nodal basis functions are included in $CR^1(\mathcal{T})$ and hence

$$\begin{aligned}\int_\Omega f v\, dx &= \int_\Omega \nabla_{\mathrm{NC}} u_{\mathrm{CR}} \cdot \nabla v_C\, dx \\ &= \int_\Omega \nabla u_C \cdot \nabla v_C\, dx \quad \text{for all } v_C \in P_1(\mathcal{T}) \cap H_0^1(\Omega).\end{aligned}$$

For comparison, we also compute the optimal $v_{\mathrm{MP1RED}} \in P_1(\operatorname{red}(\mathcal{T})) \cap C(\Omega)$ on the red-refined triangulation $\operatorname{red}(\mathcal{T})$ and the optimal piecewise quadratic $v_{\mathrm{MP2}} \in P_2(\mathcal{T}) \cap C(\Omega)$. Note that they don't have to equal the coressponding conforming solutions. To reduce the computational costs of v_{MP1RED} one might use $I_{\mathrm{MARED}} u_{\mathrm{CR}}$ as an initial guess for some iterative solver to draw near the optimal value. We use a preconditioned conjugate gradients algorithm and stop at the third iterate $v_{\mathrm{MP1RED(3)}}$. For the preconditioner we use the diagonal of the system matrix also known as Jacobi preconditioner.

5.6. Numerical example on L-shaped domain

Recall the data from the L-shaped problem from Section 4.2. Figures 7 and 8 show the efficiency indices of all estimators for uniform and adaptive mesh refinement, respectively. They vary between 1.1 for η_{MP2} and about 1.55 for

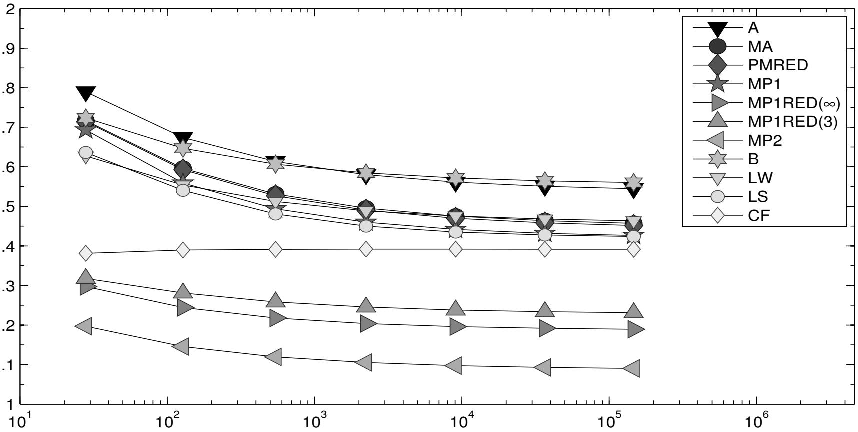


FIGURE 7. History of efficiency indices $\eta_{\mathrm{xyz}}/|||e|||$ of various a posteriori error estimators η_{xyz} labelled xyz in the figure as functions of the number of unknowns on uniform meshes in Subsection 5.6.

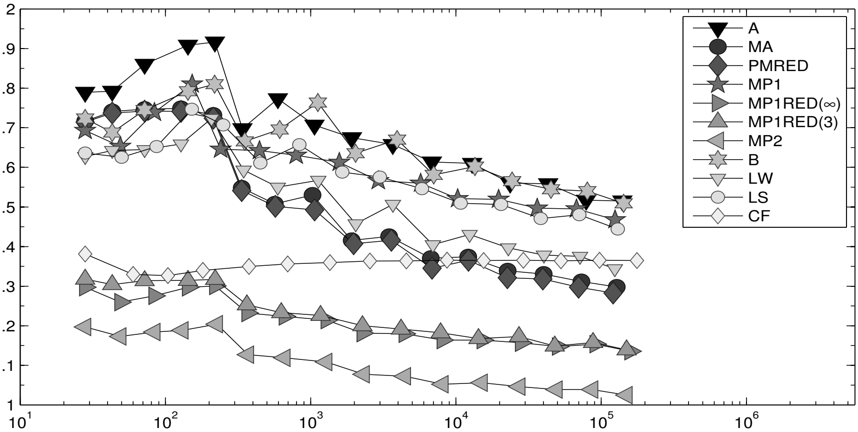


FIGURE 8. History of efficiency indices $\eta_{\mathrm{xyz}}/|||e|||$ of various a posteriori error estimators η_{xyz} labelled xyz in the figure as functions of the number of unknowns on adaptive meshes in Subsection 5.6.

η_{A} or η_{B}. The improved estimators η_{MARED} and η_{PMRED} perform significantly better. Their overestimation decreases under 35 percent which is even better than η_{MP1} or η_{LS}. The estimator η_{LW} performs similar but slightly worse compared to η_{MARED}. Figure 9 shows the convergence history of the

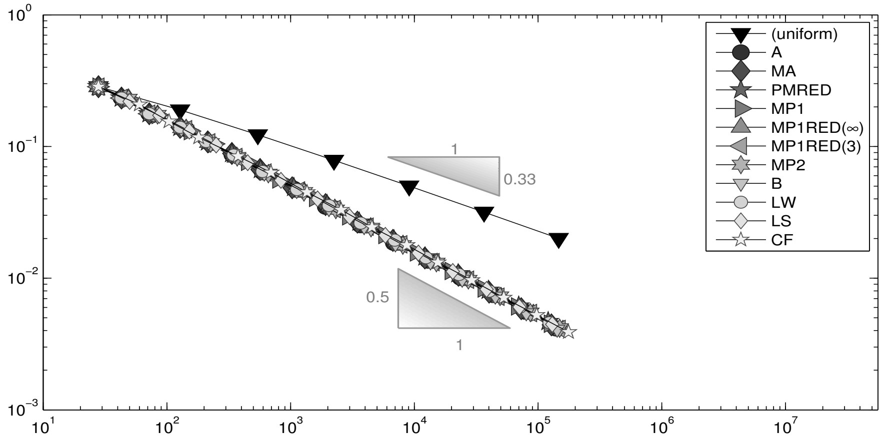


FIGURE 9. Convergence history of the energy errpr $|||e|||(\eta_{\mathrm{xyz}})$ for uniform and adaptive mesh refinement driven by various a posteriori error estimators η_{xyz} as functions of the number of unknowns in Subsection 5.6.

energy error for the adaptive meshes. The quality of the adaptive meshes is comparable for all error estimators.

6. Conforming obstacle problems

The unique exact weak solution $u \in K$ of the obstacle problem inside the closed and convex set of admissable functions,

$$K := \{v \in H^1(\Omega) \mid v = 0 \text{ on } \Gamma_D \text{ and } \chi \le v \text{ a.e. in } \Omega\} \neq \varnothing$$

satisfies

$$\int_\Omega \nabla u \cdot \nabla(u - v)\, dx \le \int_\Omega f(u - v)\, dx \text{ for all } v \in K. \tag{6.1}$$

6.1. Error control via auxiliary residual

After [12] and for a particular choice of Λ_h [23], the discrete solution of the obstacle problem u_h in

$$K(\mathcal{T}) := \{v_h \in P_1(\mathcal{T}) \cap C(\Omega) \mid v_h = 0 \text{ on } \Gamma_D \text{ and } \mathcal{I}\chi \le v_h \text{ in } \Omega\}$$

solves also the discrete version of the Poisson problem for $w \in V$ with

$$\int_\Omega \nabla w \cdot \nabla v\, dx = \int_\Omega (f - \Lambda_h) v\, dx \qquad \text{for all } v \in V. \tag{6.2}$$

The associated residual reads, for any $v \in H^1_0(\Omega)$,

$$\mathrm{Res}_{\mathrm{AUX}}(v) := \int_\Omega (f - \Lambda_h) v\, dx - \int_\Omega \nabla u_h \cdot \nabla v\, dx.$$

The energy norm difference $|||w - u_h||| = |||\mathrm{Res}_{\mathrm{AUX}}|||_\star$ between u_h and the exact solution w of the Poisson problem (6.2) can be estimated by any a posteriori error estimator from Section 3. In the conforming case $\chi \leq \mathcal{I}\chi$, [23] leads, for any a posteriori estimator η for $|||w - u_h|||$, to the reliable global upper bound (GUB) in the strict sense of

$$|||e||| \leq \mathrm{GUB}(\eta) := \left(\eta + |||\Lambda_h - J\Lambda_h|||_\star\right)/2 + \sqrt{\int_\Omega (\chi - u_h) J\Lambda_h\,dx + (\eta + |||\Lambda_h - J\Lambda_h|||_\star)^2}.$$

The patchwise oscillations

$$\mathrm{osc}(\Lambda_h, \mathcal{N}) := \left(\sum_{z\in\mathcal{N}} h_z^2 \min_{f_z\in\mathbb{R}} \|\Lambda_h - f_z\|^2_{L^2(\omega_z)}\right)^{1/2}$$

are a computable bound for

$$|||\Lambda_h - J\Lambda_h|||_\star := \sup_{v\in V\setminus\{0\}} \int_\Omega (\Lambda_h - J\Lambda_h) v\,dx / |||v||| \lesssim \mathrm{osc}(\Lambda_h, \mathcal{N}).$$

The competition in [23] compares five classes of error estimators from Section 3.

6.2. Numerical example with constant obstacle on L-shaped domain

This benchmark example from [8] mimics a typical corner singularity on the L-shaped domain $\Omega = (-2,2)^2 \setminus ([0,2]\times[-2,0])$ with constant obstacle $\chi = \mathcal{I}\chi \equiv 0$ and homogeneous Dirichlet data $u_D \equiv 0$ along $\partial\Omega$, with the right-hand side

$$\begin{aligned} f(r,\varphi) &:= -r^{2/3}\sin(2\varphi/3)\left(7/3\,(\partial g/\partial r)(r)/r + (\partial^2 g/\partial r^2)(r)\right) - H(r - 5/4),\\ g(r) &:= \max\{0, \min\{1, -6s^5 + 15s^4 - 10s^3 + 1\}\} \end{aligned}$$

for $s := 2(r - 1/4)$ and the Heaviside function H. The exact solution reads

$$u(r,\varphi) := r^{2/3} g(r)\sin(2\varphi/3).$$

The experimental convergence rate for uniform refinement is about 0.4 and adaptive refinement improves it to the optimal value 0.5 as depicted in Figure 12. Figures 10 and 11 monitor the efficiency of the upper bounds $\mathrm{GUB}(\eta_{\mathrm{xyz}})$. The efficiency of the bound associated to the standard residual-based error estimator $\mathrm{GUB}(\eta_{\mathrm{R}})$ is between 7 and 9, while all other error estimators allow efficiency indices below 2. As observed in a posteriori error estimation for Poisson Problems in Section 4.2, the upper bound $\mathrm{GUB}(\eta_{\mathrm{MP1}})$ almost arrives at effiency index 1.

7. Conclusions

The theoretical and practical results of this paper support the following observations.

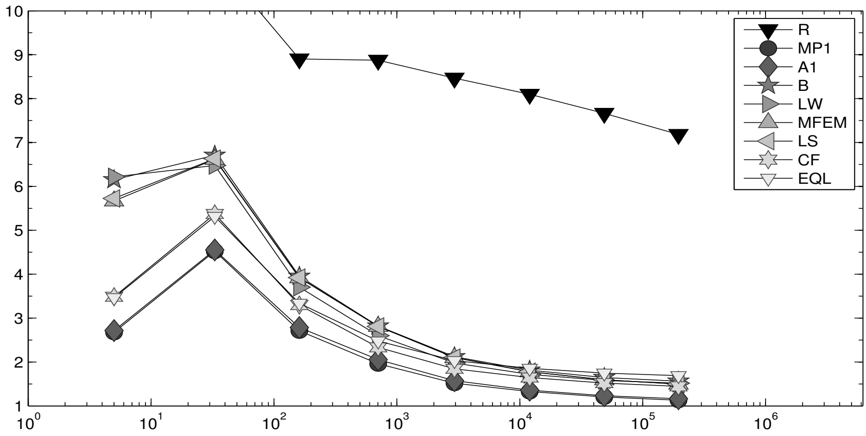


FIGURE 10. History of efficiency indices $\mathrm{GUB}(\eta_{\mathrm{xyz}})/|\!|\!|e|\!|\!|$ of various a posteriori error estimators η_{xyz} labelled xyz in the figure as functions of the number of unknowns on uniform meshes in Subsection 6.2.

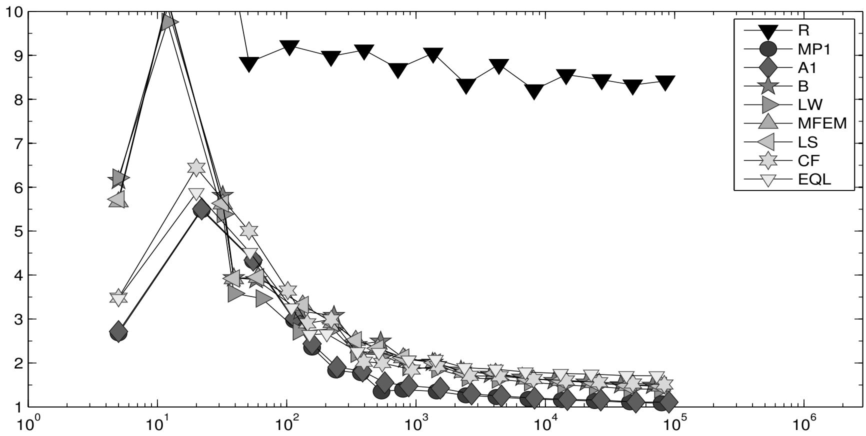


FIGURE 11. History of efficiency indices $\mathrm{GUB}(\eta_{\mathrm{xyz}})/|\!|\!|e|\!|\!|$ of various a posteriori error estimators η_{xyz} labelled xyz in the figure as functions of the number of unknowns on adaptive meshes in Subsection 6.2.

7.1. Explicit error estimators sufficient for effective mesh design

Adaptive mesh refinement may be steered by simple η_{R}-based refinement rules. It does not appear to be favourable to spend more computational time for more laborious refinement rules if the data are (relatively) smooth.

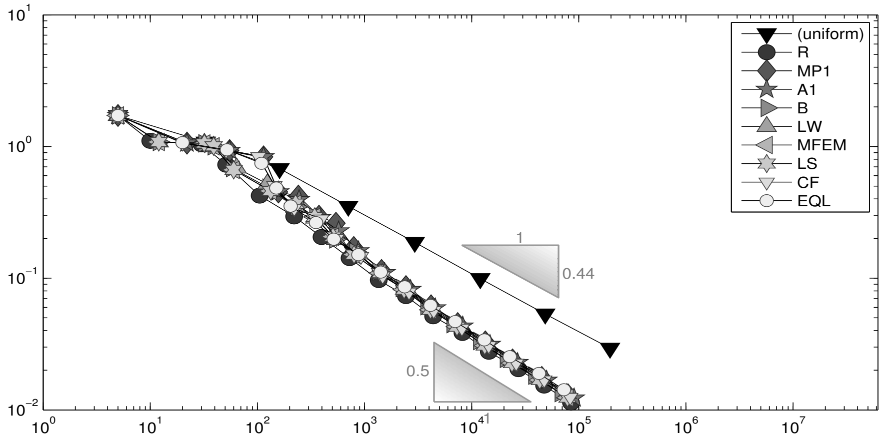


FIGURE 12. Convergence history of the energy errpr $|||e|||$ for uniform and adaptive mesh refinement driven by various a posteriori error estimators η_{xyz} as functions of the number of unknowns in Subsection 6.2.

7.2. Approximation of local problems

We found that fourth-order polynomials are sufficient enough to provide accurate approximations of the guaranteed upper bounds. However, for full reliability, this approximation error has to be controlled further. The numerical experiments in this paper leave this out and therefore are not fully reliable. This fundamental difficulty is circumvented by modern equilibration error estimators like η_{B} and η_{LW}. This suffices to conclude, that the novel techniques are superior to η_{EQL} or η_{CF}.

7.3. Robust error control via η_{CF}, η_{LS}, η_{MFEM} or η_{LW}

The estimators η_{CF}, η_{LS} or η_{MFEM} and η_{LW} seem to be the most robust estimators and are recommended as a termination criterion for guaranteed error control. The residual-based estimator η_{R} is too coarse and not appropriate as termination criterion for guaranteed error control.

7.4. Accurate error control pays off

Averaging error estimators might be an very good exact error guess but they do not guarantee to be an upper bound for the exact error to justify termination. On the other hand, relying only on cheap error estimators like η_R causes overkill refinements and might be more expensive than the computation of more laborious but sharper error estimators like the ones from Section 7.3. That is why it is is favorable to have a variety of error estimators [11].

7.5. Recomandation in practise

In the end a combination of several error estimators is recommended, e.g., η_{R} for generating refinement indicators and a simple averaging error estimator for the decision eighter to refine or to employ a fine error estimator to justify termination or the need for further refinement.

References

[1] Ainsworth, M., *Robust a posteriori error estimation for nonconforming finite element approximation*, SIAM J. Numer. Anal., **42**(2004), no. 6, 2320-2341.

[2] Ainsworth, M., Dörfler, W., *Reliable a posteriori error control for nonconformal finite element approximation of Stokes flow*, Math. Comp., **74**(2005), no. 252, 1599-1619 (electronic).

[3] Ainsworth, M., Oden, J.T., *A posteriori error estimation in finite element analysis*, Wiley, 2000.

[4] Ainsworth, M., Oden, J.T., *A posteriori error estimators for second order elliptic systems, I, Theoretical foundations and a posteriori error analysis*, Comput. Math. Appl., **25**(1993), no. 2, 101-113.

[5] Ainsworth, M., Oden, J.T., *A posteriori error estimators for second order elliptic systems, II, An optimal order process for calculating self-equilibrating fluxes*, Comput. Math. Appl., **26**(1993), no. 9, 75-87.

[6] Ainsworth, M., Oden, J.T., *A unified approach to a posteriori error estimation using element residual methods*, Numer. Math., **65**(1993), 23-50.

[7] Ainsworth, M., Rankin, R., *Fully computable bounds for the error in nonconforming finite element approximations of arbitrary order on triangular elements*, SIAM J. Numer. Anal., **46**(2008), no. 6, 3207-3232.

[8] Bartels, S., Carstensen, C., *Averaging techniques yield reliable a posteriori finite element error control for obstacle problems*, Numer. Math., **99**(2004), no. 2, 225-249.

[9] Bartels, S., Carstensen, C., Dolzmann, G., *Inhomogeneous Dirichlet conditions in a priori and a posteriori finite element error analysis*, Numer. Math., **99**(2004), no. 1, 1-24.

[10] Bartels, S., Carstensen, C., Jansche, S., *A posteriori error estimates for nonconforming finite element methods*, Numer. Math., **92**(2002), 233-256.

[11] Bartels, S., Carstensen, C., Klose, R., *An experimental survey of a posteriori Courant finite element error control for the Poisson equation*, Adv. Comput. Math., **15**(2001), no. 1-4, 79-106.

[12] Braess, D., *A posteriori error estimators for obstacle problems - Another look*, Numer. Math., **101**(2005), 415-421.

[13] Braess, D., Schöberl, J., *Equilibrated residual error estimator for edge elements*, Math. Comp., **77**(2008), no. 262, 651-672.

[14] Brandts, J., Chen, Y., Yang, J., *A note on least-squares mixed finite elements in relation to standard and mixed finite elements*, IMA J. Numer. Anal., **26**(2006), no. 4, 779-789.

[15] Brenner, S.C., Carstensen, C., *Finite element methods*, John Wiley and Sons, 2004.

[16] Brenner, S.C., Scott, L.R., *The mathematical theory of finite element methods*, Springer-Verlag, 2008.

[17] Carstensen, C., *Lectures on finite element methods at Yonsei University (Seoul) and Humboldt-Universität zu Berlin*, 2009/2010.

[18] Carstensen, C., *All first-order averaging techniques for a posteriori finite element error control on unstructured grids are effective and reliable*, Math. Comp., **73**(2004), 1153-1165.

[19] Carstensen, C., *A unifying theory of a posteriori finite element error control*, Numer. Math., **100**(2005), no. 4, 617-637.

[20] Carstensen, C., Eigel, M., Hoppe, R.H.W., Löbhard, C., *A review of unified a posteriori finite element error control*, IMA Preprint Series, **2338**(2010).

[21] Carstensen, C., Funken, S.A., *Fully reliable localised error control in the FEM*, SIAM J. Sci. Comput., **21**(1999), no. 4, 1465-1484 (electronic).

[22] Carstensen, C., Merdon, C., *Estimator competition for Poisson problems*, J. Comp. Math., **28**(2010), no. 3, 309-330.

[23] Carstensen, C., Merdon, C., *A posteriori error estimator competition for conforming obstacle problems, part 1: Theoretical preliminaries, part 2: Numerical results*, submitted.

[24] Carstensen, C., Merdon, C., *Computational survey on a posteriori error estimators for nonconforming finite element methods for the Poisson problem*, in preparation.

[25] Carstensen, C., Merdon, C., *Six remarks on the Luce-Wohlmuth a posteriori error control* (in preparation).

[26] Dietrich, B., *Finite elements - Theory, fast solvers, and applications in solid mechanics*, Cambridge University Press, 2007.

[27] Girault, V., Raviart, P.A., *Finite element methods for Navier-Stokes equations*, Springer-Verlag, 1986.

[28] Hoppe, H.W., Wohlmuth, B., *Element-oriented and edge-oriented local error estimators for nonconforming finite element methods*, 1996.

[29] Ladeveze, P., Leguillon, D., *Error estimate procedure in the finite element method and applications*, SIAM J. Numer. Anal., **20**(1983), no. 3, 485-509.

[30] Luce, R., Wohlmuth, B.I., *A local a posteori error estimator based on equilibrated fluxes*, Society for Industrial and Applied Mathematics, 2004.

[31] Marini, L.D., *An inexpensive method for the evaluation of the solution of the lowest order Raviart-Thomas mixed method*, SIAM J. Numer. Anal., **22**(1985), 493-496.

[32] Payne, L.E., Weinberger, H.F., *An optimal Poincaré inequality for convex domains*, Arch. Rat. Mech. Anal., **5**(1960), 286-292.

[33] Repin, S., *A posteriori estimates for partial differential equations*, Walter de Gruyter, 2008.

[34] Repin, S., *Two-sided estimates of deviation from exact solutions of uniformly elliptic equations*, Proceedings of the St. Petersburg Mathematical Society, **IX**(2003), 143-171.

[35] Repin, S., Sauter, S., Smolianski, A., *A posteriori error estimation for the Dirichlet problem with account of the error in the approximation of boundary conditions*, Computing, **70**(2003), no. 3, 205-233.

[36] Stoyan, G., *Towards discrete Velte decompositions and narrow bounds for inf-sup constants*, Comput. Math. Appl., **38**(1999), no. 7-8, 243-261.

[37] Valdman, J., *Minimization of functional majorant in a posteriori error analysis based on H(div) multigrid-preconditioned CG method*, Advances in Numerical Analysis, 2009.

[38] Verfürth, R., *A posteriori error estimators for the Stokes equations. II, Nonconforming discretizations*, Numer. Math., **60**(1991), no. 2, 235-249.

[39] Verfürth, R., *A posteriori error estimators for the Stokes equations*, Numer. Math., **55**(1989), no. 3, 309-325.

[40] Vohralík, M., *Guaranteed and fully robust a posteriori error estimates for conforming discretizations of diffusion problems with discontinuous coefficients*, 2009.

[41] Volker, J., *A posteriori L2-error estimates for the nonconforming P1/P0-finite element discretization of the Stokes equations*, J. Comput. Appl. Math., **96**(1998), no. 2, 99-116.

Carsten Carstensen and Christian Merdon
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin, Germany

and

Department of Computational Science and Engineering
Yonsei University
120-749 Seoul, Korea
e-mail: cc@mathematik.hu-berlin.de
merdon@mathematik.hu-berlin.de

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 295–298

On a quaternion valued Gaussian random variables

George Chelidze and Nicholas Vakhania

Abstract. In the present note we show that Polya's type characterization theorem of Gaussian distributions does not hold. This happens because in the linear form, constituted by the independent copies of quaternion random variables, a part of the quaternion coefficients is written on the right hand side and another part on the left side. This gives a negative answer to the question posed in [1].

Mathematics Subject Classification (2010): 60B15.

Keywords: Quaternion random variables, Polya's characterization theorem.

The present note is a natural extension of paper [1] where the formulation and proof of Polya's theorem on the characterization of Gaussian random variables with values in quaternion algebra is considered. We mean the following well-known theorem of Polya:

Theorem 1.1. *Let $\xi_1, \xi_2, ..., \xi_n$, $n \geq 2$ be i.i.d. random variables and $(a_1, a_2, ..., a_n)$ be nonzero reals that satisfy the condition $\sum_{h=1}^{n} a_h^2 = 1$. If the sum $\sum_{h=1}^{n} a_h \xi_h$ has the same distribution as ξ_1, then ξ_1 is a Gaussian random variable.*

If the random variable takes values in the quaternion algebra then three types of Gaussian random variables are considered: real, complex and quaternion Gaussian random variables. Let us recall the definition of complex and quaternion Gaussian random variables. The usual motivation for these definitions comes from the form of characteristic function of a centered Gaussian random variable, see e.g. [2]. For the real case this is given as

$$\exp\{-\frac{1}{2}t^2 E\xi^2\},\ \forall t \in R.$$

The authors was supported by grant GNSF/ST09_99_3-104.

For the complex (quaternion) case we would analogously expect the characteristic function to be

$$\exp\{-c|q|^2 E|\xi|^2|\}, \forall q \in C, (\forall q \in Q), c > 0. \tag{1.1}$$

The characteristic function of a complex (quaternion) random variables ξ is defined as

$$\chi_\xi(q) = E \exp(i\mathrm{Re}(\xi\overline{q}))$$

and if we want the characteristic function of centered complex (quaternion) Gaussian random variable to have the form (1.1), then the covariance matrix of real two dimensional vector $(\xi^{'}, \xi^{''})$ (four dimensional vector $(\xi^{'}, \xi^{''}, \xi^{'''}, \xi^{IV})$) should be proportional to the identity matrix. Thus the covariance matrices of complex (quaternion) Gaussian random variables have a quite specific form: they are proportional to unit matrices in R^2 (in R^4). Therefore the coordinates of corresponding two dimension (four dimension) random vector $(\xi^{'}, \xi^{''})$, $((\xi^{'}, \xi^{''}, \xi^{'''}, \xi^{IV}))$ are mutually independent and have the same variances.

In [3] there is formulated Polya's theorem for the case of complex random variables.

Theorem 1.2. *Let ξ be a complex random variable, $\xi_1, \xi_2, ..., \xi_n$, $n \geq 2$ be independent copies of ξ and $(a_1, a_2, ..., a_n)$ be nonzero complex numbers such that $\sum_{h=1}^{n} |a_h|^2 = 1$ and at least one of them is not a real number. If $\sum_{h=1}^{n} a_h \xi_h$ has the same distribution as ξ, then ξ is a complex Gaussian random variable.*

As we see in the complex case there is an additional condition on the complex coefficients $(a_1, a_2, ..., a_n)$, for the Theorem 1.2 to be true, namely one of these coefficients should be essentially complex number. In [1] there is shown that in the quaternion case, such additional condition on the quaternions $(a_1, a_2, ..., a_n)$, plays condition which we call jointly quaternion system, i.e. the following theorem is true.

Theorem 1.3. *Let ξ be a quaternion random variable, $\xi_1, \xi_2, ..., \xi_n$, $n \geq 2$, be independent copies of ξ, and $(a_1, a_2, ..., a_n)$ be nonzero quaternions that form jointly quaternion system and satisfy the condition $\sum_{h=1}^{n} |a_h|^2 = 1$. Then, if the sum $\eta = \sum_{h=1}^{n} a_h \xi_h$ has the same distribution as ξ, ξ is quaternion Gaussian random variable.*

Now let us recall the definition of jointly quaternion system.

Definition 1.4. *We say that a collection of n quaternions $(a_1, a_2, ..., a_n)$, $n \geq 2$, constitutes a jointly quaternion system (JQS) if there does not exist imaginary number $\tilde{i} = \alpha i + \beta j + \gamma k$, with real α, β, γ, such that the following expressions holds: $a_1 = a_1^{'} + a_1^{''}\tilde{i}$, $a_2 = a_2^{'} + a_2^{''}\tilde{i}$,..., $a_n = a_n^{'} + a_n^{''}\tilde{i}$, $a_i^{'}, a_i^{''} \in R$, $1 \leq i \leq n$.*

This definition has also another interpretation: let $A \equiv (a_1, a_2, \ldots, a_n)$, $n \geq 2$, be the collection of quaternions not necessarily different to each other. Denote by $A'' \equiv (a_1'', a_2'', \ldots, a_n'')$, $A''' \equiv (a_1''', a_2''', \ldots, a_n''')$, $A^{IV} \equiv (a_1^{IV}, a_2^{IV}, \ldots, a_n^{IV})$. We say that the collection A is JQS if at least one of the three pairs (A'', A'''), (A'', A^{IV}) and (A''', A^{IV}) is a pair of non-collinear vectors in R^n or, in other words, if the vectors A'', A''' and A^{IV} do not belong to an one dimensional subspace of R^n. This name is motivated by the following observation: Any (one) quaternion $a = a' + ia'' + ja''' + ka^{IV}$ can be written as a complex number with respect to some imaginary unit $\tilde{i}$, defined by the following equality

$$\tilde{i} = \frac{ia'' + ja''' + ka^{IV}}{(a''^2 + a'''^2 + a^{IV^2})^{1/2}}.$$

Indeed, we have, $\tilde{i}^2 = -1$ and $a = a' + \tilde{i}\,\widetilde{a''}$, where $\widetilde{a''} = (a''^2 + a'''^2 + a^{IV^2})^{1/2}$. However, the collection of quaternions $A \equiv (a_1, a_2, \ldots, a_n)$, $n \geq 2$, not always can be expressed as complex numbers with the common imaginary unit. This can be done if and only if A is not a JQS.

Since the multiplication of quaternions is not commutative, the following natural question was posed at the end of [1]: is the Theorem 1.3 true if in the linear form $\eta = \sum_{h=1}^{n} a_h \xi_h$, a part of the coefficients $a_h, 1 \leq h \leq n$, are written on the left side of $\xi_h, 1 \leq h \leq n$ and other part on the right? The following example shows that the answer of this question is negative, i.e. it may happen that $a_1\xi_1 + \xi_2 a_2$ has the same distribution as ξ, (a_1, a_2) form the jointly quaternion system, but ξ is not a quaternion Gaussian random variable.

Example 1.5. Let $\xi = \xi^{'} + i\xi^{''} + j\xi^{''} - k\xi^{'}$, where $\xi^{'}$ and $\xi^{''}$ are independent standard Gaussian random variables. It is clear that the covariance matrix of the random vector $(\xi^{'}, \xi^{''}, \xi^{''}, -\xi^{'})$ has the form

$$\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

Hence, ξ is not a quaternion Gaussian random variable, however using technique of characteristic functions it is not hard to show that ξ and $\eta = \frac{i}{\sqrt{2}}\xi_1 + \xi_2\frac{j}{\sqrt{2}}$ are equally distributed, and $(\frac{i}{\sqrt{2}}, \frac{j}{\sqrt{2}})$ is the jointly quaternion system.

References

[1] Vakhania, N.N., Chelidze, G.Z., *Quaternion-valued Gaussian random variables*, Theory of probability and its applications, **54**(2009), no. 2, 337-344.

[2] Vakhania, N.N., *Random variables with values in quaternion Hilbert spaces*, Theory of Probability and its Applications, **43**(1998), no. 1, 18-40.

[3] Vakhania, N.N., *Polya's characterization theorem for complex random variables*, J. Complexity, **13**(1997), no. 4, 480-488.

George Chelidze and Nicholas Vakhania
Faculty of Exact and Natural Sciences
I. Javakhshvili Tbilisi State University 2
University St., Tbilisi 0143
Georgia
e-mail: g.chelidze@math.sci.tsu.ge

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 299–303

SLAD method for cancer registration

Ioana Chiorean, Liana Lupşa and Luciana Neamţiu

Abstract. The Logical Analysis of Data (LAD) is a method extensively used in Medicine for data classification. The present paper contains a slightly modified approach of this method, called Successive Logical Analysis of Data (SLAD), more appropriate to the data registration in oncology. The corresponding algorithm is also presented.

Mathematics Subject Classification (2010): 68T30, 68T50, 68W10.

Keywords: Logical Analysis of Data (LAD), patterns, data mining, cancer registration, morphological code.

1. Introduction

Cancer registration is a continuous and systematic process of collecting data concerning the occurrence and characteristics of reportable neoplasm. The tumors may be classified according to the International Classification of Diseases for Oncology (ICD-O) and each of them has a corresponding code made by six digits. The first four digits represent the specific histological term, the fifth is the behavior code and the sixth is the grade of differentiation. The book ICD-O contains also a dictionary of codes, where to every numerical code corresponds a group-of-words-in-natural-language. For instance:

Code	Description
8500/3_	Infiltrating duct carcinoma
	Infiltrating duct adeno carcinoma
	Duct adeno carcinoma
	Duct carcinoma
	Duct cell carcinoma
	Duct carcinoma
8480/3_	Mucinous adeno carcinoma
	Gelatinous adeno carcinoma
	Mucous carcinoma
	Colloidal carcinoma

2. The data and the problem

The data

When a patient has a tumor, several investigations have to be done in order to determine its nature. The information is written in a medical record, in a "free" language, which contains, medical terms and other terms (which we call "noisy" terms).

2.1. The problem

In order to process the medical information, for establishing the corresponding code, firstly the information has to be cleaned of "noisy" words, and so the medical terms will be emphasized. Then, if the described tumor is malign, the patient has to be introduced in the cancer register with the morphological and topographical code for the tumor given according to the rules from **ICD-O**. Due to the fact that the medical terms may be consider "patterns", being recognized in the dictionary of codes, the method that we use in our paper to determine the final code for a tumor is based on Logical Analysis of Data (LAD), which is a new methodology used for detecting structural information about datasets. A specific characteristic of LAD is the detection of logical patterns which determine and predict out of a group, a class satisfying specific requirements (see [1], [2]).

Due to the fact that most of the observations in which we have to detect some code, do not contain exactly the group of words which are coded in the dictionary, but others, with the same meaning, we have to construct our own patterns. For this purpose, as in [3], we use our own method, called Successive LAD Method (SLAD), because we have to decide what code we have to give to an observation which contains groups of words belonging to different codes.

3. Constructing the patterns sets

In what follows, we denote by
PG: the set of all expressions corresponding to all codes (all patterns)
WE: the set of all words which appear in the expressions from second column of the dictionary.

The main idea of SLAD method consists in applying the classical LAD method successively, introducing patterns of different levels. They are the following:

1. *The patterns of level 0-* "does the tumor exist"?

In order to answer to this question, we construct the sets:

$PL+0 = \{exists,\ has,\ etc.\}$,

$PL-0 = \{does\ not\ exists,\ has\ not,\ etc.\}$.

2. *The patterns of level 1*, denoted by *PL1*, contains 1 key word from the dictionary, and determine the fifth position in the morphological code (e.g. metastatic, carcinoma, limphoma, etc.).

3. *The patterns of level 2, level 3, etc., using SLAD.*

They define the four digits of the morphological code. For every pattern p from PL1 we consider the set $PL2(p)$ made by those patterns $w\Box PG$ with the property that the concatenated patterns pw or wp are to be found in the dictionary.

Example. $p = carcinoma$ from PL1;

$w = duct$ from PG;

Then $pw = duct\ carcinoma$ and has the code 8500/3 in the dictionary. Therefore,

$$duct\Box PL2(carcinoma).$$

4. The algorithms

In [3], the algorithms for the following situations are given:

a) **All the key words in the record appear exactly in the order given in the pattern.**

Example. Let's consider the registration *"Invasive duct carcinoma with extensive papillary component"*.

Step 1. Transform this observation in patterns (key words):

"Invasive duct carcinoma"

Step 2. Apply algorithm *Pattern*:

- looking in the dictionary for the existing patterns, we get:

8500/3, for *duct carcinoma* and,

8503/3, for *intraductal papillary adeno carcinoma with invasion*

Step 3. Computing the final code, as the maximum:

$Code = max 8500/3, 8503/3 = 8503/3.$

Conclusion: Our registration *"Invasive duct carcinoma with extensive papillary component"* will have the code 8503/3.

b) **The words from the record are the same with those in the patterns, but their order differs**

Example. Let us consider the registration: *myxofibrosarcoma*, which is not in PG, but pattern fibromyxosarcoma is, to which corresponds the code 8811/3. Then, the myxofibrosarcoma record will receive the code 8811/3.

In the present paper we present another approach, when:

c) **The record contains key words which are not in the dictionary**

Example. Let's consider the record *"intrusive duct malignant with pap. comp."*

The following key words are not in the dictionary: *intrusive, malignant, pap., comp.* Also, we have some shortenings: *pap.=papillary; comp.=component.*

In order to solve the problem, we propose the following steps:

1. Generate lexicographic dictionary (SINO), which contains all the synonyms
2. Give weights, $w(i)$ to every key word $r(i)$ in the record, where $w(i) \in \{0, 0.1, \ldots, 0.9, 1\}$, for $i = 1$ to n, according with how close is the word to a pattern from the ICD-O dictionary
3. Compute $WI = (w(1) + \ldots + w(n))/n$

4. If $WI >= 0.80$, then the record enters in the dictionary, as another pattern, and receives the corresponding morphological code; if not (i.e. $WI < 0.80$), it has to go back to the physician. He will give the corresponding code and the record, together with this code, will be memorized in the dictionary

Algorithm NewPattern;

Begin
Generate SINO;
For $i = 1$ *to* p *do* {*take a record*}
For $j = 1$ *to* n *do* {*take a key word*}
Lookfor_in_SINO;
Give_Weight$(w(j))$*;*
Endfor;
$WI = (w(1) + \ldots + w(n))/n;$
If $WI >= 0.80$ *then Memo_in_ICD-O;*
Give_code
else Return_to_ Physician
Endif;
Endfor;
End.

Example. Let's consider the registration *"intrusive duct malignant with pap. comp."*

Key words: *intrusive, duct, malignant, pap., comp. so* $n = 5$
- suppose SINO is created, then we have:
Synonyms: $intrusive = invasive; w(1) = 1$
$malignant = carcinoma; w(3) = 1$
- *comp.* and *pap.* get $w(4) = 0.3$, $w(5) = 0.7$
- Compute $WI = (1 + 1 + 1 + 0.3 + 0.7)/5 = 0.8$
- Write in the ICD-O dictionary the new pattern
- Give the record the code 8503/3.

References

[1] Boros, E., Hammer P. L., Ibaraki T., Kogan Al., *Logical analysis of numerical data*, Mathematical Programming, 1997, 79, 163-90.

[2] Boros E., Hammer P. L., Ibaraki T., Kogan Al., Mayoraz E., Muchnik I., *An Implementation of Logical Analysis of Data*, IEEE Transactions on Knowledge and Data Engineering, 2000, 12, 292-306.

[3] Lupşa, L., Chiorean,I., Neamţiu, L., *Use of LAD in establishing morphologic code*, Proceedings of IEEE International Conference on Automation, Quality and Testing, Robots, AQTR 2010, 219-224.

Ioana Chiorean
Babeş-Bolyai University
Faculty of Mathematics and Computer Science
Str. Kogălniceanu, nr. 1
400084 Cluj-Napoca, Romania
e-mail: ioana@math.ubbcluj.ro

Liana Lupşa
Babeş-Bolyai University
Faculty of Mathematics and Computer Science
Str. Kogălniceanu, nr. 1
400084 Cluj-Napoca, Romania
e-mail: llupsa@math.ubbcluj.ro

Luciana Neamţiu
Cancer Institute "I. Chiricuţă"
Str. Republicii, nr. 34-36
400015 Cluj-Napoca, Romania
e-mail: luciana@iocn.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 305–313

On the Szasz-Inverse Beta operators

Cristina S. Cismaşiu

Abstract. In this paper, we consider a probabilitistic representation of the Szasz-Inverse Beta operators, which are an mixed summation-integral type operators, and we study some approximation properties using probabilistic methods.

Mathematics Subject Classification (2010): 41A35, 41A36, 41A25,42A61.

Keywords: Szasz-Inverse Beta operators, Szasz-Mirakjan operators, Inverse Beta operators, Phillips operators, modified Szasz-Inverse Beta operators, estimates.

1. Probabilistic representation of the Szasz-Inverse Beta operators

In this paper we consider a probabilistic representation of the Szasz-Inverse Beta operators and study some approximation properties, using probabilistic methods. These operators were defined by (1.1)-(1.5) and were investigated by Gupta V., Noor M. A., [11] and some iterative constructions of these operators were studied recently by Finta Z., Govil N. K., Gupta V. [10]:

$$\begin{aligned} L_t(f;x) &= e^{-tx}f(0) + \sum_{k=1}^{\infty} s_{t,k}(x) \int_0^{\infty} b_{t,k}(u)f(u)du \qquad (1.1) \\ &= \int_0^{\infty} J_t(u;x)f(u)du,\ x \geq 0 \end{aligned}$$

with

$$s_{t,k}(x) = e^{-tx}\frac{(tx)^k}{k!},\ t > 0,\ x \geq 0,\ k \in \mathbb{N} \cup \{0\} \qquad (1.2)$$

$$b_{t,k}(u) = \frac{1}{B(k,t+1)} \cdot \frac{u^{k-1}}{(1+u)^{t+k+1}},\ t > 0,\ u > 0, \qquad (1.3)$$

$$B(k,t+1)=\int\limits_0^\infty \frac{u^{k-1}}{(1+u)^{t+k+1}}du \tag{1.4}$$

being Inverse-Beta function

$$J_t(u;x)=e^{-tx}\delta(u)+\sum_{k=1}^{\infty} s_{t,k}(x)b_{t,k}(u), \tag{1.5}$$

$\delta(u)$ being the Dirac's delta function, for which $\int\limits_0^\infty \delta(u)f(u)du=f(0)$.

Using same ideea as Adell J. A., De la Cal J., [2], these operators can be represented as the mean value of the random variable $f\left(\frac{U_{N(tx)}}{V_{t+1}}\right)$ which has the probability density function $J_t(\cdot;x)$:

$$L_t(f;x)=E\left[f\left(Z_{tx}\right)\right]=E\left[f\left(\frac{U_{N(tx)}}{V_{t+1}}\right)\right],\ t>0,\ x\geq 0, \tag{1.6}$$

with $\{N(t):\ t\geq 0\}$ a standard Poisson process and $\{U_t:\ t\geq 0\}$, $\{V_t:\ t\geq 0\}$ two mutually independent Gamma processes defined all on the same probability space.

Note that, the Poisson process is a stochastic process starting at the origin, having stationary independent increments with probability

$$P\left(N(t)=k\right)=\frac{e^{-t}t^k}{k!},\ t\geq 0,\ k\in\mathbb{N}\cup\{0\} \tag{1.7}$$

and the Gamma process is a stochastic process starting at the origin $(U_0=0,V_0=0)$, having stationary independent increments and such that for $t>0$, U_t, V_t have the Gamma probability density function

$$\rho_t(u)=\begin{cases}\dfrac{u^{t-1}e^{-u}}{\Gamma(t)} & ,\ t>0,\ u>0,\\ 0 & ,\ u=0\end{cases} \tag{1.8}$$

and without loss of generality [17] it can be assumed that $\{U_t:\ t\geq 0\}$, $\{V_t:\ t\geq 0\}$ for each $t>0$ has a.s. no decreasing right-continuous paths.

Indeed, in our paper [4] we showed that

$$\begin{aligned}
&E\left[f\left(\frac{U_{N(tx)}}{V_{t+1}}\right)\right]=\int\limits_0^\infty f(u)\left(\int\limits_0^\infty y\rho_{U_{N(tx)}}(yu)\rho_{V_{t+1}}(y)dy\right)du\\
&=\int\limits_0^\infty f(u)\left(\int\limits_0^\infty y\sum_{k=0}^{\infty}\frac{e^{-tx}(tx)^k}{k!}\rho_{U_k}(yu)\rho_{V_{t+1}}(y)dy\right)du\\
&=e^{-tx}f(0)+\\
&\quad+\sum_{k=1}^{\infty}s_{t,k}(x)\int\limits_0^\infty f(u)\left(\int\limits_0^\infty \frac{y^{k+t}}{\Gamma(k)}\cdot\frac{u^{k-1}}{\Gamma(t+1)}e^{-y(u+1)}dy\right)du
\end{aligned}$$

$$= e^{-tx}f(0) +$$

$$+\sum_{k=1}^{\infty} s_{t,k}(x)\int_0^{\infty} f(u)\left(\int_0^{\infty}\frac{\left(\frac{v}{u+1}\right)^{k+t}}{\Gamma(k)}\cdot\frac{u^{k-1}}{\Gamma(t+1)}e^{-v}\frac{dv}{u+1}\right)du$$

$$= e^{-tx}f(0) +$$

$$+\sum_{k=1}^{\infty} s_{t,k}(x)\int_0^{\infty} f(u)\frac{b_{t,k}(u)}{\Gamma(k+t+1)}\left(\int_0^{\infty} v^{k+t}e^{-v}dv\right)du$$

$$= e^{-tx}f(0) + \sum_{k=1}^{\infty} s_{t,k}(x)\int_0^{\infty} f(u)b_{t,k}(u)du = L_t(f;x).$$

On the other hand, the Szasz-Inverse Beta operators (1.1)-(1.5) can be represented as the composition between Szasz-Mirakjan operators and Inverse-Beta operators:

$$L_t(f;x) = (S_t \circ T_t)(f;x) = S_t(T_t)(f;x),\ t > 0,\ x \geq 0 \tag{1.9}$$

with the Szasz-Mirakjian operators

$$S_t(f;x) = E\left[f\left(\frac{N_{tx}}{t}\right)\right] = \sum_{k=0}^{\infty} s_{t,k}(x)f\left(\frac{k}{t}\right) \text{ with (1.2)} \tag{1.10}$$

and the Inverse-Beta operators or the Stancu operators of second kind [19]:

$$\begin{cases} T_t(f;x) = E\left[f\left(W_{tx,t+1}\right)\right] \\ \qquad\qquad = \dfrac{1}{B(tx,t+1)}\int\limits_0^{\infty}\dfrac{u^{tx-1}}{(1+u)^{tx+t+1}}f(u)du \\ \qquad\qquad = \int\limits_0^{\infty} f(u)b_{tx,t+1}(u)du,\ t>0,\ x>0, \\ T_t(f;0) = f(0), \end{cases} \tag{1.11}$$

with $W_{tx,t+1}$ a random variable having the Inverse-Beta distribution with probability density function as

$$b_{tx,t+1}(u) = \frac{1}{B(tx,t+1)}\cdot\frac{u^{tx-1}}{(1+u)^{tx+t+1}},\ t>0,\ x>0,\ u>0 \tag{1.12}$$

and $B(tx,t+1) = \int\limits_0^{\infty}\frac{u^{tx-1}}{(1+u)^{tx+t+1}}du,\ t>0,\ x>0.$

It is known [16. IV.10.(3)] that, if we consider two independent random variables U_{tx}, V_{t+1} having Gamma distribution with probability density function (1.8) for $t := tx$ respectively $t := t+1$, then the probability density function of the ratio $\frac{U_{tx}}{V_{t+1}}$ is $b_{tx,t+1}(u) = \int\limits_0^{\infty} y\rho_{U_{tx}}(uy)\rho_{V_{t+1}}(y)dy$ a Inverse-Beta probability density function as (1.12).

Remark 1.1. The Inverse-Beta probability density function can be represented with a negative binomial probability for $t>0$ and with convention $\binom{t}{k}=\frac{t(t-1)(t-2)\cdots(t-k+1)}{k!}$, $t>0$, $k\in\mathbb{N}$, we have

$$\begin{aligned} p_{t,k-1}(u) &= \binom{t+k}{k-1}\left(\frac{u}{1+u}\right)^{k-1}\left(\frac{1}{1+u}\right)^{t+2} \\ &= \binom{t+k}{k-1}\cdot\frac{u^{k-1}}{(1+u)^{t+k+1}}, \end{aligned} \tag{1.13}$$

$t>0$, $u>0$, $k\in\mathbb{N}$, for which $\int\limits_0^\infty p_{t,k-1}(u)du=\dfrac{1}{t+1}$ and so

$$\begin{aligned} b_{t,k}(u) &= \frac{1}{B(k,t+1)}\cdot\frac{u^{k-1}}{(1+u)^{t+k+1}} \\ &= (t+1)p_{t,k-1}(u)=\frac{p_{t,k-1}(u)}{\int\limits_0^\infty p_{t,k-1}(u)du}. \end{aligned} \tag{1.14}$$

The probability density function (1.5) becomes the kernel:

$$\begin{aligned} J_t(u;x) &= e^{-tx}\delta(u)+\sum_{k=1}^\infty s_{t,k}(x)b_{t,k}(u) \\ &= e^{-tx}\delta(u)+(t+1)\sum_{k=1}^\infty s_{t,k}(x)p_{t,k-1}(u) \end{aligned}$$

and the operators (1.1) have a Durrmeyer-type construction

$$\begin{aligned} L_t(f;x) &= e^{-tx}f(0)+\sum_{k=1}^\infty s_{t,k}(x)\frac{\int\limits_0^\infty p_{t,k-1}(u)f(u)du}{\int\limits_0^\infty p_{t,k-1}(u)du} \\ &= e^{-tx}f(0)+(t+1)\sum_{k=1}^\infty s_{t,k}(x)\int_0^\infty p_{t,k-1}(u)f(u)du. \end{aligned} \tag{1.15}$$

Using the representation (1.9) and the images of the test functions $e_i(x)=x^i$, $i=0,1,2$, $x\geq 0$ with these operators (1.10) and (1.11)-(1.12), it is easy to prove that

$$\begin{aligned} L_t\left(e_i;x\right) &= e_i(x),\ i=\overline{0,1},\ x\geq 0; \\ L_t\left(e_2;x\right) &= \frac{t}{t-1}x^2+\frac{2}{t-1}x,\ t>1,\ x\geq 0; \\ L_t\left(e_2-x^2;x\right) &= L_t\left((e_1-x)^2;x\right)=D^2\left[\frac{U_{N(tx)}}{V_{t+1}}\right] \\ &= E\left[\left(\frac{U_{N(tx)}}{V_{t+1}}-x\right)^2\right]=\frac{x(2+x)}{t-1},\ t>1,\ x\geq 0. \end{aligned} \tag{1.16}$$

2. Approximation properties of Szasz-Inverse Beta operators

In view of (1.9) because a part of the properties of Szasz-Inverse Beta operators depends on the same properties of Szasz-Mirakjan operators (1.10) and of the Inverse-Beta operators (1.11)-(1.12), next time, using a probabilistic method which was presented in [1], we studied [4] the monotonic convergence under convexity for the Szasz-Inverse Beta operators (1.1)-(1.5):

Theorem 2.1. *Let $t > 1$ be fixed. For the Szasz-Inverse Beta operators (1.1)-(1.5) following:*

1. $L_t\left(e_i; x\right) = e_i(x)$, $i = \overline{0,1}$*;*
2. $L_t\left(e_2; x\right) = \dfrac{t}{t-1}x^2 + \dfrac{2}{t-1}x$*;*
3. *If f is a convex function on $(0, +\infty)$ then $L_t f$ is convex too and in addition, f is nondecreasing then for $1 < r < s$, $L_r f \geq L_s f \geq f$;*
4. *If $f \in Lip_{(0,+\infty)}(C, \alpha)$, $\alpha \in (0, 1]$ then $L_t f \in Lip_{(0,+\infty)}(C, \alpha)$, $\alpha \in (0, 1]$.*

The proof is immediately [4] using the following two lemmas:

Lemma 2.2. *If $\left(U_{tx}\right)_{t>0,\, x\geq 0}$, $\left(V_{t+1}\right)_{t>0}$ are two independent Gamma processes defined on the same probability space, then for all $1 < r \leq s$ and $x > 0$ we have*

$$E\left(\frac{U_{rx}}{V_{r+1}} \mid \frac{U_{sx}}{V_{s+1}}\right) = \frac{U_{sx}}{V_{s+1}}\ a.\,s.$$

Lemma 2.3. *Let $t > 1$ be fixed. For the Inverse-Beta operators (1.11)-(1.12) following:*

1. *If f is a real convex function on $(0, +\infty)$ then $T_t f$ is convex too.*
2. *If f is a nondecreasing and convex function on $(0, +\infty)$and $1 < r < s$ then $T_r f \geq T_s f \geq f$.*
3. *If $f \in Lip_{(0,+\infty)}(C, \alpha)$, $\alpha \in (0, 1]$ then $T_t f \in Lip_{(0,+\infty)}(C, \alpha)$, $\alpha \in (0, 1]$*

Theorem 2.4. *For any function $f \in \mathbf{C}_B\left[0, +\infty\right)$ and for any compact set $K \subset [0, +\infty)$ we have $\lim\limits_{t\to\infty} L_t(f) = f$ uniform on K.*

Proof. It follows from the Bohmann-Korovkin's theorem and from Theorem 2.1. □

In the next theorem we give in 1 and 2 an approximation using the modulus of continuity of f and of derivative f' and in 3 an asymptotic approximation of Voronovskaja type.

Theorem 2.5. **1.** *If $f \in \mathbf{C}_B\left[0, +\infty\right)$, then for every $x \in [0, +\infty)$*

$$\left|L_t(f; x) - f(x)\right| \leq \left(1 + \sqrt{x(2+x)}\right)\omega\left(f; \frac{1}{\sqrt{t-1}}\right),\ t > 1.$$

2. *If $f' \in \mathbf{C}_B[0,+\infty)$, then for every $x \in [0,+\infty)$*

$$|L_t(f;x) - f(x)| \leq$$
$$\leq \sqrt{\frac{x(2+x)}{t-1}}\left(1+\sqrt{x(2+x)}\right)\omega\left(f'; \frac{1}{\sqrt{t-1}}\right),\ t>1.$$

3. *If f is bounded on $[0,+\infty)$, differentiable in some neighborhood of x and has second derivative f" for some $x \in [0,+\infty)$, then for $t>1$*

$$\lim_{t\to\infty}(t-1)\left[L_t(f;x) - f(x)\right] = \frac{x(2+x)}{2}f''(x).$$

If $f \in \mathbf{C}^2[0,+\infty)$, then the convergence is uniform on any compact $K \subset [0,+\infty)$.

Proof. For 1 and 2 see (1.16) and a result of Shisha O., Mond B., [18] and for 3 see Cismaşiu C. [3]. □

Remark 2.6. An interesting result which was obtained by De la Cal J., Carcamo J., [7] for the operators of Bernstein-type which preserves the affine functions, namely centered Bernstein-type operators, can be used for Szasz-Inverse Beta operators (1.1)-(1.5) :

Theorem 2.7 (De la Cal J., Carcamo J., [7]**).** *If $L_1 = L_2 \circ L_3$, where L_1, L_2, L_3 are centered Bernstein-type operators ($Lf(x) = E\left[f\left(Y_x\right)\right],\ x \in I \subset \mathbb{R},\ L_1(x) = E\left[Y_x\right] = x$) over the same interval I and if $\mathbf{L}_{cx}$ is the set of all convex functions in the domain of the three operators, then $L_1 f \geq L_2 f,\ f \in \mathbf{L}_{cx}$.*

If, in addition L_3 preserves convexity, then $L_1 f \geq L_2 f \vee L_3 f,\ f \in \mathbf{L}_{cx}$ where $f \vee g$ denotes the maximum of f and g.

In view of this result and using the representation (1.9) for Szasz-Inverse-Beta operators, we have $L_t f \geq S_t f,\ f \in \mathbf{L}_{cx}[0,+\infty)$ and $L_t f \geq S_t f \vee T_t f,\ f \in \mathbf{L}_{cx}[0,+\infty)$, where S_t are the Szasz-Mirakjan operators (1.10), T_t are the Inverse-Beta operators (1.11)-(1.12) and L_t are the Szasz-Inverse Beta operators (1.1)-(1.5).

An estimate of the difference $|L_t(f;x) - S_t(f,x)|$ was given by us in [6]:

Theorem 2.8. *If $f \in \mathbf{C}_B[0,+\infty) \cap \mathbf{L}_{cx}[0,+\infty)$ then for every $x \in [0,+\infty)$ and $t>1$*

$$|L_t(f;x) - S_t(f,x)| \leq \left(1+\delta^{-2}\left(\frac{x(x+1)}{t-1} + \frac{x}{t(t-1)}\right)\right)\omega(f,\delta)$$

with $\omega(f,\delta) = \sup\{|f(x) - f(y)| : x,y \geq 0,\ |x-y| \leq \delta\}$ the modulus of continuity of f.

Using the probabilistic representation of these operators, result for $t>1$, $\delta>0$

$$\left|E\left[f\left(\frac{U_{N(tx)}}{V_{t+1}}\right)\right] - E\left[f\left(\frac{N(tx)}{t}\right)\right]\right| \leq$$
$$\leq \left(1+\delta^{-2}\left(D^2\left(\frac{U_{tx}}{V_{t+1}}\right) + \frac{1}{t-1}D^2\left(\frac{N(tx)}{t}\right)\right)\right)\omega(f,\delta)$$

3. Approximating Phillips operators by modified Szasz-Inverse Beta operators

Using the same ideea as De la Cal J. , Luquin F. [8] or as Adell J. A., De la Cal J. [2] , we consider a new operator defined as the aid of Szasz-Inverse Beta operator (1.1)-(1.5) for $r > 0$, $t > 0$, $x \geq 0$:

$$\begin{aligned}\Theta_{r,t}(f;x) &= L_{rt}\left(f(tu);\frac{x}{t}\right)=\int_0^\infty \frac{1}{t}J_{rt}\left(\frac{u}{t};\frac{x}{t}\right)f(u)du \qquad (3.1)\\ &= \int_0^\infty \frac{1}{t}\left[e^{-rx}\delta\left(\frac{u}{t}\right)+\sum_{k=1}^\infty s_{rt,k}\left(\frac{x}{t}\right)b_{rt,k}\left(\frac{u}{t}\right)\right]f(u)du\\ &= e^{-rx}f(0)+\sum_{k=1}^\infty s_{r,k}(x)\int_0^\infty \frac{1}{t}b_{rt,k}\left(\frac{u}{t}\right)f(u)du\end{aligned}$$

where f is any real function defined on $[0,\infty)$ such that $\Theta_{r,t}(|f|);x) < \infty$.

We obtain for the operators (3.1) a Durrmeyer-type construction in a similar way as for representation (1.15) with (1.14) for the Szasz-Inverse Beta operators (1.1)-(1.5):

$$\begin{aligned}\Theta_{r,t}(f;x) &= L_{rt}\left(f(tu);\frac{x}{t}\right)=\int_0^\infty \frac{1}{t}J_{rt}\left(\frac{u}{t};\frac{x}{t}\right)f(u)du \qquad (3.2)\\ &= \int_0^\infty \frac{1}{t}\left[e^{-rx}\delta\left(\frac{u}{t}\right)+\sum_{k=1}^\infty s_{rt,k}\left(\frac{x}{t}\right)b_{rt,k}\left(\frac{u}{t}\right)\right]f(u)du\\ &= e^{-rx}f(0)+\\ &\quad +\left(r+\frac{1}{t}\right)\sum_{k=1}^\infty s_{r,k}(x)\int_0^\infty p_{rt,k-1}\left(\frac{u}{t}\right)f(u)du.\end{aligned}$$

and a probabilistic representation

$$\Theta_{r,t}(f;x)=L_{rt}\left(f(tu);\frac{x}{t}\right)=E\left[f\left(t\frac{U_{N(rx)}}{V_{rt+1}}\right)\right] \qquad (3.3)$$

These operators $\Theta_{r,t}(f;\cdot)$ approximate the Phillips' operators [14] defined as

$$\begin{aligned}P_r(f;x) &= E\left[f\left(\frac{U_{N(rx)}}{r}\right)\right] \qquad (3.4)\\ &= e^{-rx}f(0)+r\sum_{k=1}^\infty s_{r,k}(x)\int_0^\infty s_{r,k-1}(u)f(u)du\\ &= \int_0^\infty H_r(u;x)f(u)du,\ r>0,\ x\geq 0,\end{aligned}$$

with $s_{r,k}(x)$ as (1.2) ,

$$H_r(u;x) \quad = \quad e^{-rx}\delta(u) + r\sum_{k=1}^{\infty} s_{r,k}(x)s_{r,k-1}(u) \tag{3.5}$$

$x \geq 0$, $k \in \mathbb{N}\cup\{0\}$, $r > 0$, δ the Dirac's Delta function and for $f : [0,\infty) \longrightarrow \mathbb{R}$ any integrabile function, such that $P_r(|f|;x) < \infty$.

The Phillips operators (3.4)-(3.5) were studied by several authors (see [9],[12], [13], [14]) and are considered "the genuine Durrmeyer-Szasz-Mirakjan operators". A generalization of these operators, using two continuous parameters was obtained by Păltănea R. [15].

Theorem 3.1. *Let $x \geq 0$, $r, t, u > 0$ be. If, f is a real bounded function on $[0,\infty)$ then*

$$\begin{aligned} |\Theta_{r,t}(f;x) - P_r(f;x)| \quad &= \quad \left|L_{rt}\left(f(tu);\frac{x}{t}\right) - P_r(f;x)\right| \\ &\leq \quad \|f\| \cdot \frac{r^2x^2 + 4rx + 2}{rt+1} \end{aligned}$$

and we have uniform convergence as $t \to \infty$ on every bounded interval $[0,a]$, $a > 0$.

Proof. We presented in detail the proof in [5] and we gave a bound for the total variation distance between the probability distributions of the random variables $t\dfrac{U_{N(rx)}}{V_{rt+1}}$ and $\dfrac{U_{N(rx)}}{r}$, respectively between $\left|\dfrac{1}{t}b_{rt,k}\left(\dfrac{u}{t}\right) - rs_{r,k-1}(u)\right|$. □

References

[1] Adell, J., De la Cal, J., San Miguel, M., *Inverse Beta and generalized Bleimann Butzer-Hahn operators*, J. Approx. Theory, **76**(1994), 54-64.

[2] Adell, J.A., De la Cal, J., *Bernstein-Durrmeyer operators*, Computers Math. Applic., **30**(1995), no. 3-6, 3-6.

[3] Cismaşiu, S.C., *Probabilistic interpretation of Voronovskaja's theorem*, Bull. Univ. Brasov, Seria C, **XXVII**(1985), 7-12.

[4] Cismaşiu, C., *Szasz-inverse Beta operators revisited* (to appear).

[5] Cismaşiu, S.C., *Approximating Phillips operators by modified Szasz-Inverse Beta operators*, Bull. of the Transilvania Univ., Braşov, **3**(2010), Series Mathematics, Informatics, Physics, 19-26.

[6] Cismaşiu, S.C., *An estimate of difference between the Szasz-Inverse Beta operators and the Szasz-Mirakjian operators*, Rev. of the Air Force Academy, Brasov, **17**(2010), no. 2, 67-70.

[7] De la Cal, J., Carcamo, J., *On the approximation of convex functions by Bernstein-type operators*, J. Math. Anal. Appl., **334**(2007), 1106-1115.

[8] De la Cal, J., Luquin, F., *Approximation Szasz and Gamma operators by Baskakov operators*, J. Math. Anal. Appl., **184**(1994), 585-593.

[9] Finta, Z., Gupta, V., *Direct and inverse estimates for Phillips type operators*, J. Math. Anal. Appl., **303**(2005), no. 2, 627-642.

[10] Finta, Z., Govil, N.K., Gupta, V., *Some results on modified Szasz-Mirakjan operators*, J. Math. Anal. Appl., **327**(2007), 1284-1296.

[11] Gupta, V., Noor, M.A., *Convergence of derivatives for certain mixed Szasz-Beta operators*, J. Math. Anal. Appl., **321**(2006), no. 1, 1-9.

[12] Mazhar, S.M., Totik, V., *Approximation by modified Szasz operator*, Acta Sci. Math., **49**(1985), 257-263.

[13] May, C.P., *On Phillips operators*, J. Approx. Theory, **20**(1977), no. 4, 315-332.

[14] Phillips, R.S, *An inversion formula for Laplace transforms and semi-groups of operators*, Annals of Mathmatics, Second Series, **59**(1954), 325-356.

[15] Păltănea, R., *Modified Szasz-Mirakjan operators of integral form*, Carpathian J. Math., **24**(2008), no. 3, 378-385.

[16] Renyi, A., *Probability theory*, Akad. Kiado, Budapest, 1970.

[17] Skorohod, A.V., *Random processes with independent increments*, Kluwer, London, 1986.

[18] Shisha, O., Mond, B., *The degree of convergence of linear positive operators*, Proc. Nat. Acad. Sci. U.S.A., **60**(1968), 1196-1200.

[19] Stancu, D.D., *On the Beta-approximating operators of second kind*, Revue d'Analyse Num. et de Theorie de l'Approx., **24**(1995), no. 1-2, 231-239.

Cristina S. Cismaşiu
"Transilvania" University
Department of Mathematics
Eroilor 29, 500 036, Braşov, Romania
e-mail: `ccismasiu@yahoo.com`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 315–325

Approximation by max-product Lagrange interpolation operators

Lucian Coroianu and Sorin G. Gal

Abstract. The aim of this note is to associate to the Lagrange interpolatory polynomials on various systems of nodes (including the equidistant and the Jacobi nodes), continuous piecewise rational interpolatory operators of the so-called max-product kind, uniformly convergent to the function f, with Jackson-type rates of approximation.

Mathematics Subject Classification (2010): 41A05, 41A25, 41A35.

Keywords: Nonlinear Lagrange interpolation operators of max-product kind, equidistant nodes, Jacobi nodes, degree of approximation.

1. Introduction

Based on the Open Problem 5.5.4, pp. 324-326 in [12], in a series of recent 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 non-truncated case), see [3], Baskakov operators (truncated and nontruncated case), Meyer-König and Zeller operators, see [4] and Bleimann-Butzer-Hahn operators, see [5].

For example, in the two recent papers [1], [2], starting from the linear Bernstein operators

$$B_n(f)(x) = \sum_{k=0}^{n} b_{n,k}(x) f(k/n),$$

where $b_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}$, written in the equivalent form

$$B_n(f)(x) = \frac{\sum_{k=0}^{n} b_{n,k}(x) f(k/n)}{\sum_{k=0}^{n} b_{n,k}(x)}$$

and then replacing the sum operator Σ by the maximum operator $\bigvee$, one obtains the nonlinear Bernstein operator of max-product kind

$$B_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} b_{n,k}(x) f\left(\frac{k}{n}\right)}{\bigvee_{k=0}^{n} b_{n,k}(x)},$$

where the notation $\bigvee_{k=0}^{n} b_{n,k}(x)$ means $max\{b_{n,k}(x); k \in \{0, ..., n\}\}$ and similarly for the numerator.

For this max-product operator nice approximation and shape preserving properties were found in e.g. [2].

In other two recent papers [9] and [10], this idea is applied to the Lagrange interpolation based on the Chebyshev nodes of second kind plus the endpoints, and to the Hermite-Fejér interpolation based on the Chebyshev nodes of first kind respectively, obtaining max-product interpolation operators which, in general, (for example, in the class of positive Lipschitz functions) approximates essentially better than the corresponding Lagrange and Hermite-Fejér interpolation polynomials.

The aim of the present paper is to use the same idea (but slightly modified to simplify the calculation) in the case of the linear interpolation polynomials of Lagrange type on general nodes. Applications to Lagrange interpolation based on equidistant knots and on the roots of orthogonal polynomials, including the Jacobi roots, are obtained.

Thus, let $I \subset \mathbb{R}$ be a bounded or unbounded interval, $f : I \to \mathbb{R}$, $x_{n,k} \in I$, $k \in \{0, ..., n\}$, $x_{n,0} < x_{n,1} < ... < x_{n,n}$, and consider the Lagrange interpolation polynomial of degree $\leq n$ attached to f and to the nodes $(x_{n,k})_k$,

$$P_n(f)(x) = \sum_{k=0}^{n} p_{n,k}(x) f(x_{n,k}),$$

with

$$p_{n,k}(x) = \frac{(x - x_{n,0})...(x - x_{n,k-1})(x - x_{n,k+1})...(x - x_{n,n})}{(x_{n,k} - x_{n,0})...(x_{n,k} - x_{n,k-1})(x_{n,k} - x_{n,k+1})...(x_{n,k} - x_{n,n})}.$$

It is well known that $\sum_{k=0}^{n} p_{n,k}(x) = 1$, for all $x \in \mathbb{R}$, which allows us to write

$$P_n(f)(x) = \frac{\sum_{k=0}^{n} p_{n,k}(x) f(x_{n,k})}{\sum_{k=0}^{n} p_{n,k}(x)}, \text{ for all } x \in I.$$

Therefore, its corresponding max-product interpolation operator will be obtained by replacing the sum operator Σ, by the maximum operator $\bigvee$, that is

$$P_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} p_{n,k}(x) f(x_{n,k})}{\bigvee_{k=0}^{n} p_{n,k}(x)}, x \in I.$$

By the property $p_{n,k}(x_{n,j}) = 1$ if $k = j$ and $p_{n,k}(x_{n,j}) = 0$ if $k \neq j$, we immediately obtain that $P_n^{(M)}(f)(x_{n,j}) = f(x_{n,j})$, for all $j \in \{0, ..., n\}$.

But because this max-product operator seems to present some difficulties in calculations, in this paper we deal with a simplified max-product operator with good approximation properties and which keeps the interpolation properties, given by

$$L_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} l_{n,k}(x) f(x_{n,k})}{\bigvee_{k=0}^{n} l_{n,k}(x)}, x \in I,$$

where

$$l_{n,k}(x) = c_{n,k} \cdot p_{n,k}(x) = (-1)^{n-k} \Pi_{i=0}^{n} (x - x_{n,i})/(x - x_{n,k}) \tag{1.1}$$

and

$$c_{n,k} = (x_{n,k} - x_{n,0})...(x_{n,k} - x_{n,k-1})(x_{n,k+1} - x_{n,k})...(x_{n,n} - x_{n,k}) > 0.$$

The plan of the paper goes as follows. In Section 2 we present some auxiliary results while in Section 3 we prove the approximation results for the max-product Lagrange interpolation operators on equidistant and Jacobi nodes.

2. Auxiliary results

Let us define the space

$$CB_+(I) = \{f : I \to \mathbb{R}_+; f \text{ is continuous and bounded on } I\}.$$

Remark. Firstly, it is clear that $L_n^{(M)}(f)(x)$ is a well-defined function for all $x \in \mathbb{R}$ and it is continuous on $\mathbb{R}$. Indeed, by $\sum_{k=0}^{n} p_{n,k}(x) = 1$, for all $x \in \mathbb{R}$, for any x there exists an index $k \in \{0, ..., n\}$ such that $p_{n,k}(x) > 0$ (which implies that $\bigvee_{k=0}^{n} p_{n,k}(x) > 0$), because contrariwise would follow that $p_{n,k}(x) \leq 0$ for all k and therefore we would obtain the contradiction $\sum_{k=0}^{n} p_{n,k}(x) \leq 0$. Therefore, as $l_{n,k}(x) = c_{n,k} \cdot p_{n,k}(x)$ with $c_{n,k} > 0$, for this k we also have $\bigvee_{k=0}^{n} l_{n,k}(x) > 0$.

Also, by the obvious property $l_{n,k}(x_{n,j}) = c_{n,j} > 0$ if $k = j$ and $l_{n,k}(x_{n,j}) = 0$ if $k \neq j$, we immediately obtain that $L_n^{(M)}(f)(x_{n,j}) = f(x_{n,j})$, for all $j \in \{0, ..., n\}$. In addition, clearly we have $L_n^{(M)}(e_0)(x) = 1$, where $e_0(x) = 1$, for all $x \in I$.

In what follows we will see that for $f \in CB_+[a, b]$, the $L_n^{(M)}(f)$ operator fulfils similar properties with those of the $B_n^{(M)}(f)$ operator in [1].

Lemma 2.1. *Let $I \subset \mathbb{R}$ be a bounded or unbounded interval.*

(i) Then $L_n^{(M)} : CB_+(I) \to CB_+(I)$, for all $n \in \mathbb{N}$:

(ii) If $f, g \in CB_+(I)$ satisfy $f \leq g$ then $L_n^{(M)}(f) \leq L_n^{(M)}(g)$ for all $n \in N$;

(iii) $L_n^{(M)}(f+g) \leq L_n^{(M)}(f) + L_n^{(M)}(g)$ *for all* $f, g \in CB_+(I)$ *;*
(iv) For all $f, g \in CB_+(I)$*,* $n \in N$ *and* $x \in I$ *we have*

$$|L_n^{(M)}(f)(x) - L_n^{(M)}(g)(x)| \leq L_n^{(M)}(|f-g|)(x);$$

(v) $L_n^{(M)}$ *is positive homogenous, that is* $L_n^{(M)}(\lambda f) = \lambda L_n^{(M)}(f)$ *for all* $\lambda \geq 0$ *and* $f \in CB_+(I)$*.*

Proof. (i) The continuity of $L_n^{(M)}(f)(x)$ on I follows from the previous Remark. Also, by the formula of definition for $L_n^{(M)}(f)(x)$, if f is bounded on I, then it easily follows that $L_n^{(M)}$ is bounded on I. It remains to prove the positivity of $L_n^{(M)}(f)$. So let $f : I \to \mathbb{R}_+$ and fix $x \in I$. Reasoning exactly as in the above Remark, there exists $k \in \{0, 1, ..., n\}$ such that $l_{n,k}(x) > 0$. Therefore, denoting $I_n^+(x) = \{k \in \{0, 1, ..., n\}; l_{n,k}(x) > 0\}$, clearly $I_n^+(x)$ is nonempty and for $f \in CB_+(I)$ we get that

$$L_n^{(M)}(f)(x) = \frac{\bigvee_{k \in I_n^+(x)} l_{n,k}(x) f(x_{n,k})}{\bigvee_{k \in I_n^+(x)} l_{n,k}(x)} \geq 0. \tag{2.1}$$

(ii) Let $f, g \in CB_+(I)$ be with $f \leq g$ and fix $x \in I$. Since $I_n^+(x)$ is independent of f and g, by (2.1) we immediately obtain $L_n^{(M)}(f)(x) \leq L_n^{(M)}(g)(x)$.

(iii) By (2.1) and by the sublinearity of $\bigvee$, it is immediate.

(iv) Let $f, g \in CB_+(I)$. We have $f = f - g + g \leq |f - g| + g$, which by $(i) - (iii)$ successively implies $L_n^{(M)}(f)(x) \leq L_n^{(M)}(|f-g|)(x) + L_n^{(M)}(g)(x)$, that is $L_n^{(M)}(f)(x) - L_n^{(M)}(g)(x) \leq L_n^{(M)}(|f-g|)(x)$.

Writing now $g = g - f + f \leq |f - g| + f$ and applying the above reasonings, it follows $L_n^{(M)}(g)(x) - L_n^{(M)}(f)(x) \leq L_n^{(M)}(|f-g|)(x)$, which combined with the above inequality gives $|L_n^{(M)}(f)(x) - L_n^{(M)}(g)(x)| \leq L_n^{(M)}(|f-g|)(x)$.

(v) By (2.1) it is immediate. □

Remark. By (2.1) it is easy to see that instead of (ii), $L_n^{(M)}$ satisfies the stronger condition

$$L_n(f \vee g)(x) = L_n(f)(x) \vee L_n(g)(x), \ f, g \in CB_+(I).$$

Corollary 2.2. *For all* $f \in CB_+(I)$*,* $n \in N$ *and* $x \in I$ *we have*

$$|f(x) - L_n^{(M)}(f)(x)| \leq \left[\frac{1}{\delta} L_n^{(M)}(\varphi_x)(x) + 1\right] \omega_1(f; \delta)_I,$$

where $\delta > 0$*,* $\varphi_x(t) = |t - x|$ *for all* $t \in I$*,* $x \in I$ *and* $\omega_1(f; \delta)_I = \max\{|f(x) - f(y)|; x, y \in I, |x - y| \leq \delta\}$*.*

Proof. Indeed, denoting $e_0(x) = 1$, from the identity

$$L_n^{(M)}(f)(x) - f(x) = [L_n^{(M)}(f)(x) - f(x) \cdot L_n^{(M)}(e_0)(x)] + f(x)[L_n^{(M)}(e_0)(x) - 1],$$

by Lemma 2.1 it easily follows

$$|f(x) - L_n^{(M)}(f)(x)| \leq$$

$$|L_n^{(M)}(f(x))(x) - L_n^{(M)}(f(t))(x)| + |f(x)| \cdot |L_n^{(M)}(e_0)(x) - 1| \leq$$
$$L_n^{(M)}(|f(t) - f(x)|)(x) + |f(x)| \cdot |L_n^{(M)}(e_0)(x) - 1|.$$

Now, since for all $t, x \in I$ we have

$$|f(t) - f(x)| \leq \omega_1(f; |t-x|)_I \leq \left[\frac{1}{\delta}|t-x| + 1\right] \omega_1(f; \delta)_I,$$

replacing above and taking into account that $L_n^{(M)}(e_0) = 1$, for all $x \in I$, we immediately obtain the estimate in the statement. □

Remark. The results in Lemma 2.1 and Corollary 2.2 remain valid if we replace the space $CB_+(I)$ by the space

$$C_+(I) = \{f : I \to \mathbb{R}_+; f \text{ is continuous on } I\}.$$

3. Approximation results for max-product Lagrange interpolation

In this section we study the approximation properties of the max-product operators $L_n^{(M)}$.

It is clear that for the approximation purpose, in the case of the operator $L_n^{(M)}$, from Corollary 2.2 it is enough to obtain a good estimate for the expression

$$E_n(x) := L_n^{(M)}(\varphi_x)(x) = \frac{\bigvee_{k=0}^{n} l_{n,k}(x)\,|x_{n,k} - x|}{\bigvee_{k=0}^{n} l_{n,k}(x)} = \frac{\bigvee_{k \in I_n^+(x)} l_{n,k}(x)|x_{n,k} - x|}{\bigvee_{k \in I_n^+(x)} l_{n,k}(x)}.$$

We present the first main approximation result.

Theorem 3.1. *Given the nodes $-\infty < a \leq x_{n,0} < x_{n,1} < ... < x_{n,n} \leq b < \infty$, $f \in C_+([a,b])$ and denoting*

$$d_n = \max\{x_{n,0} - a, \max\{x_{n,k+1} - x_{n,k}; k = 0, 1, ..., n-1\}, b - x_{n,n}\},$$

we have

$$|L_n^{(M)}(f)(x) - f(x)| \leq 2\omega_1(f; d_n)_{[a,b]}, \text{ for all } x \in [a,b],$$

where $\omega_1(f; \delta)_{[a,b]} = \sup\{|f(x) - f(y)|; x, y \in [a,b], |x-y| \leq \delta\}$.

Proof. Firstly, because $L_n^{(M)}(f)(x_{n,j}) = f(x_{n,j})$, for all $j \in \{0, 1, ..., n\}$, in all calculations and estimations we may suppose that $x \neq x_{n,j}$, for all $j \in \{0, 1, ..., n\}$.

Denote $\Omega_n(x) = \Pi_{i=0}^n (x - x_{n,i})$. It is easy to see that for any $x \in [a,b]$, with $x \neq x_{n,j}$, $j \in \{0, 1, ..., n\}$, we can write

$$E_n(x) = \frac{\bigvee_{k \in I_n^+(x)} l_{n,k}(x)|x_{n,k} - x|}{\bigvee_{k \in I_n^+(x)} l_{n,k}(x)} = \frac{|\Omega_n(x)|}{\bigvee_{k \in I_n^+(x)} l_{n,k}(x)} = \frac{1}{\bigvee_{k \in I_n^+(x)} \frac{1}{|x - x_{n,k}|}}$$

$$= \min\{|x - x_{n,k}|; k \in I_n^+(x)\}.$$

Denote $x_{n,-1} := a$ and $x_{n,n+1} := b$ and fix $j \in \{-1, 0, ..., n, n+1\}$. We have three possibilities : 1) $j = -1$; 2) $0 \le j \le n-1$; 3) $j = n$. Let $x \in (x_{n,j}, x_{n,j+1})$.

Case 1). We may suppose that $a < x_{n,0}$. We have $l_{n,0}(x) > 0$ for all $x \in [a, x_{n,0})$. Indeed, by using (1.1) we easily get that for $x \in [a, x_{n,0})$, we have $sign[l_{n,0}(x)] = (-1)^n \cdot (-1)^n = +1$. Therefore $0 \in I_n^+(x)$, for all $x \in [a, x_{n,0})$. We also get $|x - x_{n,0}| \le |x - x_{n,k}|$, for all $k \in I_n^+(x)$ and $x \in [a, x_{n,0})$, which implies $E_n(x) = |x - x_{n,0}| = x_{n,0} - x \le x_{n,0} - a \le d_n$, for all $x \in [a, x_{n,0})$.

Case 2). We have $l_{n,j}(x) > 0$ and $l_{n,j+1}(x) > 0$ for all $x \in (x_{n,j}, x_{n,j+1})$. Indeed, by using (1.1) we easily get that for $x \in (x_{n,j}, x_{n,j+1})$, we have $sign[l_{n,j}(x)] = (-1)^{n-j} \cdot (-1)^{n-j} = +1$ and $sign[l_{n,j+1}(x)] = (-1)^{n-j-1} \cdot (-1)^{n-j-1} = +1$. Therefore $j, j+1 \in I_n^+(x)$, for all $x \in (x_{n,j}, x_{n,j+1})$.

We also get $|x - x_{n,j}| \le |x - x_{n,k}|$ for all $k \in \{0.1, ..., j\}$ and $|x - x_{n,j+1}| \le |x - x_{n,k}|$ for all $k \in \{j+1, j+2, ..., n\}$, which implies $E_n(x) = \min\{|x - x_{n,j}|, |x - x_{n,j+1}|\} \le \frac{d_n}{2}$, for all $x \in (x_{n,j}, x_{n,j+1})$.

Case 3). We may suppose that $x_{n,n} < b$. We have $l_{n,n}(x) > 0$ for all $x \in (x_{n,n}, b]$. Indeed, by using (1.1) we easily get that for $x \in (x_{n,n}, b]$, we have $sign[l_{n,n}(x)] = (-1)^0 \cdot (-1)^0 = +1$. Therefore $n \in I_n^+(x)$, for all $x \in (x_{n,0}, b]$. We also get $|x - x_{n,n}| \le |x - x_{n,k}|$, for all $k \in I_n^+(x)$ and $x \in (x_{n,n}, b]$, which implies $E_n(x) = |x - x_{n,n}| = x - x_{n,n} \le b - x_{n,n} \le d_n$, for all $x \in (x_{n,n}, b]$.

Collecting all the above estimates and applying Corollary 2.2, the theorem is proved. □

Remark. The order of approximation in terms of $\omega_1(f; d_n)_{[a,b]}$ in Theorem 3.1 cannot be improved, in the sense that it easily follows from the proof of Theorem 3.1, that the estimate $E_n(x) \le \mathcal{O}(d_n)$ cannot be improved.

As applications we obtain the following two results.

Corollary 3.2. *(i) Let $I = [a, b]$, $f \in C_+([a, b])$ and the equidistant knots in $I = [a, b]$, $x_{n,k} = a + kh$, $k \in \{0, ..., n\}$, with $h = (b - a)/n$. Then we have*

$$|L_n^{(M)}(f)(x) - f(x)| \le 2\omega_1\left(f; \frac{b-a}{n}\right)_{[a,b]}, \text{ for all } x \in [a, b].$$

(ii) Let $w(x)$ be a weight function on the finite interval $I = [a, b]$, satisfying $w(x) \ge \nu > 0$, for all $x \in [a, b]$. If $a < x_{n,0} < x_{n,1} < ... < x_{n,n} < b$ are the the zeros of the associated orthonormal polynomial $p_{n+1}(x)$ of degree $\le n+1$, then for any $f \in C_+([a, b])$ we have

$$|L_n^{(M)}(f)(x) - f(x)| \le C\omega_1\left(f; \frac{ln(n+1)}{n+1}\right)_{[a,b]}, \text{ for all } x \in [a, b],$$

where $C > 0$ is a constant depending only on ν, a and b.

(iii) Let $w(x)$ be a weight function on the interval $I = [-1, 1]$, satisfying $A \le \sqrt{1 - x^2}w(x) \le B$, for all $x \in [-1, 1]$, where $A, B > 0$ are constants. If $-1 < x_{n,0} < x_{n,1} < ... < x_{n,n} < 1$ are the the zeros of the associated orthonormal polynomial $p_{n+1}(x)$ of degree $\le n+1$, then for any $f \in C_+([-1, 1])$

we have

$$|L_n^{(M)}(f)(x)-f(x)|\le C\omega_1\left(f;\frac{1}{n+1}\right)_{[-1,1]}, \text{ for all } x\in[-1,1],$$

where $C>0$ is a constant depending only on A and B.

(iv) If $-\frac{1}{2}\le\alpha\le+\frac{1}{2}$, $-\frac{1}{2}\le\beta\le+\frac{1}{2}$ and $-1<x_{n,0}<x_{n,1}<...<x_{n,n}<1$ are the the zeros of the associated orthonormal Jacobi polynomial $J_{n+1}(x)$ of degree $\le n+1$, associated to the weight $w(x)=(1-x)^{\alpha}(1+x)^{\beta}$, then for any $f\in C_+([-1,1])$ we have

$$|L_n^{(M)}(f)(x)-f(x)|\le C\omega_1\left(f;\frac{1}{n+1}\right)_{[-1,1]}, \text{ for all } x\in[-1,1],$$

where $C>0$ is a constant depending only on α and β.

Proof. (i) It is immediate from Theorem 3.1 for $d_n=\frac{b-a}{n}$.

(ii) It follows from Theorem 3.1, taking into account that by Theorem 6.11.1, pp. 112-113 in [18], we have $d_n\le c\frac{ln(n+1)}{n+1}$, with $c>0$ depending on ν,a and b only.

(iii) It follows from Theorem 3.1, taking into account that by Theorem 6.11.2, p. 114 in [18], we have $d_n\le c\frac{1}{n+1}$, with $c>0$ depending on A and B only.

(iv) It follows from Theorem 3.1, taking into account that by Theorem 6.3.1, p. 125 in [18], we have $d_n\le c\frac{1}{n+1}$, with $c>0$ depending on α and β only. □

It is of interest to have a more explicit form for the operator $L_n(f)(x)$ in Theorem 3.1. In this sense we present the following.

Theorem 3.3. *Given $f\in C_+([a,b])$ and the nodes $-\infty<a\le x_{n,0}<x_{n,1}<...<x_{n,n}\le b<\infty$, the max-product operator $L_n^{(M)}(f)(x)$ is continuous on $[a,b]$, $L_n^{(M)}(f)(x_{n,j})=f(x_{n,j})$ for all $j\in\{0,1,...,n\}$ and we can write :*

$$L_n^{(M)}(f)(x)=\bigvee_{k=0}^{n}(-1)^k\frac{x-x_{n,0}}{x-x_{n,k}}f(x_{n,k}),\ \text{for } x\in[a,x_{n,0}),$$

$$L_n^{(M)}(f)(x)$$
$$=\bigvee_{k=0}^{n}(-1)^{j-k}\frac{x-x_{n,j}}{x-x_{n,k}}f(x_{n,k}), x\in(x_{n,j},(x_{n,j}+x_{n,j+1})/2], j=\overline{0,n-1},$$

$$L_n^{(M)}(f)(x)$$
$$=\bigvee_{k=0}^{n}(-1)^{j+1-k}\frac{x-x_{n,j+1}}{x-x_{n,k}}f(x_{n,k}), x\in[(x_{n,j}+x_{n,j+1})/2,x_{n,j+1}),$$

$$j=\overline{0,n-1},$$

$$L_n^{(M)}(f)(x) = \bigvee_{k=0}^{n} (-1)^{n-k} \frac{x - x_{n,n}}{x - x_{n,k}} f(x_{n,k}), \ \textit{for } x \in (x_{n,n}, b].$$

Proof. The continuity and the interpolation properties were already established by the Remark from the beginning of Section 2. In order to get the rest of the statement in the theorem, it suffices to prove the following formulas :

$$\bigvee_{k\in I_n^+(x)} l_{n,k}(x) = l_{n,0}(x), \text{ for } x \in [a, x_{n,0}),$$

$$\bigvee_{k\in I_n^+(x)} l_{n,k}(x) = l_{n,j}(x), \text{ for } x \in (x_{n,j}, (x_{n,j} + x_{n,j+1})/2], j = \overline{0, n-1},$$

$$\bigvee_{k\in I_n^+(x)} l_{n,k}(x) = l_{n,j+1}(x), \text{ for } x \in [(x_{n,j} + x_{n,j+1})/2, x_{n,j+1}), j = \overline{0, n-1},$$

$$\bigvee_{k\in I_n^+(x)} l_{n,k}(x) = l_{n,n}(x), \text{ for } x \in (x_{n,n}, b].$$

We have three cases : 1) $x \in [a, x_{n,0})$; 2) $x \in (x_{n,j}, x_{n,j+1})$, $j \in \{0, 1, ..., n-1\}$; 3) $x \in (x_{n,n}, b]$.

Case 1). By the proof of Theorem 3.1, Case 1), we have $l_{n,0}(x) > 0$, for $x \in [a, x_{n,0})$. Also, for any $k \in I_n^+(x)$, we have

$$\frac{l_{n,0}(x)}{l_{n,k}(x)} = \frac{x_{n,k} - x}{x_{n,0} - x} \geq 1.$$

Case 2). Let $j \in \{0, 1, ..., n-1\}$ be fixed. By the proof of Theorem 3.1, Case 2), we have $l_{n,j}(x) > 0$ and $l_{n,j+1}(x) > 0$, for $x \in (x_{n,j}, x_{n,j+1})$. We have

$$\frac{l_{n,j}(x)}{l_{n,j+1}(x)} = \frac{x_{n,j+1} - x}{x - x_{n,j}}.$$

Therefore, for any $x \in (x_{n,j}, (x_{n,j} + x_{n,j+1})/2]$ we have $l_{n,j}(x) \geq l_{n,j+1}(x)$ and for any $x \in [(x_{n,j} + x_{n,j+1})/2, x_{n,j+1})$ we have $l_{n,j+1}(x) \geq l_{n,j}(x)$.

Let $k \in I_n^+(x)$. If $k \leq j$ then

$$\frac{l_{n,j}(x)}{l_{n,k}(x)} = \frac{x - x_{n,k}}{x - x_{n,j}} \geq 1$$

and if $k \geq j + 1$ then

$$\frac{l_{n,j+1}(x)}{l_{n,k}(x)} = \frac{x_{n,k} - x}{x_{n,j+1} - x} \geq 1.$$

Case 3). By the proof of Theorem 3.1, Case 3), we have $l_{n,n}(x) > 0$, for $x \in (x_{n,n}, b]$. Also, in this case, for any $k \in I_n^+(x)$, we have

$$\frac{l_{n,n}(x)}{l_{n,k}(x)} = \frac{x - x_{n,k}}{x - x_{n,n}} \geq 1$$

and the theorem is proved. □

In what follows, would be of interest to compare the approximation results for the max-product Lagrange interpolation operators, with their linear counterparts. Thus, in the case of Lagrange interpolatory polynomials, it is well-known the fact that the divergence phenomenon is very pronounced.

In this sense, let us briefly recall some results (for details, see e.g. Chapter 4 in the book Szabados-Vértesi [17]). Thus, Bernstein [6] proved that for $f(x) = |x|$, the Lagrange interpolatory polynomials attached to the system of equidistant nodes in $[-1,1]$ does not converge to $f(x)$, for any $x \in (-1,1) \setminus \{0\}$. Grümwald [13] and independently Marcinkiewicz [15], proved that when the system of interpolation nodes consists in the Chebyshev nodes of the first kind, there exists a function $f \in C([-1,1])$ such that for the attached Lagrange interpolatory polynomials $L_n(f)(x)$, we have $\limsup_{n\to\infty} |L_n(f)(x)| = +\infty$, for all $x \in [-1,1]$. More general, a similar result holds for the system of Jacobi nodes in $[-1,1]$ (see the book Szabados-Vértesi [17], relationship (4.1), p. 126). For an arbitrary system of interpolation nodes in $[-1,1]$, in Erdös-Vértesi [11] it is proved that there exists a function $f \in C([-1,1])$, such that for the attached Lagrange interpolatory polynomials we have $\limsup_{n\to\infty} |L_n(f)(x)| = +\infty$, almost everywhere $x \in [-1,1]$. By using the condensation singularities principle in Functional Analysis, Muntean [16], Cobzas-Muntean [8] proved that for any system of nodes in $[0,1]$, there exists a superdense subset $X_0 \subset C([0,1])$, such that for any $f \in X_0$, the subset of divergence points in $[0,1]$ for the attached Lagrange interpolatory polynomials $L_n(f)(x)$, is superdense in $[0,1]$ (a countable intersection of open subsets which, in addition, is infinite, uncountable and dense subset, is called superdense).

In contrast with these results, the results in Theorem 3.1 and Corollary 3.2 show that for the max-product interpolatory operator $L_n^{(M)}(f)(x)$, the situation is essentially better, having uniform convergence with good rates of convergence for some of the most important systems of interpolation nodes.

Let us note that on the other hand, in Hermann-Vértesi [14], starting from a Lagrange interpolatory process (convergent or not)

$$P_n(f)(x) = \sum_{k=0}^{n} p_{n,k}(x) f(x_{n,k}),$$

with

$$p_{n,k}(x) = \frac{(x - x_{n,0})...(x - x_{n,k-1})(x - x_{n,k+1})...(x - x_{n,n})}{(x_{n,k} - x_{n,0})...(x_{n,k} - x_{n,k-1})(x_{n,k} - x_{n,k+1})...(x_{n,k} - x_{n,n})},$$

new linear interpolatory rational operators are constructed, of the form

$$R_n(f)(x) = \frac{\sum_{k=0}^{n} f(x_{n,k})|p_{n,k}(x)|^r}{\sum_{k=0}^{n} f(x_{n,k})|p_{n,k}(x)|^r},$$

are constructed, for which in the case when $r > 2$ and $x_{n,k}$ are some Jacobi knots, the Jackson-type order of approximation

$$\|R_n(f) - f\| \le C\omega_1(f; 1/n),$$

is obtained (see Theorem 3.2 in Hermann-Vértesi [14]).

In other words, for the linear rational construction $R_n(f)(x)$, we get the same order of approximation as for the interpolatory rational max-product operator in Theorem 3.1 of the form

$$L_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} p_{n,k}(x) f\left(x_{n,k}\right)}{\bigvee_{k=0}^{n} p_{n,k}(x)}.$$

Clearly that with respect to $R_n(f)(x)$, the max-product rational operator $L_n^{(M)}(f)(x)$ present the advantage that it provides an estimate in terms of $\omega_1(f; 1/n)$ for any kind of interpolatory systems of points, with the properties that the distance between two consecutive nodes converges to zero as $n \to \infty$.

But still it is an interesting open problem, a comparison from computational point of view, between a rational max-product type product like that given by Theorem 3.1 (that is of the form $L_n^{(M)}(f)(x)$) and the linear rational one like $R_n(f)(x)$ mentioned above.

References

[1] Bede, B., Gal, S.G., *Approximation by nonlinear Bernstein and Favard-Szász-Mirakjan operators of max-product kind*, Journal of Concrete and Applicable Mathematics, **8**(2010), no. 2, 193–207.

[2] 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., vol. 2009, Article ID 590589, 26 pages, 2009. doi:10.1155/2009/590589.

[3] Bede, B., Coroianu, L., Gal, S.G., *Approximation by truncated Favard-Szasz-Mirakjan operator of max-product kind*, Demonstratio Mathematica (accepted for publication).

[4] Bede, B., Coroianu, L., Gal, S.G., *Approximation and shape preserving properties of the nonlinear Bleimann-Butzer-Hahn operators of max-product kind*, Comment. Math. Univ. Carol., **51**(2010), no. 3, 397–415.

[5] Bede, B., Coroianu, L., Gal, S.G., *Approximation and shape preserving properties of the nonlinear Meyer-Konig and Zeller operator of max-product kind*, Numerical Functional Analysis and Optimization, **31**(2010), no. 3, 232–253.

[6] Bernstein, S., *Quelques remarques sur l'interpolation*, Math. Ann., **79**(1918), no. 1-2, 1–12.

[7] Borel, E., *Sur l'interpolation*, C.R. Acad. Sci. Paris, **124**(1897), 673–676.

[8] Cobzas, S., Muntean, I., *Condensation of singularities and divergence results in approximation theory*, J. Approx. Theory, **31**(1980), no. 2, 138–153.

[9] Coroianu, L., Gal, S.G., *Approximation by nonlinear Lagrange interpolation operators of max-product kind on Chebyshev knots of second kind*, J. Comp. Anal. Appl., 2011 (accepted for publication).

[10] Coroianu, L., Gal, S.G., *Approximation by nonlinear Hermite-Fejér interpolation operators of max-product kind on Chebyshev nodes*, Revue d'Anal. Numér. Théor. Approx. (Cluj), **39**(2010), no. 1, 29–39.

[11] Erdös, P., Vértesi, P., *On the almost everywhere divergence of Lagrange interpolatory polynomials for arbitrary system of nodes*, Acta Math. Acad. Sci. Hung., **36**(1980), no. 1-2, 71–89.

[12] Gal, S.G., *Shape-Preserving Approximation by Real and Complex Polynomials*, Birkhäuser, Boston-Basel-Berlin, 2008.

[13] Grünwald, G., *Über Divergenzerscheinungen der Lagrangeschen Interpolationspolynome stetiger Functionen*, Ann. Math., **37**(1936), 908–918.

[14] Hermann, T., Vértesi, P., *On the method of Somorjai*, Acta Math. Hung., **54**(1989), no. 3-4, 253–262.

[15] Marcinkiewicz, J., *Sur la divergence des polynômes d'interpolation*, Acta Sci. Math. (Szeged), **8**(1937), 131–135.

[16] Muntean, I., *The Lagrange interpolation operators are densely divergent*, Studia Univ. Babeş-Bolyai Math., **21**(1976), 28–30.

[17] Szabados, J., Vértesi, P., *Interpolation of Functions*, World Scientific, Singapore, New Jersey, London, Hong Kong, 1990.

[18] Szegö, G., *Orthogonal Polynomials*, American Mathematical Society, Colloquium Publications, Vol. **23**, Providence, Rhode Island, 2003.

Lucian Coroianu
Department of Mathematics and Computer Science
The University of Oradea
Universitatii 1, 410087, Oradea
Romania
e-mail: `lcoroianu@uoradea.ro`

Sorin G. Gal
Department of Mathematics and Computer Science
The University of Oradea
Universitatii 1, 410087, Oradea
Romania
e-mail: `galso@uoradea.ro`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 327–334

Almost greedy uniformly bounded orthonormal bases in rearrangement invariant Banach function spaces

Ana Danelia and Ekaterine Kapanadze

Abstract. We construct uniformly bounded orthogonal almost greedy bases in rearrangement invariant Banach spaces.

Mathematics Subject Classification (2010): 41A17, 42C40.

Keywords: Greedy algorithm, non-linear approximation.

1. Introduction

Let $\{x_n\}_{n\in\mathbb{N}}$ be a semi-normalized basis in a Banach space X. This means that $\{x_n\}_{n\in\mathbb{N}}$ is a Schauder basis and is semi-normalized i.e. $0 < \inf_{n\in\mathbb{N}} \|x_n\| \leq \sup_{n\in\mathbb{N}} \|x_n\| < \infty$. For an element $x \in X$ we define the error of the best $m-$term approximation as follows

$$\sigma_m(x) = \inf\{\|x - \sum_{n\in A} \alpha_n x_n\|\},$$

where the inf is taken over all subsets $A \subset \mathbb{N}$ of cardinality at most m and all possible scalars α_n. The main question in approximation theory concerns the construction of efficient algorithms for m-term approximation. A computationally efficient method to produce m-term approximations, which has been widely investigated in recent years, is the so called greedy algorithm. We define the greedy approximation of $x = \sum_n a_n x_n \in X$ as

$$\mathcal{G}_m(x) = \sum_{n\in A} a_n x_n,$$

where $A \subset \mathbb{N}$ is any set of the cardinality m in such a way that $|a_n| \geq |a_l|$ whenever $n \in A$ and $l \overline{\in} A$. We say that a semi-normalized basis $\{x_n\}_{n\in\mathbb{N}}$ is

The authors was supported by grant GNSF/ST08/3-393.

greedy if there exists a constant C such that for all $m = 1, 2, ...$ and all $x \in X$ we have

$$\|x - \mathcal{G}_m(x)\| \leq C\sigma_m(x).$$

This notion evolved in theory of non-linear approximation (see e.g.[1],[2]). A result of Konyagin and Temlyakov [3] characterizes greedy bases in a Banach spaces X as those which are unconditional and democratic, the latter meaning that for some constant $C > 0$

$$\left\| \sum_{\alpha \in A} \frac{x_\alpha}{\|x_\alpha\|} \right\| \leq C \left\| \sum_{\alpha \in A'} \frac{x_\alpha}{\|x_\alpha\|} \right\|$$

holds for all finite sets of indices A, $A' \subset \mathbb{N}$ with the same cardinality.

Wavelet systems are well known examples of greedy bases for many function and distribution spaces. Indeed, Temlyakov showed in [1] that the Haar system is greedy in the Lebesgye spaces $L^p(\mathbb{R}^n)$ for $1 < p < \infty$. When wavelets have sufficient smoothness and decay, they are also greedy bases for the more general Sobolev and Tribel-Lizorkin classes (see e.g.[4-5]).

A bounded Schauder basis for a Banach space X is called quasi-greedy if there exists a constant C such that for $x \in X$ $\|\mathcal{G}_m(x)\| \leq C\|x\|$ for $m \geq 1$.

Wojtaszczyk [2] proved the following result which gives a more intuitive interpretation of quasi-greedy bases.

Theorem 1.1. *A bounded Schauder basis for a Banach space X is quasi-greedy if and only if* $\lim_{m\to\infty} \|x - \mathcal{G}_m(x)\|_X = 0$ *for every element $x \in X$.*

A bounded Schauder basis for a Banach space X is almost greedy if there exists a constant C such that for $x \in X$, $\|x - \mathcal{G}_m(x)\| \leq C \inf\{\|x - \sum_{n \in A} < x, x_n > x_n\| : A \subset \mathbb{N},\ |A| = m\}$.

It was proved in [6] that a basis is almost greedy if and only if it is quasi-greedy and democratic.

A Banach function space on $[0, 1]$ is said to be a rearrangement invariant (r.i) space provided $f^*(t) \leq g^*(t)$ for every $t \in [0, 1]$ and $g \in X$ imply $f \in X$ and $\|f\|_X \leq \|g\|_X$, where $f^*(t)$ denotes the decreasing rearrangement of $|f|$.

An r.i. space X with a norm $\|\cdot\|_X$ has the Fatou property if for any increasing positive sequence f_n in X with $\sup_n \|f_n\|_X < \infty$ we have that $\sup_n f_n \in X$ and $\|\sup_n f_n\|_X = \sup_n \|f_n\|_X$. We will assume that the r.i. space X has the Fatou property.

Given $s > 0$, the dilation operator σ_s given by

$$\sigma_s f(t) = f(t/s)\chi_{[0,1]}(t/s),\ t \in [0, 1]$$

(χ_A denotes the characteristic function of a measurable set $A \subset [0, 1]$) is well defined in every r.i. space X. The classical Boyd indices of X are defined by

$$p_X = \lim_{s\to\infty} \frac{\ln s}{\ln \|\sigma_s\|_{X\to X}},\ q_X = \lim_{s\to 0+} \frac{\ln s}{\ln \|\sigma_s\|_{X\to X}}.$$

In general, $1 \leq p_X \leq q_X \leq \infty$.

Any r.i. function space X on $[0,1]$ satisfies $L^\infty([0,1]) \subset X \subset L^1([0,1])$. If we have information on the Boyd indices of X then a stronger assertion is valid. Indeed for every $1 \le p < p_X$ and $q_X < q < \infty$, we have

$$L^q([0,1]) \subset X \subset L^p([0,1]) \tag{1.1}$$

with the inclusion maps being continuous. Let X' denote the associate Banach function space of X. Then X' is a r.i. Banach function space whose Boyd indices are defined as $1/p_X + 1/q_{X'} = 1$ and $1/q_X + 1/p_{X'} = 1$ (see [7]).

M. Nielsen in [8] proved that there exists a uniformly bounded orthonormal almost greedy basis in $L^p([0,1])$, $1 < p < \infty$, that shows that it is not possible to extend Orlicz's theorem, stating that there are no uniformly bounded orthonormal unconditional bases for $L^p([0,1])$, $p \neq 2$, to the class of almost greedy bases.

The purpose of this paper is to study these problems in the r.i. function spaces. Namely, the following theorem is obtained.

Theorem 1.2. *Let X be a separable r.i. Banach function space on $[0,1]$ and $1 < p_X \le q_X < 2$ or $2 < p_X \le q_X < \infty$. Then there exists a uniformly bounded orthogonal almost greedy basis in X.*

2. Proof of theorem

Let us construct some system in the following way. For $k = 1, 2, \dots$, we define the $2^k \times 2^k$ Olevskii matrix $A^k = (a_{ij}^{(k)})_{i,j=1}^{2^k}$ by the following formulas

$$a_{i1}^k = 2^{-\frac{k}{2}} \quad \text{for} \quad i = 1, 2, \dots, 2^k,$$

and for $j = 2^s + \nu$, with $1 \le \nu \le 2^s$ and $s = 0, 1, \dots, k-1$, we let

$$a_{ij}^{(k)} = \begin{cases} 2^{\frac{s-k}{2}} & \text{for} \quad (\nu-1)2^{k-s} < i \le (2\nu-1)2^{k-s-1} \\ -2^{\frac{s-k}{2}} & \text{for} \quad (2\nu-1)2^{k-s-1} < i \le \nu 2^{k-s} \\ 0 & \text{otherwise.} \end{cases}$$

It is known [16] that A^k are orthogonal matrices and there exists a finite constant C such that for all i, k we have

$$\sum_{j=1}^{2^k} |a_{i,j}^{(k)}| \le C.$$

Put $N_k = 2^{10^k}$ and define F_k such that $F_0 = 0$, $F_1 = N_1 - 1$ and $F_k - F_{k-1} = N_k - 1$, $k = 1, 2, \dots$. We consider the Walsh system $\mathcal{W} = \{W_n\}_{n=0}^\infty$ on $[0,1]$. We split $\mathcal{W}$ into two subsystems. The first subsystem $\mathcal{W}_1 = \{r_k\}_{k=1}^\infty$ is Rademacher functions with their natural ordering. The second subsystem $\mathcal{W}_2 = \{\phi_k\}_{k=1}^\infty$ is the collection of Walsh functions not in $\mathcal{W}_1$ with the ordering from $\mathcal{W}$. We now impose the ordering

$$\phi_1, r_1, r_2, \dots, r_{F_1}, \phi_2, r_{F_1+1}, \dots, r_{F_2}, \phi_3, r_{F_2+1}, \dots, r_{F_3}, \phi_4, \dots$$

The block $\mathcal{B}_k := \{\phi_k, r_{F_{k-1}+1}, ..., r_{F_k}\}$ has length N_k and we apply A^{10^k} to $\mathcal{B}_k$ to obtain a new orthonormal system $\{\psi_i^{(k)}\}_{i=1}^{N_k}$ given by

$$\psi_i^{(k)} = \frac{\phi_k}{\sqrt{N_k}} + \sum_{j=2}^{N_k} a_{ij}^{(10^k)} r_{F_{k-1}+j-1}.$$

The system ordered $\psi_1^{(1)}, ..., \psi_{N_1}^{(1)}, \psi_1^{(2)}, ..., \psi_{N_2}^{(2)}, ...$ will be denoted by $\mathcal{B} = \{\psi_k\}_{k=1}^{\infty}$. It is easy to verify that $\mathcal{B}$ is an orthonormal basis for L_2 since each matrix A^{10^k} is orthogonal and it is uniformly bounded also.

Lemma 2.1. *Let X be a r.i. Banach function space on $[0,1]$ and $1 < p_X \leq q_X < \infty$. The system $\mathcal{B} = \{\psi_k\}_{k=1}^{\infty}$ is democratic in X with*

$$\|\sum_{k\in A} \psi_k\|_X \asymp |A|^{\frac{1}{2}}.$$

Proof. Taking into account that fact that $B\|\cdot\|_{p_X} \leq \|\cdot\|_X \leq C\|\cdot\|_{q_X}$ and the estimate (see [8])

$$\|\sum_{k\in A} \psi_k\|_p \asymp |A|^{\frac{1}{2}} \text{ for any } 1 < p < \infty$$

we obtain our result. □

Lemma 2.2. *(Khintchine's inequality)Suppose that X is a r.i. Banach function space on $[0,1]$, $1 < p_X \leq q_X < \infty$,and $r_k(t), k \geq 1$, are the Rademacher functions. Then there exist A, B such that for any sequence $\{a_k\}_{k\geq 1}$,*

$$A(\sum_k |a_k|^2)^{\frac{1}{2}} \leq \|\sum_k a_k r_k(t)\|_X \leq B(\sum_k |a_k|^2)^{\frac{1}{2}}.$$

Proof. It is known that (see [10]) for $1 \leq p < \infty$ there exist A_p, B_p such that for any sequence $\{a_k\}_{k\geq 1}$,

$$A_p(\sum_k |a_k|^2)^{\frac{1}{2}} \leq \|\sum_k a_k r_k(t)\|_p \leq B_p(\sum_k |a_k|^2)^{\frac{1}{2}}.$$

Taking into account that fact that $B\|\cdot\|_{q_X} \leq \|\cdot\|_X \leq C\|\cdot\|_{p_X}$ and the above inequality we obtain Lemma 2.2. □

Lemma 2.3. *Suppose that X is a r.i. Banach function space on $[0,1]$, $1 < p_X \leq q_X < \infty$, and $r_k(t), k \geq 1$, are the Rademacher functions. Then for $f \in X$ we have*

$$(\sum_{k=1}^{\infty} |< f, r_k >|^2)^{\frac{1}{2}} \leq C\|f\|_X.$$

Proof. For any $n \geq 1$ by the Hölder inequality and Khintchine's inequality we obtain

$$\sum_{k=1}^{n} |< f, r_k >|^2 = \int_0^1 f(x)(\sum_{k=1}^{n} r_k(x) < f, r_k >)dx \leq$$
$$2\|\sum_{k=1}^{n} < f, r_k > r_k\|_{X'}\|f\|_X \leq C(\sum_{k=1}^{n} |< f, r_k >|^2)^{1/2}\|f\|_X.$$

This implies

$$(\sum_{k=1}^{n} | < f, r_k > |^2)^{\frac{1}{2}} \leq B\|f\|_X.$$

Now taking the limit when $n \to \infty$ we obtain our result. □

Lemma 2.4. *Let X be a separable r.i. Banach function space on $[0,1]$ and $1 < p_X \leq q_X < \infty$. Then the system $\mathcal{B} = \{\psi_k\}_{k=1}^{\infty}$ is a Schauder basis for X.*

Proof. Notice that $span(\mathcal{B}) = span(\mathcal{W})$ by construction, so $span(\mathcal{B})$ is dense in X, since $\mathcal{W}$ is a Schauder basis for X (see [11]).

Let $S_n(f) = \sum_{k=1}^{n} < f, \psi_k > \psi_k$ be the partial sum operator. We need to prove that the family of operators $\{S_n\}_{n=1}^{\infty}$ is uniformly bounded on X. Let $f \in L^\infty([0,1]) \subset L^2([0,1])$. For $n \in \mathbb{N}$ we can find $L \geq 1$ and $1 \leq m \leq N_L$ such that

$$S_n(f) = \sum_{k=1}^{n} < f, \psi_k > \psi_k = \sum_{k=1}^{L-1}\sum_{j=1}^{N_k} < f, \psi_j^{(k)} > \psi_j^{(k)} + \sum_{k=1}^{m} < f, \psi_k^{(L)} > \psi_k^{(L)}$$

$$:= T_1 + T_2.$$

Let us estimate T_1. If $L = 1$ then $T_1 = 0$, so we may assume $L > 1$. The construction of $\mathcal{B}$ shows that T_1 is the orthogonal projection of f onto

$$span\left(\cup_{k=1}^{L-1} \cup_{j=1}^{N_k} \psi_k^{(k)}\right) = span\{\{W_0, W_1, ..., W_{L-2}\} \cup \{r_{l_0}, r_{l_0+1}, ..., r_{F_{L-1}}\}\},$$

with $l_0 = [\log_2(L)]$. It follows that we can rewrite T_1 as

$$T_1 = \sum_{k=0}^{L-2} < f, W_k > W_k + P_R(f),$$

where $P_R(f)$ is the orthogonal projection of f onto $span\{r_{l_0}, r_{l_0+1}, ..., r_{F_{L-1}}\}$. Thus, using the fact that $\mathcal{W}$ is a Schauder basis for X, Khintchine's inequality and Lemma 2.3, we will have

$$\|T_1\|_X \leq C\|f\|_X.$$

Let us now estimate T_2.

$$T_2 = \sum_{k=1}^{m} < f, \psi_k^{(L)} > \psi_k^{(L)}$$

$$= \sum_{k=1}^{m} < f, \frac{\phi_L}{\sqrt{N_L}} + \sum_{j=2}^{N_L} a_{kj}^{(10^L)} r_{F_{L-1}+j-1} >= (\frac{\phi_L}{\sqrt{N_L}}\phi_L + \sum_{t=2}^{N_L} a_{kt}^{(10^L)} r_{F_{L-1}+t-1})$$

$$= \frac{m}{N_L} < f, \phi_L > + \frac{\phi_L}{\sqrt{N_L}} \sum_{j=2}^{N_L}(\sum_{k=1}^{m} a_{kj}^{(10^L)}) < f, r_{F_{L-1}+j-1} >$$

$$+ < f, \frac{\phi_L}{\sqrt{N_L}} > \sum_{j=2}^{N_L}(\sum_{k=1}^{m} a_{kj}^{(10^L)}) r_{F_{L-1}+j-1}$$

$$+\sum_{k=1}^{m}[\sum_{j=2}^{N_L} a_{kj}^{(10^L)} < f, r_{F_{L-1}+j-1} >][\sum_{t=2}^{N_L} a_{kt}^{(10^L)} r_{F_{L-1}+t-1}]$$

$$= G_1 + G_2 + G_3 + G_4.$$

Using that fact that $1 \le m \le N_L$ and Hölder inequality we obtain $\|G_1\|_X \le C\|f\|_X$. Using the Hölder and Khintchine's inequality, the fact that matrices A^k are orthonormal and Lemma 2.3 we obtain $\|G_i\|_X \le C\|f\|_X$ $i = 2, 3, 4$ for some constant C independent of $f \in L^\infty([0,1])$. Consequently for some constant C independent on $f \in L^\infty([0,1])$ we have $\|S_n f\|_X \le C\|f\|_X$. Since $L^\infty([0,1])$ is dense in X we deduce that $\{S_n\}_{n=1}^\infty$ is a uniformly bounded family of linear operators on X and the system $\mathcal{B}$ is a Schauder basis for X. □

Lemma 1.1 and Lemma 2.4 give the following

Theorem 2.5. *Let X be a separable r.i. Banach function space on $[0,1]$ and $1 < p_X \le q_X < \infty$. Then there exists a uniformly bounded orthonormal democratic basis in X.*

Lemma 2.6. *Let X be a separable r.i. Banach function space on $[0,1]$ and $1 < p_X \le q_X < 2$ or $2 < p_X \le q_X < \infty$. Then the system $\mathcal{B} = \{\psi_k\}_{k=1}^\infty$ is a quasi-greedy basis for X.*

Proof. First we consider $2 < p_X \le q_X < \infty$ case. Let $f \in X \subset L_2$. We have

$$f = \sum_{i=1}^{\infty} < f, \psi_i > \psi_i,$$

with $\|\{< f, \psi_i >\}\|_{l_2} \le \|f\|_2 \le C\|f\|_X$. We must prove that $\mathcal{G}_m(f)$ is convergent in X.

Let us formally write

$$f = \sum_{k=1}^{\infty}\sum_{j=1}^{N_k} < f, \psi_j^{(k)} > \psi_j^{(k)}$$

$$= \sum_{k=1}^{\infty}\sum_{j=1}^{N_k} < f, \psi_j^{(k)} > \frac{\phi_k}{\sqrt{N_k}} + \sum_{k=1}^{\infty}\sum_{i=1}^{N_k} < f, \psi_i^{(k)} > \sum_{j=2}^{N_k} a_{ij}^{(10^k)} r_{F_{k-1}+j-1}$$

$$= S_1 + S_2.$$

Consider $\varepsilon_i^k \subset \{0,1\}$. By Kchintchine's inequality and the fact that each A^{10^k} is orthogonal we conclude that S_2 converges unconditionally in X. Indeed

$$\left\|\sum_{k=1}^{\infty}\sum_{j=2}^{N_k}\left(\sum_{i=1}^{N_k}\varepsilon_i^k < f, \psi_i^{(k)} > a_{ij}^{(10^k)}\right) r_{F_{k-1}+j-1}\right\|_X$$

$$\le C\left(\sum_k\sum_{i=1}^{N_k}\varepsilon_i^k| < f, \psi_i^{(k)} > |^2\right)^{1/2}.$$

The series defining S_2 converges unconditionally, so it suffices to prove that the series defining S_1 converges in X when the coefficients $< f, \psi_{>}$ are arranged in decreasing order. Let us consider the sets

$$\Lambda_k^1 = \left\{ j : \frac{1}{N_k} < | < f, \psi_j^{(k)} > | < \frac{1}{N_k^{1/10}} \right\}$$
$$\Lambda_k^2 = \left\{ j : | < f, \psi_j^{(k)} > | \leq \frac{1}{N_k} \right\}$$
$$\Lambda_k^3 = \left\{ j : | < f, \psi_j^{(k)} > | \geq \frac{1}{N_k^{1/10}} \right\}.$$

Then

$$S_1 = \sum_{k=1}^{\infty} \sum_{j \in \Lambda_k^1} < f, \psi_j^{(k)} > \frac{\phi_k}{\sqrt{N_k}} + \sum_{k=1}^{\infty} \sum_{j \in \Lambda_k^2} < f, \psi_j^{(k)} > \frac{\phi_k}{\sqrt{N_k}} +$$
$$\sum_{k=1}^{\infty} \sum_{j \in \Lambda_k^3} < f, \psi_j^{(k)} > \frac{\phi_k}{\sqrt{N_k}} = T_1 + T_2 + T_3.$$

By the construction of sets Λ_k^i we can conclude that the series defining T_2 and T_3 converges absolutely in X.

From the definition of Λ_k^1 we get

$$| < f, \psi_i^{(k)} > | > \frac{1}{N_k} \geq \frac{1}{N_{k+1}^{1/10}} \geq | < f, \psi_j^{(k+1)} > |,$$

$i \in \Lambda_k^1$, $j \in \Lambda_{k+1}^1$, $k = 1, 2, ...$ so when we arrange T_1 by decreasing order the rearrangement can only take place inside the blocks. From the estimate

$$\sum_{j \in \Lambda_k^1} \left\| < f, \psi_j^{(k)} > \frac{\phi_k}{\sqrt{N_k}} \right\|_X \leq \left(\sum_{j \in \Lambda_k^1} | < f, \psi_j^{(k)} > |^2 \right)^{1/2} \frac{|\Lambda_k^1|^{1\backslash 2}}{\sqrt{N_k}}, \ k \geq 1$$

we obtain that the rearrangements inside blocks are well-behaved, and

$$\sum_{j \in \Lambda_k^1} \left\| < f, \psi_j^{(k)} > \frac{\phi_k}{\sqrt{N_k}} \right\|_X \to 0, \ k \to \infty.$$

We can conclude that $\mathcal{G}_m(f)$ is convergent in X.

Using Theorem 1.1 we conclude that $\mathcal{B}$ is a quasi-greedy basis and consequently almost greedy in X.

Let $1 < p_X \leq q_X < 2$. By the results proved above it follows that the system $\mathcal{B}$ is almost greedy in X. From [6, Theorem 5.4] we conclude that $\mathcal{B}$ is quasi-greedy basis and consequently almost greedy in X This completes the proof. $\square$

References

[1] Temlyakov, V.N., *The best $m-$term approximation and greedy algorithms*, Adv. Comput. Math., **8**(1998), 249-265.

[2] Wojtaszczyk, P., *Greedy algorithm for general biorthogonal system*, J. Approx. Theory, **107**(2000), 293-314.

[3] Konyagin, S.V., Temlyakov, V.N., *A remark on greedy approximation in Banach spaces*, East J. Approx., **5**(1999), 1-15.

[4] Garrigos, G., Hernandez, E., *Sharp Jackson and Bernstein inequalities for $N-$term approximation in sequence spaces with applications*, Indiana Univ. Math. J., **53**(2004), 1739-1762.

[5] Hsiao, C., Jawerth, B., Lucier, B.J., Yu, X.M., *Near optimal compression of almost optimal wavelet expansions. Wavelets: mathematics and applications*, Stud. Adv. Math., CRC, Boca Raton, FL, 1994, 425-446.

[6] Dilworth, S.J., Kalton, N.J., Kutzarova, D., Temlyakov, V.N., *The thresholding greedy algorithm, greedy bases and duality*, Constr. Approx., **19**(2003), no. 4, 575-597.

[7] Krein, S., Petunin, J., Semenov, E., *Interpolation of linear operators*, Translations Math. Monographs, vol. 55, Amer. Math. Soc., Providence, RI, (1992).

[8] Nielsen, M., *An example of an almost greedy uniformly bounded orthonormal basis in $L_p[0,1]$*, J. Approx. Theory, **149**(2007), no. 2, 188-192.

[9] Olevskii, A.M., *Fourier series with respect to general orthogonal systems*, Springer-Verlag, New York, 1975.

[10] Kashin, B.S., Saakian, A.A., *Orthogonal series*, (in Russian), Moscow, 1999.

[11] Lindenstraus, J., Tzafriri, L., *Classical Banach spaces II*, Springer-Verlag, Berlin, Heidelberg, New York, 1978.

Ana Danelia and Ekaterine Kapanadze
Faculty of Exact and Natural Sciences
I. Javakhshvili Tbilisi State University 2
University St., Tbilisi 0143
Georgia
e-mail: a.danelia@math.sci.tsu.ge
ekaterine_kapani@yahoo.com

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 335–339

Remark on Voronovskaja theorem for q-Bernstein operators

Zoltán Finta

Abstract. We establish quantitative Voronovskaja type theorems for the q-Bernstein operators introduced by Phillips in 1997. Our estimates are given with the aid of the first order Ditzian-Totik modulus of smoothness.

Mathematics Subject Classification (2010): 41A10, 41A25, 41A36.

Keywords: Voronovskaja theorem, q-integers, q-Bernstein operators, K-functional, first order Ditzian-Totik modulus of smoothness.

1. Introduction

Let $q > 0$ and n be a non-negative integer. Then the q-integers $[n]_q$ and the q-factorials $[n]_q!$ are defined by

$$[n]_q = \begin{cases} 1 + q + \ldots + q^{n-1}, & \text{if} \quad n \geq 1 \\ 0, & \text{if} \quad n = 0 \end{cases}$$

and

$$[n]_q! = \begin{cases} [1]_q [2]_q \ldots [n]_q, & \text{if} \quad n \geq 1 \\ 1, & \text{if} \quad n = 0. \end{cases}$$

For integers $0 \leq k \leq n$, the q-binomial coefficients are defined by

$$\left[\begin{array}{c} n \\ k \end{array} \right]_q = \frac{[n]_q!}{[k]_q! [n-k]_q!}.$$

The so-called q-Bernstein operators were introduced by G.M. Phillips [3] and they are defined by $B_{n,q} : C[0,1] \to C[0,1]$,

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

where

$$p_{n,k}(q,x) = \left[\begin{array}{c} n \\ k \end{array} \right]_q x^k(1-x)(1-qx)\dots(1-q^{n-k-1}x), \quad x \in [0,1],$$

and an empty product denotes 1. Note that for $q = 1$, we recover the classical Bernstein operators. It is well-known that Voronovskaja's theorem [5] deals with the asymptotic behaviour of Bernstein operators. Then naturally raises the following question: can we state a similar Voronovskaja theorem for the q-Bernstein operators? The positive answer was given in [3] as follows.

Theorem 1.1. *Let $q = q_n$ satisfy $0 < q_n < 1$ and let $q_n \to 1$ as $n \to \infty$. If f is bounded on $[0,1]$, differentiable in some neighborhood of x and has second derivative $f''(x)$ for some $x \in [0,1]$, then the rate of convergence of the sequence $\{(B_{n,q_n}f)(x)\}$ is governed by*

$$\lim_{n\to\infty} [n]_{q_n} \{(B_{n,q_n}f)(x) - f(x)\} = \frac{1}{2}x(1-x)f''(x). \tag{1.1}$$

In [4], the convergence (1.1) was given in quantitative form as follows.

Theorem 1.2. *Let $q = q_n$ satisfy $0 < q_n < 1$ and let $q_n \to 1$ as $n \to \infty$. Then for any $f \in C^2[0,1]$ the following inequality holds*

$$\left| \, [n]_{q_n} \{(B_{n,q_n}f)(x) - f(x)\} - \frac{1}{2}x(1-x)f''(x) \, \right| \le c\, x(1-x)\, \omega\left(f'', [n]_{q_n}^{-1/2}\right),$$

where c is an absolute positive constant, $x \in [0,1]$, $n = 1, 2, \dots$ and ω is the first order modulus of continuity.

The goal of this note is to obtain new quantitative Voronovskaja type theorems for the q-Bernstein operators. Our results will be formulated with the aid of the first order Ditzian-Totik modulus of smoothness (see [1]), which is given for $f \in C[0,1]$ by

$$\omega_\varphi^1(f,\delta) = \sup_{0<h\le\delta} \|\Delta^1_{h\varphi(\cdot)} f(\cdot)\|, \tag{1.2}$$

where $\varphi(x) = \sqrt{x(1-x)}$, $x \in [0,1]$, $\|\cdot\|$ is the uniform norm and

$$\Delta^1_{h\varphi(x)} f(x) = \begin{cases} f\left(x + \frac{1}{2}h\varphi(x)\right) - f\left(x - \frac{1}{2}h\varphi(x)\right), & \text{if} \quad x \pm \frac{1}{2}h\varphi(x) \in [0,1] \\ 0, & \text{otherwise.} \end{cases}$$

Further, the corresponding $K-$functional to (1.2) is defined by

$$K_{1,\varphi}(f,\delta) = \inf\{\|f-g\| + \delta\|\varphi g'\| : g \in W^1(\varphi)\},$$

where $W^1(\varphi)$ is the set of all $g \in C[0,1]$ such that g is absolutely continuous on every interval $[a,b] \subset [0,1]$ and $\|\varphi g'\| < +\infty$. Then, in view of [1, p.11], there exists $C > 0$ such that

$$K_{1,\varphi}(f,\delta) \le C\omega_\varphi^1(f,\delta). \tag{1.3}$$

Here we mention that throughout this paper C denotes a positive constant independent of n and x, but it is not necessarily the same in different cases.

2. Main result

Our result is the following.

Theorem 2.1. *Let $\{q_n\}$ be a sequence such that $0 < q_n < 1$ and $q_n \to 1$ as $n \to \infty$. Then for every $f \in C^2[0,1]$ the following inequalities hold*

$$\begin{aligned} &\left| \, [n]_{q_n} \{(B_{n,q_n} f)(x) - f(x)\} - \frac{1}{2} x(1-x) f''(x) \, \right| \\ &\leq \; C\, \omega_\varphi^1 \left(f'', \sqrt{[n]_{q_n}^{-1} x(1-x)} \right), \end{aligned} \tag{2.1}$$

$$\begin{aligned} &\left| \, [n]_{q_n} \{(B_{n,q_n} f)(x) - f(x)\} - \frac{1}{2} x(1-x) f''(x) \, \right| \\ &\leq \; C \sqrt{x(1-x)}\, \omega_\varphi^1 \left(f'', \sqrt{[n]_{q_n}^{-1}} \right), \end{aligned} \tag{2.2}$$

where $x \in [0,1]$ and $n = 1, 2, \ldots$

Proof. We recall some properties of the q-Bernstein operators (see [3]):

$$B_{n,q_n}(1,x) = 1,\; B_{n,q_n}(t,x) = x,\; B_{n,q_n}(t^2,x) = x^2 + [n]_{q_n}^{-1} x(1-x) \tag{2.3}$$

and B_{n,q_n} are positive.

Let $f \in C^2[0,1]$ be given and $t, x \in [0,1]$. Then, by Taylor's formula, $f(t) = f(x) + f'(x)(t-x) + \int_x^t f''(u)(t-u)\,du$. Hence

$$\begin{aligned} & f(t) - f(x) - f'(x)(t-x) - \frac{1}{2} f''(x)(t-x)^2 \\ &= \int_x^t f''(u)(t-u)\,du - \int_x^t f''(x)(t-u)\,du \\ &= \int_x^t [f''(u) - f''(x)](t-u)\,du. \end{aligned}$$

In view of (2.3), we obtain

$$\begin{aligned} & \left| \, B_{n,q_n}(f,x) - f(x) - \frac{1}{2}[n]_{q_n}^{-1} x(1-x) f''(x) \, \right| \\ &= \left| \, B_{n,q_n} \left(\int_x^t [f''(u) - f''(x)]\,(t-u)\,du, x \right) \right| \\ &\leq B_{n,q_n} \left(\left| \int_x^t |f''(u) - f''(x)|\,|t-u|\,du \right|, x \right). \end{aligned} \tag{2.4}$$

In what follows we estimate $\left| \int_x^t |f''(u) - f''(x)|\,|t-u|\,du \right|$. For $g \in W^1(\varphi)$, we have

$$\begin{aligned}
&\Big| \int_x^t |f''(u)-f''(x)|\,|t-u|\,du \Big| \\
\le\ & \Big| \int_x^t |f''(u)-g(u)|\,|t-u|\,du \Big| + \Big| \int_x^t |g(u)-g(x)|\,|t-u|\,du \Big| \\
&+ \Big| \int_x^t |g(x)-f''(x)|\,|t-u|\,du \Big| \\
\le\ & 2\|f''-g\|(t-x)^2 + \left| \int_x^t \Big| \int_x^u |g'(v)|\,dv \Big| \; |t-u|\,du \right| \\
\le\ & 2\|f''-g\|(t-x)^2 + \|\varphi g'\| \left| \int_x^t \Big| \int_x^u \frac{dv}{\varphi(v)} \Big| \; |t-u|\,du \right| \\
\le\ & 2\|f''-g\|(t-x)^2 \\
&+ \|\varphi g'\| \left| \int_x^t \Big| \int_x^u \frac{|u-x|^{1/2}}{\varphi(x)} \frac{dv}{|u-v|^{1/2}} \Big| \; |t-u|\,du \right| \\
=\ & 2\|f''-g\|(t-x)^2 + 2\|\varphi g'\|\varphi^{-1}(x) \left| \int_x^t |u-x|\,|t-u|\,du \right| \\
\le\ & 2\|f''-g\|(t-x)^2 + 2\|\varphi g'\|\varphi^{-1}(x)|t-x|^3, \qquad (2.5)
\end{aligned}$$

where we have used the inequality $\dfrac{|u-v|}{\varphi^2(v)} \le \dfrac{|u-x|}{\varphi^2(x)}$, v is between u and x (see [1, p. 141]).

On the other hand, by [2, p. 440], we have the following property: for any $m = 1, 2, \ldots$ and $0 < q < 1$, there exists a constant $C(m) > 0$ such that

$$|B_{n,q}((t-x)^m, x)| \le C(m)\, \frac{\varphi^2(x)}{[n]_q^{\lfloor (m+1)/2 \rfloor}}, \qquad (2.6)$$

where $\varphi(x) = \sqrt{x(1-x)}$, $x \in [0,1]$ and $\lfloor a \rfloor$ is the integer part of $a \ge 0$ (see also [4, (4.2) and (5.6)]).

Now combining (2.4), (2.5), (2.6) and the Cauchy-Schwarz inequality, we find that

$$\begin{aligned}
&\Big| (B_{n,q_n} f)(x) - f(x) - \frac{1}{2}[n]_{q_n}^{-1} x(1-x) f''(x) \Big| \\
\le\ & 2\|f''-g\|\, B_{n,q_n}((t-x)^2, x) + 2\|\varphi g'\|\, \varphi^{-1}(x)\, B_{n,q_n}(|t-x|^3, x) \\
\le\ & 2\|f''-g\|\, B_{n,q_n}((t-x)^2, x) \\
&+ 2\|\varphi g'\|\, \varphi^{-1}(x) (B_{n,q_n}((t-x)^2, x))^{1/2} (B_{n,q_n}((t-x)^4, x))^{1/2} \\
\le\ & C \left\{ \|f''-g\| \frac{1}{[n]_{q_n}}\, \varphi^2(x) + \|\varphi g'\|\varphi^{-1}(x)\, \frac{\varphi(x)}{[n]_{q_n}^{1/2}}\, \frac{\varphi(x)}{[n]_{q_n}} \right\} \\
=\ & \frac{C}{[n]_{q_n}} \left\{ \|f''-g\|\, \varphi^2(x) + \|\varphi g'\|\, \frac{\varphi(x)}{[n]_{q_n}^{1/2}} \right\}. \qquad (2.7)
\end{aligned}$$

Because $\varphi^2(x) \le \varphi(x) \le 1$, $x \in [0,1]$, we obtain, in view of (2.7),

$$\begin{aligned} &\left| \, [n]_{q_n} \{(B_{n,q_n} f)(x) - f(x)\} - \frac{1}{2} x(1-x) f''(x) \, \right| \\ &\quad \le \quad C \left\{ \|f'' - g\| + \frac{\varphi(x)}{[n]_{q_n}^{1/2}} \|\varphi g'\| \right\} \end{aligned} \tag{2.8}$$

and

$$\begin{aligned} &\left| \, [n]_{q_n} \{(B_{n,q_n} f)(x) - f(x)\} - \frac{1}{2} x(1-x) f''(x) \, \right| \\ &\quad \le \quad C\,\varphi(x) \left\{ \|f'' - g\| + \frac{1}{[n]_{q_n}^{1/2}} \|\varphi g'\| \right\}, \end{aligned} \tag{2.9}$$

respectively. Taking the infimum on the right hand side of (2.8) and (2.9) over all $g \in W^1(\varphi)$, we obtain

$$\left| \, [n]_{q_n} \{(B_{n,q_n} f)(x) - f(x)\} - \frac{1}{2} x(1-x) f''(x) \, \right| \le \begin{cases} C\, K_{1,\varphi}(f'', \varphi(x)[n]_{q_n}^{-1/2}) \\ C\,\varphi(x) K_{1,\varphi}(f'', [n]_{q_n}^{-1/2}). \end{cases}$$

Hence, by (1.3), we find the estimates (2.1) and (2.2). Thus the theorem is proved. □

References

[1] Ditzian, Z., Totik, V., *Moduli of Smoothness*, Springer, New York, 1987.

[2] Mahmudov, N., *The moments for the q-Bernstein operators in the case* $0 < q < 1$, Numer. Algor., **53**(2010), 439–450.

[3] Phillips, G.M., *Bernstein polynomials based on the q-integers*, Ann. Numer. Math., **4**(1997), 511–518.

[4] Videnskii, V.S., *On some classes of q-parametric positive linear operators*, Operator Theory, Advances and Applications, **158**(2005), 213–222.

[5] Voronovskaja, E.V., *Détermination de la forme asymptotique d'approximation des fonctions par les polynômes de M. Bernstein*, Dokl. Akad. Nauk SSSR, **4**(1932), 86–92.

Zoltán Finta
"Babeş-Bolyai" University,
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street,
400084 Cluj-Napoca,
Romania
e-mail: fzoltan@math.ubbcluj.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 341–352

Approximation by max-product type nonlinear operators

Sorin G. Gal

Abstract. The purpose of this survey is to present some approximation and shape preserving properties of the so-called nonlinear (more exactly sublinear) and positive, max-product operators, constructed by starting from any discrete linear approximation operators, obtained in a series of recent papers jointly written with B. Bede and L. Coroianu. We will present the main results for the max-product operators of: Bernstein-type, Favard-Szász-Mirakjan-type, truncated Favard-Szász-Mirakjan-type, Baskakov-type, truncated Baskakov-type, Meyer-König and Zeller-type, Bleimann-Butzer-Hahn-type, Hermite-Fejér interpolation-type on Chebyshev nodes of first kind, Lagrange interpolation-type on Chebyshev knots of second kind, Lagrange interpolation-type on arbitrary knots, generalized sampling-type, sampling sinc-type, Cardaliaguet-Euvrard neural network-type.

Mathematics Subject Classification (2010): 41A30, 41A25, 41A29, 41A20, 41A35, 41A05, 94A20, 94A12, 92B20.

Keywords: Degree of approximation, shape preserving properties, nonlinear max-product operators of: Berstein-type, Hermite-Fejér and Lagrange interpolation-type (on Chebyshev, Jacobi and equidistant nodes), Whittaker (sinc)-type, sampling-type, neural network Cardaliaguet-Euvrard-type.

1. Introduction

The idea of construction of these operators goes back to a paper of Bede, B., Nobuhara, H., Fodor, J. and Hirota K. [11], where it is applied to the rational approximation operators of Shepard. How could be applied to any linear and discrete Bernstein-type operator I have shown in my book Gal [18], pp. 324-326, Open Problem 5.5.4, where also a general form for the estimate in terms of the modulus of continuity is obtained.

The construction is based on a simple idea, exemplified for the case of Bernstein polynomials, as follows.

Let $B_n(f)(x) = \sum_{k=0}^{n} p_{n,k}(x) f(k/n)$ be with $p_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}$ and $f : [0,1] \to \mathbb{R}$. If in the obvious formula

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

we replace the $\sum$ operator with the max operator denoted by $\bigvee$, then we obtain the so-called max-product Bernstein nonlinear (sublinear), piecewise rational operator by (Gal [18], p. 325)

$$B_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} p_{n,k}(x) f(k/n)}{\bigvee_{k=0}^{n} p_{n,k}(x)}, x \in [0,1],$$

where recall

$$\bigvee_{k=0}^{n} p_{n,k}(x) f(k/n) := \max_{0 \le k \le n} \{p_{n,k}(x) f(k/n)\}.$$

The same idea of construction can be applied to any discrete linear Bernstein-type operator or to any discrete linear interpolation operator, obtaining thus the corresponding nonlinear max-product operators (well-defined because the denominators of these new operators always are strictly positive).

Surprisingly, the max-product operators do not lose the approximation properties of the corresponding linear operators to which they are attached. Moreover, for large classes of functions, they improve the order of approximation to the Jackson-type order. The most important improvement is in the case of interpolation (on any arbitrary system of nodes), when for the whole class of continuous functions the Jackson order $\omega_1(f; 1/n)$ is achieved. Also, the max-product Bernstein-type operators preserve the monotonicity and the quasi-convexity of the functions.

In this survey we will present the main results for the max-product operators of: Bernstein-type, Favard-Szász-Mirakjan-type, truncated Favard-Szász-Mirakjan-type, Baskakov-type, truncated Baskakov-type, Meyer-König and Zeller-type, Bleimann-Butzer-Hahn-type, Hermite-Fejér interpolation-type on Chebyshev nodes of first kind, Lagrange interpolation-type on Chebyshev knots of second kind, Lagrange interpolation-type on arbitrary knots, generalized sampling-type, sampling sinc-type, Cardaliaguet-Euvrard neural network-type.

2. Approximation by max-product operators of Bernstein-type

Denote

$$C_+[0,1] = \{f : [0,1] \to \mathbb{R}_+; f \text{ is continuous on } [0,1]\}.$$

This section contains the approximation and shape preserving properties for a series of important max-product Bernstein-type operators.

Theorem 2.1. *For $f \in C_+[0,1]$, define the max-product Bernstein operator by (Gal* [18]*, p. 325)*

$$B_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} p_{n,k}(x) f(k/n)}{\bigvee_{k=0}^{n} p_{n,k}(x)}, x \in [0,1].$$

(i) (Bede-Coroianu-Gal [4]*) For any $j \in \{0,1,...,n\}$ and $x \in [\frac{j}{n+1}, \frac{j+1}{n+1}]$ we have*

$$B_n^{(M)}(f)(x) = \bigvee_{k=0}^{n} f_{k,n,j}(x).$$

where $f_{k,n,j}(x) = \frac{\binom{n}{k}}{\binom{n}{j}} \left(\frac{x}{1-x}\right)^{k-j} f\left(\frac{k}{n}\right)$. This form suggested the denomination of "max-product" operator for $B_n^{(M)}$ (that is the maximum of the product of the values of f on nodes with some rational functions).

(ii) (Bede-Coroianu-Gal [4]*) $B_n^{(M)}(f)(x)$ is a continuous, piecewise convex and piecewise rational function on* $[0,1]$.

(iii) (Bede-Coroianu-Gal [4]*) For all $x \in [0,1]$, $n \in \mathbb{N}$ we have*

$$|B_n^{(M)}(f)(x) - f(x)| \leq 12\omega_1\left(f; \frac{1}{\sqrt{n+1}}\right),$$

where

$$\omega_1(f;\delta) = \sup\{|f(x) - f(y)|; x, y \in [0,1], |x-y| \leq \delta\}.$$

(iv) (Coroianu-Gal [15]*) There exists $f \in C_+[0,1]$ such that the order in (iii) is exactly $1/\sqrt{n+1}$, that is on the whole class $C_+[0,1]$, the order in (iii) cannot be improved.*

(v) (Coroianu-Gal [15]*) If $f \in C_+[0,1]$ is strictly positive on $[0,1]$ then*

$$\|B_n^{(M)}(f) - f\| \leq C_f\left\{n\left[\omega_1\left(f;\frac{1}{n}\right)\right]^2 + \omega_1\left(f;\frac{1}{n}\right)\right\}.$$

(vi) (Coroianu-Gal [15]*) If $f \in Lip 1$ then by (v)*

$$\|B_n^{(M)}(f) - f\| \leq \frac{C_f}{n}, n \in \mathbb{N}.$$

(vii) (Coroianu-Gal [15]*) If $f \in Lip\,\alpha$, then (v) gives the approximation order $1/n^{2\alpha-1}$, which for $\alpha \in (2/3, 1]$ is essentially better than the general approximation order $O[\omega_1(f; 1/\sqrt{n})] = O[1/n^{\alpha/2}]$ given by (iii).*

(viii) (Bede-Coroianu-Gal [4]*) If $f : [0,1] \to \mathbb{R}_+$ is a concave function then we have the Jackson-type estimate*

$$\left\|B_n^{(M)}(f)(x) - f(x)\right\| \leq 2\omega_1\left(f;\frac{1}{n}\right), n \in \mathbb{N}.$$

(ix) (Coroianu-Gal [15]*) If $f \in C_+[0,1]$ is strictly positive then the pointwise estimate holds*

$$|B_n^{(M)}(f)(x) - f(x)| \leq 24\omega_1\left(f, \sqrt{\frac{x(1-x)}{n}}\right),$$

for all $x \in [0, 1/(n+1)] \cup [n/(n+1), 1]$, *and*

$$\left| B_n^{(M)}(f)(x) - f(x) \right| \le \left(\frac{n\omega_1(f, \frac{1}{n})}{m_f} + 4 \right) \omega_1(f, \frac{1}{n}),$$

for all $x \in [1/(n+1), n/(n+1)]$.

(x) (Bede-Coroianu-Gal [4]*)* $f : [0,1] \to \mathbb{R}$ *is called quasi-convex (quasi-concave) on* $[0,1]$ *if it satisfies the inequality (for all* $x, y, \lambda \in [0,1]$*)*

$$f(\lambda x + (1-\lambda)y) \le (\ge) \max\{f(x), f(y)\}.$$

$B_n^{(M)}(f)$, $n \in \mathbb{N}$, *preserve the quasi-convexity, quasi-concavity and monotonicity of* f.

Remarks. 1) Comparing with the approximation by the Bernstein polynomials, clearly for large classes of functions, $B_n^{(M)}$ gives essentially better estimates.

2) The problem of finding the saturation class for $B_n^{(M)}$ is still open. Clearly it is different from the saturation class of the Bernstein polynomials.

For $f \in C_+[0, \infty)$ we define the Bleimann-Butzer-Hahn max-product operators by (Gal [18], p. 326)

$$H_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} \binom{n}{k} x^k f\left(\frac{k}{n+1-k}\right)}{\bigvee_{k=0}^{n} \binom{n}{k} x^k}.$$

Theorem 2.2. (Bede-Coroianu-Gal [8]) *(i) If* $f : [0, \infty) \to \mathbb{R}_+$ *is continuous, then for any* $n + 1 \ge \max\{1 + 2x, 16x(1+x)\}$ *we have*

$$|H_n^{(M)}(f)(x) - f(x)| \le 5\omega_1\left(f, \frac{(1+x)^{\frac{3}{2}}\sqrt{x}}{\sqrt{n+1}}\right), x \in [0, \infty),$$

where

$$\omega_1(f, \delta) = \sup\{|f(x) - f(y)|; x, y \in [0, \infty), |x - y| \le \delta\}.$$

(ii) If $f : [0, \infty) \to \mathbb{R}_+$ *is a nondecreasing concave function, then for* $x \in [0, \infty)$, $n \ge 2x$,

$$\left| H_n^{(M)}(f)(x) - f(x) \right| \le 2\omega_1\left(f; \frac{(1+x)^2}{n}\right).$$

(iii) $H_n^{(M)}(f)$, $n \in \mathbb{N}$, *preserve the monotonicity and the quasi-convexity of* f.

For $f \in C_+[0, 1)$ we define the Meyer-König and Zeller max-product operators by (Gal [18], p. 326)

$$Z_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} \binom{n+k}{k} x^k f(k/(n+k))}{\bigvee_{k=0}^{\infty} \binom{n+k}{k} x^k}, x \in [0, 1), n \in \mathbb{N}.$$

Theorem 2.3. (Bede-Coroianu-Gal [5]) *(i) If* $f : [0,1] \to \mathbb{R}_+$ *is continuous on* $[0,1]$*, then for* $n \geq 4$ *we have*

$$|Z_n^{(M)}(f)(x) - f(x)| \leq 18\omega_1\left(f, \frac{(1-x)\sqrt{x}}{\sqrt{n}}\right), x \in [0,1],$$

where

$$\omega_1(f,\delta) = \sup\{|f(x) - f(y)|; x, y \in [0,1], |x-y| \leq \delta\}.$$

(ii) If $f : [0,1] \to \mathbb{R}_+$ *is a continuous, nondecreasing concave function, then*

$$\left|Z_n^{(M)}(f)(x) - f(x)\right| \leq \omega_1\left(f; \frac{1}{n}\right), x \in [0,1], n \in \mathbb{N}.$$

(iii) $Z_n^{(M)}(f)$*,* $n \in \mathbb{N}$*, preserve the monotonicity and the quasi-convexity of* f*.*

For $f \in C_+[0,\infty)$ and $f \in C_+[0,1]$, we define the Favard-Szász-Mirakjan max-product (Gal [18], p. 326) and the truncated Favard-Szász-Mirakjan max-product operators (Bede-Coroianu-Gal [7]) by

$$F_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!} f\left(\frac{k}{n}\right)}{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!}}, x \in [0,\infty), n \in \mathbb{N}$$

and

$$T_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} \frac{(nx)^k}{k!} f\left(\frac{k}{n}\right)}{\bigvee_{k=0}^{n} \frac{(nx)^k}{k!}}, x \in [0,1], n \in \mathbb{N},$$

respectively.

Theorem 2.4. (Bede-Coroianu-Gal [10], [7]) *(i)*

$$|F_n^{(M)}(f)(x) - f(x)| \leq 8\omega_1\left(f, \frac{\sqrt{x}}{\sqrt{n}}\right), n \in \mathbb{N}, x \in [0,\infty),$$

where

$$\omega_1(f,\delta) = \sup\{|f(x) - f(y)|; x, y \in [0,\infty), |x-y| \leq \delta\},$$

and

$$|T_n^{(M)}(f)(x) - f(x)| \leq 6\omega_1\left(f, \frac{1}{\sqrt{n}}\right), n \in \mathbb{N}, x \in [0,1].$$

(ii) If $f : [0,\infty) \to \mathbb{R}_+$ *is a nondecreasing concave function on* $[0,\infty)$*, then*

$$\left|F_n^{(M)}(f)(x) - f(x)\right| \leq \omega_1\left(f; \frac{1}{n}\right), x \in [0,\infty), n \in \mathbb{N}.$$

(iii) If $f : [0,1] \to \mathbb{R}_+$ *is a nondecreasing concave function on* $[0,1]$*, then*

$$|T_n^{(M)}(f)(x) - f(x)| \leq 6\omega_1\left(f, \frac{1}{n}\right), n \in \mathbb{N}, x \in [0,1].$$

(iv) $F_n^{(M)}(f)$ *and* $T_n^{(M)}(f)$, $n \in \mathbb{N}$, *preserve the monotonicity and the quasi-convexity of* f *on the corresponding intervals.*

For $f \in C_+[0,\infty)$ and $f \in C_+[0,1]$, we define Baskakov max-product (Gal [18], p. 326) and the truncated Baskakov max-product operators (Bede-Coroianu-Gal [9]) by, respectively

$$V_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} b_{n,k}(x) f\left(\frac{k}{n}\right)}{\bigvee_{k=0}^{\infty} b_{n,k}(x)},$$

and

$$U_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} b_{n,k}(x) f\left(\frac{k}{n}\right)}{\bigvee_{k=0}^{n} b_{n,k}(x)}, x \in [0,1], n \in \mathbb{N}, n \geq 1,$$

where $b_{n,k}(x) = \binom{n+k-1}{k} x^k/(1+x)^{n+k}$.

Theorem 2.5. (Bede-Coroianu-Gal [6], [9]) *(i) For* $n \geq 3$ *and* $x \in [0,\infty)$ *we have*

$$|V_n^{(M)}(f)(x) - f(x)| \leq 12\omega_1\left(f, \sqrt{\frac{x(x+1)}{n-1}}\right),$$

where

$$\omega_1(f,\delta) = \sup\{|f(x) - f(y)|; x, y \in [0,\infty), |x-y| \leq \delta\}.$$

Also, for $n \in \mathbb{N}, n \geq 2, x \in [0,1]$ *we have*

$$|U_n^{(M)}(f)(x) - f(x)| \leq 24\omega_1\left(f, \frac{1}{\sqrt{n+1}}\right),$$

where

$$\omega_1(f,\delta) = \sup\{|f(x) - f(y)|; x, y \in [0,1], |x-y| \leq \delta\}.$$

(ii) If $f : [0,\infty) \to [0,\infty)$ *is a nondecreasing concave function on* $[0,\infty)$, *then for* $n \geq 3$, $x \in [0,\infty)$,

$$\left|V_n^{(M)}(f)(x) - f(x)\right| \leq 2\omega_1\left(f; \frac{x+1}{n-1}\right).$$

(iii) If $f : [0,1] \to [0,\infty)$ *is a nondecreasing concave function on* $[0,1]$, *then*

$$\left|U_n^{(M)}(f)(x) - f(x)\right| \leq 2\omega_1\left(f; \frac{1}{n}\right), x \in [0,1], n \in \mathbb{N}.$$

(iv) $V_n^{(M)}(f)$ *and* $U_n^{(M)}(f)$, $n \in \mathbb{N}$, *preserve the monotonicity and the quasi-convexity of* f *on the corresponding intervals.*

Remark. The estimates in Theorems 2.1, (iii), and Theorems 2.2-2.5, (i), were obtained by using the following general result:

Theorem 2.6. (Gal [18], p. 326, Bede-Gal [3]) *Let* $I \subset \mathbb{R}$ *be a bounded or unbounded interval,*

$$CB_+(I) = \{f : I \to \mathbb{R}_+; f \text{ continuous and bounded on } I\},$$

and $L_n : CB_+(I) \to CB_+(I)$, $n \in \mathbb{N}$ *be a sequence of* **positive homogenous** *operators, satisfying in addition the following properties:*

*(i) (***Monotonicity***) if* $f, g \in CB_+(I)$ *satisfy* $f \le g$ *then* $L_n(f) \le L_n(g)$ *for all* $n \in N$*;*

*(ii) (***Sublinearity***)* $L_n(f+g) \le L_n(f) + L_n(g)$ *for all* $f, g \in CB_+(I)$*.*

Then for all $f \in CB_+(I)$, $n \in N$ *and* $x \in I$ *we have*

$$|f(x) - L_n(f)(x)| \le$$

$$\left[\frac{1}{\delta} L_n(\varphi_x)(x) + L_n(e_0)(x)\right] \omega_1(f;\delta)_I + f(x) \cdot |L_n(e_0)(x) - 1|,$$

where $\delta > 0$, $e_0(t) = 1$ *for all* $t \in I$, $\varphi_x(t) = |t - x|$*.*

Remarks. 1) The above Theorem 2.6 is a generalization of the classical one for Positive Linear Operators, because the **Positivity** + **Linearity** imply the

Positivity + Sublinearity + Positive homogeneity

+Monotonicity,

but the converse implication does not hold, taking into account that the max product operators are counterexamples.

2) The Jackson-type estimates (for subclasses of functions) in Theorems 2.1-2.5, were obtained by direct reasonings.

3) The saturation results for the above max-product Bernstein-type operators are interesting open questions.

3. Approximation by interpolation max-product operators

In this section we present the approximation properties of a series of max-product interpolation operators.

Consider the Hermite-Fejér interpolation polynomial of degree $\le 2n+1$ attached to $f : [-1,1] \to \mathbb{R}$ and to the Chebyshev knots of first kind, $x_{n,k} = \cos\left(\frac{2(n-k)+1}{2(n+1)}\pi\right)$,

$$H_{2n+1}(f)(x) = \sum_{k=0}^{n} h_{n,k}(x) f(x_{n,k}),$$

with

$$h_{n,k}(x) = (1 - x x_{n,k}) \cdot \left(\frac{T_{n+1}(x)}{(n+1)(x - x_{n,k})}\right)^2,$$

$T_{n+1}(x) = \cos[(n+1)arccos(x)]$-Chebyshev polynomials. Because

$$H_{2n+1}(f)(x) = \frac{\sum_{k=0}^{n} h_{n,k}(x) f(x_{n,k})}{\sum_{k=0}^{n} h_{n,k}(x)},$$

by the max-product method the corresponding max-product Hermite-Fejér interpolation operator is

$$H_{2n+1}^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} h_{n,k}(x) f\left(x_{n,k}\right)}{\bigvee_{k=0}^{n} h_{n,k}(x)}.$$

Remark. We have $H_{2n+1}^{(M)}(f)(x_{n,j}) = f(x_{n,j})$, for all $j \in \{0, ..., n\}$.

Theorem 3.1. (Coroianu-Gal [14]) *If* $f : [-1,1] \to \mathbb{R}_+$ *is continuous on* $[-1,1]$ *then for all* $x \in [-1,1]$ *and* $n \in \mathbb{N}$

$$\|H_{2n+1}^{(M)}(f) - f\| \leq 14\omega_1\left(f, \frac{1}{n+1}\right).$$

Remark. For $f \in Lip_1[-1,1]$, we have $\|H_{2n+1}^{(M)}(f) - f\| \leq \frac{c}{n+1}$, while it is well-known that $\|H_{2n+1}(f) - f\| \sim \frac{ln(n+1)}{n+1}$.

Let now $x_{n,k} \in [-1,1]$, $k \in \{1, ..., n\}$, be arbitrary and consider the Lagrange interpolation polynomial of degree $\leq n-1$ attached to f and to the nodes $(x_{n,k})_k$,

$$L_n(f)(x) = \sum_{k=1}^{n} l_{n,k}(x) f(x_{n,k}),$$

with

$$l_{n,k}(x) = \frac{(x - x_{n,1})...(x - x_{n,k-1})(x - x_{n,k+1})...(x - x_{n,n})}{(x_{n,k} - x_{n,1})...(x_{n,k} - x_{n,k-1})(x_{n,k} - x_{n,k+1})...(x_{n,k} - x_{n,n})}.$$

Because $\sum_{k=1}^{n} l_{n,k}(x) = 1$, for all $x \in \mathbb{R}$, we can write

$$L_n(f)(x) = \frac{\sum_{k=1}^{n} l_{n,k}(x) f(x_{n,k})}{\sum_{k=1}^{n} l_{n,k}(x)}, \text{ for all } x \in I.$$

Therefore, its corresponding max-product interpolation operator will be given by

$$L_n^{(M)}(f)(x) = \frac{\bigvee_{k=1}^{n} l_{n,k}(x) f\left(x_{n,k}\right)}{\bigvee_{k=1}^{n} l_{n,k}(x)}, x \in I.$$

Remark. We have $L_n^{(M)}(f)(x_{n,k}) = f(x_{n,k})$, $k = 1, ..., n$.

Theorem 3.2. (Coroianu-Gal [12]) *If* $x_{n,k} = \cos\left(\frac{n-k}{n-1}\pi\right)$, $k = 1, ..., n$ *and* $f : [-1,1] \to \mathbb{R}_+$ *then*

$$\|L_n^{(M)}(f) - f\| \leq 28\omega_1\left(f, \frac{1}{n-1}\right), n \geq 3.$$

Remarks. 1) For the linear Lagrange polynomials we have the worst estimate

$$\|L_n(f) - f\| \leq C\omega_1\left(f; \frac{1}{n}\right) ln(n), n \in \mathbb{N}.$$

2) The case of other kind of nodes (e.g. equidistant, or roots of orthogonal polynomials, etc) can be found in the joint paper [17] with L. Coroianu published in this proceedings.

Now, consider the truncated Whittaker (sinc) series defined by

$$W_n(f)(x) = \sum_{k=0}^{n} \frac{\sin(nx - k\pi)}{nx - k\pi} \cdot f\left(\frac{k\pi}{n}\right), x \in [0, \pi],$$

and the truncated max-product Whittaker operator given by

$$W_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{n} \frac{\sin(nx-k\pi)}{nx-k\pi} \cdot f\left(\frac{k\pi}{n}\right)}{\bigvee_{k=0}^{n} \frac{\sin(nx-k\pi)}{nx-k\pi}}, x \in [0, \pi]$$

Remark. Clearly, $W_n^{(M)}(f)(j\pi/n) = f(j\pi/n)$, for all $j \in \{0, ..., n\}$.
Theorem 3.3. (Coroianu-Gal [16]) *If $f : [0, \pi] \to \mathbb{R}_+$ is continuous then*

$$|W_n^{(M)}(f)(x) - f(x)| \leq 4\omega_1\left(f; \frac{1}{n}\right)_{[0,\pi]}, n \in \mathbb{N}, x \in [0, \pi].$$

Remark. If $\lim_{n\to\infty} \omega_1(f; 1/n) \ln(n) = 0$ then $W_n(f)(x) \to f(x)$ uniformly inside of $(0, \pi)$ and pointwise in $[0, \pi]$, while it is known that $\|W_n(1) - 1\| \geq \frac{1}{3\pi}$, for all $n \geq 2$.

4. Approximation by sampling and neural networks max-prod operators

This section contains approximation results for some max-product sampling operators and for some max-product neural networks operators.
Definition 4.1. (Bardaro-Butzer-Stens-Vinti [2]) A function $\varphi \in C(\mathbb{R})$ is called a time-limited kernel (for a sampling operator), if:

(i) There exist $T_0, T_1 \in \mathbb{R}, T_0 < T_1$, such that $\varphi(t) = 0$ for all $t \notin [T_0, T_1]$;
(ii) $\sum_{k=-\infty}^{\infty} \varphi(u - k) = 1$, for all $u \in \mathbb{R}$.

If φ is a time-limited kernel and $W > 0$, then

$$S_{W,\varphi}(f)(t) = \sum_{k=-\infty}^{\infty} f\left(\frac{k}{W}\right) \varphi(Wt - k), t \in \mathbb{R},$$

will be called a generalized sampling operator.

Taking into account Definition 4.1, (ii), we can write

$$S_{W,\varphi}(f)(t) = \frac{\sum_{k=-\infty}^{\infty} f\left(\frac{k}{W}\right) \varphi(Wt - k)}{\sum_{k=-\infty}^{\infty} \varphi(Wt - k)}, t \in \mathbb{R}.$$

Remark. If e.g. $\varphi(t) = sinc(t) = \frac{sin(\pi t)}{\pi t}$, then $S_{W,\varphi}(f)(t)$ becomes the Whittaker cardinal (sinc) series.

Therefore, applying the max-product method, the corresponding max-product Whittaker operator will be given by

$$S_{W,\varphi}^{(M)}(f)(t)=\frac{\bigvee_{k=-\infty}^{\infty}\varphi(Wt-k)f\left(\frac{k}{W}\right)}{\bigvee_{k=-\infty}^{\infty}\varphi(Wt-k)}, t\in\mathbb{R}.$$

Theorem 4.2. (Coroianu-Gal [13]) *If* $\varphi(t)=sinc(t)=\frac{\sin(\pi t)}{\pi t}$ *and* $f:\mathbb{R}\to\mathbb{R}_+$ *is bounded and continuous on* $\mathbb{R}$*, then*

$$|S_{W,\varphi}^{(M)}(f)(t)-f(t)|\leq 2\omega_1\left(f;\frac{1}{W}\right)_{\mathbb{R}},\ \textit{for all } t\in\mathbb{R},$$

where $\omega_1(f;\delta)_{\mathbb{R}}=\sup\{|f(u)-f(v)|;u,v\in\mathbb{R},|u-v|\leq\delta\}$*.*

Remarks. 1) If $f\in Lip\alpha$, $\alpha\in(0,1]$, then in Theorem 4.2 we get $\|S_{W,\varphi}^{(M)}(f)-f\|=O\left(\frac{1}{W^\alpha}\right)$, while it is well-known that for the usual Whitaker cardinal series, we have the worst estimate

$$\|S_{W,\varphi}(f)-f\|=O\left(\frac{\log(W)}{W^\alpha}\right).$$

2) We get similar results for other kernels $\varphi(t)$ too.

The Cardaliaguet-Euvrard neural network is defined by

$$C_{n,\alpha}(f)(x)=\sum_{k=-n^2}^{n^2}\frac{f(k/n)}{I\cdot n^{1-\alpha}}\cdot b\left(n^{1-\alpha}\left(x-\frac{k}{n}\right)\right),$$

where $0<\alpha<1$, $n\in\mathbb{N}$ and $f:\mathbb{R}\to\mathbb{R}$ is continuous and bounded or uniformly continuous on $\mathbb{R}$.

The corresponding max-product Cardaliaguet-Euvrard network operator is formally given by

$$C_{n,\alpha}^{(M)}(f)(x)=\frac{\bigvee_{k=-n^2}^{n^2}b\left[n^{1-\alpha}\left(x-\frac{k}{n}\right)\right]f\left(\frac{k}{n}\right)}{\bigvee_{k=-n^2}^{n^2}b\left[n^{1-\alpha}\left(x-\frac{k}{n}\right)\right]}, x\in\mathbb{R}.$$

Theorem 4.3. (Anastassiou-Coroianu-Gal [1]) *Let* $b(x)$ *be a centered bell-shaped function, continuous and with compact support* $[-T,T]$*,* $T>0$ *and* $0<\alpha<1$*. In addition, suppose that the following requirements are fulfilled:*

(i) There exist $0<m_1\leq M_1<\infty$ *such that* $m_1(T-x)\leq b(x)\leq M_1(T-x)$ *for all* $x\in[0,T]$*;*

(ii) There exist $0<m_2\leq M_2<\infty$ *such that* $m_2(x+T)\leq b(x)\leq M_2(x+T)$ *for all* $x\in[-T,0]$*.*

Then for all $f \in CB_+(R)$, $x \in \mathbb{R}$ *and for all* $n \in \mathbb{N}$ *satisfying* $n > \max\{T + |x|, (2/T)^{1/\alpha}\}$, *we have the estimate*

$$|f(x) - C_{n,\alpha}^{(M)}(f)(x)| \leq c\omega_1 \left(f; n^{\alpha-1}\right)_{\mathbb{R}},$$

where

$$c = 2\left(\max\left\{\frac{TM_2}{2m_2}, \frac{TM_1}{2m_1}\right\} + 1\right).$$

Remark. Let $f \in Lip\alpha$. For $\frac{1}{2} \leq \alpha < 1$, we get the same order of approximation $O\left(\frac{1}{n^{1-\alpha}}\right)$ for both operators $C_{n,\alpha}(f)(x)$ and $C_{n,\alpha}^{(M)}(f)(x)$, while for $0 < \alpha < \frac{1}{2}$, the approximation order obtained by the max-product operator $C_{n,\alpha}^{(M)}(f)(x)$ is essentially better than that obtained by the linear operator $C_{n,\alpha}(f)(x)$.

References

[1] Anastassiou, G.A., Coroianu, L., Gal, S.G., *Approximation by a nonlinear Cardaliaguet-Euvrard neural network operator of max-product kind*, J. Comp. Analysis and Applications, **12**(2010), no. 2, 396–406.

[2] Bardaro, C., Butzer, P.L., Stens, R.L., Vinti, G., *Prediction by samples from the past with error estimates covering discontinuous signals*, IEEE Transaction on Information Theory, **56**(2010), no. 1, 614-633.

[3] Bede, B., Gal, S.G., *Approximation by nonlinear Bernstein and Favard-Szász-Mirakjan operators of max-product kind*, J. Concrete and Applicable Mathematics, **8**(2010), no. 2, 193–207.

[4] Bede, B., Coroianu, L., Gal, S.G., *Approximation and shape preserving properties of the Bernstein operator of max-product kind*, Intern. J. Math. and Math. Sci., volume 2009, Article ID 590589, 26 pages, doi:10.1155/2009/590589.

[5] Bede, B., Coroianu, L., Gal, S.G., *Approximation and shape preserving properties of the nonlinear Meyer-Konig and Zeller operator of max-product kind*, Numerical Functional Analysis and Optimization, **31**(2010), no. 3, 232–253.

[6] Bede, B., Coroianu, L., Gal, S.G., *Approximation and shape preserving properties of the nonlinear Baskakov operator of max-product kind*, Stud. Univ. Babeş-Bolyai Math., 2010 (accepted for publication).

[7] Bede, B., Coroianu, L., Gal, S.G., *Approximation by truncated Favard-Szász-Mirakjan operator of max-product kind*, Demonstratio Mathematica (accepted for publication).

[8] Bede, B., Coroianu, L., Gal, S.G., *Approximation and shape preserving properties of the nonlinear Bleimann-Butzer-Hahn operators of max-product kind*, Comment. Math. Univ. Carol. (Praga), **51**(2010), no. 3, 397–415.

[9] Bede, B., Coroianu, L., Gal, S.G., *Approximation and shape preserving properties of the truncated Baskakov operator of max-product kind*, (submitted for publication).

[10] Bede, B., Coroianu, L., Gal, S.G., *Approximation and shape preserving properties of the nonlinear Favard-Szász-Mirakjan operator of max-product kind*, Filomat (Nis), **24**(2010), no. 3, 55–72.

[11] Bede, B., Nobuhara, H., Fodor, J., Hirota, K., *Max-product Shepard approximation operators*, Journal of Advanced Computational Intelligence and Intelligent Informatics, **10**(2006), 494–497.

[12] Coroianu, L., Gal, S.G., *Approximation by nonlinear Lagrange interpolation operators of max-product kind on Chebyshev knots of second kind*, Journal Comp. Analysis and Applications (accepted for publication).

[13] Coroianu, L., Gal, S.G., *Approximation by nonlinear generalized sampling operators of max-product kind*, Sampling Theory in Signal and Image Processing (accepted for publication).

[14] Coroianu, L., Gal, S.G., *Approximation by nonlinear Hermite-Fejér interpolation operators of max-product kind on Chebyshev nodes*, Rev. d' Anal. Numér. Théor. Approx. (accepted for publication).

[15] Coroianu, L., Gal, S.G., *Classes of functions with improved estimates in approximation by the max-product Bernstein operator*, Analysis and Applications (accepted for publication).

[16] Coroianu, L., Gal, S.G., *Approximation by max-product sampling operators based on sinc-type kernels*, (in preparation).

[17] Coroianu, L., Gal, S.G., *Approximation by max-product Lagrange interpolation operators*, Stud. Univ. Babeş-Bolyai Math., this volume.

[18] Gal, S.G., *Shape-Preserving Approximation by Real and Complex Polynomials*, Birkhauser Publ. Co., Boston, Basel, Berlin, 2008.

Sorin G. Gal
University of Oradea
Department of Mathematics and Computer Science
1, Universităţii Street
410087 Oradea
Romania
e-mail: galso@uoradea.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 353–357

Rigid body time-stepping schemes in a quasi-static setting

Bogdan Gavrea

Abstract. We discuss how linear complementary problems (LCPs) can be used to simulate rigid-body systems in a quasi-static setting. LCP-based time-stepping schemes were successfully used in [1] in order to plan and control meso-scale manipulation tasks.

Mathematics Subject Classification (2010): 65K10, 90C33.

Keywords: Linear complementarity problems, rigid body simulation.

1. Introduction

In [1] we considered the canonical problem of assembling a peg into a hole. Simulation of this quasi-static system was used in order to select the control parameters. The integration step in the simulator was formulated as a *mixed linear complementarity problem* (MLCP). MLCPs should be thought of as *linear complementarity problems*(LCPs) coupled with additional linear equality constraints. A brief description of the linear complementarity problem and results concerning LCPs with copositive matrices are given in the following subsections. For a detailed analysis of these problems we refer the reader to the excellent manuscript [2].

1.1. Linear complementarity problems

In this section we present the definitions for the linear complementarity problem (LCP) and the mixed linear complementarity problem(MLCP).

Definition 1.1. *The problem of finding $z \in \mathbb{R}^n$ such that*

$$z \geq 0, \quad Mz + b \geq 0, \quad \text{and} \quad z^T(Mz + b) = 0, \tag{1.1}$$

where $b \in \mathbb{R}^n$ and $M \in \mathbb{R}^{n\times n}$ is called a linear complementarity problem.

In the above definition the inequality $z \geq 0,\ z \in \mathbb{R}^n$ is to be understood componentwise, i.e., $z_i \geq 0,\ i = \overline{1,n}$. The non-negativity and complementarity conditions (1.1) can be also written in the more compact form:

$$0 \leq z \perp w := Mz + b \geq 0.$$

We denote the problem (1.1) by $LCP(b, M)$. If in addition to the complementarity constraints we add some equality constraints we obtain a *mixed linear complementarity problem (MLCP)*. To be more precise, we follow the definition in [2] and consider the matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{m\times m}$, $C \in \mathbb{R}^{n\times m}$ and $D \in \mathbb{R}^{m\times n}$. Let $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ be given.

Definition 1.2. *The mixed linear complementarity problem is the problem of finding vectors $u \in \mathbb{R}^n$ and $v \in \mathbb{R}^m$ such that*

$$\begin{array}{lcl} a + Au + Cv & = & 0 \\ b + Du + Bv & \geq & 0 \\ v & \geq & 0 \\ v^T(b + Du + Bv) & = & 0 \end{array} \tag{1.2}$$

We note that if the matrix A in (1.2) is invertible we can write u in terms of v and use this form to reduce the problem to a standard LCP formulation.

1.2. LCPs with copositive matrices

The matrix of the underlying LCP used in the time-stepping schemes such as the one used in [1] is a copositive matrix.

Definition 1.3. *A matrix $M \in \mathbb{R}^{n\times n}$ is said to be copositive if*

$$x^T M x \geq 0 \ \text{ for all } \ x \in \mathbb{R}^n, \ \ x \geq 0.$$

In general a linear complementarity problem with a copositive matrix is not guaranteed to possess a solution. Solvability of such LCPs is discussed in the following Theorem.

Theorem 1.4 ([2], Th. 3.8.6). *Let $M \in \mathbb{R}^{n\times n}$ be a copositive matrix and let $b \in \mathbb{R}^n$ be given. If the implication*

$$\left[v \geq 0,\ Mv \geq 0,\ v^T M v = 0\right] \ \Rightarrow\ \left[v^T b \geq 0\right]$$

holds, then $LCP(b, M)$ has a solution. Lemke's algorithm with precautions taken against cycling will always find a solution of $LCP(b, M)$.

Lemke's algorithm is a pivoting method similar to the simplex method of linear programming. Cycling here refers to the possibility of using the same basis twice.

2. The quasi-static model

The continuous-time model under the rigid body assumption is given by the following *differential complementarity problem (DCP)*:

$$\dot{q}(t) = v(t), \quad (2.1)$$

$$Ev(t) - W_n(q,u,t)\lambda_n(t) - W_t(q,u,t)\lambda_t(t) = 0, \quad (2.2)$$

$$0 \leq \Psi_n(q,u,t) \quad \perp \quad \lambda_n(t) \geq 0, \quad (2.3)$$

$$\dot{s}^+_{tk}(t) - \dot{s}^-_{tk}(t) = \left(W_{tk}(q,u,t)\right)^T v(t) + \frac{\partial \Psi_{tk}}{\partial t}(q,u,t), \quad k = 1,...,n_c, \quad (2.4)$$

$$0 \leq \dot{s}^+_{tk}(t) \quad \perp \quad \mu_k \lambda_{nk}(t) + \lambda_{tk}(t) \geq 0, \quad k = 1,...,n_c, \quad (2.5)$$

$$0 \leq \dot{s}^-_{tk}(t) \quad \perp \quad \mu_k \lambda_{nk}(t) - \lambda_{tk}(t) \geq 0, \quad k = 1,...,n_c. \quad (2.6)$$

Here q denotes the generalized system position and v the generalized system velocity. The control parameters are encoded in the vector u. The quasi-static assumption is reflected by the equilibrium equation (2.2), where E is a damping matrix, assumed to be symmetric positive definite. The vectors $\lambda_n(t) \in \mathbb{R}^{n_c}$ and $\lambda_t(t) \in \mathbb{R}^{n_c}$ represent all normal and tangential forces, while $W_n(q,u,t)$ and $W_t(q,u,t)$ are the normal and tangential wrench matrices. More precisely, the k-th column of $W_n(q,u,t)$ ($W_t(q,u,t)$) is the normal (tangential) wrench vector $W_{nk}(q,u,t)$ ($W_{tk}(q,u,t)$) corresponding to contact k, $k = \overline{1,n_c}$, with n_c denoting the number of active contacts. The vector $\Psi_n(q,u,t)$ contains the normal displacements for configuration q, controls u and time t. More precisely, $\Psi_n(q,u,t) = \left[\Psi_{n1}(q,u,t),...,\Psi_{nn_c}(q,u,t)\right]^T$, where $\Psi_{nk}(q,u,t)$ represents the normal displacement function corresponding to contact k. In a similar way, one defines the vector of tangential displacements, $\Psi_t(q,u,t) = \left[\Psi_{t1}(q,u,t),...,\Psi_{tn_c}(q,u,t)\right]^T$. Equation (2.3) represents the contact and non-penetration constraints; that is whenever the normal separation at contact k is strictly positive ($\Psi_{nk}(q,u,t) > 0$), the corresponding normal force is 0 ($\lambda_{nk} = 0$), while whenever contact k is established ($\Psi_{nk}(q,u,t) = 0$), the corresponding normal force is nonnegative ($\lambda_{nk} \geq 0$).

Equation (2.4) defines the positive, $\dot{s}^+_{tk}(t)$, and negative, $\dot{s}^-_{tk}(t)$, sliding velocities at contact k. The right-hand side of (2.4) represents the (overall) sliding velocity $\dot{s}_{tk}(t) := \dot{\Psi}_{tk}(q,u,t) = \left(W_{tk}(q,u,t)\right)^T v(t) + \dfrac{\partial \Psi_{tk}}{\partial t}(q,u,t)$ at contact k. The last two equations, namely (2.5) and (2.6), represent Coulomb's friction law at contact k, with $\mu_k \in [0,1]$ being the friction coefficients.

3. The time-stepping scheme

Let t_l denote the time at which one has a solution configuration q^l and let $t_{l+1} = t_l + h$ denote the time at which one would want an estimate of the solution. We approximate the new configuration q^{l+1} using a backward Euler formula, as follows

$$q^{l+1} = q^l + hv^{l+1},$$

where v^{l+1} is an estimate for the new velocity and will be found by solving a mixed linear complementarity problem. At each integration step the unknowns $\left(hv^{l+1}, h\lambda_n^{l+1}, h\lambda_f^{l+1}, h\sigma^{l+1}\right)$ may be obtained as the solution of the following MLCP:

$$\begin{pmatrix} 0 \\ \rho_n^{l+1} \\ \rho_f^{l+1} \\ s^{l+1} \end{pmatrix} = \begin{pmatrix} E & -W_n^l & -W_f^l & 0 \\ (W_n^l)^T & 0 & 0 & 0 \\ (W_f^l)^T & 0 & 0 & E_f \\ 0 & U_f & -E_f^T & 0 \end{pmatrix} \begin{pmatrix} hv^{l+1} \\ h\lambda_n^{l+1} \\ h\lambda_f^{l+1} \\ h\sigma^{l+1} \end{pmatrix} + \begin{pmatrix} 0 \\ \Psi_n^l + h\frac{\partial \Psi_n}{\partial t}^l \\ h\frac{\partial \Psi_f}{\partial t}^l \\ 0 \end{pmatrix} \tag{3.1}$$

with $0 \leq \left[\rho_n^{l+1}, \rho_f^{l+1}, s^{l+1}\right] \quad \perp \quad \left[h\lambda_n^{l+1}, h\lambda_f^{l+1}, h\sigma^{l+1}\right] \geq 0$. Here $U_f \in \mathbb{R}^{n_c \times n_c}$, $E_f \in \mathbb{R}^{2n_c \times n_c}$ with U_f a diagonal matrix with elements on its diagonal equal to μ_k, $k = 1, ..., n_c$ and E_f a block diagonal matrix, with diagonal blocks given by the vector e (e is a two-dimensional vector of all ones). That is,

$$U_f = \begin{pmatrix} \mu_1 & ... & 0 \\ \vdots & & \vdots \\ 0 & ... & \mu_{n_c} \end{pmatrix}, \quad E_f = \begin{pmatrix} 1 & ... & 0 \\ 1 & ... & 0 \\ \vdots & & \vdots \\ 0 & ... & 1 \\ 0 & ... & 1 \end{pmatrix}.$$

The superscript l used in the MLCP (3.1) indicates that all the corresponding quantities are calculated with $q := q^l$ and $t := t_l$. For each contact k we define the 3×2 matrix $W_{fk}(q, u, t)$ by joining the column vectors $W_{tk}(q, u, t)$ and $-W_{tk}(q, u, t)$. That is,

$$W_{fk}(q, u, t) = [W_{tk}(q, u, t) \quad - W_{tk}(q, u, t)].$$

If we put all the active contacts together we obtain the "frictional" wrench matrix $W_f(q, u, t)$ appearing in formulation (3.1). In a similar way, we get the vector $\Psi_f(q, u, t)$.

Solvability and the Friction Cone. For an active contact k, we define the friction cone corresponding to that contact by

$$FC_k(q, u, t) = \left\{ z = W_{nk}\lambda_{nk} + W_{fk}\lambda_{fk} \mid \lambda_{nk} \geq 0,\ \lambda_{fk} \geq 0,\ e^T\lambda_{fk} \leq \mu_k\lambda_{nk} \right\}, \tag{3.2}$$

where $W_{nk} := W_{n,k}(q, u, t)$, $W_{fk} := W_{fk}(q, u, t)$ and $e = [1, 1]^T$. The total friction cone, $FC(q, u, t)$, which accounts for all active contacts is defined by

$$FC(q, u, t) = \sum_{k=1}^{n_c} FC_k(q, u, t).$$

Using the fact that the matrix E in the MLCP (3.1) is positive definite, we can eliminate the variables hv^{l+1} and reduce the MLCP to a standard LCP with a copositive matrix. It can be shown that the resulting LCP, is solvable whenever the total friction cone $FC(q^l, u, t_l)$ is pointed. We recall that a cone is pointed if it doesn't contain any proper subspace. The lack of pointedness for the friction cone results in jammed configurations (see [3]) and therefore this regularity assumption is very realistic and can be successfully used in devising randomized plans (see [1]).

4. Conclusions

We have discussed an LCP-based time-stepping scheme that can be used to simulate rigid body systems in a quasi-static setting. The scheme was introduced and successfully used for a particular case in [1]. Solvability of the integration step is guaranteed by the pointedness of the friction cone, an assumption that is common in dynamic settings as well (see [3] and [4] for example).

Acknowledgments. The work has been co-funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labour, Family and Social Protection through the Financial Agreement POSDRU/89/1.5/S/62557.

References

[1] Cheng, P., Cappelleri, S., Gavrea, B., Kumar, V., *Planning and Control of Mesoscale Manipulation Tasks with Uncertainties*, Proceedings of Robotics, Science and Systems (3rd: 2007, Atlanta GA), MIT PRESS, 2008.

[2] Cottle, R.W., Pang, J.S., Stone, R.E., *The Linear Complementarity Problem*, Academic Press, Boston, MA, 1992.

[3] Anitescu, M., *Optimization-based simulation of nonsmooth dynamics*, Mathematical Programming, **105**(1990), no. 1, 113-143.

[4] Gavrea, B., Anitescu, M., Potra, F.A., *Convergence of a Class of Semi-Implicit Time-Stepping schemes for Nonsmooth Rigid Multibody Dynamics*, SIAM Journal on Optimization, **19**(2008), no. 2, 969-1001.

Bogdan Gavrea
Technical University of Cluj-Napoca
Faculty of Automation and Computer Science
28, Str. Memorandumului
400114 Cluj-Napoca
Romania
e-mail: Bogdan.Gavrea@math.utcluj.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 359–367

On some quadrature formulas on the real line with the higher degree of accuracy and its applications

Ioan Gavrea

Abstract. In this paper we study quadrature formulas with the higher degree of accuracy. We study the quasi-orthogonality of orthogonal polynomials and we give some results on the location of their zeros.

Mathematics Subject Classification (2010): 41A55, 42C05.

Keywords: Orthogonal polynomials, quasi-orthogonal polynomials, zeros, quadrature formulas.

1. Introduction

Let P_n be a polynomial of degree n such that

$$\int_a^b x^k P_n(x)w(x)dx = 0, \quad k = 0, 1, \ldots, n-1,$$

where w is a positive weight function on the finite or infinite interval $[a, b]$. P_n is the polynomial of degree n belonging to the family of orthogonal polynomials on $[a, b]$ with respect to the weight function w. It is well known that the zeros of P_n are all real and distinct and lie in (a, b).

Definition 1.1. *Let R_n be a polynomial of exact degree n, $n \geq r$, r being a fixed natural number. If R_n satisfies the conditions*

$$\int_a^b x^k P_n(x)w(x)dx = \begin{cases} 0, & for \quad k = 0, 1, \ldots, n-r-1 \\ \neq 0, & for \quad k = n-r \end{cases} \tag{1.1}$$

where w is a positive weight function on $[a, b]$, then R_n is a quasi-orthogonal polynomial of order r on $[a, b]$ with respect to w.

Remark 1.2. The quasi-orthogonal polynomials R_n are only defined for $n \geq r$.

If $r = 0$ then $R_n = \lambda P_n$ where λ is a real constant.

The following result can be found in [1].

Theorem 1.3. *Let $\{P_n\}$ be the family of orthogonal polynomials on $[a,b]$ with respect to a positive weight function w. A necessary and sufficient condition for a polynomial R_n of degree n to be quasi-orthogonal of order r on $[a,b]$ with respect to w is that*

$$R_n(x) = c_0P_n(x) + c_1P_{n-1}(x) + \ldots + c_rP_{n-r}(x) \tag{1.2}$$

where c_i's are numbers which can depend on n and $c_0c_r \neq 0$.

If R_n is quasi-orthogonal of order r on $[a,b]$, then at least $n-r$ distinct zeros of R_n lie in the interval (a,b).

In [1] C. Brezinski, K. A. Driver, M. Redino-Zaglia consider quasi-orthogonal polynomials of degree $n-1$, $n-2$:

$$R_n(x) = P_n(x) + a_nP_{n-1}(x), \quad a_n \neq 0 \tag{1.3}$$

and

$$R_n(x) = P_n(x) + a_nP_{n-1}(x) + b_nP_{n-2}(x), \quad b_n \neq 0 \tag{1.4}$$

and make a study of its zeros.

The following result is well known.

Theorem 1.4. *The quadrature formula*

$$\int_a^b f(x)w(x)dx = \sum_{i=1}^{n} A_{i,n}f(x_{i,n}) + R(f) \tag{1.5}$$

has the degree of exactness $n+k$ if and only if it is of interpolatory type and the nodal polynomial

$$\Pi_n(x) = \prod_{i=1}^{n}(x - x_{i,n})$$

is quasi-orthogonal of order $n-k-1$ in $[a,b]$ with respect to w.

A. Bultheel, R. Cruz-Barroso and Marc Van Borel ([2]) consider an n point quadrature formula of Gauss-Radon type:

$$\int_a^b f(x)w(x)dx = A_\alpha f(\alpha) + \sum_{k=1}^{n-1} A_{k,n}f(x_{k,n}) + R(f) \tag{1.6}$$

where $\alpha \in [a,b]$ is a fixed point and the degree of exactness is $2n-2$.

Remark 1.5. If $P_n(\alpha) = 0$ then (1.6) is actually a Gaussian quadrature formula.

Remark 1.6. The coefficients of the quadrature formula (1.6) are positive.

In [2] the authors studied also Gauss-Lobatto-type quadrature formulas with two arbitrary prefixed nodes, α and β:

$$\int_a^b f(x)w(x)dx = A_\alpha f(\alpha) + A_\beta f(\beta) + \sum_{k=1}^{n-2} A_{k,n}f(x_{k,n}) + R_n(f) \tag{1.7}$$

the degree of exactness being $2n-3$.

From Theorem 1.3, the nodes of such a rule will be the zeros of

$$R_n(x) = P_n(x) + a_n P_{n-1}(x) + b_n P_{n-2}(x).$$

2. $P_{n,k}$-polynomials and its properties

Let w be a positive weight function on $[a, b]$ $(a > -\infty)$, $n \in \mathbb{N}^*$, $k \in \mathbb{N}$ such that $k \leq n$.

We denote by $P_{n,k}$ the polynomial of degree n which satisfies the following conditions:

$$\int_a^b (x-a)^i P_{n,k}(x) w(x) dx = \delta_{k,i}, \quad i = 0, 1, \ldots, n. \tag{2.1}$$

In the following, without loss of generality, we will consider $a = 0$.

Remark 2.1. By (2.1) it follows that $P_{n,k}$ is a quasi-orthogonal polynomial of order $n - k$ with respect to the weight function w.

Theorem 2.2. *The zeros of $P_{n,k}$ are all real, distinct and lie in $(0, b)$.*

Proof. Let us denote by $0 < x_1 < \ldots < x_i < b$ the zeros of $P_{n,k}$ where it changes the sign. Obviously $i \geq k$. Suppose $i < n$. We have

$$\int_0^b (x - x_1) \ldots (x - x_i) P_{n,k}(x) w(x) dx > 0. \tag{2.2}$$

Using the definition of $P_{n,k}$, from (2.2) we obtain

$$(-1)^{i-k} \sigma_{i-k} > 0, \tag{2.3}$$

where $(-1)^{i-k}\sigma_{i-k}$ is the coefficient of x^k of the polynomial

$$(x - x_1) \ldots (x - x_i), \quad \sigma_{i-k} > 0.$$

On the other hand we have:

$$\int_0^b x(x - x_1) \ldots (x - x_i) P_{n,k}(x) w(x) dx > 0$$

or

$$(-1)^{i-k-1} \sigma_{i-k+1} > 0. \tag{2.4}$$

The relations (2.3) and (2.4) are contradictory. □

It is easy to see that the set $\{P_{n,k}\}_{k=0}^n$ forms a base in Π_n and for every $P \in \Pi_n$ we have:

$$\begin{aligned} P &= \sum_{k=0}^n \langle e_k, P \rangle P_{n,k} \\ &= \sum_{k=0}^n e_k \langle P_{n,k} P \rangle, \end{aligned}$$

where $e_k : \mathbb{R} \to \mathbb{R}$, $e_k(x) = x^k$, and

$$\langle f, g\rangle = \int_a^b f(x)g(x)w(x)dx.$$

We denote by $K_n(x, y)$ the Christoffel-Darboux kernel

$$K_n(x, y) = \sum_{k=0}^{n} p_k(x)p_k(y)$$

where the set $\{p_k\}_{k=0}^n$ is an orthonormal set

$$\int_0^b p_k(x)p_i(x)w(x)dx = \delta_{k,i}, \quad k, i \in \{0, 1, \dots, n\}.$$

The result from the following Theorem is easily verified.

Theorem 2.3. *The following relations hold:*

$$\begin{aligned} K_n(x, y) &= \sum_{k=0}^{n} x^k P_{n,k}(y) \\ &= \sum_{k=0}^{n} y^k P_{n,k}(x) \\ &= \frac{1}{a_{n+1,n+1}} \cdot \frac{P_{n+1,n+1}(x)P_{n,n}(y) - P_{n,n}(x)P_{n+1,n+1}(y)}{x - y} \end{aligned} \tag{2.5}$$

where $a_{n+1,n+1}$ is the coefficient of x^{n+1} from $P_{n+1,n+1}$.

3. Main results

Let P be a polynomial of degree n and let m_k be the moment of order k with respect to the weight function w,

$$m_k = \langle e_k, P\rangle = \int_0^b x^k P(x)w(x)dx, \quad k = 0, 1, \dots, n.$$

Then P can be written as

$$P(x) = \sum_{k=0}^{n} m_k P_{n,k}(x).$$

Theorem 3.1. *If*

$$(-1)^k m_k \geq 0, \quad k = 0, 1, 2, \dots, n \tag{3.1}$$

then the zeros of P are all real, distinct and lie in $(0, b)$.

Proof. By (3.1) it follows that there exist at least a point x_1 where P changes the sign.

Let $x_1, \dots, x_p$ be all the zeros where P changes its sign in the interval $(0, b)$ and suppose that $p < n$.

So, the polynomial $(x - x_1) \dots (x - x_p)P(x)$ doesn't change the sign.

Suppose that

$$(x-x_1)\dots(x-x_p)P(x)\geq 0. \tag{3.2}$$

From (3.2) we get:

$$\int_0^b (x-x_1)\dots(x-x_p)P(x)w(x)dx>0 \tag{3.3}$$

$$\int_0^b (x-x_1)\dots(x-x_p)P(x)w(x)dx=(-1)^p\sum_{i=0}^{p}(-1)^{p-i}m_{p-i}\sigma_i \tag{3.4}$$

where σ_i are Vieta's sum of order i of the numbers $x_1,\dots,x_p$.

On the other hand we have:

$$\int_0^b x(x-x_1)\dots(x-x_p)P(x)w(x)dx>0 \tag{3.5}$$

$$\int_0^b x(x-x_1)\dots(x-x_p)P(x)w(x)dx=(-1)^{p+1}\sum_{i=0}^{p}(-1)^{p-i+1}m_{p-i+1}\sigma_i. \tag{3.6}$$

From (3.4) and (3.6) it follows that the inequalities (3.3) and (3.4) are contradictory and so $p=n$. □

Corollary 3.2. *Let R_n be a quasi-orthogonal polynomial of order* 1,

$$R_n(x)=P_{n,n-1}(x)-a_nP_{n,n}(x).$$

If $a_n>0$ then the zeros of R_n are all real and distinct and lie in $(0,b)$.

Remark 3.3. The condition $a_n>0$ is only sufficient.

A necessary and sufficient condition is given by

$$(-1)^n(P_{n,n-1}(0)-a_nP_{n,n}(0))(P_{n,n-1}(b)-a_nP_{n,n}(b))>0.$$

Let $\alpha\in[0,b]$ be a fixed point and let us consider the quadrature formula

$$\int_0^b f(x)w(x)x=A_\alpha f(\alpha)+\sum_{k=1}^{n-1}A_{k,n}f(x_{k,n})+R(f) \tag{3.7}$$

having the degree of exactness $2n-2$.

This means that α is a root of polynomial R_n which is of the form

$$R_n(x)=P_{n,n-1}(x)+aP_{n,n}(x).$$

The coefficients A_α, $A_{k,n}$, $k=1,2,\dots,n-1$ are positive and are given by

$$A_{k,n}=\frac{\displaystyle\int_0^b (x-\alpha)^2 l_k^2(x)w(x)dx}{(x_{k,n}-\alpha)^2},\quad A_\alpha=\frac{\displaystyle\int_0^b l^2(x)w(x)dx}{l^2(\alpha)}$$

where

$$l(x)=\prod_{k=1}^{n-1}(x-x_{k,n}),\quad l_k(x)=\frac{l(x)}{(x-x_{k,n})l'(x_{k,n})}.$$

Theorem 3.4. *The coefficients $A_{k,n}$, $k = 1, \dots, n-1$ and A_α are given by*

$$A_{k,n} = \frac{1}{K_{n-1}(x_{k,n}, x_{k,n})}, \quad k = 1, 2, \dots, n-1$$

$$A_\alpha = \frac{1}{K_{n-1}(\alpha, \alpha)}.$$

Proof. Let us denote by:

$$M_i = \int_0^b x^i (x-\alpha) l_k(x) w(x) dx.$$

We have

$$\begin{aligned} M_1 &= x_{k,n} M_0 \\ M_2 &= x_{k,n}^2 M_0 \\ &\dots \\ M_{n-1} &= x_{k,n}^{n-1} M_0. \end{aligned} \tag{3.8}$$

From (3.8) we get

$$(x-\alpha) l_k(x) = M_0 \sum_{i=0}^{n-1} x_{k,n}^i P_{n-1,i}(x). \tag{3.9}$$

By (3.9) we obtain

$$M_0 = \frac{x_{k,n} - \alpha}{K_{n-1}(x_{k,n}, x_{k,n})}$$

and so

$$A_{k,n} = \frac{1}{K_{n-1}(x_{k,n}, x_{k,n})}, \quad k = \overline{1, n-1}.$$

Similarly we get

$$A_\alpha = \frac{1}{K_{n-1}(\alpha, \alpha)}.$$

The proof of the theorem is finished. □

Corollary 3.5. *Let $P \in \Pi_{2n-2}$, $P(x) > 0$, $\forall\ x \in \mathbb{R}$. Then*

$$\int_0^b P(x) w(x) dx \geq \frac{1}{K_{n-1}(\alpha, \alpha)} P(\alpha), \ \forall\ \alpha \in \mathbb{R}.$$

Theorem 3.6. *Let R_n be a quasi-orthogonal polynomial of order 1 with the weight function w having all its zeros lie in $[0, b)$. Suppose that*

$$R_n(x) = a_n x^n + \dots$$

Then for every continuous function f, $f : [a, b] \to \mathbb{R}$, the following equality holds:

$$\int_0^b w(x) f(x) dx - \sum_{k=1}^n A_k f(x_k) = \frac{1}{a_n} [x_1, x_2, \dots, x_n; [x, x_1, \dots, x_n; f]] \tag{3.10}$$

$$+\frac{1}{a_n^2}\int_0^b [x, x_1, x_2, \ldots, x_n; [\cdot, x_1, \ldots, x_n; f]] R_n^2(x) w(x) dx$$

where x_k, $k = 1, 2, \ldots, n$, *are the zeros of* R_n *and* $A_k = \dfrac{1}{K_{n-1}(x_k, x_k)}$.

Proof. The quadrature formula

$$\int_0^b w(x) f(x) dx = \sum_{k=1}^{n} A_k f(x_k) + R(f) \tag{3.11}$$

having degree of exactness $2n - 2$ is a quadrature formula of interpolatory type, coefficients A_k, $k = 1, 2, \ldots, n$ being given by

$$\begin{aligned} A_k &= \int_0^b l_k(x) w(x) dx \\ &= \frac{1}{K_{n-1}(x_k, x_k)}. \end{aligned}$$

We have

$$f(x) - L_{n-1}(f; x_1, \ldots, x_n)(x) = \frac{1}{a_n} R_n(x)[x, x_1, \ldots, x_n; f] \tag{3.12}$$

where $L_{n-1}(f; x_1, \ldots, x_n)$ is Lagrange's polynomial of degree $n - 1$ which interpolates the function f at the points x_k, $k = \overline{1, n}$.

R_n is of the form:

$$R_n = P_{n,n-1} + \alpha P_{n,n}, \quad \alpha \in \mathbb{R}.$$

From (3.12) we obtain

$$\begin{aligned} &\int_0^b f(x) R_n(x) w(x) dx - [x_1, x_2, \ldots, x_n; f] \\ &= \frac{1}{a_n} \int_0^b R_n^2(x) [x, x_1, x_2, \ldots, x_n; f] w(x) dx \end{aligned} \tag{3.13}$$

and

$$\int_0^b f(x) w(x) dx - \sum_{k=1}^{n} A_k f(x_k) = \frac{1}{a_n} \int_0^b R_n(x)[x, x_1, \ldots, x_n; f] w(x) dx \tag{3.14}$$

From (3.13) and (3.14) we get (3.10). □

Corollary 3.7. *Let* $f \in C^1[0, b]$. *Then there exists* $\theta \in [0, b]$ *such that* $R(f)$ *from (*3.11*) can be written in the following form*

$$\begin{aligned} R(f) = &\frac{1}{a_n}[x_1, x_2, \ldots, x_n; [x, x_1, \ldots, x_n; f]] \\ &+ \frac{k_n}{a_n^2}[\theta, x_1, \ldots, x_n; [x, x_1, \ldots, x_n; f]] \end{aligned} \tag{3.15}$$

where

$$k_n = \int_0^b R_n^2(x) w(x) dx.$$

Proof. Equation (3.15) follows from (3.13) if we put instead of f the divided difference $[x, x_1, \ldots, x_n; f]$. □

Theorem 3.8. *Let x_k, $k = 1, 2, \ldots, n$ be the zeros of $P_{n,0}$ and w a positive weight such that*

$$\int_0^b w(x)dx = 1.$$

Then, for every $P \in \Pi_{n-1}$ we have:

$$\int_0^b P(x)w(x)dx = \sum_{k=1}^n \frac{P(x_k)}{K_n(x_k, x_k)} - \frac{1}{a_n}\left[x_1, \ldots, x_n; \frac{P(x)}{x}\right]$$

where a_n is the coefficient of x^n from $P_{n,0}$.

Proof. Let us consider the quadrature formula

$$\int_0^b f(x)w(x)dx = \sum_{k=1}^n A_k f(x_k) + R(f). \tag{3.16}$$

The quadrature formula (3.16) has the degree of exactness $n-1$ and A_k, $k = 1, 2, \ldots, n$ are given by

$$A_k = \int_0^b \frac{P_{n,0}(x)w(x)}{(x - x_k)P'_{n,0}(x_k)}dx.$$

Let us denote by M_i the moment of order i, $i = 0, 1, \ldots, n$ of the polynomial

$$\frac{P_{n,0}(x)}{(x - x_k)P'_{n,0}(x_k)}.$$

We get

$$M_1 - x_k M_0 = \frac{1}{P'_{n,0}(x_k)} \tag{3.17}$$

$$M_i = x_k^{i-1} M_1, \quad i = 2, 3, \ldots, n.$$

So

$$\frac{P_{n,0}(x)}{(x - x_k)P'_{n,0}(x)} = M_0 P_{n,0}(x) + M_1 P_{n,1}(x) \tag{3.18}$$

$$+\frac{M_1}{x_k}(K_n(x, x_k) - P_{n,0}(x) - x_k P_{n,1}(x)).$$

For $x = x_k$ we get

$$1 = \frac{M_1}{x_k} K_n(x_k, x_k). \tag{3.19}$$

From (3.17) and (3.19) we obtain

$$M_0 = \frac{1}{K_n(x_k, x_k)} - \frac{1}{x_k P'_{n,0}(x_k)}.$$

On the other hand $M_0 = A_k$ and the quadrature formula (3.16) becomes:

$$\int_0^b f(x)w(x)dx = \sum_{k=1}^n \frac{f(x_k)}{K_n(x_k, x_k)} - \frac{1}{a_n}\left[x_1, \ldots, x_n; \frac{f(x)}{x}\right] + R(f).$$

If $f \in \Pi_{n-1}$, $R(f) = 0$ and the theorem is proved. $\square$

Corollary 3.9. *If* $P(0) = 0$ *and* $P \in \Pi_{n-1}$ *then*

$$\int_0^b P(x)w(x)dx = \sum_{k=1}^{n} \frac{P(x_k)}{K_n(x_k, x_k)}.$$

Acknowledgment. The author is extremely thankful to the referee for his valuable comments and suggestions, leading to a better presentation of the paper.

References

[1] Brezinski, C., Driver, K. A., Redino-Zaglia, M., *Quasi-orthogonality with applications to some families of classical orthogonal polynomials*, Appl. Numer. Math., **48**(2004), 157-168.

[2] Bultheel, A., Cruz-Barroso, R., Van Borel, M., *On Gauss-type quadrature formulas with prescribed nodes anywhere on the real line*, Calcolo, **47**(2010), 21-48, DOI 10.1007s 10092-009-0013-x.

Ioan Gavrea
Technical University of Cluj-Napoca
Department of Mathematics
28 Memorandumului Street
400114 Cluj-Napoca, Romania
e-mail: ioan.gavrea@math.utcluj.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 369–380

A Q-fractional version of Itô's formula

Wilfried Grecksch and Christian Roth

Abstract. In this paper we consider a white noise calculus for fractional Brownian motion with values in a separable Hilbert space, whereby the covariance operator Q is a kernel operator (Q-fractional Brownian motion). We prove a Q-fractional version of the Itô's formula.

Mathematics Subject Classification (2010): 60H15, 60H40.

Keywords: Q-fractional Itô formula.

1. Introduction

Extending white noise analysis [9], Biagini and Øksendal [2] introduce fractional white noise calculus. They give the corresponding definition of stochastic integrals, a fractional Itô formula and Itô isometry, fractional differentiation and a fractional Malliavin calculus, using the results of Elliott and van der Hoek [4].

In [1] Grecksch, Roth and Anh introduce the Q-fractional Brownian motion, i.e., a Hilbert space-valued fractional Brownian motion defined by a kernel operator Q, and develop the Q-fractional Brownian motion framework for $\frac{1}{2} < h < 1$ as it was done in [9] for the standard Brownian motion case and in [2] for the fractional Brownian motion case in finite dimensions. Grecksch, Roth and Anh introduce Q-fractional test functions spaces and distribution spaces analogous to the way Hida [7] did and develop the Q-fractional chaos expansion. The corresponding stochastic integral and the Hilbert space-valued Wick scalar product are introduced. Furthermore they proved Q-fractional versions of Girsanov's theorem and of Clark-Haussmann-Ocone theorem.

In this paper we give a short overview of the most important notions and definitions for Q-fractional Brownian motion, see [1]. In Section 3 we prove a Q-fractional version of Itô's formula (see Theorem 3.1).

2. Q-fractional Brownian motion setup

Let $\mathcal{S}(\mathbb{R}^1)$ denote the Schwartz space of rapidly decreasing smooth functions on $\mathbb{R}^1$ and let $\mathcal{S}'(\mathbb{R}^1)$ be its dual, usually called the space of tempered distributions.

Let K and H be two separable Hilbert spaces with scalar product $(\cdot,\cdot)_K$ and $(\cdot,\cdot)_H$, and $(\Omega,\mathcal{F},P)$ a complete probability space. We denote by $L(K,H)$ the set of all linear bounded operators from K to H. Let $Q \in L(K,K)$ be a self-adjoint, non-negative operator on K. We call Q a kernel operator in K if

(i) there exists a sequence $(\lambda_n)_{n\in\mathbb{N}} \subset \mathbb{R}^1_+ = \{x \in \mathbb{R}^1 : x \geq 0\}$ with $\lambda_n \to 0$ as $n \to \infty$;

(ii) there exists a complete orthonormal system $(e_n)_{n\in\mathbb{N}} \in K$ such that

$$Q(x) := \sum_{n=1}^{\infty} \lambda_n (x, e_n) e_n \tag{2.1}$$

for all $x \in K$ and $\sum_{n=1}^{\infty} \lambda_n < \infty$.

Definition 2.1. *A K-valued continuous Gaussian process $B^h(t)_{t\in[0,T]}$ with Hurst parameter $h \in (0,1)$ is called a Q-fractional Brownian motion, if there exists a kernel operator Q in K such that*

1. $\forall x, y \in K,\ s,t \in [0,T]$,

$$E\left(\left(B^h(t),x\right)_K \left(B^h(s),y\right)_K\right) = \frac{1}{2}(Q(x),y)_K \left(t^{2h} + s^{2h} - |t-s|^{2h}\right); \tag{2.2}$$

2. $\forall x \in K$,

$$E\left(B^h(t),x\right)_K = 0. \tag{2.3}$$

Remark 2.2. (i) In view of (2.2) we say that B^h has the covariance operator $\frac{1}{2}Q\left(t^{2h} + s^{2h} - |t-s|^{2h}\right)$.

(ii) Eq. (2.3) is equivalent to $EB^h(t) = 0$, i.e., it is the zero element of K.

(iii) The case of long-range dependence, i.e. $\frac{1}{2} < h < 1$, is given by

$$E\left(\left(B^h(t),x\right)_K \left(B^h(s),y\right)_K\right) = (Q(x),y)_K \int_0^t \int_0^s \varphi(u,v)\, du\, dv,$$

where $\varphi(u,v) := h(2h-1)|u-v|^{2h-2}$.

(iv) The Hilbert space valued Wiener process is obtained for $h = \frac{1}{2}$.

Theorem 2.3. *Let*

(i) *$(e_n)_{n\in\mathbb{N}}$ be a complete orthonormal system in K;*

(ii) *$(\lambda_n)_{n\in\mathbb{N}} \subset \mathbb{R}^1_+$, $\sum_{n=1}^{\infty} \lambda_n < \infty$;*

(iii) *$(\beta_n^h(t))_{t\in[0,T]}$, $n = 1,2,\ldots$ be independent real fractional Brownian motions with*

$$E\left(\beta_n^h(t)\beta_k^h(s)\right) = \frac{1}{2}\delta_{nk}\left(t^{2h} + s^{2h} - |t-s|^{2h}\right),$$

where δ_{nk} is the Kronecker delta function.

Then $\left(B^h(t)\right)_{t\in[0,T]}$ *is a Q-fractional Brownian motion if and only if*

$$B^h(t) = \sum_{n=1}^{\infty} \sqrt{\lambda_n}\beta_n^h(t)e_n = \sum_{n=1}^{\infty} Q^{1/2}(e_n)\beta_n^h(t). \tag{2.4}$$

Proof. See Grecksch and Anh [6], or Duncan, Maslowski and Pasic-Duncan [3].

We write $B_n^h(t) = \sqrt{\lambda_n}\beta_n^h(t)$.

In the following we will discuss (a two-sided) Q-fractional Brownian motion with help of fractional white noise calculus. Therefore we assume that the underlying probability spaces for the independent real fractional Brownian motions $B_1^h(\cdot)$, $B_2^h(\cdot)$, ... are $\Omega_1 = \mathcal{S}'(\mathbb{R}^1)$, $\Omega_2 = \mathcal{S}'(\mathbb{R}^1)$, ..., that is $B^h(\cdot)$ is defined on $\Omega = \prod_{i=1}^{\infty}\Omega_i$.

We now introduce the fundamental operator $M_h(t)$ according to Elliott and van der Hoek [4].

For $0 < h < \frac{1}{2}$ and $f \in \mathcal{S}(\mathbb{R}^1)$,

$$M_h f(x) := \left(2\Gamma\left(h-\frac{1}{2}\right)\cos\left(\frac{\pi}{2}\left(h-\frac{1}{2}\right)\right)\right)^{-1}\int_{\mathbb{R}^1}\frac{f(x-t)-f(x)}{|t|^{\frac{3}{2}-h}}\,dt. \tag{2.5}$$

For $\frac{1}{2} < h < 1$ and $f \in \mathcal{S}(\mathbb{R}^1)$,

$$M_h f(x) := \left(2\Gamma\left(h-\frac{1}{2}\right)\cos\left(\frac{\pi}{2}\left(h-\frac{1}{2}\right)\right)\right)^{-1}\int_{\mathbb{R}^1}\frac{f(t)}{|t-x|^{\frac{3}{2}-h}}\,dt. \tag{2.6}$$

For $h = \frac{1}{2}$ we put $M_h f(x) = f(x)$, the identity map.

When $f(x) = I(0,t)(x)$ we write

$$M_h f(x) = M_h(0,t)(x). \tag{2.7}$$

Now we want to characterize the Hilbert space valued fractional Brownian motion with white noise calculus. We define

$$\tilde{B}_h(t,\omega) = \sum_{n=1}^{\infty}\sqrt{\lambda_n} < M_h(0,t),\omega_n > e_n, \tag{2.8}$$

with $< M_h(0,t),\omega_n >= \int_{\mathbb{R}^1} M_h(0,t)(s)\,d\beta_n(s)$ and β_n are independent real Brownian motions.

Again, $\tilde{B}_h(t)$ is a Gaussian random variable with

$$E\left[\left(\tilde{B}_h(t),x\right)_K\right] = 0 \tag{2.9}$$

and for $s < t$, we get using the independence of ω_i

$$\begin{aligned} &E\left[\left(\tilde{B}_h(t),x\right)_K\left(\tilde{B}_h(s),y\right)_K\right] \\ &= E\left[\sum_{i=1}^{\infty}\sqrt{\lambda_i} < M_h(0,t),\omega_i > (x,e_i)_K \sum_{k=1}^{\infty}\sqrt{\lambda_k} < M_h(0,s),\omega_k > (y,e_k)_K\right] \\ &= C_h\left(|t|^{2h}+|s|^{2h}-|t-s|^{2h}\right)(Qx,y). \end{aligned} \tag{2.10}$$

The process $\tilde{B}^h(t)$ has a continuous version in K, which we denote by $B^h(t)$.

We extend the definition of M_h to Hilbert space valued functions $f : \mathbb{R}^1 \to K$. Then M_h is defined by

$$M_h f(x) := \sum_{n=1}^{\infty} e_n M_h \left(f, e_n\right)_K (x) \tag{2.11}$$

for all $x \in \mathbb{R}^1$ and all

$$\begin{aligned} f \in L_h^2(\mathbb{R}^1, K) &:= \\ &\left\{ f : \mathbb{R} \to K, M_h f = \sum_{i=1}^{\infty} M_h \left((f, e_i)_K\right) e_i \in L^2(\mathbb{R}^1, K) \right\}, \end{aligned} \tag{2.12}$$

where $M_h \left(f, e_i\right)_K$ is defined by applying (2.5) and (2.6) to the real functions $(f(\cdot), e_i)_K$.

The Hermite functions $\{\xi_n\}_{n=1}^{\infty}$, i.e.

$$\xi_n = \pi^{-\frac{1}{4}}((n-1)!)^{-\frac{1}{2}} h_{n-1}(\sqrt{2}x) e^{\frac{x^2}{2}}, \tag{2.13}$$

where $h_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} \left(e^{\frac{-x^2}{2}}\right)$ form a basis of $L^2(\mathbb{R}^1, \mathbb{R}^1)$. Define

$$\eta_n(x) = M_h^{-1} \xi_n(x); \quad n = 1, 2... \tag{2.14}$$

Then it follows from [4]

$$(f(x), e_n) = \sum_{j=1}^{\infty} c_{jn} \eta_j(x) \tag{2.15}$$

that η_j is an orthonormal basis of $L_h^2(\mathbb{R}^1, \mathbb{R}^1)$. Consequently $\eta_j(x) e_n$, $(j = 1, 2, ..., \ n = 1, 2...)$ defines an orthonormal basis of $L_h^2(\mathbb{R}^1, K)$.

Let $\mathcal{H}_r$, $r = 1, 2, ...$, be the Hermite polynomials of order r. Evidently we have

$$\mathcal{H}_1(\langle B^h, \eta_j e_n \rangle) = \frac{1}{2} \langle B^h, \eta_j e_n \rangle = \frac{1}{2} \langle B_n^h, \eta_j \rangle = \frac{1}{2} \langle \sqrt{\lambda_n} \beta_n^h, \eta_j \rangle.$$

Furthermore we define

$$\mathcal{H}_\alpha \left(B_n^h\right) := \mathcal{H}_{\alpha_1} \left(B_n^h \left(\eta_1\right)\right) \cdot ... \cdot \mathcal{H}_{\alpha_j} \left(B_n^h \left(\eta_j\right)\right),$$

and α is a multi-index, that is, $\alpha = (\alpha_1, ..., \alpha_j), \ \alpha_i \in \mathbb{N}$. In particular $\varepsilon^{(n)}$ denotes the multi-index with 1 at the place n and 0 else.

Remark 2.4. In view of the representation Theorem 2.3, Eq. (2.4) for Q-fractional Brownian motions, we have for a deterministic function F with values in $L^2[0, T]$

$$\int_0^T F(s) \, dB^h(s) \quad = \quad \sum_{n=1}^{\infty} \int_0^T \sqrt{\lambda_n} F(s) e_n \, d\beta_n^h(s) \tag{2.16}$$

in mean square in H.

We can write the expansion of $B^h(t)$ as

$$B^h(t) \;=\; \sum_{n=1}^{\infty}\sqrt{\lambda_n}\beta_n^h(t)e_n = \sum_{j=1}^{\infty}\sum_{n=1}^{\infty}\int_0^t \eta_j(s)\,ds\mathcal{H}_{\varepsilon(j)}(B_n^h)e_n. \quad (2.17)$$

We introduce the notation

$$B^h(\eta_j e_n) := \langle B^h, \eta_j e_n\rangle e_n = \int_{\mathbb{R}^1}\eta_j(x)\,dB_n^h(x)e_n. \quad (2.18)$$

Furthermore $\int_0^T \eta_j(t)\,dB_n^h(t)e_n$ is defined by $\int_{\mathbb{R}^1} I_{[0,T]}(t)\eta_j(t)\,dB_n^h(t)e_n$.

Therefore we have

$$E\left(B^h(\eta_j e_n)\right)^2 \;=\; \int_{\mathbb{R}^1}\lambda_n|M_h\left(\eta_j(t)\right)|^2\,dt = \lambda_n. \quad (2.19)$$

Remark 2.5. (i) Let $F(s)$ be a deterministic operator function. Then we get

$$\int_0^T \left(F(s)e_n, h_k\right)_K\; dB_n^h(s) \;=\; \sum_{j=1}^{\infty} c_{knj}\sqrt{\lambda_n}\mathcal{H}(\beta_n^h(I_{[0,T]}\eta_j)). \quad (2.20)$$

(ii) Especially, if $H=\mathbb{R}^1$ and $F(s)=\gamma(s)\in L_h^2([0,T],K)$ and $\|\gamma(s)\|\le C$ $\forall\; s\in[0,T]$. Then

$$\int_0^T \left(\gamma(s), dB^h(s)\right)_K \;=\; \sum_{n=1}^{\infty}\sum_{j=1}^{\infty} c_{nj}\mathcal{H}_1(B_n^h(I_{[0,T]}\eta_j)). \quad (2.21)$$

(iii) Using the properties of Hermite polynomials the expansion of $Exp\{b_j\eta_j\}$ $(b_j\in\mathbb{R}^1)$ is given by

$$\begin{aligned} Exp\left\{b_j\eta_j\right\} &= exp\left\{b_j\int_{\mathbb{R}^1}\sqrt{\lambda_n}\eta_j(t)\,d\beta_n^h(t) - \frac{b_j^2\lambda_n}{2}\|M_h\eta_j\|^2_{L^2(\mathbb{R})}\right\} \\ &= \sum_{l=1}^{\infty}\frac{b_j^l}{l!}\mathcal{H}_l(B_n^h(\eta_j)) = \sum_{l=1}^{\infty}\frac{b_j^l}{l!}\mathcal{H}_l(B_n^h(\eta_j))), \end{aligned} \quad (2.22)$$

(see [7], [8] or [10]).

Example 2.6. Now let us consider the expansion of $Exp\{\gamma\}$ for $\gamma\in L_Q\left(\mathbb{R}^1,K\right)$ with respect to $e_n\eta_j(t)$, $j=1,2...$, $n=1,2,...$ see (2.21). We can write the exponential of γ as

$$\begin{aligned} Exp\left\{\gamma\right\} &= \exp\left\{\int_{\mathbb{R}^1}\left(\gamma(t),\, dB^h(t)\right) - \frac{1}{2}\|M_h\gamma\|^2_{L_Q^2(\mathbb{R}^1,K)}\right\} \\ &= \exp\left\{\sum_{n=1}^{\infty}\sum_{j=1}^{\infty}\sqrt{\lambda_n}c_{nj}\mathcal{H}_1(\beta_n^h(\eta_j)) - \frac{1}{2}\sum_{n=1}^{\infty}\sum_{j=1}^{\infty}\lambda_n c_{nj}^2\|M_h\eta_j\|^2_{L^2(\mathbb{R})}\right\} \\ &=: \sum_{\alpha\in\mathcal{I}}\prod_{n,j=1}^{\infty} c_{\alpha nj}\mathcal{H}_\alpha\left(B_n^h(\eta_j)\right), \end{aligned} \quad (2.23)$$

where $\mathcal{H}_\alpha(B_n^h) := \mathcal{H}_{\alpha_1}(B_n^h(\eta_j)) \cdot ... \cdot \mathcal{H}_{\alpha_j}(B_n^h(\eta_j))$ and

$$c_{\alpha nj} := \prod_{l=1}^{\infty} \frac{(c_{nj})^{\alpha_l}}{\alpha_l!}, \ \alpha = (\alpha_1, ..., \alpha_j) .$$

Here, $\mathcal{I}$ denotes the set of all multi-indices α, $\mathcal{I} = \{(\alpha_1, ..., \alpha_n) : \alpha_1, ..., \alpha_n \in \mathbb{N}_0, n \in \mathbb{N}\}$.

We obtain for $Exp\{\gamma(t)\}$

$$Exp\{\gamma(t)\} = \sum_{\alpha\in\mathcal{I}} \prod_{n,j=1}^{\infty} c_{\alpha nj} \mathcal{H}_\alpha \left(B_n^h(I_{[0,T]}\eta_j)\right) . \tag{2.24}$$

Now we want to develop a fractional white noise integration theory for $h \in (0,1)$. Grecksch, Roth and Anh [1] define the Q-fractional version of the Hida test function space and the Hida distribution space for $h \in (\frac{1}{2}, 1)$. Inspired by (2.23) we make the definitions as follows:

Let V be a separable Hilbert space with a complete orthonormal system $(v_k) \subseteq V$.

Definition 2.7. *The Q-fractional test function space $S_Q^h(V)$ is the space of all V-valued random functions with expansion*

$$\Psi(\omega) = \sum_{k=1}^{\infty} \left[\sum_{\alpha\in\mathcal{I}} \prod_{n,j=1}^{\infty} c_{\alpha nj}^{(k)} \mathcal{H}_\alpha(B_n^h) \right] v_k,$$

for which

$$\|\Psi\|_{h,r} := \sum_{k=1}^{\infty} \sum_{\alpha\in\mathcal{I}} \prod_{n,j=1}^{\infty} \alpha!(c_{\alpha nj}^{(j)})^2 (2\mathbb{N})^{r\alpha} < \infty, \ \forall r \in \mathbb{N},$$

and $(2\mathbb{N})^\alpha := \prod_{j=1}^{\infty}(2j)^{\alpha_j}$ *if* $\alpha = (\alpha_1, ..., \alpha_m)$.

Definition 2.8. *The Q-fractional distribution space $(S_Q^h(V))^*$ is the space of all V-valued random functions with expansion*

$$G(\omega) = \sum_{k=1}^{\infty} \left[\sum_{\beta\in\mathcal{I}} \prod_{n,j=1}^{\infty} b_{\beta nj}^{(k)} \mathcal{H}_\beta(B_n^h) \right] v_k,$$

for which

$$\|G\|_{h,-q} := \sum_{k=1}^{\infty} \sum_{\beta\in\mathcal{I}} \prod_{n,j=1}^{\infty} \beta!(b_{\beta nj}^{(k)})^2 (2\mathbb{N})^{-q\beta} < \infty \ \textit{ for some } q \in \mathbb{N}.$$

Remark 2.9. If $V = \mathbb{R}^1$, then $\Psi(\omega) \in S_Q^h(V)$ (or $\Psi(\omega) \in (S_Q^h(V))^*$) has the following representation

$$\Psi(\omega) = \sum_{\alpha\in\mathcal{I}} \prod_{n,j=1}^{\infty} c_{\alpha nj} \mathcal{H}_\alpha(B_n^h).$$

Furthermore if the fractional noise is only one-dimensional, we find the well-known representation

$$\Psi(\omega) = \sum_{\alpha\in\mathcal{I}} c_\alpha \mathcal{H}_\alpha(B^h).$$

Consider the following duality relation between $S_Q^h(V)$ and $(S_Q^h(V))^*$. For $G \in (S_Q^h(V))^*$ and $\psi \in S_Q^h(V) \subset L_V^2(\Omega)$ we define

$$\langle\langle G, \psi\rangle\rangle := \sum_{k=1}^{\infty}\sum_{\alpha\in\mathcal{I}}\prod_{n,j=1}^{\infty} \alpha! c_{\alpha nj}^{(k)} b_{\alpha nj}^{(k)}. \tag{2.25}$$

Example 2.10. If $G \in L_V^2(\Omega)$ and $\psi \in S_Q^h(V) \subset L_V^2(\Omega)$, then we have

$$\langle\langle G, \psi\rangle\rangle = E(G,\psi)_V = (G,\psi)_{L_V^2(\Omega)}. \tag{2.26}$$

Definition 2.11. *Let $Z : [0,T] \to (S_Q^h(V))^*$ with*

$$\int_0^T |\langle\langle Z(t), \psi\rangle\rangle| \, dt < \infty, \quad \forall\psi \in S_Q^h(V).$$

Then $\int_0^T Z(t)\, dt \in (S_Q^h(V))^*$ is uniquely determined by the relation

$$\langle\langle \int_0^T Z(t)\, dt, \psi\rangle\rangle = \int_0^T \langle\langle Z(t), \psi\rangle\rangle \, dt.$$

We say that Z is $(S_Q^h(V))^*$-integrable.

Definition 2.12. (Wick scalar product)
Let F, $G \in (S_Q^h(K))^*$ with

$$\begin{aligned} F(\omega) &= F(B^h) = \sum_{k=1}^{\infty}\left[\sum_{\alpha\in\mathcal{I}}\prod_{n,j=1}^{\infty} a_{\alpha nj}^{(k)} \mathcal{H}_\alpha(B_n^h)\right] v_k, \\ G(\omega) &= G(B^h) = \sum_{k=1}^{\infty}\left[\sum_{\beta\in\mathcal{I}}\prod_{l,m=1}^{\infty} b_{\beta lm}^{(k)} \mathcal{H}_\beta(B_l^h)\right] v_k, . \end{aligned}$$

We define

$$\begin{aligned} (F,G)_{\diamond V} := & \sum_{k=1}^{\infty}\sum_{\alpha,\beta\in\mathcal{I}}\prod_{n,j=1}^{\infty} a_{\alpha nj}^{(k)} b_{\beta nj}^{(k)} \mathcal{H}_{\alpha+\beta}(B_n^h) \\ = & \sum_{k=1}^{\infty}\left[\sum_{\gamma\in\mathcal{I}}\sum_{\alpha+\beta=\gamma}\prod_{n,j=1}^{\infty} a_{\alpha nj}^{(k)} b_{\beta nj}^{(k)} \mathcal{H}_{\alpha+\beta}(B_n^h)\right]. \end{aligned} \tag{2.27}$$

Remark 2.13. If $V = \mathbb{R}^1$ then $(\cdot,\cdot)_{\diamond V}$ is the usual Wick product.

Now we introduce a fractional stochastic integral with stochastic integrands.

Definition 2.14. $Y : [0,T] \to (S_Q^h(V))^*$ *is* (dB^h-)*integrable if*

$$(Y(t), W^h(t))_{\diamond V} = \sum_{n=1}^{\infty} \sqrt{\lambda_n}\,(Y(t), e_n)_V \diamond W_n^h(t)$$

is integrable with respect to t *in the sense of Definition* 2.11*. We define*

$$\int_0^T \big(Y(t), dB^h(t)\big) := \int_0^T (Y(t), W^h(t))_{\diamond V}\, dt.$$

3. A Q-fractional version of Itô's formula

In this section we prove a Q-fractional version of Itô's formula the way Biagini, Øksendal and al. presented it for a usual fractional Brownian motion, see [2].

$C^{1,2}([0,T]\times K, \mathbb{R}^1)$ denotes the space of all functions $f : [0,T]\times K \to \mathbb{R}^1$, such that the first Fréchet derivative $\nabla_s f(s,x)$ with respect to $s \in [0,T]$ and the first and second Fréchet derivatives $\nabla_x f(s,x)$ and $\nabla_{xx} f(s,x)$ exist continuously.

Theorem 3.1. *Let* $f(s,x) : [0,T]\times K \to \mathbb{R}$ *belong to* $C^{1,2}([0,T]\times K, \mathbb{R}^1)$*. Furthermore assume that there are constants* $C \geq 0$ *and* $0 < \lambda < \frac{1}{4T^{2h}}$ *such that for all* $(t,x) \in [0,T]\times K$

$$\max\Big\{|f(t,x)|,\ |\nabla_t f(t,x)|, \|\nabla_x f(t,x)\|_K,\\ \|\nabla_{xx} f(t,x)\|_{L(K,K)}\Big\} \leq Ce^{\lambda x^2}. \tag{3.1}$$

Then

$$\begin{aligned} f(t,B^h(t)) &= f(0,0) + \int_0^t \nabla_s f(s,B^h(s))\,ds \\ &\quad + \int_0^t \big(\nabla_x f(s,B^h(s)), dB^h(s)\big)_K \\ &\quad + h\sum_{i=1}^{\infty}\int_0^t \big(\nabla_{xx} f(s,B^h(s))e_i, e_i\big)_K \lambda_i s^{2h-1}\,ds, \end{aligned} \tag{3.2}$$

whereby

$$\begin{aligned} \nabla_s f(s,B^h(s)) &= \nabla_u f(u,B^h(s))\big|_{u=s}, \\ \nabla_x f(s,x) &= \nabla_x f(s,x)\big|_{x=B^h(s)}, \\ \nabla_{xx} f(s,x) &= \nabla_{xx} f(s,x)\big|_{x=B^h(s)}. \end{aligned}$$

Proof. Define

$$g(t,x) = \exp\{(a,x)_K + \beta(t)\}, \tag{3.3}$$

whereby $a \in K$ is a constant, $\beta \in C^1([0,T], \mathbb{R}^1)$ is a deterministic function, and put

$$Y(t) = g(t, B^h(t)),\ \text{i.e. } x = B^h(t). \tag{3.4}$$

With

$$(a, B^h(s))_K = \sum_{i=1}^{\infty} \sqrt{\lambda_i}(a, e_i)\beta_i^h(t)$$

we can rewrite

$$Y(t) = \exp\left\{\sum_{i=1}^{\infty} \sqrt{\lambda_i}(a, e_i)_K \beta_i^h(t)\right\} \exp\{\beta(t)\}$$

$$= \exp^{\diamond}\left\{\sum_{i=1}^{\infty} \sqrt{\lambda_i}(a, e_i)_K \beta_i^h(t) + \frac{1}{2}\sum_{i=1}^{\infty} \lambda_i (a, e_i)_K^2 t^{2h}\right\} \exp\{\beta(t)\}. \quad (3.5)$$

Therefore, by applying Wick calculus, we have

$$\begin{aligned}
&\frac{d}{dt}Y(t) \\
&= \exp^{\diamond}\left\{\sum_{i=1}^{\infty} \sqrt{\lambda_i}(a, e_i)_K \beta_i^h(t) + \frac{1}{2}\sum_{i=1}^{\infty} \lambda_i (a, e_i)_K^2 t^{2h}\right\} \exp\{\beta(t)\} \\
&\quad \diamond \left[(a, W^h(t))_K + h\sum_{i=1}^{\infty} \lambda_i (a, e_i)_K^2 t^{2h-1}\right] \\
&\quad + \exp^{\diamond}\left\{\sum_{i=1}^{\infty} \sqrt{\lambda_i}(a, e_i)_K \beta_i^h(t) + \frac{1}{2}\sum_{i=1}^{\infty} \lambda_i (a, e_i)_K^2 t^{2h}\right\} \exp\{\beta(t)\}\, \beta'(t) \\
&= Y(t)\cdot\beta'(t) + Y(t)\diamond(a, W^h(t))_K + Y(t)\cdot h\sum_{i=1}^{\infty} \lambda_i (a, e_i)_K^2 t^{2h-1}. \quad (3.6)
\end{aligned}$$

Hence we have found the following representation

$$\begin{aligned}
Y(t) &= Y(0) + \int_0^t Y(s)\cdot\beta'(s)\,ds + h\int_0^t Y(s)\cdot\sum_{i=1}^{\infty} \lambda_i (a, e_i)_K^2 s^{2h-1}\,ds \\
&\quad + \int_0^t Y(s)\diamond(a, W^h(s))_K\,ds. \quad (3.7)
\end{aligned}$$

Remembering (3.3) this can be written as

$$\begin{aligned}
g(t, B^h(t)) &= g(0,0) + \int_0^t \nabla_s g(s, B^h(s))\,ds + \int_0^t \left(\nabla_x g(s, B^h(s)), dB^h(s)\right)_K \\
&\quad + h\sum_{i=1}^{\infty}\int_0^t \left(\nabla_{xx} g(s, B^h(s))e_i, e_i\right)_K \lambda_i s^{2h-1}\,ds, \quad (3.8)
\end{aligned}$$

which is (3.2).

Now let $f(t,x)$ be as demanded above. Every function $f \in C^{1,2}\left([0,T]\times K, \mathbb{R}^1\right)$ can be approximated by a sequence of linear combinations of type (3.3), hence we can find a sequence of linear combinations

$f_n(t,x)$ of functions $g(t,x)$ of the form (3.3) such that

$$f_n(t,x) \to f(t,x), \;\; \nabla_t f_n(t,x) \to \nabla_t f(t,x), \;\; \nabla_x f_n(t,x) \to \nabla_x f(t,x), \\ \nabla_{xx} f_n(t,x) \to \nabla_{xx} f(t,x)$$

pointwise dominatedly as $n \to \infty$. By (3.8) we have for all n

$$f_n(t, B^h(t)) = f_n(0,0) + \int_0^t \left(\nabla_x f_n(s, B^h(s)), dB^h(s)\right)_K \\ +h \sum_{i=1}^{\infty} \int_0^t \left(\nabla_{xx} f_n(s, B^h(s)) e_i, e_i\right)_k \lambda_i s^{2h-1}\, ds + \int_0^t \nabla_s f_n(s, B^h(s))\, ds \quad (3.9)$$

Taking the limit of (3.9) in $L^2_Q(K, \mathbb{R}^1)$ (and therefore also in $(S^h_Q(\mathbb{R}^1))^*$) we get

$$f(t, B^h(t)) = f(0,0) + \lim_{n\to\infty} \int_0^t \left(\nabla_x f_n(s, B^h(s)), dB^h(s)\right)_K \\ +h \sum_{i=1}^{\infty} \int_0^t \left(\nabla_{xx} f(s, B^h(s)) e_i, e_i\right)_K \lambda_i s^{2h-1}\, ds + \int_0^t \nabla_s f(s, B^h(s))\, ds. \quad (3.10)$$

Since the mapping $s \to \nabla_x f(s, B^h(s))$ is continuous in $(S^h_Q(\mathbb{R}^1))^*$ we get

$$\begin{aligned} \int_0^t \left(\nabla_x f_n(s, B^h(s)), dB^h(s)\right)_K &= \int_0^t \left(\nabla_x f_n(s, B^h(s)), W^h(s)\right)_K ds \\ &\to \int_0^t \left(\nabla_x f(s, B^h(s)), W^h(s)\right)_K ds \end{aligned}$$

for $n \to \infty$ in $(S^h_Q(\mathbb{R}^1)^*)$. The last relation and (3.10) show (3.2). □

Example 3.2. Now let $f(s,x) : [0,T] \times K \to \mathbb{R}$ be defined as follows:

$$f(t,x) := \exp(t+x),$$

then we have

$$\nabla_t f(t,x) = \nabla_x f(t,x) = \nabla_{xx} f(t,x) = \exp(t+x),$$

and therefore we have by (3.2)

$$\begin{aligned}
f(t,B^h(t)) &= 1+\int_0^t \exp(s+B^h(s))\,ds \\
&\quad +\int_0^t \left(\exp(s+B^h(s)),dB^h(s)\right)_K \\
&\quad +h\sum_{i=1}^{\infty}\int_0^t \left(\exp(s+B^h(s))e_i,e_i\right)_K \lambda_i s^{2h-1}\,ds \\
&= 1+\int_0^t \exp(s+B^h(s))\,ds \\
&\quad +\int_0^t \left(\exp(s+B^h(s)),W^h(s)\right)_{\diamond K}\,ds \\
&\quad +h\sum_{i=1}^{\infty}\int_0^t \left(\exp(s+B^h(s))e_i,e_i\right)_K \lambda_i s^{2h-1}\,ds.
\end{aligned}$$

Example 3.3. Now let $f(s,x):[0,T]\times K\to \mathbb{R}$ be defined as follows:

$$f(t,x):=\ln\left(1+x^2\right),$$

then we have

$$\nabla_t f(t,x)=0,\quad \nabla_x f(t,x)=\frac{2x}{1+x^2}\text{ and }\nabla_{xx}f(t,x)=\frac{2-2x^2}{(1+x^2)^2},$$

and therefore we have by (3.2)

$$\begin{aligned}
f(t,B^h(t)) &= 0+\int_0^t \left(\frac{2B^h(s)}{1+\left(B^h(s)\right)^2},W^h(s)\right)_{\diamond K} ds \\
&\quad +h\sum_{i=1}^{\infty}\int_0^t \left(\frac{2-2\left(B^h(s)\right)^2}{\left(1+\left(B^h(s)\right)^2\right)^2}e_i,e_i\right)_K \lambda_i s^{2h-1}\,ds.
\end{aligned}$$

References

[1] Anh, V.V., Grecksch, W., Roth, C., *Q-fractional Brownian Motion in Infinite Dimensions with Applications to the Fractional Black-Scholes Market*, Stoch. Anal. Appl., **27**(2009), no. 1, 149–175.

[2] Biagini, F., Øksendal, B., Sulem, A., Wallner, N., *An introduction to white noise theory and Malliavin calculus for fractional Brownian motion*, Proc. R. Soc. Lond., Ser. A, Math. Phys. Eng. Sci., **460**(2004), no. 2041, 347–372.

[3] Duncan, T.E., Maslowski, B., Pasic-Duncan, B., *Fractional Brownian motion and stochastic equations in a Hilbert space,* Stochastics and Dynamics, **2**(2002), no. 2, 225-250.

[4] Elliott, R.J., van der Hoek, J., *A general fractional white noise theory and apllications to finance*, Mathematical Finance, **13**(2003), no. 2, 301–330.

[5] Grecksch, W., Roth, C., *Galerkin approximation for nonlinear parabolic equation with fractional white noise,* Institute of Mathematics, Martin-Luther-University Halle-Wittenberg, 2008.

[6] Grecksch, W., Anh, V.V., *A parabolic stochastic differential equation with fractional Brownian motion input,* Statist. Probab. Letters, **41**(1999), 337-345.

[7] Hida, T., *Brownian Motion*, Springer, New York, 1980.

[8] Hille, E.E., Phillips, R.S., *Functional analysis and semigroups,* Amer. Math. Soc. Colloq. Publ., **31**(1957).

[9] Holden, H., Øksendal, B., Ubøe, J., Zhang, T., *Stochastic Partial Differential Equations*, Birkhäuser, Boston, 1996.

[10] Hu, Y., Øksendal, B., *Fractional white noise calculus and applications to finance*, Infinite Dimens. Anal. Quant. Probab. Rel. Top., **6**(2003), 1–32.

[11] Hu, Y., Øksendal, B., Sulem, A., *Optimal consumption and portfolio in a Black-Scholes market driven by fractional Brownian motion*, Infin. Dimens. Anal. Quantum Probab. Relat. Top., **6**(2003), 519–536.

Wilfried Grecksch
Martin-Luther-University of Halle-Wittenberg
Faculty of Natural Sciences III
Institute of Mathematics
06099 Halle (Saale), Germany
e-mail: `grecksch@mathematik.uni-halle.de`

Christian Roth
Martin-Luther-University of Halle-Wittenberg
Faculty of Natural Sciences III
Institute of Mathematics
06099 Halle (Saale), Germany
e-mail: `roth@mathematik.uni-halle.de`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 381–391

Stochastic Schrödinger equation driven by cylindrical Wiener process and fractional Brownian motion

Wilfried Grecksch and Hannelore Lisei

Abstract. In this paper we study the properties of the solution of a stochastic nonlinear equation of Schrödinger type, which is perturbed by a cylindrical Wiener process and an additive cylindrical fractional Brownian motion with Hurst parameter in the interval $(\frac{1}{2}, 1)$. The existence of the solution and the existence of the Malliavin derivative are proved.

Mathematics Subject Classification (2010): 60H15, 60H07, 35Q41.

Keywords: Nonlinear stochastic Schrödinger equations, cylindrical Wiener process, cylindrical fractional Brownian motion, Malliavin derivative.

1. Introduction

In physics, specifically in quantum mechanics, the Schrödinger equation is an equation that describes how the quantum state of a physical system changes in time.

We describe the Schrödinger equation for a harmonic oscillator subject to a periodic electric field: a particle of mass m, electric charge Q, is displaced along the x-axis ($x \in \mathbb{R}$) and subject to a force $-m\omega_0^2 x$ (for all $t > 0$) and to an electric field $E\sin(\omega t)$ directed along the x-axis

$$i\hbar\frac{\partial}{\partial t}X(x,\, t) = \Big(-\frac{\hbar^2}{2m}\nabla^2 + \frac{1}{2}m\omega_0^2 x^2 + QEx\sin(\omega t)\Big)X(x,\, t), \quad x \in \mathbb{R}, t > 0,$$

$$X(\cdot, 0) = X_0$$

where i is the imaginary unit, $-\dfrac{\hbar^2}{2m}\nabla^2$ is the kinetic energy operator, $\hbar$ is Planck's constant, the complex valued function X is the wave function at position x at time t, X_0 is the initial condition (see [8], p. 639).

Many authors investigated stochastic equations of Schrödinger type: The case of additive noise is considered in [11], [13], while the case of multiplicative noise is discussed in [2], [9], [10], [16]. In these papers the existence of a mild solution is investigated. Different approaches to linear and nonlinear stochastic Schrödinger equations perturbed by cylindrical Brownian motions are given in [14] and [15].

In this paper we study the properties of the solution of a stochastic nonlinear equation of Schrödinger type, which is perturbed by a cylindrical Wiener process and an additive cylindrical fractional Brownian motion. Consequently, this model respects as well fluctations of a Brownian motion as additive disturbances with long range dependence. This paper completes the results about stochastic equations of Schrödinger type given in [5] by considering also a cylindrical fractional Brownian motion with Hurst parameter in the interval $(\frac{1}{2}, 1)$. We use the framework of stochastic evolution equations driven by fractional noise developed by T.E. Duncan, B. Pasik-Duncan, B. Maslowski [12] and M. Röckner and Y. Wang [17]. The existence results are derived by using the properties of Schrödinger type equations developed in [5]. Smoothness properties such as the existence of the Malliavin derivative are also proved. The Malliavin derivatives can be used to calculate conditional expectations or chaos decompositions of stochastic processes (see [3], [7]).

This paper has the following structure: In Section 2 we introduce the list of assumptions and give the definition of the solution. In Section 3 we briefly present the two stochastic integrals that appear in the equation which is investigated. The existence of the solution is derived in Section 4. Section 5 contains results about infinite dimensional Malliavin derivatives and the existence of the Malliavin derivative of the solution is proved.

2. Assumptions and formulation of the problem

We consider $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ to be a filtered complete probability space. Let $(V, (\cdot,\cdot)_V)$ and $(H, (\cdot,\cdot))$ be separable complex Hilbert spaces, such that (V, H, V^*) forms a triplet of rigged Hilbert spaces. Let K be a separable real Hilbert space. We consider $\Big(W(t)\Big)_{t\geq 0}$ to be a K-valued cylindrical Wiener process adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ and $(B^h(t))_{t\geq 0}$ to be a K-valued cylindrical fractional Brownian motion with Hurst index $h \in (\frac{1}{2}, 1)$ adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$.

We study the properties of the *variational solution* X of the following stochastic nonlinear evolution equation of Schrödinger type

$$
\begin{aligned}
(X(t), v) = (X_0, v) - i\int_0^t \langle AX(s), v\rangle ds + i\int_0^t (f(s, X(s)), v)ds \qquad (2.1)\\
+ i(\int_0^t g(s, X(s))dW(s), v) + i(\int_0^t b(s)dB^h(s), v)
\end{aligned}
$$

for a.e. $\omega \in \Omega$ and all $t \in [0,T], v \in V$.

We assume that:

[I] X_0 is $\mathcal{F}_0$-measurable, $X_0 \in L^2(\Omega; V)$;

[A] $A : V \to V^*$ has the following properties:

• A is linear and continuous $\|Au\|_{V^*} \leq c_A \|u\|_V$ for all $u \in V$;

• $\langle Au, v\rangle = \overline{\langle Av, u\rangle}$ for all $u, v \in V$;

• there exists constants $\alpha_1 \in \mathbb{R}$ and $\alpha_2 > 0$, such that for all $v \in V$ it holds

$$\langle A(v), v\rangle \geq \alpha_1 \|v\|^2 + \alpha_2 \|v\|_V^2.$$

• Let $(h_n)_n \subset H$ be the eigenvectors of the operator A, for which we assume that $Ah_n \in H$ for all $n \in \mathbb{N}$ and $(h_n)_n$ is a complete orthonormal system in H.

[f] $f : \Omega \times [0,T] \times H \to H$ is a measurable function, which is $\mathcal{F}_t$-adapted for each $t \in [0,T]$:

(1) there exists a constant $c_f > 0$ such that for a.e. $\omega \in \Omega$ it holds

$$\|f(t,u) - f(t,v)\|^2 \leq c_f \|u - v\|^2 \quad \text{for all } t \in [0,T], u, v \in H;$$

(2) for a.e. $\omega \in \Omega$ and all $t \in [0,T], u \in V$ we have $f(t,u) \in V$ and there exists $k_f > 0$ such that

$$\|f(t,u)\|_V^2 \leq k_f (1 + \|u\|_V^2);$$

[g] $g : \Omega \times [0,T] \times H \to L_2(K,H)$ is a measurable function, which is $\mathcal{F}_t$-adapted for each $t \in [0,T]$:

(1) there exists a constant $c_g > 0$ such that for a.e. $\omega \in \Omega$ it holds

$$\|g(t,u) - g(t,v)\|_{L_2(K,H)}^2 \leq c_g \|u - v\|^2 \text{ for all } t \in [0,T], u, v \in H;$$

(2) for a.e. $\omega \in \Omega$ and all $t \in [0,T]$, $u \in V$ we have $g(t,u) \in L_2(K,V)$ and there exists $k_g > 0$ such that

$$\|g(t,u)\|_{L_2(K,V)}^2 \leq k_g (1 + \|u\|_V^2);$$

[b] $b : [0,T] \to L_2(K,V)$ and for each $u \in K$ we have $b(\cdot)u \in L^p([0,T];V)$ for some $p > \frac{1}{h}$ and it holds

$$\int_0^T \int_0^T \|b(r)\|_{L_2(K,V)} \|b(s)\|_{L_2(K,V)} |r - s|^{2h-2} dr ds < \infty.$$

3. The stochastic integrals

In this section we briefly present the definitions of the stochastic integrals we considered in (2.1). Let $(e_n)_n$ be an orthonormal basis in K.

For the K-valued cylindrical Wiener process $(W(t))_{t\geq 0}$ and for $g : \Omega \times [0,T] \times H \to L_2(K,H)$ satisfying [g]-(1) the stochastic integral

$\int\limits_0^T g(s,v)dW(s)$ ($v \in H$ fixed) is defined as a zero mean H-valued Gaussian random variable given by

$$\int\limits_0^T g(s,v)dW(s) := \sum_{n=1}^{\infty} \int\limits_0^T g(s,v)e_n dw_n(s),$$

where the series above converges in $L^2(\Omega;H)$ and $((w_n(t))_{t\geq 0})_n$ is a sequence of mutually independent real-valued Brownian motions. One can prove that

$$E\Big\|\int\limits_0^T g(s,v)dW(s)\Big\|^2 = \sum_{n=1}^{\infty} E\Big\|\int\limits_0^T g(s,v)e_n dw_n(s)\Big\|^2$$

$$= \sum_{n=1}^{\infty} E\int\limits_0^T \|g(s,v)e_n\|^2 ds = E\int\limits_0^T \|g(s,v)\|^2_{L_2(K,H)} ds < \infty.$$

For $0 < r < 1/(2-2h)$ the function $\phi : [0,T] \to \mathbb{R}$ defined by $\phi(u) = h(2h-1)|u|^{2h-2}$ belongs to the space $L^r([0,T];\mathbb{R})$.

If $p > 1/h$, then by Theorem 3.9.4 in [4], there exists $C_T > 0$ such that for any function $\eta, \varphi \in L^p([0,T];\mathbb{R})$ it holds

$$\int\limits_0^T\int\limits_0^T |\eta(u)\varphi(v)\phi(u-v)|dudv \leq C_T\|\varphi\|_{L^p([0,T];\mathbb{R})}\|\eta\|_{L^p([0,T];\mathbb{R})}.$$

If $(\beta^h(t))_{t\geq 0}$ is a real-valued fractional Brownian motion with Hurst index $h \in (\frac{1}{2},1)$, and $\varphi \in L^p([0,T];\mathbb{R})$, then the stochastic integral $\int\limits_0^T \varphi(s)d\beta^h(s) \in L^2(\Omega;\mathbb{R})$ is defined as a zero mean real-valued Gaussian random variable, such that

$$E\left(\int\limits_0^T \varphi(s)d\beta^h(s)\int\limits_0^T \varphi(s)d\beta^h(s)\right) = E\int\limits_0^T\int\limits_0^T \varphi(u)\varphi(v)\phi(u-v)dudv.$$

If $\varphi \in L^p([0,T];\mathbb{R})$ with $p > \frac{1}{h}$, then the process $\Big(\int\limits_0^t \varphi(s)d\beta^h(s)\Big)_{t\geq 0}$ has P-a.s. continuous sample paths (see [18] Lemma 2.0.17).

Let $(k_n)_n$ be an orthonormal basis in K.

For the K-valued cylindrical fractional Brownian motion $(B^h(t))_{t\geq 0}$ and for $b : [0,T] \to L_2(K,V)$ satisfying assumption [b] the stochastic integral $\int\limits_0^T b(s)dB^h(s)$ is defined as a zero mean V-valued Gaussian random variable given by

$$\int_0^T b(s)dB^h(s) := \sum_{n=1}^{\infty} \int_0^T b(s)k_n d\beta_n^h(s),$$

where the series above converges in $L^2(\Omega; V)$ and $\left((\beta_n^h(t))_{t\geq 0}\right)_n$ is a sequence of mutually independent real-valued fractional Brownian motions each with Hurst parameter h. One can prove that

$$E\Big\|\int_0^T b(s)dB^h(s)\Big\|_V^2 = \sum_{n=1}^{\infty} E\Big\|\int_0^T b(s)k_n d\beta_n^h(s)\Big\|_V^2$$

$$= \sum_{n=1}^{\infty} \int_0^T \int_0^T (b(r)k_n, b(s)k_n)_V \phi(r,s)drds$$

$$\leq \int_0^T \int_0^T \|b(r)\|_{L_2(K,V)} \|b(s)\|_{L_2(K,V)} \phi(r,s)drds < \infty.$$

For more details see for example [12],[18].

For a.e. $\omega \in \Omega$ and for each $t \in [0,T]$ we denote by

$$Z(t) := \int_0^t b(s)dB^h(s),$$

which is obviously a V-valued process adapted to $(\mathcal{F}_t)_{t\geq 0}$.

Proposition 3.1. [18, Corollary 2.0.16, Lemma 2.0.17] *The process $(Z(t))_{t\in[0,T]}$ has a continuous version in V and in H and*

$$E\int_0^T \|Z(s)\|_V^2 ds < \infty.$$

Remark 3.2. *The stochastic integral $Z(t)$ can also be represented by a stochastic integral with respect to the cylindrical Wiener process W (see* [3], [6]*). For $f : \mathbb{R} \to \mathbb{C}$ and $\frac{1}{2} < h < 1$ we introduce the operator*

$$(M^h f)(x) = c_h \int_{\mathbb{R}} \frac{f(t)}{|t-x|^{3/2-h}} dt,$$

where $c_h = [2\Gamma(h-1/2)\cos(1/2\pi(h-1/2))]^{-1}(\Gamma(2h+1)\sin(\pi h))^{1/2}$ and f is chosen in such a manner that $(M^h f) \in L^2(\mathbb{R})$. If f is concentrated on $[0,T]$, then we consider $[0,T]$ instead of $\mathbb{R}$. If

$$\sum_{n=1}^{\infty}\sum_{j=1}^{\infty} \int_0^T \left(\left(M^h\left(b(\cdot)k_n, h_j\right)\right)(s)\right)^2 ds < \infty,$$

then

$$\int_0^t b(s)dB^h(s) = \sum_{j=1}^{\infty}\sum_{n=1}^{\infty} \int_0^t \left(M^h\left(b(\cdot)k_n, h_j\right)\right)(s)dw_n(s)h_j.$$

4. Existence of the solution

Theorem 4.1. *Assume that* [I], [A], [f], [g], [b] *are satisfied. Equation* (2.1) *admits a unique solution* $X \in L^2(\Omega \times [0,T];V) \cap L^2(\Omega; C([0,T];H))$.

In order to prove the existence of the solution of (2.1), we first transform it equivalently into an equation of Schrödinger type studied in [5]. For a.e. $\omega \in \Omega$ and for each $t \in [0,T], v \in H$ we denote by

- $U(t) := X(t) - iZ(t)$.
- $F(\omega,t,v) := f(\omega,t,v+iZ(\omega,t))$,
- $G(\omega,t,v) := g(\omega,t,v+iZ(\omega,t))$.

Observe that for a.e. $\omega \in \Omega$ and all $t \in [0,T], u, v \in H$ it holds

$$\|F(t,u) - F(t,v)\|^2 \le c_f \|u-v\|^2$$

$$\|G(t,u) - G(t,v)\|^2_{L_2(K,H)} \le c_g \|u-v\|^2$$

and for all $u \in V$

$$\|F(t,u)\|^2_V \le 2k_f(1 + \|u\|^2_V + \|Z(t)\|^2_V);$$

$$\|G(t,u)\|^2_{L_2(K,V)} \le 2k_g(1 + \|u\|^2_V + \|Z(t)\|^2_V).$$

We rewrite (2.1) equivalently as

$$\begin{aligned}(U(t),v) &= (X_0,v) - i\int_0^t \langle AU(s),v\rangle ds + i\int_0^t (F(s,U(s)),v)ds \qquad (4.1)\\ &\quad + i(\int_0^t G(s,U(s))dW(s),v) + i\int_0^t \langle AZ(s),v\rangle ds \text{ for all } v \in V.\end{aligned}$$

(2.1) admits a unique solution $X \in L^2(\Omega \times [0,T];V) \cap L^2(\Omega; C([0,T];H))$ if and only if (4.1) admits a unique solution $U \in L^2(\Omega \times [0,T];V) \cap L^2(\Omega; C([0,T];H))$.

The proof of the existence of a unique solution U for (4.1) is similar to the proof of Theorem 1 in [5]. For this reason one introduces Galerkin approximations: For each $n \in \mathbb{N}$ we consider the finite dimensional spaces $H_n := \mathrm{sp}\{h_1, h_2, \ldots, h_n\}$ (equipped with the norm induced from H) and $K_n := \mathrm{sp}\{e_1, e_2, \ldots, e_n\}$ (equipped with the norm induced from K). We define $\pi_n : H \to H_n$ the orthogonal projection of H on H_n by $\pi_n h := \sum_{j=1}^{n} (h,h_j)h_j$. Let $A_n : H_n \to H_n$, $F_n : \Omega \times [0,T] \times H_n \to H_n, G_n : \Omega \times [0,T] \times H_n \to L(K_n, H_n)$ be defined respectively by

$$A_n u = \sum_{j=1}^{n} \langle Au, h_j\rangle h_j, \quad F_n(t,u) = \sum_{j=1}^{n} (F(t,u),h_j)h_j,$$

$$G_n(t,u)v = \sum_{j=1}^{n} (G(t,u)v,h_j)h_i \text{ for } v \in K_n$$

$$Z_n(t) = \sum_{j=1}^{n} (Z(t),h_j)h_j$$

and we denote $X_{0n} = \pi_n X_0$ and $W_n(t) = \sum_{j=1}^{n} e_j w_j(t) \in K_n$. For a.e. $\omega \in \Omega$ and all $t \in [0,T]$ and all $j = \overline{1,n}$ we consider the finite dimensional equations corresponding to (4.1)

$$\begin{aligned}(U_n(t), h_j) \quad = \quad & (X_{0n}, h_j) - i\int_0^t (A_n U_n(s), h_j)ds \qquad (4.2)\\ & + i\int_0^t (F_n(s, U_n(s)), h_j)ds + i(\int_0^t G_n(s, U_n(s))dW_n(s), h_j)\\ & + i\int_0^t (A_n(s) Z_n(s), h_j)ds.\end{aligned}$$

One can show similar as in the proof of Theorem 1 in [5] (see also Remark 3 in [5]) that for all $t \in [0,T]$ it holds

$$\lim_{n\to\infty} E\|U_n(t) - U(t)\|^2 = 0$$

and

$$\lim_{n\to\infty} E\int_0^t \|U_n(s) - U(s)\|^2 ds = 0.$$

5. The existence of Malliavin derivative of the solution

We briefly present some results about infinite dimensional Malliavin derivatives: We consider the random variable Y with values in a complex Hilbert space H. Y with $E\|Y\|^2 < \infty$ is called a smooth random variable and we denote $Y \in \mathcal{S}$, if

$$Y = \sum_{j=1}^{n} f_j \left(\int_0^T (\gamma_{1,j}(s), dW(s))_K, \dots, \int_0^T \left(\gamma_{n_j,j}(s), dW(s)\right)_K \right) h_j,$$

where $\gamma_{1,j}, \dots, \gamma_{n_j,j} \in L^2([0,T];K)$ for $j = 1, \dots n$, $h_j \in H$, $f_j \in C^\infty(R^{n_j})$ and f_j and all its derivatives have polynomial growth for $j = 1, \dots, n$.

The Malliavin derivative $D_t Y$, ($t \in [0,T]$) of $Y \in \mathcal{S}$ is a random variable with values in $L_2(K,H)$ defined by

$$D_t Y = \sum_{j=1}^{n}\sum_{k=1}^{n_j} \frac{\partial f_j}{\partial x_k} \left(\int_0^T (\gamma_{1,j}(s), dW(s))_K, \dots, \int_0^T \left(\gamma_{n_j,j}(s), dW(s)\right)_K \right) \cdot \\ \cdot h_j \otimes \gamma_{k,j}(t).$$

The Malliavin derivative D_t as defined for H-valued smooth random variables is closable on $L^2(\Omega; L_2(K,H))$ (see Proposition 5.1 in [7]).

Consequently, if Y is the $L^2(\Omega; H)$ limit of a sequence $(Y_n)_n \subset \mathcal{S}$ so that the sequence $(D_t Y_n)_n$ convergences in $L^2(\Omega; L_2(K,H))$, we can define $D_t Y$ as

$$D_t Y = \lim_{n\to\infty} D_t Y_n.$$

We use the notation $H(K)$ for the subspace of $L^2(\Omega; H)$, where the derivative D_t can be defined. This subspace is a separable Hilbert space equipped with the graph norm

$$\|Y\|_{H(K)}^2 = E\|Y\|^2 + E\|D_t Y\|_{L_2(K,H)}^2.$$

The following result is known (see Lemma 5.2 in [7]):

Lemma 5.1. *Let $Y_n \to Y$ in $L^2(\Omega; H)$ and suppose that there is a constant $C > 0$ such that for all n we have*

$$E\|D_t Y\|_{L_2(K,H)}^2 < C.$$

Then, the random variable Y is in the domain $H(K)$ of the Malliavin derivative D_t.

By using Proposition 5.2 in [7] the following chain rule holds:

Proposition 5.2. *Let M be a further separable Hilbert space. Given a random variable $Y \in H(K)$ and a Fréchet differentiable function $\eta : H \to M$. Then,*

$$D_t \eta(Y) = \nabla \eta D_t Y.$$

We will use the following well-known properties of D_t (see, for example [7], [3]):

Proposition 5.3. (1) *If Y is $\mathcal{F}_s$-measurable and $Y \in H(K)$, then $D_t Y = 0$ a.e. $\omega \in \Omega$ and for all $t > s$.*

(2) *Let $a(s)$, $s \in [0, T]$ an $\mathcal{F}_s$-adapted $L_2(K, H)$-valued process which fulfills the assumptions of the Skorochod integral definition in* [7]*. Then, for all $r > t$ it holds*

$$D_t \int_0^r a(s) dW(s) = a(t) + \int_t^r D_t a(s) dW(s).$$

Further in this section we assume:

1. The assumption in Remark 3.2 is valid for the process b.
2. The functions f and g are deterministic.
3. The functions f and g are Fréchet differentiable with respect to $x \in H$ for all $t \in [0, T]$ and the Fréchet derivatives $\nabla_x f(t, x)$ and $\nabla_x g(t, x)$ are bounded in the following sense: There exists a positive constant c such that
$$\|\nabla_x f(t, x)\|_{L(H,H)},\ \|\nabla_x g(t, x)\|_{L(H,L_2(K,H))} \le c$$
for all $t \in [0, T]$, $x \in H$.
4. The initial condition X_0 is deterministic.

Theorem 5.4. *There exists $D_r U(t)$ as an $L_2(K, H)$-valued random variable for all $r, t \in [0, T]$.*

Proof. We process the proof in two steps:
Step 1: It follows from the above assumption 3 that the functions f and g are globally Lipschitz continuous. Consequently, we can consider directly

the Galerkin equations (4.2). Similar to Remark 3 in [5] we have for the variational solution U

$$\lim_{n\to\infty} E\|U_n(t)-U(t)\|^2 = 0 \text{ and } \lim_{n\to\infty} E\int_0^t \|U_n(s)-U(s)\|^2 ds = 0 \quad (5.1)$$

for all $t \in [0,T]$. Equation (4.2) is an Itô equation in V_n and H_n and its solution can be approximated by the method of successive approximations

$$\begin{aligned} U_n^{m+1}(t) &= X_{0n} - i\int_0^t A_n U_n^m(s)ds \qquad (5.2)\\ &\quad +i\int_0^t F_n(s,U_n^m(s))ds + i\int_0^t G_n(s,U_n^m(s))dW_n(s)\\ &\quad +i\int_0^t A_n(s)Z_n(s)ds. \end{aligned}$$

for $m = 0, 1, \ldots$ with $U^0(s) \equiv X_{0n}$.

The finite dimensional theory shows

$$\lim_{m\to\infty} E\|U_n^m(t)-U_n(t)\|^2 = 0. \quad (5.3)$$

Now we calculate $D_r U_n^{m+1}(t)$. Since U_n^{m+1} is $\mathcal{F}_t$-measurable we get also the $\mathcal{F}_r$-measurability for $r \geq t$. In this case it follows from Proposition 5.3 $D_r U_n^{m+1}(t) = 0$. We now consider $r < t$. Then, by Proposition 5.2, Proposition 5.3 and Remark 3.2 we get

$$\begin{aligned} D_r U_n^{m+1}(t) &= -i\int_r^t A_n D_r U_n^m(s)ds \qquad (5.4)\\ &\quad +i\int_r^t \nabla_x F_n(s,U_n^m(s))D_r U_n^m(s)ds\\ &\quad +i\int_r^t \nabla_x F_n(s,U_n^m(s))D_r Z_n(s)ds\\ &\quad +i\int_r^t \nabla_x G_n(s,U_n^m(s))D_r U_n^m(s)dW_n(s)\\ &\quad +i\int_r^t \nabla_x G_n(s,U_n^m(s))D_r Z_n(s)dW_n(s)\\ &\quad +iG_n(r,U_n^m(r)) + i\int_r^t A_n(s)D_r Z_n(s)ds \end{aligned}$$

where $D_r Z_n(t) : K_n \to H_n$ is the linear operator defined by

$$(D_r Z_n(t)x, y) = \left(M^h\left(b_n(\cdot)x, y\right)\right)(s).$$

$D_r Z_n(t)$ has values in $L(K_n, V_n)$ and $L(K_n, H_n)$. Since the spaces are finite dimensional, the operators are also Hilbert-Schmidt operators. If we use the energy equality in the space $L_2(K_n, H_n)$, then we get by the assumptions of this section and by Gronwall's lemma that there is a positive constant C with

$$E\|D_r U_n^m(t)\|^2_{L_2(K,H)} \leq C$$

for all m, r, t and fixed n, since from equation (5.3) the boundedness of $E\|U_n^m(t)\|^2$ follows for all m, r, t and fixed n. The constant C does not depend on n. Then we get by Lemma 5.1, from the last inequality and from equation (5.3) that $D_r U_n(t)$ exists and

$$E\|D_r U_n(t)\|^2_{L_2(K,H)} \leq C. \tag{5.5}$$

Step 2: Since the relations (5.5) and (5.1) hold, we can use again Lemma 5.1 and get

$$E\|D_r U(t)\|^2_{L_2(K,H)} \leq C.$$

□

Theorem 5.5. *Consider that the assumptions of this section hold. Then, for $t > r$ we have*

$$D_r X(t) = D_r U(t) + i(M^h b(\cdot))(r),$$

where $(M^h b(\cdot))(r) \in L_2(K, H)$ is defined by the bilinearform

$$(M^h(b(\cdot)x, y))(r) \text{ for all } x \in K,\ x \in H.$$

Proof. Theorem 5.4 shows the existence of $D_r U(t)$ and it holds $D_r X(t) = D_r U(t) + iD_r Z(t)$. Since b is deterministic, we get by Proposition 5.3 and Remark 3.2 for $t > r$

$$D_r Z(t) = (M^h b(\cdot))(r).$$

□

Remark 5.6. The Malliavin derivative is used for example to define Skorochod integrals [12] and in the optimal control theory [1]. Optimal control problems for stochastic Schrödinger equations are under preparation.

Acknowledgement. This work was supported by CNCSIS - UEFISCSU (Romania) project number PN II IDEI ID 2162/nr. 501/2008 "Nonsmooth phenomena in nonlinear elliptic problems" and by Deutsche Forschungsgemeinschaft (Germany) project GR 1525/10-1.

References

[1] Anh, V.V., Grecksch, W., Yong, J., *Regularity of Backward Stochastic Volterra Integral Equations*, Stochastic Analysis and Applications, **29**(2011), no. 1, 146–168.

[2] Bang, O., Christiansen, P.L., If, F., Rasmussen, K.O., Gaididei, Y.B., *Temperature Effects in a Nonlinear Model of Monolayer Scheibe Aggregates*, Phys. Rev. E, **49**(1994), 4627-4636.

[3] Biagini, F., Oksendal, B., Sulem, A., Wallner, N., *An Introduction to White Noise Theory and Malliavin Calculus for Fractional Brownian Motion*, Proc. Royal Soc. London, A, **460**(2004), 347-372.

[4] Bogachev, V.I., *Measure Theory*, Vol. I, Springer-Verlag, New-York, 2007.

[5] Grecksch, W., Lisei, H., *Stochastic Nonlinear Equations of Schrödinger Type*, to appear in Stochastic Analysis and Applications.

[6] Grecksch, W., Roth, C., Anh, V.V., *Q-Fractional Brownian Motion in Infinite Dimensions with Applications to Fractional Black-Scholes Market*, Stochastic Analysis and Applications, **27**(2009), no. 1, 149-175.

[7] Carmona, R., Tehrani, M., *Interest Rate Models: an Infininite Dimensional Stochastic Analysis Perspective*, Springer Verlag, Berlin - Heidelberg, 2006.

[8] Dautray, R., Lions, J.L., *Mathematical Analysis And Numerical Methods For Science And Technology*, Volume 5, Evolution Problems I. Springer-Verlag, Berlin, 1992.

[9] De Bouard, A., Debussche, A., *A Stochastic Nonlinear Schrödinger Equation with Multiplicative Noise*, Commun. Math. Phys., **205**(1999), 161–181.

[10] De Bouard, A., Debussche, A., *A Semi-discrete Scheme for the Stochastic Nonlinear Schrödinger Equation,* Numer. Math., **96**(2004), 733–770.

[11] Debussche, A., Odasso, C., *Ergodicity for a Weakly Damped Stochastic Nonlinear Schrödinger Equation*, J. Evol. Eq., **5**(2005), 317-356.

[12] Duncan, T.E., Maslowski, B., Pasic-Duncan, B., *Fractional Brownian Motion and Stochastic Equations in Hilbert Space,* Stochast. Dyn., **2**(2002), 225–250.

[13] Falkovich, G.E., Kolokolov, I., Lebedev, V., Turitsyn, S.K., *Statistics of Soliton-Bearing Systems with Additive Noise*, Phys. Rev. E, **63**(2001).

[14] Mora, C.M., Rebolledo, R., *Regularity of Solutions to Linear Stochastic Schrödinger Equations*, Infin. Dimens. Anal. Quantum Probab. Relat. Top., **10**(2007), 237-259.

[15] Mora, C.M., Rebolledo, R., *Basic Properties of Nonlinear Stochastic Schrödinger Equations Driven by Brownian Motions,* Ann. Appl. Probab., **18**(2008), no. 2, 591-619.

[16] Rasmussen, K.O., Gaididei, Y.B., Bang, O., Christiansen, P.L., *The Influence of Noise on Critical Collapse in the Nonlinear Schrödinger Equation*, Physics Letters A, **204**(1995), 121–127.

[17] Röckner, M., Wang, Y., *A Note on Variational Solutions to SPDE Perturbed by Gaussian Noise in a General Class*, Infin. Dimens. Anal. Quantum Probab. Relat. Top., **12**(2009), no. 2, 353–358.

[18] Wang, Y., *Variational Solutions to SPDE Perturbed by a General Gaussian Noise,* PhD thesis, Purdue University, 2009.

Wilfried Grecksch
Martin-Luther University Halle-Wittenberg
Faculty of Sciences III, Institute of Mathematics
D - 06099 Halle
Germany
e-mail: `wilfried.grecksch@mathematik.uni-halle.de`

Hannelore Lisei
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Science
1, Kogălniceanu Street
400084 Cluj-Napoca
Romania
e-mail: `hanne@math.ubbcluj.ro`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 393–410

Stochastic simulation of the gradient process in semi-discrete approximations of diffusion problems

Flavius Guiaş

Abstract. We analyze a stochastic version of the so-called diffusion-velocity method. For moving particles with velocities depending on the gradient of their density function we introduce a stochastic scheme based on the simulation of the gradient process where the values of the density are recovered by a numerical integration method. We apply this method to the diffusion equation and show a convergence result.

Mathematics Subject Classification (2010): 65C20, 68U20, 65M06.

Keywords: Stochastic simulation, diffusion, gradient process.

1. Introduction

Numerical particle methods are suited to approximate the time evolution of a density function by simulating trajectories of the (interacting) particles. We assume that the velocity vector depends (linearly or nonlinearly) on the gradient of this density function. For example in the so-called *deterministic diffusion* of particles one assigns to every particle a velocity vector which is proportional to the logarithmic derivative of the density function. This principle can be derived as follows: consider in a volume element $U \subset \mathbb{R}^d$ the *continuity equation*: $u_t = -\nabla \cdot (uv)$, which describes the motion with velocity v of a quantity with density function u. Pure formally, if we put $v = -\frac{\nabla u}{u}$ we obtain nothing else than the diffusion equation: $u_t = \Delta u$. This type of velocity is often called in the physical literature as *osmotic velocity.* A discussion of this deterministic particle method together with several applications is presented in [4] and [5] and a rigorous convergence result is proved in [3].

Our goal is to construct an analogous scheme in a stochastic framework. The main motivation is that this scheme may be used in combination with usual Monte Carlo methods for kinetic equations in a spatially inhomogeneous setting, with spatial motion and local interaction.

After moving a particle from a location to another the density configuration changes, so one has also to update the corresponding values of the gradient. In a stochastic framework for particle simulation, the use of standard discretization schemes for computing the gradient can lead to strong oscillations in the density values. In order to avoid this problem we consider the simulation of the gradient process. Given an initial data in terms of the density, one computes the gradient by a usual discretization scheme (e.g. finite differences). The scheme follows then only the dynamics of the gradient process which are derived from the original dynamics of the particle system. The density is recovered by a numerical integration method.

In this paper we present an application of the principle described above to the diffusion equation. By this example we intend also to develop a formalism and a methodology which can be applied in more general cases.

In Section 2 introduce the stochastic counterpart of the diffusion-velocity method. We point out that this direct approach in the stochastic framework leads to strong fluctuations of the density profile which we want to approximate. In order to overcome this problem we introduce in Section 3 a stochastic scheme for the gradient process. Convergence results for the density computed by numerical integration are presented in Section 4.

2. The diffusion-velocity method in a stochastic framework

In this section we will present an approach to approximate the diffusion equation in the one-dimensional case, which can be extended easily to higher dimensions. The goal is to approximate on the interval $(0,1)$ the solution of:

$$u_t = u_{xx}, \text{ with the boundary condition: } \begin{cases} u(0) = u(1) = 0 \ \ (D) \\ \text{or} \\ u_x(0) = u_x(1) = 0 \ \ (N) \end{cases} \tag{2.1}$$

and initial condition $u(0,x) = u_0(x) \in H^1(0,1)$ for all $x \in [0,1]$.

In this section we will assume that $u_0 \geq 0$. Let M be an integer, denote $\varepsilon = M^{-1}$ and consider the discrete set of sites $G_\varepsilon = \{k\varepsilon, k = \overline{1, M-1}\}$. Assume that we have N particles distributed in the locations of G_ε, and denote by $n^k(t)$ the number of particles present at the moment t in the location $k\varepsilon$. We introduce the scaling parameter $h = M/N = \varepsilon^{-1}N^{-1}$, which means that h^{-1} is the average number of particles per site. The density function corresponding to this particle system is defined in the points $k\varepsilon$ of the discretization grid by $u^k(t) = hn^k(t)$. In analogy to the formula $v = -\frac{\nabla u}{u}$ we assign to the particles located at $k\varepsilon$, $k = \overline{1, M-1}$, the velocity

$$v^k(t) = \frac{u^{k-1}(t) - u^{k+1}(t)}{2\varepsilon u^k(t)}.$$

Sometimes we will consider formal function values at the *boundary sites* 0 and $M\varepsilon = 1$ (which contain no particles), in order to model the boundary conditions. These values influence the transitions in the sites ε and $(M-1)\varepsilon$.

Every particle situated at $k\varepsilon$ can jump one site to the left or to the right (depending on the sign of the velocity). Our interest is however to follow the time evolution of the density function and from this viewpoint all particles present in the same location are indistinguishable. That is, if any particle from the site $k\varepsilon$ jumps with the rate $Mv^k = \varepsilon^{-1}|v^k|$, the density function in the new state will be the same, independently on which particle jumped. We can thus consider a single transition of this type and multiply the rate with $n^k(t)$, i.e. with the number of particles present at time t at the site $k\varepsilon$.

2.1. Construction of the Markov jump process

Based on the previous considerations, we will construct two $\mathbb{R}^{M-1}$-valued Markov jump processes as follows. Given the time moment t and the state $u(t) = (u^k(t))_{k=1}^{M-1}$, we define

$$w^k(t) = hn^k(t) \cdot v^k(t) = \frac{u^{k-1}(t) - u^{k+1}(t)}{2\varepsilon} =: -\nabla_\varepsilon u^k(t) \tag{2.2}$$

for $k = \overline{1, M-1}$.

The transitions in the interior sites are given by:

$$u(t) \to u(t) - he_k + he_{k+\zeta(w^k)} \text{ at rate } h^{-1}\varepsilon^{-1}|w^k(t)| \tag{2.3}$$

where e_i denotes the i-th unit vector in $\mathbb{R}^{M-1}$, while $\zeta(\cdot)$ denotes the *signum* function. This corresponds to the jump of a particle from the site k in the site $k + \zeta(w^k)$, for $k = \overline{1, M-1}$.

The quantity $w^k(t)$ defined in (2.2) represents the discrete derivative of the density function $u^k(t)$ and the transition (2.3) changes this function in the locations $k, k + \zeta(w^k), k + 2\zeta(w^k), k - \zeta(w^k)$ as follows:

$$\begin{aligned}
w^k &\to -\zeta(w^k)\frac{u^{k+\zeta(w^k)} + h - u^{k-\zeta(w^k)}}{2\varepsilon} = w^k - \frac{h}{2\varepsilon}\zeta(w^k) \\
w^{k+\zeta(w^k)} &\to -\zeta(w^k)\frac{u^{k+2\zeta(w^k)} - u^k + h}{2\varepsilon} = w^{k+\zeta(w^k)} - \frac{h}{2\varepsilon}\zeta(w^k) \\
w^{k+2\zeta(w^k)} &\to -\zeta(w^k)\frac{u^{k+3\zeta(w^k)} - u^{k+\zeta(w^k)} - h}{2\varepsilon} = w^{k+2\zeta(w^k)} + \frac{h}{2\varepsilon}\zeta(w^k) \\
w^{k-\zeta(w^k)} &\to -\zeta(w^k)\frac{u^k - h - u^{k-2\zeta(w^k)}}{2\varepsilon} = w^{k-\zeta(w^k)} + \frac{h}{2\varepsilon}\zeta(w^k)
\end{aligned} \tag{2.4}$$

at rate $h^{-1}\varepsilon^{-1}|w^k(t)|$.

We will discuss next the **situation at the boundary**.

In the case of *zero boundary conditions* (D), we consider formally $u^i(t) = 0$ for all $i \notin \{1, \ldots, M-1\}$ in all expressions from (2.3). This value does not change after any possible transition, that is, the particle which leaves the interior of the domain is 'killed'. Outside the range $\overline{1, M-1}$ the function w is not defined. Consider $w^1(t) = -\frac{1}{2\varepsilon}u^2(t)$ and $w^{M-1}(t) = \frac{1}{2\varepsilon}u^{M-2}(t)$. Note that always holds $w^1 \leq 0$ and $w^{M-1} \geq 0$. This implies that in the case $k = 1$ or $k = M-1$, the changes of $w^{k-\zeta(w^k)}$ are well defined in (2.4), while $w^{k+\zeta(w^k)}$ and $w^{k+2\zeta(w^k)}$ are not present, the indices being out of range.

In the case of *Neumann boundary conditions* (N) we take

$$w^1(t) = w^{M-1}(t) = 0.$$

2.2. Remarks

As the following considerations will show, the density function u may exhibit strong oscillations. Suppose that we have an even number of interior locations $\{1, \dots, 2M\}$, while the sites 0 and $2M+1$ correspond to the boundary. Expressing u in terms of w in the case of 0 boundary conditions we obtain

$$u^{2k} - u^{2k+1} = -2\varepsilon \left(\sum_{i=1}^{k} w^{2i-1} + \sum_{i=k+1}^{M} w^{2i} \right). \tag{2.5}$$

The value of the difference in (2.5) approaches the integral $-\int_0^1 w dx$. Since the expected limits satisfy $w = -u_x$, the difference will be close to $u(1) - u(0)$, up to a factor of $O(\varepsilon)$. The same statement holds for Neumann boundary conditions, where the computations are similar, but the expressions for the values of u in the sites $\overline{2, 2M-1}$ involve also the values of u^1 and u^{2M} (which cannot be computed directly from w). In this situation, especially in the case of asymmetric initial data, the value of the difference can be of $O(1)$. If the integral is nonzero, the difference $u^{2k} - u^{2k+1}$ will have basically a constant sign, which means that one can observe a strongly oscillating pattern. Only in the case that the integral vanishes (in our setting only for zero boundary conditions or symmetric data) the oscillations have a smaller amplitude. In this case, after some elementary computations, the difference (2.5) can be estimated by $\varepsilon(\|w\|_\infty + \frac{1}{2}\|w'\|_\infty) + O(\varepsilon^2)$.

3. The particle scheme for the gradient process

Based on the previous considerations, we will present next a particle scheme for the one-dimensional diffusion equation. We consider a discretized version of the initial condition u_0 of the equation (2.1), from which we derive the values of $w(0)$ according to (2.2) and the settings at the boundary. In the interior of the domain we simulate the time evolution of the gradient process w according to (2.4). The state changes of the process which affect the values of w at the 'near boundary' sites are chosen such that for given ε, the deterministic difference equation obtained in the limit proves to be consistent with the corresponding diffusion equation for $w = u_x$, where u is a sufficiently smooth solution of (2.1).

For *zero boundary conditions* (D) we construct the dynamics at the 'near-boundary' sites ε and $(M-1)\varepsilon$ according to the following natural conservation principle. Since we expect $w = -u_x$ and thus

$$\int_0^1 w(x)dx = u(0) - u(1) = 0,$$

we impose that $\sum_{i=1}^{M-1} w^i(t) = 0$ for all t. That is, the total sum of the changes after each transition should vanish. In the interior this condition is fulfilled in all situations, as one can see from (2.4).

In order to construct the approximate solution for equation (2.1) we have to perform a numerical integration. An accurate result is delivered for example by the computation of

$$u(t,x) := \frac{1}{2}\left(-\int_0^x w(t,y)dy + \int_x^1 w(t,y)dy + u(t,0) + u(t,1) \right) \tag{3.1}$$

with the trapezoidal rule.

For *zero boundary conditions* (D) we can compute u by knowing only the gradient process w, since the density function vanishes at the boundary. In order to perform the integration, we need the values of w in the boundary sites 0 and $M\varepsilon = 1$. For zero boundary conditions we have (formally): $w_x(t,0) \approx -u_{xx}(t,0) \approx -\frac{d}{dt}u(t,0) = 0$. We take thus $w(0) = w(\varepsilon)$ and $w(1) = w((M-1)\varepsilon)$.

In the case of *Neumann boundary conditions* (N) we do not know the values $u(t,0)$ and $u(t,1)$ (except if we simulate also the density process). But we can recover u by using the conservation property $\int_0^1 u(t,x)dx = \int_0^1 u_0(x)dx$. If we let $f(t,x) = -\int_0^x w(t,y)dy + \int_x^1 w(t,y)dy$, we then have

$$u(t,x) := \frac{1}{2}\left(f(t,x) - \int_0^1 f(t,y)dy \right) + \int_0^1 u_0(x)dx.$$

By a discrete version of the above formula (computed by the trapezoidal rule) we can recover the desired approximation for the solution of (2.1) by knowing only the initial data and the time evolution of the gradient process w.

3.1. Dynamics in terms of the infinitesimal generator

We will express the dynamics of the Markov process w given by the transitions (2.4) in terms of its generator, by using the characterization from [1], p.162 f. If we have an E-valued Markov jump process with a set of transitions $\{x(\cdot) \to y(\cdot)\}$ and the corresponding rates $r_{x\to y}$, the *waiting time parameter function* $\lambda(t) = \sum r_{x\to y}$ is given by the sum of all possible transition rates. The *infinitesimal generator* Λ is an operator acting on the bounded, measurable functions on E and is given by $(\Lambda f)(x) = \sum_{x\to y}(f(y) - f(x))r_{x\to y}$.

We note that for fixed M and N the process w has bounded components, i.e. there exists a constant $L_{M,N}$ such that $max|w^k| \leq L_{M,N}$ for all times. The waiting time parameter function λ is also bounded, which implies that the process is well-defined for all t, i.e. the jumps do not accumulate.

For a vector $w \in \mathbb{R}^{M-1}$ and $3 \leq k \leq M-3$ denote $\eta_k^w := e_{k-\zeta(w^k)} - e_k - e_{k+\zeta(w^k)} + e_{k+2\zeta(w^k)}$. In order to define the values in the sites near the boundary, we take into account the requirements of conservation and consistency.

In the case of *zero boundary condition* (D) we define:

$$\eta_1^w = \begin{cases} -e_1 + e_2 & \text{if } w^1 < 0 \\ -e_2 + e_3 & \text{if } w^1 > 0 \end{cases} \tag{3.2}$$

$$\eta_2^w = \begin{cases} -e_2 + e_3 & \text{if } w^2 < 0 \\ e_1 - e_2 - e_3 + e_4 & \text{if } w^2 > 0. \end{cases}$$

At the other end of the interval we define similarly η_{M-i}^w for $i = 1, 2$ by replacing w^i with $-w^{M-i}$ and e_j with e_{M-j} for $j = 1 \dots 4$.

For *Neumann boundary condition* (N) we consider $\eta_1^w = \eta_{M-1}^w = 0$ and for $k = 2, 3$ (or $k = M-3, M-2$) we suppress the term e_1 (respectively e_{M-1}) if it appears in the formula $\eta_k^w = e_{k-\zeta(w^k)} - e_k - e_{k+\zeta(w^k)} + e_{k+2\zeta(w^k)}$.

Taking in account (2.4), the transitions of the process w are therefore given by

$$w \longrightarrow w + \zeta(w^k)\frac{h}{2\varepsilon}\eta_k^w$$

at rate $h^{-1}\varepsilon^{-1}|w^k(t)|$.

Define the linear operator $\Delta_\varepsilon^{\zeta(w)}\phi$ by:

$$(\Delta_\varepsilon^{\zeta(w)}\phi)^k := \frac{1}{2\varepsilon^2}\langle \eta_k^w, \phi\rangle \tag{3.3}$$

for all $\phi \in \mathbb{R}^{M-1}$, where $\langle\cdot,\cdot\rangle$ denotes the scalar product on $\mathbb{R}^{M-1}$.

For $3 \le k \le M-3$ we thus have

$$(\Delta_\varepsilon^{\zeta(w)}\phi)^k = \frac{1}{2\varepsilon^2}[\phi^{k-\zeta(w^k)} - \phi^k - \phi^{k+\zeta(w^k)} + \phi^{k+2\zeta(w^k)}]. \tag{3.4}$$

For a fixed element $\phi \in \mathbb{R}^{M-1}$ consider on $\mathbb{R}^{M-1}$ a bounded smooth function f_ϕ which on the set $\{x : max|x^k| \le L_{M,N}\}$ has the form $f_\phi(x) = \langle x, \phi\rangle = \sum_{i=1}^{M-1} x^i\phi^i$. Outside this set the values of the function are in our case not of interest, only the boundedness is essential. From [1], p.162 we have that the process w satisfies the identity

$$f_\phi(w(t)) = f_\phi(w(0)) + \int_0^t (\Lambda^w f_\phi)(w(s))ds + M_\phi(t) \tag{3.5}$$

where $M_\phi(\cdot)$ is a martingale with respect to the filtration generated by the process w and Λ^w is the infinitesimal generator. The value $\Lambda^w f_\phi$ is given by

$$\begin{aligned} (\Lambda^w f_\phi)(w(t)) &= \frac{h}{2\varepsilon}\sum_k \langle \zeta(w^k)\eta_k^w, \phi\rangle h^{-1}\varepsilon^{-1}|w^k(t)| \\ &= \frac{1}{2\varepsilon^2}\sum_k [\phi^{k-\zeta(w^k)} - \phi^k - \phi^{k+\zeta(w^k)} + \phi^{k+2\zeta(w^k)}]w^k(t) \\ &= \langle \Delta_\varepsilon^{\zeta(w)}\phi, w(t)\rangle. \end{aligned} \tag{3.6}$$

Equation (3.5) becomes thus

$$\langle w(t), \phi\rangle = \langle w(0), \phi\rangle + \int_0^t \langle \Delta_\varepsilon^{\zeta(w)}\phi, w(s)\rangle ds + M_\phi^w(t). \tag{3.7}$$

3.2. The deterministic scheme as limit of the family of stochastic processes

By standard techniques as in [1] or [2] one can show that for fixed ε the stochastic processes converge in probability for $N \to \infty$ to the solution of the ODE-system obtained by suppressing the martingale term in (3.7). We may note that the stochastic method proposed here delivers for fixed ε an approximation of the solution of the ODE-system by computing all transitions of the stochastic process at the 'microscopic level'. However, this ODE-system is not meant to be approximated by a deterministic time-discretization scheme, but its solution is approximated directly by the stochastic simulations. The convergence for $\varepsilon \to 0$ of the difference scheme provided by the spatially discretized system to the solution of the corresponding spatially continuous diffusion equation will be analyzed subsequently. The system of ODE's which is obtained for $N \to \infty$ and fixed ε is therefore given by:

$$\langle v_\varepsilon(t), \phi\rangle = \langle v_\varepsilon(0), \phi\rangle + \int_0^t \langle \Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi, v_\varepsilon(s)\rangle ds \tag{3.8}$$

for all $\phi \in \mathbb{R}^{M-1}$, where for any $w \in \mathbb{R}^{M-1}$, $\Delta_\varepsilon^{\zeta(w)}\phi$ was defined in (3.3) as

$$(\Delta_\varepsilon^{\zeta(w)}\phi)^k := \frac{1}{2\varepsilon^2}\langle \eta_k^w, \phi\rangle.$$

The vectors η_k^w can be written in the more convenient form:

$$\eta_k^w = -e_{k-2}\zeta(w^k \wedge 0) + e_{k-1}\zeta(w^k) - e_k - e_{k+1}\zeta(w^k) + e_{k+2}\zeta(w^k \vee 0) \tag{3.9}$$

for $3 \le k \le M-3$ and, for zero boundary conditions,

$$\begin{aligned} \eta_2^w &= e_1\zeta(w^2 \vee 0) - e_2 - e_3\zeta(w^2) + e_4\zeta(w^2 \vee 0) \\ \eta_1^w &= e_1\zeta(w^1 \wedge 0) - e_2\zeta(w^1) + e_3\zeta(w^1 \vee 0). \end{aligned}$$

By $x \wedge 0$, $x \vee 0$ we denote respectively the minimum between x and 0 and the maximum between x and 0. The terms $\eta_{M-1}^w, \eta_{M-2}^w$ are defined analogous to η_1^w and η_2^w. With this form we obtain:

$$\begin{aligned} (\Delta_\varepsilon^{\zeta(w)}\phi)^k = \frac{1}{2\varepsilon^2}\big[&-\phi^{k-2}\zeta(w^k \wedge 0) + \phi^{k-1}\zeta(w^k) - \phi^k - \phi^{k+1}\zeta(w^k) \\ &+ \phi^{k+2}\zeta(w^k \vee 0)\big] \end{aligned}$$

for $3 \le k \le M-3$ and

$$(\Delta_\varepsilon^{\zeta(w)}\phi)^2 = \frac{1}{2\varepsilon^2}\left[\phi^1\zeta(w^2 \vee 0) - \phi^2 - \phi^3\zeta(w^2) + \phi^4\zeta(w^2 \vee 0)\right]$$

$$(\Delta_\varepsilon^{\zeta(w)}\phi)^1 = \frac{1}{2\varepsilon^2}\left[\phi^1\zeta(w^1 \wedge 0) - \phi^2\zeta(w^1) + \phi^3\zeta(w^1 \vee 0)\right].$$

An explicit form of the deterministic equations is given by letting $\phi = e_i$ in (3.8), for $i = \overline{1, M-1}$. We obtain then the system

$$\begin{aligned} v_\varepsilon^i(t) &= v_\varepsilon^i(0) + \int_0^t \langle \Delta_\varepsilon^{\zeta(v)} e_i, v_\varepsilon(s)\rangle ds \\ &= v_\varepsilon^i(0) + \int_0^t F_\varepsilon^i(v_\varepsilon(s)) ds \end{aligned} \tag{3.10}$$

for $i = \overline{1, M-1}$, where for $3 \le i \le M-3$ we have the explicit form:

$$F_\varepsilon^i(v_\varepsilon) = \frac{1}{2\varepsilon^2}\left[v_\varepsilon^{i-2} \vee 0 - |v_\varepsilon^{i-1}| - v_\varepsilon^i + |v_\varepsilon^{i+1}| + v_\varepsilon^{i+2} \wedge 0\right].$$

The terms corresponding to the sites near the boundary are computed similarly by using the corresponding values of η_k^w.

4. Convergence results

In this section we analyze the approximation properties of the diffusion equation in the case of *zero boundary conditions* (D) by the ODE system (3.10).

Set $u_\varepsilon^0 = u_\varepsilon^M = 0$ and for $i \in \{1, \dots, M-1\}$ define

$$u_\varepsilon^i(t) = -\varepsilon \sum_{k=1}^{i} v_\varepsilon^k(t). \tag{4.1}$$

Note that this corresponds to a discrete integration scheme for computing $u(t,x) = -\int_0^x v(t,x)dx$. For the sake of computations we will treat the theoretical estimates with this construction, but in practice we will use a scheme for computing the integrals in (3.1), that is we integrate in both directions.

Define the piecewise linear function $u_\varepsilon(t,\cdot) : [0,1] \to \mathbb{R}$ by:

$$u_\varepsilon(t,x) := -v_\varepsilon^{i+1}(t)(x - i\varepsilon) + u_\varepsilon^i(t) \tag{4.2}$$

for $x \in [i\varepsilon, (i+1)\varepsilon]$. We take $u_\varepsilon^0 = u_\varepsilon^M = v_\varepsilon^M = 0$. This is the linear interpolant between the values u_ε^i at the sites $i\varepsilon$. Let us first show some properties of the solutions v_ε of (3.10) and of u_ε defined in (4.1).

Lemma 4.1. *(i) We have for all $T > 0$:*

$$\sup_{t\in[0,T]} \sum_{i=1}^{M-1} (u_\varepsilon^i(t))^2 \le \sum_{i=1}^{M-1} (u_\varepsilon^i(0))^2.$$

(ii) Suppose $v_\varepsilon(t)$ has the following properties for all $t \in [0,T]$: $v_\varepsilon^1(t) \le 0$, $v_\varepsilon^{M-1}(t) \ge 0$ and if $v_\varepsilon^i(t) \cdot v_\varepsilon^{i+2}(t) \ge 0$ then we have also $v_\varepsilon^i(t) \cdot v_\varepsilon^{i+1}(t) \ge 0$. Under these assumptions we have:

$$\sup_{t\in[0,T]} \sum_{i=1}^{M-1} (v_\varepsilon^i(t))^2 \le \sum_{i=1}^{M-1} (v_\varepsilon^i(0))^2.$$

Proof. We recall the form (3.8) of the deterministic equation system:

$$\langle \frac{d}{dt} v_\varepsilon(t), \phi\rangle = \langle \Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi, v_\varepsilon(t)\rangle \tag{4.3}$$

which holds for all vectors $\phi = (\phi^i)_{i=1}^{M-1} \in \mathbb{R}^{M-1}$. In order to derive a similar equation for u_ε, we take as test vectors $\phi = -\varepsilon\phi_{(i)} := -\varepsilon(\phi^i, \phi^i, \dots \phi^i, 0, \dots 0)$ where the first i components are equal to ϕ^i and the rest are 0. We obtain then:

$$\frac{d}{dt} u_\varepsilon^i(t)\phi^i = -\varepsilon\langle \Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi_{(i)}, v_\varepsilon(t)\rangle. \tag{4.4}$$

In particular, if we take $\phi = -\varepsilon\theta_{(i)} := -\varepsilon(1,1,\dots 1,0,\dots 0)$ (the first i components are equal to 1) we obtain:

$$\frac{d}{dt}u_\varepsilon^i(t) = -\varepsilon\langle\Delta_\varepsilon^{\zeta(v_\varepsilon)}\theta_{(i)}, v_\varepsilon(t)\rangle. \tag{4.5}$$

For $3 \le i \le M-3$ we have:

$$\frac{d}{dt}u_\varepsilon^i(t) = -\frac{1}{2\varepsilon}\sum_{k=1}^{M-1}[\theta_{(i)}^{k-\zeta(v_\varepsilon^k)} - \theta_{(i)}^k - \theta_{(i)}^{k+\zeta(v_\varepsilon^k)} + \theta_{(i)}^{k+2\zeta(v_\varepsilon^k)}]v_\varepsilon^k. \tag{4.6}$$

Taking in account the structure of $\theta_{(i)}$ we can note that the terms in the brackets vanish, except in the situation that $i-2 \le k \le i+2$. In this case the values depend on the sign of the corresponding v_ε^k, and we obtain easily:

$$\frac{d}{dt}u_\varepsilon^i(t) = \frac{1}{2\varepsilon}\left[(v_\varepsilon^{i-1}\vee 0) + (v_\varepsilon^i \wedge 0) - (v_\varepsilon^{i+1}\vee 0) - (v_\varepsilon^{i+2}\wedge 0)\right]. \tag{4.7}$$

Let us consider further a general form for ϕ. By summing up the equations (4.4) with respect to i we obtain:

$$\langle\frac{d}{dt}u_\varepsilon(t),\phi\rangle = -\varepsilon\sum_{i=1}^{M-1}\langle\Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi_{(i)}, v_\varepsilon(t)\rangle. \tag{4.8}$$

Let us compute the r.h.s. by rearranging the terms in a convenient form. We have:

$$\sum_{i=1}^{M-1}\langle\Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi_{(i)}, v_\varepsilon(t)\rangle = \sum_{i=1}^{M-1}\sum_{k=1}^{M-1}(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi_{(i)})^k\cdot v_\varepsilon^k = \sum_{k=1}^{M-1}v_\varepsilon^k\cdot\sum_{i=1}^{M-1}(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi_{(i)})^k.$$

Taking in account formula (3.10) we thus have for $3 \le k \le M-3$:

$$(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi_{(i)})^k =$$
$$= \frac{1}{2\varepsilon^2}\left[-\phi_{(i)}^{k-2}\zeta(v_\varepsilon^k\wedge 0) + \phi_{(i)}^{k-1}\zeta(v_\varepsilon^k) - \phi_{(i)}^k - \phi_{(i)}^{k+1}\zeta(v_\varepsilon^k) + \phi_{(i)}^{k+2}\zeta(v_\varepsilon^k\vee 0)\right].$$

Since $\phi_{(i)} = (\phi^i,\phi^i,\dots\phi^i,0,\dots 0)$, it follows imediately that the expression vanishes for $i < k-2$ and $i \ge k+2$. By analyzing all possibilities with respect to the sign of v_ε^k we obtain for $3 \le k \le M-3$ the expression:

$$\sum_{i=1}^{M-1}(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi_{(i)})^k = \frac{1}{2\varepsilon^2}\left[(\phi^{k-1}-\phi^{k+1})\zeta(v_\varepsilon^k\vee 0) + (\phi^k - \phi^{k-2})\zeta(v_\varepsilon^k\wedge 0)\right]. \tag{4.9}$$

In the case of the the terms corresponding to the sites near the boundary we obtain similarly

$$\sum_{i=1}^{M-1}(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi_{(i)})^1 = \frac{1}{2\varepsilon^2}[-\phi^2\zeta(v_\varepsilon^1\vee 0) + \phi^1\zeta(v_\varepsilon^1\wedge 0)]$$

and

$$\sum_{i=1}^{M-1}(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\phi_{(i)})^2 = \frac{1}{2\varepsilon^2}[(\phi^1-\phi^3)\zeta(v_\varepsilon^2\vee 0) + \phi^2\zeta(v_\varepsilon^2\wedge 0)].$$

Similar equations are derived at the other end of the interval. We note that these equations can be also reduced to the form (4.9) by setting $\phi^j = 0$ if $j \notin \{1, \ldots, M-1\}$.

Since all computations are nothing more than rearrangements of the terms, we can obtain the same results by multiplying the explicit equations with time dependent test functions $\phi(t)$. We can take now $\phi = u_\varepsilon(t)$ and use the corresponding $\phi_{(i)}$. Note that $u_\varepsilon^{j-1} - u_\varepsilon^{j+1} = \varepsilon(v_\varepsilon^{j+1} + v_\varepsilon^j)$ and that we can set $u_\varepsilon^j = v_\varepsilon^j = 0$ if $j \notin \{1, \ldots, M-1\}$, due to the considered boundary conditions. We then have:

$$\sum_{k=1}^{M-1} v_\varepsilon^k \sum_{i=1}^{M-1} (\Delta_\varepsilon^{\zeta(v_\varepsilon)} \phi_{(i)})^k =$$

$$= \frac{1}{2\varepsilon} \sum_{k=1}^{M-1} v_\varepsilon^k [(v_\varepsilon^k + v_\varepsilon^{k+1})\zeta(v_\varepsilon^k \vee 0) - (v_\varepsilon^k + v_\varepsilon^{k-1})\zeta(v_\varepsilon^k \wedge 0)]$$

$$= \frac{1}{2\varepsilon} \sum_{k=1}^{M-1} [(v_\varepsilon^k)^2 + v_\varepsilon^k v_\varepsilon^{k+1}\zeta(v_\varepsilon^k \vee 0) - v_\varepsilon^k v_\varepsilon^{k-1}\zeta(v_\varepsilon^k \wedge 0)] =: \frac{1}{2\varepsilon} \sum_{k=1}^{M-1} a_k.$$

The terms a_k have the form:

$$a_k = \begin{cases} (v_\varepsilon^k)^2 + v_\varepsilon^k v_\varepsilon^{k+1} & \text{if } v_\varepsilon^k > 0 \\ (v_\varepsilon^k)^2 + v_\varepsilon^k v_\varepsilon^{k-1} & \text{if } v_\varepsilon^k \le 0. \end{cases}$$

We claim that $\sum_k a_k \ge 0$. In order to show this, we proceed inductively. Denote $S_m = \sum_{k=1}^m a_k$. We have $S_1 = a_1 = (v_\varepsilon^1)^2 + v_\varepsilon^1 v_\varepsilon^2 \zeta(v_\varepsilon^1 \vee 0)$. If $v_\varepsilon^1 \le 0$ then $S_1 = (v_\varepsilon^1)^2 \ge 0$. If $v_\varepsilon^1 \ge 0$ and $v_\varepsilon^2 \le 0$ then $S_2 = (v_\varepsilon^1)^2 + 2v_\varepsilon^1 v_\varepsilon^2 + (v_\varepsilon^2)^2 \ge 0$. If $v_\varepsilon^1 > 0, v_\varepsilon^2 > 0, \ldots, v_\varepsilon^{p-1} > 0, v_\varepsilon^p \le 0$, then we have $a_1 + a_2 + \cdots + a_p = (v_\varepsilon^1)^2 + v_\varepsilon^1 v_\varepsilon^2 + \cdots + (v_\varepsilon^{p-1})^2 + 2v_\varepsilon^p v_\varepsilon^{p-1} + (v_\varepsilon^p)^2 \ge 0$. We have thus ($S_p \ge 0$ and $v_\varepsilon^p \le 0$).

The first step leads thus to a situation on the type ($S_p \ge 0$ and $v_\varepsilon^p \le 0$).

If $p = M-1$ we are done. If not, we repeat the procedure. Suppose that we have shown that ($S_{k-1} \ge 0$ and $v_\varepsilon^{k-1} \le 0$). If $v_\varepsilon^k \le 0$, then we have $a_k = (v_\varepsilon^k)^2 + v_\varepsilon^k v_\varepsilon^{k-1} \ge 0$ and thus ($S_k \ge 0$ and $v_\varepsilon^k \le 0$).

If $v_\varepsilon^k > 0$ and $v_\varepsilon^{k+1} \le 0$, then we have $a_k + a_{k+1} = (v_\varepsilon^k)^2 + 2v_\varepsilon^k v_\varepsilon^{k+1} + (v_\varepsilon^{k+1})^2 \ge 0$. This implies ($S_{k+1} \ge 0$ and $v_\varepsilon^{k+1} \le 0$).

If $v_\varepsilon^k > 0$ and $v_\varepsilon^{k+1} > 0, \ldots, v_\varepsilon^{p-1} > 0, v_\varepsilon^p \le 0$, then we have $a_k + a_{k+1} + \cdots + a_p = (v_\varepsilon^k)^2 + v_\varepsilon^k v_\varepsilon^{k+1} + \cdots + (v_\varepsilon^{p-1})^2 + 2v_\varepsilon^p v_\varepsilon^{p-1} + (v_\varepsilon^p)^2 \ge 0$. We have thus ($S_p \ge 0$ and $v_\varepsilon^p \le 0$).

If starting with some index j we have $v_\varepsilon^k \ge 0$ for $k \ge j$, then we are done, since we add only positive terms a_k for $k \ge j$. In the case of the last term we have then only $(v_\varepsilon^{M-1})^2$. The other alternative is to obtain the situation ($S_{M-1} \ge 0$ and $v_\varepsilon^{M-1} \le 0$), when we are also done.

The fact that for $\phi = u_\varepsilon$ we have $S_{M-1} \ge 0$, together with equation (4.8) imply $\frac{d}{dt}\langle u_\varepsilon(t), u_\varepsilon(t)\rangle = 2\langle \frac{d}{dt} u_\varepsilon(t), u_\varepsilon(t)\rangle = -S_{M-1} \le 0$ which proves the first part of the lemma.

For the second part we take $\phi = v_\varepsilon(t)$ in (4.3) and we have to show that $\langle \Delta_\varepsilon^{\zeta(v_\varepsilon)} v_\varepsilon(t), v_\varepsilon(t)\rangle \le 0$. This can be written also as $\sum_{k=1}^{M-1} T_k \le 0$, where

$$T_k = [-v_\varepsilon^{k-2}\zeta(v_\varepsilon^k \wedge 0) + v_\varepsilon^{k-1}\zeta(v_\varepsilon^k) - v_\varepsilon^k - v_\varepsilon^{k+1}\zeta(v_\varepsilon^k) + v^{k+2}\zeta(v_\varepsilon^k \vee 0)] \cdot v_\varepsilon^k$$

if $3 \le k \le M-3$, and

$$\begin{aligned} T_1 &= [v_\varepsilon^1\zeta(v_\varepsilon^1 \wedge 0) - v_\varepsilon^2\zeta(v_\varepsilon^1) + v_\varepsilon^3\zeta(v_\varepsilon^1 \vee 0)] \cdot v_\varepsilon^1 \\ T_2 &= [v_\varepsilon^1\zeta(v_\varepsilon^2 \vee 0) - v_\varepsilon^2 - v_\varepsilon^3\zeta(v_\varepsilon^2) + v^4\zeta(v_\varepsilon^2 \vee 0)] \cdot v_\varepsilon^2, \end{aligned}$$

while T_{M-1}, T_{M-2} are computed analogous to T_1, T_2. We will structure the proof in an algorithmic fashion.

From the hypothesis we have that $v_\varepsilon^1 \le 0$. Define $T := 0$.

0. If we have $v_\varepsilon^2 \le 0$ let $T := T_1 + T_2$. We have thus

$$T = -(v_\varepsilon^1)^2 + v_\varepsilon^1 v_\varepsilon^2 - (v_\varepsilon^2)^2 + v_\varepsilon^2 v_\varepsilon^3 \le -\frac{1}{2}(v_\varepsilon^2)^2 + v_\varepsilon^2 v_\varepsilon^3.$$

Let $q = 1$ and GOTO **2.**

1. Else, if we have $v_\varepsilon^2 > 0$, then the hypothesis implies that we have also $v_\varepsilon^3 \ge 0$. Let $T := T_1 + T_2$. We have thus

$$T = -(v_\varepsilon^1)^2 + 2v_\varepsilon^1 v_\varepsilon^2 - (v_\varepsilon^2)^2 - v_\varepsilon^2 v_\varepsilon^3 + v_\varepsilon^2 v_\varepsilon^4 \le -(v_\varepsilon^2)^2 + v_\varepsilon^2 v_\varepsilon^4.$$

Let $p = 2$ and GOTO **3.**

2. Suppose we have $v_\varepsilon^q \le 0, v_\varepsilon^{q+1} \le 0, \dots v_\varepsilon^{p-1} \le 0, v_\varepsilon^p > 0$ and

$$T \le -\frac{1}{2}(v_\varepsilon^{q+1})^2 + v_\varepsilon^{q+1} v_\varepsilon^{q+2}.$$

The hypothesis on v_ε implies that we must have also $v_\varepsilon^{p+1} \ge 0$. We observe that for $q+2 \le k \le p-1$ in the sum $T_{k-1} + T_k$ appear the cancelling terms $-v_\varepsilon^k v_\varepsilon^{k-1}\zeta(v_\varepsilon^{k-1}) + v_\varepsilon^k v_\varepsilon^{k-1}\zeta(v_\varepsilon^k)$, since v_ε^{k-1} and v_ε^k have the same sign. If $p = q+2$ we do not have such terms. We thus have:

$$\begin{aligned} \sum_{k=q+2}^{p} T_k &= -\sum_{k=q+2}^{p} (v_\varepsilon^k)^2 + \sum_{k=q+2}^{p-1} v_\varepsilon^k v_\varepsilon^{k-2} + 2v_\varepsilon^{p-1} v_\varepsilon^p - v_\varepsilon^p v_\varepsilon^{p+1} + v_\varepsilon^p v_\varepsilon^{p+2} \\ &\le -\frac{1}{2}(v_\varepsilon^{q+2})^2 \cdot \chi_{\{p>q+2\}} - (v_\varepsilon^p)^2 + v_\varepsilon^p v_\varepsilon^{p+2}. \end{aligned}$$

We grouped the terms in order to obtain nonpositive quantities like $[(-v_\varepsilon^k)^2 + 2v_\varepsilon^k v_\varepsilon^{k-2} - (v_\varepsilon^{k-2})^2]/2$, together with $2v_\varepsilon^{p-1}v_\varepsilon^p$ and $-v_\varepsilon^p v_\varepsilon^{p+1}$ which are also ≤ 0.

Let $T := T + \sum_{k=q+2}^{p} T_k$. If $p = q+2$ we obtain:

$$T \le -\frac{1}{2}(v_\varepsilon^{q+1})^2 + v_\varepsilon^{q+1} v_\varepsilon^{q+2} - (v_\varepsilon^{q+2})^2 + v_\varepsilon^{q+2} v_\varepsilon^{q+4} \le -(v_\varepsilon^{q+2})^2 + v_\varepsilon^{q+2} v_\varepsilon^{q+4}$$

since $v_\varepsilon^{q+1} v_\varepsilon^{q+2} \le 0$. If $p > q+2$ we have:

$$T \le -\frac{1}{2}(v_\varepsilon^{q+1})^2 + v_\varepsilon^{q+1} v_\varepsilon^{q+2} - \frac{1}{2}(v_\varepsilon^{q+2})^2 - (v_\varepsilon^p)^2 + v_\varepsilon^p v_\varepsilon^{p+2} \le -(v_\varepsilon^p)^2 + v_\varepsilon^p v_\varepsilon^{p+2}.$$

(stopping condition) If $p = M-2$ we are done, since the term $v_\varepsilon^p v_\varepsilon^{p+2}$ does not appear, while T_{M-1} equals $-(v_\varepsilon^{M-1})^2 - v_\varepsilon^{M-1} v_\varepsilon^{M-2}$, which can be grouped together with $-(v_\varepsilon^{M-2})^2$ in order to obtain a nonpositive quantity. If $p = M-1$ we are also done.

3. Suppose we have $v_\varepsilon^p \geq 0, v_\varepsilon^{p+1} \geq 0, \dots v_\varepsilon^{q-1} \geq 0, v_\varepsilon^q \leq 0$ and

$$T \leq -(v_\varepsilon^p)^2 + v_\varepsilon^p v_\varepsilon^{p+2}.$$

The hypothesis implies that we have $v_\varepsilon^{q+1} \leq 0$. Similarly as in **2.** (dropping nonpositive terms which are not needed and using the cancelling property) we compute:

$$\sum_{k=p+1}^{q+1} T_k =$$

$$= - \sum_{k=p+1}^{q+1} (v_\varepsilon^k)^2 + \sum_{k=p+1}^{q-1} v_\varepsilon^k v_\varepsilon^{k+2} + v_\varepsilon^q v_\varepsilon^{q-2} + 2 v_\varepsilon^q v_\varepsilon^{q-1} + v_\varepsilon^{q+1} v_\varepsilon^{q-1} + v_\varepsilon^{q+1} v_\varepsilon^{q+2}$$

$$\leq -\frac{1}{2}(v_\varepsilon^{p+2})^2 - \frac{1}{2}(v_\varepsilon^q)^2 - \frac{1}{2}(v_\varepsilon^{q+1})^2 + v_\varepsilon^q v_\varepsilon^{q-2} + v_\varepsilon^{q+1} v_\varepsilon^{q+2}.$$

Let $T := T + \sum_{k=p+1}^{q+1} T_k$. If $q = p+2$ we obtain:

$$T \leq -(v_\varepsilon^p)^2 + v_\varepsilon^p v_\varepsilon^{p+2} - (v_\varepsilon^{p+2})^2 - \frac{1}{2}(v_\varepsilon^{p+3})^2 + v_\varepsilon^p v_\varepsilon^{p+2} + v_\varepsilon^{p+3} v_\varepsilon^{p+4}$$

$$\leq -\frac{1}{2}(v_\varepsilon^{p+3})^2 + v_\varepsilon^{p+3} v_\varepsilon^{p+4},$$

since $v_\varepsilon^p v_\varepsilon^{p+2} \leq 0$.

If $q > p+2$ we have:

$$T \leq -(v_\varepsilon^p)^2 + v_\varepsilon^p v_\varepsilon^{p+2} - \frac{1}{2}(v_\varepsilon^{p+2})^2 - \frac{1}{2}(v_\varepsilon^q)^2 - \frac{1}{2}(v_\varepsilon^{q+1})^2 + v_\varepsilon^q v_\varepsilon^{q-2} + v_\varepsilon^{q+1} v_\varepsilon^{q+2}$$

$$\leq -\frac{1}{2}(v_\varepsilon^{q+1})^2 + v_\varepsilon^{q+1} v_\varepsilon^{q+2},$$

since $v_\varepsilon^q v_\varepsilon^{q-2} \leq 0$. If $q+1 = M-1$ and $v_\varepsilon^{M-1} = 0$ we don't have the term $v_\varepsilon^{q+1} v_\varepsilon^{q+2}$ and we are done. Otherwise, since $v_\varepsilon^{M-1} \geq 0$, we cannot end in the situation $q = M-2$. Thus, GOTO **2.**

The above algorithm clearly stops in step **2.** arriving in the final situation with $T = \sum_{k=1}^{M-1} T_k \leq 0$. The proof is thus completed. $\square$

Remark. The assumptions on v_ε in (ii) hold true, at least on a given time interval, if $v_\varepsilon(0)$ is constructed by taking the finite differences of a positive, piecewise Lipschitz continuous function u_0 and ε is chosen small enough. As it will be shown further, for the convergence of the method we will need bounds for $\sum \varepsilon (v_\varepsilon^k(t))^2$ independent on ε. This condition is not fulfilled if we choose an arbitrary initial data $v_\varepsilon(0)$. Numerical computations show that the sum will blow up in a short time if we take e.g. $v_\varepsilon^k(0) = (-1)^k$ if $k \in \{1, \dots, 2M\} \setminus \{M\}$ and $v_\varepsilon^M(0) = 2(-1)^M$. In general we cannot expect for v_ε a similar inequality as for u_ε in (i), but in most practically relevant situations this property holds true, as shown for example in (ii).

Lemma 4.2. *(Energy estimates). Assume that for all ε we have*

$$\|u_\varepsilon(0,\cdot)\|_{H_0^1(0,1)} \le C_0\|u_0\|_{H_0^1(0,1)}$$

for a given function $u_0 \in H_0^1(0,1)$, with a positive constant C_0. Suppose further that the functions $v_\varepsilon(t)$ satisfy the inequality $\sup_{t\le T}\sum_k(v_\varepsilon^k(t))^2 \le C_1\sum_k(v_\varepsilon^k(0))^2$ with a positive constant C_1, independent on ε (in particular, if the assumption from Lemma 4.1 (ii) holds). Then there exists a constant $C>0$, independent on ε, such that:

$$\sup_{t\in[0,T]}\big[\|u_\varepsilon(t,\cdot)\|_{H_0^1(0,1)} + \|\frac{d}{dt}u_\varepsilon(t,\cdot)\|_{H^{-1}(0,1)}\big] \le C\|u_0\|_{H_0^1(0,1)}.$$

Proof. We have:

$$\begin{aligned}
&\sup_{t\in[0,T]}\big[\int_0^1 u_\varepsilon^2(t,x)dx\big]\\
&=\sup_{t\in[0,T]}\big[\sum_{i=0}^{M-1}\int_{i\varepsilon}^{(i+1)\varepsilon}\big(-v_\varepsilon^{i+1}(t)(x-i\varepsilon)+u_\varepsilon^i(t)\big)^2dx\big]\\
&\le 2\sup_{t\in[0,T]}\big[\sum_{i=0}^{M-1}\int_{i\varepsilon}^{(i+1)\varepsilon}\big((v_\varepsilon^{i+1}(t))^2(x-i\varepsilon)^2+(u_\varepsilon^i(t))^2\big)dx\big]\\
&=2\sup_{t\in[0,T]}\big[\sum_{i=0}^{M-1}\big(\frac{(v_\varepsilon^{i+1}(t))^2}{3}(x-i\varepsilon)^3\Big|_{i\varepsilon}^{(i+1)\varepsilon}+\varepsilon(u_\varepsilon^i(t))^2\big)\big]\\
&=2\sup_{t\in[0,T]}\big[\sum_{i=0}^{M-1}\big(\frac{\varepsilon^3}{3}(v_\varepsilon^{i+1}(t))^2+\varepsilon(u_\varepsilon^i(t))^2\big)\big]\\
&\le 2C_1\sum_{i=0}^{M-1}\big(\frac{\varepsilon^3}{3}(v_\varepsilon^{i+1}(0))^2+\varepsilon(u_\varepsilon^i(0))^2\big)\\
&=2C_1\frac{\varepsilon^2}{3}\|(u_\varepsilon(0,\cdot))_x\|^2_{L^2(0,1)}+2C_1\|u_\varepsilon(0,\cdot)\|^2_{L^2(0,1)}.
\end{aligned}$$

We made use of Lemma 4.1 and on the hypothesis on v_ε. Using now the estimate for $u_\varepsilon(0,\cdot)$ from the hypothesis we obtain: $\sup_{t\in[0,T]}\|u_\varepsilon(t)\|^2_{L^2(0,1)} \le C'\|u_0\|^2_{H^1}$. Note that the equation for v_ε implies that we always have

$$\sup_{t\in[0,T]}\sum_{i=0}^{M-1}\varepsilon^3(v_\varepsilon^{i+1}(t))^2 \le C(T)\sum_{i=0}^{M-1}\varepsilon(v_\varepsilon^{i+1}(0))^2.$$

We can show thus that $\sup_{t\in[0,T]}\|u_\varepsilon(t)\|_{L^2(0,1)} \le C'\|u_0\|_{H^1}$ by using only the H^1-bounds for $u_\varepsilon(0)$, independent on any additional assumptions on $v_\varepsilon(t)$.

We further have:

$$\sup_{t\in[0,T]}\int_0^1 (u_\varepsilon(t,x)_x)^2dx = \sup_{t\in[0,T]}\sum_{i=0}^{M-1}\varepsilon(v_\varepsilon^{i+1}(t))^2$$
$$\le C_1\sum_{i=0}^{M-1}\varepsilon(v_\varepsilon^{i+1}(0))^2 = C_1\|(u_\varepsilon(0,\cdot))_x\|^2_{L^2(0,1)}\le C_1C_0\|u_0\|^2_{H^1}.$$

In the previously used notation, where $\langle\cdot,\cdot\rangle$ denotes the usual scalar product on $\mathbb{R}^{M-1}$, we have for $\psi\in C_0^\infty(0,1)$ with $\|\psi\|_{H_0^1}=1$:

$$\sup_{t\in[0,T]}\int_0^1 \frac{d}{dt}u_\varepsilon(t,x)\psi(x)dx$$
$$=\sup_{t\in[0,T]}\sum_{i=0}^{M-1}\int_{i\varepsilon}^{(i+1)\varepsilon}\big(-\frac{d}{dt}v_\varepsilon^{i+1}(t)(x-i\varepsilon)+\frac{d}{dt}u_\varepsilon^i(t)\big)\psi(x)dx$$
$$=\sup_{t\in[0,T]}[-\langle\frac{d}{dt}v_\varepsilon(t),\tilde\Psi_1\rangle+\langle\frac{d}{dt}u_\varepsilon(t),\tilde\Psi_2\rangle]$$

where $\tilde\Psi_1,\tilde\Psi_2\in\mathbb{R}^{M-1}$ are given by

$$(\tilde\Psi_1)^k=\int_{(k-1)\varepsilon}^{k\varepsilon}(x-(k-1)\varepsilon)\psi(x)dx$$

respectively

$$(\tilde\Psi_2)^k=\int_{k\varepsilon}^{(k+1)\varepsilon}\psi(x)dx.$$

By (4.3) we have

$$\langle\frac{d}{dt}v_\varepsilon(t),\tilde\Psi_1\rangle=\langle\Delta_\varepsilon^{\zeta(v_\varepsilon)}\tilde\Psi_1,v_\varepsilon(t)\rangle$$

where $\Delta_\varepsilon^{\zeta(v_\varepsilon)}\tilde\Psi_1$ is computed like in (3.10) and has the form

$$(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\tilde\Psi_1)^k=\frac{1}{2\varepsilon^2}\sum_{j\in I_k}\pm(\tilde\Psi_1^{j+1}-\tilde\Psi_1^j), \tag{4.10}$$

since we always can group the terms which arise in the r.h.s. of (3.10) in pairs with opposite signs. The index set I_k has two elements, except for sites $k\varepsilon$ near the boundary, when we have only one element.

By partial integration we have:

$$\tilde\Psi_1^j=\int_{(j-1)\varepsilon}^{j\varepsilon}(x-(j-1)\varepsilon)\psi(x)dx=\frac{\varepsilon^2}{2}\psi(j\varepsilon)-\frac{1}{2}\int_{(j-1)\varepsilon}^{j\varepsilon}(x-(j-1)\varepsilon)^2\psi'(x)dx.$$

A similar formula holds also for $\tilde\Psi_1^{j+1}$. By substraction we obtain easily the estimate

$$\frac{1}{2\varepsilon^2}|\tilde\Psi_1^{j+1}-\tilde\Psi_1^j|\le\frac{1}{4}(2\int_{j\varepsilon}^{(j+1)\varepsilon}|\psi'(x)|dx+\int_{(j-1)\varepsilon}^{j\varepsilon}|\psi'(x)|dx)$$

$$\leq \frac{1}{2}\int_{(j-1)\varepsilon}^{(j+1)\varepsilon} |\psi'(x)|dx,$$

which in conjunction with the Cauchy-Schwartz inequality implies:

$$T_j := \frac{1}{4\varepsilon^4}|\tilde{\Psi}_1^{j+1} - \tilde{\Psi}_1^j|^2 \leq \frac{\varepsilon}{2}\int_{(j-1)\varepsilon}^{(j+1)\varepsilon} |\psi'(x)|^2 dx. \tag{4.11}$$

Returning to (4.10), where the index j has at most two values, we thus have the estimate: $[(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\tilde{\Psi}_1)^k]^2 \leq \varepsilon\sum_{j\in I_k} T_j$ with T_j given in (4.11). It can be readily seen that there exists a constant $C' > 0$ such that

$$\sum_k \sum_{j\in I_k} T_j \leq C'\|\psi'\|^2_{L^2(0,1)} \leq C',$$

since every index j appears in the above sums maximally a given number of times. We thus have:

$$|\langle \frac{d}{dt}v_\varepsilon(t), \tilde{\Psi}_1\rangle|^2 = |\langle \Delta_\varepsilon^{\zeta(v_\varepsilon)}\tilde{\Psi}_1, v_\varepsilon(t)\rangle|^2 \leq \sum_k (v_\varepsilon^k(t))^2 \sum_k [(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\tilde{\Psi}_1)^k]^2$$

$$\leq C'\sum_k \varepsilon(v_\varepsilon^k(t))^2 \leq C'C_1\sum_k \varepsilon(v_\varepsilon^k(0))^2 \leq C'C_1C_0\|u_0\|^2_{H^1} \tag{4.12}$$

where the constants C', C_1, C_0 do not depend on ψ and on ε.

By (4.8) and the subsequent computations we have:

$$\langle \frac{d}{dt}u_\varepsilon(t), \tilde{\Psi}_2\rangle = -\varepsilon\sum_{i=1}^{M-1}\langle \Delta_\varepsilon^{\zeta(v_\varepsilon)}(\tilde{\Psi}_2)_{(i)}, v_\varepsilon(t)\rangle = -\varepsilon\sum_{k=1}^{M-1} v_\varepsilon^k(t)T'_k$$

where from (4.9), we have that

$$T'_k := \sum_{i=1}^{M-1}(\Delta_\varepsilon^{\zeta(v_\varepsilon)}(\tilde{\Psi}_2)_{(i)})^k = \pm\frac{1}{2\varepsilon^2}(\tilde{\Psi}_2^{j_k} - \tilde{\Psi}_2^{j_k-2})$$

with $j_k = k$ or $j_k = k+1$, depending on the sign on $v_\varepsilon^k(t)$.

By partial integration we have:

$$\tilde{\Psi}_2^{j_k} = \int_{j_k\varepsilon}^{(j_k+1)\varepsilon}\psi(x)dx = \varepsilon\psi((j_k+1)\varepsilon) - \int_{j_k\varepsilon}^{(j_k+1)\varepsilon}(x - j_k\varepsilon)\psi'(x)dx$$

and for $\tilde{\Psi}_2^{j_k-2}$ we obtain a similar equation. Proceeding similarly as in the case of $\tilde{\Psi}_1$ we arrive at

$$\varepsilon(T'_k)^2 \leq 3\int_{(j_k-2)\varepsilon}^{(j_k+1)\varepsilon}|\psi'(x)|^2 dx$$

with $\sum_k \varepsilon(T'_k)^2 \leq C''\|\psi'\|^2_{L^2(0,1)} \leq C''$.

Similarly as in the previous computations we have:

$$|\langle \frac{d}{dt}u_\varepsilon(t), \tilde{\Psi}_2\rangle|^2 \leq$$

$$\leq (\sum_{k=1}^{M-1} |\varepsilon^{\frac{1}{2}} v_\varepsilon^k(t)||\varepsilon^{\frac{1}{2}} T_k'|)^2 \leq \sum_{k=1}^{M-1} \varepsilon |v_\varepsilon^k(t)|^2 \cdot \sum_{k=1}^{M-1} \varepsilon (T_k')^2 \leq C_1 C'' C_0 \|u_0\|_{H^1}^2.$$

This completes the proof of the lemma. □

Theorem 4.3. *If on the time interval $[0, T]$ the hypotheses of Lemma 4.2 hold, then the family $u_\varepsilon(\cdot, \cdot)$ has a weakly convergent subsequence in $L^2(0, T; H_0^1(0, 1)) \cap H^1(0, T; H^{-1}(0, 1))$ and the limit function, denoted by u, lies in $C(0, T; L^2(0, 1))$. Moreover, if we have $\|u_\varepsilon(0) - u_0\|_{L^2(0,1)} \to 0$ as $\varepsilon \to 0$, then u satisfies for all test functions $\phi \in C([0, T] \times [0, 1]) \cap C^\infty((0, T) \times (0, 1))$ with $supp\phi \subset [0, T) \times (0, 1)$ the equation:*

$$-\int_0^T (u(t), \frac{d}{dt}\phi(t))dt = (u_0, \phi(0)) - \int_0^T (u_x(t), \phi_x(t))dt.$$

By $(\cdot, \cdot)$ we denote the usual duality pairing in the corresponding function spaces.

Proof. The weak convergence property is implied by the apriori estimates from Lemma 4.2. By a result from [6], p.379 we have further $u \in C(0, T; L^2(0, 1))$. We denote the convergent subsequence again by u_ε. For a test function ϕ like in the hypothesis we have:

$$\begin{aligned}
&-\int_0^T (u(t), \frac{d}{dt}\phi(t))dt - (u_0, \phi(0)) + \int_0^T (u_x(t), \phi_x(t))dt = \\
&= -\int_0^T (u(t) - u_\varepsilon(t), \frac{d}{dt}\phi(t))dt - (u_0 - u_\varepsilon(0), \phi(0)) + \\
&+ \int_0^T (u_x(t) - (u_\varepsilon)_x(t), \phi_x(t))dt - \int_0^T (u_\varepsilon(t), \frac{d}{dt}\phi(t))dt \\
&-(u_\varepsilon(0), \phi(0)) + \int_0^T ((u_\varepsilon)_x(t), \phi_x(t))dt.
\end{aligned}$$

The first three terms converge to 0 as $\varepsilon \to 0$ due to the weak convergence property. In order to prove the statement of the theorem, we will show that the rest of the sum can be made arbitrarily small for ε small enough. For this it suffices to show that the term

$$\sup_{t \leq T} [(\frac{d}{dt}u_\varepsilon(t), \phi(t)) + ((u_\varepsilon)_x(t), \phi_x(t))] \tag{4.13}$$

can be proved to be arbitrarily small by taking ε small enough.

Similarly like in Lemma 4.2 we obtain:

$$\begin{aligned}
&(\frac{d}{dt}u_\varepsilon(t), \phi(t)) + ((u_\varepsilon)_x(t), \phi_x(t)) = \qquad (4.14) \\
&= -\langle \frac{d}{dt}v_\varepsilon(t), \tilde{\Phi}_1(t)\rangle + \langle \frac{d}{dt}u_\varepsilon(t), \tilde{\Phi}_2(t)\rangle + \langle v_\varepsilon(t), \tilde{\Phi}_3(t)\rangle \\
&= \langle v_\varepsilon(t), \Delta_\varepsilon^{\zeta(v_\varepsilon)} \tilde{\Phi}_1(t)\rangle + \langle v_\varepsilon(t), \nabla_\varepsilon^{\zeta(v_\varepsilon)} \tilde{\Phi}_2(t)\rangle + \langle v_\varepsilon(t), \tilde{\Phi}_3(t)\rangle
\end{aligned}$$

where the vectors $\tilde{\Phi}_i(t) \in \mathbb{R}^{M-1}$ have the components

$$\tilde{\Phi}_1^k(t) = \int_{(k-1)\varepsilon}^{k\varepsilon} (y-(k-1)\varepsilon)\phi(t,y)dy,$$

$$\tilde{\Phi}_2^k(t) = \int_{k\varepsilon}^{(k+1)\varepsilon} \phi(t,y)dy,$$

$$\tilde{\Phi}_3^k(t) = \int_{(k-1)\varepsilon}^{k\varepsilon} \phi_x(t,y)dy$$

and where

$$\begin{aligned}
(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\tilde{\Phi}_1(t))^k &= \frac{1}{2\varepsilon^2}[-\tilde{\Phi}_1^{k-2}\zeta(v_\varepsilon^k \wedge 0) + \tilde{\Phi}_1^{k-1}\zeta(v_\varepsilon^k) - \tilde{\Phi}_1^k \\
&\qquad -\tilde{\Phi}_1^{k+1}\zeta(v_\varepsilon^k) + \tilde{\Phi}_1^{k+2}\zeta(v_\varepsilon^k \vee 0)] \\
(\nabla_\varepsilon^{\zeta(v_\varepsilon)}\tilde{\Phi}_2(t))^k &= \frac{1}{2\varepsilon}[(\tilde{\Phi}_2^{k-1} - \tilde{\Phi}_2^{k+1})\zeta(v_\varepsilon^k \vee 0) + (\tilde{\Phi}_2^k - \tilde{\Phi}_2^{k-2})\zeta(v_\varepsilon^k \wedge 0)].
\end{aligned}$$

By taking ε small enough, since ϕ has compact support in $[0,T) \times (0,1)$, we may consider only indices k for which the above formulas hold for all t, disregarding the sites near the boundary where ϕ vanishes. By the same reason (neglecting the sites close to the boundary), we may note that for $i = 1,2,3$ we can write $(\tilde{\Phi}_i(t))^k = \Phi_i(t,k\varepsilon)$ where the functions Φ_i are defined on $[0,T] \times (0,1)$ by

$$\Phi_1(t,x) = \int_{x-\varepsilon}^{x} (y-x+\varepsilon)\phi(t,y)dy,$$

$$\Phi_2(t,x) = \int_{x}^{x+\varepsilon} \phi(t,y)dy,$$

$$\Phi_3(t,x) = \int_{x-\varepsilon}^{x} \phi_x(t,y)dy = \phi(t,x) - \phi(t,x-\varepsilon).$$

The derivatives with respect to x of these functions are given by

$$\Phi_{1,x}(t,x) = \varepsilon\phi(t,x) - \int_{x-\varepsilon}^{x} \phi(t,y)dy,$$

$$\Phi_{1,xx}(t,x) = \varepsilon\phi_x(t,x) - \phi(t,x) + \phi(t,x-\varepsilon),$$

$$\Phi_{2,x}(t,x) = \phi(t,x+\varepsilon) - \phi(t,x).$$

By the Taylor formula, using the form of Φ_i and the bounds of the derivatives of second and third order of $\phi(t,\cdot)$ it is easy to see that if $v_\varepsilon^k > 0$ we have:

$$\begin{aligned}
(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\tilde{\Phi}_1(t))^k &= \Phi_{1,xx}(t,k\varepsilon) + O(\varepsilon^2) \\
(\nabla_\varepsilon^{\zeta(v_\varepsilon)}\tilde{\Phi}_2(t))^k &= -\Phi_{2,x}(t,k\varepsilon) + O(\varepsilon^2)
\end{aligned} \tag{4.15}$$

while for $v_\varepsilon^k < 0$ we have:

$$\begin{aligned}
(\Delta_\varepsilon^{\zeta(v_\varepsilon)}\tilde{\Phi}_1(t))^k &= \Phi_{1,xx}(t,(k-1)\varepsilon) + O(\varepsilon^2) \\
(\nabla_\varepsilon^{\zeta(v_\varepsilon)}\tilde{\Phi}_2(t))^k &= -\Phi_{2,x}(t,(k-1)\varepsilon) + O(\varepsilon^2).
\end{aligned} \tag{4.16}$$

Plugging (4.15), (4.16) together with the formulae for Φ_3, $\Phi_{1,x}$, $\Phi_{1,xx}$, $\Phi_{2,x}$ into (4.14), we obtain:

$$(\frac{d}{dt}u_\varepsilon(t), \phi(t)) + ((u_\varepsilon)_x(t), \phi_x(t)) = \sum_k v_\varepsilon^k(t) U_k(t) \tag{4.17}$$

where for $v_\varepsilon^k(t) > 0$ we have:

$$\begin{aligned} U_k(t) &= \varepsilon\phi_x(t, k\varepsilon) - \phi(t, k\varepsilon) + \phi(t, (k-1)\varepsilon) \\ &\quad -\phi(t, (k+1)\varepsilon) + \phi(t, k\varepsilon) + \phi(t, k\varepsilon) - \phi(t, (k-1)\varepsilon) + O(\varepsilon^2) \\ &= \varepsilon\phi_x(t, k\varepsilon) - \phi(t, (k+1)\varepsilon) + \phi(t, k\varepsilon) + O(\varepsilon^2) \\ &= -\frac{\varepsilon^2}{2}\phi_{xx}(t, \xi_k) + O(\varepsilon^2) \end{aligned}$$

while for $v_\varepsilon^k(t) < 0$ we have similarly $U_k(t) = -\frac{\varepsilon^2}{2}\phi_{xx}(t, \eta_k) + O(\varepsilon^2)$.

The regularity of ϕ implies that if $v_\varepsilon^k(t) \neq 0$ we have $U_k(t)$ of magnitude $O(\varepsilon^2)$, otherwise we have $v_\varepsilon^k(t)U_k(t) = 0$. Using the L^2-boundedness property of $v_\varepsilon(t)$ we conclude that the expression in (4.13) can be made thus arbitrary small for ε small enough. The proof is completed. □

References

[1] Ethier, S., Kurtz, T.G., *Markov Processes: Characterization and Convergence*, Wiley, 1986.

[2] Kurtz, T.G., *Limit theorems for sequences of jump Markov processes approximating ordinary differential processes*, J. Appl. Prob., **8**(1971), 344–356.

[3] Lions, P.L., Mas-Gallic, S., *Une méthode particulaire déterministe pour des équations diffusives non linéaires*, C.R. Acad. Sci. Paris, **332**(2001), no. 1, 369–376.

[4] Mas-Gallic, S., *The diffusion velocity method as a tool for general diffusion equations*, in Nonlinear problems in mathematical physics and related topics II. In honour of Professor O. A. Ladyzhenskaya. (Birman, Michael Sh. (ed.) et al.) Kluwer Academic Publishers, New York, Int. Math. Ser., N.Y. 2, 2002, 285–303.

[5] Russo, G., *Deterministic diffusion of particles*, Comm. Pure Appl. Math., **XLIII**(1990), 697–733.

[6] Renardy, M., Rogers, R., *An Introduction to Partial Differential Equations*, Springer, New York, 1993.

Flavius Guiaş
Dortmund University of Technology
Faculty of Mathematics
Vogelpothsweg 87
44221 Dortmund
Germany
e-mail: `flavius.guias@mathematik.tu-dortmund.de`

Volume 56, Number 2, June, 2011

S T U D I A
UNIVERSITATIS BABEŞ-BOLYAI
MATHEMATICA

CONTENTS

NUMBER 2A

NUMBER 2B

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 413–422

Uniform approximation in weighted spaces using some positive linear operators

Adrian Holhoş

Abstract. We characterize the functions defined on a weighted space, which are uniformly approximated by the Post-Widder, Gamma, Weierstrass and Picard operators and we obtain the range of the weights which can be used for uniform approximation. We give, also, an estimation of the rate of the approximation in terms of the usual modulus of continuity. Some well-known results are obtained, as limit cases.

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

Keywords: Positive linear operator, rate of approximation, weighted spaces, modulus of continuity.

1. Introduction

In the survey paper [2], the authors present some ideas related to the approximation of functions in weighted spaces and enounced some unsolved problems in weighted approximation theory. Three such problems are:
1. Let $\mathcal{F}$ be a linear subspace of $\mathbb{R}^I$ and $A_n\colon \mathcal{F} \to C(I)$ a sequence of positive linear operators. For which weights ρ, does A_n map $C_\rho(I) \cap \mathcal{F}$ onto $C_\rho(I)$ with uniformly bounded norms?
2. For which functions $f \in C_\rho(I)$ do we have $\|A_n - f\|_\rho \to 0$, as $n \to \infty$?
3. Which moduli of smoothness are appropriate for weighted approximation?

In [6], we presented a result that give an answer to this questions. Below, in Theorem 1.1 we recall this result. In the same paper, we analized the particular cases of Szász-Mirakjan and Baskakov operators. In this paper, we continue the applications of the general result in the case of some integral-type positive linear operators, namely: the Post-Widder, Gamma, Gauss-Weierstrass and Picard operators. Firstly, we introduce the basic notations.

Let $I \subseteq \mathbb{R}$ be a noncompact interval and let $\rho\colon I \to [1,\infty)$ be an increasing and differentiable function called weight. Let $B_\rho(I)$ be the space of all functions $f\colon I \to \mathbb{R}$ such that $|f(x)| \le M \cdot \rho(x)$, for all $x \in I$, where $M > 0$ is a constant depending on f and ρ, but independent of x. The space

$B_\rho(I)$ is called weighted space and it is a Banach space endowed with the ρ-norm

$$\|f\|_\rho = \sup_{x\in I} \frac{|f(x)|}{\rho(x)}.$$

Let $C_\rho(I) = C(I) \cap B_\rho(I)$ be the subspace of $B_\rho(I)$ containing continuous functions.

Let $(A_n)_{n\geq 1}$ be a sequence of positive linear operators defined on the weighted space $C_\rho(I)$. It is known (see [4]) that A_n maps $C_\rho(I)$ onto $B_\rho(I)$ if and only if $A_n\rho \in B_\rho(I)$.

Theorem 1.1. *Let $A_n : C_\rho(I) \to B_\rho(I)$ be positive linear operators reproducing constant functions and satisfying the conditions*

$$\sup_{x\in I} A_n(|\varphi(t) - \varphi(x)|, x) = a_n \to 0, \quad (n \to \infty) \tag{1.1}$$

$$\sup_{x\in I} \frac{A_n(|\rho(t) - \rho(x)|, x)}{\rho(x)} = b_n \to 0. \quad (n \to \infty) \tag{1.2}$$

If $A_n(f,x)$ is continuously differentiable and there is a constant $K(f,\rho,n)$ such that

$$\frac{|(A_nf)'(x)|}{\varphi'(x)} \leq K(f,\rho,n)\cdot\rho(x), \quad \textit{for every } x \in I, \tag{1.3}$$

and ρ and φ are such that there exists a constant $\alpha > 0$ with the property

$$\frac{\rho'(x)}{\varphi'(x)} \leq \alpha\cdot\rho(x), \quad \textit{for every } x \in I, \tag{1.4}$$

then, the following statements are equivalent

(*i*) $\|A_nf - f\|_\rho \to 0$ *as* $n \to \infty$.

(*ii*) $\dfrac{f}{\rho}\circ\varphi^{-1}$ *is uniformly continuous on* J.

Furthermore, we have

$$\|A_nf - f\|_\rho \leq b_n \cdot \|f\|_\rho + 2\cdot\omega\left(\frac{f}{\rho}\circ\varphi^{-1}, a_n\right), \quad \textit{for every } n \geq 1.$$

Remark 1.2. The relation (1.4) give us the connection between the function φ and the weight ρ. We must have

$$\rho(x) \leq Me^{\alpha\cdot\varphi(x)}, \qquad \text{for every } x \in I,$$

where $M, \alpha > 0$ are constants independent of x. So, we have obtained the range of the weights ρ, for which Theorem 1.1 is valid. In the case of the maximal class of weights: $\rho(x) = e^{\alpha\varphi(x)}$, instead of proving the conditions (1.1) and (1.2) we prove

$$\lim_{n\to\infty} \sup_{x\in I} A_n(|\varphi(t) - \varphi(x)|^2, x) = 0. \tag{1.5}$$

For the estimation of the sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ we use the inequalities

$$a_n \le \sup_{x\in I} \sqrt{A_n(|\varphi(t)-\varphi(x)|^2,x)}$$
$$b_n \le \frac{\alpha}{2}\sqrt{\|A_n\rho^2\|_{\rho^2} + 2\,\|A_n\rho\|_\rho + 1}\cdot \sup_{x\in I}\sqrt{A_n(|\varphi(t)-\varphi(x)|^2,x)}.$$

2. Main results

The Post-Widder operators.

Lemma 2.1. *For $I=(0,\infty)$ and for $\rho(x)=1+x^\alpha$, for some $\alpha>0$, the Post-Widder operators (*[9], [14]*)*

$$P_n(f,x) = \frac{1}{(n-1)!}\left(\frac{n}{x}\right)^n \int_0^\infty e^{-\frac{nu}{x}} u^{n-1} f(u)\,du, \quad x>0,$$

have the property that $P_n f\in C_\rho(0,\infty)$ for every $f\in C_\rho(0,\infty)$.

Proof. Setting $t=nu/x$, we get

$$P_n(\rho,x) = 1+\frac{1}{(n-1)!}\int_0^\infty e^{-t}t^{n-1}\left(\frac{xt}{n}\right)^\alpha dt = 1+\frac{x^\alpha\Gamma(n+\alpha)}{n^\alpha(n-1)!}.$$

Using the formula (see [1, formula 6.1.46])

$$\lim_{n\to\infty}\frac{\Gamma(n+\alpha)}{n^\alpha\Gamma(n)}=1,$$

we deduce the existence of a constant $C>0$, independent of n, such that $\Gamma(n+\alpha)\le Cn^\alpha(n-1)!$, for every $n\ge 1$. We obtain

$$P_n(\rho,x)\le C\rho(x), \quad x>0,$$

which proves the mapping property of P_n. □

Theorem 2.2. *For $\alpha>0$ and $\rho(x)=1+x^\alpha$, the Post-Widder operators $P_n\colon C_\rho(0,\infty)\to C_\rho(0,\infty)$ have the property*

$$\|P_nf-f\|_\rho \to 0, \quad \textit{whenever } n\to\infty$$

if and only if

$$f(e^x)e^{-\alpha x} \quad \textit{is uniformly continuous on } (0,\infty).$$

Moreover, for every $f\in C_\rho(0,\infty)$ and every $n\ge 2$, we have

$$\|P_nf-f\|_\rho \le \|f\|_\rho\frac{\alpha C}{\sqrt{n-1}} + 2\cdot\omega\left(f(e^t)e^{-\alpha t}, \frac{1}{\sqrt{n-1}}\right),$$

where $C=\sup_{n\in\mathbb{N}}\frac{1}{2}\sqrt{\|P_n\rho^2\|_{\rho^2}+2\,\|P_n\rho\|_\rho+1}<\infty$ is a constant depending only on α.

Proof. Using the Geometric-Logarithmic-Arithmetic Mean Inequality (see [8, p. 40])
$$\sqrt{u \cdot v} \leq \frac{u-v}{\ln u - \ln v} < \frac{u+v}{2}, \quad 0 < v < u, \tag{2.1}$$
for the function $\varphi(x) = \ln x$, we obtain
$$|\varphi(t) - \varphi(x)| \leq \left| \sqrt{\frac{t}{x}} - \sqrt{\frac{x}{t}} \right|, \quad t, x > 0.$$
Because $P_n(1,x) = 1$, $P_n(t,x) = x$ and $P_n\left(\frac{1}{t}, x\right) = \frac{n}{(n-1)x}$, $n \geq 2$, we deduce
$$\sup_{x>0} P_n(|\varphi(t) - \varphi(x)|^2, x) \leq \sup_{x>0} \left[P_n\left(\frac{t}{x}, x\right) + P_n\left(\frac{x}{t}, x\right) - 2 \right] = \frac{1}{n-1},$$
which proves (1.5)

Now, using the equality (see [12])
$$P_n((t-x)^2, x) = \frac{x^2}{n}$$
and the Cauchy-Schwarz inequality for positive linear operators, we have
$$P_n(|t-x|\rho(t), x) \leq \sqrt{P_n((t-x)^2, x) \cdot P_n(\rho^2, x)} \leq \frac{x}{\sqrt{n}} \cdot C_1 \rho(x).$$
Estimating the absolute value of the derivative
$$\begin{aligned}
|(P_n f)'(x)| &= \frac{n}{x^2} \left| \int_0^\infty \left(\frac{n}{x}\right)^n \frac{1}{(n-1)!} e^{-\frac{nu}{x}} u^{n-1} (u-x) f(u)\, du \right| \\
&\leq \frac{n}{x^2} P_n(|t-x| \cdot |f(t)|, x)| \leq \frac{n}{x^2} \|f\|_\rho P_n(|t-x|\rho(t), x) \\
&\leq \|f\|_\rho \frac{\sqrt{n}}{x} C_1 \rho(x),
\end{aligned}$$
we obtain
$$\frac{|(P_n f)'(x)|}{\varphi'(x)} \leq C_2 \rho(x), \quad \text{for every } x > 0,$$
which proves (1.3) The relation (1.4) is true because
$$\frac{\rho'(x)}{\varphi'(x)} = \alpha x^\alpha \leq \alpha(1 + x^\alpha) = \alpha \rho(x).$$
Using the Theorem 1.1, the convergence $\|P_n f - f\|_\rho \to 0$ is true if and only if the function $\frac{f}{\rho} \circ \varphi^{-1}$ is uniformly continuous on $(0, \infty)$. The equality
$$\frac{f(e^x)}{e^{\alpha x}} = \frac{f(e^x)}{1 + e^{\alpha x}} \cdot \left(1 + e^{-\alpha x}\right),$$
the boundedness of the function $1 \leq 1 + e^{-\alpha x} \leq 2$ and the uniform continuity of the functions $1 + e^{-\alpha x}$ and $(1 + e^{-\alpha x})^{-1}$ prove that $\frac{f}{\rho} \circ \varphi^{-1}$ is uniformly continuous, if and only if $f(e^x)e^{-\alpha x}$ is uniformly continuous. □

Remark 2.3. The result of the Theorem 2.2 for the limit case, $\alpha = 0$, was obtained in [12] and in [3].

The Gamma operators.

Lemma 2.4. *For $I = (0, \infty)$ and for $\rho(x) = 1 + x^\alpha$, for some $\alpha > 0$ the Gamma operators ([7])*

$$G_n(f, x) = \frac{x^{n+1}}{n!} \int_0^\infty e^{-xu} u^n f\left(\frac{n}{u}\right) du, \quad x > 0,\ n \geq 1,$$

have the property that $G_n f \in C_\rho(0, \infty)$ for every $f \in C_\rho(0, \infty)$ and $n \geq [\alpha]$.

Proof. Setting $xu = t$, we get

$$G_n(\rho, x) = 1 + \frac{1}{n!} \int_0^\infty e^{-t} t^n \left(\frac{nx}{t}\right)^\alpha dt = 1 + \frac{(nx)^\alpha \Gamma(n + 1 - \alpha)}{n!}.$$

Using the formula (see [1, formula 6.1.46])

$$\lim_{n\to\infty} \frac{n^\alpha \Gamma(n + 1 - \alpha)}{\Gamma(n + 1)} = 1,$$

we deduce the existence of a constant $C > 0$, independent of n, such that $n^\alpha \Gamma(n + 1 - \alpha) \leq C n!$, for every $n \geq [\alpha]$. We obtain

$$G_n(\rho, x) \leq C\rho(x), \quad x > 0,$$

which proves the property of G_n stated in the lemma. □

Theorem 2.5. *For $\alpha > 0$ and $\rho(x) = 1 + x^\alpha$, the Gamma operators $G_n \colon C_\rho(0, \infty) \to C_\rho(0, \infty)$ have the property*

$$\|G_n f - f\|_\rho \to 0, \quad \textit{whenever } n \to \infty$$

if and only if

$$f(e^x)e^{-\alpha x} \quad \textit{is uniformly continuous on } (0, \infty).$$

Moreover, for every $f \in C_\rho(0, \infty)$ and every $n \geq [2\alpha]$, we have

$$\|G_n f - f\|_\rho \leq \|f\|_\rho \frac{\alpha C}{\sqrt{n}} + 2 \cdot \omega\left(f(e^t)e^{-\alpha t}, \frac{1}{\sqrt{n}}\right),$$

where $C = \sup_{n\in\mathbb{N}} \frac{1}{2}\sqrt{\|G_n \rho^2\|_{\rho^2} + 2\|G_n \rho\|_\rho + 1} < \infty$ is a constant depending only on α.

Proof. As in the proof of the Theorem 2.2, let $\varphi(x) = \ln x$. We have

$$|\ln t - \ln x| \leq \left|\sqrt{\frac{t}{x}} - \sqrt{\frac{x}{t}}\right|, \quad t, x > 0.$$

Because $G_n(e_0, x) = 1,\ G_n(t, x) = x$ and

$$G_n\left(\frac{1}{t}, x\right) = \frac{n+1}{nx},$$

we deduce

$$\sup_{x>0} G_n(|\varphi(t) - \varphi(x)|^2, x) \leq \sup_{x>0} \left[G_n\left(\frac{t}{x}, x\right) + G_n\left(\frac{x}{t}, x\right) - 2\right] = \frac{1}{n},$$

which proves (1.5).

Estimating the derivative

$$\begin{aligned}|(G_nf)'(x)| &= \left|\frac{n+1}{x}G_n(f,x) - \frac{n+1}{x}G_{n+1}\left(f\left(\frac{nt}{n+1}\right),x\right)\right| \\ &\le \frac{n+1}{x}\left\|f\right\|_\rho |G_n(\rho,x)+G_{n+1}(\rho,x)| \\ &\le \left\|f\right\|_\rho \frac{n+1}{x} C_1\rho(x),\end{aligned}$$

we deduce

$$\frac{|(G_nf)'(x)|}{\varphi'(x)} \le C_2\rho(x), \quad \text{for every } x>0,$$

which proves (1.3). The relation (1.4) is true, because

$$\frac{\rho'(x)}{\varphi'(x)} = \alpha x^\alpha \le \alpha(1+x^\alpha) = \alpha\rho(x).$$

Using the Theorem 1.1, the convergence $\left\|P_nf-f\right\|_\rho \to 0$ is true if and only if the function $\frac{f}{\rho}\circ\varphi^{-1}$ is uniformly continuous on $(0,\infty)$. The equality

$$\frac{f(e^x)}{e^{\alpha x}} = \frac{f(e^x)}{1+e^{\alpha x}}\cdot\left(1+e^{-\alpha x}\right),$$

the boundedness of the function $1 \le 1+e^{-\alpha x} \le 2$ and the uniform continuity of the functions $1+e^{-\alpha x}$ and $(1+e^{-\alpha x})^{-1}$ prove that $\frac{f}{\rho}\circ\varphi^{-1}$ is uniformly continuous, if and only if $f(e^x)e^{-\alpha x}$ is uniformly continuous. □

Remark 2.6. The result of the Theorem 2.5 for the limit case, $\alpha = 0$, was obtained in [11].

The Gauss-Weierstrass operators.

Lemma 2.7. *For $I=\mathbb{R}$ and for $\rho(x)=e^{\alpha x}$, for some $\alpha>0$, the Gauss-Weierstrass operators (*[13]*)*

$$W_nf(x) = \frac{\sqrt{n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-n\frac{(u-x)^2}{2}} f(u)\,du, \quad x\in(-\infty,\infty),$$

have the property that $W_nf\in C_\rho(\mathbb{R})$ for $f\in C_\rho(\mathbb{R})$.

Proof. We have

$$\begin{aligned}\frac{W_n(\rho,x)}{\rho(x)} &= \frac{\sqrt{n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-n\frac{(u-x)^2}{2}+\alpha(u-x)}\,du \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{n}{2}\left(u-x-\frac{\alpha}{n}\right)^2}\cdot e^{\frac{\alpha^2}{2n}}\,du = e^{\frac{\alpha^2}{2n}} \le e^{\frac{\alpha^2}{2}},\end{aligned}$$

which proves the statement from the lemma. □

Theorem 2.8. *For $\alpha>0$ and for $\rho(x)=e^{\alpha x}$ the Gauss-Weierstrass operators $W_n\colon C_\rho(\mathbb{R})\to C_\rho(\mathbb{R})$ have the property*

$$\left\|W_nf-f\right\|_\rho \to 0, \quad \textit{whenever } n\to\infty,$$

if and only if

$$f(x)e^{-\alpha x} \quad \text{is uniformly continuous on } \mathbb{R}.$$

Moreover, for every $f \in C_\rho(\mathbb{R})$ and for every $n \geq 1$, we have

$$\|W_n f - f\|_\rho \leq \|f\|_\rho \frac{\alpha C}{\sqrt{n}} + 2 \cdot \omega\left(f(t)e^{-\alpha t}, \frac{1}{\sqrt{n}}\right),$$

where $C = e^{\frac{\alpha^2}{2}}\sqrt{1 + \frac{\alpha^2}{4}\left(1 + e^{\frac{\alpha^2}{2}}\right)^2}$.

Proof. Set $\varphi(x) = x$. Because $W_n(e_0, x) = 1$ and $W_n((t-x)^2, x) = \frac{1}{n}$ (see [10]), we get

$$W_n(|\varphi(t) - \varphi(x)|^2, x) = W_n((t-x)^2, x) = \frac{1}{n},$$

which proves (1.5). Using the relation

$$W_n(e^{\alpha t}, x) = e^{\alpha x} \cdot e^{\frac{\alpha^2}{2n}}$$

we deduce

$$\begin{aligned}
b_n &= \sup_{x \in I} \frac{W_n(|\rho(t) - \rho(x)|, x)}{\rho(x)} = \frac{W_n\left(\left|e^{\alpha t} - e^{\alpha x}\right|, x\right)}{e^{\alpha x}} \\
&\leq \frac{\sqrt{W_n(e^{2\alpha t}, x) - 2e^{\alpha x} W_n(e^{\alpha t}, x) + e^{2\alpha x}}}{\rho(x)} \\
&= \frac{\sqrt{e^{2\alpha x} \cdot e^{\frac{4\alpha^2}{2n}} - 2e^{2\alpha x} \cdot e^{\frac{\alpha^2}{2n}} + e^{2\alpha x}}}{e^{\alpha x}} = \sqrt{e^{\frac{4\alpha^2}{2n}} - 2e^{\frac{\alpha^2}{2n}} + 1}.
\end{aligned}$$

Using the equality $x^4 - 2x + 1 = (x-1)[(x-1)(x+1)^2 + 2x]$ and the inequality $e^t - 1 \leq te^t$, for $t = \frac{\alpha^2}{2n}$, we obtain

$$\begin{aligned}
b_n &\leq \sqrt{\left(e^{\frac{\alpha^2}{2n}} - 1\right)} \cdot \sqrt{\left(e^{\frac{\alpha^2}{2n}} - 1\right)\left(e^{\frac{\alpha^2}{2n}} + 1\right)^2 + 2e^{\frac{\alpha^2}{2n}}} \\
&\leq \frac{\alpha}{\sqrt{2n}} e^{\frac{\alpha^2}{2}} \sqrt{2 + \frac{\alpha^2}{2}\left(1 + e^{\frac{\alpha^2}{2}}\right)^2} \leq \frac{\alpha C}{\sqrt{n}}.
\end{aligned}$$

The estimation of the derivative

$$\begin{aligned}
|(W_n f)'(x)| &= n|W_n((t-x)f(t), x)| \leq n\|f\|_\rho W_n(|t-x|\rho(t), x) \\
&\leq n\|f\|_\rho \sqrt{W_n((t-x)^2, x)}\sqrt{W_n(e^{2\alpha t}, x)} \\
&= \sqrt{n}\|f\|_\rho e^{\frac{2\alpha^2}{n}} \rho(x)
\end{aligned}$$

proves the relation

$$\frac{|(W_n f)'(x)|}{\varphi'(x)} \leq C_1 \rho(x), \quad \text{for every } x \in \mathbb{R}.$$

$\square$

Remark 2.9. The result of the Theorem 2.8 for the limit case, $\alpha = 0$, was obtained in [5] and partially in [10].

The Picard Operators.

Lemma 2.10. *For $I = \mathbb{R}$ and for $\rho(x) = e^{\alpha x}$, for some $\alpha > 0$, the Picard operators*

$$\mathcal{P}_n(f, x) = \frac{n}{2} \int_{-\infty}^{\infty} e^{-n|u-x|} f(u)\, du, \quad x \in \mathbb{R},\ n \geq [\alpha] + 2,$$

have the property that $\mathcal{P}_n \rho \in C_\rho(\mathbb{R})$ for every $f \in C_\rho(\mathbb{R})$.

Proof. The evaluation

$$\begin{aligned}
\frac{\mathcal{P}_n(\rho, x)}{\rho(x)} &= \frac{n}{2} \int_{-\infty}^{x} e^{\alpha u - nx + nu - \alpha x}\, du + \frac{n}{2} \int_{x}^{\infty} e^{\alpha u + nx - nu - \alpha x}\, du \\
&= \frac{n}{2} e^{-nx - \alpha x} \left. \frac{e^{u(\alpha+n)}}{\alpha + n} \right|_{-\infty}^{x} + \frac{n}{2} e^{nx - \alpha x} \left. \frac{e^{u(\alpha - n)}}{\alpha - n} \right|_{x}^{\infty} \\
&= \frac{n^2}{n^2 - \alpha^2} \leq 1 + \alpha,
\end{aligned}$$

proves the statement from the lemma. □

Theorem 2.11. *For $\alpha > 0$ and for $\rho(x) = e^{\alpha x}$ the Picard operators $\mathcal{P}_n \colon C_\rho(\mathbb{R}) \to C_\rho(\mathbb{R})$, $n \geq [2\alpha] + 2$, have the property*

$$\|\mathcal{P}_n f - f\|_\rho \to 0, \quad \textit{whenever } n \to \infty,$$

if and only if

$$f(x) e^{-\alpha x} \quad \textit{is uniformly continuous on } \mathbb{R}.$$

Furthermore, for every $f \in C_\rho(\mathbb{R})$ and for every $n \geq [2\alpha] + 2$, it is true the estimation

$$\|\mathcal{P}_n f - f\|_\rho \leq \|f\|_\rho \frac{\alpha C}{n} + 2 \cdot \omega \left(f(t) e^{-\alpha t}, \frac{\sqrt{2}}{n} \right),$$

where $C > 0$ is a constant dependening on α, but independent of n.

Proof. Set $\varphi(x) = x$. Using the relations $\mathcal{P}_n(e_0, x) = 1$, $\mathcal{P}_n(e_1, x) = x$ and $\mathcal{P}_n(e_2, x) = x^2 + \frac{2}{n^2}$, we obtain

$$a_n = \sup_{x \in \mathbb{R}} \mathcal{P}_n(|\varphi(t) - \varphi(x)|, x) \leq \sup_{x \in \mathbb{R}} \sqrt{\mathcal{P}_n((t-x)^2, x)} = \frac{\sqrt{2}}{n},$$

which proves (1.1). Using the equality

$$\mathcal{P}_n(e^{\alpha t}, x) = \frac{n^2 e^{\alpha x}}{n^2 - \alpha^2},$$

obtained in the previous lemma, we get

$$\begin{aligned}
b_n &= \sup_{x\in\mathbb{R}} \frac{\mathcal{P}_n(|\rho(t)-\rho(x)|,x)}{\rho(x)} = \sup_{x\in\mathbb{R}} \frac{\mathcal{P}_n\left(\left|e^{\alpha t}-e^{\alpha x}\right|,x\right)}{e^{\alpha x}} \\
&\le \sup_{x\in\mathbb{R}} \frac{\sqrt{\mathcal{P}_n(e^{2\alpha t},x) - 2e^{\alpha x}\mathcal{P}_n(e^{\alpha t},x) + e^{2\alpha x}}}{\rho(x)} \\
&= \sqrt{\frac{n^2}{n^2-4\alpha^2} - \frac{2n^2}{n^2-\alpha^2} + 1} = \alpha\sqrt{\frac{2(n^2+2\alpha^2)}{(n^2-4\alpha^2)(n^2-\alpha^2)}} \le \frac{\alpha C}{n},
\end{aligned}$$

where

$$C^2 = \max_{n\ge[2\alpha]+2} \frac{2n^2(n^2+2\alpha^2)}{(n^2-4\alpha^2)(n^2-\alpha^2)}.$$

Using the relation

$$\mathcal{P}_n(f,x) = \frac{n}{2}\int_{-\infty}^{x} f(u)e^{-n(x-u)}\,du + \frac{n}{2}\int_{x}^{\infty} f(u)e^{-n(u-x)}\,du$$

we can compute the derivative

$$\begin{aligned}
\mathcal{P}_n'(f,x) &= \frac{n^2}{2}\left(\int_{x}^{\infty} f(u)e^{-n(u-x)}\,du - \int_{-\infty}^{x} f(u)e^{-n(x-u)}\,du\right) \\
&= \frac{n^2}{2}\int_0^\infty \left[f(x+t)-f(x-t)\right]e^{-nt}\,dt
\end{aligned}$$

and obtain the estimation

$$\begin{aligned}
|\mathcal{P}_n'(f,x)| &\le \frac{n^2}{2}\int_0^\infty |f(x+t)-f(x-t)|\,e^{-nt}\,dt \\
&\le \|f\|_\rho \frac{n^2}{2}\int_0^\infty \left[e^{\alpha(x+t)} + e^{\alpha(x-t)}\right]e^{-nt}\,dt \\
&\le e^{\alpha x}\,\|f\|_\rho \frac{n^3}{n^2-\alpha^2}.
\end{aligned}$$

This proves the inequality

$$\frac{|\mathcal{P}_n'(f,x)|}{\varphi'(x)} \le C_{n,\alpha}\rho(x), \qquad \text{for every } x\in\mathbb{R}.$$

□

Corollary 2.12. *For a continuous and bounded function $f:\mathbb{R}\to\mathbb{R}$, it is true the equivalence*

$$\|\mathcal{P}_n f - f\| \to 0,\ (n\to\infty) \text{ if and only if } f \text{ is uniformly continuous on } \mathbb{R}.$$

Moreover,

$$\|\mathcal{P}_n f - f\| \le 2\cdot\omega\left(f,\frac{\sqrt{2}}{n}\right),\ n\ge 2.$$

References

[1] Abramowitz, M., Stegun, I.A., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York, Dover Publications, 1972.

[2] Bustamante, J., Morales, L., de la Cruz, *Positive linear operators and continuous functions on unbounded intervals*, Jaen J. Approx., **1**(2009), no. 2, 145–173.

[3] de la Cal, J., Cárcamo, J., *On uniform approximation by some classical Bernstein-type operators*, J. Math. Anal. Appl., **279**(2003), 625–638.

[4] Gadjiev, A.D., *Theorems of Korovkin type*, Mat. Zametki., **20**(1976), no. 5, 781–786, (in Russian), Math. Notes, **20**(1976), no. 5, 996–998.

[5] Holhoş, A., *Uniform Approximation by Positive Linear Operators on Noncompact Intervals*, Automat. Comput. Appl. Math., **18**(2009), no. 1, 121–132.

[6] Holhoş, A., *Uniform weighted approximation by positive linear operators*, Stud. Univ. Babeş-Bolyai Math., submitted.

[7] Lupaş, A., Müller, M., *Approximationseigenschaften der Gammaoperatoren*, Math. Z., **98**(1967), 208–226.

[8] Mitrinović, D.S., Pečarić, J., Fink, A.M., *Classical and New Inequalities in Analysis*, Kluwer Academic Publishers, Dordrecht, 1993.

[9] Post, E.L., *Generalized differentiation*, Trans. Amer. Math. Soc., **32**(1930), 723–781.

[10] Stancu, D.D., *Use of probabilistic methods in the theory of uniform approximation of continuous functions*, Rev. Roumaine Math. Pures Appl., **14**(1969), 673–691.

[11] Totik, V., *Uniform approximation by positive operators on infinite intervals*, Anal. Math., **10**(1984), 163–182.

[12] Totik, V., *Uniform approximation by Exponential-Type Operators*, J. Math. Anal. Appl., **132**(1988), 238–246.

[13] Weierstrass, K., *Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen*, Sitzungsber. Akad. Berlin, 1885, 633–639, 789–805.

[14] Widder, D.V., *The inversion of the Laplace integral and the related moment problem*, Trans. Amer. Math. Soc., **36**(1934), 107–200.

Adrian Holhoş
Technical University of Cluj-Napoca
28, Memorandumului
400114 Cluj-Napoca
Romania
e-mail: adrian.holhos@math.utcluj.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 423–436

Recent results on Chlodovsky operators

Harun Karsli

Abstract. We take a view on the results concerning the Bernstein–Chlodovsky operators obtained especially in the last five years. The list presented in this paper is not exhaustive. We apologise all authors possessing papers on the Chlodovsky operators and are not referred in this paper.

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

Keywords: Bernstein-Chlodovsky polynomials, positive linear operators.

1. Introduction

For a function f defined on $[0,\infty)$ and bounded on every finite interval $[0,b] \subset [0,\infty)$, the classical Bernstein-Chlodovsky operators are defined by

$$(C_n f)(x) := \sum_{k=0}^{n} f\left(\frac{b_n}{n}k\right) p_{k,n}\left(\frac{x}{b_n}\right), \tag{1.1}$$

where $p_{k,n}$ denotes as usual

$$p_{k,n}(x) = \left(\begin{array}{c} n \\ k \end{array}\right) x^k (1-x)^{n-k}, \quad 0 \leq x \leq 1,$$

and $(b_n)_{n=1}^{\infty}$ is a positive increasing sequence of reals with the properties

$$\lim_{n\to\infty} b_n = \infty \quad , \quad \lim_{n\to\infty} \frac{b_n}{n} = 0. \tag{1.2}$$

These polynomials were introduced by I. Chlodovsky [11] in **1937** to generalize the Bernstein polynomials $(B_n f)(x)$, for the case $b_n = 1$, $n \in \mathbb{N}_0$, which approximate the function f on the interval $[0,1]$ (or, suitably modified on any fixed finite interval $[-b,b]$). His main result is the following:

Chlodovsky's Theorem. *Let* (b_n) *satisfy* (1.2) *and, for* $b > 0$, *let* $M(b;f) := \sup\limits_{0\leq t\leq b} |f(t)|$. *If*

$$\lim_{n\to\infty} M(b_n;f)\exp(-\sigma n/b_n) = 0 \qquad for\ each\ \sigma > 0, \tag{1.3}$$

then

$$\lim_{n\to\infty} (C_n f)(x) = f(x)$$

at each point x of continuity of the function f.

As a corollary he states that if a function f belonging to $C[0,\infty)$ is of order $f(x) = \mathcal{O}(\exp x^p)$ for some $p > 0$, and if the sequence $\{b_n\}$ satisfies the condition

$$b_n \leq n^{\frac{1}{p+1+\eta}},$$

where $\eta > 0$, no matter how small, then $(C_n f)(x)$ converges to $f(x)$ at each point $x \in \mathbb{R}^+$.

The first part of the next and very important lemma is due to Chlodovsky [11].

For $t \in [0,1]$ the inequality

$$0 \leq z \leq \frac{3}{2}\sqrt{nt(1-t)}$$

implies

$$\sum_{|k-nt|\geq 2z\sqrt{nt(1-t)}} p_{k,n}(t) \leq 2\exp\left(-z^2\right).$$

In particular, for $0 < \delta \leq x < b_n$ and sufficiently large n,

$$\sum_{\left|\frac{kb_n}{n}-x\right|\geq\delta} p_{k,n}\left(\frac{x}{b_n}\right) \leq 2\exp\left(-\frac{\delta^2}{4x}\frac{n}{b_n}\right). \tag{1.4}$$

The proof of (1.4) is given in the **1960** by Albrycht and Radecki [2].

Chlodovsky showed more, namely the simultaneous convergence of the derivative $(C_n f)'(x)$ to $f'(x)$ at points x where it exists, a result taken up by Butzer [6].

Next question concerning Chlodovsky operators was the rate of approximation by $(C_n f)(x)$ to $f(x)$, which is the counterpart of the classical questions for Bernstein polynomials answered by Voronovskaya [29] in **1932**. She showed that for bounded f on $[0,1]$, one has the asymptotic formula

$$\lim_{n\to\infty} n[(B_n f)(x_0) - f(x_0)] = \frac{x_0(1-x_0)}{2} f''(x_0) \tag{1.5}$$

at each fixed point $x_0 \in [0,1]$ for which there exists $f''(x_0) \neq 0$.

The following relations of the Voronovskaya-type for the Chlodovsky operators and their derivatives are presented in [2].

If a function f satisfies

$$\lim_{n\to\infty} \frac{n}{b_n} \exp\left(-\sigma\frac{n}{b_n}\right) M(b_n; f) = 0 \quad \textit{for each } \sigma > 0,$$

then the Voronovskaya-type theorems for Chlodovsky operators read

$$\lim_{n\to\infty} \frac{n}{b_n}[C_n f(x) - f(x)] = \frac{x}{2} f''(x)$$

at each point $x \geq 0$ for which $f''(x)$ exists.

2002 [3] In their introduction the authors write that " as far as we know, [a Voronovskaya-type formula] cannot be stated for the classical C_n". For this purpose they introduced the "more flexible" polynomials

$$C_n^* f(x) = \sum_{k=0}^{n} f\left(\frac{c_n}{n}k\right)\left(\begin{array}{c} n \\ k \end{array}\right)\left(\frac{x}{b_n}\right)^k\left(1-\frac{x}{b_n}\right)^{n-k}$$

for which $b_n \leq c_n$ for all $n \geq 1$, and $b_n \to 0$, $b_n/n \to 0$, with $b_n - c_n \to 0$, all as $n \to \infty$. They worked in the weighted (polynomial) space. Their main theorem stated that

$$\lim_{n\to\infty} \rho_n[C_n^* f(x) - f(x)] = axf''(x) + bxf'(x),$$

where $\{\rho_n\}$ is a divergent increasing sequence of reals such that $\rho_n c_n/n \to 2a$ and $\rho_n(c_n/b_n - 1) \to b$ as $n \to \infty$, $a, b \geq 0$.

It is a pity these authors were not aware of the paper [2].

2003 [4] In this paper it is presented the extension of (1.5) to derivatives of the Bernstein polynomials. The result states that for bounded f for which $f'''(x)$ exists at $x \in [0,1]$, one has

$$\lim_{n\to\infty} n[(B_n f)'(x) - f'(x)] = \frac{1-2x}{2} f''(x) + \frac{x(1-x)}{2} f'''(x). \qquad (1.6)$$

2. A brief history of the recent results on Chlodovsky operators (2005-...)

We present below, in chronological order, a list of papers dealing with the Bernstein-Chlodovsky Polynomials.

2005 [13] We introduce a Chlodovsky Type Durrmeyer operator as follows: $D_n : BV[0,\infty) \to \mathcal{P}$,

$$\left(D_n f\right)(x) = \frac{(n+1)}{b_n} \sum_{k=0}^{n} p_{k,n}\left(\frac{x}{b_n}\right) \int_0^{b_n} f(t) p_{k,n}\left(\frac{t}{b_n}\right) \; dt, \;\; 0 \leq x \leq b_n \quad (2.1)$$

where $\mathcal{P} := \{P : [0,\infty) \to R\}$, is a polynomial functions set, and $p_{k,n}(x) = \binom{n}{k} x^k (1-x)^{n-k}$ is the Bernstein basis. We estimated the rate of convergence of operators D_n, for functions of bounded variation at the points which one sided limit exist, for functions of bounded variation on the interval $[0,\infty)$, by means of the techniques of probability theory.

2006 [8] The authors establish two inverse theorems for Bernstein-Chlodovsky type polynomials of two variables in a rectangular and a triangular domain.

2006 [15] The aim of this paper is to study the problem of the approximation of functions of two variables by means of Bernstein-Chlodovsky polynomials in a rectangular domain.

2006 [1] The concern of this note is to introduce a general class of linear positive operators of discrete type acting on the space of real valued functions defined on a plane domain. These operators preserve some test functions of Bohman-Korovkin theorem. As a particular class, a modified variant of the bivariate Bernstein-Chlodovsky operators is presented.

2007 [17] We estimate the rate of pointwise convergence of the Chlodovsky-type Bernstein operators $(C_nf)(x)$ for functions defined on the interval $[0, b_n]$, for $b_n \to \infty$ as $n \to \infty$, which are of bounded variation on $[0, \infty)$. At those points for which one-sided limits exists, we shall prove that the operators $(C_nf)(x)$ converge to the limit $\dfrac{f(x+) + f(x-)}{2}$.

2007 [18] Denote by $DBV(I)$, the class of differentiable functions defined on a set $I \subset R$, whose derivatives are with bounded variation on I:

$$DBV(I) = \{f : f' \in BV(I)\}.$$

The aim of this paper is to estimate the rate of convergence of D_nf defined in (2.1) toward f, which is a function that has a derivative with bounded variation on $[0, b_n]$, where $b_n \to \infty$ as n goes to infinity. $(D_nf)(x)$ converges to $f(x)$ in every point x of discontinuity of the first kind of the derivative of f.

2008 [19] We define a new kind of MKZD operators for functions defined on $[0, b_n]$, named Chlodovsky-type MKZD operators as

$$(M_n^*f)(x) = \sum_{k=0}^{\infty} \frac{n+k}{b_n} m_{n,k}\left(\frac{x}{b_n}\right) \int_0^{b_n} f(t) b_{n,k}\left(\frac{t}{b_n}\right) dt, \;\; 0 \le x \le b_n,$$

where $m_{n,k}(x) = \binom{n+k-1}{k} x^k (1-x)^n$ and $b_{n,k}(t) = n\binom{n+k}{k} t^k (1-t)^{n-1}$. The aim of this paper is to study the behavior of the M_n^* operators for functions of bounded variation and give an estimate, by means of the techniques of probability theory, of the rate of convergence of the operators on the interval $[0, b_n]$, $(n \to \infty)$ extending infinity.

2008 [20] The concern of this paper is to study the rate of convergence of C_nf to f for $f \in DBV([0, b_n])$, $(n \to \infty)$ extending infinity. At the point x, which is a discontinuity of the first kind of the derivative, we shall prove that $(C_nf)(x)$ converge to the limit $f(x)$.

2008 [23] For $\alpha \geq 1$, we now introduce Chlodovsky-Bézier operators $C_{n,\alpha}$ as follows:

$$(C_{n,\alpha} f)(x) = \sum_{k=0}^{n} f\left(\frac{k\, b_n}{n}\right) Q_{n,k}^{(\alpha)}\left(\frac{x}{b_n}\right) \ , (\, 0 \leq x \leq b_n\,), \tag{2.2}$$

where $Q_{n,k}^{(\alpha)}(\frac{x}{b_n}) = \left(J_{n,k}(\frac{x}{b_n})\right)^{\alpha} - \left(J_{n,k+1}(\frac{x}{b_n})\right)^{\alpha}$ and $J_{n,k}(\frac{x}{b_n}) = \sum\limits_{j=k}^{n} p_{j,n}(\frac{x}{b_n})$ be the Bézier basis functions. Obviously, $C_{n,\alpha}$ is a positive linear operator and $C_{n,\alpha}(1,x) = 1$. In particular when $\alpha = 1$, the operators (2.2) reduce to the operators (1.1) In this paper, we estimate the rate of pointwise convergence of the Bézier Variant of Chlodovsky operators $C_{n,\alpha}$ for functions, defined on the interval extending infinity, of bounded variation.

2008 [24] We introduce the following q-Chlodovsky polynomials defined as

$$(C_{n,q} f)(x) = \sum_{k=0}^{n} f\left(\frac{[k]_q}{[n]_q} b_n\right) \left[\begin{array}{c} n \\ k \end{array}\right]_q \left(\frac{x}{b_n}\right)^k \prod_{s=0}^{n-k-1} \left(1 - q^s \frac{x}{b_n}\right), 0 \leq x \leq b_n$$

where (b_n) is a positive increasing sequence with the property (1.2). We study some approximation properties of these new operators, which include the well-known Bohman-Korovkin-type theorem, degree of pointwise and uniform convergence and investigation of the monotonocity property of $q-$Chlodovsky operators.

2009 [9] The author introduce the positive linear operators q-Bernstein-Chlodovsky polynomials on a rectangular domain and obtain their Korovkin type approximation properties. The rate of convergence of this generalization is obtained by means of the modulus of continuity, and also by using the K-functional of Peetre. He obtains weighted approximation properties for these positive linear operators and their generalizations.

2009 [16] Approximation on an unbounded interval is studied in this work by means of a new-defined two-parameter polynomial operator based on Chlodovsky polynomials. The operator's properties including convergence rate are investigated using the weighted modulus of continuity.

2009 [7] This paper is first of all devoted to the counterpart of (1.6) for the Chlodovsky polynomials, namely the Voronovskaya-type theorem for $(C_n f)'(x)$. The Theorem states that:

For a function f, defined on $[0, \infty)$

$$\lim_{n\to\infty} \frac{n}{b_n}[(C_n f)'(x) - f'(x)] = \frac{f''(x) + x f'''(x)}{2} \tag{2.3}$$

at each fixed point $x \geq 0$ for which $f'''(x)$ exists, provided that the growth condition (1.3) *is satisfied.*

The second aim of this paper is to study Voronovskaya-type theorems for the derivatives of this operator and to compare the effectiveness of the Szász-Mirakyan operator with the Bernstein-Chlodovsky polynomials in general.

The only way to fully match the assertion of (2.3) is to work with the Szász-Chlodovsky operator

$$\exp\left(-\frac{nx}{b_n}\right)\sum_{k=0}^{\infty} f\left(\frac{kb_n}{n}\right)\left(\frac{nx}{b_n}\right)^k \frac{1}{k!} := (L_n f)(x),$$

defined and studied by Stypinski [28].

In the same paper, given a function f locally integrable on the interval $[0,\infty)$ we define the Kantorovich variant of the Chlodovsky-Bernstein polynomials as

$$(K_n f)(x) := \frac{n+1}{b_{n+1}}\sum_{k=0}^{n} p_{k,n}\left(\frac{x}{b_{n+1}}\right)\int\limits_{\frac{kb_{n+1}}{n+1}}^{\frac{(k+1)b_{n+1}}{n+1}} f(u)du \quad \text{if } 0 \le x \le b_{n+1}, \tag{2.4}$$

where (b_n) satisfies conditions (1.2).

If F denotes the indefinite integral of f, i.e., $F(x) = \int\limits_0^x f(u)du$, then we have $(C_{n+1}F)'(x) = (K_n f)(x)$ for almost all $x \in [0, b_{n+1}]$, in particular for every $x \in [0, b_{n+1}]$ at which f is continuous.

We set

$$M^I(b;f) := \sqrt{\int\limits_0^b |f(u)|^2\, du}.$$

The following result is a corollary of our Theorem on the Voronovskaya-type theorems for the derivatives of $(C_n f)(x)$.

If one has

$$\lim_{n\to\infty} \frac{n}{\sqrt{b_n}} \exp\left(-\alpha\frac{n}{b_n}\right) M^I(b_n;f) = 0$$

for every $\alpha > 0$, *then*

$$\lim_{n\to\infty} \frac{n+1}{b_{n+1}}\left[(K_n f)(x) - f(x)\right] = \frac{f'(x) + xf''(x)}{2}$$

at each fixed point $x \ge 0$ *for which* $f''(x)$ *exists.*

2009 [26] In this paper we introduce the Bézier variant of the Chlodovsky-Kantorovich operators (2.4) of order $(n-1)$ for $f \in L_{loc}[0,\infty)$ as

$$K_{n-1,\alpha} f(x) := \frac{n}{b_n}\sum_{k=0}^{n-1} Q_{n-1,k}^{(\alpha)}\left(\frac{x}{b_n}\right)\int\limits_{\frac{kb_n}{n}}^{\frac{(k+1)b_n}{n}} f(u)du \quad \text{if } 0 \le x \le b_n, \tag{2.5}$$

where $\alpha > 0$, $Q_{n-1,k}^{(\alpha)}(t) = J_{n-1,k}^{\alpha}(t) - J_{n-1,k+1}^{\alpha}(t)$ and $J_{n-1,k}(t)$ are the Bézier basis functions defined for $t \in [0,1]$ as

$$J_{n-1,k}(t) = \sum_{j=k}^{n-1} p_{j,n-1}(t) \quad \text{if } \; k = 0,1,\ldots,n-1,$$

$J_{n-1,n}(t) = 0$. Clearly, if $\alpha = 1$ then $K_{n-1,\alpha}f$ reduce to operators (2.4) with n replaced by $n-1$. Our paper is concerned with the rate of pointwise convergence of operators (2.5) when $f \in M_{loc}[0,\infty)$,i.e. f is measurable and locally bounded on $[0,\infty)$. By using the Chanturiya modulus of variation we present estimations for the rate of convergence of $K_{n-1,\alpha}f(x)$ at the points x of continuity of f and at the discontinuity points of the first kind of f. We will formulate our results for $K_{n-1,\alpha}f$ with $\alpha > 0$. The corresponding estimations for the Chlodovsky-Kantorovich polynomials $K_{n-1}f$ follow immediately as a special case $\alpha = 1$.

2009 [27] The author estimate the rates of convergence of Chlodovsky-Kantorovich polynomials in classes of locally integrable functions. Namely,

if $f \in L_{loc}[0,\infty)$ *and if*

$$\lim_{n\to\infty} \int_0^{b_n} |f(u)|\,du \exp(-\sigma \frac{n}{b_n}) = 0 \quad \text{for each } \sigma > 0,$$

then

$$\lim_{n\to\infty} (K_n f)(x) = f(x) \;\; \text{almost everywhere on } [0,\infty),$$

i.e. at every $x > 0$ *at which* $F'(x) = f(x)$.

Some modified Chlodovsky-Kantorovich operators are considered also in [14].

2009 [22] For $f \in X_{loc}[0,\infty)$ and $\alpha \geq 1$, we introduce the Bézier variant of Chlodovsky-Durrmeyer operators $D_{n,\alpha}$ as follows:

$$(D_{n,\alpha}f)(x) = \frac{n+1}{b_n} \sum_{k=0}^{n} Q_{n,k}^{(\alpha)}\left(\frac{x}{b_n}\right) \int_0^{b_n} f(t) p_{k,n}\left(\frac{t}{b_n}\right) dt, \;\; 0 \leq x \leq b_n, \quad (2.6)$$

Obviously, $D_{n,\alpha}$ is a positive linear operator and $D_{n,\alpha}(1,x) = 1$. Particularly, when $\alpha = 1$ the operators (2.6) reduce to the operators (2.1).

The paper is concerned with the rate of pointwise convergence of the operators (2.6) when f belong to $X_{loc}[0,\infty)$. By using the Chanturiya modulus of variation we examine the rate of pointwise convergence of $(D_{n,\alpha}f)(x)$ at the points of continuity and at the discontinuity points of the first kind of f.

It is necessary to point out that in the present paper we extend and improve the earlier result of [13] for Chlodovsky-Durrmeyer operators.

At first, we give the following definition.

Definition. *Let* f *be a bounded function on a compact interval* $I = [a,b]$. *The* *modulus of variation* $\nu_n(f;[a,b])$ *of the function* f *is defined for nonnegative integers* n *as follows:*

$$\nu_0(f;[a,b]) = 0$$

and for $n \geq 1$

$$\nu_n(f;[a,b]) = \sup_{\Pi_n} \sum_{k=0}^{n-1} |f(x_{2k+1}) - f(x_{2k})|,$$

where Π_n *is an arbitrary system of* n *disjoint intervals* (x_{2k}, x_{2k+1})*, where* $k = 0, 1, ..., n-1$*, i.e.,* $a \leq x_0 < x_1 \leq x_2 < x_3 ... \leq x_{2n-2} < x_{2n-1} \leq b$.

If $f \in BV_p(I),\ p \geq 1, i.e.$, if f is of bounded *pth* power variation on I, then for every $k \in N$,

$$\nu_k(f;I) \leq k^{1-1/p} V_p(f,I),$$

where $V_p(f,I)$ denotes the total *pth* power variation of f on I, defined as the upper bound of the set of numbers $\left(\sum\limits_j |f(k_j) - f(l_j)|^p\right)^{1/p}$ over all finite systems of non-overlapping intervals $(k_j, l_j) \subset I$. We also consider the class $BV_{loc}^p[0,\infty),\ p \geq 1$, consisting of all functions of bounded pth power variation on every compact interval $I \subset [0,\infty)$.

Theorem 2.1. *Let* $f \in X_{loc}[0,\infty)$ *and assume that the one-sided limits* $f(x+)$, $f(x-)$ *exist at a fixed point* $x \in (0,\infty)$*. Then, for all integers* n *such that* $b_n > 2x$ *and* $4b_n \leq n$ *one has*

$$\left| D_{n,\alpha}(f;x) - \frac{f(x+) + \alpha f(x-)}{\alpha+1} \right| \leq 2\nu_1(g_x; H_x(x\sqrt{b_n/n}))$$

$$+\frac{32\alpha}{x^2}\left(x\left(1-\frac{x}{b_n}\right)+\frac{b_n}{n}\right)\left[\sum_{j=1}^{m-1}\frac{\nu_j(g_x;H_x(jx\sqrt{b_n/n}))}{j^3}+\frac{\nu_m(g_x;H_x(x))}{m^2}\right]$$

$$+\frac{2\alpha c_q}{x^{2q}}\mu(b_n;f)\left(\frac{b_n}{n}\right)^q\left(x\left(1-\frac{x}{b_n}\right)+\frac{b_n}{n}\right)^q+\frac{2\alpha\,|f(x+)-f(x-)|}{\sqrt{\frac{nx}{b_n}\left(1-\frac{x}{b_n}\right)}},$$

where $m := [\sqrt{n/b_n}]$, $H_x(u) = [x-u, x+u]$ *for* $0 \leq u \leq x$, $\mu(b;f) := \sup\limits_{0\leq t\leq b} |f(t)|$,

$$g_x(t) := \begin{cases} f(t) - f(x+) & if \quad t > x, \\ 0 & if \quad t = x, \\ f(t) - f(x-) & if \quad 0 \leq t < x, \end{cases}$$

q *is an arbitrary positive integer and* c_q *is a positive constant depending only on* q.

From Theorem 2.1 we get

Theorem 2.2. *Let $f \in BV^p_{loc}[0,\infty)$, $p \geq 1$, and let $x \in (0,\infty)$. Then, for all integers n such that $b_n > 2x$ and $4b_n \leq n$ we have*

$$\left| D_{n,\alpha}(f;x) - \frac{f(x+) + \alpha f(x-)}{\alpha+1} \right| \leq 2V_p(g_x; H_x(x\sqrt{b_n/n}))$$

$$+\frac{2^{7+1/p}\alpha}{x^2 m^{1+1/p}} \left(x\left(1-\frac{x}{b_n}\right) + \frac{b_n}{n} \right) \sum_{k=1}^{(m+1)^2-1} \frac{V_p(g_x; H_x(\frac{x}{\sqrt{k}}))}{\left(\sqrt{k}\right)^{1-1/p}}$$

$$+\frac{2\alpha c_q}{x^{2q}} \mu(b_n; f) \left(\frac{b_n}{n}\right)^q \left(x\left(1-\frac{x}{b_n}\right) + \frac{b_n}{n}\right)^q + \frac{2\alpha \left| f(x+) - f(x-) \right|}{\sqrt{\frac{nx}{b_n}\left(1-\frac{x}{b_n}\right)}}.$$

So, we get the following approximation theorem.

Corollary 2.3. *Suppose that $f \in X_{loc}[0,\infty)$ (in particular, $f \in BV^p_{loc}[0,\infty)$, $p \geq 1$) and that there exists a positive integer q such that*

$$\lim_{n\to\infty} \left(\frac{b_n}{n}\right)^q \mu(b_n; f) = 0.$$

Then, at every point $x \in (0,\infty)$ at which the limits $f(x+)$, $f(x-)$ exist, we have

$$\lim_{n\to\infty} D_{n,\alpha}(f;x) = \frac{f(x+) + \alpha f(x-)}{\alpha+1}.$$

Obviously, the above relations hold true for every measurable function f bounded on $[0,\infty)$, in particular for every function f of bounded *pth* power variation ($p \geq 1$) on the whole interval $[0,\infty)$.

2010 [10] In this paper, the author investigates convergence and approximation properties of a Chlodovsky type generalization of Stancu polynomials.

2010 [25] The authors estimate the rates of pointwise approximation of certain King-type positive linear operators for functions with derivative of bounded variation. We also extend our results to the statistical approximation process via the concept of statistical convergence.

2010 [5] In this work, they state a Chlodovsky variant of a multivariate beta operator to be called hereafter the multivariate beta-Chlodovsky operator. They show that the multivariate beta-Chlodovsky operator can preserve properties of a general function of modulus of continuity and also the Lipschitz constant of a Lipschitz continuous function.

2010 [12] Another recent result concerning uniform approximation by the Chlodovsky operators is due to A. Holhoş.

2010 [21] Let $J_{n,k}(t)=\sum\limits_{j=k}^{n}p_{j,n}(t)\,,\ t\in[0,1]$, be the Bézier basis functions. For $f\in X_{loc}[0,\infty)$ and $\alpha>0$, the Bézier modification $C_{n,\alpha}f$ of operators (1.1) is defined as

$$C_{n,\alpha}f(x)=\sum_{k=0}^{n}f\left(\frac{kb_n}{n}\right)Q_{n,k}^{(\alpha)}\left(\frac{x}{b_n}\right)\qquad for\quad x\in[0,b_n], \tag{2.7}$$

where $Q_{n,k}^{(\alpha)}(t)=J_{n,k}^{\alpha}(t)-J_{n,k+1}^{\alpha}(t)$ for $t\in[0,1]$ ($J_{n,l}(x)\equiv 0$ if $l>n$).

If $\alpha=1$, then $C_{n,\alpha}f$ reduce to the operators (1.1).

Recently, Karsli and Ibikli [17],[23] gave some estimates for the rates of convergence of operators (1.1) and (2.7) (with $\alpha\geq 1$) for functions $f\in BV[0,\infty)$. In this paper:

1− we essentially improve those estimates,

2− we extend those results to some wider classes of functions, in particular for classes $BV^p[0,\infty)$ with $p>1$,

3− we extend them to all parameters $\alpha>0$.

If $x\in(0,\infty)$, the following intervals $H_x(u):=[x-u,x+u]$ for $0<u\leq x$ will be used.

Theorem 2.4. *Let $f\in X_{loc}[0,\infty)$ and assume that the one-sided limits $f(x+)$, $f(x-)$ exist at a fixed point $x\in(0,\infty)$. Then, for all integers n such that $b_n>2x$ and $n/b_n\geq\max\{4,21/x\}$ we have*

$$\left|C_{n,\alpha}f(x)-\frac{1}{2^\alpha}f(x+)-\left(1-\frac{1}{2^\alpha}\right)f(x-)\right|\leq 2\nu_1(g_x;H_x(x\sqrt{b_n/n}))$$

$$+\frac{16\lambda_\alpha}{x^2}\left(x\left(1-\frac{x}{b_n}\right)+\frac{b_n}{n}\right)\left[\sum_{j=1}^{m-1}\frac{\nu_j(g_x;H_x(jx\sqrt{b_n/n}))}{j^3}+\frac{\nu_m(g_x;H_x(x))}{m^2}\right]$$

$$+\kappa_\alpha\sqrt{\frac{b_n}{n}}\sqrt{\frac{b_n}{x\,(b_n-x)}}\left(|f(x+)-f(x-)|+|f(x)-f(x-)|\,e_n\left(\frac{x}{b_n}\right)\right)$$

$$+4\kappa_\alpha M(b_n;f)\exp\left(-\rho_\alpha\frac{nx}{4b_n}\right),$$

where $m:=[\sqrt{n/b_n}]$, $\kappa_\alpha=\max\{1,\alpha\}\,,\ \rho_\alpha=\min\{1,\alpha\}\,,\ \lambda_\alpha$ is a positive constant depending only on α (if $\alpha\geq 1$ then $\lambda_\alpha=\alpha$), $e_n(x/b_n)=1$ if there exists a $k'\in\{0,1,...,n\}$ such that $nx=k'b_n,\ e_n(x/b_n)=0$ otherwise, $M(b;f):=\sup\limits_{0\leq t\leq b}|f(t)|$.

Here we note that, under the Chlodovsky condition (1.3), Theorem 2.4 is also an approximation theorem. To see this we must verify that the right-hand side of the inequality given in this theorem converges to zero as $n \to \infty$. In view of (1.2) we have $b_n/n \to 0$ and $m = \left[\sqrt{n/b_n}\right] \to \infty$ as $n \to \infty$. Clearly,

$$\frac{\nu_m(g_x; H_x(x))}{m^2} \leq \frac{2}{m} M(2x; f) \to 0 \quad as \quad n \to \infty.$$

Therefore it is enough to consider only the term

$$\Lambda_m(x) := \sum_{j=1}^{m-1} \frac{\nu_j(g_x; H_x(jxd_n))}{j^3} \quad where \quad d_n = \sqrt{b_n/n}.$$

It is easy to see that

$$\begin{aligned}\Lambda_m(x) &\leq \sum_{j=1}^{m-1} \frac{\nu_1(g_x; H_x(jxd_n))}{j^2} \leq 4d_n \int_{d_n}^{md_n} \frac{\nu_1(g_x; H_x(xt))}{t^2} dt \\ &\leq 4d_n \int_1^{m+1} \nu_1\left(g_x; H_x\left(\frac{x}{s}\right)\right) ds \leq \frac{4}{m} \sum_{k=1}^{m} \nu_1\left(g_x; H_x\left(\frac{x}{k}\right)\right).\end{aligned}$$

Since the function g_x is continuous at x and $\nu_1(g_x; H_x(x/k))$ denotes the oscillation of g_x on the interval $[x - x/k, x + x/k]$, we have $\lim_{k\to\infty} \nu_1(g_x; H_x(x/k)) = 0$. Consequently $\lim_{m\to\infty} \Lambda_m(x) = 0$, by the well-known theorem on the limit of the sequence of arithmetic means. Hence we get the following

Corollary 2.5. *Suppose that $f \in X_{loc}[0, \infty)$ and that the Chlodovsky condition (1.3) is satisfied. Then*

$$\lim_{n\to\infty} C_{n,\alpha} f(x) = \frac{1}{2^\alpha} f(x+) + \left(1 - \frac{1}{2^\alpha}\right) f(x-) \tag{2.8}$$

at every point $x \in (0, \infty)$ at which the limits $f(x+)$, $f(x-)$ exist.

Of course, relation (2.8) holds true for every function f bounded on the interval $[0, \infty)$. In particular, if $\alpha = 1$ and x is the point of continuity of f, our Corollary 2.5 coincides with the above mentioned theorem of Chlodovsky.

Retaining the symbols used in Theorem 2.4 we also get

Theorem 2.6. *Let $f \in BV^p_{loc}[0, \infty)$, $p \geq 1$, and let $x \in (0, \infty)$. Then*

$$\left| C_{n,\alpha} f(x) - \frac{1}{2^\alpha} f(x+) - \left(1 - \frac{1}{2^\alpha}\right) f(x-)\right| \leq 2V_p(g_x; H_x(x\sqrt{b_n/n}))$$

$$+\frac{2^{6+1/p}\lambda_\alpha}{x^2 m^{1+1/p}}\left(x\left(1-\frac{x}{b_n}\right)+\frac{b_n}{n}\right)\sum_{k=1}^{(m+1)^2-1}\frac{V_p\left(g_x;H_x\left(\frac{x}{\sqrt{k}}\right)\right)}{\left(\sqrt{k}\right)^{1-1/p}}$$

$$+\kappa_\alpha\sqrt{\frac{b_n}{n}}\sqrt{\frac{b_n}{x\left(b_n-x\right)}}\left(\left|f(x+)-f(x-)\right|+\left|f(x)-f(x-)\right|e_n\left(\frac{x}{b_n}\right)\right)$$

$$+4\kappa_\alpha M(b_n;f)\exp\left(-\rho_\alpha\frac{nx}{4b_n}\right),$$

for all integers n such that $b_n>2x$ and $n/b_n\geq\max\{4,21/x\}$.

It is easy to verify that, in view of continuity of g_x at x,

$$\lim_{m\to\infty}\frac{1}{m^{1+1/p}}\sum_{k=1}^{(m+1)^2-1}\frac{1}{\left(\sqrt{k}\right)^{1-1/p}}V_p\left(g_x;H_x\left(\frac{x}{\sqrt{k}}\right)\right)=0.$$

Hence from Theorem 2.6 we have

Corollary 2.7. *If f belongs to the class $BV^p_{loc}[0,\infty)$, $p\geq 1$, and if it satisfies condition (1.3), the relation (2.8) holds true at every $x\in(0,\infty)$. In particular, (2.8) remains valid for every function f of class $BV^p[0,\infty)$, $p\geq 1$.*

Corollary 2.8. *Let us consider now the special case $p=1$, $\alpha\geq 1$, and let us suppose that $f\in BV[0,\infty)$. Then at every $x>0$ and for all integers n such that $b_n>2x$ and $n/b_n\geq 4$, we have*

$$\left|C_{n,\alpha}f(x)-\frac{1}{2^\alpha}f(x+)-\left(1-\frac{1}{2^\alpha}\right)f(x-)\right|\leq 2V(g_x;H_x(x\sqrt{b_n/n}))$$

$$+\frac{2^9\alpha b_n}{n}\left(\frac{1}{x}-\frac{1}{b_n}\right)\sum_{k=1}^{2[n/b_n]}V\left(g_x;H_x\left(\frac{x}{\sqrt{k}}\right)\right)+4\alpha M\exp\left(-\frac{nx}{4b_n}\right)$$

$$+\alpha\sqrt{\frac{b_n}{n}}\sqrt{\frac{b_n}{x\left(b_n-x\right)}}\left(\left|f(x+)-f(x-)\right|+\left|f(x)-f(x-)\right|e_n\left(\frac{x}{b_n}\right)\right),$$

where $M=\sup\limits_{0\leq x<\infty}|f(x)|$ and $V(g_x;H)$ denotes the Jordan variation of g_x on the interval H.

The above estimate is essentially better than the estimates presented in [17] ($\alpha=1$) and [23] ($\alpha\geq 1$).

Acknowledgement. The author would like to express his sincere thanks to Professor Paul L. Butzer and Professor Paulina Pych-Taberska, who are true experts of the great Aachen and Poznań School, respectively, for their very helpful, careful and intensive studies on Chlodovsky operators. Their exceptional kindness are greatly appreciated by the author.

References

[1] Agratini, O., *On a class of linear positive bivariate operators of King type*, Stud. Univ. Babeş-Bolyai Math., **51**(2006), no. 4, 13-22.

[2] Albrycht, J., Radecki, J., *On a generalization of the theorem of Voronovskaya*, Zeszyty Naukowe UAM, Zeszyt 2, Poznan, 1960, 1-7.

[3] Attalienti, A., Campiti, M., *Bernstein-type operators on the half line*, Czechoslovak Mathematical Journal, **52**(2002), no. 4, 851-860.

[4] Bardaro, C., Butzer, P.L., Stens, R.L., Vinti, G., *Convergence in variation and rates of approximation for Bernstein-type polynomials and singular convolution integrals*, Analysis (Munich), **23**(2003), no. 4, 299-346.

[5] Bascanbaz, G.T., Tuncer, Y., *On a Chlodovsky variant of a multivariate beta operator*, Journal of Comp. and Appl. Math., in press, 2010.

[6] Butzer, P.L., *On the extension of Bernstein polynomials to the infinite interval*, Proc. Amer. Math. Soc., **5**(1954), 547-553.

[7] Butzer, P.L., Karsli, H., *Voronovskaya-type theorems for derivatives of the Bernstein-Chlodovsky polynomials and the Szasz-Mirakyan operator*, Comment. Math., **49**(2009), no. 1, 33-57.

[8] Büyükyazici, I., Ibikli, E., *Inverse theorems for Bernstein-Chlodovsky type polynomials*, J. Math. Kyoto Univ., **46**(2006), no. 1, 21-29.

[9] Büyükyazici, I., *On the approximation properties of two-dimensional q-Bernstein-Chlodovsky polynomials*, Math. Commun., **14**(2009), no. 2, 255-269.

[10] Büyükyazici, I., *Approximation by Stancu-Chlodovsky polynomials*, Comp. Math.Appl., **59**(2010), no. 1.

[11] Chlodovsky I., *Sur le développement des fonctions définies dans un intervalle infini en séries de polynomes de M. S. Bernstein*, Compositio Math., **4**(1937), 380-393.

[12] Holhoş, A., *The rate of approximation of functions in an infinite interval by positive linear operators*, Stud. Univ. Babeş-Bolyai Math., **55**(2010), no. 2, 133-142.

[13] Ibikli, E., Karsli, H., *Rate of convergence of Chlodowsky type Durrmeyer operators*, JIPAM. J. Inequal. Pure Appl. Math., **6**(2005), no. 4, Article 106, 12 pp.

[14] Ispir, N., Aral, A., Doğru, O., *On Kantorovich process of a sequence of the generalized linear positive operators*, Numer. Funct. Anal. Optim., **29**(2008), no. 5-6, 574–589.

[15] Izgi, A., Büyükyazici, I., *On a generalization of Bernstein–Chlodovsky polynomials for two variables*, Int. Math. Forum, **21-24**(2006), 1001–1015.

[16] Izgi, A., Büyükyazici, I., Ibikli, E., *Rate of convergence by Chlodovsky–Taylor polynomials*, Appl. Math. Comp. **213**(2009), no. 2, 426-431.

[17] Karsli, H., Ibikli, E., *Rate of convergence of Chlodovsky type Bernstein operators for functions of bounded variation*, Numer. Func. Anal. Optim., **28**(2007), no. 3-4, 367–378.

[18] Karsli, H., *Order of convergence of Chlodovsky type Durrmeyer operators for functions with derivatives of bounded variation*, Indian J. Pure Appl. Math., **38**(2007), no. 5, 353-363.

[19] Karsli, H., *On convergence of Chlodovsky-type MKZD operators,* Commun. Fac. Sci. Univ. Ank. Ser. A1 Math. Stat., **57**(2008), no. 1, 1-12.

[20] Karsli, H., *Rate of convergence of Chlodovsky operators for functions with derivatives of bounded variation*, Applied Math. e-Notes, **8**(2008), 203-213.

[21] Karsli, H., Pych-Taberska, P., *On the rates of convergence of Bernstein-Chlodovsky polynomials and their Bézier-type variants,* Applicable Analysis, 90, 2011, no. 3-4, 403-416.

[22] Karsli, H., Pych-Taberska, P., *On the rates of convergence of Chlodovsky-Durrmeyer operators and their Bézier variant*, Georgian Math. J., **16**(2009), no. 4, 693–704.

[23] Karsli, H., Ibikli, E., *Convergence rate of a new Bezier variant of Chlodovsky operators to bounded variation functions*, Journal of Comp. and Appl. Math., **212**(2008), no. 2, 431-443.

[24] Karsli, H., Gupta, V., *Some approximation properties of q-Chlodovsky operators*, Appl. Math. Comp., **195**(2008), no. 1, 220-229.

[25] Özarslan, M.A., Duman, O., Kaanoglu, C., *Rates of convergence of certain King-type operators for functions with derivative of bounded variation*, Math. Comp. Modelling, **52**(2010), 334-345.

[26] Pych-Taberska, P., Karsli, H., *On the rates of convergence of Chlodovsky-Kantorovich operators and their Bézier variant*, Comment. Math., **49**(2009), no. 2, 189–208.

[27] Pych-Taberska, P., *Rates of convergence of Chlodovsky-Kantorovich polynomials in classes of locally integrable functions*, Discuss. Math. Differ. Incl. Control Optim., **29**(2009), 53–66.

[28] Stypinski, Z., *Theorem of Voronovskaya for Szász-Chlodovsky operators*, Funct. Approximatio Comment. Math., **1**(1974), 133-137.

[29] Voronovskaya, E., *Détermination de la forme asymptotique dapproximation des functions par polynomes de M. Bernstein*, C. R. Acad. Sci. URSS, **79**(1932), 79-85.

Harun Karsli
Abant Izzet Baysal University
Faculty of Science and Arts
Department of Mathematics
14280, Gölköy-Bolu, Turkey
e-mail: karsli_h@ibu.edu.tr

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 437–447

Multifractional Brownian motion in vehicle crash tests

Diana Keller

Abstract. Different crash tests are carried out in the car industry to measure the acceleration dependent on time. With the aim of improving the airbag-system a discussion of crash processes was raised. Experimental studies approve the modelling of the crash tests as a multifractional Brownian motion which will be introduced as a generalisation of the fractional case (including the Wiener process). Based on the ideas of Coeurjolly [1] an estimation of the significant time-dependent Hurst parameter $H(t)$ will be developed. Its interpretation as a measure of deformation of the crash car leads to interesting results. So the Hurst index' value is important for supporting the fire-decision [4].

Mathematics Subject Classification (2010): 60H05, 60H30, 62P30.

Keywords: Multifractional Brownian motion, Itô integral, Hurst index, vehicle crash tests, airbag-control-model.

1. Motivation of the model

The car industry has performed extensive crash tests for sensitizing and improving the airbag-system. They have measured the acceleration dependent on time with different sensors installed on characteristic positions in the vehicle, especially in the front part of the cars. The activation of the restraint-system is implemented in the airbag-control-unit which is mounted on the middle tunnel. On the basis of mechanical models in a crash situation the airbag-algorithms will be specifically adapted and optimized for each new car. To further improve the accident detection a more general mathematical discussion of the crash process should be conducted.

Currently the crucial criterion for activating the airbags is the velocity calculated by the integral over the acceleration. But these results are not sufficient for a distinction between different crash cases and situations. The aim is to identify the type of crash so that selected airbags will fire only if they are necessary.

The researches are premised on data like in FIGURE 1 whose character changes in time. Here a head-on collision with 56 km/h against a solid wall is presented. There arises the question whether crash test situations suffice a stochastic process. This assumption can be affirmed because the progress of acceleration is significant: wild fluctuations at the beginning which rapidly decrease after 50 ms. These fluctuations can be described by the fractional Brownian motion with a Hurst index H greater than 0 but less than 1/2. If H converges to 1 the fractional Brownian motion will tend to a random variable. This supports the interpretation of the crash process as a multifractional Brownian motion with a time-dependent Hurst parameter $H(t)$.

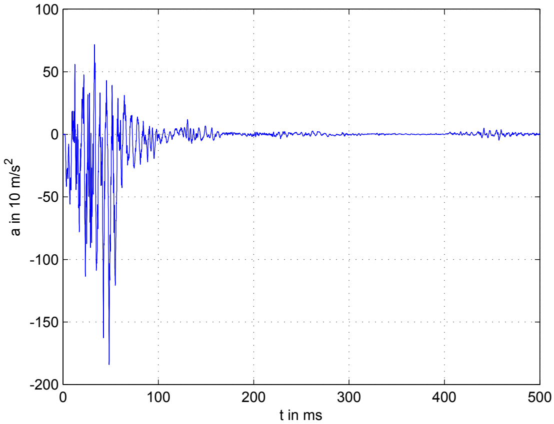


FIGURE 1. Head-on collision with 56 km/h against a solid wall

2. The multifractional Brownian motion

2.1. Definition and representation

First the fractional Brownian motion will be defined as a Brownian motion with a constant parameter H:

Definition 2.1. *A real-valued random process $(B_H(t), t \geq 0)$ is called fractional Brownian motion with Hurst parameter $H \in (0,1)$ provided that*

(i) $B_H(t)$ *is a Gaussian process;*
(ii) $B_H(0) = 0$ *a.s.;*
(iii) $I\!E(B_H(t)) = 0,\ \forall\, t \geq 0$, *that means the process is centered;*
(iv) $I\!E(B_H(t)B_H(s)) = \frac{1}{2}\,\mathrm{Var}\,(B_H(1))\left[|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right]$.

Especially the case $H = 0.5$ leads to the Brownian motion also known as Wiener process [4]. A generalisation of the fractional Brownian motion is the multifractional Brownian motion where the constant Hurst index H will be substituted by a time-dependent Hurst exponent $H(t)$:

Definition 2.2. *A real-valued random process* $(B_{H_t}(t), t \geq 0)$ *is said to be a multifractional Brownian motion if the following conditions are fulfilled*

(i) $B_{H_t}(t)$ *is a Gaussian process;*
(ii) $B_{H_0}(0) = 0$ *a.s.;*
(iii) $I\!\!E(B_{H_t}(t)) = 0,\ \forall\, t \geq 0,$ *that means the process is centered;*
(iv) $I\!\!E(B_{H_t}(t) B_{H_s}(s)) = \frac{1}{2}\, C(H_t, H_s) \left[|t|^{H_t+H_s} + |s|^{H_t+H_s} - |t-s|^{H_t+H_s} \right],$ *with* $C(H_t, H_s) =$ *const. dependent on* H_t *and* H_s*;*
(v) $H : [0,\infty) \mapsto (0,1)$ *is Hölder continuous with exponent* $\beta > 0$.

This definition of the multifractional case is equivalent to a representation as an Itô integral [5]

$$B_{H_t}(t) = \frac{1}{\Gamma(H_t + \frac{1}{2})} \left\{ \int_{-\infty}^{0} \left[(t-s)^{H_t - \frac{1}{2}} - (-s)^{H_t - \frac{1}{2}} \right] dB(s) + \int_{0}^{t} (t-s)^{H_t - \frac{1}{2}}\, dB(s) \right\}$$

for all $t \geq 0$ where $H : [0,\infty) \mapsto (0,1)$ is a Hölder continuous function with exponent $\beta > 0$ and B marks the ordinary two-sided Brownian motion.

A process $\left(B(t),\, t \in \mathbb{R}^1\right)$ denotes a two-sided Brownian motion if

$$B(t) = \begin{cases} B_1(t) & : \text{for } t \geq 0, \\ B_2(-t) & : \text{for } t < 0, \end{cases}$$

where $B_1(t)$ and $B_2(t)$ are two independent Brownian motions for $t \geq 0$.

2.2. Typical properties

Because of zero mean and the Itô isometry [3] of the stochastic integral all the properties listed in Definition 2.2 can be proved from the equivalent integral representation, explicitly shown in [4]. Furthermore two important theorems will be presented but not proved, only the main idea will be mentioned.

Theorem 2.3. *The multifractional Brownian motion* $B_{H_t}(t)$ *is a continuous process for all* $t \in [0,\infty)$ *with probability* 1.

It is possible to show this with the help of skilful splittings of the Itô integral representation, some fundamental inequalities and the Kolmogorov criterion [5], detailed in [4].

Theorem 2.4. *It exists a positive continuous function* $t \mapsto \sigma_t$ *so that for all* $t \geq 0$ *the following asymptotic distribution holds*

$$\frac{B_{H_{t+h}}(t+h) - B_{H_t}(t)}{h^{H_t}} \xrightarrow[h \to 0]{\mathcal{L}} N(0, \sigma_t^2).$$

Evidently the mean is 0 but the variance is harder to predict. Again skilful splittings and useful inequalities yield the result [5], explicitly in [4]. Hence a standard multifractional Brownian motion can be introduced.

2.3. Hurst index' estimation

An estimation of the significant time-dependent Hurst parameter $H(t)$ is based on the ideas of Coeurjolly [1], [2]. It is a kind of parameter estimator harking back to the asymptotic behaviour of the k-th absolute moment. Here $k \leq 2$ is considered. A particularity is that only one realisation is necessary for the estimation which actually is a well-known method for the fractional case with constant H. First the raw data have to be filtered, here with the so called Daubechies-filter. Then the procedure will be extended from the fractional Brownian motion to the multifractional one. That means the estimation does not happen over the entire time range, but rather over a defined time period so that a time-dependent $H(t)$ will be obtained (see also in the next chapter).

With the help of the trajectory filtered by the Daubechies-filter a of length $l+1$ (in detail [1], [2])

$$V^a\left(\frac{i}{n}\right) = \sum_{q=0}^{l} a_q B_H\left(\frac{i-q}{n}\right), \qquad \text{for } i = l, \dots, n-1,$$

the covariance function π_H^a of this series will be calculated by

$$\pi_H^a(j) = I\!\!E\left(V^a\left(\frac{i}{n}\right) V^a\left(\frac{i+j}{n}\right)\right) = -\frac{1}{2}\sum_{q,r=0}^{l} a_q a_r \left|q-r+j\right|^{2H}.$$

The k-th empirical absolute moment of the discrete variations of the fractional Brownian motion has the following representation

$$S_n(k,a) = \frac{1}{n-l}\sum_{i=l}^{n-1}\left|V^a\left(\frac{i}{n}\right)\right|^k.$$

Finally Coeurjolly estimates the Hurst parameter H by

$$\hat{H}_n(k,a) = g_{k,a,n}^{-1}\left(S_n(k,a)\right),$$

where the function $g_{k,a,n}^{-1}(t)$ is defined as the inverse of

$$g_{k,a,n}(t) = \frac{1}{n^{kt}}\{\pi_t^a(0)\}^{\frac{k}{2}} E_k$$

and the indices k, a and n denote the order of the moment, the filter and the number of partition points. The factor E_k depends on the used order k of the moment and is explained by

$$E_k = 2^{\frac{k}{2}}\Gamma\left(k+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right).$$

3. Crash test analysis

3.1. Application of the estimation

Experimental studies have shown that the Hurst index depends on time. Figure 1 represents the acceleration measured over 500 ms in 10.000 data points. That means for 1 ms 20 data points are available. But if the airbags are necessary to protect the inmates they have to fire empirically by no later than 25 ms. So it suffices to consider only the first 500 measured points.

Now the described method to estimate the Hurst index $H(t)$ can be applied using the Daubechies-filter of order 6 and a time period of 10 ms containing 200 data points. Practically the first approximation of H results from considering the interval $(1, 200)$. Then all intervals from $(2, 201)$ to $(301, 500)$ will be examined. Because the fire-decision is usually made after 25 ms there are 15 ms available for interpretation.

The Hurst parameter is a measure of deformation of the crash car with a small H corresponding to a big deformation and a big one to a small deformation. Please note $0 < H < 1$. If the passenger cabin is affected by deformation there will be a high risk of injury for the occupants. That is why the activation of the airbags is essential.

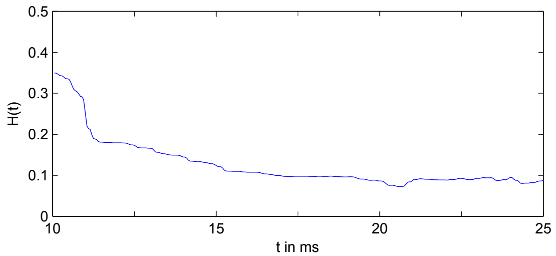


Figure 2. Corresponding Hurst parameter to head-on collision

The corresponding Hurst index to the head-on collision in Figure 1 is illustrated in Figure 2, the first estimation after 10 ms and the last one after 25 ms. With small values of $H(t)$ the airbags have to activate because a big deformation is associated and the inmates are in jeopardy.

3.2. Introduction and evaluation of the test cases

Four different crash cases depicted in Figure 3 were investigated. The first one is the head-on collision against a solid wall with velocities between 16 and 56 km/h. This crash situation will be abbreviated with *frontal*. In the picture at the top on the right a car is overlapping a barrier by only 40 %. The barrier is a deformable obstacle (that is where the name *deform* comes from) and the car collides with the obstacle with 40 to 64 km/h. The third one is called *angle10* and illustrates the crash with only 15 km/h against a solid wall at an angle of 10 degrees and a 40 % overlap. Finally at the bottom on the right there is a collision against a solid wall at an angle of 30 degrees with velocities of 32 or 40 km/h which will be abbreviated with *angle30*.

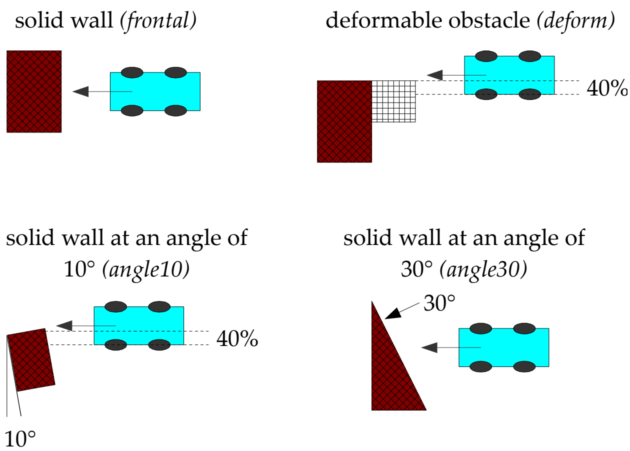


FIGURE 3. Distinction between crash cases

As a measure of deformation of the crash car the Hurst index will be considered for each situation and velocity. This leads to very interesting results. FIGURE 4 shows the Hurst parameters for some selected cases estimated with the method above using the Daubechies-filter of order 6 and a time period of 10 ms realised in 200 data points.

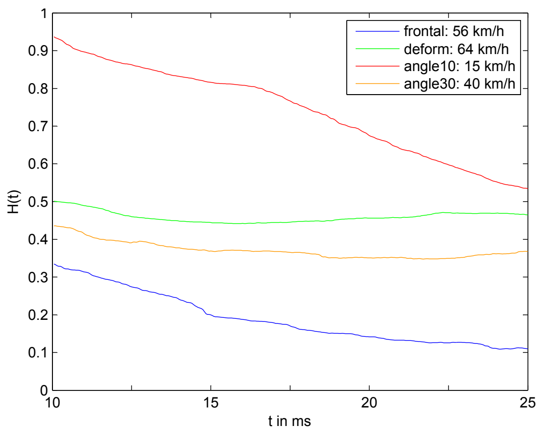


FIGURE 4. Hurst parameters for selected cases

The red line at the top represents a collision at an angle of 10 degrees and 15 km/h against a solid wall (overlapping 40 %). With a big monotonically decreasing Hurst index between 1 and 0.5 the deformation of the car body is very small. There is only an almost unnoticeable danger for the inmates and therefore the airbags are unnecessary. It is the biggest Hurst index of the four observed cases in FIGURE 4, thus the lowest damage. In consideration of a velocity of only 15 km/h this result is easily comprehensible.

Beneath, the crash case with the deformable obstacle proceeds nearly constantly at 0.5 and the velocity of 64 km/h suggests the use of the airbags. It is the biggest test velocity and a huge deformation is accompanied by a high risk of injury for the vehicle occupants. To grant the best possible protection the airbags have to fire.

The orange Hurst index belongs to a car which collides with a solid wall at an angle of 30 degrees and a velocity of 40 km/h. The car slides along the wall because of the angle of contingence. With values of about 0.4 the Hurst parameter is smaller than in the previous cases. That means the deformation is greater due to the rough impact. So the airbags are essential because of the imminent danger.

Last but not least the blue line characterises a head-on collision against a solid wall with 56 km/h. Monotonically decreasing values between 0.35 and 0.1 illustrate the crash situation with the smallest Hurst index. Hence the biggest deformation of the vehicle takes place and the occupants could be seriously injured. Such a head-on collision can entail severe consequences and therefore require the airbags to be deployed.

In FIGURE 4 two of the curves are monotonically decreasing while the other two are nearly constant. Perhaps more information to support the fire-decision are conceivable by use of the monotonicity of the trajectories. Moreover the estimation of the Hurst index in the case *angle10* is much greater than in all the other cases. The airbags do not have to fire because there is only a small deformation contrary to the three other cases. That is why the airbags are necessary to guarantee the safety of the passengers.

Looking at the mentioned figure a boundary at about 0.5 seperating the case with airbags from these without can be supposed. This boundary is well-motivated since $H = 0.5$ forms the characteristic change between wild fluctuations of the acceleration and the levelling values which tend to a random variable. The special case $H = 0.5$ realises the Brownian motion.

3.3. Further results in detail

Considering the four presented crash situations and averaging over the Hurst parameters of these crashes with the same case and the same velocity there are the following outcomes.

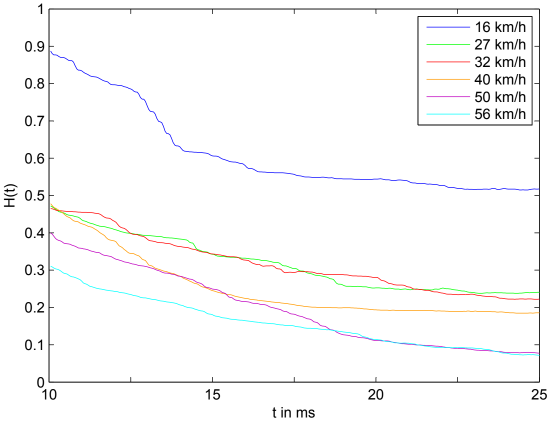


FIGURE 5. Hurst parameter for the case *frontal*

In FIGURE 5 the estimation on top (head-on collision with a velocity of 16 km/h) differs with values greater than 0.5 from all the other velocities. A big Hurst index is interpreted as a small deformation of the car body and a small risk for the occupants. That is why the airbags are unnecessary. This result is very catchy because a velocity of 16 km/h is so slow that big damages are unbelievable. But the tests with all the other velocities show with values less than 0.5 that the deformation is getting greater and so the risk of injury is growing. The airbags have to activate to protect the inmates optimally.

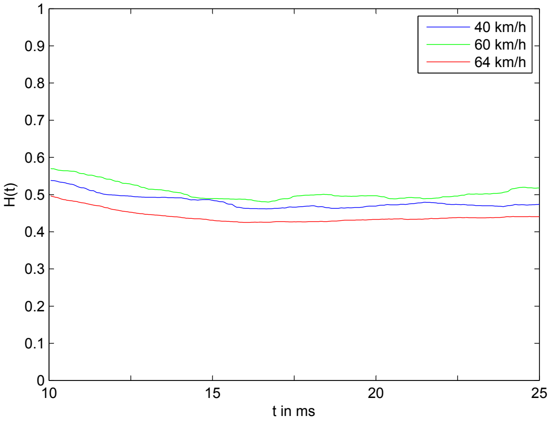


FIGURE 6. Hurst parameter for the case *deform*

The estimations of the case *deform* are close together and their progress is nearly identically. But it is conspicuous that the Hurst parameter is decreasing with growing velocities. That means the deformation keeps on entering into the passenger cabin and the occupants are increasingly threatened. To give maximum shelter to the inmates the use of the airbags is essential at all presented velocities.

Nevertheless Figure 6 requires to raise the boundary between the cases with and without airbags from 0.5 to 0.6 because the collision with 40 and 60 km/h against a deformable obstacle - where the activation of the airbags can not be abandoned - have Hurst parameters just under 0.6. Such an enlargement does not contradict all the previous figures since in all crashes with a lower Hurst index the airbags have to fire and in all crashes with a greater Hurst parameter the airbags are not necessary.

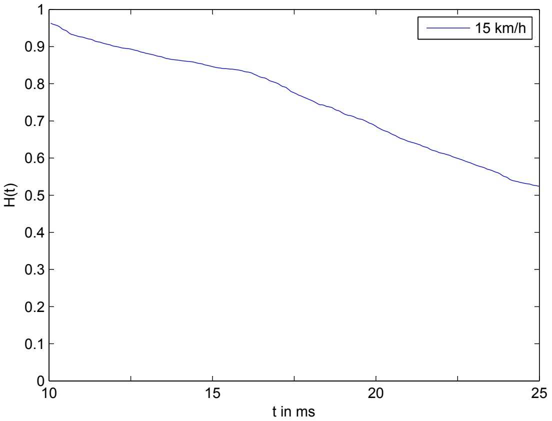


Figure 7. Hurst parameter for the case *angle10*

The trajectory of the estimated Hurst index of the case *angle10* with 15 km/h in Figure 7 is the same as in Figure 4 because there were no other velocities to analyse.

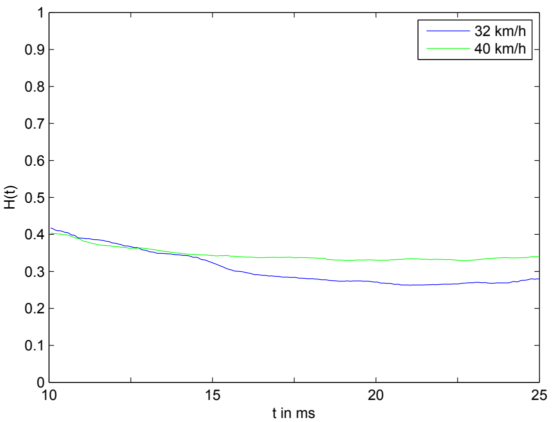


Figure 8. Hurst parameter for the case *angle30*

Finally the case *angle30* is mapped in FIGURE 8 whose curves are similar to the estimations of the case *frontal*. If the case is unknown one can interchange them. But considering the known velocities the distinction is easier since the estimations of the case *frontal* start with greater values at about 0.5 and finish with lower values at about 0.2 after 25 ms, the moment the fire-decision has to be made.

There exists a characteristic estimation for the Hurst index in each crash situation so that on the one hand different crash cases and situations can be distinguished due to progress and dimension of $H(t)$ and on the other hand there are some similarities. Referring to the averages of crash cases with the same situation and the same velocity a strict boundary at about 0.6 is recognisable - a boundary between cases where the airbags have to fire and those where they are unnecessary. All these results are heuristically and have to be tested with more data to cover a bigger spectrum of crash cases and velocities.

One difficulty in all well-known methods of the past was to differentiate the case *deform* from the case *angle10*. Now a distinction between these two cases is obvious. It is harder to differ between the cases *frontal* and *angle30*. Perhaps a symbiosis of old and new methods is promising.

4. Conclusion

In sum, the Hurst index' value is important for supporting the fire-decision. It exists a characteristic estimation of the Hurst parameter in progression and dimension for each crash situation so that a strict distinction is possible. In certain circumstances only special airbags have to fire. With huge values of $H(t)$ the collision at an angle of 10 degrees - requiring no activation of the airbags - contrasts with all the other cases with Hurst parameters less than 0.6. The airbags are essential for the security of the inmates. All in all there is a distinct boundary at about 0.6 between non-activating and activating the airbags. But this is only an assumption, perhaps this boundary has to be corrected by investigating more statistical series, other crash cases and velocities.

A boundary of 0.5 would be motivated very well because $H = 0.5$ is the characteristic change between wild fluctuations and the levelling values of the acceleration which tend to a random variable. It is the special case of the well-known Brownian motion.

An interesting question arises: Is it possible to make the fire-decision based only on the knowledge of the estimated Hurst index? This would be a very great result but requires any more researches.

References

[1] Coeurjolly, J.-F., *Estimating the Parameters of a Fractional Brownian motion by Discrete Variations of its Sample Paths*, Statistical Inference for Stochastic Processes, **4**(2001), 199-227.

[2] Coeurjolly, J.-F., *Identification of Multifractional Brownian Motion*, 2004.

[3] Deck, T., *Der Itô-Kalkül: Einführung und Anwendungen*, Springer, 2006.

[4] Keller, D., *Zur multifraktalen Brownschen Bewegung: Modellierung und Anwendungen*, Bachelor Thesis, Martin-Luther-University Halle-Wittenberg, 2009.

[5] Peltier, R.F., Lévy Véhel, J., *Multifractional Brownian motion: definition and preliminary results*, INRIA, no. 2645, 1995.

Diana Keller
Martin-Luther-University Halle-Wittenberg
Faculty of Natural Sciences II
Institute for Mathematics
Theodor-Lieser-Straße 5
06120 Halle (Saale)
Germany
e-mail: `diana.math@web.de`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 449–464

Numerical quadratures and orthogonal polynomials

Gradimir V. Milovanović

Abstract. Orthogonal polynomials of different kinds as the basic tools play very important role in construction and analysis of quadrature formulas of maximal and nearly maximal algebraic degree of exactness. In this survey paper we give an account on some important connections between orthogonal polynomials and Gaussian quadratures, as well as several types of generalized orthogonal polynomials and corresponding types of quadratures with simple and multiple nodes. Also, we give some new results on a direct connection of generalized Birkhoff-Young quadratures for analytic functions in the complex plane with multiple orthogonal polynomials.

Mathematics Subject Classification (2010): 33C45, 41A55, 65D30, 65D32.

Keywords: Quadrature formula, node, weight, maximal degree of exactness, orthogonal polynomial, quasi-orthogonal polynomial, s-orthogonal polynomial, σ-orthogonal polynomial, multiple orthogonal polynomial.

1. Introduction

Let $\mathcal{P}_n$ be the set of all algebraic polynomials of degree at most n and $d\sigma$ be a finite positive Borel measure on the real line $\mathbb{R}$ such that its support $\operatorname{supp}(d\sigma)$ is an infinite set, and all its moments $\mu_k = \int_{\mathbb{R}} t^k \, d\sigma(t)$, $k = 0, 1, \ldots$, exist and are finite.

The n-point quadrature formula

$$\int_{\mathbb{R}} f(t) d\sigma(t) = \sum_{k=1}^{n} \sigma_k f(\tau_k) + R_n(f), \tag{1.1}$$

which is exact on the set $\mathcal{P}_{2n-1}$ is known as the Gauss-Christofell quadrature formula (cf. [14, p. 29], [20, p. 324]). It is a quadrature formula of the maximal algebraic degree of exactness, i.e., $R_n(\mathcal{P}_{d_{\max}}) = 0$, where $d_{\max} = 2n - 1$.

This famous method of numerical integration, for the Legendre measure $d\sigma(t) = dt$ on $[-1, 1]$, was discovered in 1814 by C.F. Gauss [11], using his theory of continued fractions associated with hypergeometric series. It is interesting to mention that Gauss determined numerical values of quadrature parameters, the nodes τ_k and the weights σ_k, $k = 1, \ldots, n$, for all $n \le 7$. An elegant alternative derivation of this method was provided by Jacobi, and a significant generalization to arbitrary measures was given by Christoffel. The error term $R_n(f)$ and convergence were proved by Markov and Stieltjes, respectively. A nice survey of Gauss-Christoffel quadrature formulae was written by Gautschi [12].

In this survey paper we give an account on some important connections between orthogonal polynomials and Gaussian quadratures, as well as several types of generalized orthogonal polynomials and corresponding types of quadratures. The paper is organized as follows. Section 2 is devoted to quadratures of Gaussian type (with maximal or nearly maximal degree of exactness) and quasi-orthogonal polynomials. A connection between s- and *σ-orthogonal polynomials* and quadratures with multiple nodes is presented in Section 3. Finally, in Section 4 we consider the so-called *multiple orthogonal polynomials* and give two applications. First, we show a direct connection of Borges quadratures [3] with multiple orthogonal polynomial. Second application is related to a generalization of the Birkhoff-Young quadratures [2] for analytic functions in the complex plane. We give a characterization of such generalized quadratures in terms of multiple orthogonal polynomials and prove the existence and uniqueness of these quadratures.

2. Orthogonal and quasi-orthogonal polynomials and Gaussian type of quadratures

The construction of quadrature formulae of the maximal (Gauss-Christoffel), or nearly maximal, algebraic degree of exactness for integrals involving a positive measure $d\sigma$ is closely connected to polynomials orthogonal on the real line with respect to the inner product

$$(f, g) = (f, g)_{d\sigma} = \int_{\mathbb{R}} f(t)g(t)\, d\sigma(t) \quad (f, g \in L^2(d\sigma)). \tag{2.1}$$

The monic polynomials $\pi_\nu = \pi_\nu(d\sigma; \cdot\,)$, $\nu = 0, 1, \ldots$, orthogonal with respect to (2.1) satisfy the three-term recurrence relation (cf. [20, p. 97])

$$\begin{aligned} \pi_{\nu+1}(t) &= (t - \alpha_\nu)\pi_\nu(t) - \beta_\nu \pi_{\nu-1}(t), \quad \nu = 0, 1, \ldots, \\ \pi_0(t) &= 1,\ \pi_{-1}(t) = 0, \end{aligned} \tag{2.2}$$

with recurrence coefficients $\alpha_\nu = \alpha_\nu(d\sigma)$ and $\beta_\nu = \beta_\nu(d\sigma) > 0$, and $\beta_0 = \mu_0 = \int_{\mathbb{R}} d\sigma(t)$ (by definition).

The following theorem is due to Jacobi (cf. [20, p. 322]):

Theorem 2.1. *Given a positive integer m ($\le n$), the quadrature formula* (1.1) *has degree of exactness $d = n - 1 + m$ if and only if the following conditions are satisfied:*

$1°$ *Formula* (1.1) *is interpolatory;*

2° *The node polynomial* $q_n(t) = (t - \tau_1)(t - \tau_2) \cdots (t - \tau_n)$ *satisfies*

$$(\forall p \in \mathcal{P}_{m-1}) \qquad (p, q_n) = \int_{\mathbb{R}} p(t) q_n(t)\, d\sigma(x) = 0.$$

According to this theorem, an n-point quadrature formula (1.1) has the maximal degree of exactness $2n - 1$, i.e., $m = n$ is optimal, because the higher m ($> n$) is impossible. Namely, the condition 2° in Theorem 2.1 for $m = n + 1$ requires the orthogonality $(p, q_n) = 0$ for all $p \in \mathcal{P}_n$, which is impossible when $p = q_n$.

Thus, in the case $m = n$, the orthogonality condition 2° from Theorem 2.1 shows that the node polynomial q_n must be (monic) orthogonal polynomial with respect to the measure $d\sigma$, and therefore the nodes τ_k must be zeros of the polynomial $q_n(t) = \pi_n(d\sigma; t)$. The corresponding weights σ_k (Christoffel numbers) can be expressed in terms of orthogonal polynomials as values of the Christoffel function $\lambda_n(d\sigma; t)$ at these zeros (cf. [20, p. 324]).

Computationally, today there are very stable methods for generating Gauss-Christoffel rules. The most popular of them is one due to Golub and Welsch [18]. Their method is based on determining the eigenvalues and the first components of the eigenvectors of a symmetric tridiagonal Jacobi matrix $J_n(d\sigma)$, with elements formed from the coefficients in the three-term recurrence relation (2.2).

Theorem 2.2. *The nodes* τ_k *in the Gauss-Christoffel quadrature rule* (1.1), *with respect to a positive measure* $d\sigma$, *are the eigenvalues of the* n*-th order Jacobi matrix*

$$J_n(d\sigma) = \begin{bmatrix} \alpha_0 & \sqrt{\beta_1} & & & \mathrm{O} \\ \sqrt{\beta_1} & \alpha_1 & \sqrt{\beta_2} & & \\ & \sqrt{\beta_2} & \alpha_2 & \ddots & \\ & & \ddots & \ddots & \sqrt{\beta_{n-1}} \\ \mathrm{O} & & & \sqrt{\beta_{n-1}} & \alpha_{n-1} \end{bmatrix},$$

where α_ν *and* β_ν, $\nu = 0, 1, \ldots, n-1$, *are the coefficients in the three-term recurrence relation for the monic orthogonal polynomials* $\pi_\nu(d\sigma;\, \cdot\,)$, *and the weights* σ_k *are given by*

$$\sigma_k = \beta_0 v_{k,1}^2, \qquad k = 1, \ldots, n,$$

where $\beta_0 = \mu_0 = \int_{\mathbb{R}} d\sigma(t)$ *and* $v_{k,1}$ *is the first component of the normalized eigenvector* $\mathbf{v}_k$ *corresponding to the eigenvalue* τ_k,

$$J_n(d\sigma)\mathbf{v}_k = \tau_k \mathbf{v}_k, \qquad \mathbf{v}_k^{\mathrm{T}} \mathbf{v}_k = 1, \qquad k = 1, \ldots, n.$$

If we put a smaller value of m, say $m = n - r$, in Theorem 2.1, the node polynomial can be expressed in terms of orthogonal polynomials π_ν as

$$q_n(t) = q_{n,r}(t) = \pi_n(t) + \varrho_1 \pi_{n-1}(t) + \cdots + \varrho_r \pi_{n-r}(t), \tag{2.3}$$

where $\varrho_1, \ldots, \varrho_r$ are real numbers and $n > r$. For $r = 0$ we put $q_{n,0} = \pi_n$.

Such polynomials $\{q_{n,r}\}$ are known as *quasi-orthogonal* polynomials and they play very important role in the study of interpolatory quadratures with exactness $d = 2n - r - 1$, $0 \le r < n$. Notice that for $r = n$, i.e., $m = 0$, the quadrature (1.1) is only interpolatory, without the orthogonality condition 2° in Theorem 2.1.

It is clear if τ_k, $k = 1, \ldots, n$, are nodes of the quadrature formula (1.1), with exactness $d = 2n - r - 1$, then these nodes are zeros of a quasi-polynomial of the form (2.3). Contrary, if a quasi-polynomial $q_{n,r}$ has n real distinct zeros τ_k, $k = 1, \ldots, n$, then there exists a quadrature rule of the form (1.1), with exactness $d = 2n - r - 1$ and non-zero weights σ_k, $k = 1, \ldots, n$. Such kind of quadratures have been studied by several authors (cf. [4, 5, 10, 21, 43]). Quadratures with positive weigts are of particular interest and they are known as *positive quadrature formulae*. Their convergence and some characterizations were studied by several authors (cf. [10, 26, 27, 44]). For example, Xu [44] showed that the quasi-orthogonal polynomials that lead to the positive quadratures can all be expressed as characteristic polynomials of a symmetric tridiagonal matrix with positive subdiagonal entries. Also, as a consequence, for a fixed n, Xu [44] obtained that every positive quadrature is a Gaussian quadrature formula for some another nonnegative measure.

Positive quadrature formulas on the real line with the highest degree of exactness and with one or two prescribed nodes anywhere on the interval of integration have been recently characterized in [5]. The simplest kinds of such formulas are well known Gauss-Radau and Gauss-Lobatto quadratures with one or both (finite) endpoints being fixed nodes, respectively (cf. [20, p. 328]). Their nodes and weights can be obtained by a little modification of the Golub-Welsch Theorem 2.2. Also, some cases with one or two additional prescribed nodes inside the interval of integration can be analyzed by considering certain modified Jacobi matrices (see [5]).

3. Power orthogonality and quadrature with multiple nodes

The first idea of numerical integration involving multiple nodes appeared in the middle of the last century (Chakalov [6, 7, 8], Turán [40], Popoviciu [28], Ghizzetti and Ossicini [15, 16], etc.).

Let $\eta_1, \ldots, \eta_m$ ($\eta_1 < \cdots < \eta_m$) be given *fixed* (or *prescribed*) nodes, with multiplicities $m_1, \ldots, m_m$, respectively, and $\tau_1, \ldots, \tau_n$ ($\tau_1 < \cdots < \tau_n$) be *free* nodes, with given multiplicities $n_1, \ldots, n_n$, respectively. Interpolation quadrature formulae of a general form

$$I(f) = \int_{\mathbb{R}} f(t)\, d\sigma(t) \cong \sum_{\nu=1}^{n} \sum_{i=0}^{n_\nu - 1} A_{i,\nu} f^{(i)}(\tau_\nu) + \sum_{\nu=1}^{m} \sum_{i=0}^{m_\nu - 1} B_{i,\nu} f^{(i)}(\eta_\nu), \quad (3.1)$$

with an algebraic degree of exactness at least $M + N - 1$, were investigated by Stancu [31, 35, 38].

Using fixed and free nodes we introduce two polynomials

$$q_M(t) \colon = \prod_{\nu=1}^{m} (t - \eta_\nu)^{m_\nu} \quad \text{and} \quad Q_N(t) \colon = \prod_{\nu=1}^{n} (t - \tau_\nu)^{n_\nu},$$

where $M=\sum_{\nu=1}^{m} m_\nu$ and $N=\sum_{\nu=1}^{n} n_\nu$. Choosing the free nodes to increase the degree of exactness leads to the so-called Gaussian type of quadratures. If the free (or *Gaussian*) nodes $\tau_1,\ldots,\tau_n$ are such that the quadrature rule (3.1) is exact for each $f\in\mathcal{P}_{M+N+n-1}$, then we call it the *Gauss-Stancu formula*. Stancu [36] proved that $\tau_1,\ldots,\tau_n$ *are the Gaussian nodes if and only if*

$$\int_{\mathbb{R}} t^k Q_N(t) q_M(t)\, d\sigma(t) = 0, \quad k=0,1,\ldots,n-1. \tag{3.2}$$

Under some restrictions of node polynomials $q_M(t)$ and $Q_N(t)$ on the support interval of the measure $d\sigma(t)$ we can give sufficient conditions for the existence of Gaussian nodes (cf. Stancu [36] and [17]). For example, *if the multiplicities of the Gaussian nodes are odd, e.g., $n_\nu = 2s_\nu+1$, $\nu=1,\ldots,n$, and if the polynomial with fixed nodes $q_M(t)$ does not change its sign in the support interval of the measure $d\sigma(t)$, then, in this interval, there exist real distinct nodes τ_ν, $\nu=1,\ldots,n$.*

The last condition for the polynomial $q_M(t)$ means that the multiplicities of the internal fixed nodes must be even. Defining a new (nonnegative) measure $d\hat\sigma(t) := |q_M(t)|\, d\sigma(t)$, the "orthogonality conditions" (3.2) can be expressed in a simpler form

$$\int_{\mathbb{R}} t^k Q_N(t)\, d\hat\sigma(t) = 0, \quad k=0,1,\ldots,n-1.$$

This means that the general quadrature problem (3.1), under these conditions, can be reduced to a problem with only Gaussian nodes, but with respect to another modified measure. Computational methods for this purpose are based on Christoffel's theorem and described in details in [13] (see also [17]).

Let $\pi_n(t) \colon = \prod_{\nu=1}^{n}(t-\tau_\nu)$. Since $Q_N(t)/\pi_n(t) = \prod_{\nu=1}^{n}(t-\tau_\nu)^{2s_\nu} \ge 0$ over the support interval, we can make an additional reinterpretation of the "orthogonality conditions" (3.2) in the form

$$\int_{\mathbb{R}} t^k \pi_n(t)\, d\mu(t) = 0, \quad k=0,1,\ldots,n-1, \tag{3.3}$$

where

$$d\mu(t) = \left(\prod_{\nu=1}^{n}(t-\tau_\nu)^{2s_\nu}\right) d\hat\sigma(t). \tag{3.4}$$

This means that $\pi_n(t)$ is a polynomial orthogonal with respect to the new nonnegative measure $d\mu(t)$ and, therefore, all zeros $\tau_1,\ldots,\tau_n$ are simple, real, and belong to the support interval. As we see the measure $d\mu(t)$ involves the nodes $\tau_1,\ldots,\tau_n$, i.e., the unknown polynomial $\pi_n(t)$, which is implicitly defined. This polynomial $\pi_n(t)$ belongs to the class of the so-called *σ-orthogonal polynomials* $\{\pi_{n,\sigma}(t)\}_{n\in\mathbb{N}_0}$, which correspond to the sequence $\sigma=(s_1,s_2,\ldots)$ connected with multiplicities of Gaussian nodes. Namely, the solution $(\hat\tau_1,\ldots,\hat\tau_n)$ of the previous (nonlinear) system of equations (3.3) gives the σ-orthogonal polynomial

$$\pi_n(t) = \pi_{n,\sigma}(t) = (t-\hat\tau_1)\cdots(t-\hat\tau_n),$$

which is also the unique solution of the extremal problem

$$\min_{\tau_1<\cdots<\tau_n} \int_{\mathbb{R}} |t-\tau_1|^{2s_1+2}\cdots|t-\tau_n|^{2s_n+2}d\hat{\sigma}(t) = \int_{\mathbb{R}} |\pi_{n,\sigma}(t)|^2 d\hat{\mu}(t), \quad (3.5)$$

where $d\hat{\mu}$ is of the form (3.4) with $\hat{\tau}_\nu$ instead of τ_ν, $\nu = 1,\dots,n$.

If $\sigma = (s, s, \dots)$, these polynomials reduce to the *s-orthogonal polynomials* and the corresponding extremal problem (3.5) becomes

$$\min_{p\in\mathcal{P}_{n-1}} \int_{\mathbb{R}} |t^n + p(t)|^{2s+2}d\hat{\sigma}(t) = \int_{\mathbb{R}} |\pi_n(t)|^2 d\hat{\mu}(t) = \|\pi_n\|^2_{d\hat{\mu}},$$

where $d\hat{\mu}(t) = \pi_n(t)^{2s}d\hat{\sigma}(t)$. (For details see Milovanović [22].)

Quadratures with only Gaussian nodes ($m = 0$),

$$\int_{\mathbb{R}} f(t)\,d\sigma(t) = \sum_{\nu=1}^{n}\sum_{i=0}^{2s_\nu} A_{i,\nu} f^{(i)}(\tau_\nu) + R(f),$$

which are exact for all algebraic polynomials of degree at most $d_{\max} = 2\sum_{\nu=1}^{n} s_\nu + 2n - 1$, are known as *Chakalov-Popoviciu quadrature formulas* (see [6, 7, 8], [28]). A deep theoretical progress in this subject was made by Stancu (see [38] and [32]–[37]). In the special case of the Legendre measure on $[-1, 1]$, when all multiplicities are mutually equal, these formulas reduce to the well-known *Turán quadrature* [40]. A connection between quadratures, s and σ-orthogonality and moment-preserving approximation with defective splines was given in survey paper [22]. A very efficient method for constructing quadratures with multiple nodes was given recently by Milovanović, Spalević and Cvetković [24]. We mention also a nice recent book by Shi [30].

4. Multiple orthogonality

In this section we consider applications of multiple orthogonal polynomials to some special type of quadratures. Otherwise, multiple orthogonal polynomials are intimately related to Hermite-Padé approximants and, because of that, they are known as Hermite-Padé polynomials. A nice survey on these polynomials, as well as some their applications to various fields of mathematics (number theory, special functions, etc.) and in the study of their analytic, asymptotic properties, was given by Aptekarev [1].

4.1. Multiple orthogonal polynomials

Multiple orthogonal polynomials are a generalization of standard orthogonal polynomials in the sense that they satisfy m orthogonality conditions.

Let $m \ge 1$ be an integer and let w_j, $j = 1,\dots,m$, be weight functions on the real line so that the support of each w_j is a subset of an interval E_j. Let $\vec{n} = (n_1, n_2, \dots, n_m)$ be a vector of m nonnegative integers, which is called a *multi-index* with the length $|\vec{n}| = n_1 + n_2 + \cdots + n_m$. There are two types of multiple orthogonal polynomials, but here we consider only the so-called *type II multiple orthogonal polynomials* $\pi_{\vec{n}}(t)$ of degree $|\vec{n}|$. Such monic polynomials are defined by the m orthogonality relations

$$\left.\begin{aligned}
&\int_{E_1} \pi_{\vec{n}}(t)\, t^{\ell} w_1(t)\, dt = 0, \qquad \ell = 0, 1, \ldots, n_1 - 1,\\
&\int_{E_2} \pi_{\vec{n}}(t)\, t^{\ell} w_2(t)\, dt = 0, \qquad \ell = 0, 1, \ldots, n_2 - 1,\\
&\qquad\qquad\vdots\\
&\int_{E_m} \pi_{\vec{n}}(t)\, t^{\ell} w_m(t)\, dt = 0, \qquad \ell = 0, 1, \ldots, n_m - 1.
\end{aligned}\right\} \tag{4.1}$$

Evidently, for $m = 1$ they reduce to the ordinary orthogonal polynomials.

The conditions (4.1) give $|\vec{n}|$ linear equations for the $|\vec{n}|$ unknown coefficients $a_{k,\vec{n}}$ of the polynomial $\pi_{\vec{n}}(t) = \sum_{k=0}^{|\vec{n}|} a_{k,\vec{n}}\, t^k$, where $a_{|\vec{n}|,\vec{n}} = 1$. However, the matrix of coefficients of this system of equations can be singular and we need some additional conditions on the m weight functions to provide the uniqueness of the multiple orthogonal polynomials. If the polynomial $\pi_{\vec{n}}(t)$ is unique, then we say that $\vec{n}$ is a *normal multi-index* and if all multi-indices are normal then we have a *complete system.*

One important complete system is the AT system, in which all weight functions are supported on the same interval E $(= E_1 = E_2 = \cdots = E_m)$ and the following $|\vec{n}|$ functions:

$$w_1(t),\, tw_1(t),\, \ldots,\, t^{n_1-1}w_1(t),\quad w_2(t),\, tw_2(t),\, \ldots, t^{n_2-1}w_2(t),\\ \ldots,\, w_m(t),\, tw_m(t),\, \ldots,\, t^{n_m-1}w_m(t)$$

form a *Chebyshev system* on E for each multi-index $\vec{n}$. This means that every linear combination

$$\sum_{j=1}^{m} Q_{n_j-1}(t) w_j(t),$$

where Q_{n_j-1} is a polynomial of degree at most $n_j - 1$, has at most $|\vec{n}| - 1$ zeros on E.

In 2001 Van Assche and Coussement [42] proved the following result:

Theorem 4.1. *In an AT system the type II multiple orthogonal polynomial $\pi_{\vec{n}}(x)$ has exactly $|\vec{n}|$ zeros on E.*

For these multiple orthogonal polynomials with *nearly diagonal multi-index* there is an interesting recurrence relation of order $m + 1$. Let $n \in \mathbb{N}$ and write it as $n = km + j$, with $k = [n/m]$ and $0 \le j < m$. The nearly diagonal multi-index $\vec{s}(n)$ corresponding to n is given by

$$\vec{s}(n) = (\underbrace{k+1, k+1, \ldots, k+1}_{j \text{ times}}, \underbrace{k, k, \ldots, k}_{m-j \text{ times}}).$$

Denote the corresponding type II multiple (monic) orthogonal polynomials by $\pi_n(t) = \pi_{\vec{s}(n)}(t)$. Then, the following recurrence relation

$$x\pi_k(t) = \pi_{k+1}(t) + \sum_{i=0}^{m} \alpha_{k,m-i}\pi_{k-i}(t)\,, \quad k \geq 0, \tag{4.2}$$

holds, with initial conditions $\pi_0(t) = 1$ and $\pi_i(t) = 0$ for $i = -1, -2, \dots, -m$ (see [41]).

Setting $k = 0, 1, \dots, n-1$ in the recurrence relation (4.2), we get

$$t \begin{bmatrix} \pi_0(t) \\ \pi_1(t) \\ \vdots \\ \pi_{n-1}(t) \end{bmatrix} = H_n \begin{bmatrix} \pi_0(t) \\ \pi_1(t) \\ \vdots \\ \pi_{n-1}(t) \end{bmatrix} + \pi_n(t) \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

i.e.,

$$H_n \mathbf{p}_n(t) = t\, \mathbf{p}_n(t) - \pi_n(t)\mathbf{e}_n, \tag{4.3}$$

where

$$\mathbf{p}_n(t) = \begin{bmatrix}\pi_0(t) & \pi_1(t) & \dots & \pi_{n-1}(t)\end{bmatrix}^T, \qquad \mathbf{e}_n = [0 \;\; 0 \;\; \dots \;\; 0 \;\; 1]^T,$$

and $H_n = [h_{ij}]_{i,j=1}^{n}$ is a lower (banded) Hessenberg matrix of order n, where

$$\begin{aligned} h_{i,i+1} &= 1, \quad i = 1, \dots, n-1; \\ h_{i,i-r} &= \alpha_{i-1,m-r}, \quad i = r+1, \dots, n, \; r = 0, 1, \dots, m. \end{aligned}$$

It is easy to see that $\pi_n(t) = \det(tI_n - H_n)$, where I_n is the identity matrix of the order n. In [25] we presented an effective numerical method for constructing the Hessenberg matrix H_n using a form of the discretized Stieltjes-Gautschi procedure.

These multiple orthogonal polynomials can be applied to some kinds of quadratures. Here, we consider such two applications.

4.2. Quadratures of C.F. Borges

In 1994 Borges [3] considered a problem that arises in evaluation of computer graphics illumination models. Starting with that problem, he examined the problem of numerically evaluating a set of m definite integrals taken with respect to distinct weight functions w_j, $j = 1, 2, \dots, m$, but related by a common integrand and interval of integration

$$\int_E f(t) w_j(t)\, dt, \quad j = 1, 2, \dots, m.$$

It was shown that it is not efficient to use a set of m Gauss-Christoffel quadrature formulas because valuable information is wasted.

In [3] Borges introduced a performance ratio as

$$R = \frac{\text{Overall degree of exactness} + 1}{\text{Number of integrand evaluation}}\,.$$

For example, for a set of m Gauss-Christoffel n-point quadrature formulas, this performance index gives

$$R = \frac{(2n-1)+1}{mn} = \frac{2}{m},$$

i.e., $R < 1$ for all $m > 2$.

Borges [3] proposed quadratures of the following form

$$\int_E f(t)\, w_j(t)\, dt \approx \sum_{\nu=1}^{n} A_{j,\nu} f(\tau_\nu), \quad j = 1, 2, \dots, m. \tag{4.4}$$

If $A_{j,\nu}$ are determined so that (4.4) are interpolatory quadratures (degree of exactness $\le n-1$), then $R = n/n = 1$. However, this performance ratio can be improved taking an AT system of the weights $W = \{w_1, w_2, \dots, w_m\}$ supported on the same interval E. For a multi-index $\vec{n} = (n_1, n_2, \dots, n_m)$ we put $n = |\vec{n}|$.

Following [3, Definition 3] an optimal set of quadratures with respect to $(W, \vec{n})$ was introduced in [25]. In that sense, the Borges set of quadratures (4.4) is optimal if and only if their weight coefficients $A_{j,\nu}$ and nodes τ_ν satisfy the following system of equations

$$\sum_{\nu=1}^{n} A_{j,\nu}\, \tau_\nu^k = \int_E t^k w_j(t)\, dt, \quad k = 0, 1, \dots, n + n_j - 1,$$

for $j = 1, 2 \dots, m$.

Regarding this facts, the following characterization of Borges quadratures in terms of multiple orthogonal polynomials can be given (see [25]):

Theorem 4.2. *Let W be an AT system of weight functions supported on the interval E, $\vec{n} = (n_1, n_2, \dots, n_m)$, and $n = |\vec{n}|$. The Borges quadrature formulae* (4.4) *form an optimal set with respect to* $(W, \vec{n})$ *if and only if:*

1° *They are exact for all polynomials of degree $\le n-1$;*

2° *The node polynomial $q_n(t) = (t-\tau_1)(t-\tau_2)\cdots(t-\tau_n)$ is the type II multiple orthogonal polynomial $\pi_{\vec{n}}$ with respect to W.*

Notice that the performance ratio for such quadratures is $R > 1$. Evidently, the nodes τ_ν, $\nu = 1, \dots, n$, as a zeros of the type II multiple orthogonal polynomial $\pi_{\vec{n}}$, are distinct and located in E (see Theorem 4.1). The weight coefficients satisfy m systems of linear equations with Vandermonde matrix

$$V(\tau_1, \tau_2, \dots, \tau_n) \begin{bmatrix} A_{j,1} \\ A_{j,2} \\ \vdots \\ A_{j,n} \end{bmatrix} = \begin{bmatrix} \mu_0^{(j)} \\ \mu_1^{(j)} \\ \vdots \\ \mu_{n-1}^{(j)} \end{bmatrix}, \quad j = 1, 2, \dots, m,$$

where

$$\mu_\nu^{(j)} = \int_E t^\nu w_j(t)\, dt, \quad \nu = 0, 1, \dots, n-1.$$

This Vandermonde matrix is non-singular and each of the previous systems always has the unique solution.

For the case of the nearly diagonal multi-indices $\vec{s}(n)$ we can compute the nodes τ_ν, $\nu = 1, \dots, n$, as eigenvalues of the corresponding banded Hessenberg matrix H_n. Then, from (4.3) it follows that the eigenvector associated with τ_ν is given by $\mathbf{p}_n(\tau_\nu)$, where $\mathbf{p}_n(t) = \big[\pi_0(t) \;\; \pi_1(t) \;\; \dots \;\; \pi_{n-1}(t)\big]^T$. We can use now this fact to compute the weight coefficients $A_{j,\nu}$ by requiring that each rule correctly generate the first n modified moments

$$\hat{\mu}_\nu^{(j)} = \int_E \pi_\nu(t) w_j(t)\, dt, \qquad \nu = 0, 1, \dots, n-1.$$

Let V_n be the matrix of the eigenvectors of matrix H_n, each normalized so that the first component is equal to 1, i.e.,

$$V_n = \big[\mathbf{p}_n(\tau_1) \; \mathbf{p}_n(\tau_2) \; \dots \; \mathbf{p}_n(\tau_n)\big].$$

Thus, for determining the weight coefficients we should solve the following m systems of equations

$$V_n \begin{bmatrix} A_{j,1} \\ A_{j,2} \\ \vdots \\ A_{j,n} \end{bmatrix} = \begin{bmatrix} \hat{\mu}_0^{(j)} \\ \hat{\mu}_1^{(j)} \\ \vdots \\ \hat{\mu}_{n-1}^{(j)} \end{bmatrix}, \qquad j = 1, 2, \dots, m.$$

This efficient and stable algorithm for constructing Borges quadratures, as well as several numerical examples, were given in [25].

4.3. Birkhoff-Young quadratures and improvements

For numerical integration of analytic functions over a line segment in the complex plane, Birkhoff and Young [2] proposed a quadrature formula of the form

$$\begin{aligned} I(f) \;\; &= \int_{z_0-h}^{z_0+h} f(z)\, dz = \frac{h}{15}\big\{24 f(z_0) + 4\,[f(z_0+h) + f(z_0-h)] \\ &\qquad - [f(z_0+\mathrm{i}h) + f(z_0-\mathrm{i}h)]\big\} + R_5^{BY}(f). \end{aligned} \tag{4.5}$$

For the error term $R_5^{BY}(f)$ the following estimate [45] (see also Davis and Rabinowitz [9, p. 136])

$$|R_5^{BY}(f)| \le \frac{|h|^7}{1890} \max_{z\in S} |f^{(6)}(z)|$$

holds, where S denotes the square with vertices $z_0 + \mathrm{i}^k h$, $k = 0, 1, 2, 3$. This error estimate is about four tenths as large as the corresponding error $R_5^{ES}(f)$ for the so-called extended Simpson's rule (cf. [29, p. 124])

$$I(f) \approx \frac{h}{90}\big\{114 f(z_0) + 34\big[f(z_0+h) + f(z_0-h)\big] - \big[f(z_0+2h) + f(z_0-2h)\big]\big\},$$

for which we have

$$|R_5^{ES}(f)| \sim \frac{|h|^7}{756} |f^{(6)}(\zeta)|, \qquad 0 < \frac{\zeta - (z_0 - 2h)}{4h} < 1.$$

Without loss of generality we can consider the integration over $[-1,1]$ for analytic functions in a unit disk $\Omega=\{z\,:\,|z|\le 1\}$, so that the previous Birkhoff-Young formula (4.5) becomes

$$\int_{-1}^{1} f(z)\,dz=\frac{8}{5}f(0)+\frac{4}{15}\bigl[f(1)+f(-1)\bigr]-\frac{1}{15}\bigl[f(\mathrm{i})+f(-\mathrm{i})\bigr]+R_5(f). \quad (4.6)$$

In 1976 Lether [19] pointed out that the three point Gauss-Legendre quadrature

$$\int_{-1}^{1} f(z)\,dz=\frac{8}{9}f(0)+\frac{5}{9}\bigl[f(\sqrt{3/5})+f(-\sqrt{3/5})\bigr]+R_3(f), \quad (4.7)$$

which is also exact for all polynomials of degree at most five, is more precise than (4.6) and he recommended it for numerical integration. However, Tošić [39] improved the quadrature (4.6) in the form

$$\int_{-1}^{1} f(z)\,dz=Af(0)+B\bigl[f(r)+f(-r)\bigr]+C\bigl[f(\mathrm{i}r)+f(-\mathrm{i}r)\bigr]+R_5^T(f;r), \quad (4.8)$$

where

$$A=2\left(1-\frac{1}{5r^4}\right),\quad B=\frac{1}{6r^2}+\frac{1}{10r^4},\quad C=-\frac{1}{6r^2}+\frac{1}{10r^4},\quad 0<r<1,$$

and the error-term is given by the expression

$$R_5^T(f;r)=\left(-\frac{2}{3}r^4+\frac{2}{7}\right)\frac{f^{(6)}(0)}{6!}+\left(-\frac{2}{5}r^4+\frac{2}{9}\right)\frac{f^{(8)}(0)}{8!}+\cdots\,. \quad (4.9)$$

Evidently, for $r=1$ this formula reduces to (4.6) and for $r=\sqrt{3/5}$ to the Gauss-Legendre formula (4.7) (then $C=0$). Moreover, for $r=\sqrt[4]{3/7}$ the first term on the right-hand side in (4.9) vanishes and (4.8) reduces to the modified Birkhoff-Young formula of maximum accuracy (named *MF* in [39]), with the coefficients

$$A=\frac{16}{15},\quad B=\frac{1}{6}\left(\frac{7}{5}+\sqrt{\frac{7}{3}}\right),\quad C=\frac{1}{6}\left(\frac{7}{5}-\sqrt{\frac{7}{3}}\right),$$

and with the error-term

$$R_5^{MF}(f)=R_5^T(f;\sqrt[4]{3/7})=\frac{1}{793800}f^{(8)}(0)+\frac{1}{61122600}f^{(10)}(0)+\cdots\,.$$

This formula was extended by Milovanović and Đorđević [23] to the following quadrature formula of interpolatory type

$$\begin{aligned}\int_{-1}^{1} f(z)\,dz=Af(0)&+C_{11}\bigl[f(r_1)+f(-r_1)\bigr]+C_{12}\bigl[f(\mathrm{i}r_1)+f(-\mathrm{i}r_1)\bigr]\\&+C_{21}\bigl[f(r_2)+f(-r_2)\bigr]+C_{22}\bigl[f(\mathrm{i}r_2)+f(-\mathrm{i}r_2)\bigr]+R_9(f;r_1,r_2),\end{aligned}$$

where $0<r_1<r_2<1$. They proved that for

$$r_1=r_1^*=\sqrt[4]{\frac{63-4\sqrt{114}}{143}}\quad\text{and}\quad r_2=r_2^*=\sqrt[4]{\frac{63+4\sqrt{114}}{143}},$$

this formula reduces to a quadrature rule of the algebraic exactness $p = 13$, with the error-term

$$R_9(f; r_1^*, r_2^*) = \frac{1}{28122661066500} f^{(14)}(0) + \cdots \approx 3.56 \cdot 10^{-14} f^{(14)}(0).$$

4.4. Generalized Birkhoff-Young quadratures

In this subsection we consider a kind of generalized Birkhoff-Young quadrature formulas and give a connection with multiple orthogonal polynomials. We introduce N-point quadrature formula for weighted integrals of analytic functions in $\Omega = \{z : |z| \le 1\}$,

$$I(f) := \int_{-1}^{1} f(z)w(z)\,dz = Q_N(f) + R_N(f),$$

where $w : (-1,1) \to \mathbb{R}^+$ is an *even* positive weight function, for which all moments $\mu_k = \int_{-1}^{1} z^k w(z)\,dz$, $k = 0, 1, \ldots$, exist.

For a given fixed integer $m \ge 1$ and for each $N \in \mathbb{N}$, we put $N = 2mn+\nu$, where $n = [N/2m]$ and $\nu \in \{0, 1, \ldots, 2m-1\}$. We define the node polynomial

$$\omega_N(z) = z^\nu p_{n,\nu}(z^{2m}) = z^\nu \prod_{k=1}^{n} (z^{2m} - r_k), \quad 0 < r_1 < \cdots < r_n < 1, \quad (4.10)$$

and consider the corresponding interpolatory quadrature rule Q_N of the form

$$Q_N(f) = \sum_{j=0}^{\nu-1} C_j f^{(j)}(0) + \sum_{k=1}^{n} \sum_{j=1}^{m} A_{k,j} \left[f\bigl(x_k e^{\mathrm{i}\theta_j}\bigr) + f\bigl(-x_k e^{\mathrm{i}\theta_j}\bigr) \right],$$

where

$$x_k = \sqrt[2m]{r_k}, \quad k = 1, \ldots, n; \quad \theta_j = \frac{(j-1)\pi}{m}, \quad j = 1, \ldots, m.$$

If $\nu = 0$, the first sum in $Q_N(f)$ is empty.

Theorem 4.3. *Let m be a fixed positive integer and w be an even positive weight function w on $(-1,1)$, for which all moments $\mu_k = \int_{-1}^{1} z^k w(z)\,dz$, $k \ge 0$, exist. For any $N \in \mathbb{N}$ there exists a unique interpolatory quadrature $Q_N(f)$ with a maximal degree of exactness $d_{\max} = 2(m+1)n + s$, where*

$$n = \left[\frac{N}{2m}\right], \quad \nu = N - 2mn, \quad s = \begin{cases} \nu - 1, & \nu \ \text{even}, \\ \nu, & \nu \ \text{odd}. \end{cases} \quad (4.11)$$

The node polynomial (4.10) *is characterized by the following orthogonality relations*

$$\int_0^1 t^k p_{n,\nu}(t^m) t^{s/2} w(\sqrt{t})\,dt = 0, \quad k = 0, 1 \ldots, n-1. \quad (4.12)$$

Proof. For a given $N \in \mathbb{N}$ and a fixed $m \in \mathbb{N}$, suppose that $f \in \mathcal{P}_d$, where $d \ge N = 2mn + \nu$, with $n = [N/2m]$ and $\nu = N - 2mn$. Then, it can be expressed in the form

$$f(z) = u(z)\omega_N(z) + v(z) = u(z) z^\nu p_{n,\nu}(z^{2m}) + v(z), \quad u \in \mathcal{P}_{d-N}, \ v \in \mathcal{P}_{N-1},$$

from which, by an integration with respect to the weight function w, we get

$$I(f) = \int_{-1}^{1} u(z) z^{\nu} p_{n,\nu}(z^{2m}) w(z)\, dz + I(v).$$

Since our quadrature is interpolatory and $v(z) = f(z)$ at the zeros of ω_N, we have $I(v) = Q_N(v) = Q_N(f)$. Thus, the quadrature formula $Q_N(f)$ has a maximal degree of precision if and only if

$$\int_{-1}^{1} u(z) z^{\nu} p_{n,\nu}(z^{2m}) w(z)\, dz = 0$$

for a maximal degree of the polynomial $u \in \mathcal{P}_{d-N}$. According to the values of ν, this "orthogonality condition" can be represented in the form

$$\int_{-1}^{1} h(z^2) z^{s+1} p_{n,\nu}(z^{2m}) w(z)\, dz = 0, \quad h \in \mathcal{P}_{n-1}, \tag{4.13}$$

which means that the maximal degree of the polynomial $u \in \mathcal{P}_{d-N}$ is

$$d_{\max} - N = \begin{cases} 2n-1, & \nu \text{ is even}, \\ 2n, & \nu \text{ is odd}, \end{cases}$$

i.e., $d_{\max} = 2(m+1)n + s$, where s is defined by (4.11).

Finally, by substitution $z^2 = t$, the orthogonality conditions (4.13) can be expressed in the form (4.12). $\square$

Regarding (4.12) the polynomial $t \mapsto p_{n,\nu}(t^m)$ (of degree mn) is orthogonal to $\mathcal{P}_n$ with respect to the weight function $t^{s/2} w(\sqrt{t})$ on $(0,1)$, and it can be interpreted in terms of multiple orthogonal polynomials.

Theorem 4.4. *Under conditions of the previous theorem, for any $N \in \mathbb{N}$ there exists a unique interpolatory quadrature $Q_N(f)$, with a maximal degree of exactness $d_{\max} = 2(m+1)n + s$, if and only if the polynomial $p_{n,\nu}(t)$ is the type II multiple orthogonal polynomial $\pi_{\vec{n}}(t)$, with respect to the weights $w_j(t) = t^{(s+2j)/(2m)-1} w(t^{1/(2m)})$, with $n_j = 1 + \left[\frac{n-j}{m}\right]$, $j = 1, \ldots, m$.*

Proof. Evidently, the conditions (4.12) are equivalent to

$$\int_0^1 t^{k/m} p_{n,\nu}(t) t^{(s+2)/(2m)-1} w(t^{1/(2m)})\, dt = 0, \quad k = 0, 1, \ldots, n-1.$$

Now, putting $k = m\ell + j - 1$, $\ell = [k/m]$, we get for each $j = 1, \ldots, m$,

$$\int_0^1 t^{\ell} p_{n,\nu}(t) w_j(t)\, dt = 0, \quad \ell = 0, 1 \ldots, n_j - 1,$$

where

$$w_j(t) = t^{(s+2j)/(2m)-1} w(t^{1/(2m)}) \quad \text{and} \quad n_j = 1 + \left[\frac{n-j}{m}\right].$$

Notice that these weight functions, defined on the same interval $E_1 = E_2 = \cdots = E_m = E = (0,1)$, can be expressed in the form $w_j(t) = t^{(j-1)/m} w_1(t)$, $j = 1, \ldots, m$, where $w_1(t) = t^{(s+2)/(2m)-1} w(t^{1/(2m)})$. Since the Müntz system $\{t^{k+(j-1)/m}\}$, $k = 0, 1, \ldots, n_j - 1$; $j = 1, \ldots, m$, is a Chebyshev system on $[0,\infty)$, and also on $E=(0,1)$, and $w_1(t)>0$ on E, we conclude that $\{w_j,\, j=1,\ldots,m\}$ is an AT system on E.

Therefore, according to Theorem 4.1, the unique type II multiple orthogonal polynomial $p_{n,\nu}(t)=\pi_{\vec{n}}(t)$ has exactly

$$|\vec{n}|:=\sum_{j=1}^{m} n_j=\sum_{j=1}^{m}\left(1+\left[\frac{n-j}{m}\right]\right)=n$$

zeros in $(0,1)$. □

References

[1] Aptekarev, A.I., *Multiple orthogonal polynomials*, J. Comput. Appl. Math., **99**(1998), 423–447.

[2] Birkhoff, G., Young, D.M., *Numerical quadrature of analytic and harmonic functions*, J. Math. Phys., **29**(1950), 217–221.

[3] Borges, C.F., *On a class of Gauss-like quadrature rules*, Numer. Math., **67**(1994), 271–288.

[4] Brezinski, C., Driver, K.A., Redivo-Zaglia, M., *Quasi-orthogonality with applications to some families of classical orthogonal polynomials*, Appl. Numer. Math., **48**(2004), 157–168.

[5] Bultheel, A., Cruz-Barroso, R., Van Barel, M., *On Gauss-type quadrature formulas with prescribed nodes anywhere on the real line*, Calcolo, **47**(2010), 21–48.

[6] Chakalov, L., *Über eine allgemeine Quadraturformel*, C.R. Acad. Bulgar. Sci., **1**(1948), 9–12.

[7] Chakalov, L., *General quadrature formulae of Gaussian type*, Bulgar. Akad. Nauk Izv. Mat. Inst., **1**(1954), 67–84 (Bulgarian); English transl.: East J. Approx., **1**(1995), 261–276.

[8] Chakalov, L., *Formules générales de quadrature mécanique du type de Gauss*, Colloq. Math., **5**(1957), 69–73.

[9] Davis, P.J., Rabinowitz, P., *Methods of Numerical Integration*, Academic Press, New York, 1975.

[10] Freud, G., *Orthogonal Polynomials*, Pergamon Press, Oxford, 1971.

[11] Gauss, C.F., *Methodus nova integralium valores per approximationem inveniendi*, Commentationes Societatis Regiae Scientarium Göttingensis Recentiores, **3**(1814) 163–196.

[12] Gautschi, W., *A survey of Gauss-Christoffel quadrature formulae*, In: *E.B. Christoffel – The Influence of his Work on Mathematics and the Physical Sciences* (P.L. Butzer, F. Fehér, eds.), Birkhäuser, Basel, 1981, 72–147.

[13] Gautschi, W., *Orthogonal polynomials: applications and computation*, Acta Numerica, 1996, 45–119.

[14] Gautschi, W., *Orthogonal Polynomials: Computation and Approximation*, Clarendon Press, Oxford, 2004.

[15] Ghizzetti, A., Ossicini, A., *Quadrature Formulae*, Akademie Verlag, Berlin, 1970.

[16] Ghizzetti, A., Ossicini, A., *Sull' esistenza e unicità delle formule di quadratura gaussiane*, Rend. Mat., **8**(1975), no. 6, 1–15.

[17] Golub, G.H., Kautsky, J., *Calculation of Gauss quadratures with multiple free and fixed knots*, Numer. Math., **41**(1983), 147–163.

[18] Golub, G.H., Welsch, J.H., *Calculation of Gauss quadrature rule*, Math. Comp., **23**(1969), 221–230.

[19] Lether, F., *On Birkhoff-Young quadrature of analytic functions*, J. Comput. Appl. Math., **2**(1976), 81–84.

[20] Mastroianni, G., Milovanović, G.V., *Interpolation Processes – Basic Theory and Applications*, Springer Monographs in Mathematics, Springer – Verlag, Berlin – Heidelberg, 2008.

[21] Micchelli, C.A., Rivlin, T.J., *Numerical integration rules near Gaussian formulas*, Israel J. Math., **16**(1973), 287–299.

[22] Milovanović, G.V., *Quadratures with multiple nodes, power orthogonality, and moment-preserving spline approximation*, J. Comput. Appl. Math., **127**(2001), 267–286.

[23] Milovanović, G.V., Đorđević, R.Ž., *On a generalization of modified Birkhoff-Young quadrature formula*, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fis., **735-762**(1982), 130–134.

[24] Milovanović, G.V., Spalević, M.M., Cvetković, A.S., *Calculation of Gaussian type quadratures with multiple nodes*, Math. Comput. Modelling, **39**(2004), 325–347.

[25] Milovanović, G.V., Stanić, M., *Construction of multiple orthogonal polynomials by discretized Stieltjes-Gautschi procedure and corresponding Gaussian quadrature*, Facta Univ. Ser. Math. Inform., **18**(2003), 9–29.

[26] Peherstorfer, F., *Characterization of positive quadrature formulas*, SIAM J. Math. Anal., **12**(1981), 935–942.

[27] Peherstorfer, F., *Characterization of positive quadrature formulas II*, SIAM J. Math. Anal., **15**(1984), 1021–1030.

[28] Popoviciu, T., *Sur une généralisation de la formule d'intégration numérique de Gauss*, (Romanian), Acad. R. P. Romîne Fil. Iaşi Stud. Cerc. Şti., **6**(1955), 29–57.

[29] Scarborough, J.B., *Numerical Mathematical Analysis*, The Johns Hopkins Press, Baltimore, 1930.

[30] Shi, Y.G., *Power Orthogonal Polynomials*, Nova Science Publishers, Inc., New York, 2006.

[31] Stancu, D.D., *On a class of orthogonal polynomials and on some general quadrature formulas with minimum number of terms*, Bull. Math. Soc. Sci. Math. Phys. R. P. Romîne (N.S), **49**(1957), no. 1, 479–498.

[32] Stancu, D.D., *On the interpolation formula of Hermite and some applications of it*, (Romanian), Acad. R. P. Romîne Fil. Cluj Stud. Cerc. Mat., **8**(1957), 339–355.

[33] Stancu, D.D., *Generalization of the quadrature formula of Gauss-Christoffel*, (Romanian), Acad. R.P. Romîne Fil. Iaşi Stud. Cerc. Şti. Mat., **8**(1957), 1–18.

[34] Stancu, D.D., *A method for constructing quadrature formulas of higher degree of exactness*, (Romanian), Com. Acad. R. P. Romîne, **8**(1958) 349–358.

[35] Stancu, D.D., *On certain general numerical integration formulas*, (Romanian), Acad. R. P. Romîne. Stud. Cerc. Mat., **9**(1958), 209–216.

[36] Stancu, D.D., *Sur quelques formules générales de quadrature du type Gauss-Christoffel*, Mathematica (Cluj), **24**(1959), no. ,1 167–182.

[37] Stancu, D.D., Stroud, A.H., *Quadrature formulas with simple Gaussian nodes and multiple fixed nodes*, Math. Comp., **17**(1963), 384–394.

[38] Stroud, A.H., Stancu, D.D., *Quadrature formulas with multiple Gaussian nodes*, J. SIAM Numer. Anal. Ser. B **2**(1965), 129–143.

[39] Tošić, D.Đ., *A modification of the Birkhoff-Young quadrature formula for analytic functions*, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fis., **602-633**(1978), 73–77.

[40] Turán, P., *On the theory of the mechanical quadrature*, Acta Sci. Math. Szeged, **12**(1950), 30–37.

[41] Van Assche, W., *Non-symetric linear difference equations for multiple orthogonal polynomials*, CRM Proceedings and Lecture Notes, **25**, Amer. Math. Soc., Providence, RI, 2000, 391–405.

[42] Van Assche, W., Coussement, E., *Some classical multiple orthogonal polynomials*, J. Comput. Appl. Math., **127**(2001), 317–347.

[43] Xu, Y., *Quasi-orthogonal polynomials, quadrature, and interpolation*, J. Math. Anal. Appl., **182**(1994), 779–799.

[44] Xu, Y., *A characterization of positive quadrature formulae*, Math. Comp., **62**(1994), 703–718.

[45] Young, D.M., *An error bound for the numerical quadrature of analytic functions*, J. Math. Phys., **31**(1952), 42–44.

Gradimir V. Milovanović
Faculty of Computer Sciences, Megatrend University
Bulevar Umetnosti 29, 11070 Novi Beograd
Serbia
e-mail: `gvm@megatrend.edu.rs`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 465–472

On the pointwise convergence of the Chebyshev best approximation on Jacobi nodes

Alexandru I. Mitrea

Abstract. This paper is devoted to obtain estimates and to point out convergence-type results and the superdense unbounded divergence for some pointwise approximation formulas, related to the Chebyshev best approximation on Jacobi node matrix.

Mathematics Subject Classification (2010): 41A05, 41A10, 41A17.

Keywords: Chebyshev best approximation, pointwise approximation, superdense unbounded divergence, error-approximation.

1. Introduction

Denote by C the Banach space of all continuous functions $f : [-1, 1] \to \mathbb{R}$, endowed with the uniform norm and let C^r, $r \geq 1$, be the subspace of C which contains the functions f whose derivatives up to the order r belong to C; we admit $C^0 = C$.

Let us consider, also, a strictly increasing sequence of positive integers m_n, with $m_n \geq n + 1$, $\forall\ n \geq 1$, and the node matrix

$$\mathcal{M} = \{x^k_{m_n} :\ n \geq 1,\ 1 \leq k \leq m_n\}, \tag{1.1}$$

where $-1 \leq x^1_{m_n} < x^2_{m_n} < x^3_{m_n} < \ldots < x^{m_n}_{m_n} \leq 1$.

Define the operators $U_n : C \to \mathcal{P}_n$, $n \geq 1$, as follows: for each f in C, let $U_n f$ be the unique polynomial of $\mathcal{P}_n$ for which the infimum of the set

$$\{\max\{|f(x^k_{m_n}) - P(x^k_{m_n})| :\ 1 \leq k \leq m_n\} :\ P \in \mathcal{P}_n\} \tag{1.2}$$

is attained, [1], [4]; in this paper, $\mathcal{P}_n$ is the usual notation for the set of all algebraic polynomials of degree at most $n \in \mathbb{N}$.

The polynomial $U_n f = U_n(f; \mathcal{M}) \in \mathcal{P}_n$, that provides the best approximation of f in the Chebyshev sense, with respect to the finite point set

$$J_n = \{x^k_{m_n} :\ 1 \leq k \leq m_n\}, \quad n \geq 1, \tag{1.3}$$

is said to be *the $\mathcal{M}$-projection* of f on the space $\mathcal{P}_n$.

We associate to each row J_n, $n \geq 1$, and $f \in C$, the Lagrange polynomial $L_{m_n} f$ which interpolates f at the nodes of J_n, namely

$$(L_{m_n} f)(x) = \sum_{k=1}^{m_n} f(x_{m_n}^k) l_{m_n}^k(x), \quad x \in [-1,1], \tag{1.4}$$

and the Lebesgue function $\Lambda_{m_n} : [-1,1] \to [0,\infty)$,

$$\Lambda_{m_n}(x) = \sum_{k=1}^{m_n} |l_{m_n}^k(x)|, \quad x \in [-1,1], \tag{1.5}$$

where

$$l_{m_n}^k(x) = \frac{u_{m_n}(x)}{(x - x_{m_n}^k) u'_{m_n}(x_{m_n}^k)},\ 1 \leq k \leq m_n;\ u_{m_n}(x) = \prod_{k=1}^{m_n} (x - x_{m_n}^k),\ n \geq 1. \tag{1.6}$$

Clearly, if $m_n = n+1$, $n \geq 1$, then the operators U_n coincide with the classical Lagrange projection operators, $f \mapsto L_{m_n} f$.

On the other hand, assuming that each row J_n of $\mathcal{M}$ contains exactly $n+2$ points, i.e. $m_n = n+2$, $\forall\ n \geq 1$, Ph. C. Curtis Jr., [4], has proved that the corresponding $\mathcal{M}$-projection operators U_n, $n \geq 1$, are linear and continuous operators and there exists a function $g \in C$ for which the sequence $(U_n g)_{n \geq 1}$ fails to converge uniformly on $[-1,1]$. As we proved in [5], the set of all functions $f \in C$ with the property that $\limsup_{n\to\infty} \|U_n f\| = \infty$ is, in fact, a superdense set in the Banach space $(C, \|\cdot\|)$; this superdense unbounded divergence remains valid if $m_n = n+3$ and the nodes of J_n are symmetric with respect to the origin, $\forall\ n \geq 1$, [6]. We recall that a subset $\mathcal{S}$ of a topological space $\mathcal{T}$ is said to be *superdense* in $\mathcal{T}$ if it is residual (i.e. its complement is of first Baire category), uncountable and dense in $\mathcal{T}$. These results of divergence type contrast with the well-known theorem concerning the uniform convergence of the best approximation polynomials in supremum norm, which states that the operators $Q_n : C \to \mathcal{P}_n$, defined by $\|f - Q_n f\| = \inf\{\|f - P\| :\ P \in \mathcal{P}_n\}$, $f \in C$, are continuous nonlinear projections and the sequence $(Q_n f)_{n\geq 1}$ is uniformly convergent to f, for each $f \in C$.

In the next sections, we consider the case $m_n = n+2$, $n \geq 1$. Our aim is to point out estimates, results of convergence type and the phenomenon of condensation of singularities for some pointwise approximation formulas associated to the Chebyshev best approximation on the Jacobi node matrix.

The paper is organized as follows. In the second section, we introduce the point-functionals that define the pointwise approximation formulas for an arbitrary node matrix $\mathcal{M}$ in (1.1) and we derive an estimate of the corresponding approximation error. In the third and fourth sections we establish results of convergence type and we prove the superdense unbounded divergence, respectively, for the pointwise approximation formulas corresponding to the Jacobi matrix. To this goal, we use the following *principle of condensation of singularities* from Functional Analysis.

Theorem 1.1. [2], [3]. *If X is a Banach space, Y is a normed space and $(A_n)_{n\geq 1}$ is a sequence of continuous linear operators from X into Y so that the set of norms $\{\|A_n\| : \ n \geq 1\}$ is unbounded, then the set of singularities of the family $\{A_n : \ n \geq 1\}$, namely*

$$\mathcal{S} = \left\{ x \in X : \ \limsup_{n\to\infty} \|A_n x\| = \infty \right\},$$

is superdense in X.

In this paper, the notations m, M, M_k, $k \geq 1$, stand for some generic positive constants, which do not depend on n. If (a_n) and (b_n) are sequences of real numbers with $b_n \neq 0$, we write $a_n \sim b_n$ if $0 < m \leq |a_n/b_n| \leq M$, for all $n \geq 1$. Also, $\omega(f;\cdot)$ denotes the modulus of continuity of a function $f \in C$.

2. Estimates for pointwise approximation formulas

Firstly, let us derive, according to [4], the formula of computing $U_n f$, for a given $n \geq 1$. Let $\sigma_{n+2} \in C$ be a function satisfying the conditions $\sigma_{n+2}(x^k_{n+2}) = (-1)^k$, $1 \leq k \leq n+2$. By means of Theorem of Charles de la Vallée-Poussin, [1], [8], and taking into account (1.2), we get:

$$U_n f = L_{n+2} f - \frac{a_{n+1}(f)}{a_{n+1}(\sigma_{n+2})} L_{n+2}\sigma_{n+2}; \quad f \in C, \ n \geq 1, \tag{2.1}$$

where $a_{n+1}(f)$ is the leading-coefficient of $L_{n+2} f$.

Further, by introducing the notation

$$\tau^k_{n+2} = (u'_{n+2}(x^k_{n+2}))^{-1}, \quad 1 \leq k \leq n+2, \tag{2.2}$$

and remarking that $\mathrm{sign}\tau^k_{n+2} = (-1)^{n-k}$, $1 \leq k \leq n+2$, we have:

$$a_{n+1}(f) = \sum_{k=1}^{n+2} \tau^k_{n+2} f(x^k_{n+2}) \tag{2.3}$$

and

$$a_{n+1}(\sigma_{n+2}) = (-1)^{n+2} \sum_{k=1}^{n+2} |\tau^k_{n+2}|. \tag{2.4}$$

The relations (2.1), (1.4) and the definition of σ_{n+2} lead to:

$$(U_n f)(x) = \sum_{k=1}^{n+2} d^k_{n+2}(f) l^k_{n+2}(x); \quad f \in C, \ |x| \leq 1, \ n \geq 1, \tag{2.5}$$

where the linear functionals $d^k_{n+2} : C \to \mathbb{R}$, are given by:

$$d^k_{n+2}(f) = f(x^k_{n+2}) + (-1)^{k+1} \frac{a_{n+1}(f)}{a_{n+1}(\sigma_{n+2})}, \quad 1 \leq k \leq n+2. \tag{2.6}$$

The relations (2.3) and (2.4) give $|a_{n+1}(f)| \le |a_{n+1}(\sigma_{n+2})| \cdot \|f\|$ which, combined with (2.5) and (2.6), yield $|d^k_{n+2}(f)| \le 2\|f\|$, so:

$$|(U_n f)(x)| \le 2\Lambda_{n+2}(x) \cdot \|f\|; \quad f \in C,\ |x| \le 1,\ n \ge 1. \tag{2.7}$$

Now, for a given point $t \in [-1,1]$, let us define the point-functionals $T^t_n : C \to \mathbb{R}$ by

$$T^t_n(f) = (U_n f)(t) = \sum_{k=1}^{n+2} d^k_{n+2}(f) \cdot l^k_{n+2}(t); \quad f \in C,\ n \ge 1 \tag{2.8}$$

and let us consider the approximation-errors $R^t_n f$, of the pointwise approximation formulas

$$f(t) = T^t_n(f) + R^t_n(f); \quad f \in C,\ n \ge 1, \tag{2.9}$$

associated to the Chebyshev discrete best approximation on the nodes (1.3) of J_n.

By using the relation $U_n P = P$, $\forall\ P \in \mathcal{P}_n$, that follows from (2.1), we obtain, taking into account (2.9):

$$|R^t_n f| = |R^t_n(f-P)| \le |f(t) - P(t)| + |T^t_n(f-P)|, \quad f \in \mathcal{P}_n.$$

The last inequality, combined with (2.7), leads to:

$$|R^t_n f| \le (1 + 2\Lambda_{n+2}(t)) \cdot \|f - P\|; \quad f \in C,\ P \in \mathcal{P}_n. \tag{2.10}$$

Further, let $f \in C^r$, $r \ge 0$. It follows from the inequality of Gopengauz, [9], the existence of a polynomial $\widetilde{P} \in \mathcal{P}_n$ so that:

$$\|f - \widetilde{P}\| \le M_1 n^{-r} \omega\left(f^{(r)}; \frac{1}{n}\right); \quad n \ge 1, \tag{2.11}$$

where $M_1 = M_1(r)$.

We derive from (2.10) and (2.11):

$$|R^t_n f| \le M_1 n^{-r}(1 + 2\Lambda_{n+2}(t))\omega\left(f^{(r)}; \frac{1}{n}\right); \quad f \in C^r,\ n \ge 1. \tag{2.12}$$

Finally, (2.12) leads to the following statement.

Theorem 2.1. *The pointwise approximation formulas (2.8) and (2.9), with respect to an arbitrary point $t \in [-1,1]$, are convergent on C^r, $r \ge 0$, i.e.*

$$\lim_{n\to\infty} T^t_n(f) = f(t),\ \forall\ f \in C^r,$$

if the corresponding Lebesgue functions satisfy the condition

$$\Lambda_{n+2}(t) = O(n^r).$$

3. Results of convergence-type for pointwise Jacobi approximation formulas

In this section and the next section, we take as node matrix $\mathcal{M}$ the Jacobi ultraspherical matrix $\mathcal{M}^{(\alpha)}$, $\alpha > -1$, whose n-th row contains the roots of the Jacobi ultraspherical polynomial $P_{n+2}^{(\alpha)}$, $n \geq 1$. In this framework, the formulas (2.8) and (2.9) will be refereed to as *pointwise Jacobi approximation formulas*, associated to the Chebyshev discrete best approximation.

The following estimate is valid, [7]:

$$\Lambda_n(t) - 1 \sim |P_n^{(\alpha)}(t)|\sqrt{n} \cdot k_n(\alpha); \quad n \geq 2, \ t \in [-1,1], \tag{3.1}$$

with

$$k_n(\alpha) = \begin{cases} 1 + (1-t)^{\alpha/2+1/4} \ln n, & \text{if} \quad \alpha > -1/2 \\ \ln n, & \text{if} \quad \alpha = -1/2 \\ \dfrac{\ln(2 + n\sqrt{1-t})}{(1-t)^{-\alpha/2-1/4} + n^{\alpha+1/2}}, & \text{if} \quad \alpha < -1/2. \end{cases} \tag{3.2}$$

It follows from (2.12), (3.1) and (3.2):

$$|R_n^t f| \leq M_2 n^{-r}(1 + |P_{n+2}^{(\alpha)}(t)|\sqrt{n+2}\ k_{n+2}(\alpha))\omega\left(f^{(r)}; \frac{1}{n}\right); \tag{3.3}$$

$f \in C^r, \ n \geq 2, \ t \in [-1,1]$.

3.1. First case

Suppose that $t \in (-1,1)$. The following statement holds.

Theorem 3.1. *Let consider the Jacobi pointwise approximation formulas with respect to an arbitrary point $t \in (-1,1)$.*

$1°$. *If $r > \alpha + 1/2 > 0$* **or** *$\alpha \leq -\dfrac{1}{2}$ and $r \geq 1$, then these formulas are convergent on the space C^r, namely*

$$\lim_{n\to\infty} T_n^t(f) = f(t), \ \forall \ f \in C^r.$$

$2°$. *If $\alpha + \dfrac{1}{2} \in \mathbb{N}^*$ and $r = \alpha + \dfrac{1}{2}$* **or** *$\alpha \leq -\dfrac{1}{2}$ and $r = 0$, then these formulas are convergent on the subset of all $f \in C^r$, whose r-th derivatives satisfy the Dini-Lipschitz condition*

$$\lim_{\delta \searrow 0} \omega(f^{(r)}; \delta) \ln \delta = 0.$$

Proof. The estimate $\|P_n^{(\alpha)}\| \sim n^q$, with $q = \max\{\alpha, -1/2\}$, [10], together with (3.3), yields:

$$\begin{cases} |R_n^t f| \leq M_3 n^{\alpha - r + 1/2}(\ln n)\omega\left(f^{(r)}; \dfrac{1}{n}\right), & \text{if} \quad \alpha > -1/2 \\ |R_n^t f| \leq M_3 n^{-r}(\ln n)\omega\left(f^{(r)}; \dfrac{1}{n}\right), & \text{if} \quad \alpha \leq -1/2, \end{cases} \tag{3.4}$$

for each $f \in C^r$, $n \geq 2$ and $t \in (-1,1) \setminus J_n$.

If $t \in J_n$, then $P_{n+2}^{(\alpha)}(t) = 0$ and (2.12) provides:

$$|R_n^t f| \leq 3M_1 n^{-r} \omega\left(f^{(r)}; \frac{1}{n}\right),$$

so the formulas (3.4) are valid for $t \in (-1,1)$.

The estimates (3.4) and the properties of ω imply the validity of the assertions 1° and 2° of this theorem, which completes the proof. □

3.2. Second case

Let us examine the remaining cases $t = \pm 1$.

Theorem 3.2. *Let consider the Jacobi pointwise approximation formulas with respect to the end points $t = -1$ and $t = 1$.*

1°. If $r \geq \alpha + 1/2 > 0$ **or** *$\alpha = -1/2$ and $r \geq 1$* **or** *$\alpha < -1/2$ and $r \geq 0$, then these formulas are convergent on the space C^r.*

2°. If $\alpha = -1/2$ and $r = 0$, then these formulas are convergent on the subset of all functions f in C satisfying the Dini-Lipschitz condition

$$\lim_{\delta \searrow 0} \omega(f;\delta) \ln \delta = 0.$$

Proof. Using the estimate $|P_n^{(\alpha)}(\pm 1)| \sim n^\alpha$, $\alpha > -1$, [10], we derive from (3.1) and (3.2)

$$\Lambda_n(1) \sim \begin{cases} n^{\alpha+1/2}, & \text{if} \quad \alpha > -1/2 \\ \ln n, & \text{if} \quad \alpha = -1/2 \\ 1, & \text{if} \quad \alpha < -1/2. \end{cases} \tag{3.5}$$

The relations (2.12) and (3.5) yield:

$$|R_n^{\pm 1} f| \leq M_4 n^{\alpha - r + 1/2} \omega\left(f^{(r)}; \frac{1}{n}\right), \quad \text{if} \quad \alpha > -1/2$$

$$|R_n^{\pm 1} f| \leq M_5 n^{-r} \ln n \omega\left(f^{(r)}; \frac{1}{n}\right), \quad \text{if} \quad \alpha = -1/2$$

$$|R_n^{\pm 1} f| \leq M_6 n^{-r} \omega\left(f^{(r)}; \frac{1}{n}\right), \quad \text{if} \quad \alpha < -1/2,$$

which proves the assertions 1° and 2° of this theorem. □

4. Superdense unbounded divergence for a class of Jacobi pointwise approximation formulas

In this section, we emphasize the phenomenon of condensation of singularities for the family of the pointwise approximating functionals $\{T_n^0 : n \geq 1\}$.

Theorem 4.1. *The set of all functions $f \in C$ for which the Jacobi pointwise approximation formulas* (2.8) *and* (2.9) *with respect to the origin are unboundedly divergent, i.e.* $\limsup_{n\to\infty} |T_n^0 f| = \infty$, *is superdense in the Banach space* $(C, \|\cdot\|)$.

Proof. Define $f_{n+2} \in C$ by

$$f_{n+2}(x) = \begin{cases} \mathrm{sign} l_{n+2}^k(0), & \text{if } x \in J_n \\ 1, & \text{if } x \in \{-1, 1\} \\ \text{linear}, & \text{otherwise.} \end{cases}$$

We obtain from (2.5), (2.6) and (2.8):

$$T_{4n-2}^0(f_{4n}) = \sum_{k=1}^{4n} \left[1 + (-1)^{k+1} \frac{a_{4n-1}(f)}{a_{4n-1}(\sigma_{4n})} \mathrm{sign} l_{4n}^k(0) \right] |l_{4n}^k(0)|. \tag{4.1}$$

On the other hand, the relations $\mathrm{sign} l_{4n}^k(0) = (-1)^k$, $1 \le k \le 2n$ and $\mathrm{sigm} l_{4n}^k(0) = (-1)^{k+1}$, $2n+1 \le k \le 4n$, show that f_{4n} is an even function, so we derive from (2.2) and (2.3):

$$a_{4n-1}(f_{4n}) = \sum_{k=1}^{4n} \tau_{4n}^k f_{4n}(x_{4n}^k) = \sum_{k=1}^{4n} \tau_{4n}^{4n-k+1} f_{4n}(x_{4n}^{4n-k+1})$$

$$= \sum_{k=1}^{4n} \frac{f_{4n}(-x_{4n}^k)}{u_{4n}'(-x_{4n}^k)} = -\sum_{k=1}^{4n} \frac{f_{4n}(x_{4n}^k)}{u_{4n}'(x_{4n}^k)} = -a_{4n-1}(f_{4n}) \tag{4.2}$$

because the nodes of J_n in $\mathcal{M}^{(\alpha)}$ are symmetric with respect to the origin and u_{4n} is an even function. So, we obtain from (4.2):

$$a_{4n-1}(f_{4n}) = 0, \quad n \ge 1. \tag{4.3}$$

The equalities (4.1) and (4.3) leads to:

$$T_{4n-2}^0(f_{4n}) = \Lambda_{4n}(0), \quad n \ge 1. \tag{4.4}$$

Using the estimates (3.1), (3.2) and taking into account that

$$|P_{2n}^{(\alpha)}(0)| \sim 1/\sqrt{n},$$

[10], we infer:

$$\Lambda_{4n}(0) - 1 \sim \ln n, \ \forall\, \alpha > -1. \tag{4.5}$$

The relations (4.4) and (4.5) give:

$$|T_{4n-2}^0(f_{4n})| \sim \ln n, \quad n \ge 2, \ \alpha > -1. \tag{4.6}$$

Finally, apply Theorem 1.1, with $X = C$, $Y = \mathbb{R}$, $A_n = T_n^0$ and remark that:

$$\sup\{\|A_n\| : \ n \ge 1\} \ge \sup\{\|T_{4n-2}^0\| : \ n \ge 1\} \ge \sup\{|T_{4n-2}^0(f_{4n})| : \ n \ge 1\},$$

which together with (4.6), proves the unboundedness of the set of norms $\{\|A_n\| : \ n \ge 1\}$. This completes the proof. $\square$

References

[1] Cheney, E. W., Light, W. A., *A Course in Approximation Theory*, Amer. Math. Soc., Pacific Grove, 2000.

[2] Cobzaş, S., Muntean, I., *Condensation of Singularities and Divergence Results in Approximation Theory*, J. Approx. Theory, **31**(1981), 138-153.

[3] Cobzaş, S., Muntean, I., *Superdense A.E. Unbounded Divergence of some Approximation Processes of Analysis*, Real Analysis Exchange, **25**(1999/2000), 501-512.

[4] Curtis Jr., P. C., *Convergence of Approximating Polynomials*, Proc. Amer. Math. Soc., **13**(1962), 385-387.

[5] Mitrea, A. I., *Convergence and superdense unbounded divergence in Approximation Theory*, Transilvania Press, Cluj-Napoca, Romania, 1998.

[6] Mitrea, A. I., *On the unbounded divergence of some projections operators in the discrete best approximation*, General Math., **18**(2010), no. 1, 81-88.

[7] Natanson, G. I., *Two-sided Estimates for Lebesgue functions of Lagrange interpolation processes based on Jacobi nodes* (Russian), Izv. Vysš. Ucebn. Zaved. (Matematika), **11**(1967), 67-74.

[8] Schönhage, A., *Approximationstheorie*, Berlin, Walter de Gruyter, 1971.

[9] Szabados, J., Vértesi, P., *Interpolation of Functions*, World Sci. Publ. Co., Singapore, 1990.

[10] Szegö, G., *Orthogonal Polynomials*, Amer. Math. Soc. Providence, 1975.

Alexandru I. Mitrea
Technical University of Cluj-Napoca
Department of Mathematics
28 Memorandumului Street
400114, Cluj-Napoca, Romania
e-mail: alexandru.ioan.mitrea@math.utcluj.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 473–487

On elliptic partial differential equations with random coefficients

Antje Mugler and Hans-Jörg Starkloff

Abstract. We consider stationary diffusion equations with random coefficients which cannot be bounded strictly away from zero and infinity by constants. We prove the existence of a unique solution to the corresponding weak formulation with different solution and test function spaces. Furthermore, the convergence of the Stochastic Galerkin solution is established under certain conditions.

Mathematics Subject Classification (2010): 35R60, 65N12, 65N30.

Keywords: Random partial differential equation, Stochastic Galerkin Methods, generalized polynomial chaos.

1. Introduction

In recent years there has been a growing interest in quantifying uncertainty in complex systems which are modeled via algebraic, ordinary or partial differential equations with random input data. For example, the stationary diffusion equation with a random coefficient is an instructive model problem. Thus, we consider the boundary value problem consisting of the random partial differential equation

$$-\nabla \cdot (\kappa \nabla u) = f$$

and some suitable boundary conditions. Thereby, the coefficient κ and also the forcing f are random functions. In previous works (see for example Babuška et al. [1, 3, 4] or Schwab et al. [5, 6, 14]) it is often assumed that there exist constants $\underline{\kappa}, \overline{\kappa} > 0$, such that

$$0 < \underline{\kappa} \leq \kappa(x, \omega) \leq \overline{\kappa} \qquad \text{a.e. and a.s.}$$

Then the theorem of Lax-Milgram can be used to prove the existence of a unique weak solution. In a first step towards a generalization of the problem setting Galvis and Sarkis [9] as well as Gittelson [11] investigate this random

This work was supported by the Deutsche Forschungsgemeinschaft Priority Programme 1324.

partial differential equation where the coefficient is modeled as a lognormal random field. That is, $\kappa(x) = \exp(G(x))$ with a Gaussian random field $G(x)$. In this case, however, there do not exist constants $\underline{\kappa}, \overline{\kappa} > 0$ as above and thus the Lax-Milgram theorem is not applicable. For this reason, the authors employ alternative techniques to prove the existence and uniqueness of the weak solution and to obtain a priori error estimates of the Stochastic Galerkin approximation to this solution. In the following we generalize these results to arbitrary random input fields which can be bounded by random variables $\kappa_{min}, \kappa_{max} > 0$ a.s., that is,

$$0 < \kappa_{min}(\omega) \le \kappa(x,\omega) \le \kappa_{max}(\omega) < \infty \qquad \text{a.e. and a.s.}$$

2. Setting and problem formulation

Let $D \subset \mathbb{R}^d,\ d \in \mathbb{N}$, be a bounded Lipschitz domain and $(\Omega, \mathfrak{A}, \mathbf{P})$ a probability space. We consider the following boundary value problem

$$\begin{aligned} -\nabla \cdot (\kappa(x,\omega)\nabla u(x,\omega)) &= f(x,\omega) && x \in D,\ \omega \in \Omega \\ u(x,\omega) &= 0 && x \in \partial D,\ \omega \in \Omega \end{aligned} \tag{2.1}$$

with random coefficient κ and random forcing f. We assume that the coefficient function $\kappa : D \times \Omega \to \mathbb{R}$ is a strongly measurable random variable with values in $\mathrm{L}^\infty(D)$ and that there exist real-valued random variables κ_{min} and κ_{max} such that

$$0 < \kappa_{min}(\omega) \le \kappa(x,\omega) \le \kappa_{max}(\omega) < \infty \qquad \text{a.e. and a.s.} \tag{2.2}$$

We define the pathwise bilinear form $b(\cdot,\cdot;\omega) : H^1(D) \times H^1(D) \to \mathbb{R}$ by

$$b(u,v;\omega) = \int\limits_D \kappa(x,\omega)\nabla u(x) \cdot \nabla v(x)\, dx$$

for $\omega \in \Omega$ and we denote by $\langle g, v\rangle_{H^{-1},\mathring{H}^1}$ the duality pairing between $g \in H^{-1}(D)$ and $v \in \mathring{H}^1(D)$. Now, assuming that f is a random variable with values in $H^{-1}(D)$, we consider a pathwise weak formulation of the boundary value problem:

Problem 2.1 (Pathwise Weak Formulation). Find a random variable $\tilde{u}$ with values in $\mathring{H}^1(D)$, such that

$$b(\tilde{u}(\omega), v; \omega) = \langle f(\omega), v\rangle_{H^{-1},\mathring{H}^1} \quad \text{for all } v \in \mathring{H}^1(D) \tag{2.3}$$

holds almost surely.

Remark 2.2. In Problem 2.1 we look for a random variable $\tilde{u}$ with values in $\mathring{H}^1(D)$, such that

$$\mathbf{P}\left(b(\tilde{u}(\omega), v; \omega) = \langle f(\omega), v\rangle_{H^{-1},\mathring{H}^1} \quad \text{for all } v \in \mathring{H}^1(D)\right) = 1. \tag{2.3a}$$

Due to the separability of $\mathring{H}^1(D)$ this problem is equivalent to the weaker problem formulation: Find a random variable $\tilde{u}$ with values in $\mathring{H}^1(D)$, such that for all $v \in \mathring{H}^1(D)$ there holds

$$\mathbf{P}\left(b(\tilde{u}(\omega), v; \omega) = \langle f(\omega), v\rangle_{H^{-1},\mathring{H}^1}\right) = 1. \tag{2.4}$$

Since every realization of the coefficient κ is bounded by assumption (2.2) and f is a random variable with values in $H^{-1}(D)$, by the theorem of Lax-Milgram (see e.g. [7] Theorem 2.7.7) there exists a mapping $\tilde{u} : \Omega \to \mathring{H}^1(D),\ \omega \mapsto \tilde{u}(\omega)$ satisfying

$$b(\tilde{u}(\omega), v; \omega) = \langle f(\omega), v\rangle_{H^{-1},\mathring{H}^1} \quad \text{for all } v \in \mathring{H}^1(D)$$

for almost all $\omega \in \Omega$. Furthermore, the estimate

$$\|\tilde{u}(\omega)\|_{H^1(D)} \leq C\frac{\|f(\omega)\|_{H^{-1}(D)}}{\kappa_{min}(\omega)} \qquad \text{a.s.} \tag{2.5}$$

holds, where $C > 0$ is a suitable constant which does not depend on $\omega \in \Omega$. This mapping $\tilde{u}$ is a.s. uniquely defined and measurable as is proved in the next Lemma.

Lemma 2.3. *Assume $\kappa : D \times \Omega \to \mathbb{R}$ is a strongly measurable random variable in $L^\infty(D)$ satisfying*

$$0 < \kappa_{min}(\omega) \leq \kappa(x, \omega) \leq \kappa_{max}(\omega) < \infty \quad \textit{a.e. and a.s.}$$

for real-valued random variables $\kappa_{min}, \kappa_{max}$, and f is a random variable with values in $H^{-1}(D)$. Then the mapping $\tilde{u} : \Omega \to \mathring{H}^1(D)$ is a random variable in $\mathring{H}^1(D)$ which is measurable with respect to the σ-algebra $\sigma(f, \kappa)$, generated by f and κ, and solves Problem 2.1.

Proof. From the assumptions on κ and f it follows that there exist sequences $(\kappa_n)_{n\in\mathbb{N}}$ and $(f_n)_{n\in\mathbb{N}}$ of $\sigma(f, \kappa)$-measurable, simple random variables with values in $\mathrm{L}^\infty(D)$ and $H^{-1}(D)$, respectively, satisfying

$$\|\kappa - \kappa_n\|_{\mathrm{L}^\infty(D)} \to 0, \text{ a.s.} \quad \text{and} \quad \|f - f_n\|_{H^{-1}(D)} \to 0, \text{ a.s.} \quad \text{for } n \to \infty.$$

Then the result follows immediately from the properties of the pathwise bilinear form b and the convergence of the simple random variables. □

In analogy to variational formulations of boundary value problems with purely deterministic input data we want to study also the corresponding variational formulation for random input data which is sometimes referred to as “stochastic variational formulation”. Such a formulation is obtained by defining a suitable bilinear form on a Hilbert space of random variables in $\mathring{H}^1(D)$, e.g. $a(u(\cdot), v(\cdot)) = \mathbf{E}_\mathbf{P} b(u(\cdot), v(\cdot); \cdot)$, and correspondingly by defining a linear form. However, since the coefficient κ is not bounded by constants but random variables we cannot directly use the Lax-Milgram theorem to prove existence and uniqueness of the weak solution. To address this problem we will define suitable solution and test function spaces to formulate the problem and to ensure the existence of a unique weak solution. The key observation is obtained as follows: Squaring inequality (2.5) and taking the expectation $\mathbf{E_P}$ with respect to the probability measure $\mathbf{P}$ yields

$$\mathbf{E_P}\left(\|\tilde{u}\|^2_{H^1(D)}\right) \leq C^2 \mathbf{E_P}\left(\frac{\|f\|^2_{H^{-1}(D)}}{\kappa^2_{min}}\right). \tag{2.6}$$

Hence, the pathwise solution $\tilde{u}$ is a second-order random variable in $\mathring{H}^1(D)$ if the second-order moment of the H^{-1}-norm of f, weighted with the reciprocal of the real-valued random variable κ^2_{min}, is finite. Thus, we need weighted function spaces in order to formulate the stochastic variational problem. Given a general real-valued random variable $\varrho > 0$ a.s. we introduce the spaces

$$\begin{aligned} U^m_\varrho &:= \mathrm{L}^2(\Omega, \mathfrak{A}, \varrho d\mathbf{P}; H^m(D)), \quad m \in \mathbb{Z}, \quad \text{and} \\ \mathring{U}^m_\varrho &:= \mathrm{L}^2(\Omega, \mathfrak{A}, \varrho d\mathbf{P}; \mathring{H}^m(D)), \quad m \in \mathbb{N}_0, \end{aligned}$$

where the ϱ-weighted L^2-spaces are defined by

$$\mathrm{L}^2(\Omega, \mathfrak{A}, \varrho d\mathbf{P}; V) := \left\{\xi : \Omega \to V \ \text{ measurable } : \mathbf{E_P}\left(\|\xi\|^2_V \varrho\right) < \infty\right\}$$

with $V = H^m(D)$ or $\mathring{H}^m(D)$, respectively. Endowing the spaces U^m_ϱ and $\mathring{U}^m_\varrho$ with the inner product

$$(u, v)_{U^m_\varrho} = \mathbf{E_P}\left((u, v)_{H^m(D)}\varrho\right), \quad u, v \in U^m_\varrho$$

and the induced norm

$$\|u\|_{U^m_\varrho} = \sqrt{\mathbf{E_P}\left(\|u\|^2_{H^m(D)}\varrho\right)}, \quad u \in U^m_\varrho$$

these spaces are also Hilbert spaces and there exist isomorphisms to the corresponding tensor product spaces (see e.g. [13])

$$U^m_\varrho \cong H^m(D) \otimes \mathrm{L}^2(\Omega, \mathfrak{A}, \varrho d\mathbf{P}) \quad \text{and} \quad \mathring{U}^m_\varrho \cong \mathring{H}^m(D) \otimes \mathrm{L}^2(\Omega, \mathfrak{A}, \varrho d\mathbf{P}),$$

if $\mathrm{L}^2(\Omega, \mathfrak{A}, \varrho d\mathbf{P})$ is separable. Furthermore, we note that the seminorm

$$|u|_{U^1_\varrho} = \sqrt{\mathbf{E_P}\left(|u|^2_{H^1(D)}\varrho\right)} = \sqrt{\int\limits_{D\times\Omega} |\nabla u(x,\omega)|^2 \varrho(\omega)\, dx\, d\mathbf{P}(\omega)}$$

is equivalent to the norm $\|\cdot\|_{U^1_\varrho}$ in $\mathring{U}^1_\varrho$ and that the dual space of $\mathring{U}^m_\varrho$ can be identified with the space $U^{-m}_{\varrho^{-1}}$. For convenience we denote by U^m or $\mathring{U}^m$ the spaces $H^m(D) \otimes \mathrm{L}^2(\Omega, \mathfrak{A}, \mathbf{P})$ or $\mathring{H}^m(D) \otimes \mathrm{L}^2(\Omega, \mathfrak{A}, \mathbf{P})$, respectively. On occasion we will replace $\mathbf{P}$ by another probability measure $\mathbf{Q}$ and write $U^m_{\mathbf{Q}} := H^m(D) \otimes \mathrm{L}^2(\Omega, \mathfrak{A}, \mathbf{Q})$ and $\mathring{U}^m_{\mathbf{Q}} := \mathring{H}^m(D) \otimes \mathrm{L}^2(\Omega, \mathfrak{A}, \mathbf{Q})$.

Then for a given $f \in U^{-1}_{\frac{1}{\kappa^2_{min}}}$ the stochastic weak formulation reads as follows:

Problem 2.4 (Stochastic Weak Formulation). Find $\hat{u} \in \mathring{U}^1$, such that

$$a(\hat{u}, v) = \langle f, v\rangle \quad \text{for all } v \in \mathring{U}^1_{\kappa^2_{min}}, \tag{2.7}$$

where the bilinear form a is given by

$$a(u,v) = \mathbf{E_P}\left(\int_D \kappa(x)\nabla u(x)\cdot\nabla v(x)\,dx\right) = \int_\Omega b(u(\omega),v(\omega);\omega)\,d\mathbf{P}(\omega) \quad (2.8)$$

and the duality pairing between $f \in U^{-1}_{\frac{1}{\kappa^2_{min}}}$ and $v \in \mathring{U}^1_{\kappa^2_{min}}$ is given by

$$\langle f,v\rangle = \mathbf{E_P}\left(\langle f,v\rangle_{H^{-1},\mathring{H}^1}\right) = \int_\Omega \langle f(\omega),v(\omega)\rangle_{H^{-1},\mathring{H}^1}\,d\mathbf{P}(\omega).$$

It is important to note that the solution and test function spaces are now different spaces. Furthermore, the domain of the bilinear form a is a proper subset of $\mathring{U}^1 \times \mathring{U}^1_{\kappa^2_{min}}$, i.e., the bilinear form a is not defined or finite for all pairs $(u,v) \in \mathring{U}^1 \times \mathring{U}^1_{\kappa^2_{min}}$. Thus, an implicit requirement of the weak formulation is to find a solution $\hat{u}$ such that the related bilinear form $a(\hat{u},\cdot)$ is defined and finite for all test functions.

3. Existence and uniqueness of weak solution

In this section, we will present two alternative proofs of existence and uniqueness of a solution to the weak formulation (2.7). Both approaches have benefits and drawbacks but when combined appropriately they are a powerful tool to study weak solutions and their properties. First we state a theorem which is a generalization of the Lax-Milgram theorem where the bilinear form is not defined on a cartesian product.

Theorem 3.1. *Let Hilbert spaces X_1, X_2, Y_1, Y_2 with dense and continuous embeddings $X_2 \subset X_1$ and $Y_2 \subset Y_1$ and a bilinear form $a : X_1 \times Y_1 \supsetneq \mathcal{D}_a \to \mathbb{R}$ be given such that*

(i) the restricted bilinear forms $a\big|_{X_1\times Y_2} : X_1 \times Y_2 \to \mathbb{R}$ and $a\big|_{X_2\times Y_1} : X_2 \times Y_1 \to \mathbb{R}$ are continuous,

(ii) there holds the inf-sup condition with a constant $c > 0$

$$\inf_{u\in X_1\setminus\{0\}}\sup_{v\in Y_1\setminus\{0\}} \frac{|a(u,v)|}{\|u\|_{X_1}\|v\|_{Y_1}} \geq c > 0, \qquad \textit{and}$$

(iii) for any $v \in Y_1 \setminus \{0\}$ there exists $u \in X_2$ such that $a(u,v) > 0$.

Then for any $f \in Y_1^*$ there exists a unique $u \in X_1$ satisfying

$$a(u,v) = \langle f,v\rangle \qquad \textit{for all } v \in Y_1.$$

Proof. The operator $T_a : X_1 \to Y_2^*$, $u \mapsto a(u,\cdot)$, is linear and continuous. The restricted operator $\hat{T}_a : X_1 \supsetneq \mathcal{D}(\hat{T}_a) \to Y_1^* \subset Y_2^*$ associated with T_a is densely defined, since $X_2 \subset \mathcal{D}(\hat{T}_a) \subset X_1$ is densely embedded, and injective and closed, because of the inf-sup condition (ii). Therefore it follows with Banach's closed range theorem (see e.g. [17] p. 205) that

$$\mathcal{R}(\hat{T}_a) = \mathcal{N}(\hat{T}_a^*)^\perp$$

where $\hat{T}_a^*$ is the adjoint operator of $\hat{T}_a$. Condition (iii) yields $\mathcal{N}(\hat{T}_a^*) = \{0\}$, thus $\mathcal{R}(\hat{T}_a) = Y_1^*$, which completes the proof. □

Corollary 3.2. *For any $f \in U^{-1}_{\frac{1}{\kappa^2_{min}}}$ there exists a unique $\hat{u} \in \mathring{U}^1$ satisfying the stochastic weak formulation* (2.7) *and the estimate*

$$\|\hat{u}\|_{U^1} \leq C \|f\|_{U^{-1}_{\frac{1}{\kappa^2_{min}}}} .$$

Proof. The Hilbert spaces $X_1 = \mathring{U}^1, X_2 = \mathring{U}^1_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}, Y_1 = \mathring{U}^1_{\kappa^2_{min}}$ and $Y_2 = \mathring{U}^1_{\kappa^2_{max}}$ and the bilinear form a defined in (2.8) satisfy all conditions in Theorem 3.1. The continuous and dense embeddings and the continuity of the bilinear forms $a|_{\mathring{U}^1 \times \mathring{U}^1_{\kappa^2_{max}}} : \mathring{U}^1 \times \mathring{U}^1_{\kappa^2_{max}} \to \mathbb{R}$ and $a|_{\mathring{U}^1_{\frac{\kappa^2_{max}}{\kappa^2_{min}}} \times \mathring{U}^1_{\kappa^2_{min}}} : \mathring{U}^1_{\frac{\kappa^2_{max}}{\kappa^2_{min}}} \times \mathring{U}^1_{\kappa^2_{min}} \to \mathbb{R}$ follow immediately from the definition of the spaces. To verify the inf-sup condition (ii), we define for $u \in \mathring{U}^1$ the random variable v_R with values in $\mathring{H}^1(D)$ by

$$v_R := \begin{cases} \frac{u}{\kappa_{min}}, & \frac{\kappa_{max}}{\kappa_{min}} \leq R, \\ 0, & \text{otherwise}, \end{cases}$$

and denote by B_R the set

$$B_R := \left\{ \omega \in \Omega \, : \, \frac{\kappa_{max}(\omega)}{\kappa_{min}(\omega)} \leq R \right\}.$$

Thus we obtain $v_R \in \mathring{U}^1_{\kappa^2_{min}}$, since

$$|v_R|^2_{U^1_{\kappa^2_{min}}} = \int_{B_R} |u(\omega)|^2_{H^1(D)} \, d\mathbf{P}(\omega) \leq |u|^2_{U^1} < \infty,$$

and by assumption (2.2) on the coefficient κ there holds

$$|a(u, v_R)| = \int_{D \times B_R} \frac{\kappa(x,\omega)}{\kappa_{min}(\omega)} |\nabla u(x,\omega)|^2 \, dx \, d\mathbf{P}(\omega) \geq \int_{B_R} |u(\omega)|^2_{H^1(D)} \, d\mathbf{P}(\omega).$$

Since $\mathbf{P}\left(\Omega \setminus \bigcup_{R>0} B_R\right) = 0$, there exists for every $\delta > 0$ a $R > 0$ such that

$$\int_{B_R} |u(\omega)|^2_{H^1(D)} \, d\mathbf{P}(\omega) \geq (1 - \delta) |u|^2_{U^1}$$

and thus

$$\sup_{v\in \mathring{U}^1_{\kappa^2_{min}}\setminus\{0\}} \frac{|a(u,v)|}{|v|_{U^1_{\kappa^2_{min}}}} \geq \frac{|a(u,v_R)|}{|v_R|_{U^1_{\kappa^2_{min}}}} \geq \frac{(1-\delta)|u|^2_{U^1}}{|u|_{U^1}} = (1-\delta)|u|_{U^1}.$$

Because $\delta > 0$ can be chosen arbitrarily the inf-sup condition holds with constant $c = 1$. Condition (iii) is satisfied, since for any $v \in \mathring{U}^1_{\kappa^2_{min}} \setminus \{0\}$ we can define

$$u_R := \begin{cases} v\kappa_{min}, & \frac{\kappa_{max}}{\kappa_{min}} \leq R, \\ 0, & \text{otherwise,} \end{cases} \qquad \text{and the set } B_R \text{ as above,}$$

which satisfies $u_R \in \mathring{U}^1$ and $a(u_R, v) > 0$ for R large enough. Hence, by Theorem 3.1 the statement follows. □

Obviously, Corollary 3.2 is also true for problems with other boundary conditions as long as the seminorm is a norm in the corresponding function spaces.

An alternative method to prove existence and uniqueness of the solution to Problem 2.4 where the coefficient κ is a lognormal random field, is given in the work of Gittelson [11]. For this special case it can be shown that the unique pathwise solution $\tilde{u}$ is also the unique solution of the stochastic variational problem if it belongs to the solution space. Below we prove an analogous result for the more general assumptions (2.2) on the random coefficient.

Theorem 3.3. *For $f \in U^{-1}_{\frac{1}{\kappa^2_{min}}}$ the unique solution $\tilde{u}$ of Problem* 2.1 *belongs to $\mathring{U}^1$ and it solves also Problem* 2.4. *Furthermore, any solution $\hat{u} \in \mathring{U}^1$ of Problem* 2.4 *is $\sigma(f,\kappa)$-measurable and there holds*

$$\hat{u}(x,\omega) = \tilde{u}(x,\omega) \qquad \textit{a.e. and a.s.}$$

Proof. Recalling that $f \in U^{-1}_{\frac{1}{\kappa^2_{min}}}$ and utilizing the estimate (2.6) we obtain

$$\|\tilde{u}\|^2_{U^1} = \mathbf{E_P}\|\tilde{u}\|^2_{H^1(D)} \leq C^2 \mathbf{E_P}\frac{\|f\|^2_{H^{-1}(D)}}{\kappa^2_{min}} = C^2\|f\|^2_{U^{-1}_{\frac{1}{\kappa^2_{min}}}} < \infty.$$

Since $\tilde{u}$ satisfies equation (2.3), there holds for all $v \in \mathring{U}^1_{\kappa^2_{min}}$

$$b(\tilde{u}(\omega), v(\omega); \omega) = \langle f(\omega), v(\omega)\rangle_{H^{-1},\mathring{H}^1} \qquad \text{a.s.}$$

Taking the expectation yields $a(\tilde{u}, v) = \langle f, v\rangle$ for all $v \in \mathring{U}^1_{\kappa^2_{min}}$ and hence $\tilde{u}$ solves Problem 2.4. Now, we consider a random variable $\hat{u} \in \mathring{U}^1$ satisfying

$$a(\hat{u}, v) = \langle f, v\rangle \qquad \text{for all } v \in \mathring{U}^1_{\kappa^2_{min}}.$$

Then we define for $w \in \overset{\circ}{H}{}^1(D)$ and $A \in \mathfrak{A}$ the functions $v_{w,A}(x,\omega) := w(x)\frac{\mathbf{1}_A(\omega)}{\kappa_{min}(\omega)}$. It follows $v_{w,A} \in \overset{\circ}{U}{}^1_{\kappa^2_{min}}$ and we get

$$\mathbf{E_P}\frac{\mathbf{1}_A}{\kappa_{min}}b(\hat{u},w;\cdot) = a(\hat{u},v_{w,A}) = \langle f, v_{w,A}\rangle = \mathbf{E_P}\frac{\mathbf{1}_A}{\kappa_{min}}\langle f,w\rangle_{H^{-1},\overset{\circ}{H}{}^1}.$$

Since $A \in \mathfrak{A}$ can be chosen arbitrarily this implies for any $w \in \overset{\circ}{H}{}^1(D)$

$$b(\hat{u}(\omega),w;\omega) = \langle f(\omega),w\rangle_{H^{-1},\overset{\circ}{H}{}^1} \quad \text{a.s.}$$

Hence, the random variable $\hat{u}$ with values in $\overset{\circ}{H}{}^1(D)$ solves problem (2.4) and since its solution is almost surely unique and $\sigma(f,\kappa)$-measurable (cf. Lemma 2.3), there holds

$$\hat{u}(x,\omega) = \tilde{u}(x,\omega) \quad \text{a.e. and a.s.},$$

i.e., the random variable $\hat{u}$ is measurable with respect to the σ-algebra $\sigma(f,\kappa)$. □

4. Stochastic Galerkin discretization

Let $\xi := (\xi_i)_{i\in I_\xi}$ with index set $I_\xi \subseteq \mathbb{N}$ be a sequence of real-valued so called "basic" random variables, such that there are measurable functions $\kappa_\xi,\ f_\xi : D\times\mathbb{R}^{|I_\xi|} \to \mathbb{R}$ satisfying

$$\kappa(x,\omega) = \kappa_\xi(x,\xi(\omega)) \qquad \text{and} \qquad f(x,\omega) = f_\xi(x,\xi(\omega)) \quad \text{a.e. and a.s.}$$

Thereby the index set I_ξ can be finite, i.e., $I_\xi = \{1,\dots,M\}$, $M\in\mathbb{N}$, or the set of the natural numbers, i.e., $I_\xi = \mathbb{N}$. Sequences of basic random variables can be obtained with the help of Karhunen-Loève expansions (see e.g. [12]) or other series expansions (see e.g. [10]) of the input data.

Then according to Theorem 3.3 the solution $\hat{u}$ of variational formulation (2.7) belongs to $\mathrm{L}^2(\Omega,\sigma(\xi),\mathbf{P};\overset{\circ}{H}{}^1(D))$ since κ and f are $\sigma(\xi)$-measurable. In the following we assume that the random variable $\xi = (\xi_i)_{i\in I_\xi}$ on the probability space $(\Omega,\mathfrak{A},\mathbf{P})$ has the distribution $F^{\mathbf{P}}_\xi$ and that any ξ_i, $i\in I_\xi$, possesses finite moments of arbitrary order, i.e., $\mathbf{E_P}|\xi_i|^n < \infty$, $n\in\mathbb{N}$, and a continuous distribution function $F^{\mathbf{P}}_{\xi_i}$.

In order to apply the Stochastic Galerkin Method we define the space

$$U_{N,K,p} := U_p \otimes U_{N,K} \subset \overset{\circ}{U}{}^1$$

which serves as solution space for the Stochastic Galerkin approximation. The space U_p is a finite-dimensional subspace of $\overset{\circ}{H}{}^1(D)$ obtained by a uniform p version of the Finite Element Method and $U_{N,K}$ is a finite-dimensional subspace of $\mathrm{L}^2(\Omega,\sigma(\xi_1,\dots,\xi_K),\mathbf{P}) \subseteq \mathrm{L}^2(\Omega,\sigma(\xi),\mathbf{P})$ with $\{1,\dots,K\}\subseteq I_\xi$. Since we want to use generalized polynomial chaos (see e.g. [15, 16]), i.e. polynomials in the underlying basic random variables ξ, we construct the finite dimensional space $U_{N,K}$ as follows,

$$U_{N,K} := \operatorname{span}\left\{\xi^\alpha := \prod_{i\in I_\xi}\xi_i^{\alpha_i},\ \alpha\in\Lambda_{N,K}\right\}.$$

We choose the index set

$$\Lambda_{N,K} \subset \Lambda := \{\alpha \in \mathbb{N}_0^{|I_\xi|} \,:\, \alpha \text{ has only finitely many non-zero entries}\}$$

such that the total degree of the multivariate polynomials is bounded,

$$\Lambda_{N,K} = \{\alpha \in \Lambda \,:\, \alpha_i = 0 \;\forall\; i > K,\; |\alpha| \le N\;\}, \qquad |\alpha| := \sum_{i \in I_\xi} \alpha_i.$$

As discretized test function space we choose

$$V_{N,K,p} := \left\{ \frac{u}{\kappa_{min}} \,:\, u \in U_{N,K,p} \right\}.$$

Then for a given $f \in U^{-1}_{\frac{1}{\kappa^2_{min}}}$ the discrete version of the weak formulation (2.7) reads as follows:

Problem 4.1 (Discrete Weak Formulation). Find $\hat{u}_{N,K,p} \in U_{N,K,p}$, such that

$$a(\hat{u}_{N,K,p}, v) = \langle f, v\rangle \quad \text{for all } v \in V_{N,K,p}. \tag{4.1}$$

The existence of a unique Stochastic Galerkin solution $\hat{u}_{N,K,p} \in U_{N,K,p}$ to problem (4.1) can be proved under the assumptions in the following lemma.

Lemma 4.2. *If $\frac{\kappa^2_{max}}{\kappa^2_{min}} \in L^r(\Omega, \mathfrak{A}, \mathbf{P})$ for some $r > 1$ then for any $f \in U^{-1}_{\frac{1}{\kappa^2_{min}}}$ there exists a unique $\hat{u}_{N,K,p} \in U_{N,K,p}$ such that*

$$a(\hat{u}_{N,K,p}, v) = \langle f, v\rangle \qquad \textit{for all } v \in V_{N,K,p}.$$

Proof. The result follows from Theorem 3.1 with the Hilbert spaces

$$X_1 = U_{N,K,p} \subset \mathring{U}^1, \qquad X_2 = U_{N,K,p} \subset \mathring{U}^1_{\frac{\kappa^2_{max}}{\kappa^2_{min}}},$$

$$Y_1 = V_{N,K,p} \subset \mathring{U}^1_{\kappa^2_{min}} \quad \text{and} \quad Y_2 = V_{N,K,p} \subset \mathring{U}^1_{\kappa^2_{max}}$$

due to $\frac{\kappa^2_{max}}{\kappa^2_{min}} \in \mathrm{L}^r(\Omega, \mathfrak{A}, \mathbf{P})$ for some $r > 1$ and a discrete version of the inf-sup condition for the bilinear form a. □

Now, we want to investigate the approximation error of this Stochastic Galerkin solution $\hat{u}_{N,K,p}$. Employing the discrete inf-sup condition we get a quasi-optimal result for the Galerkin solution, i.e., the error can be bounded by a best approximation error in another – a stronger – norm.

Lemma 4.3. *If $\frac{\kappa^2_{max}}{\kappa^2_{min}} \in L^r(\Omega, \mathfrak{A}, \mathbf{P})$ for some $r > 1$ and $\hat{u} \in \mathring{U}^1_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}$ then the following estimate holds*

$$|\hat{u} - \hat{u}_{N,K,p}|_{U^1} \le \widetilde{C} \inf_{z \in U_{N,K,p}} |\hat{u} - z|_{U^1_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}}$$

with a constant $\widetilde{C} > 0$ (independent of N, K and p) for the solutions $\hat{u}$ and $\hat{u}_{N,K,p}$ of the weak formulation (2.7) *and the discrete weak formulation (*4.1*), respectively.*

Proof. Utilizing $a(\hat{u} - z, v) = a(\hat{u}_{N,K,p} - z, v)$ for all $v \in V_{N,K,p}$ and the discrete inf-sup condition we obtain

$$\begin{aligned}|\hat{u} - \hat{u}_{N,K,p}|_{U^1} &\leq |\hat{u} - z|_{U^1} + |\hat{u}_{N,K,p} - z|_{U^1} \\ &\leq |\hat{u} - z|_{U^1} + |\hat{u} - z|_{U^1_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}} \leq 2|\hat{u} - z|_{U^1_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}}\end{aligned}$$

for all $z \in U_{N,K,p}$. □

Consequently we measure the error in the stronger $U^1_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}$-norm and we assume the following.

Assumption 4.4. *Let* $q := \mathbf{E}_{\mathbf{P}} \frac{\kappa^2_{max}}{\kappa^2_{min}} < \infty$ *and assume* $\frac{\kappa^2_{max}}{\kappa^2_{min}}$ *is* $\sigma(\xi)$*-measurable, i.e., there exists a measurable transformation* $t_{\frac{\kappa^2_{max}}{\kappa^2_{min}}} : \mathbb{R}^{|I_\xi|} \to \mathbb{R}^+$ *with* $\frac{\kappa^2_{max}}{\kappa^2_{min}} = t_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}(\xi)$.

Then the measure $\mathbf{Q}$ with $d\mathbf{Q} = \frac{1}{q}\kappa^2_{max}\kappa^{-2}_{min}d\mathbf{P}$ is a probability measure. In the following we consider the function spaces $U^m_{\mathbf{Q}}$ and $\mathring{U}^m_{\mathbf{Q}}$ instead of $U^m_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}$ and $\mathring{U}^m_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}$, $m \in \mathbb{Z}$, which coincide with $U^m_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}$ and $\mathring{U}^m_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}$ but are much easier to handle due to the corresponding probability space $(\Omega, \mathfrak{A}, \mathbf{Q})$ at hand.

Corollary 4.5. *If* $\frac{\kappa^2_{max}}{\kappa^2_{min}} \in L^r(\Omega, \mathfrak{A}, \mathbf{P})$ *for some* $r > 1$ *and Assumption* 4.4 *is fulfilled there holds for* $\hat{u} \in \mathring{U}^1_{\mathbf{Q}}$ *with a suitable constant* $C > 0$ *(independent of* N, K *and* p*)*

$$|\hat{u} - \hat{u}_{N,K,p}|_{U^1} \leq C \inf_{z \in U_{N,K,p}} |\hat{u} - z|_{U^1_{\mathbf{Q}}} \tag{4.2}$$

for the solutions $\hat{u}$ *and* $\hat{u}_{N,K,p}$ *of the corresponding weak formulation* (2.7) *and discrete weak formulation* (4.1).

Proof. This result follows immediately from Lemma 4.3 and Assumption 4.4. □

By choosing a suitable $z \in U_{N,K,p}$ and applying the triangle inequality to the right-hand side of (4.2) we can identify different sources of the approximation error. To see this, we introduce some notations: We denote by

$$\Pi_{\mathring{U}^1_{\mathbf{Q}},N,K,p} : \mathring{U}^1_{\mathbf{Q}} \to U_{N,K,p}$$

the orthogonal projection onto $U_{N,K,p}$, and by

$$\Pi_{\mathring{U}^1_{\mathbf{Q}},N,K} : \mathring{U}^1_{\mathbf{Q}} \to \mathring{H}^1(D) \otimes U_{N,K}$$

the orthogonal projection onto $\mathring{H}^1(D) \otimes U_{N,K}$, both with respect to the $U^1_{\mathbf{Q}}$-norm. Assuming $\hat{u} \in \mathring{U}^1_{\mathbf{Q}}$ the approximation error of the Stochastic Galerkin approximation to the exact solution can be estimated using (4.2) with $z = \Pi_{\mathring{U}^1_{\mathbf{Q},N,K,p}} \hat{u}$ as

$$|\hat{u} - \hat{u}_{N,K,p}|_{U^1} \leq C \left[|\hat{u} - \Pi_{\mathring{U}^1_{\mathbf{Q},N,K}} \hat{u}|_{U^1_{\mathbf{Q}}} + |\Pi_{\mathring{U}^1_{\mathbf{Q},N,K}} \hat{u} - \Pi_{\mathring{U}^1_{\mathbf{Q},N,K,p}} \hat{u}|_{U^1_{\mathbf{Q}}} \right]. \quad (4.3)$$

Hence this error has two components, namely an approximation error due to discretizing in the stochastic dimension and an approximation error due to discretizing in the spatial dimension.

The spatial approximation error can be bounded using standard arguments from the theory of Finite Element Methods (FEMs). Here, we have employed a p version of the FEM (see e.g. [2]). Under the assumptions of Corollary 2.2 in [2] there holds the following.

Corollary 4.6. *If $\frac{\kappa^2_{max}}{\kappa^2_{min}} \in L^r(\Omega, \mathfrak{A}, \mathbf{P})$ for some $r > 1$ and Assumption 4.4 is satisfied then for $\hat{u} \in U^k_{\frac{\kappa^2_{max}}{\kappa^3_{min}}} \cap \mathring{U}^1_{\mathbf{Q}}$ with constant $\widetilde{C} > 0$ (independent of N, K, p and $\hat{u}$) there holds*

$$|\Pi_{\mathring{U}^1_{\mathbf{Q},N,K}} \hat{u} - \Pi_{\mathring{U}^1_{\mathbf{Q},N,K,p}} \hat{u}|_{U^1_{\mathbf{Q}}} \leq \widetilde{C} p^{-(k-1)} \|\hat{u}\|_{U^k_{\frac{\kappa^2_{max}}{\kappa^3_{min}}}}.$$

Proof. From Corollary 2.2 in [2] it follows

$$\sqrt{\kappa_{min}(\omega)} \left| \Pi_{\mathring{U}^1_{\mathbf{Q},N,K}} \hat{u}(\omega) - \Pi_{\mathring{U}^1_{\mathbf{Q},N,K,p}} \hat{u}(\omega) \right|_{H^1(D)} \leq \widetilde{C} p^{-(k-1)} \|\hat{u}(\omega)\|_{H^k(D)}$$

with a constant $\widetilde{C}$ independent of N, K, p, $\omega \in \Omega$ and $\hat{u}$. Squaring and taking the expectation $\mathbf{E}_{\mathbf{Q}}$ with respect to $\mathbf{Q}$ leads to

$$\mathbf{E}_{\mathbf{Q}} \left| \Pi_{\mathring{U}^1_{\mathbf{Q},N,K}} \hat{u} - \Pi_{\mathring{U}^1_{\mathbf{Q},N,K,p}} \hat{u} \right|^2_{H^1(D)} \leq \widetilde{C}^2 p^{-2(k-1)} \mathbf{E}_{\mathbf{Q}} \frac{\|\hat{u}\|^2_{H^k(D)}}{\kappa_{min}}.$$

□

We note that analogous results to Corollary 4.6 can be obtained for h or h-p versions of the FEM by using Theorem 2.1 in [2].

The first term on the right-hand side of inequality (4.3) can be estimated with the help of generalized polynomial chaos expansions. In view of Assumption 4.4 the random variable $\xi = (\xi_i)_{i \in I_\xi}$ as a random variable on the probability space $(\Omega, \mathfrak{A}, \mathbf{Q})$ has the distribution $F^{\mathbf{Q}}_\xi(dy) = \frac{1}{q} t_{\frac{\kappa^2_{max}}{\kappa^2_{min}}}(y) F^{\mathbf{P}}_\xi(dy)$. Assuming $\mathbf{E}_{\mathbf{Q}}|\xi_i|^n < \infty$ for all $i \in I_\xi$ and $n \in \mathbb{N}$ the multivariate orthonormal polynomials $\{q_\alpha(\xi),\ \alpha \in \Lambda\}$ in $\mathrm{L}^2(\Omega, \mathfrak{A}, \mathbf{Q})$ exist. Hence, in order to expand any random variable $u \in \mathrm{L}^2(\Omega, \sigma(\xi), \mathbf{Q}; \mathring{H}^1(D))$ in this generalized polynomial chaos the polynomials $\{q_\alpha(\xi),\ \alpha \in \Lambda\}$ have to be dense in $\mathrm{L}^2(\Omega, \sigma(\xi), \mathbf{Q})$. Some necessary conditions to establish this property are discussed in [8]. If the polynomials lie dense and $\hat{u} \in \mathrm{L}^2(\Omega, \sigma(\xi), \mathbf{Q}; \mathring{H}^1(D))$ then the solution possesses a generalized polynomial chaos expansion $\{q_\alpha(\xi),\ \alpha \in \Lambda\}$, i.e.,

$$\hat{u}(x,\omega) = \sum_{\alpha\in\Lambda} \hat{u}_\alpha(x) q_\alpha(\xi(\omega)), \quad \text{where} \quad \hat{u}_\alpha(x) = \mathbf{E}_{\mathbf{Q}} \hat{u}(x) q_\alpha(\xi).$$

Furthermore, the projection $\Pi_{\mathring{U}^1_{\mathbf{Q},N,K}} \hat{u}$ is given by the truncated expansion

$$\Pi_{\mathring{U}^1_{\mathbf{Q},N,K}} \hat{u}(x,\omega) = \sum_{\alpha\in\Lambda_{N,K}} \hat{u}_\alpha(x) q_\alpha(\xi(\omega)).$$

Corollary 4.7. *If the polynomials* $\{q_\alpha(\xi),\ \alpha \in \Lambda\}$ *are dense in* $L^2(\Omega, \sigma(\xi), \mathbf{Q})$ *and* $\hat{u} \in \mathring{U}^1_{\mathbf{Q}}$ *then the approximation error*

$$|\hat{u} - \Pi_{\mathring{U}^1_{\mathbf{Q},N,K}} \hat{u}|_{U^1_{\mathbf{Q}}} \to 0 \quad (K, N \to \infty).$$

Proof. The multivariate polynomials $\{q_\alpha(\xi),\ \alpha \in \Lambda\}$ form an orthonormal basis of $\mathrm{L}^2(\Omega, \sigma(\xi), \mathbf{Q})$ because they are dense in $\mathrm{L}^2(\Omega, \sigma(\xi), \mathbf{P})$. Since the weak solution $\hat{u}$ is $\sigma(\xi)$-measurable (according to Theorem 3.3) and

$$\bigcup_{N\geq 0,\, K\geq 1} \Lambda_{N,K} = \Lambda$$

there holds that $\quad \Pi_{\mathring{U}^1_{\mathbf{Q},N,K}} \hat{u} \to \hat{u} \quad$ in $\mathring{U}^1_{\mathbf{Q}}$ for $K \to \infty,\ N \to \infty$. □

Hence in view of Corollary 4.6 and Corollary 4.7 the approximation error $|\hat{u} - \hat{u}_{N,K,p}|_{U^1}$ converges to zero if the solution $\hat{u} \in U^2_{\frac{\kappa^2_{max}}{\kappa^3_{min}}} \cap \mathring{U}^1_{\mathbf{Q}}$ and the orthonormal polynomials $\{q_\alpha(\xi),\ \alpha \in \Lambda\}$ are complete in $\mathrm{L}^2(\Omega, \sigma(\xi), \mathbf{Q})$.

5. Numerical example

Now, we turn to a specific application, namely the approximation of the solution of an one-dimensional differential equation with random data. Consider the boundary value problem

$$\begin{aligned} -(\kappa(x,\omega) u'(x,\omega))' &= f(x), \quad x \in (0,1),\ \omega \in \Omega \\ u(0,\omega) &= 0, \quad \omega \in \Omega \\ \kappa(1,\omega) u'(1,\omega) &= F, \quad \omega \in \Omega \end{aligned}$$

where forcing $f \in H^{-1}(D)$ is a deterministic function, F a given constant and κ a strongly measurable random variable in $\mathrm{L}^\infty(D)$ satisfying

$$0 < \kappa_{min}(\omega) \leq \kappa(x,\omega) \leq \kappa_{max}(\omega) < \infty \qquad \text{a.e. and a.s.}$$

for some real-valued random variables κ_{min} and κ_{max}. Then the exact solution is given by

$$u(x,\omega) = \int_0^x \frac{1}{\kappa(y,\omega)} \left(F + \int_y^1 f(z)\, dz \right) dy.$$

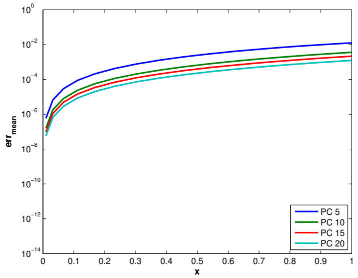
If the coefficient κ is modeled as an exponential function of the absolute value of one standard Gaussian distributed random variable, that is,

$$\kappa(x,\omega) := \exp(|\zeta(\omega)|x) \quad \text{with } \zeta \sim \mathcal{N}(0,1)$$

then κ is bounded by

$$0 < 1 \leq \kappa(x,\omega) \leq \exp(|\zeta(\omega)|) < \infty \qquad \text{a.e. and a.s.}$$

The random variable $\kappa_{max}^2/\kappa_{min}^2{=}\exp(2|\zeta|)$ is in $\mathrm{L}^r(\Omega,\mathfrak{A},\mathbf{P})$ for all $r{\geq}1$. As basic random variable we choose the standard Gaussian distributed random variable ζ, i.e., $\xi = \zeta$, and employ the Stochastic Galerkin Method using orthonormal polynomials, i.e. polynomial chaos, in ξ. Figure 1 shows the relative errors of the mean and second-order moment of the Stochastic Galerkin approximation to the exact solution as a function of the spatial variable x.



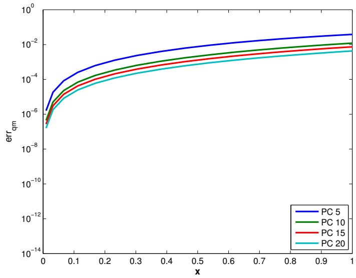


FIGURE 1. Relative errors of mean (left) and second moment (right) of the Stochastic Galerkin approximation to the solution with $f \equiv 1$ and $F = 1$ using polynomials of different orders in ξ.

Thereby we have chosen the forcing $f \equiv 1$ and the boundary value $F = 1$ and we use a p version of the Finite Element Method, precisely, a single Gauss-Lobatto-Legendre spectral finite element of degree $p = 20$ for the spatial discretization. In the stochastic dimension we use orthonormal polynomials in ξ up to degree 5, 10, 15 and 20. Obviously, the error decays, which agrees with the theory developed in Section 4. On the other hand it is also possible to choose as basic random variable $\eta = |\zeta|$, a chi-distributed random variable with one degree of freedom. Thus, we can use orthonormal polynomials, i.e. generalized polynomial chaos, in η within the Stochastic Galerkin Method, in particular orthonormal polynomials in η up to degree 2 and 5. In the spatial dimension we again use a single Gauss-Lobatto-Legendre spectral finite element of degree $p = 20$. In Figure 2 we observe that the associated relative errors of the mean and second-order moment tend to zero much faster than for the standard Gaussian basic random variable ξ. Notably, we obtain much better approximation results by using polynomials up to order 2 and 5 in $\eta = |\zeta|$ as compared to polynomials up to order 20 in $\xi = \zeta$. Hence, the approximation error, more precisely the rate of convergence, and thus the

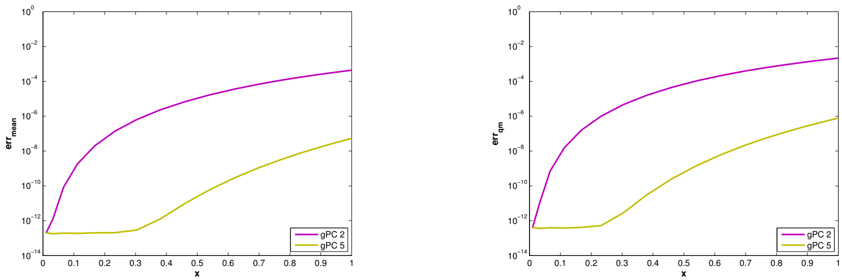


Figure 2. Relative errors of mean (left) and second moment (right) of the Stochastic Galerkin approximation to the solution with $f \equiv 1$ and $F = 1$ using polynomials of different orders in η.

approximation quality depends on the set of basic random variables. This relation is currently being investigated in ongoing research.

References

[1] Babuška, I., Nobile, F., Tempone, R., *A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data*, SIAM J. Numer. Anal., **45**(2007), no. 3, 1005–1034.

[2] Babuška, I., Suri, M., *The p and h-p versions of the finite element method, basic priniciples and properties*, SIAM Review, **36**(1994), no. 4, 578–632.

[3] Babuška, I., Tempone, R., Zouraris, G. E., *Galerkin Finite Element Approximations of Stochastic Elliptic Partial Differential Equations*, SIAM J. Numer. Anal., **42**(2004), no. 2, 800–825.

[4] Babuška, I., Tempone, R., Zouraris, G. E., *Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation*, Comput. Methods Appl. Mech. Engrg., **194**(2005), 1251–1294.

[5] Bieri, M., Andreev, R., Schwab, C., *Sparse tensor discretization of elliptic SPDEs*, SIAM J. Sci. Comput., **31**(2009), no. 6, 4281–4304.

[6] Bieri, M., Schwab, C., *Sparse high order FEM for elliptic sPDEs*, Comput. Methods Appl. Mech. Engrg., **198**(2009), no. 13-14, 1149–1170.

[7] Brenner, S. C., Scott, L. R., *The Mathematical Theory of Finite Element Methods*, 2nd ed., Texts in Applied Mathematics, vol. 15, Springer-Verlag, New York, 2002.

[8] Ernst, O., Mugler, A., Ullmann, E., Starkloff, H.-J., *On the convergence of generalized polynomial chaos*, http://www.dfg-spp1324.de/publications.php, 2010.

[9] Galvis, J., Sarkis, M., *Approximating infinity-dimensional stochastic Darcy's Equations without uniform ellipticity*, SIAM J. Numer. Anal., **47**(2009), no. 5, 3624–3651.

[10] Gittelson, C.J., *Representation of Gaussian fields in series with independent coefficients*, ETH Zürich, Research Report (2010), no. 2010-15.

[11] Gittelson, C.J., *Stochastic Galerkin discretization of the log-normal isotropic diffusion problem*, Math. Models Methods Appl. Sci., **20**(2010), no. 2, 237–263.

[12] Loève, M., *Probability Theory II. 4th Edition*, Springer-Verlag, New York, Heidelberg, Berlin, 1978.

[13] Reed, M., Simon, B., *Methods of Modern Mathematical Physics. 1. Functional analysis*, Academic Press, New York, 1972.

[14] Todor, R.A., Schwab, C., *Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients*, IMA Journal of Numerical Analysis, **27**(2007), 232–261.

[15] Xiu, D., Karniadakis, G.E., *Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos*, Comput. Methods Appl. Mech. Engrg., **191**(2002), no. 43, 4927–4948.

[16] Xiu, D., Karniadakis, G.E., *The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations*, SIAM J. Sci. Comput., **24**(2002), no. 2, 619–644.

[17] Yosida, K., *Functional Analysis*, Reprint of the 1980 ed., Classics in Mathematics, Springer, Berlin, 1995.

Antje Mugler
University of Applied Sciences, Zwickau
Faculty of Physical Technology/Computer Sciences, Group of Mathematics
2A, Dr. Friedrichs-Ring, PSF 201037
08012 Zwickau
Germany
e-mail: `antje.mugler@fh-zwickau.de`

Hans-Jörg Starkloff
University of Applied Sciences, Zwickau
Faculty of Physical Technology/Computer Sciences, Group of Mathematics
2A, Dr. Friedrichs-Ring, PSF 201037
08012 Zwickau
Germany
e-mail: `hans.joerg.starkloff@fh-zwickau.de`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 489–495

Note on q-Bernstein-Schurer operators

Carmen-Violeta Muraru

Abstract. In this paper, we introduce a generalization of the Bernstein-Schurer operators based on q-integers and get a Bohman-Korovkin type approximation theorem of these operators. We also compute the rate of convergence by using the first modulus of smoothness.

Mathematics Subject Classification (2010): 41A36.

Keywords: q-integers, positive linear operator, Bernstein operator, modulus of continuity.

1. Preliminaries

Lupaş [18] introduced in 1987 a q-type of the Bernstein operators and in 1997 another generalization of these operators based on q-integers was introduced by Phillips [20]. He obtained the rate of convergence and a Voronovskaja type asymptotic formula for the new Bernstein operators. After this, many authors studied new classes of q-generalized operators. To show the extend of this research direction, we mention in the following some achievements in this field. In [5] Bărbosu introduced a Stancu type generalization of two dimensional Bernstein operators based on q-integers. In [1] O. Agratini introduced a new class of q-Bernstein-type operators which fix certain polynomials and studied the limit of iterates of Lupaş q-analogue of the Bernstein operators. In [4] Aral and Doğru obtained the uniform approximation of q-Bleimann-Butzer-Hahn (BBH) operators and in [9] O. Doğru and V. Gupta studied the monotonicity properties and the Voronovskaja type asymptotic estimate of these operators. See also the recent paper [2].

T. Trif [21] investigated Meyer-König and Zeller (MKZ) operators based on q-integers. Some approximation properties of q-MKZ operators were investigated by W. Heping in [16]. O. Doğru and O. Duman introduced also a new generalization of Meyer-König and Zeller operators and studied some statistical approximation properties in [7]. O. Doğru and Gupta [8] constructed a q-type generalization of Meyer-König and Zeller operators in bivariate case. A new q-generalization of Meyer-König and Zeller type operators was constructed by Doğru and Muraru for improve the rate of convergence, see [10].

O. Doğru and M. Orkcu proved in [11] that a new modification of q-MKZ operators provides a better estimation on the $[\alpha_n, 1] \subset [1/2, 1)$ by means of the modulus of continuity.

An extension in q-Calculus of Szász-Mirakyan operators was constructed by Aral [3] who formulated also a Voronovskaya theorem related to q-derivatives for these operators.

Durrmeyer type generalization of the operators based on q-integers was studied by Derriennic in [6]. Gupta and Heping introduced a q-analoque of Bernstein-Durrmeyer operators in [13] and in 2009 Gupta and Finta [14] studied some local and global approximation properties for q-Durrmeyer operators. See also [12]. In [15] Gupta and Radu constructed a q-analoque of Baskakov-Kantorovich operators and investigated their weighted statistical approximation properties. Also, N. Mahmudov introduced in [19] new classes of q-Baskakov and q-Baskakov-Kantorovich operators.

First of all, we recall elements of q-Calculus, see, e.g., [17]. For any fixed real number $q > 0$, the q-integer $[k]_q$, for $k \in \mathbb{N}$ is defined as

$$[k]_q = \begin{cases} (1-q^k)/(1-q), & q \neq 1, \\ k, & q = 1. \end{cases}$$

Set $[0]_q = 0$. The q-factorial $[k]_q!$ and q-binomial coefficients $\begin{bmatrix} n \\ k \end{bmatrix}_q$ are defined as follows

$$[k]_q! = \begin{cases} [k]_q[k-1]_q \dots [1]_q, & k = 1, 2, \dots, \\ 1, & k = 0, \end{cases}$$

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{[n]_q!}{[k]_q![n-k]_q!} \quad (0 \le k \le n).$$

The q-analogue of $(x-a)^n$ is the polynomial

$$(x-a)_q^n = \begin{cases} 1 & \text{if} \quad n = 0, \\ (x-a)(x-qa)\dots(x-q^{n-1}a) & \text{if} \quad n \ge 1. \end{cases}$$

$C([a,b])$ represents the space of all real valued continuous functions defined on $[a,b]$. The space is endowed with usual norm $\|\cdot\|$ given by

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

Let $p \in \mathbb{N}$ be fixed. In 1962 Schurer [22] introduced and studied the operators $\widetilde{B}_{m,p} : C([0, p+1]) \to C([0,1])$ defined for any $m \in \mathbb{N}$ and any function $f \in C([0, p+1])$ as follows

$$\widetilde{B}_{m,p}(f;x) = \sum_{k=0}^{m+p} \binom{m+p}{k} x^k (1-x)^{m+p-k} f\left(\frac{k}{m}\right), \quad x \in [0,1].$$

Our aim is to introduce a q-analogue of the above operators. We investigate the approximation properties of this class and we estimate the rate of convergence by using modulus of continuity.

2. Construction of generalized q-Bernstein-Schurer and approximation properties

Throughout the paper we consider $q \in (0,1)$.
For any $m \in \mathbb{N}$ and $f \in C([0,p+1])$, p is fixed, we construct the class of generalized q-Bernstein-Schurer operators as follows

$$\widetilde{B}_{m,p}(f;q;x) = \sum_{k=0}^{m+p} \begin{bmatrix} m+p \\ k \end{bmatrix}_q x^k \prod_{s=0}^{m+p-k-1} (1-q^s x) f\left(\frac{[k]_q}{[m]_q}\right), \; x \in [0,1]. \tag{2.1}$$

From here on, an empty product is taken to be equal 1. Clearly, the operator defined by (2.1) is linear and positive.

Lemma 2.1. *Let $B_{m,p}(\cdot;q;\cdot)$ be given by (2.1). The following identities*

1° $\widetilde{B}_{m,p}(e_0;q;x) = 1,$

2° $\widetilde{B}_{m,p}(e_1;q;x) = \dfrac{x[m+p]_q}{[m]_q},$

3° $\widetilde{B}_{m,p}(e_2;q;x) = \dfrac{[m+p]_q}{[m]_q^2}([m+p]_q x^2 + x(1-x))$

hold, where $e_j(x) = x^j$, $j = 0,1,2$.

Proof. 1° We use the known identity

$$\sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q x^k (1-x)_q^{n-k} = 1,$$

which can be proved by induction with respect to n. Actually, the left hand side represents $(B_{n,q}e_0)(x)$ where $B_{n,q}$ is the q-analogue of Bernstein operator introduced by G. M. Phillips [20]. Phillips proved $B_{n,q}e_0 = e_0$.

In the above we choose $n := m+p$.

Since

$$(1-x)_q^{m+p-k} = \prod_{s=0}^{m+p-k-1} (1-q^s x),$$

we get

$$\sum_{k=0}^{m+p} \begin{bmatrix} m+p \\ k \end{bmatrix}_q x^k \prod_{s=0}^{m+p-k-1} (1-q^s x) = 1.$$

Consequently, we obtain $\widetilde{B}_{m,p}(e_0;q;x) = 1$.

$$2^\circ \; \widetilde{B}_{m,p}(e_1;q;x) = \sum_{k=1}^{m+p} \begin{bmatrix} m+p \\ k \end{bmatrix}_q x^k \prod_{s=0}^{m+p-k-1} (1-q^s x) \frac{[k]_q}{[m]_q}$$

$$\overset{k \to k+1}{=} x \cdot \frac{[m+p]_q}{[m]_q} \sum_{k=0}^{m+p-1} \frac{[m+p-1]_q!}{[k]![m+p-k-1]_q!} x^k \prod_{s=0}^{m+p-k-2} (1-q^s x)$$

$$= x \cdot \frac{[m+p]_q}{[m]_q}.$$

$$3^\circ\ \widetilde{B}_{m,p}(e_2;q;x) = \sum_{k=1}^{m+p} \begin{bmatrix} m+p \\ k \end{bmatrix}_q x^k \prod_{s=0}^{m+p-k-1}(1-q^s x)\frac{[k]_q^2}{[m]_q^2}$$

$$= \sum_{k=1}^{m+p} \frac{[k]_q}{[m]_q}\cdot\frac{[k]_q}{[m]_q}\cdot\frac{[m+p]_q!}{[m+p-k]_q![k]_q!}\cdot x^k \prod_{s=0}^{m+p-k-1}(1-q^s x).$$

Taking into account that $[k]_q = q[k-1]_q + 1$, we obtain

$$\widetilde{B}_{m,p}(e_2;q;x) = \frac{[m+p]_q}{[m]_q^2}\sum_{k=2}^{m+p}\frac{q[k-1]_q[m+p-1]_q!}{[k-1]_q![m+p-k]_q!}\, x^k \prod_{s=0}^{m+p-k-1}(1-q^s x)$$

$$+ \frac{[m+p]_q}{[m]_q^2}\sum_{k=1}^{m+p}\frac{[m+p-1]_q!}{[k-1]_q![m+p-k]_q!}\, x^k \prod_{s=0}^{m+p-k-1}(1-q^s x).$$

Replacing $k \to k+2$ in first sum and $k \to k+1$ in the second, we have

$$\widetilde{B}_{m,p}(e_2;q;x) = \frac{[m+p-1]_q[m+p]_q}{[m]_q^2}\, q \sum_{k=0}^{m+p-2}\frac{[m+p-2]_q!}{[k]_q![m+p-k-2]_q}$$

$$\cdot\, x^{k+2}\prod_{s=0}^{m+p-k-3}(1-q^s x)$$

$$+ \frac{[m+p]_q}{[m]_q^2}\sum_{k=0}^{m+p-1}\frac{[m+p-1]_q!}{[k]_q![m+p-k-1]_q} x^{k+1}\prod_{s=0}^{m+p-k-2}(1-q^s x)$$

$$= \frac{[m+p-1]_q[m+p]_q}{[m]_q^2} q x^2 + \frac{[m+p]_q}{[m]_q^2} x.$$

Since $[m+p-1]_q q x^2 + x = [m+p]_q x^2 + x(1-x)$, the conclusion follows. □

We can give now the following result, a theorem of Korovkin type.

Theorem 2.2. *Let $q = q_m$ satisfy $0 < q_m < 1$, $\lim_{m\to\infty} q_m = 1$ and $\lim_{m\to\infty} q_m^m = a$, $a < 1$. Then, for any $f \in C([0,p+1])$, the following relation holds*

$$\lim_{m\to\infty}\widetilde{B}_{m,p}(f;q_m) = f \text{ uniformly on } [0,1].$$

Proof. The proof is based on the well known Korovkin theorem regarding the convergence of a sequence of linear and positive operators. So, it is enough to prove the conditions

$$\lim_{m\to\infty}\widetilde{B}_{m,p}(e_i;q_m;x) = x^i, \quad i = 0,1,2,$$

uniformly on $[0,1]$.

To prove the theorem we take into account the next relations obtained by simple calculations, where p is a fixed natural number.

$$\lim_{m\to\infty}\frac{[m+p]_{q_m}}{[m]_{q_m}} = 1, \quad \lim_{m\to\infty}\frac{[m+p]_{q_m}}{[m]_{q_m}^2} = 0. \tag{2.2}$$

Taking into account Lemma 2.1 and the relations (2.2), our statement is proved. □

3. On the rate of convergence

We will estimate the rate of convergence in terms of the modulus of continuity. Let $f \in C([0,b])$. The modulus of continuity of f denoted by $\omega_f(\delta)$ gives the maximum oscillation of f in any interval of length not exceeding $\delta > 0$ and it is given by relation

$$\omega_f(\delta) = \sup_{|y-x|\le\delta} |f(y) - f(x)|, \quad x, y \in [0,b].$$

It is known that $\lim\limits_{\delta\to 0^+} \omega_f(\delta) = 0$ for $f \in C([0,b])$, and for any $\delta > 0$ one has

$$|f(y) - f(x)| \le \omega_f(\delta)\left(\frac{|y-x|}{\delta} + 1\right). \tag{3.1}$$

Our result will be read as follows.

Theorem 3.1. *If $f \in C([0,1+p])$, then*

$$|\widetilde{B}_{m,p}(f;q;x) - f(x)| \le 2\omega_f(\delta_m)$$

takes place, where

$$\delta_m = \frac{1}{\sqrt{[m]_q}}\left(p + \frac{1}{2\sqrt{1-q^m}}\right), \quad q \in (0,1). \tag{3.2}$$

Proof. Since $B_{m,p}e_0 = e_0$, we have

$$|\widetilde{B}_{m,p}(f;q;x) - f(x)|$$
$$\le \sum_{k=0}^{m+p} \left| f\left(\frac{[k]_q}{[m]_q}\right) - f(x)\right| \frac{[m+p]_q!}{[m+p-k]_q![k]_q!}\; x^k \prod_{s=0}^{m+p-k-1}(1-q^s x).$$

In view of (3.1) we get

$$|\widetilde{B}_{m,p}(f;q;x) - f(x)|$$
$$\le \omega_f(\delta)\Bigg\{\frac{1}{\delta}\sum_{k=0}^{m+p}\left|\frac{[k]_q}{[m]_q} - x\right| \frac{[m+p]_q!}{[m+p-k]_q![k]_q!}\; x^k \prod_{s=0}^{m+p-k}(1-q^s x)$$
$$+ \sum_{k=0}^{m+p}\frac{[m+p]_q!}{[m+p-k]_q![k]_q!}\; x^k \prod_{s=0}^{m+p-k}(1-q^s x)\Bigg\}$$
$$= \omega_f(\delta)\left\{\frac{1}{\delta}\sum_{k=0}^{m+p}\left|\frac{[k]_q}{[m]_q} - x\right| \frac{[m+p]_q!}{[m+p-k]_q![k]_q!} x^k \prod_{s=0}^{m+p-k}(1-q^s x) + (\widetilde{B}_{m,p,q}e_0)(x)\right\}.$$

Using Cauchy-Schwartz inequality and Lemma 2.1 we can write

$$|\widetilde{B}_{m,p}(f;q;x) - f(x)|$$
$$\le \omega_f(\delta)\left\{\frac{1}{\delta}\left(\sum_{k=0}^{m+p}\left(\frac{[k]_q}{[m]_q} - x\right)^2 \frac{[m+p]_q!}{[m+p-k]_q![k]_q!} x^k \prod_{s=0}^{m+p-k}(1-q^s x)\right)^{1/2} + 1\right\}$$
$$= \omega_f(\delta)\left\{\frac{1}{\delta}((\widetilde{B}_{m,p,q}e_2)(x) - 2x(\widetilde{B}_{m,p,q}e_1)(x) + x^2(\widetilde{B}_{m,p,q}e_0)(x))^{1/2} + 1\right\}$$

$$=\omega_f(\delta)\left\{\frac{1}{\delta}\left(\frac{[m+p]_q}{[m]_q^2}([m+p]_q x^2+x(1-x))-2x^2\frac{[m+p]_q}{[m]_q}+x^2\right)^{1/2}+1\right\}$$

$$=\omega_f(\delta)\left\{\frac{1}{\delta}\left(x^2\left(\frac{[m+p]_q}{[m]_q}-1\right)^2+x(1-x)\frac{[m+p]_q}{[m]_q^2}\right)^{1/2}+1\right\}.$$

On the basis of the relation $(a^2+b^2)^{1/2}\le|a|+|b|$, the above inequality implies

$$|\widetilde{B}_{m,p}(f;q;x)-f(x)|$$

$$\le\omega_f(\delta)\left\{\frac{1}{\delta}\left(x\left|\frac{[m+p]_q}{[m]_q}-1\right|+\sqrt{\frac{x(1-x)}{[m]_q}}\sqrt{\frac{[m+p]_q}{[m]_q}}\right)+1\right\}. \tag{3.3}$$

Since

$$x\left|\frac{[m+p]_q}{[m]_q}-1\right|\le\frac{p}{\sqrt{[m]_q}},\qquad\sqrt{\frac{[m+p]_q}{[m]_q}}\le\frac{1}{\sqrt{1-q^m}}$$

and $\max\limits_{x\in[0,1]} x(1-x)=1/4$, choosing $\delta=\delta_m$ as in (3.2), we obtain the desired result. □

References

[1] Agratini, O., *On certain q-analogues of the Bernstein operators*, Carpathian J. Math., **24** (2008), no. 3, 281-286.

[2] Agratini, O., Nowak, G., *On a generalization of Bleimann, Butzer and Hahn operators based on q-integers*, Mathematical and Computer Modelling, **53**(2011), no. 5-6, 699-706.

[3] Aral, A., *A generalization of Szász-Mirakyan operators based on q-integers*, Mathematical and Computer Modelling, **47**(2008), 1052-1062.

[4] Aral, A., Doğru, O., *Bleimann, Butzer, and Hahn Operators Based on the q-Integers*, J. Inequal. Appl., vol. 2007, Art. ID 79410, 12 pp.

[5] Bărbosu, D., *Some generalized bivariate Bernstein operators*, Math. Notes, Miskolc, **1**(2000), 3-10.

[6] Derriennic, M.M., *Modified Bernstein polynomials and Jacobi polynomials in q-calculus*, Rend. Mat. Palermo, **76**(2005), 269-290.

[7] Doğru, O., Duman, O., *Statistical approximation of Meyer-König and Zeller operators based on q-integers*, Publ. Math. Debrecen, **68**(2006), 199-214.

[8] Doğru, O., Gupta, V., *Korovkin-type approximation properties of bivariate q-Meyer-Konig and Zeller operators*, Calcolo, **43**(2006), 51-63.

[9] Doğru, O., Gupta, V, *Monotonicity and the asymptotic estimate of Bleimann-Butzer and Hann operators based on q-integers*, J. Inequal. Appl., 2007, 1-12.

[10] Doğru, O., Muraru, C. V., *Statistical approximations by a Stancu type bivariate generalization of Meyer-König and Zeller type operators*, Mathematical and Computer Modelling, **48**(2008), no. 5-6, 961-968.

[11] Doğru, O., Orkcu, M., *Statistical approximation by a modification of q-Meyer-Konig-Zeller operators*, Applied Mathematics Letters, **23**(2010), 261-266.

[12] Gupta, V., *Some approximation properties of q-Durrmeyer operators*, Appl. Math. Comput., **197**(2008), 172-178.

[13] Gupta, V., Heping, W., *The rate of convergence of q-Durrmeyer operators for* $0 < q < 1$, Math. Methods Appl. Sci., **31(16)**(2008), 1946-1955.

[14] Gupta, V., Finta, Z., *On certain q-Durrmeyer operators*, Appl. Math. Comput., **209**(2009), 415-420.

[15] Gupta, V., Radu, C., *Statistical approximation properties of q-Baskakov-Kantorovich operators*, Cent. Eur. J. Math., **7**(2009), no. 4, 809-818.

[16] Heping, W., *Properties of convergence for q-Meyer-Konig and Zeller operators*, J. Math. Anal. Appl., **335**(2007), 1360-1373.

[17] Kac, V., Cheung, P., *Quantum Calculus*, Universitext, Springer, 2002.

[18] Lupaş, A., *A q-analogue of the Bernstein operator*, University of Cluj-Napoca, Seminar on Numerical and Statistical Calculus, Preprint **9**(1987), 85-92.

[19] Mahmudov, N., *Statistical approximation of Baskakov and Baskakov-Kantorovich operators based on the q-integers*, Cent. Eur. J. Math., **8(4)**(2010), 816-826.

[20] Philips, G. M., *On generalized Bernstein polynomials*, in D. F. Griffits, G. A. Watson (Eds.), 1996, 263-269.

[21] Trif, T., *Meyer-König and Zeller operators based on q-integers*, Rev. Anal. Numer. Theor. Approx., **29**(2000), 221-229.

[22] Schurer, F., *Linear positive operators in approximation theory*, Math. Inst. Techn. Univ. Delft Report, 1962.

Carmen-Violeta Muraru
University "Vasile Alecsandri" of Bacău
Department of Mathematics and Informatics
e-mail: carmen_7419@yahoo.com, cmuraru@ub.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 497–504

Estimates for general positive linear operators on non-compact interval using weighted moduli of continuity

Radu Păltănea

Abstract. We give estimates with explicit constants of the degree of approximation by general positive linear operators on the interval $[0, \infty)$, using a weighted modulus of continuity. In particular we obtain a quantitative version of a result of Totik concerning Szász-Mirakjan operators.

Mathematics Subject Classification (2010): 41A36.

Keywords: Positive linear operators, weighted modulus of continuity.

1. Introduction

The moduli of continuity or smoothness of different kinds play a crucial role in estimating the degree of approximation by using linear methods. In approximation on non-compact intervals more convenient are the weighted moduli. There are several types of constructions of weighted moduli of first order. A very short list of contributions in this directions are given in References.

In this paper we introduce a class of first order weighted moduli of continuity constructed starting from a family of "admissible" functions and we deduce estimates for general positive operators. These estimates are with explicit constants. Such type of estimates are already obtained for weighted moduli on a compact interval, for the Ditzian-Totik modulus of second order, (see [9], [8], [12]).

Finally we remark that, in the case of a certain admissible function, our modulus is equivalent to the usual modulus applied to a certain modification of the function. This last modulus was used by Totik [14] for Szász-Mirakjan operators.

2. A general estimate with the modulus ω^φ

Denote $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For $k \in \mathbb{N}$ denote by Π_k, the space of polynomials of degree at most k and for $j \in \mathbb{N}_0$ consider the monomial functions $e_j(t) = t^j$, $t \in [0,\infty)$. Denote by $[a]$, the integer part of a number $a \in \mathbb{R}$. Denote also by $\mathcal{F}(I)$, the space of real functions defined on an interval I.

We adopt the following

Definition 2.1. *A function $\varphi \in C([0,\infty))$ is named admissible if it satisfies the following conditions:*

i) $\varphi(t) > 0$, *for* $t \in (0,\infty)$;
ii) $\frac{1}{\varphi}$ *is convex on interval* $(0,\infty)$;
iii) *we have*

$$\lim_{a \to +0} \int_a^x \frac{\mathrm{d}t}{\varphi(t)} < \infty \quad \textit{for all } x > 0; \tag{2.1}$$

iv) *we have*

$$\int_0^\infty \frac{\mathrm{d}t}{\varphi(t)} = +\infty. \tag{2.2}$$

In this definition we use the Riemann improper integral. Using an admissible function φ we introduce the following first order weighted modulus.

Definition 2.2. *For $f \in \mathcal{F}([0,\infty))$, and $h > 0$ set:*

$$\omega^\varphi(f,h) = \sup\left\{ |f(v) - f(u)| \ : \ u, v \in [0,\infty),\ |v-u| \le h\varphi\left(\frac{u+v}{2}\right) \right\}. \tag{2.3}$$

We admit in this definition that the supremum could be equal to $+\infty$.

Remark 2.3. Function e_0 is admissible and for $\varphi = e_0$ we obtain $\omega^\varphi = \omega$, where ω denotes the usual first order modulus.

Property iii) allows to take φ with condition $\frac{1}{\varphi(x)} = \mathrm{O}(x^\alpha)$ $(x \to 0)$, with $\alpha > -1$. Very suitable for applications is the case $\varphi(x) \sim \sqrt{x}$ $(x \to 0)$, when the dependence of modulus $\omega^\varphi(f,\cdot)$ on the values taken by a function f in a neighbourhood of the point $x = 0$ is similar with the dependence of the first order Ditzian-Totik modulus on the values taken by a function near the end points of the interval $[0,1]$. However if we take $\varphi(x) = \sqrt{x}$, for $x \ge 0$, then $\omega^\varphi(f,h)$ is finite for any $h > 0$ only if f satisfies the restrictive condition $f(x) = \mathrm{O}(\sqrt{x})$ $(x \to \infty)$. This fact can be deduced, for instance, from Remark 2.6 in Section 2.

In order to enlarge the class of functions for which $\omega^\varphi(f,h) < \infty$, for any $h > 0$, by condition iv), we have the possibility to take φ rapidly decreasing to 0 when $x \to \infty$. For instance an admissible function is $\varphi(x) = \frac{\sqrt{x}}{1+x^m}$, $x \ge 0$, for $m \in \mathbb{N}$, $m \ge 2$. Then we have $\omega^\varphi(f,h) < \infty$, for any differentiable function f such that $|f'(x)| \le M x^{m-\frac{1}{2}}$.

Given an admissible function φ, we consider the following corresponding function

$$\Phi(x) = \int_0^x \frac{\mathrm{d}t}{\varphi(t)},\ x \in (0,\infty). \tag{2.4}$$

Lemma 2.4. *Let $f \in \mathcal{F}([0,\infty))$, $h > 0$ and $0 \le a < b$, such that $\Phi(b) - \Phi(a) = h$. Then for all points c, d such that $a \le c \le d \le b$, we have*

$$|f(d) - f(c)| \le \omega^{\varphi}(f,h). \tag{2.5}$$

Proof. We have to show that $d - c \le h\varphi\left(\frac{c+d}{2}\right)$.

From condition iii) of Definition 2.1 we deduce, using Jensen inequality:

$$\frac{d-c}{\varphi\left(\frac{c+d}{2}\right)} \le \int_c^d \frac{\mathrm{d}t}{\varphi(t)}.$$

But

$$\int_c^d \frac{\mathrm{d}t}{\varphi(t)} \le \int_a^b \frac{\mathrm{d}t}{\varphi(t)} = \Phi(b) - \Phi(a) = h.$$

□

Lemma 2.5. *Let $f \in \mathcal{F}([0,\infty))$, $x > 0$ and $h > 0$. We have*

$$|f(t) - f(x)| \le \left(1 + \frac{1}{h^2}\left(\Phi(t) - \Phi(x)\right)^2\right)\omega^{\varphi}(f,h). \tag{2.6}$$

Proof. We may consider only the case $\omega^{\varphi}(f,h) < \infty$. Note that function $\Phi : (0,\infty) \to (0,\infty)$ is a strictly increasing bijection. Therefore it admits an inverse $\Phi^{-1} : (0,\infty) \to (0,\infty)$.

Put $p = \left[\frac{\Phi(x)}{h}\right]$. Define the sequence $(u_j)_{j \ge -p}$ by

$$u_j = \Phi^{-1}(jh + \Phi(x)),\ j \ge -p.$$

From this it immediately follows that

$$\Phi(u_{j+1}) - \Phi(u_j) = h,\ j \ge -p.$$

Consider the decomposition

$$[0,\infty) = [0,u_{-p}) \cup \bigcup_{j=-p}^{\infty} [u_j, u_{j+1}),$$

where $[0,u_{-p}) = \emptyset$, if $u_{-p} = 0$. Let $t \in [0,\infty)$. We have to consider several cases.

<u>Case 1: $t \in [x,\infty)$.</u> Then there is an index $n \in \mathbb{N}_0$, such that $t \in [u_n, u_{n+1})$. We have

$$|f(t) - f(x)| \le |f(t) - f(u_n)| + \sum_{j=0}^{n-1} |f(u_{j+1}) - f(u_j)|,$$

where the last sum is 0 if $n = 0$. Using Lemma 2.4 we have $|f(t) - f(u_n)| \le \omega^{\varphi}(f,h)$ and $|f(u_{j+1}) - f(u_j)| \le \omega^{\varphi}(f,h)$, for $0 \le j \le n-1$. Hence

$$|f(t) - f(x)| \le (n+1)\omega^{\varphi}(f,h).$$

If $n=0$, from this we obtain directly relation (2.6). If $n \geq 1$ we have successively:

$$\begin{aligned} 1+n &= 1+\frac{1}{h}\sum_{j=0}^{n-1}(\Phi(u_{j+1})-\Phi(u_j)=1+\frac{1}{h}(\Phi(u_n)-\Phi(x)) \\ &\leq 1+\frac{1}{h}|\Phi(t)-\Phi(x)| \leq 1+\frac{1}{h^2}\cdot(\Phi(t)-\Phi(x))^2 \end{aligned}$$

It follows relation (2.6).

Case 2: $t \in [u_{-p}, x)$. This implies that $p \geq 1$. Then there is $n \in \mathbb{N}$, such that $t \in [u_{-n-1}, u_{-n})$. We have

$$|f(t)-f(x)| \leq |f(t)-f(u_{-n})| + \sum_{j=0}^{n-1}|f(u_{-j})-f(u_{-j-1})|,$$

where the last sum is 0 if $n=0$. Using Lemma 2.4 we have $|f(t)-f(u_{-n})| \leq \omega^{\varphi}(f,h)$ and $|f(u_{-j})-f(u_{-j-1})| \leq \omega^{\varphi}(f,h)$, for $0 \leq j \leq n-1$. Hence

$$|f(t)-f(x)| \leq (n+1)\omega^{\varphi}(f,h).$$

If $n=0$, from this we obtain directly relation (2.6). If $n \geq 1$ we have successively, similarly as in Case 1:

$$\begin{aligned} 1+n &= 1+\frac{1}{h}\sum_{j=0}^{n-1}(\Phi(u_{-j})-\Phi(u_{-j-1}))=1+\frac{1}{h}(\Phi(x)-\Phi(u_{-n})) \\ &\leq 1+\frac{1}{h}|\Phi(x)-\Phi(t)| \leq 1+\frac{1}{h^2}\cdot(\Phi(t)-\Phi(x))^2 \end{aligned}$$

Case 3: $t \in [0, u_{-p})$. We have

$$|f(t)-f(x)| \leq |f(t)-f(u_{-p})| + \sum_{j=0}^{p-1}|f(u_{-j})-f(u_{-j-1})|,$$

where the last sum is 0 if $p=0$. Let show that $|f(t)-f(u_{-p})| \leq \omega^{\varphi}(f,h)$. We must to prove $u_{-p}-t \leq h\varphi\left(\frac{u_{-p}+t}{2}\right)$. But from the convexity of function $\frac{1}{\varphi}$ we obtain

$$\frac{u_{-p}-t}{\varphi\left(\frac{u_{-p}+t}{2}\right)} \leq \int_t^{u_{-p}} \frac{\mathrm{d}s}{\varphi(s)} = \Phi(u_{-p})-\Phi(t) \leq \Phi(u_{-p}).$$

Since function Φ^{-1} is strictly increasing and $\Phi(x)-ph<h$ it follows that $u_{-p} \leq \Phi^{-1}(h)$. Hence $\Phi(u_{-p}) \leq h$. Then we continue like in Case 2, for $n=p$. □

Remark 2.6. From the proof of Lemma 2.5 it follows that for $f \in \mathcal{F}([0,\infty))$, $x>0$ and $h>0$, we have also

$$|f(t)-f(x)| \leq \left(1+\frac{1}{h}|\Phi(t)-\Phi(x)|\right)\omega^{\varphi}(f,h). \tag{2.7}$$

The main result of this section is the following

Theorem 2.7. *Let W be a linear subspace of $\mathcal{F}([0,\infty))$ and let $F : W \to \mathbb{R}$ be a positive linear functional. Let $x \in [0,\infty)$ and let φ be an admissible function. Suppose that $(\Phi - \Phi(x)e_0)^2 \in W$ and $e_0 \in W$. Then, for all $f \in W$ and all $h > 0$ we have*

$$\begin{aligned} |F(f) - f(x)| &\le |f(x)| \cdot |F(e_0) - 1| \\ &\quad + \Big(F(e_0) + h^{-2}F((\Phi - \Phi(x)e_0)^2)\Big)\omega^{\varphi}(f,h). \quad (2.8) \end{aligned}$$

Proof. The theorem follows from Lemma 2.5 and the inequality:

$$|F(f) - f(x)| \le |f(x)| \cdot |F(e_0) - 1| + F(|f - f(x)e_0|).$$

□

Corollary 2.8. *Let W be a linear subspace of $\mathcal{F}([0,\infty))$ and let $L : W \to \mathcal{F}([0,\infty))$ be a positive linear operator. Let φ an admissible function. Suppose that $(\Phi - \Phi(x)e_0)^2 \in W$ for each $x \in [0,\infty)$ and also $e_0 \in W$. Then for all $f \in W$, all $x \in [0,\infty)$ and $h > 0$ we have*

$$\begin{aligned} &|L(f,x) - f(x)| \le |f(x)| \cdot |L(e_0,x) - 1| \\ &+ \Big(L(e_0,x) + h^{-2}L((\Phi - \Phi(x)e_0)^2, x)\Big)\omega^{\varphi}(f,h). \quad (2.9) \end{aligned}$$

Remark 2.9. In the case $\varphi = e_0$, we have $\Phi = e_1$ and relation (2.9) becomes the well-known estimate of Mond [11].

3. Estimates for the weight $\varphi(x) = \sqrt{x}$

Theorem 3.1. *Let $W \subset \mathcal{F}([0,\infty))$ be a linear subspace, such that $\Pi_2 \in W$. If $L : W \to \mathcal{F}((0,\infty))$ is a positive linear operator, then for any $f \in W$, any $x \in (0,\infty)$ and any $h > 0$ we have*

$$\begin{aligned} &|L(f,x) - f(x)| \le |f(x)| \cdot |L(e_0,x) - 1| \\ &\qquad + \left(L(e_0,x) + \frac{4}{h^2 x}L((e_1 - xe_0)^2, x)\right)\omega^{\varphi}(f,h). \quad (3.1) \end{aligned}$$

In the particular case $L(e_0) = e_0$ and $h = \sqrt{\frac{L((e_1 - xe_0)^2, x)}{x}}$ we have

$$|L(f,x) - f(x)| \le 5 \cdot \omega^{\varphi}\left(f, \sqrt{\frac{L((e_1 - xe_0)^2, x)}{x}}\right). \quad (3.2)$$

Proof. We apply Corollary 2.8 by taking into account the estimate:

$$\left(\int_x^t \frac{\mathrm{d}u}{\sqrt{u}}\right)^2 = (2(\sqrt{t} - \sqrt{x}))^2 = 4 \cdot \left(\frac{t - x}{\sqrt{x} + \sqrt{t}}\right)^2 \le \frac{4(t - x)^2}{x}.$$

□

In the following theorem we give the connections between the modulus $\omega^{\varphi}(f, \bullet)$, for $\varphi(x) = \sqrt{x}$ and the usual modulus of function $f(x^2)$.

Theorem 3.2. *For any $f \in \mathcal{F}([0,\infty))$ and $h > 0$ we have*

$$\omega^{\varphi}(f, \sqrt{2}h) \le \omega(f \circ e_2, h) \le \omega^{\varphi}(f, 2h). \tag{3.3}$$

Proof. Let $x, y \in [0,\infty)$, such that $|x^2 - y^2| \le \sqrt{2}h\sqrt{\frac{x^2+y^2}{2}}$, which is equivalent to the inequality $|x - y| \le \frac{h\sqrt{x^2+y^2}}{x+y}$. But $\sqrt{x^2+y^2} \le x + y$. Hence $|x - y| \le h$. It follows $|f(x^2) - f(y^2)| \le \omega(f \circ e_2, h)$. Therefore

$$\sup_{x,y,\ |x^2-y^2| \le \sqrt{2}h\sqrt{\frac{x^2+y^2}{2}}} |f(x^2) - f(y^2)| \le \omega(f \circ e_2, h).$$

But

$$\sup_{x,y,\ |x^2-y^2| \le \sqrt{2}h\sqrt{\frac{x^2+y^2}{2}}} |f(x^2) - f(y^2)| = \sup_{u,v,\ |u-v| \le \sqrt{2}h\sqrt{\frac{u+v}{2}}} |f(u) - f(v)|$$
$$= \omega^{\varphi}(f, \sqrt{2}h).$$

Therefore

$$\omega^{\varphi}(f, \sqrt{2}h) \le \omega(f \circ e_2, h).$$

Conversely, let $x, y \in [0,\infty)$, such that $|\sqrt{x} - \sqrt{y}| \le h$, which is equivalent to $|x-y| \le h(\sqrt{x}+\sqrt{y})$. But $\sqrt{x}+\sqrt{y} \le 2\sqrt{\frac{x+y}{2}}$. Hence $|x-y| \le 2\sqrt{\frac{x+y}{2}}$ and consequently $|f(y) - f(x)| \le \omega^{\varphi}(f, 2h)$. Since x, y are arbitrarily chosen, we have

$$\sup_{x,y,\ |\sqrt{x}-\sqrt{y}| \le h} |f(y) - f(x)| \le \omega^{\varphi}(f, 2h).$$

But

$$\sup_{x,y,\ |\sqrt{x}-\sqrt{y}| \le h} |f(y) - f(x)| = \sup_{u,v,\ |u-v| \le h} |f(u^2) - f(v^2)|$$
$$= \omega(f \circ e_2, h).$$

Therefore

$$\omega(f \circ e_2, h) \le \omega^{\varphi}(f, 2h).$$

Corollary 3.3. *For $\varphi(x) = \sqrt{x}$, $x \in [0,\infty)$ and a function $f \in \mathcal{F}([0,\infty))$, the following are equivalent:*

i) $\lim_{h \to 0} \omega^{\varphi}(f, h) = 0$,

ii) *the function $f(x^2)$, $x \in [0,\infty)$ is uniformly continuous.*

We exemplify for the Szász-Mirakjan operators

$$S_n(f, x) = \sum_{k=0}^{\infty} f\left(\frac{k}{n}\right) e^{-nx} \frac{(nx)^k}{k!}, \tag{3.4}$$

$x \in [0,\infty)$, $n \in \mathbb{N}$ and $f \in W$, where $W \subset \mathcal{F}([0,\infty))$ is the linear subspace of the functions f for which the series above is convergent.

We have $S_n(e_0, x) = 1$, $S_n((e_1 - xe_0)^2, x) = \frac{x}{n}$. Also we have $S_n(f, 0) = f(0)$ for any $f \in W$. Hence we obtain:

Theorem 3.4. *Let* $\varphi(x) = \sqrt{x}$. *Let* $f \in W$, $x \in [0,\infty)$, $n \in \mathbb{N}$. *Then*

$$|S_n(f,x) - f(x)| \le 5 \cdot \omega^{\varphi}\left(f, \frac{1}{\sqrt{n}}\right). \tag{3.5}$$

□

Remark 3.5. In view of Corollary 3.3, relation (3.5) gives a quantitative version of a result of Totik [14] which states that, if $f(x^2)$ is a uniformly continuous function, $x \in [0,\infty)$, then the sequence of functions $(S_n f)_n$ is uniformly convergent on $[0,\infty)$ to function f.

4. Estimates for the weight $\varphi(x) = \frac{\sqrt{x}}{1+x^m}$, $m \in \mathbb{N}$, $m \ge 2$

Theorem 4.1. *Let* $W \subset \mathcal{F}([0,\infty))$ *be a linear subspace, such that* $\Pi_{2m} \in W$. *If* $L : W \to \mathcal{F}([0,\infty))$ *is a positive linear operator, then for any* $f \in W$, *any* $x \in (0,\infty)$ *and any* $h > 0$ *we have*

$$\begin{aligned}
&|L(f,x) - f(x)| \le |f(x)| \cdot |L(e_0,x) - 1| \\
&+\Big(L(e_0,x) + \frac{4}{h^2 x} L((e_1 - xe_0)^2(2e_0 + x^{2m}e_0 + e_{2m}), x)\Big)\omega^{\varphi}(f,h).
\end{aligned}$$

Proof. We apply Corollary 2.8 and use the estimate:

$$\begin{aligned}
\left(\int_x^t \frac{(1+u^m)\mathrm{d}u}{\sqrt{u}}\right)^2 &= 4\left(\sqrt{t} - \sqrt{x} + \frac{(\sqrt{t})^{2m+1} - (\sqrt{x})^{2m+1}}{2m+1}\right)^2 \\
&\le 8(\sqrt{t} - \sqrt{x})^2 \left[1 + \left(\frac{\sum_{k=0}^{2m} (\sqrt{t})^k (\sqrt{x})^{2m-k}}{2m+1}\right)^2\right] \\
&\le 8\frac{(t-x)^2}{x}\left[1 + \left(\frac{t^m + x^m}{2}\right)^2\right] \\
&\le 4\frac{(t-x)^2}{x}(2 + t^{2m} + x^{2m}).
\end{aligned}$$

□

Acknowledgments. We are grateful to the anonymous referee for suggestions made for a better presentation of the paper.

References

[1] Amanov, N.T., *On the weighted approximation by Szasz-Mirakjan operators*, Anal. Math., **18**(1992), no. 3, 167–184.

[2] Becker, M., *Global approximation theorems for Szász-Mirakjan and Baskakov type operators*, Indiana Univ. Math. J., **27**(1978), 127-138.

[3] Bustamente, J., *Estimates of positive linear operators in terms of second-order moduli*, J. Math. Anal. Appl., **345**(2008), 203–212.

[4] Bustamente, J., Morales de la Cruz, L., *Positive linear operators and continuous functions on unbounded interval*, Jaen J. Approx., **1**(2009), no. 2, 145–173.

[5] Bustamente, J., Quesada, J. M., Morales de la Cruz, L., *Direct estimate for positive linear operators in polynomial weighted spaces*, J. Approx. Theory, **160**(2010), 1495–1508.

[6] Felten, M., *Direct and inverse estimates for Bernstein polynomials*, Constr. Approx., **14**(1998), 459468.

[7] Gadjiev, A. D., Aral, A., *The estimates of approximation by using new type of weighted modulus of continuity*, Comput. Math. Appl., **54**(2007), 127–135.

[8] Gavrea, I., Gonska, H., Păltănea, R., Tachev, G., *General Estimates for the Ditzian–Totik Modulus*, East Journal of Approx., **9**(2003), no. 2, 175–194.

[9] Gonska, H., Tachev, G., *On the constants in* ω_2^{φ} *- inequalities*, Rendiconti Circ. Mat. Palermo Suppl., **68**(2002), 467–477.

[10] Lopez-Moreno, A.J., *Weighted simultaneous approximation with Baskakov type operators*, Acta. Math. Hungar., **104**(2004), no. 1–2, 143-151.

[11] Mond, B., *Note: On the degree of approximation by linear positive operators*, J. Approx. Theory, **18**, (1976), 304–306.

[12] Păltănea, R., *Approximation Theory Using Positive Linear Operators*, Birkhäuser, Boston, 2004.

[13] Păltănea, R., *Weighted second order moduli on a simplex*, Results in Mathematics, **53**(2009), no. 3-4, 361–369.

[14] Totik, V., *Uniform approximation by Szász-Mirakjan type operators*, Acta Math. Hungar., **41**(1983), no. 3-4, 291–307.

[15] Walczak, Z., *On certain linear positive operators in polynomial weighted spaces*, Acta Math. Hungar., **101**(2003), no. 3, 179–191.

Radu Păltănea
Transilvania University of Braşov,
Faculty of Mathematics and Computer Sciences
50, Iuliu Maniu Street,
5000091 Braşov,
Romania
e-mail: `radupaltanea@yahoo.com`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 505–513

Steffensen type methods for approximating solutions of differential equations

Flavius Pătrulescu

Abstract. The implicit methods for numerical solving of ODEs lead to nonlinear equations which are usually solved by the Newton method. We study the use of a Steffensen type method instead, and we give conditions under which this method provides bilateral approximations for the solution of these equations; this approach offers a more rigorous control of the errors. Moreover, the method can be applied even in the case when certain functions are not differentiable on the definition domain. The convergence order is the same as for Newton method.

Mathematics Subject Classification (2010): 65L04, 65L05, 65H05.

Keywords: Initial value problems, stiff equations, Steffensen method, Newton method, convergence order.

1. Introduction

The mathematical modeling of many problems in physics, engineering, chemistry, biology, etc. gives rise to ordinary differential equations or systems of ordinary differential equations.

It is known that a high-order initial value problem (IVP) for differential equations or systems of equations can be rewritten as a first-order IVP system (see e.g. [4], [5]) so that the standard IVP can be written in the form:

$$\begin{cases} y' = f(x, y), \quad x \in I \\ y(a) = y_0, \end{cases} \tag{1.1}$$

where: $y_0 \in \mathbb{R}^m$, $I \subseteq \mathbb{R}$, $f : I \times \mathbb{R}^m \to \mathbb{R}^m$ and $a \in I$.

A solution is sought on the interval $[a, b] \subset I$, where a, b are finite. In this paper we consider only the scalar case, i.e., $m = 1$.

In practice, the number of cases where an exact solution can be found by analytical means is very limited, so that one uses numerical methods for the approximation of the solution.

Integrating (1.1), for $m = 1$, using an implicit linear multistep method with step-size h, leads to the solving at each step of an equation of the form:

$$y = hA\phi(x, y) + \psi. \tag{1.2}$$

Here A is a constant determined by the numerical method and ψ is a known value.

This equation can be solved by the *fixed point iteration*

$$y^{(\nu+1)} = hA\phi(x, y^{(\nu)}) + \psi, \quad y^{(0)} \quad \text{arbitrary}, \quad \nu = 0, 1, \dots \tag{1.3}$$

which converges to the unique solution of (1.2) provided that:

$$h < \tfrac{1}{|A|L}, \tag{1.4}$$

where L is the Lipschitz constant of ϕ with respect to the second variable.

Condition (1.4) becomes too restrictive for stiff problems. Thus, if we use an explicit method to solve a stiff equation, we have to use an excessively small step-size to avoid instability; if we use an implicit method with an absolute stability region large enough to impose no stability restriction, we can choose a step-size as large as we want, but we will not be able to solve the implicit equation (1.2) by the iteration (1.3) unless the step-size is excessively small.

In order to overcome this difficulty one uses the Newton iteration instead of the fixed point iteration. Newton iteration applied to the equation:

$$F(y) = 0, \tag{1.5}$$

where $F : [c, d] \to \mathbb{R}$, $c, d \in \mathbb{R}$, $c < d$, has the form:

$$y^{(\nu+1)} = y^{(\nu)} - F(y^{(\nu)})/F'(y^{(\nu)}), \quad \nu = 0, 1, 2, \dots, \quad y^{(0)} \in [c, d]. \tag{1.6}$$

When applied to the equation (1.2), where $F(y) = y - hA\phi(x, y) - \psi$, we get:

$$y^{(\nu+1)} = y^{(\nu)} - (y^{(\nu)} - hA\phi(x, y^{(\nu)}) - \psi)/(1 - hA\phi'_y(x, y^{(\nu)})), \tag{1.7}$$

i.e.,

$$y^{(\nu+1)} = (hA(\phi(x, y^{(\nu)}) - y^{(\nu)}\phi'_y(x, y^{(\nu)})) + \psi)/(1 - hA\phi'_y(x, y^{(\nu)})). \tag{1.8}$$

One step of Newton iteration requests considerably more computing time than one step of fixed point iteration. Each step of the latter costs one function evaluation, whereas each step of the former calls for the updating of the derivative.

In this paper we approximate the solution of equation (1.5) using the Steffensen type method:

$$y^{(\nu+1)} = y^{(\nu)} - \frac{F(y^{(\nu)})}{[y^{(\nu)}, g(y^{(\nu)}); F]}, \quad \nu = 0, 1, \dots \tag{1.9}$$

where $g : [c, d] \to [c, d]$ is an auxiliary function such that the equation:

$$y - g(y) = 0 \tag{1.10}$$

is equivalent to (1.5), and $[u, v; F]$ represents the first order divided difference of F at the points $u, v \in [c, d]$. This method does not require the calculation of the derivative of the function F.

Let $y^* \in (c, d)$ be the root of equation (1.5). If the elements of the sequence $(y^{(\nu)})_{\nu \geq 0}$ belong to the interval $[c, d]$ then from Newton identity and (1.9) we obtain:

$$y^* - y^{(\nu+1)} = -\frac{[y^*, y^{(\nu)}, g(y^{(\nu)}); F](y^* - y^{(\nu)})(y^* - g(y^{(\nu)}))}{[y^{(\nu)}, g(y^{(\nu)}); F]},$$

where $[u, v, w; F]$ represents the second order divided difference of F at the points $u, v, w \in [c, d]$. If g is Lipschitz on $[c, d]$ with constant L and if we assume that there exist the real numbers $M, m > 0$ such that:

$$|[u, v, w; F]| < M \quad \text{and} \quad |[u, v; F]| > m,$$

for all u, v, $w \in [c, d]$, then:

$$|y^* - y^{(\nu+1)}| \leq \frac{ML|y^* - y^{(\nu)}|^2}{m},$$

which shows that the q-convergence order for the method (1.9) is 2, i.e., the same as for the Newton method.

In [7] are given conditions for the convergence of the sequences generated by relation (1.9), and the function g is defined such that the sequences $(y^{(\nu)})_{\nu \geq 0}$ and $(g(y^{(\nu)}))_{\nu \geq 0}$ approximate bilaterally the exact solution y^*. Thus, we have an a posteriori error control.

For the functions F and g we suppose the following hypothesis:

(α) the equations (1.5) and (1.10) are equivalent;
(β) the function g is continuous and decreasing on $[c, d]$;
(γ) the equation (1.5) has a unique solution $y^* \in (c, d)$.

The following theorem holds (see [7]):

Theorem 1.1. *If the functions F and g satisfy the conditions $(\alpha) - (\gamma)$ and moreover the following conditions hold:*

(i) F is increasing and convex on $[c, d]$;
(ii) $F(y_0) < 0$;
(iii) $g(y_0) \leq d$,

then the elements of the sequences $(y^{(\nu)})_{\nu \geq 0}$ and $(g(y^{(\nu)}))_{\nu \geq 0}$ belong to the interval $[c, d]$ and the following properties hold:

(j) the sequence $(y^{(\nu)})_{\nu \geq 0}$ is increasing and convergent;
(jj) the sequence $(g(y^{(\nu)}))_{\nu \geq 0}$ is decreasing and convergent;
(jjj) $y^{(\nu)} \leq y^* \leq g(y^{(\nu)}), \quad \nu = 0, 1, \ldots.$
(jv) $\lim_{\nu \to \infty} y^{(\nu)} = \lim_{\nu \to \infty} g(y^{(\nu)}) = y^*$;
(vj) $|y^* - y^{(\nu)}| \leq |g(y^{(\nu)}) - y^{(\nu)}|, \quad \nu = 0, 1, \ldots$.

In the above theorem the auxiliary function g can be taken as: $g(y) = y - \frac{F(y)}{F'(c)}$. Similar results have been obtained in [7] if F verifies the properties:

-F is increasing and concave; g can be taken as $g(y) = y - \frac{F(y)}{F'(d)}$;
-F is decreasing and concave; g can be taken as $g(y) = y - \frac{F(y)}{F'(c)}$;
-F is decreasing and convex; g can be taken as $g(y) = y - \frac{F(y)}{F'(d)}$.

The interval $[a, b]$ is partitioned by the point set $\{x_n\}$ defined by $x_n = a + nh$, $n = 0, 1, \ldots, N$, $h = (b - a)/N$, and y_n denotes an approximation to the exact solution y of (1.1) at x_n.

If we use an implicit linear multistep method then y_n, $n = 1, \ldots, N$, are the solutions of the equation:

$$y = hA\phi(x_n, y) + \psi_n, \tag{1.11}$$

where $\psi_n = \psi_n(a, h, y_{n-1}, y_{n-2}, \ldots, y_0)$. We call this equation as *approximant equation* and we denote by $y_n^* \in (c, d)$, $n = 1, \ldots, N$, the exact solution.

For each $n = 1, \ldots, N$ let $F_n : [c, d] \to \mathbb{R}$ be defined by

$$F_n(y) = y - hA\phi(x_n, y) - \psi_n. \tag{1.12}$$

Then equation (1.11) can be rewritten in the form $F_n(y) = 0$.

To approximate bilaterally the solution y_n^*, $n = 1, \ldots, N$, we generate the sequence $(y_n^{(\nu)})_{\nu \geq 0}$, by:

$$y_n^{(\nu+1)} = y_n^{(\nu)} - \frac{F_n(y_n^{(\nu)})}{[y_n^{(\nu)}, g(y_n^{(\nu)}); F_n]}, \quad \nu = 0, 1, \ldots \tag{1.13}$$

or, using (1.12),

$$y_n^{(\nu+1)} = \frac{hA(\phi(x_n, y_n^{(\nu)}) - y_n^{(\nu)}[y_n^{(\nu)}, g(y_n^{(\nu)}); \phi(x_n, \cdot)]) + \psi_n}{1 - hA[y_n^{(\nu)}, g(y_n^{(\nu)}); \phi(x_n, \cdot)]}. \tag{1.14}$$

From Theorem 1.1, if F_n is increasing and convex, and the initial guess $y_n^{(0)}$ satisfy $F_n(y_n^{(0)}) < 0$, then the sequences $(y_n^{(\nu)})_{\nu \geq 0}$, $(g(y_n^{(\nu)}))_{\nu \geq 0}$ converge to y_n^* and we have the inequalities:

$$y_n^{(\nu)} \leq y_n^* \leq g(y_n^{(\nu)}), \quad \nu = 0, 1, \ldots \quad .$$

The rest of the cases can be treated in a similar fashion.

2. Application to the trapezoidal rule

We consider the *trapezoidal rule* to integrate the initial value problem (1.1), for $m = 1$, and the Steffensen method described above to solve the *approximant equation* (1.11).

The *trapezoidal rule* is a 1-step Adams-Moulton method (an implicit method), and for (1.1) is defined by:

$$y_n = y_{n-1} + \tfrac{h}{2}(f(x_n, y_n) + f(x_{n-1}, y_{n-1})), n = 1, \ldots, N.$$

It is known that the trapezoidal rule is an A-stable method and has order 2 (see [5]).

For any point x_n, $n = 1, \ldots, N$, we have:

$$y_n - \tfrac{h}{2} f(x_n, y_n) - \tfrac{h}{2} f(x_{n-1}, y_{n-1}) - y_{n-1} = 0 \tag{2.1}$$

and in this case $F_n(y) = y - \frac{h}{2} f(x_n, y) - \frac{h}{2} f(x_{n-1}, y_{n-1}) - y_{n-1}$. Thus, in (1.11) we have $A = \frac{1}{2}$, $\phi(x_n, y) = f(x_n, y)$ and $\psi_n = \frac{h}{2} f(x_{n-1}, y_{n-1}) + y_{n-1}$, $n = 1, \ldots, N$.

For simplicity we consider only the autonomous case, i.e. $f = f(y)$, and in this case equation (2.1) becomes:

$$y_n - \tfrac{h}{2} f(y_n) - \tfrac{h}{2} f(y_{n-1}) - y_{n-1} = 0 \tag{2.2}$$

and $F_n(y) = y - \frac{h}{2} f(y) - \psi_n$, $\psi_n = \frac{h}{2} f(y_{n-1}) + y_{n-1}$, $n = 1, \ldots, N$.

Using the fact that

$$[u, v; F_n] = 1 - \tfrac{h}{2} [u, v; f], \text{ for all } u, v \in [c, d], \tag{2.3}$$

and

$$[u, v, w; F_n] = -\tfrac{h}{2} [u, v, w; f], \text{ for all } u, v, w \in [c, d], \tag{2.4}$$

$n = 1, \ldots, N$, we obtain that the auxiliary function g can be taken as (see [10]):

$$g(y) = \frac{\frac{h}{2}(f(y) - y[d - \varepsilon, d; f]) + \psi_n}{1 - \frac{h}{2}[d - \varepsilon, d; f]};$$

or

$$g(y) = \frac{\frac{h}{2}(f(y) - y[c, c + \varepsilon; f]) + \psi_n}{1 - \frac{h}{2}[c, c + \varepsilon; f]},$$

where ε is sufficiently small such that the exact solution y_n^* of the equation $F_n(y_n) = 0$, $n = 1, \ldots, N$, belongs to the interval $[c + \varepsilon, d - \varepsilon]$.

For each $n = 1, \ldots, N$ we denote:

$$\psi_{\max}^n = \max\{y_k + \tfrac{h}{2} f(y_k) | k = 0, \ldots, n-1\},$$
$$\psi_{\min}^n = \min\{y_k + \tfrac{h}{2} f(y_k) | k = 0, \ldots, n-1\}.$$

We are lead to the main results of this work:

Theorem 2.1. *If the function f, the step-size h, and the initial guesses $y_n^{(0)}$, $n = 1, \ldots, N$, satisfy the following conditions:*

(i) $[u, v, w, f] \le 0$, *for all* $u, v, w \in [c, d]$;

(ii) ($m \le [u, v, f] \le M \le 0$, *for all* $u, v \in [c, d]$) *or* ($0 \le m \le [u, v, f] \le M$, *for all* $u, v \in [c, d]$, *and* $h \le \frac{2}{M}$);

(iii) $y_n^{(0)} - \frac{h}{2} f(y_n^{(0)}) < \psi_{\min}^n$;

(iv) $y_n^{(0)} M - f(y_n^{(0)}) \ge \frac{2}{h}[d(M\frac{h}{2} - 1) + \psi_{\max}^n]$,

then the elements of the sequences $(y_n^{(\nu)})_{\nu \ge 0}, (g(y_n^{(\nu)}))_{\nu \ge 0}$, $n = 1, \ldots, N$, *belong to the interval* $[c, d]$ *and the following properties hold:*

(j) $(y_n^{(\nu)})_{\nu \ge 0}$ *is increasing and convergent;*

(jj) $(g(y_n^{(\nu)}))_{\nu \ge 0}$ *is decreasing and convergent;*

(jjj) $y_n^{(\nu)} \le y_n^* \le g(y_n^{(\nu)})$, $\nu = 0, 1, \ldots$;

(jv) $\lim\limits_{\nu \to \infty} y_n^{(\nu)} = \lim\limits_{\nu \to \infty} g(y_n^{(\nu)}) = y_n^*$;

(v) $|y_n^* - y_n^{(\nu)}| \le |g(y_n^{(\nu)}) - y_n^{(\nu)}|, \quad \nu = 0, 1, \ldots$.

Proof. From (2.3), (2.4) and (i), (ii) we have $[u, v; F_n] \ge 0$, $[u, v, w; F_n] \ge 0$, $n = 1, \ldots, N$, for all $u, v, w \in [c, d]$, and we deduce that F_n is increasing and convex.

Also, from (iii) and (iv) we obtain that the initial guesses satisfy the inequalities: $F(y_n^{(0)}) < 0$ and $g(y_n^{(0)}) \le d$, $n = 1, \ldots, N$.

Using Theorem 1.1 we deduce that the properties $(j) - (v)$ hold. □

The following theorems can be proved in a similar manner:

Theorem 2.2. *If the function f, the step-size h, and the initial guesses $y_n^{(0)}$, $n = 1, \ldots, N$, satisfy the following conditions:*

(i) $[u, v, w, f] \le 0$, *for all* $u, v, w \in [c, d]$;
(ii) $0 \le m \le [u, v, f] \le M$, *for all* $u, v \in [c, d]$;
(iii) $y_n^{(0)} - \frac{h}{2} f(y_n^{(0)}) < \psi_{\min}^n$;
(iv) $y_n^{(0)} m - f(y_n^{(0)}) \ge \frac{2}{h}[c(m\frac{h}{2} - 1) + \psi_{\max}^n]$;
(v) $\frac{2}{m} \le h$,

then the elements of the sequences $(y_n^{(\nu)})_{\nu \ge 0}, (g(y_n^{(\nu)}))_{\nu \ge 0}$, $n = 1, \ldots, N$, belong to the interval $[c, d]$ and the following properties hold:

(j) $(y_n^{(\nu)})_{\nu \ge 0}$ *is decreasing and convergent;*
(jj) $(g(y_n^{(\nu)}))_{\nu \ge 0}$ *is increasing and convergent;*
(jjj) $g(y_n^{(\nu)}) \le y_n^* \le y_n^{(\nu)}$, $\nu = 0, 1, \ldots$;
(jv) $\lim_{\nu \to \infty} y_n^{(\nu)} = \lim_{\nu \to \infty} g(y_n^{(\nu)}) = y_n^*$;
(v) $|y_n^* - y_n^{(\nu)}| \le |g(y_n^{(\nu)}) - y_n^{(\nu)}|, \quad \nu = 0, 1, \ldots$.

Theorem 2.3. *If the function f, the step-size h and the initial guesses $y_n^{(0)}$, $n = 1, \ldots, N$, satisfy the following conditions:*

(i) $[u, v, w, f] \ge 0$, *for all* $u, v, w \in [c, d]$;
(ii) *(*$m \le [u, v, f] \le M \le 0$, *for all* $u, v \in [c, d]$*) or (*$0 \le m \le [u, v, f] \le M$, *for all* $u, v \in [c, d]$, *and* $h \le \frac{2}{M}$*);*
(iii) $y_n^{(0)} - \frac{h}{2} f(y_n^{(0)}) > \psi_{\max}^n$;
(iv) $y_n^{(0)} M - f(y_n^{(0)}) \le \frac{2}{h}[c(M\frac{h}{2} - 1) + \psi_{\min}^n]$,

then the elements of the sequences $(y_n^{(\nu)})_{\nu \ge 0}, (g(y_n^{(\nu)}))_{\nu \ge 0}$, $n = 1, \ldots, N$, belong to the interval $[c, d]$ and the following properties hold:

(j) $(y_n^{(\nu)})_{\nu \ge 0}$ *is decreasing and convergent;*
(jj) $(g(y_n^{(\nu)}))_{\nu \ge 0}$ *is increasing and convergent;*
(jjj) $g(y_n^{(\nu)}) \le y_n^* \le y_n^{(\nu)}$, $\nu = 0, 1, \ldots$;
(jv) $\lim_{\nu \to \infty} y_n^{(\nu)} = \lim_{\nu \to \infty} g(y_n^{(\nu)}) = y_n^*$;
(v) $|y_n^* - y_n^{(\nu)}| \le |g(y_n^{(\nu)}) - y_n^{(\nu)}|, \quad \nu = 0, 1, \ldots$.

Theorem 2.4. *If the function f, the step-size h and the initial guesses $y_n^{(0)}$, $n = 1, \ldots, N$, satisfy the following conditions:*

(*i*) $[u,v,w,f] \geq 0$, *for all* $u,v,w \in [c,d]$*;*
(*ii*) $0 \leq m \leq [u,v,f] \leq M$, *for all* $u,v \in [c,d]$*;*
(*iii*) $y_n^{(0)} - \frac{h}{2} f(y_n^{(0)}) > \psi_{\max}^n$*;*
(*iv*) $y_n^{(0)} m - f(y_n^{(0)}) \leq \frac{2}{h}[d(m\frac{h}{2} - 1) + \psi_{\min}^n]$*;*
(*v*) $\frac{2}{m} \leq h$*,*

then the elements of the sequences $(y_n^{(\nu)})_{\nu \geq 0}, (g(y_n^{(\nu)}))_{\nu \geq 0}$*,* $n = 1, \ldots, N$*, belong to the interval* $[c,d]$ *and the following properties hold:*

(*j*) $(y_n^{(\nu)})_{\nu \geq 0}$ *is increasing and convergent;*
(*jj*) $(g(y_n^{(\nu)}))_{\nu \geq 0}$ *is decreasing and convergent;*
(*jjj*) $y_n^{(\nu)} \leq y_n^* \leq g(y_n^{(\nu)})$*,* $\nu = 0, 1, \ldots$*;*
(*jv*) $\lim\limits_{\nu \to \infty} y_n^{(\nu)} = \lim\limits_{\nu \to \infty} g(y_n^{(\nu)}) = y_n^*$*;*
(*v*) $|y_n^* - y_n^{(\nu)}| \leq |g(y_n^{(\nu)}) - y_n^{(\nu)}|, \quad \nu = 0, 1, \ldots$.

3. Numerical example

We consider the autonomous initial value problem:

$$\begin{cases} y'(x) = \cos^2(y(x)), & x \in [0,1], \\ y(0) = 0. \end{cases} \tag{3.1}$$

The exact solution is $y : [0,1] \to \mathbb{R}$, $y(x) = \arctan(x)$, and it is plotted in Figure $1(a)$ with continuous line.

If we use the *trapezoidal rule* to integrate the above initial value problem we must solve for each mesh point $x_n = nh$, $n = 1, \ldots, N$, $h = 1/N$, $N \in \mathbb{N}$, the nonlinear equation:

$$y_n = y_{n-1} + \tfrac{h}{2}(\cos^2 y_n + \cos^2 y_{n-1}), \tag{3.2}$$

where $x_0 = 0$ and we choose $y_0 = 0$.

According to the above sections we can write (3.2) in the form

$$F_n(y) = 0,$$

where $F_n(y) = y - \frac{h}{2}\cos^2(y) - \psi_n$, and $\psi_n = y_{n-1} + \frac{h}{2}\cos^2(y_{n-1})$, $n = 1, \ldots, N$. It is easy to show that equation (3.2) has a unique solution $y_n^* \in (0, \frac{\pi}{4})$, $n = 1, \ldots, N$, and we will use a *Steffensen* type method to obtain a numerical approximation, $\tilde{y}_n$, for this solution.

From $F_n'(y) = 1 + \frac{h}{2}\sin(2y) \geq 0$ and $F_n''(y) = h\cos(2y) \geq 0$, $y \in [0, \frac{\pi}{4}]$, $n = 1, \ldots, N$, we deduce that F_n is increasing and convex. Thus, we can define the decreasing function g as:

$$g(y) = y - \frac{F_n(y)}{F_n'(0)} = y - F_n(y) = \tfrac{h}{2} cos^2(y) + \psi_n, \quad n = 1, \ldots, N.$$

Also, from Theorem 2.1, choosing for each $n = 1, \ldots, N$ the initial guesses $y_n^{(0)}$ such that it verifies the conditions (*iii*) and (*iv*) we obtain bilateral approximations of the solution y_n^* and an a posteriori error control.

The numerical solution, obtained with the method described above, for the step size $h = 0.05$ is also plotted in Figure 1(a) with circle marker. The values of the errors $\varepsilon_n = |y(x_n) - \tilde{y}_n|$, $n = 1, \ldots, N$, are presented in the following table. They are also plotted in Figure 1(b). We observe a very good agreement when we compare the numerical with the analytical solution.

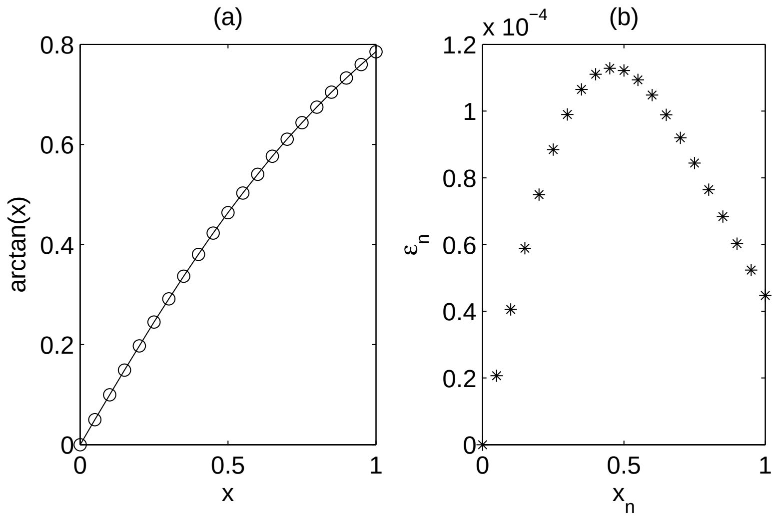


Figure 1. (a) The exact solution (continuous line) and the numerical solution (circle marker). (b) The values of the errors

Table 1. The values of the errors

x_n	ε_n	x_n	ε_n
0.05	0.00002068828052	0.55	0.00010932962999
0.1	0.00004056160110	0.6	0.00010478854028
0.15	0.00005887122616	0.65	0.00009889278012
0.2	0.00007499149969	0.7	0.00009200032239
0.25	0.00008845999793	0.75	0.00008443386474
0.3	0.00009899709371	0.8	0.00007647332226
0.35	0.00010650491201	0.85	0.00006835314397
0.4	0.00011104895260	0.9	0.00006026315926
0.45	0.00011282772086	0.95	0.00005235176793
0.5	0.00011213634983	1	0.00004473048874

References

[1] Agratini, O., Chiorean, I., Coman, Gh., Trîmbiţas, R., *Numerical Analysis and Approximation Theory*, Vol. *III*, Presa Universitară Clujeană, 2002 (in Romanian).

[2] Butcher, J.C., *Numerical Methods for Ordinary Differential Equations*, John Willey & Sons, 2003.

[3] Cătinaş, E., *Methods of Newton and Newton-Krylov Type*, Risoprint, Cluj-Napoca, 2007.

[4] Crouzeix, M., Mignot, A.L., *Analyse numérique des equations différentielles*, Masson, Paris, 1989.

[5] Lambert, J.D., *Numerical Methods for Ordinary Differential Systems-The Initial Value Problem*, John Wiley & Sons, 1990.

[6] Matheij, R., Molennar, J., *Ordinary Differential Equations in Theory and Practice*, SIAM, Philadelphia, 2002.

[7] Păvăloiu, I., *On the monotonicity of the sequences of approximations obtained by Steffensen's method*, Mathematica, **35(58)**(1993), no. 1, 71-77.

[8] Păvăloiu, I., *Bilateral approximation for the solution of scalar equations*, Rev. Anal. Numér. Théor. Approx., **23**(1994), no. 1, 95–101.

[9] Păvăloiu, I., *Approximation of the roots of equations by Aitken-Steffensen-type monotonic sequences*, Calcolo, **32**(1995), 69–82.

[10] Păvăloiu, I., *Aitken-Steffensen type methods for nonsmooth functions(III)*, Rev. Anal. Numér. Théor. Approx., **32**(2003), no. 1, 73–77.

Flavius Pătrulescu
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street
400084 Cluj-Napoca
Romania
e-mail: `fpatrulescu@ictp.acad.ro`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 515–526

Approximation methods for second order nonlinear polylocal problems

Daniel N. Pop and Radu T. Trîmbiţaş

Abstract. Consider the problem:

$$\begin{aligned} y''(x) + f(x, y) &= 0, \qquad x \in [0, 1] \\ y(a) &= \alpha \\ y(b) &= \beta, \qquad a, b \in (0, 1). \end{aligned}$$

This is not a two-point boundary value problem since $a, b \in (0, 1)$. It is possible to solve this problem by dividing it into the three problems: a two-point boundary value problem (BVP) on $[a, b]$ and two initial-value problems (IVP), on $[0, a]$ and $[b, 1]$. The aim of this work is to present two solution procedures: one based on B-splines of order $k + 2$ and the other based on a combination of B-splines (order $k + 2$) with a $(k + 1)$-order Runge-Kutta method. Then, we give two numerical examples and compare the methods experimentally.

Mathematics Subject Classification (2010): 65D07, 34B15, 65F50, 49M15.

Keywords: B-splines, nonlinear boundary value problems, sparse matrices, Newton method.

1. Introduction

Consider the problem (PVP - **P**olylocal **V**alue **P**roblem):

$$y''(x) + f(x, y) = 0, \qquad x \in [0, 1] \tag{1.1}$$

$$y(a) = \alpha \tag{1.2}$$

$$y(b) = \beta, \qquad a, b \in (0, 1),\ a < b. \tag{1.3}$$

where a, b, α, $\beta \in \mathbb{R}$. *This is not a two-point boundary value problem*, since $a, b \in (0, 1)$.

We try to solve the problem using two methods:

- a collocation method based on B-splines of order $k + 2$;
- a combined method based on B-splines (order $k+2$) and a Runge-Kutta method(order $k + 1$).

The methods are new in this context: the conditions are stated at interior points. Also it is shown that the Runge-Kutta method does not degrade the accuracy provided by the collocation method for the BVP.

Our choice to use these methods is based on the following reasons :

1. We write the code using the function `spcol` in MATLAB Spline Toolbox.
2. It is the most suitable method, for a general purpose code, among the finite element ones. See [2, 17, 21], where complexity comparisons which support the above claim are made and collocation, when efficiently implemented, is shown to be competitive with finite differences using extrapolation.
3. Theoretical results on the convergence of collocation method are given in [6, 16].
4. Several representative test problems demonstrate the stability and flexibility [7].
5. For each Newton iteration, the resulting linear algebraic system of equations (after using Newton method with quasilinearization) is solved using methods given in [8].

We also consider the BVP :

$$y''(x) + f(x,y) = 0, \qquad x \in [a,b] \tag{1.4}$$

$$y(a) = \alpha \tag{1.5}$$

$$y(b) = \beta, \tag{1.6}$$

To apply the collocation theory, we need to have an isolated solution $y(x)$ of the problem (1.4)+(1.5)+(1.6), and this occurs if the above linearized problem for $y(x)$ is uniquely solvable. R.D Russel and L.F.Shampine [22] study the existence and the uniqueness of the isolated solution.

Theorem 1.1. [22] *Suppose that $y(x)$ is a solution of the boundary value problem (1.4)+(1.5)+(1.6), that the functions*

$$f(x,z) \text{ and } \frac{\partial f(x,z)}{\partial y}$$

are defined and continuous for $a \le x \le b$, and $|z-y| \le \delta$, $\delta > 0$, and the homogeneous equation $y''(x) = 0$ subject to the homogeneous boundary conditions (1.5)+(1.6) has only the trivial solution. If the linear homogeneous equation

$$z''(x) + \frac{\partial f(x,y)}{\partial y} z(x) = 0$$

has only trivial solution, then this is sufficient to guarantee that there exists a $\sigma > 0$ such $y(x)$ is the unique solution of problem BVP in the sphere:

$$\{w : \|w - y''\| \le \sigma\}.$$

For the existence and uniqueness of an IVP, we recall the following result.

Theorem 1.2. [15, pp. 112-113]*Suppose that* $D = \{a \le x \le b, -\infty < y < \infty\}$ *and* $f(x, y)$ *is continuous on* D*. If* f *satisfies a Lipschitz condition on* D *in the variable* y*, then the initial value problem (IVP)*

$$\begin{cases} y' &= z, \\ z' &= -f(x, y), \ a \le x \le b, \\ y(a) &= \alpha, \\ y'(a) &= \varsigma, \end{cases} \tag{1.7}$$

has a unique solution $y(x)$ *for* $a \le x \le b$*.*

If the problem BVP has the unique solution, the requirement $y(x) \in C^2[0, 1]$ ensure the existence and the uniqueness of the solution of PVP.

2. The collocation method for solving the polylocal problem using B-splines

2.1. B-splines bases of degree k (order $k + 1$)

For reason of efficiency, stability, flexibility in order, and continuity, we choose B-splines as the basis functions. Efficient algorithms for calculating with B-splines are given by deBoor [9, 10] and Risler[20].

Consider a sequence of knots $t_0, \ldots, t_m$, such that $t_i \le t_{i+1}$ for all i.

Definition 2.1. Let $t = (t_0, \ldots, t_m)$. For $x \in \mathbb{R}$, $0 \le i \le m - k - 1$, we define B-splines of degree k as follows:

$$\begin{cases} B_{i,0} = \begin{cases} 1, \ \text{if } t_i \le x < t_{i+1} \\ 0, \ \text{otherwise} \end{cases} \\ B_{i,k}(x) = w_{i,k}(x) B_{i,k-1}(x) + (1 - w_{i+1,k}(x)) B_{i+1,k-1}(x), \end{cases} \tag{2.1}$$

where

$$w_{i,k}(x) = \begin{cases} \frac{x - t_i}{t_{i+k} - t_i}, & \text{if } t_i < t_{i+k} \\ 0, & \text{otherwise.} \end{cases} \tag{2.2}$$

If $s(x) = \sum_{r=0}^{m-k-1} c_r B_{r,k}(x)$, then its derivatives can be found for $x \in (t_j, t_{j+k})$ from (see for more details [4, pp. 62]):

$$s^{(i)}(x) = \sum_{l=j-k+i-1}^{j} c_{l,i+1} B_{l,k-i}(x), \tag{2.3}$$

where

$$c_{l,i+1} := \begin{cases} c_l, \ \text{if } i = 0 \\ (k - i) \frac{c_{l,i} - c_{l-1,i}}{t_{l+k-i} - t_l}, \ \text{if } i > 0. \end{cases} \tag{2.4}$$

To evaluate $B^{(i)}_{j,k}(x)$, we take $c_r = \delta_{rj}$, for $r = 0, \ldots, m - k - 1$, in (2.3) and (2.4).

2.2. Principles of the method

First we are interested to a global approach for the solution of problem (1.1) + (1.2) + (1.3). Let Δ be a partition of $[0,1]$ like

$$\Delta : 0 = x_0 < x_1 < \cdots < x_{N-1} = 1. \tag{2.5}$$

We insert the points a and b into the partition. Suppose $x_l = a$ and $x_{l+p} = b$, $0 < l < N+1$, $1 < l+p < N+1$. The multiplicity of each point inner point is k, and the multiplicity of endpoints is $k+2$. Let

$$H_i := x_{i+1} - x_i, \qquad i = 0, \dots, N \tag{2.6}$$

be the step sizes.

We construct the following collocation points

$$\xi_{ij} := x_i + H_i \rho_j; \; i = 0, 1, \dots, N-1, \; j = 1, 2, \dots, k, \tag{2.7}$$

on each subinterval $[x_i, x_{i+1}]$, $i = 0, 1, \dots, N-1$, where

$$0 < \rho_1 < \rho_2 < \cdots < \rho_k < 1 \tag{2.8}$$

are the roots of k-th Legendre polynomial (see [5] for more details). We insert the points a and b into the set of collocation point, so we obtain $n = Nk + 2$ points.

Remark 2.2. If a or b coincide with one of the previously computed collocation point, then we increment N.

One renumbers the collocation points, such that the first is $\xi_0 := x_0 + H_0 \rho_0$, and the last is $\xi_{n-1} := x_N + H_N \rho_k$, where $n = Nk + 2$. The dimension of our spline space must be $n = Nk + 2$. Using notations in section 2.1, we have $m = (N+1)k + 4$. Therefore, the partition of $[0,1]$ becomes: [23, pp. 65]:

$$\overline{\Delta} : 0 = x_0 \le x_1 \le \cdots \le x_m = 1. \tag{2.9}$$

Definition 2.3. A function $v(x)$ is in the family $L(\overline{\Delta}, k, p)$ if $v(x)$ is a polynomial of degree k on each subinterval of $\overline{\Delta}$ and $v \in C^p[0,1]$. The subfamily $L'(\overline{\Delta}, k, p)$ consists of all functions in $L(\overline{\Delta}, k, p)$ which satisfy the boundary conditions (1.5) + (1.6).

Suppose a partition (2.9) of $[0,1]$ and a sequence of partitions $\overline{\Delta}_n (n = 1, 2, 3, \dots)$ satisfying

$$\lim_{n \to \infty} h(\overline{\Delta}_n) = 0$$

are given. If we form a set of points $S_n (n = 1, 2, \dots)$ like in (2.9), then, for a large n (see [25] for more details), there is a unique element $u_{\overline{\Delta}_n}(x)$ of $L'(\overline{\Delta}_n, k+1, 1)$ satisfying (1.4) at each point of S_n and

$$\left\| u_{\overline{\Delta}_n}(x) - y(x) \right\| \le \delta. \tag{2.10}$$

The approximate solution $y_n(x)$ and its derivatives up to order two converge uniformly to $y(x)$ and to its derivatives of corresponding orders. Moreover the rate of convergence is bounded by

$$\left\| u_{\overline{\Delta}_n}(x)^{(k)} - y^{(k)}(x) \right\| \leq \theta F_n(u''),\ k = 0, 1, \tag{2.11}$$

where θ is a constant independent of n and $F_n(u'')$ is the error of the best uniform approximation to $y''(x)$ in $L(\overline{\Delta}_n, k-1, 0)$.

We wish to find an approximate solution of the problem (1.1)+(1.2)+(1.3) in $L(\overline{\Delta}, k+1, 1)$, having the following form:

$$u_{\overline{\Delta}}(x) = \sum_{i=0}^{n-1} c_i B_{i,k+1}(x), \tag{2.12}$$

where $B_{i,k+1}(x)$ is a B-spline of order $(k+2)$ with knots $\{x_i\}_{i=0}^{m}$.

Remark 2.4. Our approximation method is inspired from ([11], chap. 2,5)

Let

$$J = \{0, \ldots, n-1\} \backslash \{l, l+p\}.$$

We impose the conditions:

(c1) The approximate solution (2.12) satisfies the differential equation (1.1) at ξ_j, $j \in J$, where ξ_j are the collocation points.

(c2) The solution satisfies $u_{\overline{\Delta}}(\xi_l) = \alpha$, $u_{\overline{\Delta}}(\xi_{l+p}) = \beta$ (we recall that $a = \xi_l$, $b = \xi_{l+p}$).

The conditions **(c1)** and **(c2)** yield a nonlinear system with n equations:

$$\begin{cases} \sum\limits_{i=0}^{n-1} c_i B_{i,k+1}(a) = \alpha, & j = l, \\ \sum\limits_{i=0}^{n-1} c_i B''_{i,k+1}(\xi_j) + f\left(\xi_j, \sum\limits_{i=0}^{n-1} c_i B_{i,k+1}(\xi_j)\right) = 0, & j \in J, \\ \sum\limits_{i=0}^{n-1} c_i B_{i,k+1}(b) = \beta, & j = l+p, \end{cases} \tag{2.13}$$

with unknowns $(c_i)_{i=0}^{n-1}$. If $F = [F_0, F_1, \ldots, F_{n-1}]^T$ are the functions defined by the equations of the nonlinear systems, using the quasilinearization of Newton method [4, pp. 52-55], we find the next approximation by means of

$$c^{(k+1)} = c^{(k)} - w^{(k)}, \tag{2.14}$$

where $c^{(k)}$ is the vector of unknowns obtained at the k-th step, and $w^{(k)}$ is the solution of the linear system:

$$F'(c^{(k)})w = F(c^{(k)}). \tag{2.15}$$

The Jacobian matrix $F' = (J_{ij})$ is banded and it is given by

$$J_{ij} = \begin{cases} B_{j,k+1}(a), & \text{for } i = l \\ B_{j,k+1}(b), & \text{for } i = l+p \\ B''_{j,k+1}(\xi_i) + \frac{\partial f}{\partial y}\left(\xi_i, \sum_{i=1}^{n-1} c_i B_{j,k+1}(\xi_i)\right) B_{j,k+1}(\xi_i), & \text{for } i \in J. \end{cases} \tag{2.16}$$

To solve (1.1)+(1.2)+(1.3) we use the method presented in [7, pp. 670-674] and [24, pp. 771-795]. An initial approximation $u^{(0)} \in C^1[0,1]$ is required.

The successful **stopping criterion** [1] is

$$\left\| u^{(k+1)} - u^{(k)} \right\| \le abstol + \left\| u^{(k+1)} \right\| reltol,$$

where, *abstol* and *reltol* is the absolute and the relative error tolerance, respectively, and the norm is the usual uniform convergence norm. The reliability of the error-estimation procedure being used for stopping criterion was verified in [3]. Papers on this topics exploit the almost block diagonal structure of collocation matrix and recommend an LU factorization (see [8, 3]).

3. A combined method using B-splines and Runge-Kutta methods

Our second method consists of the decomposition of the problem (1.1) + (1.2) + (1.3) into three problems:

1. A BVP on $[a, b]$ (problem (1.4)+(1.5)+(1.6));
2. Two IVPs on $[0, a]$ and $[b, 1]$.

Also we suppose that the problem (1.4)+(1.5)+(1.6) satisfies hypothesis of the Theorem 1.1, which ensures a sufficient condition to guarantee that there exists a $\sigma > 0$ such that $y(x)$ is the unique solution of problem BVP in the sphere

$$\{w : \|w - y''\| \le \sigma\}.$$

Due to conditions in Theorems 1.1 and 1.2, the problem (1.1)+(1.2)+(1.3) has a unique solution. To solve the problem (1.4)+(1.5)+(1.6), we use the collocation method presented in Section 2. This time, we consider a partition of $[a, b]$ as follows

$$\overline{\Delta} : a = x_0 < x_1 < \cdots < x_N = b.$$

The multiplicity of a and b is $k+2$ and the multiplicity of inner points is k. The dimension of spline space is again $Nk + 2$, and the nonlinear system is analogous to (2.13).

For the solution of the two initial value problems, we use a Runge-Kutta method of appropriate order. This needs good approximations of $y'(a)$ and $y'(b)$, which could be obtained with no additional effort during the collocation. Let $u_{\overline{\Delta}}(x)$ be the approximation computed by the combined method.

Theorem 3.1. *If u is an isolated solution of* (1.1)+ (1.2)+(1.3)*, f has continuous second order partial derivatives and the initial guess is sufficiently close to u, then the combined method is convergent to u and its accuracy is $O(h^{k+1})$, where h is the norm of the partition $\overline{\Delta}$ given by (*2.9*).*

Proof. For the problem (1.4)+(1.5)+(1.6) we apply Theorem 5.147, page 257 in [4]. We conclude that Newton method, applied to Δ, converges quadratically to the restriction of u to Δ, and the accuracy for the approximation and its derivative is $O(h^{k+1})$, that is

$$|u_{\Delta}^{(j)}(x) - y^{(j)}(x)| = O(h^{k+1}), \quad x \in [a,b],\ j = 0, 1.$$

We extend convergence and the accuracy to the whole interval $[0,1]$ by using the stability and the convergence of Runge-Kutta methods. A $(k+1)$-order explicit Runge-Kutta method is consistent and stable, so it is convergent, and its accuracy is $O(h^{k+1})$. Thus the final solution has the same accuracy. The stability and convergence of Runge-Kutta method are guaranteed by Theorems 5.3.1, page 285 and 5.3.2, page 288 in [13]. □

4. Some considerations on complexity

We will give a rough estimation of the complexity of our methods. We start with the first method. In the sequel, B will be the cost for B-spline evaluation and f the time for a function evaluation.

The time required to construct the collocation matrix is $C_0 = 2(Nk + 1)(k+2)B$.

To construct the Jacobian we need $Nk(k+2)(B+f)$. The construction of the right-hand side requires $(Nk+2)B + NkB + Nkf$. So, for the linear system construction, we obtain

$$W_1 = ((B+f)k^2 + (4B+3f)k)N + B$$

For a banded linear system with bandwidth w the total cost for solution, using LU with pivoting is $n(\frac{w^2}{2} + w)$ (see [12, pp. 79-80]). In our case, $n = Nk + 2$, and $w = \frac{3}{2}(k+2)$, and the cost for the solution of the linear system will be

$$W_2 = \left(\frac{21}{2}k + \frac{9}{8}k^3 + \frac{15}{2}k^2\right) N + 15k + 21 + \frac{9k^2}{4}.$$

The cost of Newton step is $W_s = W_1 + W_2$, that is,

$$W_s = \left[\frac{9k^3}{8} + \left(f + B + \frac{15}{2}\right) k^2 + \left(3f + 4B + \frac{21}{2}\right) k\right] N + 2B + 15k + 21 + \frac{9k^2}{4}.$$

The total cost is $IW_s + C_0$, where I is the number of steps required in Newton methods. Since the convergence is quadratic, if the final tolerance is ε, assuming $\delta_{i+1} = c\delta_i^2$, where δ_i is the error at the ith step, we obtain [18, pp. 295-297]

$$I = \frac{1}{\log 2} \log \frac{\log|c| + \log \varepsilon}{\log|c| + \log|\delta_0|}.$$

For the second method, the same analysis works for BVP solution part. We have an additional amount of work for Runge-Kutta method. If the number of stages is s and the number of points is p, the cost is $O(psf)$.

5. Implementation and numerical examples

We implemented the ideas from previous sections in MATLAB 2010a. Our code uses MATLAB Spline Toolbox and sparse matrices (see [26]). The function `spcol` allows us to compute easily the collocation matrix. For IVPs the solver `ode45` works fine. To avoid the error propagation, we chose for (BVP) B-splines of order 4 (degree 3) or order 5 (degree 4).

We implemented two functions: `polycollocnelin`, global B-spline collocation, and `polycalnlinRK`, the combined method (B-spline collocation + Runge-Kutta).

Consider the following examples:

1. [14] Consider the PVP
$$\begin{aligned} &y''(x) + y^3(x) + \frac{4-(x-x^2)^3}{(x+1)^3} = 0;\ x \in (0,1) \\ &y(1/4) = 3/20;\ y(1/2) = 1/6 \end{aligned} \tag{5.1}$$
with exact solution
$$y(x) = \frac{x-x^2}{x+1}.$$
2. $$\begin{aligned} &y''(x) + e^{-y(x)} = 0;\ x \in [0,1] \\ &y(\pi/6) = \ln(3/2),\ y(\pi/4) = \ln((2+\sqrt{2})/2) \end{aligned} \tag{5.2}$$
with the exact solution
$$y = \ln(\sin(x)+1).$$

We applied both methods to each example.

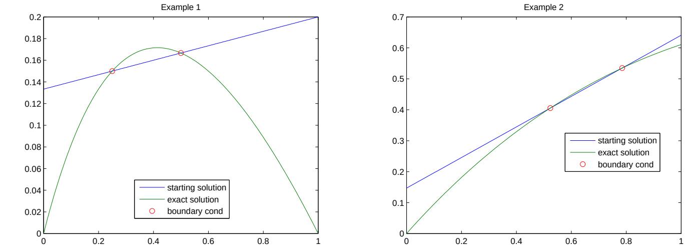
Figure 1 shows the exact solutions and the starting functions. The error plots for both methods, in semi-logarithmic scale, are given in Figure 2 for the first example and in Figure 3 for the second example, respectively.

We chose as starting function the Lagrange interpolation polynomial that takes the values α and β at a and b.

Table 1 gives the residuals $e_\Delta^{(j)}\|y^{(j)} - y_\Delta^{(j)}\|$, for $j = 0,1,2$, for the global method based on B-splines. For the residuals it holds $\|y^{(j)} - y_\Delta^{(j)}\| = O(|\Delta\|)^{k+2-j}$, for $j = 0,1,2$. To check this experimentally, we plot the residuals versus $1/\Delta$, for various values of N in a log-log scale (see Figure 4, the left column, for Example (5.1) and Figure 4, the right column, for Example (5.2)).

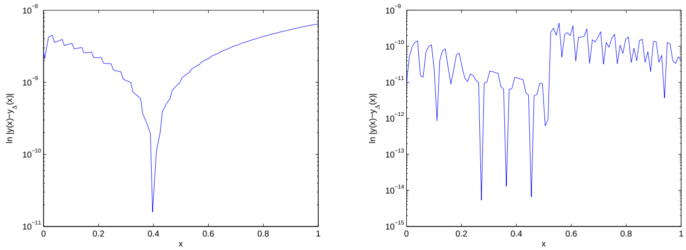
In order to compare the costs (run-times) experimentally we used MATLAB functions `tic` and `toc`. The results are given in Table 2.

The time for combined method is a bit larger.



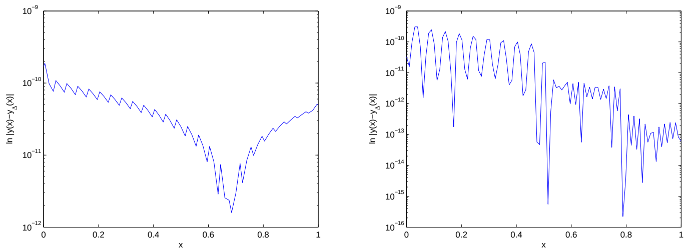
(a) Example 1 (b) Example 2

FIGURE 1. Exact solution and starting approximation



(a) Global collocation (b) Combined method

FIGURE 2. Error plot for example (5.1)



(a) Global collocation (b) Combined method

FIGURE 3. Error plot for example (5.2)

The next numerical experiment compares the running time of our methods to the running time of a pseudospectral method (see [19] for implementation details of the latter). As example, we consider a variant of Bratu's

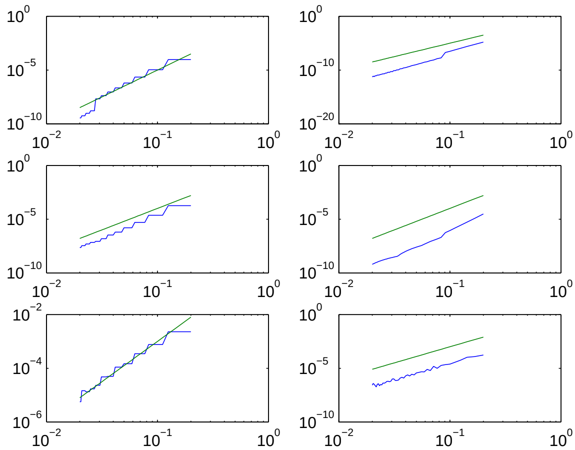


Figure 4. Order estimation for Example (5.1) (left) and Example (5.2) (right): $e_{\Delta}^{(0)}$ - up, $e_{\Delta}^{(1)}$ - middle, and $e_{\Delta}^{(2)}$ - bottom

	Example (5.1)			Example (5.2)		
N	$\|y-y_\Delta\|$	$\|y'-y'_\Delta\|$	$\|y''-y''_\Delta\|$	$\|y-y_\Delta\|$	$\|y'-y'_\Delta\|$	$\|y''-y''_\Delta\|$
5	6e-05	0.000123	0.00229	1.04e-05	2.15e-05	0.000172
6	6e-05	0.000123	0.00229	3.72e-06	8.47e-06	0.000124
7	6e-05	0.000123	0.00229	1.59e-06	3.96e-06	0.000106
8	6e-05	0.000123	0.00229	7.37e-07	2.02e-06	5.65e-05
9	6.9e-06	1.63e-05	0.000769	3.84e-07	1.16e-06	3.61e-05
10	6.9e-06	1.63e-05	0.000769	2.04e-07	6.81e-07	2.42e-05
11	6.9e-06	1.63e-05	0.000769	1.21e-07	4.46e-07	2.14e-05
12	6.9e-06	1.63e-05	0.000769	1.82e-08	2.02e-07	1.81e-05
13	1.4e-06	3.64e-06	0.000346	1.3e-08	1.44e-07	9.77e-06
14	1.4e-06	3.64e-06	0.000346	7.54e-09	1.09e-07	1.51e-05
15	1.4e-06	3.64e-06	0.000346	5.54e-09	8.32e-08	6.08e-06
16	1.4e-06	3.64e-06	0.000346	3.54e-09	6.22e-08	7.92e-06
17	3.99e-07	1.22e-06	0.000148	2.81e-09	4.65e-08	4.65e-06
18	3.99e-07	1.22e-06	0.000148	1.89e-09	3.51e-08	4.76e-06
19	3.99e-07	1.22e-06	0.000148	1.43e-09	2.98e-08	4.14e-06
20	3.99e-07	1.22e-06	0.000148	1.02e-09	2.5e-08	3.7e-06

Table 1. Error table for Examples (5.1) and (5.2)

problem [4, page 491] for $\lambda = 1$

$$
\begin{aligned}
&y'' + e^y = 0, \qquad x \in (0,1)\\
&y(0.2) = y(0.8) = 0.08918993462883.
\end{aligned}
$$

	Method 1	Method 2
First example	0.017501	0.022751
Second example	0.016387	0.021535

TABLE 2. Run times

ε	PseudoS	BS+RK	Global BS
10^{-5}	0.054	0.035	0.021
10^{-6}	0.077	0.043	0.023
10^{-7}	0.049	0.025	0.024
10^{-8}	0.055	0.031	0.031
10^{-9}	0.054	0.036	0.030
10^{-10}	0.058	0.026	0.028

TABLE 3. Running times for Bratu's problem

We chose 128 collocation points, and as starting function $y_0(t) = \frac{39}{70}x(x-1)$. The running times for various tolerances are given in Table 3.

References

[1] Ascher, U., *Solving Boundary Value Problems with a Spline Collocation Code*, Journal of Computational Physics, **34**(1980), no. 3, 401–413.

[2] Ascher, U., *Discrete least squares approximations for ordinary differential equations*, SIAM J. Numer. Anal., **15**(1978), 478–496.

[3] Ascher, U., Christiansen, J., Russel, R. D., *A collocation solver for mixed order systems of boundary value problems*, Math. Comp., **33**(1979), 659–679.

[4] Ascher, U., Mattheij, R.M., Russel, R.D., *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, SIAM, 1997.

[5] Ascher, U., Pruess S., Russel, R.D., *On spline basis selection for solving differential equations*, SIAM J. Numer. Anal. **20**(1983), no. 1, 121–142.

[6] Cerruti, J., *Collocation for Systems of Ordinary Differential Equations*, Comp. Sci. Tech. Rep. 230, Univ. Wisconsin-Madison, 1974.

[7] Christiansen, J., Ascher, U., Russel, R. D., *A Collocation Solver for Mixed Order Systems of Boundary Value Problems*, Math. Comp., **13**(1979) no. 146, 659–679.

[8] de Boor, C., Weiss, R., *Solveblock: A Package for Solving Almost Block Diagonal Linear Systems, with Applications to Spline Approximations and the Numerical Solution of Ordinary Differential Equations*, MRC TSR #1625, Madison, Wisconsin,1976.

[9] de Boor C., *On calculating with B-splines*, J. Approximation Theory, **6**(1972), 50–62.

[10] de Boor, C., *Good approximation by B-splines with variable knots II*, Lecture Notes in Math., vol.363, Springer-Verlag, Berlin, New York, 1973.

[11] de Boor, C., *A Practical Guide to Splines*, Springer-Verlag, Berlin, Heidelberg, New York, 1978.

[12] Demmel, J.W., *Applied Numerical Linear Algebra*, SIAM, 1997.

[13] Gautschi, W., *Numerical Analysis. An Introduction*, Birkhäuser, Boston, Basel, Berlin, 1997.

[14] Goldner, G., Trîmbiţaş, R. T., *A combined method for two point boundary value problem*, P.U.M.A, **11**(2000), no. 2, 255–264.

[15] Henrici, P., *Discrete variable methods in ordinary differential equations*, John Wiley, New York, 1962.

[16] Houstis, E., *A collocation method for systems of nonlinear ordinary differential equations*, J. Math. Anal. Appl., **62**(1978), 24–37.

[17] Russel, R.D., *Efficients of B-splines methods for solving differential equations*, Proc. Fifth Conference on Numerical Mathematics, Utilitas Math. Winnipeg, Manitoba, 1975, 599–617.

[18] Pizer, S.M., Wallace, V.L., *To Compute Numerically. Concepts and Strategies*, Little, Brown and Company, 1983.

[19] Pop, D.N., Trîmbiţaş, R., Păvăloiu, I. *Solution of a polylocal problem with a pseudospectral method*, Annals of the Tiberiu Popoviciu Seminar, **8**(2010), 53-63.

[20] Risler, J. J., *Méthodes mathematiques pour la CAO*, Masson, Paris, Milan, Barcelone, Bonn, 1991.

[21] Russel, R.D., *A comparation of collocation and finite differences for two point boundary value problems*, SIAM J. Numer. Anal., **14**(1977), 19–39.

[22] Russel, R.D., Shampine, L.F., *A Collocation Method for Boundary Value Problems*, Numer. Math., **19**(1972), 1–28.

[23] Schultz, M.H., *Spline Analysis*, Prentice Hall, Englewood Cliffs, N.J., 1972.

[24] Stover, R., *Collocation methods for solving linear for solving linear differential-algebraic boundary value problems*, Numer. Math., **88**(2001), 771–795.

[25] Vainniko, G.M., *On the stability and convergence of the collocation method*, Differentsial'nye Uravneniya, **1**(1965), 244–254.

[26] MathWorks Inc., MATLAB 2010a, *Set of manuals*, Natick, MA, 2010.

Daniel N. Pop
Romanian-German University
Sibiu
Romania
e-mail: `danielnicolaepop@yahoo.com`

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`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 527–535

On the rate of convergence of a new q-Szász-Mirakjan operator

Cristina Radu, Saddika Tarabie and Andreea Veţeleanu

Abstract. In the present paper we introduce a new q-generalization of Szász-Mirakjan operators and we investigate their approximation properties. By using a weighted modulus of smoothness, we give local and global estimations for the error of approximation.

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

Keywords: q-calculus, Szász-Mirakjan operator, Bohman-Korovkin type theorem, weighted space, weighted modulus of smoothness.

1. Introduction

The aim of this paper is to study the approximation properties of a new Szász-Mirakjan type operator constructed by using q-Calculus. Firstly, we recall some basic definitions and notations used in quantum calculus, see, e.g., [6, pp. 7-13].

Let $q > 0$. For any $n \in \mathbb{N}_0 := \{0\} \cup \mathbb{N}$ the q-integer $[n]_q$ is defined by

$$[n]_q := 1 + q + \dots q^{n-1} \quad (n \in \mathbb{N}), \quad [0]_q := 0,$$

and the q-factorial $[n]_q!$ by

$$[n]_q! := [1]_q[2]_q \dots [n]_q \quad (n \in \mathbb{N}), \quad [0]_q! := 1.$$

Also, the q-binomial coefficients are denoted by $\left[\begin{array}{c} n \\ k \end{array}\right]_q$ and are defined by

$$\left[\begin{array}{c} n \\ k \end{array}\right]_q = \frac{[n]_q!}{[k]_q!\,[n-k]_q!}, \quad k = 0, 1, \dots, n.$$

The q-derivative of a function $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$D_q f(x) = \frac{f(x) - f(qx)}{(1-q)x}, \quad x \neq 0, \quad D_q f(0) := \lim_{x \to 0} D_q f(x),$$

and the high q-derivatives $D_q^0 f := f$, $D_q^n f := D_q\left(D_q^{n-1} f\right)$, $n \in \mathbb{N}$.

The product rule is

$$D_q\left(f(x)g(x)\right) = D_q\left(f(x)\right)g(x) + f(qx)D_q\left(g(x)\right). \tag{1.1}$$

We recall the q-Taylor theorem as it is given in [4, p. 103].

Theorem 1.1. *If the function $g(x)$ is capable of expansion as a convergent power series and q is not a root of unity, then*

$$g(x) = \sum_{r=0}^{\infty} \frac{(x-a)_q^r}{[r]_q!} D_q^r g(a),$$

where

$$(x-a)_q^r = \prod_{s=0}^{r-1}(x - q^s a) = \sum_{k=0}^{r} \left[\begin{array}{c} r \\ k \end{array}\right]_q q^{\frac{k(k-1)}{2}} x^{r-k}(-a)^k.$$

2. Auxiliary results

Throughout the paper we consider $q \in (0,1)$.

We define a suitable q-difference operator as follows

$$\Delta_q^0 f_{k,s} = f_{k,s}, \tag{2.1}$$

$$\Delta_q^{r+1} f_{k,s} = q^r \Delta_q^r f_{k+1,s} - \Delta_q^r f_{k,s-1}, \quad r \in \mathbb{N}_0, \tag{2.2}$$

where $f_{k,s} = f\left(\frac{[k]_q}{q^s[n]_q}\right)$, $k \in \mathbb{N}_0$, $s \in \mathbb{Z}$.

The following lemma gives an expression for the r-th q-differences $\Delta_q^r f_{k,s}$ as a sum of multiplies of values of f.

Lemma 2.1. *The q-difference operator Δ_q^r defined by (2.1)-(2.2)satisfies*

$$\Delta_q^r f_{k,s} = \sum_{j=0}^{r} (-1)^{r-j} q^{j(j-1)/2} \left[\begin{array}{c} r \\ j \end{array}\right]_q f_{k+j,j+s-r} \quad \textit{for} \quad r,k \in \mathbb{N}_0, \quad s \in \mathbb{Z}. \tag{2.3}$$

Taking into account the relations (2.1)-(2.2) and the formula

$$\left[\begin{array}{c} r+1 \\ j+1 \end{array}\right]_q = q^{r-j} \left[\begin{array}{c} r \\ j \end{array}\right]_q + \left[\begin{array}{c} r \\ j+1 \end{array}\right]_q,$$

the identity (2.3) can be easily obtained by induction over $r \in \mathbb{N}_0$.

In what follows, the monomial of m degree is denoted by e_m, $m \in \mathbb{N}_0$.

Let us denote by $[x_0, x_1, \ldots, x_n; f]$ the divided difference of the function f with respect to the points $x_0, x_1, \ldots, x_n$.

Lemma 2.2. *For all $k, r \in \mathbb{N}_0$, $s \in \mathbb{Z}$, we have*

$$[x_{k,s-1}, \ldots, x_{k+r,s+r-1}; f] = \frac{q^{r(r+2s-1)/2}[n]_q^r}{[r]_q!} \Delta_q^r f_{k,r+s-1}, \tag{2.4}$$

where $x_{k,s-1} = \frac{[k]_q}{q^{s-1}[n]_q}$.

Proof. We use the mathematical induction with respect to r. For $r = 0$ the equality (2.4) follows immediately from (2.1). Let us assume that (2.4) holds true for some $r \geq 0$ and all $k \in \mathbb{N}_0$, $s \in \mathbb{Z}$.

We have

$$\begin{aligned}&[x_{k,s-1}, \dots, x_{k+r+1,s+r}; f] \\ &= \frac{[x_{k+1,s}, \dots, x_{k+r+1,s+r}; f] - [x_{k,s-1}, \dots, x_{k+r,s+r-1}; f]}{x_{k+r+1,s+r} - x_{k,s-1}}.\end{aligned}$$

Since $x_{k+r+1,s+r} - x_{k,s-1} = \frac{[r+1]_q}{q^{r+s}[n]_q}$, by using (2.2) we get

$$\begin{aligned}&[x_{k,s-1}, \dots, x_{k+r+1,s+r}; f] \\ &= \frac{q^{(r+1)(r+2s)/2}[n]_q^{r+1}}{[r+1]_q!} \left(q^r \Delta_q^r f_{k+1,r+s} - \Delta_q^r f_{k,r+s-1}\right) \\ &= \frac{q^{(r+1)(r+2s)/2}[n]_q^{r+1}}{[r+1]_q!} \Delta_q^{r+1} f_{k,r+s}.\end{aligned}$$

□

3. Construction of the operators

In 1987 A. Lupaş [9] introduced the first q-analogue of Bernstein operator and investigated its approximating and shape-preserving properties. Another q-generalization of the classical Bernstein polynomials is due to G. Phillips [13]. More properties of these two q-extensions were obtained over time in several papers such as [3], [10], [11], [1]. We mention that the comprehensive survey [12] due to S. Ostrovska gives a good perspective of the most important achievements during a decade relative to these operators.

Two of the known expansions in q-calculus of the exponential function are given as follows (see, e.g., [6, p. 31])

$$E_q(x) = \sum_{k=0}^{\infty} q^{k(k-1)/2} \frac{x^k}{[k]_q!}, \quad x \in \mathbb{R}, \quad |q| < 1,$$

$$e_q(x) = \sum_{k=0}^{\infty} \frac{x^k}{[k]_q!}, \quad |x| < \frac{1}{1-q}, \quad |q| < 1.$$

It is obvious that $\lim\limits_{q \to 1^-} E_q(x) = \lim\limits_{q \to 1^-} e_q(x) = e^x$.

For $q \in (0,1)$, in [2] A. Aral introduced the first q-analogue of the classical Szász-Mirakjan operators given by

$$S_n^q(f; x) = E_q\left(-[n]_q \frac{x}{b_n}\right) \sum_{k=0}^{\infty} f\left(\frac{[k]_q b_n}{[n]_q}\right) \frac{\left([n]_q x\right)^k}{[k]_q!(b_n)^k},$$

where $0 \leq x < \frac{b_n}{1-q^n}$, $(b_n)_n$ is a sequence of positive numbers such that $\lim\limits_n b_n = \infty$.

The operator S_n^q reproduces linear functions and

$$S_n^q(e_2;x) = qx^2 + \frac{b_n}{[n]_q}x, \quad 0 \le x < \frac{b_n}{1-q^n}.$$

Motivated by this work, for $q \in (0,1)$ we give another q-analogue of the same class of operators as follows

$$S_{n,q}(f;x) = \sum_{k=0}^{\infty} \frac{\left([n]_q x\right)^k}{[k]_q!} q^{k(k-1)} E_q\left(-[n]_q q^k x\right) f\left(\frac{[k]_q}{[n]_q q^{k-1}}\right), \quad x \ge 0, \tag{3.1}$$

where $f \in \mathcal{F}(\mathbb{R}_+) := \{f : \mathbb{R}_+ \to \mathbb{R}, \ \text{ the series in } \ (3.1) \ \text{ is convergent}\}$.

Since $E_q(x)$ is convergent for every $x \in \mathbb{R}$, by using Theorem 1.1 and the property $D_q^r E_q(x) = q^{\frac{r(r-1)}{2}} E_q(q^r x)$ we obtain

$$\sum_{r=0}^{\infty} \frac{(-x)^r}{[r]_q!} q^{r(r-1)} E_q(q^r x) = E_q(0) = 1, \quad x \in \mathbb{R},$$

which yields that the operator $S_{n,q}$ is well defined.

For $q \to 1^-$, the above operators reduce to the classical Szász-Mirakjan operators. In this case, the approximation function $S_{n,q}f$ is defined on $\mathbb{R}_+$ for each $n \in \mathbb{N}$.

Theorem 3.1. *Let $q \in (0,1)$ and $S_{n,q}$, $n \in \mathbb{N}$, be defined by (3.1). For any $f \in \mathcal{F}(\mathbb{R}_+)$ we have*

$$S_{n,q}(f;x) = \sum_{r=0}^{\infty} \frac{([n]_q x)^r}{[r]_q!} q^{\frac{r(r-1)}{2}} \Delta_q^r f_{0,r-1}, \quad x \ge 0. \tag{3.2}$$

Proof. Let $f \in \mathcal{F}(\mathbb{R}_+)$.

By using (2.1), the operator $S_{n,q}$ can be expressed as follows

$$S_{n,q}(f;x) = \sum_{k=0}^{\infty} \frac{\left([n]_q x\right)^k}{[k]_q!} q^{k(k-1)} E_q\left(-[n]_q q^k x\right) \Delta_q^0 f_{k,k-1}.$$

Applying q-derivative operator to $S_{n,q}f$ and taking into account the product rule (1.1) and the property $D_q E_q(ax) = aE_q(aqx)$, (see e.g. [6, pp. 29-32]), we have

$$\begin{aligned} & D_q S_{n,q}(f;x) \\ & = [n]_q \sum_{k=0}^{\infty} \frac{\left([n]_q x\right)^k}{[k]_q!} q^{k(k+1)} E_q\left(-[n]_q q^{k+1} x\right) \left(\Delta_q^0 f_{k+1,k} - \Delta_q^0 f_{k,k-1}\right) \\ & = [n]_q \sum_{k=0}^{\infty} \frac{\left([n]_q x\right)^k}{[k]_q!} q^{k(k+1)} E_q\left(-[n]_q q^{k+1} x\right) \Delta_q^1 f_{k,k}. \end{aligned}$$

For $n \in \mathbb{N}$ and $x \in \mathbb{R}_+$, by induction with respect to $r \in \mathbb{N}$, we can prove

$$D_q^r S_{n,q}(f;x)$$
$$= [n]_q^r q^{\frac{r(r-1)}{2}} \sum_{k=0}^{\infty} \frac{\left([n]_q\, x\right)^k}{[k]_q!} q^{k(2r+k-1)} E_q\left(-[n]_q\, q^{k+r} x\right) \Delta_q^r f_{k,k+r-1}.$$

Choosing $x = 0$, we deduce $D_q^r S_{n,q}(f;0) = [n]_q^r q^{\frac{r(r-1)}{2}} \Delta_q^r f_{0,r-1}$.
Choosing $a = 0$ in Theorem 1.1, we obtain

$$S_{n,q}(f;x) = \sum_{r=0}^{\infty} \frac{([n]_q x)^r}{[r]_q!} q^{\frac{r(r-1)}{2}} \Delta_q^r f_{0,r-1},$$

which completes the proof. □

Corollary 3.2. *Let $q \in (0,1)$ and $S_{n,q}$, $n \in \mathbb{N}$, be defined by (3.1). For any $f \in \mathcal{F}(\mathbb{R}_+)$ we have*

$$S_{n,q}(f;x) = \sum_{r=0}^{\infty} x^r \left[0, \frac{1}{[n]_q}, \frac{[2]_q}{q[n]_q}, \ldots, \frac{[r]_q}{q^{r-1}[n]_q}; f\right], \quad x \geq 0. \tag{3.3}$$

Proof. The identity (3.3) is obtained from the above theorem and (2.4) by choosing $k = s = 0$. □

Corollary 3.3. *For all $n \in \mathbb{N}$, $x \in \mathbb{R}_+$ and $0 < q < 1$, we have*

$$S_{n,q}(e_0;x) = 1, \tag{3.4}$$
$$S_{n,q}(e_1;x) = x, \tag{3.5}$$
$$S_{n,q}(e_2;x) = x^2 + \frac{1}{[n]_q} x. \tag{3.6}$$

Moreover, for $m \in \mathbb{N}_0$ and $0 < q < 1$, the operator $S_{n,q}$ defined by (3.1) can be expressed as

$$S_{n,q}(e_m;x) = \sum_{r=0}^{m} x^r \left[0, \frac{1}{[n]_q}, \frac{[2]_q}{q[n]_q}, \ldots, \frac{[r]_q}{q^{r-1}[n]_q}; e_m\right], \quad x \geq 0. \tag{3.7}$$

Proof. Since for any distinct points $x_0, \ldots, x_r$, the divided difference

$$[x_0, \ldots, x_r; e_m] = \begin{cases} 0 & if \quad m < r, \\ 1 & if \quad m = r, \\ x_0 + \ldots + x_r & if \quad m = r+1, \end{cases}$$

(see e.g. [5, p.63]), the identities (3.4)-(3.7) are obvious. □

Lemma 3.4. *For $m \in \mathbb{N}_0$ and $q \in (0,1)$ we have*

$$S_{n,q}(e_m;x) \leq A_{m,q}(1 + x^m), \quad x \geq 0, \quad n \in \mathbb{N}, \tag{3.8}$$

where $A_{m,q}$ is a positive constant depending only on q and m.

Proof. Let $m \in \mathbb{N}$. From (3.7) we get

$$S_{n,q}(e_m;x) \leq (1+x^m)\sum_{r=1}^{m}\left[0,\frac{1}{[n]_q},\ldots,\frac{[r]_q}{q^{r-1}[n]_q};e_m\right].$$

Applying the well known Lagrange's Mean Value Theorem, we can write

$$S_{n,q}(e_m;x) \leq (1+x^m)\sum_{r=1}^{m}\binom{m}{r}(\xi_r)^{m-r},$$

where $0<\xi_r<\frac{[r]_q}{q^{r-1}[n]_q}$, $0<r\leq m$.

Consequently, we have

$$\begin{aligned} S_{n,q}(e_m;x) &\leq (1+x^m)\sum_{r=1}^{m}\binom{m}{r}\frac{[r]_q^{m-r}}{q^{(r-1)(m-r)}[n]_q^{m-r}} \\ &\leq (1+x^m)[m]_q^{m-1}\sum_{r=1}^{m}\binom{m}{r}\frac{1}{q^{(r-1)(m-r)}q^{m-r+r^2}} \\ &\leq A_{m,q}(1+x^m), \end{aligned}$$

where

$$A_{m,q} := [m]_q^{m-1}\left(1+\frac{1}{q^m}\right)^m, \quad m \geq 1. \tag{3.9}$$

For $m=0$ we can take $A_{0,q}=\frac{1}{2}$. □

Examining relation (3.6) it is clear that the sequence of the operators $(S_{n,q})_n$ does not satisfies the conditions of Bohman-Korovkin theorem.

Further on, we consider a sequence $(q_n)_n$, $q_n \in (0,1)$, such that

$$\lim_n q_n = 1. \tag{3.10}$$

The condition (3.10) guarantees that $[n]_{q_n} \to \infty$ for $n \to \infty$.

Theorem 3.5. *Let $(q_n)_n$ be a sequence satisfying (3.10) and let the operators S_{n,q_n}, $n \in \mathbb{N}$, be defined by (3.1). For any compact $J \subset \mathbb{R}_+$ and for each $f \in C(\mathbb{R}_+)$ we have*

$$\lim_{n\to\infty} S_{n,q_n}(f;x) = f(x), \quad \textit{uniformly in } x \in J.$$

Proof. Replacing q by a sequence $(q_n)_n$ with the given conditions, the result follows from (3.4)-(3.6) and the well-known Bohman-Korovkin theorem (see [7], pp. 8-9). □

4. Error of approximation

Let $\alpha \in \mathbb{N}$. We denote by $B_\alpha(\mathbb{R}_+)$ the weighted space of real-valued functions f defined on $\mathbb{R}_+$ with the property $|f(x)| \leq M_f(1+x^\alpha)$ for all $x \in \mathbb{R}_+$, where M_f is a constant depending on the function f. We also consider the weighted subspace $C_\alpha(\mathbb{R}_+)$ of $B_\alpha(\mathbb{R}_+)$ given by

$$C_\alpha(\mathbb{R}_+) := \{f \in B_\alpha(\mathbb{R}_+) : f \text{ continuous on } \mathbb{R}_+\}.$$

Endowed with the norm $\|\cdot\|_\alpha$, where $\|f\|_\alpha := \sup\limits_{x\in\mathbb{R}_+} \frac{|f(x)|}{1+x^\alpha}$, both $B_\alpha(\mathbb{R}_+)$ and $C_\alpha(\mathbb{R}_+)$ are Banach spaces.

We can give estimates of the error $|S_{n,q}(f;\cdot) - f|$, $n \in \mathbb{N}$, for unbounded functions by using a weighted modulus of smoothness associated to the space $B_\alpha(\mathbb{R}_+)$.

We consider

$$\Omega_\alpha(f;\delta) := \sup_{\substack{x\geq 0\\ 0<h\leq\delta}} \frac{|f(x+h)-f(x)|}{1+(x+h)^\alpha}, \delta > 0,\ \alpha \in \mathbb{N}. \tag{4.1}$$

It is evident that for each $f \in B_\alpha(\mathbb{R}_+)$, $\Omega_\alpha(f;\cdot)$ is well defined and

$$\Omega_\alpha(f;\delta) \leq 2\|f\|_\alpha, \delta > 0, \quad f \in B_\alpha(\mathbb{R}_+),\ \alpha \in \mathbb{N}.$$

The weighted modulus of smoothness $\Omega_\alpha(f;\cdot)$ possesses the following properties ([8]).

$$\Omega_\alpha(f;\lambda\delta) \leq (\lambda+1)\Omega_\alpha(f;\delta), \quad \delta > 0, \lambda > 0, \tag{4.2}$$
$$\Omega_\alpha(f;n\delta) \leq n\Omega_\alpha(f;\delta), \quad \delta > 0, n \in \mathbb{N},$$
$$\lim_{\delta\to 0^+} \Omega_\alpha(f;\delta) = 0.$$

Theorem 4.1. *Let $(q_n)_n$ be a sequence satisfying (3.10). Let $q_0 = \inf\limits_{n\in\mathbb{N}} q_n$ and $\alpha \in \mathbb{N}$. For each $n \in \mathbb{N}$ and every $f \in B_\alpha(\mathbb{R}_+)$ one has*

$$|S_{n,q_n}(f;x) - f(x)| \leq C_{\alpha,q_0}(1+x^{\alpha+1})\Omega_\alpha\left(f;\sqrt{1/[n]_{q_n}}\right), \quad x \geq 0, \tag{4.3}$$

where C_{α,q_0} is a positive constant independent of f and n.

Proof. Let $n \in \mathbb{N}$, $f \in B_\alpha(\mathbb{R}_+)$ and $x \geq 0$ be fixed. Setting $\mu_{x,\alpha}(t) := 1+(x+|t-x|)^\alpha$ and $\psi_x(t) := |t-x|$, $t \geq 0$, relations (4.1) and (4.2) imply

$$\begin{aligned}|f(t)-f(x)| &\leq (1+(x+|t-x|)^\alpha)\left(1+\frac{1}{\delta}|t-x|\right)\Omega_\alpha(f;\delta)\\ &= \mu_{x,\alpha}(t)\left(1+\frac{1}{\delta}\psi_x(t)\right)\Omega_\alpha(f;\delta), \quad t \geq 0.\end{aligned}$$

By using the Cauchy inequality for linear positive operators which preserve the constants, we obtain

$$\begin{aligned}|S_{n,q_n}(f;x) - f(x)| &\leq S_{n,q_n}(|f-f(x)|;x) \\ &\leq \left(S_{n,q_n}(\mu_{x,\alpha};x) + \frac{1}{\delta}S_{n,q_n}(\mu_{x,\alpha}\psi_x;x)\right)\Omega_\alpha(f;\delta)\\ &\leq \sqrt{S_{n,q_n}(\mu_{x,\alpha}^2;x)}\left(1+\frac{1}{\delta}\sqrt{S_{n,q_n}(\psi_x^2;x)}\right)\Omega_\alpha(f;\delta).\end{aligned} \tag{4.4}$$

Since

$$\begin{aligned}\mu_{x,\alpha}^2(t) &= (1+(x+|t-x|)^\alpha)^2 \leq 2\left(1+(2x+t)^{2\alpha}\right)\\ &\leq 2\left(1+2^{2\alpha}\left((2x)^{2\alpha}+t^{2\alpha}\right)\right),\end{aligned}$$

and taking into account (3.4) and (3.8) we get

$$S_{n,q_n}(\mu_{x,\alpha}^2;x) \le B_{\alpha,q_n}^2(1+x^{2\alpha}), \tag{4.5}$$

where $B_{\alpha,q_n}^2 = 2^{\alpha+1}\left(2^{2\alpha}+A_{2\alpha,q_n}\right)$.
According to (3.4)-(3.6) we have $S_{n,q_n}(\psi_x^2;x) = \frac{1}{[n]_{q_n}}x$.
By choosing $\delta := \sqrt{\frac{1}{[n]_{q_n}}}$ in (4.3), from (4.5) follows

$$|S_{n,q_n}(f;x)-f(x)| \le B_{\alpha,q_n}\sqrt{1+x^{2\alpha}}(1+\sqrt{x})\Omega_\alpha\left(f;\sqrt{\frac{1}{[n]_{q_n}}}\right).$$

Finally, since $1+\sqrt{x} \le \sqrt{2}\sqrt{1+x}$ and $(1+x^{2\alpha})(1+x) \le 4(1+x^{\alpha+1})$ for $x \ge 0$ and $\alpha \in \mathbb{N}$, we obtain

$$|S_{n,q_n}(f;x)-f(x)| \le C_{\alpha,q_0}(1+x^{\alpha+1})\Omega_\alpha\left(f;\sqrt{1/[n]_{q_n}}\right), \quad x \ge 0,$$

where $q_0 := \inf_{n\in\mathbb{N}} q_n$ and $C_{\alpha,q_0} := 2\sqrt{2}B_{\alpha,q_0}$. □

On the basis of Theorem 4.1 we give the following global estimate.

Corollary 4.2. *Let $(q_n)_n$ be a sequence satisfying (3.10) and $\alpha \in \mathbb{N}$. For each $n \in \mathbb{N}$ and every $f \in B_\alpha(\mathbb{R}_+)$ one has*

$$\|S_{n,q_n}(f;\cdot)-f\|_{\alpha+1} \le C_{\alpha,q_0}\Omega_\alpha\left(f;\sqrt{1/[n]_{q_n}}\right),$$

where C_{α,q_0} is a positive constant independent of f and n.

Remark 4.3. For any function $f \in B_\alpha(\mathbb{R}_+)$, $\alpha \in \mathbb{N}$, the rate of convergence of the operators $S_{n,q_n}(f;\cdot)$ to f in weighted norm is $\sqrt{\frac{1}{[n]_{q_n}}}$ which is faster than $\sqrt{\frac{b_n}{[n]_{q_n}}}$ obtained in [2].

References

[1] Agratini, O., *On certain q-analogues of the Bernstein operators*, Carpathian J. Math., **24**(2008), no. 3, 281-286.

[2] Aral, A., *A generalization of Szász-Mirakjan operators based on q-integers*, Math. Comput. Model., **47**(2008), 1052–1062.

[3] Bărbosu, D., *Some generalized bivariate Bernstein operators*, Math. Notes (Miskolc), **1**(2000), 3–10.

[4] Ernst, T., *The history of q-calculus and a new method*, U.U.D.M. Report 2000, 16, Uppsala, Department of Mathematics, Uppsala University, 2000.

[5] Ivan, M., *Elements of Interpolation Theory*, Mediamira Science Publisher, Cluj-Napoca, 2004.

[6] Kac, V., Cheung, P., *Quantum Calculus*, Universitext, Springer-Verlag, New York, 2002.

[7] Lorentz, G. G., *Bernstein Polynomials*, Math. Expo. Vol. 8, Univ. of Toronto Press, Toronto, 1953.

[8] López-Moreno, A.-J., *Weighted silmultaneous approximation with Baskakov type operators*, Acta Mathematica Hungarica, **104**(2004), 143–151.

[9] Lupaş, A., *A q-analogue of the Bernstein operator*, Seminar on Numerical and Statistical Calculus, University of Cluj-Napoca, **9**(1987), 85–92.

[10] Ostrovska, S., *On the Lupas q-analogue of the Bernstein operator*, Rocky Mountain Journal of Mathematics, **36**(2006), 1615–1629.

[11] Ostrovska, S., *On the improvement of analytic properties under the limit q-Bernstein operator*, J. Approx. Theory, **138**(2006), 37–53.

[12] Ostrovska, S., *The first decade of the q-Bernstein polynomials :results and perspectives*, J. Math. Anal. Approx. Theory, **2** (2007), 35–51.

[13] Phillips, G.M., *Bernstein polynomials based on the q-integers*, Ann. Numer. Math., **4**(1997), 511–518.

Cristina Radu
Schuller Eh Klar,
47, Calea Baciului,
400230 Cluj-Napoca,
Romania
e-mail: `radu.cristina@schuller.ro`

Saddika Tarabie
Tishrin University,
Faculty of Sciences
1267 Latakia,
Syria
e-mail: `sadikatorbey@yahoo.com`

Andreea Veţeleanu
"Babeş-Bolyai" University,
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street,
400084 Cluj-Napoca,
Romania
e-mail: `andreeav@math.ubbcluj.ro`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 537–544

Discrete operators associated with certain integral operators

Ioan Raşa

Abstract. We associate to a given sequence of positive linear integral operators a sequence of discrete operators and investigate the relationship between the two sequences. Several examples illustrate the general results.

Mathematics Subject Classification (2010): 41A36.

Keywords: Positive linear operators.

1. Introduction

Let $I_n : C[a,b] \longrightarrow C[a,b],\ n \geq 1$, be a sequence of positive linear operators of the form

$$I_n(f;x) = \sum_{k=0}^{n} h_{n,k}(x) A_{n,k}(f),\ f \in C[a,b],\ x \in [a,b],$$

where $h_{n,k} \in C[a,b],\ h_{n,k} \geq 0$ and

$$A_{n,k}(f) = \int_a^b f(t) d\mu_{n,k}(t)$$

with $\mu_{n,k}$ probability Borel measures on $[a,b], n \geq 1, k = 0,1,\ldots,n$.

Let $x_{n,k} \in [a,b]$ be the barycenter of $\mu_{n,k}$, i.e.,

$$x_{n,k} = \int_a^b t d\mu_{n,k}(t).$$

We associate with the sequence (I_n) the sequence of operators

$$D_n(f;x) = \sum_{k=0}^{n} h_{n,k}(x) f(x_{n,k}).$$

Generally speaking, the operators D_n are simpler than I_n. We investigate the properties of D_n in relation with those of I_n.

2. Some examples

For $n \geq 1$ and $k = 0, 1, \dots, n$ let

$$p_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k},\ x \in [0,1].$$

Example 2.1. Let $U_n : C[0,1] \longrightarrow C[0,1]$ be the genuine Bernstein-Durrmeyer operators (see [3] and the references therein) defined by

$$U_n(f;x) := f(0)p_{n,0}(x) + f(1)p_{n,n}(x) +$$
$$+(n-1)\sum_{k=1}^{n-1} p_{n,k}(x) \int_0^1 p_{n-2,k-1}(t)f(t)dt.$$

It is easy to see that the associated operators are the classical Bernstein operators

$$B_n(f;x) = \sum_{k=0}^{n} p_{n,k}(x) f\Big(\frac{k}{n}\Big).$$

Example 2.2. Consider the sequences of real numbers a_n and b_n such that $0 \leq a_n < b_n \leq 1, n \geq 1$. In [1] the authors introduced and investigated the operators

$$C_n(f;x) = \sum_{k=0}^{n} p_{n,k}(x) \Big(\frac{n+1}{b_n - a_n} \int_{\frac{k+a_n}{n+1}}^{\frac{k+b_n}{n+1}} f(t)dt \Big),$$

where $f \in C[0,1]$ and $x \in [0,1]$.

The associated operators are the Stancu type operators (see [15])

$$S_n(f;x) = \sum_{k=0}^{n} p_{n,k}(x) f\Big(\frac{2k + a_n + b_n}{2(n+1)} \Big).$$

In particular, for $a_n = 0$ and $b_n = 1, (C_n)$ becomes the sequence of classical Kantorovich operators.

Example 2.3. Let $a, b > -1$ and $\alpha \geq 0$. Consider the positive linear functionals $T_{n,k} : C[0,1] \longrightarrow \mathbb{R}$,

$$T_{n,k}(f) := \frac{\int_0^1 f(t) t^{ck+a} (1-t)^{c(n-k)+b} dt}{B(ck + a + 1, c(n-k) + b + 1)},$$

where $c := c_n := [n^\alpha]$ and B is the Beta function.

For $f \in C[0,1]$ and $x \in [0,1]$ let

$$P_n(f;x) := \sum_{k=0}^{n} p_{n,k}(x) T_{n,k}(f), ; n \geq 1.$$

The sequence of positive linear operators (P_n) was introduced by D. Mache (see [5], [6]); it represents a link between the Durrmeyer operators with Jacobi weights (obtained for $\alpha = 0$) and the Bernstein operators (obtained as a limiting case when $\alpha \longrightarrow \infty$). Concerning the properties of the operators P_n and their relationship with Durrmeyer, Bernstein, and other operators, see [5], [6], [8], [9], [10], [11]. The semigroup of operators, represented in terms of iterates of P_n, is investigated in [2], [9], [10], [11], [12].

Let $e_i(x) = x^i$, $x \in [0,1]$, $i = 0,1,\ldots$. Then $T_{n,k}(e_0) = 1$ and the barycenter of the probability Radon measure $T_{n,k}$ is

$$T_{n,k}(e_1) = \frac{ck+a+1}{cn+a+b+2}.$$

As in Section 1, we associate with the sequence (P_n) the simpler sequence of positive linear operators (V_n) defined, for $f \in C[0,1]$ and $x \in [0,1]$, by

$$V_n(f;x) := \sum_{k=0}^{n} p_{n,k}(x) f\Big(\frac{ck+a+1}{cn+a+b+2}\Big).$$

When $a = b = -1$, or when $\alpha \longrightarrow \infty$, we get the classical Bernstein operators; when $\alpha = 0$, the operators V_n reduce to the operators considered by D.D. Stancu in [15].

In the next sections we investigate the properties of the operators (V_n) in connection with the properties of (P_n); see also [7].

3. Approximation properties

By direct computation we get

$$V_n e_0 = e_0,$$

$$V_n e_1 = \frac{cne_1 + (a+1)e_0}{cn+a+b+2},$$

$$V_n e_2 = \frac{c^2 n(n-1)e_2 + cn(c+2a+2)e_1 + (a+1)^2 e_0}{(cn+a+b+2)^2}.$$

Let us remark that

$$\lim_{n\to\infty} V_n e_i = e_i,\ i = 0,1,2,$$

uniformly on $[0,1]$.

From the classical Korovkin Theorem we infer:

Proposition 3.1. *For all* $f \in C[0,1]$,

$$\lim_{n\to\infty} V_n f = f,\ \textit{uniformly on}\ [0,1].$$

In the sequel we shall use the inequality

$$|L(f) - f(b)| \le (L(e_2) - b^2)\frac{\|f''\|}{2}, \quad f \in C^2[0,1],$$

where L is a probability Radon measure on $[0,1]$, $b = L(e_1)$ is the barycenter of L, and $\|\cdot\|$ is the uniform norm. To prove this inequality, it suffices to apply the *barycenter inequality*

$$L(h) \ge h(b),\ h \in C[0,1] \text{ convex},$$

to the convex functions $\dfrac{\|f''\|}{2} e_2 \pm f$.

Theorem 3.2. *For* $n \geq 1,\ x \in [0,1],$ *and* $f \in C^2[0,1]$ *we have*

$$|P_n(f;x) - V_n(f;x)| \leq$$
$$\frac{c^2 n(n-1)x(1-x) + cn(b-a)x + cn(a+1) + (a+1)(b+1)}{2(cn+a+b+2)^2(cn+a+b+3)} \|f''\|.$$

Proof. Since the barycenter of $T_{n,k}$ is

$$\frac{ck+a+1}{cn+a+b+2},$$

we have

$$|T_{n,k}(f) - f\Big(\frac{ck+a+1}{cn+a+b+2}\Big)| \leq \Big(T_{n,k}(e_2) - \Big(\frac{ck+a+1}{cn+a+b+2}\Big)^2\Big)\frac{\|f''\|}{2}$$
$$= \frac{(ck+a+1)(c(n-k)+b+1)}{(cn+a+b+2)^2(cn+a+b+3)}\frac{\|f''\|}{2}.$$

Consequently,

$$|P_n(f;x) - V_n(f;x)| \leq \frac{\|f''\|}{2}\sum_{k=0}^{n} p_{n,k}(x)\frac{(ck+a+1)(c(n-k)+b+1)}{(cn+a+b+2)^2(cn+a+b+3)}$$
$$= \frac{\|f''\|}{2}\frac{c^2 n(n-1)x(1-x) + cn(b-a)x + cn(a+1) + (a+1)(b+1)}{(cn+a+b+2)^2(cn+a+b+3)}.$$

□

Let us remark that for $\alpha = a = b = 0$ the operators P_n reduce to the classical Durrmeyer operators M_n. Consequently, the previous theorem yields

Corollary 3.3. *For* $n \geq 1, x \in [0,1]$ *and* $f \in C^2[0,1]$ *we have*

$$|M_n(f;x) - \sum_{k=0}^{n} p_{n,k}(x) f\Big(\frac{k+1}{n+2}\Big)| \leq$$
$$\leq \frac{n(n-1)x(1-x)+n+1}{2(n+2)^2(n+3)}||f''||.$$

4. Asymptotic formulae

The moments of the operator V_n are defined by

$$M_{n,m}(x) := V_n((e_1 - xe_0)^m; x) = \sum_{k=0}^{n}\Big(\frac{ck+a+1}{cn+a+b+2} - x\Big)^m p_{n,k}(x).$$

Let us remark that

$$M'_{n,m}(x) = \sum_{k=0}^{n}\Big(\frac{ck+a+1}{cn+a+b+2} - x\Big)^m p'_{n,k}(x) - mM_{n,m-1}(x).$$

Since

$$x(1-x)p'_{n,k}(x) = (k-nx)p_{n,k}(x),$$

we get

$$
\begin{aligned}
x(1-x)M'_{n,m}(x) &= \sum_{k=0}^{n}\Big(\frac{ck+a+1}{cn+a+b+2}-x\Big)^{m}(k-nx)p_{n,k}(x)\\
&- mx(1-x)M_{n,m-1}(x) =\\
&= \frac{cn+a+b+2}{c}\sum_{k=0}^{n}\Big(\frac{ck+a+1}{cn+a+b+2}-x\Big)^{m+1}p_{n,k}(x)\\
&- \frac{a+1-(a+b+2)x}{c}\sum_{k=0}^{n}\Big(\frac{ck+a+1}{cn+a+b+2}-x\Big)^{m}p_{n,k}(x)\\
&- mx(1-x)M_{n,m-1}(x).
\end{aligned}
$$

Consequently, the following recurrence formula for the moments of V_n is valid:

Theorem 4.1. *For all $n \geq 1$ and $m \geq 1$,*

$$
\begin{aligned}
&(cn+a+b+2)M_{n,m+1}(x) = cx(1-x)M'_{n,m}(x)+\\
&+(a+1-(a+b+2)x)M_{n,m}(x) + cmx(1-x)M_{n,m-1}(x).
\end{aligned}
$$

It is easy to verify that

$$M_{n,0}(x) = 1,\ M_{n,1}(x) = \frac{a+1-(a+b+2)x}{cn+a+b+2}.$$

By using the recurrence formula we get

$$M_{n,2}(x) = \frac{c^2nx(1-x)+(a+1-(a+b+2)x)^2}{(cn+a+b+2)^2}.$$

The same recurrence formula can be used in order to verify that

$$M_{n,m}(x) = O(n^{-[\frac{m+1}{2}]}),\ m \geq 0,$$

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

Now the assumptions of Sikkema's theorem [14] are fulfilled; consequently, we have the following Voronovskaja type formula:

Theorem 4.2.

$$\lim_{n\to\infty} n(V_n(f;x)-f(x)) = \begin{cases} \frac{x(1-x)}{2}f''(x) + (a+1-(a+b+2)x)f'(x),\ \alpha = 0\\ \frac{x(1-x)}{2}f''(x),\ \alpha > 0,\end{cases}$$

for all $f \in C[0,1]$ such that $f''(x)$ exists and is finite.Moreover, if $f \in C^2[0,1]$, the convergence is uniform on $[0,1]$.

Concerning the (similar) Voronovskaja formula for the operators P_n, see [10] and the references given there.

5. Iterates of V_n

Let r be a non-negative integer, $r \le n$. It is well-known (see, e.g., [4] and the references given there) that

$$B_n e_r = \frac{n(n-1)\dots(n-r+1)}{n^r} e_r \ + \textit{terms of lower degree},$$

where B_n are the classical Bernstein operators.

Let

$$\varphi_r := \Big(\frac{cne_1 + (a+1)e_0}{cn+a+b+2}\Big)^r.$$

Then, for $k = 0, 1, \dots, n$,

$$\varphi_r(\frac{k}{n}) = \Big(\frac{ck+a+1}{cn+a+b+2}\Big)^r,$$

so that

$$\begin{aligned} V_n e_r = B_n \varphi_r &= \Big(\frac{cn}{cn+a+b+2}\Big)^r B_n e_r + \text{terms of lower degree} \\ &= \frac{n(n-1)\dots(n-r+1)}{(cn+a+b+2)^r} c^r e_r + \text{terms of lower degree.} \end{aligned}$$

It follows that:

Theorem 5.1. *The numbers*

$$\lambda_r := \frac{n(n-1)\dots(n-r+1)}{(cn+a+b+2)^r} c^r, \quad r = 0, 1, \dots, n,$$

are eigenvalues of V_n, and the eigenfunction corresponding to λ_r can be chosen as a monic polynomial of degree r.

Now let us describe V_n as

$$V_n(f;x) = \sum_{k=0}^{n} p_{n,k}(x) f\Big(\frac{k + \frac{a+1}{c}}{n + \frac{a+b+2}{c}}\Big)$$

Under this form we see that V_n coincides with the operator $S_n^{<0,\beta,\gamma>}$ defined in [4;(1)], if

$$\beta := \frac{a+1}{c}\ ,\ \gamma := \frac{a+b+2}{c}.$$

Now the above Theorem 5.1 can be considered also as a consequence of Theorem 1 in [4].

The over-iterates of V_n can be studied by using the results of [4] or [13]. Indeed, let

$$a_j := \frac{j+\beta}{n+\gamma} = \frac{cj+a+1}{cn+a+b+2}\ ,\ j = 0, 1, \dots, n.$$

From [4;(9), (11), (12)] or from [13; Th. 5.3] we deduce for $f \in C[0,1]$:

$$\lim_{m\to\infty} V_n^m f = e_0 \sum_{j=0}^{n} d_j f\Big(\frac{cj+a+1}{cn+a+b+2}\Big),$$

uniformly on $[0,1]$, where $(d_0, d_1, \dots, d_n)$ is the unique solution of the system

$$\begin{pmatrix} p_{n,0}(a_0) & \dots & p_{n,0}(a_n) \\ \dots & \dots & \dots \\ p_{n,n}(a_0) & \dots & p_{n,n}(a_n) \end{pmatrix} \begin{pmatrix} d_0 \\ \dots \\ d_n \end{pmatrix} = \begin{pmatrix} d_0 \\ \dots \\ d_n \end{pmatrix}$$

satisfying $d_0 \geq 0, \dots, d_n \geq 0, d_0 + \dots + d_n = 1$.

6. Shape preserving properties

For each $m \geq 0$ consider the function

$$\varphi_m(t) = \Big(\frac{cnt + a + 1}{cn + a + b + 2}\Big)^m, \ t \in [0,1].$$

Let B_n be the classical Bernstein operators on $C[0,1]$. Then we have

$$V_n e_m = B_n \varphi_m, \ n \geq 1.$$

Consequently, the technique used in [16, Section 25.2] can be applied; as in [16, Cor.25.2] we get

Theorem 6.1. *If $0 \leq m \leq n$ and $f \in C[0,1]$ is convex of order m, then $V_n f$ is convex of order m.*

For convex functions of order 1, i.e., usual convex functions, we have also

Theorem 6.2. *If $f \in C[0,1]$ is convex, then*

$$P_n(f;x) \geq V_n(f;x) \geq f\Big(\frac{cnx + a + 1}{cn + a + b + 2}\Big), \ x \in [0,1].$$

Proof. Let $f \in C[0,1]$ be convex, and $x \in [0,1]$. From the barycenter inequality we know that

$$T_{n,k}(f) \geq f\Big(\frac{ck + a + 1}{cn + a + b + 2}\Big), \ k = 0, 1, \dots, n,$$

which immediately yields

$$P_n(f;x) \geq V_n(f;x).$$

On the other hand, consider the probability Radon measure

$$g \longrightarrow V_n(g;x), \ g \in C[0,1].$$

The corresponding barycenter is

$$V_n(e_1;x) = \frac{cnx + a + 1}{cn + a + b + 2}.$$

Again by the barycenter inequality we get

$$V_n(f;x) \geq f\Big(\frac{cnx + a + 1}{cn + a + b + 2}\Big). \qquad \square$$

Acknowledgment. The author is grateful to the referee for many helpful suggestions.

References

[1] Altomare, F., Leonessa, V., *On a sequence of positive linear operators associated with a continuous selection of Borel measures*, Mediterr. J. Math., **3**(2006), 363-382.

[2] Altomare, F., Raşa, I., *On some classes of difussion equations and related approximation processes*, In: Trends and Applications in Constructive Approximation, (Eds.) M.G. de Bruin, D.H. Mache and J. Szabados, ISNM vol. 151, 13-26, Birkhäuser Verlag, Basel, 2005.

[3] Gonska, H., Kacsó, D., Raşa, I., *On genuine Bernstein-Durrmeyer operators*, Result. Math., **50**(2007), 213-225.

[4] Gonska, H., Pitul, P., Raşa, I., *Over-iterates of Bernstein-Stancu operators,* Calcolo, **44**(2007), 117-125.

[5] Mache, D.H., *Gewichtete Simultanapproximation in der L_p-Metrik durch das Verfahren der Kantorovič Operatoren,* Dissertation, Univ. Dortmund, 1991.

[6] Mache, D.H., *A link between Bernstein polynomials and Durrmeyer polynomials with Jacobi weights*, In: Approx. Theory VIII, Vol. 1: Approximation and Interpolation, Ch.K. Chui and L.L. Schmaker (Eds.), 403-410, World Scientific Publ., 1995.

[7] Mache, D.H., Raşa, I., *Relations between polynomial operators*, Preprint.

[8] Mache, D.H., Zhou, D.X., *Characterization theorems for the approximation by a family of operators*, J. Approx. Theory, **84**(1996), 145-161.

[9] Mache, D.H., Raşa, I., *Some C_0-semigroups related to polynomial operators*, Rend. Circ. Mat. Palermo Suppl. II, **76**(2005), 459-467.

[10] Raşa, I., *Semigroups associated to Mache operators*, In: Advanced Problems in Constructive Approximation, (Eds.) M.D. Buhmann and D.H. Mache, ISNM vol.142, 143-152, Birkhäuser Verlag, Basel, 2002.

[11] Rasa, I., *Semigroups associated to Mache operators (II)*, In: Trends and Applications in Constructive Approximation, (Eds.) M.G. de Bruin, D.H. Mache and J. Szabados, ISNM vol. 151, 225-228, Birkhäuser Verlag, Basel, 2005.

[12] Raşa, I., *Positive operators, Feller semigroups and difussion equations associated with Altomare projections,* Conf. Sem. Mat. Univ. Bari, **284**(2002), 1-26.

[13] Rus, I.A., *Iterates of Stancu operators (via fixed point principles) revisited*, Fixed Point Theory, **11**(2010), 369-374.

[14] Sikkema, P.C., *On some linear positive operators*, Indag. Math., **32**(1970), 327-337.

[15] Stancu, D.D., *Asupra unei generalizari a polinoamelor lui Bernstein,* Stud. Univ. Babeş-Bolyai Math., **14**(1969), 31-45.

[16] Vladislav, T., Raşa, I., *Analiză Numerică. Aproximare, problema lui Cauchy abstractă, proiectori Altomare,* Editura Tehnică, Bucureşti, 1999.

Ioan Raşa
Technical University of Cluj-Napoca
Faculty of Automation and Computer Science
15, Daicoviciu Street
400020 Cluj-Napoca
Romania
e-mail: `ioan.rasa@math.utcluj.ro`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 545–549

Area preserving maps from rectangles to elliptic domains

Daniela Roşca

Abstract. We construct a bijection from $\mathbb{R}^2$ to $\mathbb{R}^2$, which maps, for each $\alpha \in (0, \infty)$, rectangles of arbitrary edges $2\alpha L_1, 2\alpha L_2$ onto ellipses with semi-axes $\alpha a, \alpha b$, with a, b satisfying $4L_1L_2 = \pi ab$. This bijection preserves area and thus allows us to construct uniform and refinable grids on elliptic domains starting from uniform and refinable grids on rectangles.

Mathematics Subject Classification (2010): 65M50, 65N50.

Keywords: Uniform grid, refinable grid, hierarchical grid, equal area projection, ellipse.

1. Introduction

Uniform and refinable grids (UR) are useful in many applications, like construction of multiresolution analysis and wavelets, or for solving numerically partial differential equations. While on a rectangle or on other polygonal domains the construction of UR grids is trivial, it is not immediate on an elliptic domain or on a disc.

In this paper we construct an area preserving bijection from $\mathbb{R}^2$ to $\mathbb{R}^2$, which maps rectangles of arbitrary edges $2\alpha L_1, 2\alpha L_2$ onto ellipses with semi-axes $\alpha a, \alpha b$, with a, b satisfying $4L_1L_2 = \pi ab$. This allows us to transport a rectangular grid to an elliptic grid, preserving the area of the cells. In particular, any uniform[1] rectangular grid is mapped into a uniform elliptic grid. A refinement process is needed when a grid is not fine enough to solve a problem accurately. A uniform refinement consists in dividing a cell into a given number of smaller cells with the same area. With the procedure described here, any uniform refinement of a rectangular grid leads to a uniform refinement of the corresponding elliptic grid.

[1] A grid is uniform if its cells have the same area.

In the particular case of the disc, such a bijection was constructed in a previous paper [1] and helped us to construct uniform grids on the sphere.

The particular case when $2L_1 = \sqrt{\pi}a$ and $2L_2 = \sqrt{\pi}b$ (the semi-axes a, b of the ellipse are proportional to the edges $2L_1, 2L_2$ of the rectangle) was considered in [2]. Here we consider the general case when the edges of the rectangle are arbitrary and the semi-axes satisfy the condition $4L_1L_2 = \pi ab$, implied by the fact that the rectangle and the ellipse have the same area. Also, we use another method of construction than the one in [1, 2].

2. Construction of an area preserving bijection in $\mathbb{R}^2$

Consider the ellipse $\mathcal{E}_{a,b}$ of semi-axes a and b, $a, b > 0$, of equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

and the rectangle $\mathcal{R}_{L_1,L_2}$ with edges of lengths $2L_1$ and $2L_2$, defined as

$$\mathcal{R}_{L_1,L_2} = \{(x, y) \in \mathbb{R}^2,\ |x| = L_1, |y| = L_2\}.$$

The domains enclosed by $\mathcal{E}_{a,b}$ and $\mathcal{R}_{L_1,L_2}$ will be denoted by $\overline{\mathcal{E}}_{a,b}$ and $\overline{\mathcal{R}}_{L_1,L_2}$, respectively. We will construct a bijection $T^{a,b}_{L_1,L_2} : \mathbb{R}^2 \to \mathbb{R}^2$ which maps each rectangle $\mathcal{R}_{\alpha L_1,\alpha L_2}$ onto the ellipse $\mathcal{E}_{\alpha a,\alpha b}$ and has the area preserving property

$$\mathcal{A}(D) = \mathcal{A}(T^{a,b}_{L_1,L_2}(D)),\ \text{for every domain } D \subseteq \mathbb{R}^2. \tag{2.1}$$

Here $\mathcal{A}(D)$ denotes the area of D. Thus, $\mathcal{A}(\overline{\mathcal{R}}_{L_1,L_2}) = \mathcal{A}(\overline{\mathcal{E}}_{a,b})$ implies

$$\pi ab = 4L_1L_2.$$

We focus for the moment on the first octant I of the plane,

$$I = I_{L_1,L_2} = \{(x, y) \in \mathbb{R}^2,\ 0 \le L_1 y \le L_2 x\}.$$

The map $T^{a,b}_{L_1,L_2}$ will be defined in such a way that each half-line $d_m \subset I$ of equation $y = mx$ ($0 \le m \le \frac{L_2}{L_1}$) is mapped onto the half-line $d_{\varphi(m)}$ of equation $Y = \varphi(m)X$, such that

$$0 \le \varphi(m) \le \frac{b}{a},\quad \text{for } 0 \le m \le \frac{L_2}{L_1}.$$

Let $Q = Q(L_1, mL_1)$ and let $Q' = Q'(L_1, 0)$ be its projection on Ox. The area of the triangle OQQ' is

$$\mathcal{A}_\triangle = \frac{mL_1^2}{2} = \frac{yL_1^2}{2x}.$$

We denote by $(X, \varphi(m)X)$ the coordinates of the point $P = T^{a,b}_{L_1,L_2}(Q) \in \mathcal{E}_{a,b}$. The area of the portion of the elliptic domain $\overline{\mathcal{E}}_{a,b}$ located between the axis OX and the line $Y = \varphi(m)X$ will be

$$\mathcal{A}_e = \frac{ab\theta}{2},$$

where

$$\theta = \arctan \frac{a\varphi(m)}{b}$$

is the angle between the axis OX and OP. Next, we impose the area preserving property $\mathcal{A}_\Delta = \mathcal{A}_e$, which yields

$$\theta = \frac{\pi L_1 y}{4L_2 x},$$

and therefore

$$\varphi(m) = \frac{b}{a} \tan \frac{\pi L_1 y}{4L_2 x}.$$

It is easy to see that φ has the following properties:

$$\varphi(0) = 0, \quad \varphi\left(\frac{L_2}{L_1}\right) = \frac{b}{a}, \quad \text{and}$$

$$0 \le \varphi(m) \le \frac{b}{a}, \quad \text{for } 0 \le m \le \frac{L_2}{L_1}.$$

Consider now $M = M(x, mx)$ and $N = T^{a,b}_{L_1,L_2}(M) = (X, \varphi(m)X)$, which belongs to an ellipse $\mathcal{E}_{\alpha a, \alpha b}$ for a certain α. The portion of the elliptic domain $\overline{\mathcal{E}}_{\alpha a, \alpha b}$, located between ON and OX, has the area

$$\mathcal{A}_{e,\alpha} = \frac{ab\theta\alpha^2}{2},$$

whereas the area of the triangle OMM', with $M' = M'(x, 0)$ is $mx^2/2$.

Again, the area preserving property implies this time

$$\alpha = x\sqrt{\frac{m}{ab\theta}} = 2x\sqrt{\frac{L_2}{L_1} \cdot \frac{1}{\pi ab}}.$$

Finally, from $N \in \mathcal{E}_{\alpha a, \alpha b}$ we obtain

$$\frac{X^2}{a^2} + \frac{X^2 \varphi^2(m)}{b^2} = \alpha^2,$$

and therefore

$$\begin{aligned} X &= \frac{ab\alpha}{\sqrt{b^2 + a^2\varphi^2(m)}} = \frac{a\alpha}{\sqrt{1 + \tan^2\theta}} = a\alpha\cos\theta = 2x\sqrt{\frac{aL_2}{bL_1\pi}} \cos \frac{\pi L_1 y}{4L_2 x}, \\ Y &= \varphi(m)X = 2x\sqrt{\frac{bL_2}{aL_1\pi}} \sin \frac{\pi L_1 y}{4L_2 x}. \end{aligned}$$

A simple calculation shows that the Jacobian of $T^{a,b}_{L_1,L_2}$ is 1 and therefore relation (2.1) is fulfilled for domains $D \subseteq I$.

By similar arguments for the other seven octants, we find that the function $T^{a,b}_{L_1,L_2} : \mathbb{R}^2 \to \mathbb{R}^2$ which maps rectangles onto ellipses and preserves areas is defined as follows:

- For $L_1|y| \le L_2|x|$,

$$(x, y) \longmapsto (X, Y) = \left(2x\sqrt{\frac{aL_2}{bL_1\pi}} \cos \frac{\pi L_1 y}{4L_2 x},\ 2x\sqrt{\frac{bL_2}{aL_1\pi}} \sin \frac{\pi L_1 y}{4L_2 x}\right);$$

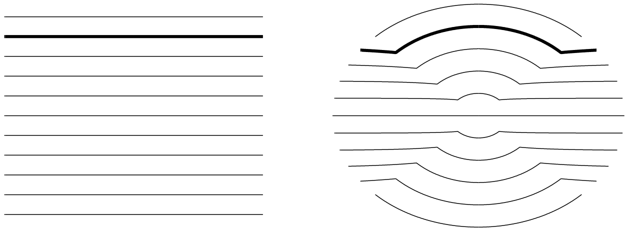


FIGURE 1. A horizontal grid and its image grid on the elliptic domain. The image of the bold line on the left is the bold curve on the right.

- For $L_2|x| \le L_1|y|$,

$$(x,y) \longmapsto (X,Y) = \left(2y\sqrt{\frac{aL_1}{bL_2\pi}} \sin \frac{\pi L_2 x}{4L_1 y},\ 2y\sqrt{\frac{bL_1}{aL_2\pi}} \cos \frac{\pi L_2 x}{4L_1 y} \right).$$

For the origin we take $T^{a,b}_{L_1,L_2}(0,0) = (0,0)$. We can prove that $T^{a,b}_{L_1,L_2}$ is continuous and bijective and its inverse is given by the following formulas:

- For $a|Y| \le b|X|$,

$$(X,Y) \longmapsto (x,y) = \mathrm{sign}(X)\sqrt{X^2 + \frac{a^2}{b^2}Y^2} \left(\frac{\sqrt{\pi}}{2},\ \frac{2b}{a\sqrt{\pi}} \arctan \frac{aY}{bX} \right);$$

- For $b|X| \le a|Y|$,

$$(X,Y) \longmapsto (x,y) = \mathrm{sign}(Y)\sqrt{\frac{b^2}{a^2}X^2 + Y^2} \left(\frac{2a}{b\sqrt{\pi}} \arctan \frac{bX}{aY},\ \frac{\sqrt{\pi}}{2} \right).$$

3. Uniform and refinable grids

The area preserving maps constructed in the previous section can be used for the construction of UR grids on elliptic domains, by mapping any UR rectangular grid.

Figure 1 shows the image of horizontal lines by an application $T^{a,b}_{L_1,L_2}$. In Figures 2 and 3 we show two grids on an elliptic domain and its refinement, both images of a rectangular grid.

Of course, other 2D uniform grids on a rectangle can be constructed, including triangular grids with different types of refinements.

Acknowledgement. The work has been co-funded by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Romanian Ministry of Labor, Family and Social Protection through the Financial Agreement POSDRU/89/1.5/S/62557.

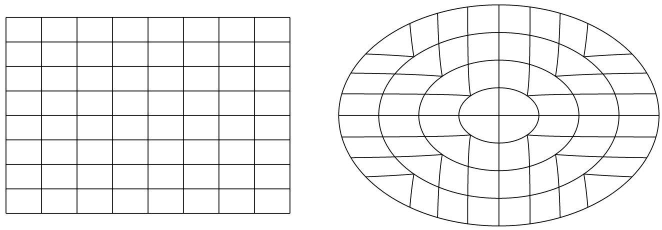


FIGURE 2. A uniform grid on a rectangle and its image - a uniform grid on the elliptic domain.

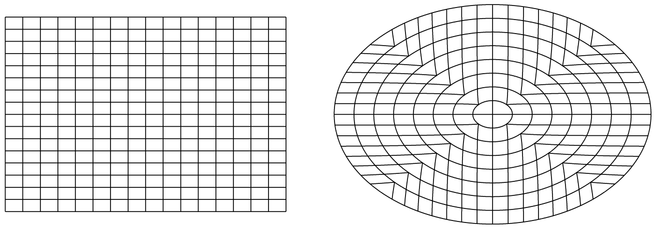


FIGURE 3. A refinement of the grid in Figure 2 and its image - a refinement of the elliptic grid in Figure 2.

References

[1] Roşca, D., *New uniform grids on the sphere*, Astronomy and Astrophysics, vol. 520, A63, 2010.

[2] Roşca, D., *Uniform and refinable grids on elliptic domains and on some surfaces of revolution*, Appl. Math. Comput, accepted doi:10.1016/j.amc.2011.02.095

Daniela Roşca
Technical University of Cluj-Napoca
Department of Mathematics
28, Memorandumului Street
400114 Cluj-Napoca
Romania
e-mail: `Daniela.Rosca@math.utcluj.ro`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 551–565

On Grüss-type inequalities for positive linear operators

Maria-Daniela Rusu

Abstract. The classical form of Grüss' inequality gives an estimate of the difference between the integral of the product and the product of the integrals of two functions in $C[a, b]$. It was first published by G. Grüss in [7]. The aim of this article is to discuss Grüss-type inequalities in $C(X)$, the set of continuous functions defined on a compact metric space X. We consider a functional $L(f) := H(f; x)$, where $H : C(X) \to C(X)$ is a positive linear operator and $x \in X$ is fixed. Generalizing a result of Acu et al. [1], a quantitative Grüss-type inequality is obtained in terms of the least concave majorant of the classical modulus of continuity. The interest is in the degree of non-multiplicativity of the functional L. Moreover, for the case $X = [a, b]$ we improve the inequality and apply it to various known operators, in particular those of Bernstein-, convolution- and Shepard-type.

Mathematics Subject Classification (2010): 47A63, 41A25, 47B38.

Keywords: Grüss-type inequality, compact metric space, least concave majorant of the modulus of continuity, convolution-type operator, Shepard interpolation operator.

1. Introduction

The classical form of Grüss' inequality gives an estimate of the difference between the integral of the product and the product of the integrals of two functions in $C[a, b]$. It was first published by G. Grüss in [7]. The aim of this article is to discuss Grüss-type inequalities in $C(X)$, the set of continuous functions defined on a compact metric space X. We consider a functional $L(f) := H(f; x)$, where $H : C(X) \to C(X)$ is a positive linear operator and $x \in X$ is fixed. Generalizing a result of Acu et al. [1], a quantitative Grüss-type inequality is obtained in terms of the least concave majorant of the classical modulus of continuity. The interest is in the degree of non-multiplicativity of the functional L. Moreover, for the case $X = [a, b]$ we improve the inequality and apply it to various known operators, in particular those of Bernstein-, convolution- and Shepard-type.

2. Auxiliary results

Before giving our main results, we need some introductory notions that will be used in the sequel. Let $C(X) = C_{\mathbb{R}}((X,d))$ represent the Banach lattice of real-valued continuous functions defined on the compact metric space (X,d). Then we have the following definition:

Definition 2.1. *Let $f \in C(X)$. If, for $t \in [0,\infty)$, the quantity*

$$\omega_d(f;t) := \sup\{|f(x) - f(y)|,\ d(x,y) \le t\}$$

is the usual modulus of continuity, then its least concave majorant is given by

$$\widetilde{\omega_d}(f,t) = \begin{cases} \sup\limits_{0\le x\le t\le y\le d(X), x\ne y} \frac{(t-x)\omega_d(f,y)+(y-t)\omega_d(f,x)}{y-x} & \text{for } 0 \le t \le d(X)\ , \\ \omega_d(f,d(X)) & \text{if } t > d(X)\ , \end{cases}$$

and $d(X) < \infty$ is the diameter of the compact space X.

For $0 < r \le 1$, let Lip_r be the set of all functions $g \in C(X)$ with the property that

$$|g|_{Lip_r} := \sup_{d(x,y)>0} |g(x) - g(y)|\,/d^r(x,y) < \infty.$$

Lip_r is a dense subspace of $C(X)$ equipped with the supremum norm $\|\cdot\|_\infty$, and $|\cdot|_{Lip_r}$ is a seminorm on Lip_r.

We also need to define the $K-$functional with respect to $(Lip_r, |\cdot|_{Lip_r})$, which is given by

$$K\left(t, f; C(X), Lip_r\right) := \inf_{g\in Lip_r} \left\{\|f-g\|_\infty + t\cdot |g|_{Lip_r}\right\},$$

for $f \in C(X)$ and $t \ge 0$.

Another tool for some proofs that follow is a lemma of Brudnyĭ (see [10]) that gives the relationship between the K-functional and the least concave majorant of the modulus of continuity.

Lemma 2.2. *Every continuous function f on X satisfies*

$$K\left(\frac{t}{2}, f; C(X), Lip_1\right) = \frac{1}{2}\cdot\widetilde{\omega_d}(f,t),\ 0 \le t \le d(X).$$

In the case $X = [a,b]$, we also have

$$\begin{aligned} K\left(\frac{t}{2}, f; C[a,b], C^1[a,b]\right) &:= \inf_{g\in C^1[a,b]} \left\{\|f-g\|_\infty + \frac{t}{2}\cdot\|g'\|_\infty\right\} \\ &= \frac{1}{2}\cdot\widetilde{\omega}(f;t),\ t \ge 0. \end{aligned}$$

3. Grüss-type inequalities in a compact metric space

What we do here is generalize Theorem 4 in [1] in the case of a compact metric space.

We consider (X, d) a compact metric space, $x \in X$ fixed, with diameter $d(X) > 0$. Now let $H : C(X) \to C(X)$ be a positive linear operator reproducing constant functions. We define the positive linear functional $H(\cdot; x)$ and consider the positive bilinear functional

$$D(f, g) := H(f \cdot g; x) - H(f; x) \cdot H(g; x).$$

It was remarked after Theorem 4 in [1] that the assertion given there can be generalized by replacing $([a, b], |\cdot|)$ by a compact metric space (X, d), the second moment $H((e_1 - x)^2; x)$ by $H(d^2(\cdot, x); x)$, and the K-functional $K(\cdot, f; C[a, b], C^1[a, b])$ by $K(\cdot, f; C(X), Lip_1)$.

We then obtain the following result:

Theorem 3.1. *If $f, g \in C(X)$, (X, d) a compact metric space and $x \in X$ fixed, then the inequality*

$$|D(f, g)| \leq \frac{1}{4}\widetilde{\omega_d}\left(f; 4\sqrt{H(d^2(\cdot, x); x)}\right) \cdot \widetilde{\omega_d}\left(g; 4\sqrt{H(d^2(\cdot, x); x)}\right) \tag{3.1}$$

holds.

Proof. Let $f, g \in C[a, b]$ and $r, s \in Lip_1$. We use the Cauchy-Schwarz inequality for positive linear functionals:

$$|H(f; x)| \leq H(|f|; x) \leq \sqrt{H(f^2; x) \cdot H(1; x)} = \sqrt{H(f^2; x)},$$

so we have

$$D(f, f) = H(f^2; x) - H(f; x)^2 \geq 0.$$

Hence D is a positive bilinear form on $C(X)$. Using Cauchy-Schwarz for D gives us

$$|D(f, g)| \leq \sqrt{D(f, f)D(g, g)} \leq \|f\|_\infty \cdot \|g\|_\infty .$$

Because $H : C(X) \to C(X)$ is a positive linear operator reproducing constant functions, $H(f; x)$, with fixed $x \in X$, is a positive linear functional that we can represent as follows

$$H(f; x) := \int_X f(t)d\mu_x(t),$$

where μ_x is a Borel probability measure on X, i.e., $\int_X d\mu_x(t) = 1$. For r as above, we have

$$\begin{aligned} D(r, r) = H(r^2; x) - H(r; x)^2 &= \int_X r^2(t)d\mu_x(t) - \left(\int_X r(u)d\mu_x(u)\right)^2 \\ &= \int_X \left(r(t) - \int_X r(u)d\mu_x(u)\right)^2 d\mu_x(t) \\ &= \int_X \left(\int_X (r(t) - r(u))\, d\mu_x(u)\right)^2 d\mu_x(t) \end{aligned}$$

$$\leq \int_X \left(\int_X (r(t) - r(u))^2 \, d\mu_x(u) \right) d\mu_x(t)$$

$$\leq |r|^2_{Lip_1} \int_X \left(\int_X d^2(t,u) d\mu_x(u) \right) d\mu_x(t)$$

$$\leq |r|^2_{Lip_1} \int_X \left(\int_X [d(t,x) + d(x,u)]^2 \, d\mu_x(u) \right) d\mu_x(t)$$

$$= |r|^2_{Lip_1} \int_X \int_X \left\{ d^2(t,x) + 2 \cdot d(t,x) \cdot d(x,u) + d^2(x,u) \right\} d\mu_x(u) d\mu_x(t)$$

$$= |r|^2_{Lip_1} \left[\int_X d^2(t,x) d\mu_x(t) + 2 \int_X \int_X d(t,x) d(x,u) d\mu_x(u) d\mu_x(t) + \int_X d^2(x,u) d\mu_x(u) \right]$$

$$= |r|^2_{Lip_1} \left[H(d^2(\cdot,x);x) + 2 \left(\int_X d(t,x) d\mu_x(t) \right) \left(\int_X d(u,x) d\mu_x(u) \right) + H(d^2(\cdot,x);x) \right]$$

$$= |r|^2_{Lip_1} \left[H(d^2(\cdot,x);x) + 2H(d(\cdot,x);x) \cdot H(d(\cdot,x);x) + H(d^2(\cdot,x);x) \right]$$

$$= |r|^2_{Lip_1} \left[2H(d^2(\cdot,x);x) + 2H(d(\cdot,x);x)^2 \right]$$

$$\leq |r|^2_{Lip_1} \left[2H(d^2(\cdot,x);x) + 2H(d^2(\cdot,x);x) \right]$$

$$= 4 \, |r|^2_{Lip_1} \cdot H(d^2(\cdot,x);x).$$

For r, s as above, we have the estimate

$$|D(r,s)| \leq \sqrt{D(r,r)D(s,s)} \leq 4 \, |r|_{Lip_1} \cdot |s|_{Lip_1} \cdot H(d^2(\cdot,x);x).$$

Moreover, for $f \in C(X)$ and $s \in Lip_1$, the inequality

$$|D(f,s)| \leq \sqrt{D(f,f)D(s,s)} \leq 2 \, \|f\|_\infty \cdot |s|_{Lip_1} \cdot \sqrt{H(d^2(\cdot,x);x)}$$

holds. Similarly, if $r \in Lip_1$ and $g \in C(X)$, we have

$$|D(r,g)| \leq \sqrt{D(r,r)D(g,g)} \leq 2 \, \|g\|_\infty \cdot |r|_{Lip_1} \cdot \sqrt{H(d^2(\cdot,x);x)}.$$

Now let $f, g \in C(X)$ be fixed and $r, s \in Lip_1$ arbitrary. Then

$$\begin{aligned}
&|D(f,g)| \\
&= |D(f - r + r, g - s + s)| \\
&\leq |D(f - r, g - s)| + |D(f - r, s)| + |D(r, g - s)| + |D(r,s)| \\
&\leq \|f - r\|_\infty \cdot \|g - s\|_\infty + 2 \, \|f - r\|_\infty \cdot |s|_{Lip_1} \cdot \sqrt{H(d^2(\cdot,x);x)} \\
&+ 2 \, \|g - s\|_\infty \cdot |r|_{Lip_1} \cdot \sqrt{H(d^2(\cdot,x);x)} + 4 \, |r|_{Lip_1} \cdot |s|_{Lip_1} \cdot H(d^2(\cdot,x);x) \\
&= \|f - r\|_\infty \cdot \{\|g - s\|_\infty + 2 \, |s|_{Lip_1} \cdot \sqrt{H(d^2(\cdot,x);x)}\} \\
&+ 2 \, |r|_{Lip_1} \cdot \sqrt{H(d^2(\cdot,x);x)} \cdot \{\|g - s\|_\infty + 2 \, |s|_{Lip_1} \cdot \sqrt{H(d^2(\cdot,x);x)}\} \\
&= \{\|f - r\|_\infty + 2 \, |r|_{Lip_1} \sqrt{H(d^2(\cdot,x);x)}\} \\
&\cdot \{\|g - s\|_\infty + 2 \, |s|_{Lip_1} \sqrt{H(d^2(\cdot,x);x)}\}.
\end{aligned}$$

We now pass to the infimum over r and s, respectively, which leads us to

$$\begin{aligned}
&|D(f,g)| \\
&\leq K\left(\sqrt{4H(d^2(\cdot,x);x)}, f; C(X), Lip_1\right) K\left(\sqrt{4H(d^2(\cdot,x);x)}, g; C(X), Lip_1\right) \\
&= \frac{1}{2}\widetilde{\omega}\left(f; 2\cdot\sqrt{4H(d^2(\cdot,x);x)}\right)\cdot\frac{1}{2}\widetilde{\omega}\left(g; 2\cdot\sqrt{4H(d^2(\cdot,x);x)}\right) \\
&= \frac{1}{4}\widetilde{\omega}\left(f; 4\sqrt{H(d^2(\cdot,x);x)}\right)\cdot\widetilde{\omega}\left(g; 4\sqrt{H(d^2(\cdot,x);x)}\right).
\end{aligned}$$

This ends our proof. □

3.1. Shepard-type operators

The latter result from Theorem 3.1 can be applied to Shepard-type operators defined in the general setting. An example of such Shepard-type operators goes back to the work of I.K. Crain and B.K. Bhattacharyya [3] and D. Shepard [11] and was first investigated by W.J. Gordon and J.A. Wixom [6]. Other important references are e.g. the Habilitationsschrift [4] and the paper [5], both by H. Gonska.

In both of the latter references, we have the following:

Definition 3.2. *Let (X,d) be a metric space and let $x_1,\ldots,x_n$ be a finite collection of distinct points in X. We further suppose that for each n-tuple $(x_1,\ldots,x_n)$ we have a finite given sequence $(\mu_1,\ldots,\mu_n)$ of real numbers $\mu_i > 0$. Then the Crain-Bhattacharyya-Shepard (CBS) operator is given by*

$$S_n(f;x) := S_{x_1,\ldots,x_n}^{\mu_1,\ldots,\mu_n}(f,x)$$

$$:= \begin{cases} \sum_{i=1}^n f(x_i)\cdot\frac{d(x,x_i)^{-\mu_i}}{\sum_{l=1}^n d(x,x_l)^{-\mu_l}} & , x \notin \{x_1,\ldots,x_n\} \\ f(x_i) & , \textit{otherwise.} \end{cases}$$

Here $x\in X$ and f is a real-valued function defined on X.

Remark 3.3. From the above definition, we can state that S_n is a positive linear operator on $C(X)$ that satisfies $S_n(1_X,x) = 1$ for all $x\in X$. Also it holds that $S_n(f,x_i) = x_i$, for all $x_i, 1\leq i\leq n$.

We now restrict ourselves to the simpler case $1\leq\mu=\mu_1=\ldots=\mu_n$ and denote the corresponding operator by S_n^μ. Now let $H := S_n^\mu$. Then we have the following main result:

Theorem 3.4. *Let $f,g\in C(X)$ be two given functions. Then the inequality*

$$|D(f,g)| \leq \frac{1}{4}\widetilde{\omega_d}\left(f; 4\sqrt{\sum_{i=1}^n\frac{d(x,x_i)^{2-\mu}}{\sum_{l=1}^n d(x,x_l)^{-\mu}}}\right)\widetilde{\omega_d}\left(g; 4\sqrt{\sum_{i=1}^n\frac{d(x,x_i)^{2-\mu}}{\sum_{l=1}^n d(x,x_l)^{-\mu}}}\right)$$

holds, for $x\notin\{x_1,\ldots,x_n\}$. For $x = x_i$, $|D(f,g)| = 0$.

Proof. If we substitute the CBS operator S_n^μ in the result of Theorem 3.1, the following inequality

$$\begin{aligned}|D(f,g)| &= |S_n^\mu(f\cdot g;x) - S_n^\mu(f;x)\cdot S_n^\mu(g;x)|\\ &\leq \frac{1}{4}\widetilde{\omega_d}\left(f;4\sqrt{S_n^\mu(d^2(\cdot,x);x)}\right)\cdot\widetilde{\omega_d}\left(g;4\sqrt{S_n^\mu(d^2(\cdot,x);x)}\right)\end{aligned}$$

holds. The second moment of the CBS-operator can be written as

$$S_n^\mu(d^2(\cdot,x);x) = \begin{cases}\sum_{i=1}^n \frac{d(x,x_i)^{2-\mu}}{\sum_{l=1}^n d(x,x_l)^{-\mu}} & , x\notin\{x_1,\ldots,x_n\},\\ 0 & , \text{otherwise.}\end{cases} \tag{3.2}$$

Using (3.2) in the previous estimate, we get the claimed result and this ends our proof. □

Remark 3.5. We can also apply the Grüss-type inequality for the CBS operator defined on $X=[a,b]$, but we are not doing this here. What will be done in the sequel is improve the inequality from Theorem 4 in [1] and then apply it to different types of operators.

4. Grüss-type inequalities in $C[a,b]$

In a recent paper [1], Grüss-type inequalities in $C[a,b]$ were treated. The degree of non-multiplicativity of a positive linear operator $H : C[a,b]\to C[a,b]$ that reproduces constant functions was examined. For fixed $x\in[a,b]$ and two functions $f,g\in C[a,b]$, the positive linear functional $H(\cdot;x)$ was defined and the positive bilinear functional

$$D(f,g) := H(f\cdot g;x) - H(f;x)\cdot H(g;x)$$

was considered. We improve a result from the above stated article (see Theorem 4) by removing the constant $\sqrt{2}$ in the arguments of the least concave majorants. The idea of the proof was given by two of the authors of the article, namely H. Gonska and I. Raşa.

We state and prove the following:

Theorem 4.1. *If $f,g\in C[a,b]$ and $x\in[a,b]$ is fixed, then the inequality*

$$|D(f,g)| \leq \frac{1}{4}\widetilde{\omega}\left(f;2\sqrt{H((e_1-x)^2;x)}\right)\cdot\widetilde{\omega}\left(g;2\sqrt{H((e_1-x)^2;x)}\right)$$

holds, where e_1 denotes the first monomial given by $e_1(t)=t$, $t\in[a,b]$.

Proof. Let $f,g\in C[a,b]$ and $r,s\in C^1[a,b]$. Just like in the proof of Theorem 4 in [1], we use the Cauchy-Schwarz inequality for positive linear functionals:

$$|H(f;x)| \leq H(|f|;x) \leq \sqrt{H(f^2;x)\cdot H(1;x)} = \sqrt{H(f^2;x)},$$

so we have

$$D(f,f) = H(f^2;x) - H(f;x)^2 \geq 0.$$

Then we can say that D is a positive bilinear form on $C[a,b]$. Using Cauchy-Schwarz for D, we obtain

$$|D(f,g)| \leq \sqrt{D(f,f)D(g,g)} \leq \|f\|_\infty\cdot\|g\|_\infty\,.$$

As stated before, $H : C[a,b] \to C[a,b]$ is a positive linear operator that reproduces constant functions, so that $H(\cdot;x)$, with fixed $x \in [a,b]$, is a positive linear functional that can be represented as

$$H(f;x) = \int_a^b f(t)d\mu_x(t),$$

where μ_x is a probability measure on $[a,b]$, i.e., $\int_a^b d\mu_x(t) = 1$. The interest is in finding an upper bound for the following:

$$\begin{aligned} |D(f,g)| &= |D(f-r+r, g-s+s)| \\ &\le |D(f-r,g-s)| + |D(f-r,s)| + |D(r,g-s)| + |D(r,s)|\,. \end{aligned}$$

What is different from Theorem 4 in [1] is that we replace a part of the proof with the following results. We first consider Theorem 12 from the same paper [1]. Let the function h in this theorem be equal to e_1. Then we can write

$$|D(r,s)| \le \|r'\|_\infty \cdot \|s'\|_\infty \cdot |D(e_1,e_1)|$$

and we know that

$$0 \le |D(e_1,e_1)| = H(e_2;x) - H(e_1;x)^2 \le H((e_1-x)^2;x).$$

This last inequality is true, because

$$\begin{aligned} H((e_1-x)^2;x) &= H(e_2 - 2\cdot e_1 \cdot x + x^2;x) \\ &= H(e_2;x) - 2\cdot x \cdot H(e_1;x) + x^2 \cdot H(e_0;x) \\ &\ge H(e_2;x) - H(e_1;x)^2 \end{aligned}$$

is equivalent to

$$x^2 - 2\cdot x\cdot H(e_1;x) + H(e_1;x)^2 = (x - H(e_1;x))^2 \ge 0.$$

We then get

$$|D(r,s)| \le \|r'\|_\infty \cdot \|s'\|_\infty \cdot H((e_1-x)^2;x).$$

For $f-r \in C[a,b]$ and $g-s \in C[a,b]$ we have

$$|D(f-r,g-s)| \le \|f-r\|_\infty \cdot \|g-s\|_\infty\,.$$

Moreover, if $f-r \in C[a,b]$ and $s \in C^1[a,b]$, then

$$\begin{aligned} |D(f-r,s)| &\le \sqrt{D(f-r,f-r)\cdot D(s,s)} \\ &\le \|f-r\|_\infty \cdot \|s'\|_\infty \cdot \sqrt{H((e_1-x)^2;x)} \end{aligned}$$

and similarly, for $r \in C^1[a,b]$, $g-s \in C[a,b]$, we obtain

$$|D(r,g-s)| \le \|r'\|_\infty \cdot \|g-s\|_\infty \cdot \sqrt{H((e_1-x)^2;x)}.$$

If we combine all these inequalities, we have

$$
\begin{aligned}
&|D(f,g)| \le \|f-r\|_\infty \cdot \|g-s\|_\infty + \|f-r\|_\infty \cdot \|s'\|_\infty \cdot \sqrt{H((e_1-x)^2;x)} \\
&+ \|r'\|_\infty \cdot \|g-s\|_\infty \cdot \sqrt{H((e_1-x)^2;x)} + \|r'\|_\infty \cdot \|s'\|_\infty \cdot H((e_1-x)^2;x) \\
&= \|f-r\|_\infty \cdot \left\{ \|g-s\|_\infty + \|s'\|_\infty \cdot \sqrt{H((e_1-x)^2;x)} \right\} \\
&+ \|r'\|_\infty \cdot \sqrt{H((e_1-x)^2;x)} \cdot \left\{ \|g-s\|_\infty + \|s'\|_\infty \cdot \sqrt{H((e_1-x)^2;x)} \right\} \\
&= \left\{ \|f-r\|_\infty + \|r'\|_\infty \cdot \sqrt{H((e_1-x)^2;x)} \right\} \\
&\cdot \left\{ \|g-s\|_\infty + \|s'\|_\infty \cdot \sqrt{H((e_1-x)^2;x)} \right\}.
\end{aligned}
$$

We now pass to the infimum with respect to each of r, s and we obtain the wanted result:

$$
\begin{aligned}
&|D(f,g)| \\
&\le K\left(\sqrt{H((e_1-x)^2;x)}, f; C^0, C^1\right) \cdot K\left(\sqrt{H((e_1-x)^2;x)}, g; C^0, C^1\right) \\
&= \frac{1}{2}\widetilde{\omega}\left(f; 2\sqrt{H((e_1-x)^2;x)}\right) \cdot \frac{1}{2}\widetilde{\omega}\left(g; 2\sqrt{H((e_1-x)^2;x)}\right) \\
&= \frac{1}{4}\widetilde{\omega}\left(f; 2\sqrt{H((e_1-x)^2;x)}\right) \cdot \widetilde{\omega}\left(g; 2\sqrt{H((e_1-x)^2;x)}\right).
\end{aligned}
$$

This ends our proof. □

At present it is an open problem if the improved inequality in Theorem 4.1 can be generalized to $C(X)$ with (X, d) a compact metric space.

5. Applications

We can now apply the above improved result for different kinds of operators, like Bernstein- , convolution- and a special kind of Shepard-type operators.

5.1. Bernstein operator

As a first example, we have the following remark:

Remark 5.1. We consider $H := B_n$, the Bernstein operator defined by

$$
B_n(f;x) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \cdot \binom{n}{k} \cdot x^k (1-x)^{n-k},
$$

where $f \in C[0,1]$ and $x \in [0,1]$, $n = 1, 2, \ldots$. It is well known that the second moment of the Bernstein polynomial is equal to

$$
B_n((e_1-x)^2;x) = \frac{x(1-x)}{n}.
$$

Using Theorem 4.1, we get the Grüss-type inequality for the Bernstein operator as follows:

$$
\begin{aligned}
&|B_n(fg;x) - B_n(f;x)B_n(g;x)| \\
&\leq \frac{1}{4}\widetilde{\omega}\left(f; 2\sqrt{B_n((e_1-x)^2;x)}\right) \cdot \widetilde{\omega}\left(g; 2\sqrt{B_n((e_1-x)^2;x)}\right) \\
&= \frac{1}{4}\widetilde{\omega}\left(f; 2\sqrt{\frac{x(1-x)}{n}}\right) \cdot \widetilde{\omega}\left(g; 2\sqrt{\frac{x(1-x)}{n}}\right) \\
&\leq \frac{1}{4} \cdot \widetilde{\omega}\left(f; \frac{1}{\sqrt{n}}\right) \cdot \widetilde{\omega}\left(g; \frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

for two functions $f, g \in C[0,1]$.

5.2. Convolution-type operators

These types of operators were treated by many authors, like J.-D. Cao, H. Gonska and H.-J. Wenz (see [2]). One of the first authors to give the following definition was H.G. Lehnhoff in [8]:

Definition 5.2. *For the case $X = [-1,1]$, given a function $f \in C(X)$ and any natural number n, the convolution operator $G_{m(n)}$ is given by*

$$G_{m(n)}(f,x) := \frac{1}{\pi} \cdot \int_{-\pi}^{\pi} f(\cos(\arccos(x)+v)) \cdot K_{m(n)}(v)dv,$$

where the kernel $K_{m(n)}$ is a positive and even trigonometric polynomial of degree $m(n)$ satisfying

$$\int_{-\pi}^{\pi} K_{m(n)}(v)dv = \pi,$$

meaning that $G_{m(n)}(1,x) = 1$ for $x \in X$.

It is clear that $G_{m(n)}(f,\cdot)$ is an algebraic polynomial of degree $m(n)$ and the kernel $K_{m(n)}$ has the following form:

$$K_{m(n)}(v) = \frac{1}{2} + \sum_{k=1}^{m(n)} \rho_{k,m(n)} \cdot \cos(kv),$$

for $v \in [-\pi, \pi]$.

We also need another result that goes back to H.G. Lehnhoff [8]:

Lemma 5.3. *For $x \in X$ the inequality*

$$
\begin{aligned}
&G_{m(n)}((e_1-x)^2, x) \\
&= x^2 \left\{\frac{3}{2} - 2\cdot\rho_{1,m(n)} + \frac{1}{2}\cdot\rho_{2,m(n)}\right\} + (1-x^2)\cdot\left\{\frac{1}{2} - \frac{1}{2}\cdot\rho_{2,m(n)}\right\}
\end{aligned}
$$

holds. Here e_1 denotes the first monomial given by $e_1(t) = t$ for $|t| \leq 1$.

This lemma gives the second moment of the convolution-type operator, which we will need in the sequel.

Furthermore, we take into account different degrees $m(n)$, different convolution operators and Grüss-type inequalities, respectively.

5.2.1. Convolution-type operator with Fejér-Korovkin kernel. If we consider degree $m(n) = n - 1$, for $n \in \mathbb{N}$, the Fejér-Korovkin kernel is given by

$$K_{n-1}(v) = \frac{1}{n+1} \left(\frac{\sin\left(\frac{\pi}{n+1}\right) \cdot \cos\left((n+1)\frac{v}{2}\right)}{\cos(v) - \cos\left(\frac{\pi}{n+1}\right)} \right)^2$$

with

$$\rho_{1,n-1} = \cos\left(\frac{\pi}{n+1}\right), \; \rho_{2,n-1} = \frac{n}{n+1}\cos\left(\frac{2\pi}{n+1}\right) + \frac{1}{n+1}.$$

Using the latter relations, we get

$$\begin{aligned}
G_{n-1}\left((e_1 - x)^2; x\right) &\leq \left| \frac{3}{2} - 2 \cdot \rho_{1,n-1} + \frac{1}{2}\rho_{2,n-1} \right| + \frac{1}{2}\left|1 - \rho_{2,n-1}\right| \\
&\leq \left| \frac{3}{2} - 2\cos\left(\frac{\pi}{n+1}\right) + \frac{1}{2(n+1)} + \frac{n}{2(n+1)}\cos\left(\frac{2\pi}{n+1}\right) \right| \\
&+ \frac{1}{2} \cdot \left| 1 - \frac{1}{n+1} - \frac{n}{n+1} \cdot \cos\left(\frac{2\pi}{n+1}\right) \right| \\
&\leq 3 \cdot \left(\frac{\pi}{n+1}\right)^2 + \left(\frac{\pi}{n+1}\right)^2 \\
&= 4 \cdot \left(\frac{\pi}{n+1}\right)^2.
\end{aligned}$$

Having this preamble, we can now state the following result.

Theorem 5.4. *If we consider $f, g \in C(X)$ and the convolution-type operator of degree $n-1$ with the Fejér-Korovkin kernel, we have*

$$\begin{aligned}
|D(f,g)| &= |G_{n-1}(f \cdot g; x) - G_{n-1}(f; x) \cdot G_{n-1}(g; x)| \\
&\leq \frac{1}{4}\widetilde{\omega}\left(f; \frac{4\pi}{n+1}\right) \cdot \widetilde{\omega}\left(g; \frac{4\pi}{n+1}\right) \\
&= \mathcal{O}\left(\widetilde{\omega}\left(f; \frac{1}{n}\right) \cdot \widetilde{\omega}\left(g; \frac{1}{n}\right)\right).
\end{aligned}$$

5.2.2. Convolution-type operator with de La Vallée Poussin kernel. We now have degree $m(n) = n \in \mathbb{N}_0$ and we define the de La Vallée Poussin kernel by

$$V_n(v) = \frac{(n!)^2}{(2n)!} \cdot \left(2\cos\left(\frac{v}{2}\right)\right)^{2n}$$

with

$$\rho_{1,n} = \frac{n}{n+1}, \; \rho_{2,n} = \frac{(n-1)n}{(n+1)(n+2)}.$$

Using the two relations, we have the second moment:

$$\begin{aligned}G_n\left((e_1-x)^2;x\right) &\le \left|\frac{3}{2}-\frac{2n}{n+1}+\frac{1}{2}\cdot\frac{n(n-1)}{(n+1)(n+2)}\right| \\ &\quad+\frac{1}{2}\left|1-\frac{n(n-1)}{(n+1)(n+2)}\right| \\ &\le \left|\frac{3}{(n+1)(n+2)}\right|+\left|\frac{2n+1}{(n+1)(n+2)}\right| \\ &\le \frac{2}{n+1}.\end{aligned}$$

Taking this into account, we give the following theorem:

Theorem 5.5. *If we consider the convolution-type operator with the de La Vallée Poussin kernel we have*

$$\begin{aligned}|D(f,g)| &= |G_n(f\cdot g;x)-G_n(f;x)\cdot G_n(g;x)| \\ &\le \frac{1}{4}\widetilde{\omega}\left(f;\frac{2\sqrt{2}}{\sqrt{n+1}}\right)\cdot\widetilde{\omega}\left(g;\frac{2\sqrt{2}}{\sqrt{n+1}}\right) \\ &= \mathcal{O}\left(\widetilde{\omega}\left(f;\frac{1}{\sqrt{n}}\right)\cdot\widetilde{\omega}\left(g;\frac{1}{\sqrt{n}}\right)\right).\end{aligned}$$

5.2.3. Convolution-type operator with Jackson kernel. Finally, the last operator we consider is of degree $m(n)=2n-2$, with $n\in\mathbb{N}$. For this, the Jackson kernel has the form

$$J_{2n-2}(v)=\frac{3}{2n(2n^2+1)}\cdot\left(\frac{\sin\left(n\frac{v}{2}\right)}{\sin\left(\frac{v}{2}\right)}\right)^4$$

with

$$\rho_{1,2n-2}=\frac{2n^2-2}{2n^2+1},\ \rho_{2,2n-2}=\frac{2n^3-11n+9}{n(2n^2+1)}$$

and the second moment

$$\begin{aligned}G_{2n-2}\left((e_1-x)^2;x\right) &\le \left|\frac{3}{2}-\frac{4n^2-4}{2n^2+1}+\frac{1}{2}\cdot\frac{2n^3-11n+9}{n(2n^2+1)}\right| \\ &\quad+\frac{1}{2}\cdot\left|1-\frac{2n^3-11n+9}{n(2n^2+1)}\right| \\ &\le \left|\frac{9}{2n(2n^2+1)}\right|+\left|\frac{12n-9}{2n(2n^2+1)}\right| \\ &\le \frac{6}{2n^2+1}\le\frac{3}{n^2}.\end{aligned}$$

The result is as follows:

Theorem 5.6. *If we consider the convolution-type operator with the Jackson kernel we have*

$$\begin{aligned}|D(f,g)| &= |G_{2n-2}(f\cdot g;x) - G_{2n-2}(f;x)\cdot G_{2n-2}(g;x)|\\ &\le \frac{1}{4}\widetilde{\omega}\left(f;\frac{2\sqrt{3}}{n}\right)\cdot\widetilde{\omega}\left(g;\frac{2\sqrt{3}}{n}\right)\\ &= \mathcal{O}\left(\widetilde{\omega}\left(f;\frac{1}{n}\right)\cdot\widetilde{\omega}\left(g;\frac{1}{n}\right)\right).\end{aligned}$$

As we can see, the best degrees of approximation are obtained when dealing with the Grüss-type inequality for convolution operators in the cases of Fejér-Korovkin and Jackson kernels.

Remark 5.7. Another possibility is to apply the above obtained Grüss inequality for the Shepard-type operator defined on $C[0,1]$. But this result, just like in the case of the Hermite-Fejér operator, is disappointing (see Remark 7 in [1]).

6. A pre-Grüss-type inequality for the CBS operator

We now try to find a pre-Grüss inequality for the CBS-operator. Just like in the case of the pre-Grüss-type inequality for the Hermite-Fejér operator, obtained in [1] (see Theorem 8), the idea is to find a different approach. We consider the special case $X=[0,1]$, $d(x,y)=|x-y|$. Then, taking $H:=S_{n+1}^{\mu}$ the CBS operator based on $n+1$ equidistant points $x_i=\frac{i}{n}$, for $0\le i\le n$ and $1\le\mu\le 2$, we get:

Theorem 6.1. *Let $f,g\in C[0,1]$. Then the inequality*

$$\begin{aligned}&|D(f,g)|\\ &\le\frac{1}{2}\min\left\{\|f\|_\infty\,\widetilde{\omega_d}\left(g;4S_{n+1}^{\mu}\left(|e_1-x|\,;x\right)\right);\|g\|_\infty\,\widetilde{\omega_d}\left(f;4S_{n+1}^{\mu}\left(|e_1-x|\,;x\right)\right)\right\}\end{aligned}$$

holds.

Proof. We want to estimate

$$|D(f,g)| = \left|S_{n+1}^{\mu}(f\cdot g;x) - S_{n+1}^{\mu}(f;x)\cdot S_{n+1}^{\mu}(g;x)\right|.$$

For two fixed functions $f,g\in C[0,1]$ and an arbitrary $s\in C^1[0,1]$, we have

$$|D(f,g)| = |D(f,g-s+s)| \le |D(f,g-s)| + |D(f,s)|. \tag{6.1}$$

First, if we have $f \in C[0,1]$ and $s \in C^1[0,1]$, we continue with

$$\begin{aligned}
|D(f,s)| &= \left|S^{\mu}_{n+1}(f \cdot s;x) - S^{\mu}_{n+1}(f;x) \cdot S^{\mu}_{n+1}(s;x)\right| \\
&= \left|S^{\mu}_{n+1}(f(s - S^{\mu}_{n+1}(s;x));x)\right| \\
&= \left|S^{\mu}_{n+1,t}(f(t)(s(t) - s(x) + s(x) - S^{\mu}_{n+1}(s;x));x)\right| \\
&\le \|f\|_\infty \cdot S^{\mu}_{n+1,t}(|s(t)-s(x)| + \left|s(x) - S^{\mu}_{n+1}(s;x)\right|;x) \\
&\le \|f\|_\infty \cdot S^{\mu}_{n+1}(\|s'\|_\infty \cdot |e_1 - x| + \|s'\|_\infty \cdot S^{\mu}_{n+1}(|e_1 - x|;x);x) \\
&= 2 \cdot \|f\|_\infty \cdot \|s'\|_\infty \cdot S^{\mu}_{n+1}(|e_1 - x|;x).
\end{aligned}$$

If we now use this result in (6.1), we get

$$\begin{aligned}
|D(f,g)| &\le \|f\|_\infty \cdot \|g-s\|_\infty + 2 \cdot \|f\|_\infty \cdot \|s'\|_\infty \cdot S^{\mu}_{n+1}(|e_1 - x|;x) \\
&= \|f\|_\infty \left\{\|g-s\|_\infty + 2 \cdot \|s'\|_\infty \cdot S^{\mu}_{n+1}(|e_1 - x|;x)\right\}.
\end{aligned}$$

Passing to the infimum over $s \in C^1[0,1]$, it follows

$$\begin{aligned}
|D(f,g)| &\le \|f\|_\infty \cdot K(2 \cdot S^{\mu}_{n+1}(|e_1 - x|;x), g; C[0,1], C^1[0,1]) \\
&= \frac{1}{2} \cdot \|f\|_\infty \cdot \widetilde{\omega}\left(g, 4 \cdot S^{\mu}_{n+1}(|e_1 - x|;x)\right).
\end{aligned}$$

The same estimate holds if we interchange f and g. Putting both inequalities together, we get the result we were looking for. $\square$

In the above result, the first absolute moment of the CBS operator appears, which can be represented by

$$S^{\mu}_{n+1}(|e_1 - x|;x) = \begin{cases} \sum_{i=0}^{n} \frac{\left|x - \frac{i}{n}\right|^{1-\mu}}{\sum_{l=0}^{n} \left|x - \frac{l}{n}\right|^{-\mu}} & , x \notin \{x_0, \ldots, x_n\} \\ 0 & , \text{otherwise.} \end{cases}$$

The idea is to further estimate this quantity. For that, we use an idea from [5] (see proof of Theorem 4.3).

We distinguish three important cases for different values of μ.
The first case is $\mu = 1$. The first absolute moment of the CBS operator becomes

$$\begin{aligned}
S^{1}_{n+1}(|e_1 - x|;x) &= \begin{cases} \sum_{i=0}^{n} \frac{1}{\sum_{l=0}^{n} \left|x - \frac{l}{n}\right|^{-1}} & , x \notin \{x_0, \ldots, x_n\} \\ 0 & , \text{otherwise} \end{cases} \\
&= \begin{cases} (n+1) \left(\sum_{l=0}^{n} \frac{1}{\left|x - \frac{l}{n}\right|}\right)^{-1} & , x \notin \{x_0, \ldots, x_n\} \\ 0 & , \text{otherwise.} \end{cases}
\end{aligned}$$

Let now l_0 be defined by $\frac{l_0}{n} < x < \frac{l_0+1}{n}$. Then we have

$$\begin{aligned}\frac{1}{n+1}\cdot\left(\sum_{l=0}^{n}\frac{1}{\left|x-\frac{l}{n}\right|}\right) &\geq \frac{n}{n+1}\cdot\left\{\sum_{l=0}^{l_0}\frac{1}{l_0+1-l}+\sum_{l=l_0+1}^{n}\frac{1}{l-l_0}\right\}\\ &\geq \frac{n}{n+1}\left\{\int_1^{l_0+2}\frac{1}{x}dx+\int_1^{n-l_0+1}\frac{1}{x}dx\right\}\\ &= \frac{n}{n+1}\ln((l_0+2)\cdot(n-l_0+1))\\ &\geq \frac{n}{n+1}\cdot\ln(2n+2),\end{aligned}$$

and the second absolute moment is then

$$S^1_{n+1}(|e_1-x|\,;x)\leq\frac{n+1}{n\cdot\ln(2n+2)},$$

for $x\notin\{x_0,\dots,x_n\}$. In the end we get

$$\begin{aligned}&|D(f,g)|\\ &\leq\frac{1}{2}\min\left\{\|f\|_\infty\cdot\widetilde{\omega_d}\left(g;\frac{4(n+1)}{n\cdot\ln(2n+2)}\right),\|g\|_\infty\cdot\widetilde{\omega_d}\left(f;\frac{4(n+1)}{n\cdot\ln(2n+2)}\right)\right\}.\end{aligned}$$

For the other two cases we will consider, first let l_0 defined by

$$\left|x-\frac{l_0}{n}\right|=\min\left\{\left|x-\frac{l}{n}\right|:0\leq l\leq n\right\}.$$

Then for the case $x\notin\{x_0,\dots,x_n\}$, we have

$$\begin{aligned}S^\mu_{n+1}(|e_1-x|\,;x)&\leq|x-x_{l_0}|^\mu\cdot\sum_{i=0}^{n}|x-x_i|^{1-\mu}\\ &\leq\frac{1}{n}+\left(\frac{1}{n}\right)\cdot\left\{\sum_{i<l_0}|x-x_i|^{1-\mu}+\sum_{i>l_0}|x-x_i|^{1-\mu}\right\}\\ &\leq\frac{1}{n}+\left(\frac{1}{n}\right)\cdot\left\{\sum_{k=0}^{l_0-1}\left(\frac{1}{2}+k\right)^{1-\mu}+\sum_{k=0}^{n-l_0-1}\left(\frac{1}{2}+k\right)^{1-\mu}\right\},\end{aligned}$$

with $0\leq l_0\leq n$. Either of the two last sums may be empty. Estimating the result in the accolades from above, we get

$$S^\mu_{n+1}(|e_1-x|\,;x)\leq\begin{cases}\frac{1}{n}+\frac{1}{n}\cdot\left[2^\mu+\frac{2}{2-\mu}\cdot\left(\frac{n+1}{2}\right)^{2-\mu}\right] & \text{, for } 1<\mu<2\\ \frac{1}{n}+\frac{1}{n}\cdot[4+2\cdot\ln(n+1)] & \text{, for } \mu=2\end{cases}. \tag{6.2}$$

For $1<\mu<2$, we obtain

$$\begin{aligned}&|D(f,g)|\\ &\leq\frac{1}{2}\min\left\{\|f\|_\infty\,\widetilde{\omega_d}\left(g;4S^\mu_{n+1}(|e_1-x|\,;x)\right),\|g\|_\infty\,\widetilde{\omega_d}\left(f;4S^\mu_{n+1}(|e_1-x|\,;x)\right)\right\}\end{aligned}$$

where the first absolute moment can be estimated from above as in (6.2). For $\mu = 2$ we obtain

$$|D(f,g)|$$
$$\leq \frac{1}{2}\min\left\{ \|f\|_\infty \widetilde{\omega_d}\left(g; \frac{20 + 8\cdot\ln(n+1)}{n}\right), \|g\|_\infty \widetilde{\omega_d}\left(f; \frac{20 + 8\cdot\ln(n+1)}{n}\right)\right\}.$$

One can also obtain results for $\mu > 2$. This was done by G. Somorjai [12](see also J. Szabados [13] for $\mu > 4$), but we are not treating other cases in this article.

References

[1] Acu, A., Gonska, H., Raşa, I., *Grüss- and Ostrowski-type inequalities in Approximation Theory*, submitted for publication.

[2] Cao, J.-D., Gonska, H., Wenz, H.-J., *Approximation by Boolean sums of positive linear operators VII: Fejér-Korovkin kernels of higher order*, Acta Math. Hungar., **73**(1996), no. 1-2, 71–85.

[3] Crain, I.K., Bhattacharyya, B.K., *Treatment of non-equispaced two-dimensional data with a digital computer*, Geoexplorations, **5**(1967), 173–194.

[4] Gonska, H., *Quantitative Approximation in $C(X)$*, Habilitationsschrift, Univ. Duisburg, 1986.

[5] Gonska, H., *On approximation in spaces of continuous functions*, Bull. Austral. Math. Soc., **28**(1983), 411–432.

[6] Gordon, W.J., Wixom, J.A., *Shepard's method of 'metric interpolation' to bivariate and multivariate interpolation*, Math. Comp., **32**(1978), 253–264.

[7] Grüss, G., *Über das Maximum des absoluten Betrages von* $\frac{1}{b-a}\int_a^b f(x)g(x)dx - \frac{1}{(b-a)^2}\int_a^b f(x)dx \cdot \int_a^b g(x)dx$, Math. Z., **39**(1935), 215–226.

[8] Lehnhoff, H.G., *A simple proof of A.F. Timan's theorem*, J. Approx. Theory, **38**(1983), 172–176.

[9] Matsuoka, Y., *On the degree of approximation of functions by some positive linear operators*, Sci. Rep. Kagoshima Univ, **9**(1960), 11–16.

[10] Mitjagin, B.S., Semenov, E.M., *Lack of interpolation of linear operators in spaces of smooth functions*, Math. USSR-Izv., **11**(1977), 1229–1266.

[11] Shepard, D., *A two-dimensional interpolation function for irregularly spaced data*, Proc. ACM National Conference, 1968, 517–524.

[12] Somorjai, G., *On a saturation problem*, Acta Math. Acad. Sci. Hungar., **32**(1978), 377–381.

[13] Szabados, J., *On a problem of R. DeVore*, Acta Math. Acad. Sci. Hungar, **27**(1976), 219–223.

Maria-Daniela Rusu
University of Duisburg-Essen
Faculty of Mathematics
Forsthausweg 2
D-47057 Duisburg
Germany
e-mail: maria.rusu@uni-due.de

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 567–585

Bernstein quasi-interpolants on triangles

Paul Sablonnière

Abstract. The aim of this paper is to provide some results on Bernstein quasi-interpolants of different types applied to functions defined on a triangle. Classical multivariate Bernstein operators and their extensions have been studied for about 25 years by various authors. Based on their representation as differential operators, we extend our previous results on the univariate case to the multivariate one and we define new families of Bernstein quasi-interpolants. Then we compare their approximation properties on various types of functions. Our approach seems to be distinct from another interesting extension given in [5, 6].

Mathematics Subject Classification (2010): 41A35, 41A63.

Keywords: Bernstein operators.

1. Introduction and notations

The aim of this paper is to provide some results on Bernstein quasi-interpolants of different types applied to functions defined on a triangle. Classical multivariate Bernstein operators and their extensions have been studied for about 25 years by various authors (see references). These extensions are of Kantorovitch and Durrmeyer types. We only consider the latter together with the genuine case studied e.g. in [24, 27, 39, 47].

On the unit triangle $T := \{(x,y) \,|\, x,y \geq 0,\, 0 \leq x+y \leq 1\}$, the classical Bernstein quasi-interpolants are defined by

$$\mathcal{B}_n f(x,y) := \sum_{0\leq i+j\leq n} f\left(\frac{i}{n},\frac{j}{n}\right)\frac{n!}{i!j!k!}\,x^i y^j z^k, \quad z := 1-x-y, \quad k := n-i-j.$$

Using the notation $\alpha := (i,j) \in \Delta_n := \{(i,j)|0 \leq i+j \leq n\}$, we often write them as

$$\mathcal{B}_n f := \sum_{\alpha\in\Delta_n} f\left(\frac{\alpha}{n}\right) B^n_\alpha, \quad B^n_\alpha(x,y,z) := \frac{n!}{i!j!k!}x^i y^j z^k$$

where $\{B^n_\alpha, \alpha \in \Delta_n\}$ is the Bernstein basis of $\mathbb{P}_n$. The Durrmeyer extension has been first developed by Derriennic [13][14] in the case of the Legendre weight and later by various authors in the general case of Jacobi weights [7][8]. With the sacalar product

$$\langle f,g\rangle := \int_T w(x,y)f(x,y)g(x,y)dxdy, \quad w(x,y)=x^p y^q z^r,\ \ p,q,r>-1$$

the multivariate Bernstein-Durrmeyer (abbr. BD) operator is defined by

$$\mathcal{M}_n f := \sum_{\gamma\in\Delta_n} \langle \tilde{B}^n_\gamma, f\rangle B^n_\gamma, \quad \text{where} \quad \tilde{B}^n_\gamma := B^n_\gamma/\langle 1, B^n_\gamma\rangle$$

The genuine Bernstein-Durrmeyer (abbr. GBD) case corresponds to the limit weight $w(x,y)=1/xyz$ and has been studied e.g. in [47]. Its definition involves line integrals along the sides of the triangle T.
Using the representation of the above operators as differential operators in the space $\mathbb{P}$ of bivariate polynomials, we extend our previous results on univariate operators [40, 42, 44, 45, 46] to the bivariate ones and we define new families of Bernstein quasi-interpolants (partial results are given in [41, 44]). Then we compare their approximation properties on various types of functions. Our approach seems to be distinct from another interesting extension given by Berdysheva, Jetter and Stöckler in [3]-[6].
Here is a brief outline of the paper. In sections 2 and 3, we compute the differential forms of the operator $\mathcal{B}_n$ and its inverse $\mathcal{A}_n$ on the space $\mathbb{P}_n$ of polynomials of total degree at most n and we define the associated quasi-interpolants $\mathcal{B}_n^{(r)}$, $0\le r\le n$ (abbr. QIs). Then, in sections 4 and 5 (resp. 6 and 7), we follow the same program for Bernstein-Durrmeyer operators $\mathcal{M}_n$ with Legendre weight $w=1$ (resp. the genuine Bernstein-Durrmeyer operators $\mathcal{G}_n$). In section 8, we give some partial results on the asymptotic expansions and convergence orders of these various quasi-interpolants. In section 9, we give some results on numerical experiments done on the approximations of two functions by Bernstein and genuine Bernstein-Durrmeyer operators. Finally, in Section 10, we set some open problems that would be useful to solve for the applications of those QIs to various problems in approximation theory and numerical analysis.

2. The classical Bernstein operator

2.1. $\mathcal{B}_n$ and its inverse $\mathcal{A}_n=\mathcal{B}_n^{-1}$ as operators on $\mathbb{P}_n$

The classical Bernstein operator

$$\mathcal{B}_n f := \sum_{\alpha\in\Delta_n} f\left(\frac{\alpha}{n}\right) B^n_\alpha$$

where $\{B^n_\alpha, \alpha\in\Delta_n\}$ is the Bernstein basis of $\mathbb{P}_n$, is an **isomorphism** of the space $\mathbb{P}_n$ of bivariate polynomials of total degree at most n. This can be proved in various ways. For example, let $\{\ell^n_\alpha, \alpha\in\Delta_n\}$ be the **Lagrange basis** of $\mathbb{P}_n$ (see e.g. Ciarlet [11], chapter 2) based on points $\{\frac{\alpha}{n}, \alpha\in\Delta_n\}$, then $\mathcal{B}_n\ell^n_\alpha = B^n_\alpha$ for $\alpha\in\Delta_n$. Similarly, let $\{\nu^n_\alpha, \alpha\in\Delta_n\}$ be the **Newton basis** of

$\mathbb{P}_n$ based on the same points $\{\frac{\alpha}{n}, \alpha \in \Delta_n\}$, defined for $|\alpha| = i + j = p \leq n$ and using the Pochammer symbol $(n)_p = n(n-1)\dots(n-p+1)$, by

$$\nu_\alpha^n = \prod_{k=0}^{i-1}(nx - k)\prod_{\ell=0}^{j-1}(ny - \ell)/(n)_p$$

then $\mathcal{B}_n \nu_\alpha^n = m_\alpha$ where $m_\alpha(x,y) = m_{i,j}(x,y) := x^i y^j$ are the monomials of $\mathbb{P}_n$. So the image of the Lagrange (resp. Newton) basis is the Bernstein (resp. monomial) basis.
Denoting $\mathcal{A}_n = \mathcal{B}_n^{-1}$ the inverse operator of $\mathcal{B}_n$ on $\mathbb{P}_n$, then we have $\mathcal{A}_n B_\alpha^n = \ell_\alpha^n$ and $\mathcal{A}_n m_\alpha = \nu_\alpha^n$ for all $\alpha \in \Delta_n$. These properties are used below for the computation of the coefficients of $\mathcal{A}_n$ expressed as a differential operator.

2.2. $\mathcal{B}_n$ as a differential operator

As in the univariate case (see e.g. [33], chapter 1, and [45]), the operator $\mathcal{B}_n$ has the following representation in $\mathbb{P}_n$:

$$\mathcal{B}_n = Id + \sum_{r=2}^{n}\sum_{k+\ell=r}\beta_{k,\ell}D^{k,\ell}$$

Note that the polynomial coefficients $\beta_{k,\ell}$ should be denoted $\beta_{k,\ell}^{(n)}$ since they depend on n. However, we omit the upper index for the sake of clarity.
Theorem. *The polynomial coefficients $\beta_{k,\ell}$ satisfy the recurrence relation, for $k, \ell \geq 1$*

$$\begin{aligned} &n\left((k+1)\beta_{k+1,\ell} + (\ell+1)\beta_{k,\ell+1}\right) \\ &= (1-x-y)\left(x(\partial_{10}\beta_{k,\ell} + \beta_{k-1,\ell}) + y(\partial_{01}\beta_{k,\ell} + \beta_{k,\ell-1})\right). \end{aligned}$$

with $\beta_{0,0} = 1, \beta_{1,0} = \beta_{0,1} = 0$, and for $k, \ell \geq 1$

$$n(k+1)\beta_{k+1,0} = x(1-x)(\partial_{10}\beta_{k,0} + \beta_{k-1,0})$$

$$n(\ell+1)\beta_{0,\ell+1} = y(1-y)(\partial_{01}\beta_{0,\ell} + \beta_{0,\ell-1})$$

Proof. Using Taylor's formula

$$f(s,t) = f(x,y) + \sum_{r\geq 1}\frac{1}{r!}\left(\sum_{k+\ell=r}\binom{n}{k}(s-x)^k(t-y)^\ell D^{k,\ell}f(x,y)\right)$$

and applying the Bernstein operator

$$\mathcal{B}_n f(x,y) = f(x,y) + \sum_{n\geq 1}\frac{1}{n!}\left(\sum_{k+\ell=n}\binom{n}{k}B_n[(.-x)^k(.-y)^\ell](x,y)D^{k,\ell}f(x,y)\right)$$

we first obtain

$$\beta_{k,\ell}(x,y) := \frac{1}{n!}\binom{n}{k}B_n[(.-x)^k(.-y)^\ell](x,y).$$

or, setting $\phi_{k,\ell} = (.-x)^k(.-y)^\ell$ and $m := n-k-\ell$:

$$\beta_{k,\ell} = \frac{1}{k!\ell!(n-k-\ell)!}\sum_{i+j\leq n}\phi_{k,\ell}\left(\frac{i}{n},\frac{j}{n}\right)B_{i,j}^n$$

Let us compute the expression

$$z(xD^{1,0}+yD^{0,1})\beta_{k,\ell}=\frac{xzD^{1,0}+yzD^{0,1}}{k!\ell!m!}\mathcal{B}_n\phi_{k,\ell}$$

First we get

$$D^{1,0}\mathcal{B}_n\phi_{k,\ell}=-k\sum_{i+j\le n}\phi_{k-1,\ell}\left(\frac{i}{n},\frac{j}{n}\right)B^n_{i,j}+\sum_{i+j\le n}\phi_{k,\ell}\left(\frac{i}{n},\frac{j}{n}\right)D^{1,0}B^n_{i,j},$$

with

$$D^{1,0}B^n_{i,j}=n\left(B^{n-1}_{i-1,j}-B^{n-1}_{i,j}\right)$$

Moreover, we have

$$nxzB^n_{i-1,j}=izB^n_{i,j},\quad\text{and}\quad nxzB^n_{i,j}=(n-i-j)B^n_{i,j}$$

therefore

$$xzD^{1,0}\mathcal{B}_n\phi_{k,\ell}=-kxz\mathcal{B}_n\phi_{k-1,\ell}+z\sum i\phi_{k,\ell}\left(\frac{i}{n},\frac{j}{n}\right)B^n_{i,j}$$
$$-x\sum(n-i-j)\phi_{k,\ell}\left(\frac{i}{n},\frac{j}{n}\right)B^n_{i,j}.$$

Now, using the identities:

$$i=n\left(\frac{i}{n}-x\right)+nx,\quad\text{and}\quad i=n\left(z-n\left(\frac{i}{n}-x\right)-n\left(\frac{j}{n}-x\right)\right)$$

we obtain

$$xzD^{1,0}\mathcal{B}_n\phi_{k,\ell}=-kz\mathcal{B}_n\phi_{k-1,\ell}+n(1-y)\mathcal{B}_n\phi_{k+1,\ell}+nx\mathcal{B}_n\phi_{k,\ell+1}$$

In the same way, we also have

$$yzD^{0,1}\mathcal{B}_n\phi_{k,\ell}=-kz\mathcal{B}_n\phi_{k,\ell-1}+n(1-x)\mathcal{B}_n\phi_{k,\ell+1}+ny\mathcal{B}_n\phi_{k+1,\ell}$$

and finally

$$z(xD^{1,0}+yD^{0,1})\mathcal{B}_n\phi_{k,\ell}=-kz(x\mathcal{B}_n\phi_{k-1,\ell}+y\mathcal{B}_n\phi_{k,\ell-1})$$
$$+n(\mathcal{B}_n\phi_{k+1,\ell}+\mathcal{B}_n\phi_{k,\ell+1}),$$

which gives the following recurrence relation on the polynomial coefficients:

$$n(k+1)\beta_{k+1,\ell}+n(\ell+1)\beta_{k,\ell+1}=z\left(x(D^{1,0}\beta_{k,\ell}+\beta_{k-1,\ell})+y(D^{0,1}\beta_{k,\ell}+\beta_{k,\ell-1})\right).$$

□

Examples. Using the notations $X:=x(1-x),Y:=y(1-y)$, the first beta polynomials (depending on n) are given by

$$2n\beta_{2,0}=X,\qquad n\beta_{1,1}=-xy$$
$$6n^2\beta_{3,0}=X(1-2x),\qquad 2n^2\beta_{2,1}=-3xy(1-2x),$$
$$24n^3\beta_{4,0}=X(1+3(n-2)X),\qquad 6n^3\beta_{3,1}=-4xy(1+3(n-2)X),$$
$$4n^3\beta_{2,2}=xy(n-1-(n-2)(x+y)+3(n-2)xy)$$
$$5!n^4\beta_{5,0}=(1-2x)X(1+2(5n-6)X),\quad 24n^4\beta_{4,1}=-xy(1+2(5n-6)X)$$
$$12n^5\beta_{3,2}=10xy((n-1)(1-6x)-(n-2)y-(5n-6)x(x+3y-4xy))$$

2.3. $\mathcal{A}_n := \mathcal{B}_n^{-1}$ as a differential operator

2.3.1. First method: long recursion. The operator $\mathcal{A}_n$ has also the following representation in $\mathbb{P}_n$:

$$\mathcal{A}_n = Id + \sum_{p=2}^{n} \sum_{i+j=p} \alpha_{i,j} D^{i,j}$$

A first method, giving a long recursion, consists in deducing the polynomial coefficients from the identities $\mathcal{A}_n m_{k,\ell} = \nu^n_{k,\ell}$ for $0 \leq i+j \leq n$.

$$\nu^n_{k,\ell} = x^k y^\ell + \sum_{p=2}^{k+\ell} \sum_{i+j=p} \frac{k!}{(k-i)!} \frac{\ell!}{(\ell-j)!} x^{k-i} y^{\ell-j} \alpha_{i,j}$$

giving the (long) recursion

$$\alpha_{k,\ell} = \frac{\nu^n_{k,\ell} - m_{k,\ell}}{k!\ell!} - \sum_{(0,0)<(i,j)<(k,\ell)} \frac{x^{k-i}}{(k-i)!} \frac{y^{\ell-j}}{(\ell-j)!} \alpha_{i,j}.$$

2.3.2. Second method : expansion in the Newton basis. From the Taylor expansion of $f \in \mathbb{P}_n$:

$$f(.,.) = f(x,y) + \sum_{p=1}^{n} \sum_{k+\ell=p} \frac{(.-x)^k (.-y)^\ell}{k!\ell!} D^{k,l} f(x,y),$$

we deduce

$$\mathcal{A}_n f = f + \sum_{p=1}^{n} \left(\sum_{k+\ell=p} \mathcal{A}_n \left[\frac{(.-x)^k (.-y)^\ell}{k!\ell!} \right] D^{k,l} f(x,y) \right)$$

giving

$$\alpha_{k,\ell}(x,y) = \mathcal{A}_n \left[\frac{(.-x)^k (.-y)^\ell}{k!\ell!} \right]$$

and since $\mathcal{A}_n m_{ij} = \nu_{i,j}$, we obtain the compact form :

$$\alpha_{k,\ell}(x,y) = \frac{(-1)^p}{k!\ell!} \sum_{i=0}^{k} \sum_{j=0}^{\ell} \binom{k}{i} \binom{\ell}{j} (-1)^{i+j} x^{k-i} y^{\ell-j} \nu_{i,j}(x,y).$$

2.3.3. Third method : direct short recursion. At least for polynomials $\alpha_{k,0}$ and $\alpha_{0,\ell}$, we have the short recursions [45]

$$(k+1)(n-k)\alpha_{k+1,0} = -k(1-2x)\alpha_{k,0} - X\alpha_{k-1,0}.$$

$$(\ell+1)(n-\ell)\alpha_{0,\ell+1} = -k(1-2y)\alpha_{0,\ell} - Y\alpha_{0,\ell-1}.$$

Following the model of beta-polynomials:

$$(k+1)n\beta_{k+1,\ell} + n(\ell+1)\beta_{k,\ell+1} = z\left(x(D^{1,0}\beta_{k,\ell} + \beta_{k-1,\ell}) + y(D^{0,1}\beta_{k,\ell} + \beta_{k,\ell-1})\right).$$

it would be possible to get a recursion for the computation of these polynomials. However, it is still an open question.

2.3.4. A table of polynomials alpha. With the notations $X = x(1-x), Y = y(1-y), n_k := (n-1)\dots(n-k), [i,j] := \alpha_{i,j}$, here are the first polynomials alpha

$$2n_1[2,0] = X, \quad n_1[1,1] = xy, \quad 2n_1[2,0] = Y$$

$$3n_2[3,0] = (1-2x)X \quad n_2[2,1] = -xy(1_2 x),$$

$$n_2[1,2] = -xy(1-2y), \quad 3n_2[0,3] = (1-2y)Y$$

$$8n_3[4,0] = -X(2-(n+6)X), \quad 2n_3[3,1] = xy(2-(n+6)X),$$

$$4n_3[2,2] = xy(n-(n+6)(x+y-3xy))$$

$$30n_4[5,0] = (1-2x)X(6-(5n+12)X),$$

$$6n_4[4,1] = -xy(1-2x)(6-(5n+12)X)$$

$$6n_4[3,2] = -xy(n-6nx-(n+12)y+(5n+12)x(x+3y-4xy))$$

3. Bernstein quasi-interpolants

3.1. Quasi-interpolants of order r

Given $0 \leq r \leq n$, define the truncated inverse of order r

$$\mathcal{A}_n^{(r)} = Id + \sum_{p=2}^{r} \sum_{i+j=p} \alpha_{i,j} D^{i,j}$$

Then the Bernstein-quasi-interpolant (abbr. BQI) of order r is defined by

$$\mathcal{B}_n^{(r)} = \mathcal{A}_n^{(r)} \mathcal{B}_n$$

In other words, for all polynomial $p \in \mathbb{P}_n$, we have

$$\mathcal{B}_n^{(r)} p = \mathcal{B}_n p + \sum_{p=2}^{r} \sum_{i+j=p} \alpha_{i,j} D^{i,j} \mathcal{B}_n p$$

Theorem. *The operator $\mathcal{B}_n^{(r)}$ is exact on $\mathbb{P}_r$, for all $0 \leq r \leq n$.*
Proof. As $p = \mathcal{A}_n \mathcal{B}_n p = \mathcal{B}_n^{(n)} p$, we can write

$$p - \mathcal{B}_n^{(r)} p = \sum_{p=r+1}^{n} \sum_{i+j=p} \alpha_{i,j} D^{i,j} \mathcal{B}_n p$$

As $p \in \mathbb{P}_r$, we have $\mathcal{B}_n p \in \mathbb{P}_r$, thus $D^{i,j} \mathcal{B}_n p = 0$ for all (i,j) satisfying $i + j = p \geq r + 1$, thus $p - \mathcal{B}_n^{(r)} p = 0$. □

Therefore, we have constructed **a chain of intermediate operators** between the classical Bernstein operator and the identity operator which can be written in the form of the Lagrange interpolation operator $\mathcal{L}_n$ since $\mathcal{A}_n B_\alpha^n = \ell_\alpha^n$:

$$p = \mathcal{A}_n \mathcal{B}_n p = \sum_{\alpha \in \Delta_n} f\left(\frac{\alpha}{n}\right) \mathcal{A}_n B_\alpha^n = \sum_{\alpha \in \Delta_n} f\left(\frac{\alpha}{n}\right) \ell_\alpha^n = \mathcal{L}_n p$$

3.2. Some open questions on BQIs

Among the open questions relative to the BQIs, the following seem particularly interesting:
1) Prove, as in the univariate case [50], that for $r \in \mathbb{N}$ fixed, the BQIs of order r are uniformly bounded, i.e. there exists a constant C_r such that

$$\|\mathcal{B}_n^{(r)}\|_\infty \le C_r \quad \text{for all } \; n \ge r$$

2) Numerical experiments show that some functions f (e.g. of Runge type) are better approximated by intermediate polynomials $\mathcal{B}_n^{(r)} f$ rather than by their Lagrange interpolant. This is not quite surprising in view of the fact that $\|\mathcal{L}_n\|_\infty$ goes to infinity rather fastly when $n \to \infty$ (see e.g. [9]). Therefore the approximating polynomials generated in this way can be useful in practice, in approximation as well as in CAGD.
3) It would be interesting to have a direct formula giving the polynomial coefficients $\alpha_{i,j}$, or at least a short recursive formula.

4. Bernstein-Durrmeyer operators

For the sake of simplicity, we take $w = 1$ (Legendre) and we only consider Bernstein Durrmeyer quasi-interpolants (abbr. BDQIs) in that case. Of course, the same technique can be extended to general BDQIs with an arbitrary Jacobi weight. It would be also interesting to study the generalizations recently proposed in [3, 4]. Setting

$$\langle f, g \rangle := \int_T f(x,y)g(x,y)dxdy$$

since area$(T) = 1/2$, we have

$$\int_T B_\gamma^n = \frac{1}{(n+1)(n+2)}$$

whence the definition of the BD operator:

$$\mathcal{M}_n f := (n+1)(n+2) \sum_{\gamma \in \Delta_n} \langle B_\gamma^n, f \rangle B_\gamma^n$$

4.1. $\mathcal{M}_n$ and $\mathcal{K}_n = \mathcal{M}_n^{-1}$ as operators on $\mathbb{P}_n$

Consider a family of orthogonal polynomials $\{P_{k,\ell}, \;\; 0 \le |\gamma| = k + \ell \le n\}$ on T (see e.g. [12, 21, 22, 48]) whose expansion in the BB basis is the following:

$$P_\gamma = \sum_{\delta \in \Delta_n} p(\delta, \gamma) B_\delta^n$$

It is known (see e.g. [13]) that for $\gamma \in \Delta_s$, with $0 \le s \le n$, one has

$$\mathcal{M}_n P_\gamma = \rho_\gamma(n) P_\gamma,$$

where the eigenvalue is given by

$$\rho_\gamma(n) = \frac{[n]_s}{(n+3)_s} = \frac{\Gamma(n+1)}{\Gamma(n-s+1)} \frac{\Gamma(n+3)}{\Gamma(n+s+3)}$$

We use here the Pochhammer symbol defined by

$$(n)_s := n(n+1)\dots(n+s-1) = \frac{(n+s-1)!}{(n-1)!} = \frac{\Gamma(n+s)}{\Gamma(n)}$$

and we set

$$[n]_s := n(n-1)\dots(n-s+1) = \frac{n!}{(n-s)!} = \frac{\Gamma(n+1)}{\Gamma(n-s+1)}$$

Thus $\mathcal{M}_n$ is an automorphism of $\mathbb{P}_n$. Denoting $\mathcal{K}_n = \mathcal{M}_n^{-1}$, we have

$$\mathcal{K}_n P_\gamma = \rho_\gamma^{-1}(n) P_\gamma, \quad \gamma \in \Delta_n$$

4.2. $\mathcal{M}_n$ as a differential operator on $\mathbb{P}_n$

Like the classical Bernstein operator, the BD operator $\mathcal{M}_n$ can be expressed as a differential operator on $\mathbb{P}_n$:

$$\mathcal{M}_n = \sum_{r=0}^{n} \sum_{\delta \in \Delta_r} \mu_\delta^{(n)} D^\delta, \quad \mu_\delta^{(n)} \in \mathbb{P}_r$$

Therefore, for $|\gamma| = m \le n$:

$$\mathcal{M}_n P_\gamma = \sum_{r=0}^{m} \sum_{\delta \in \Delta_r} \mu_\delta^{(n)} D^\delta P_\gamma = \rho_\gamma(n) P_\gamma$$

As in Section 2.2, a direct expression of the polynomials $\mu_\delta^{(n)}$ for $\delta = (k, \ell) \in \Delta_r$, can be deduced from Taylor's formula:

$$\mu_\delta^{(n)} = \frac{1}{r!} \binom{r}{k} \mathcal{M}_n[(. - x)^k (. - y)^\ell]$$

4.3. $\mathcal{K}_n := \mathcal{M}_n^{-1}$ as a differential operator

One can also write $\mathcal{K}_n$ as a differential operator on $\mathbb{P}_n$:

$$\mathcal{K}_n = \sum_{r=0}^{n} \sum_{\delta \in \Delta_r} \kappa_\delta^{(n)} D^\delta, \quad \kappa_\delta^{(n)} \in \mathbb{P}_r$$

Therefore, for $|\gamma| = m \le n$, we have the long recursion:

$$\mathcal{K}_n P_\gamma = \sum_{r=0}^{m} \sum_{\delta \in \Delta_r} \kappa_\delta^{(n)} D^\delta P_\gamma = \rho_\gamma^{-1}(n) P_\gamma$$

For the computation of the polynomial coefficients κ, we did not use this method. Rather, we compute the polynomials $p_\gamma := \mathcal{M}_n m_\gamma$ from which we deduce $\mathcal{K}_n p_\gamma = m_\gamma$ as follows.

4.4. The polynomials p_γ

In order to find the polynomial p_γ whose image by $\mathcal{M}_n$ is the monomial $m_\gamma := x^i y^j$, i.e. such that $\mathcal{K}_n m_\gamma = p_\gamma$, we write

$$p_\gamma := \sum_{\delta \in \Delta_n} c(\gamma, \delta) B_\delta^{(n)}$$

Setting

$$B_\gamma^n := B_{i,j}^n := \frac{n!}{i!j!k!} x^i y^j z^k, \quad k := n - i - j, \quad \text{for } \gamma := (i,j) \in \Delta_n,$$

$$B_\delta^n := B_{p,q}^n := \frac{n!}{p!q!r!} x^p y^q z^r, \quad r := n - p - q, \quad \text{for } \delta := (p,q) \in \Delta_n$$

and introducing the Gram matrix

$$G[\gamma, \delta] := \langle B_\gamma^n, B_\delta^n \rangle = \frac{1}{(n+1)^2} \frac{\binom{i+p}{i}\binom{j+q}{j}\binom{k+r}{k}}{\binom{2n+2}{n+1}}$$

we obtain

$$\mathcal{M}_n p_\gamma = \sum_{\delta \in \Delta_n} c(\gamma, \delta) \mathcal{M}_n B_\delta^{(n)} = \frac{1}{2}(n+1)(n+2) \sum_{\delta \in \Delta_n} c(\gamma, \delta) \left(\sum_{\theta \in \Delta_n} G[\delta, \theta] B_\theta^n \right)$$

$$\mathcal{M}_n p_\gamma = \frac{1}{2}(n+1)(n+2) \sum_{\theta \in \Delta_n} \left(\sum_{\delta \in \Delta_n} G[\theta, \delta] c(\gamma, \delta) \right) B_\theta^n$$

Now, we need the representation of the monomial m_γ in the BB basis:

$$m_{i,j} = \sum_{\theta \in \Delta_n} \frac{\binom{i}{r}\binom{j}{s}}{\binom{n}{r,s}} B_\theta^n, \quad \theta := (r,s)$$

By identification, we compute $c(\gamma, \delta)$ as the solution of the system of linear equations

$$\frac{1}{2}(n+1)(n+2) \sum_{\delta \in \Delta_n} G[\theta, \delta] c(\gamma, \delta) = \frac{\binom{i}{r}\binom{j}{s}}{\binom{n}{r,s}}, \quad \theta \in \Delta_n$$

4.5. A table of the first polynomials kappa

The list of the first kappa polynomials shows that they are more complex than alpha polynomials of section 2.3.4 :

$$n\kappa_{1,0}^{(n)} = 3x - 1, \quad n\kappa_{0,1}^{(n)} = 3y - 1$$

$$(n)_2\, \kappa_{2,0}^{(n)} = (n+9)x^2 - (n+7)x + 1$$

$$(n)_2\, \kappa_{1,1}^{(n)} = 2(n+9)xy - 4(x+y) + 1$$

$$(n)_2\, \kappa_{0,2}^{(n)} = (n+9)y^2 - (n+7)y + 1$$

$$(n)_3\, \kappa_{3,0}^{(n)} = 5(n+5)x^3 - (7n+31)x^2 + (2n+11)x - 1$$

$$(n)_3\, \kappa_{2,1}^{(n)} = 15(n+5)x^2 y - (n+13)x^2 - 4(2n+11)xy + (n+8)x + 5y - 1$$

$$(n)_3\, \kappa_{1,2}^{(n)} = 15(n+5)xy^2 - (n+13)y^2 - 4(2n+11)xy + 5x + (n+8)y - 1$$

$$(n)_3\,\kappa_{0,3}^{(n)} = 5(n+5)y^3 - (7n+31)y^2 + (2n+11)y - 1$$

$$(n)_4\,\kappa_{4,0}^{(n)} = \frac{1}{2}((n+4)(n+33)x^4 - 2(n^2+34n+113)x^3$$

$$+(n+4)(n+33))x^2 - 6(n+5)x + 2)$$

$$(n)_4\,\kappa_{3,1}^{(n)} = 2(n+4)(n+33)x^2y(x-1) - 2(3n+19)x^3$$

$$+(8n+39)x(x+2y) - 2(n+6)x - 6y + 1$$

$$(n)_4\,\kappa_{2,2}^{(n)} = 3(n+4)(n+33)x^2y^2 - (n^2+46n+189)xy(x+y) + (n+18)(x^2+y^2)$$

$$+(n+5)(n+18)xy - (n+9)(x+y) + 1$$

$$\kappa_{1,3}^{(n)}(x,y) = \kappa_{3,1}^{(n)}(y,x), \qquad \kappa_{0,4}^{(n)}(x,y) = \kappa_{4,0}^{(n)}(y,x).$$

5. Bernstein-Durrmeyer quasi-interpolants

5.1. Bernstein-Durrmeyer quasi-interpolants of order r

Given $0 \le r \le n$, define the truncated inverse of order r

$$\mathcal{K}_n^{(r)} = Id + \sum_{p=2}^{r} \sum_{i+j=p} \kappa_{i,j} D^{i,j}$$

Then the Bernstein-Durrmeyer quasi-interpolant (abbr. BDQI) of order r is defined by

$$\mathcal{M}_n^{(r)} = \mathcal{K}_n^{(r)} \mathcal{M}_n$$

In other words, for all polynomial $p \in \mathbb{P}_n$, we have

$$\mathcal{M}_n^{(r)} p = \mathcal{M}_n p + \sum_{p=2}^{r} \sum_{i+j=p} \kappa_{i,j} D^{i,j} \mathcal{M}_n p$$

Theorem. *The operator $\mathcal{M}_n^{(r)}$ is exact on $\mathbb{P}_r$, for all $0 \le r \le n$.*
The proof is the same as for BQIs.
Therefore, we have constructed a chain of intermediate operators between the Bernstein-Durrmeyer operator and the identity operator. The latter can be written in the form of the **orthogonal projector** $\mathcal{P}_n$ on the space $\mathbb{P}_n$. Indeed, since $\mathcal{M}_n$ is a self-adjoint isomorphism in that space, we have, for all $p \in \mathbb{P}_n$:

$$0 = \langle f - \mathcal{P}_n f, \mathcal{M}_n p \rangle = \langle \mathcal{M}_n(f - \mathcal{P}_n f), p \rangle$$

As $\mathcal{M}_n(f - \mathcal{P}_n f) \in \mathbb{P}_n$, this implies first that $\mathcal{M}_n f = \mathcal{M}_n \mathcal{P}_n f$, i.e. $\mathcal{M}_n \mathcal{K}_n \mathcal{M}_n f = \mathcal{M}_n \mathcal{P}_n f$ and second that $\mathcal{K}_n \mathcal{M}_n f = \mathcal{P}_n f$, in other words $\mathcal{K}_n \mathcal{M}_n = \mathcal{K}_n \mathcal{M}_n$, q.e.d. □

5.2. Some open questions on BDQIs

Among the open questions relative to the BDQIs, the following seem particularly interesting:
1) Prove that for $r \in \mathbb{N}$ fixed, the BDQIs of order r are uniformly bounded for L^p norms i.e. there exists constants $C(r,p)$ such that

$$\|\mathcal{B}_n^{(r)}\|_p \leq C(r,p) \quad \text{for all } \; n \geq r$$

2) As for BQIs, numerical experiments show that some functions f (e.g. of Runge type) are better approximated by intermediate polynomials $\mathcal{M}_n^{(r)} f$ rather than by their L^2-orthogonal projection $\mathcal{P}_n f$ on $\mathbb{P}_n$. (This is not quite surprising in view of the fact that $\|\mathcal{L}_n\|_\infty$ goes to infinity fastly when $n \to \infty$). Therefore the approximating polynomials generated in this way can be useful in practice, both in approximation and in CAGD.
3) It would be interesting to have a direct formula giving the polynomial coefficients $\kappa_{i,j}$, or at least a recursive formula allowing their fast computation.
4) From the computational point of view, it would be also interesting to have a fast algorithm for the effective computation of scalar products $\langle B_\gamma^n, f\rangle$. Even though the Bernstein polynomials are Jacobi weights (up to a constant), using the corresponding Gauss-Jacobi cubature formulas seem rather complicated since weights and data points are distinct.

6. Genuine Bernstein-Durrmeyer operators

Let f_s denote the restriction of f to the edge opposite to the vertex $A_s = (e_s)$ (barycentric coordinates :s $e_1 = (1,0,0), e_2 = (0,1,0)$ and $e_3 = (0,0,1)$), let b_{k-1}^{n-2} be the univariate Bernstein polynomials on that edge, and let Δ_n^* be the set of indices $\gamma \in \Delta_n$ with no null component. Then the genuine Bernstein-Durrmeyer (abbr. GBD) operators are defined by

$$\mathcal{G}_n f := \sum_{r=1}^{3} f(e_r) B_{ne_r}^n + (n-1) \sum_{s=1}^{3} \sum_{k=1}^{n-1} \langle f_s, b_{k-1}^{n-2}\rangle B_k^n$$

$$+(n-1)(n-2) \sum_{\gamma \in \Delta_n^*} \langle f, B_\alpha^{n-3}\rangle B_\alpha^n$$

Note that in the second sum, $\langle f_s, b_{k-1}^{n-2}\rangle$ is a univariate sacalar product along the edge, and B_k^n is an abbreviation for B_α^n when $\alpha = (k, n-k, 0)$, $(k, 0, n-k)$ or $(0, k, n-k)$.
Like the classical Bernstein and the BD operators, the GBD operator $\mathcal{G}_n$ can be expressed as a differential operator on $\mathbb{P}_n$:

$$\mathcal{G}_n = \sum_{r=0}^{n} \sum_{\delta \in \Delta_r} \theta_\delta^{(n)} D^\delta, \quad \bar{\beta}_\delta^{(n)} \in \mathbb{P}_r$$

The inverse operator $\mathcal{H}_n := \mathcal{G}_n^{-1}$ can also be expressed as a differential operator on $\mathbb{P}_n$:

$$\mathcal{H}_n = \sum_{r=0}^{n} \sum_{\delta \in \Delta_r} \eta_\delta^{(n)} D^\delta, \quad \bar{\alpha}_\delta^{(n)} \in \mathbb{P}_r$$

7. Genuine Bernstein-Durrmeyer quasi-interpolants

7.1. Genuine Bernstein-Durrmeyer quasi-interpolants of order r

Given $0 \le r \le n$, define the truncated inverse of order r

$$\mathcal{H}_n^{(r)} = Id + \sum_{p=2}^{r} \sum_{i+j=p} \theta_{i,j} D^{i,j}$$

Then the Genuine Bernstein-Durrmeyer quasi-interpolant (abbr. GBDQI) of order r is defined by

$$\mathcal{G}_n^{(r)} = \mathcal{H}_n^{(r)} \mathcal{G}_n$$

$$\mathcal{G}_n^{(r)} := \sum_{|\gamma|=0}^{r} \eta_\gamma^{(n)} D^\gamma G_n, \quad 0 \le r \le n$$

Theorem. *The operator $\mathcal{G}_n^{(r)}$ is exact on $\mathbb{P}_r$, for all $0 \le r \le n$.*
The proof is the same as for BQIs and BDQIs.

7.2. A table of the first polynomials eta

With the notation $n_k := (n-1)\ldots(n-k)$, here are the first polynomials

$$n_1 \eta_{20}^{(n)} = -X, \quad n_1 \bar{\eta}_{11}^{(n)} = 2xy$$

$$n_2 \eta_{30}^{(n)} = (1-2x)X, \quad n_2 \bar{\eta}_{21}^{(n)} = -3xy(1-2x)$$

$$2n_3 \eta_{40}^{(n)} = X((n+7)X-2), \quad n_3 \bar{\eta}_{31}^{(n)} = -2xy((n+7)X-2)$$

$$n_3 \eta_{22}^{(n)} = xy((n+7)(3xy-x-y)+n+1)$$

$$n_4 \eta_{5,0} := (1-2x)X(1-(n+3)X), \quad n_4 \eta_{4,1} := 5(2x-1)(1-(n+3)X)xy$$

$$n_4 \eta_{3,2} := (5(n+3)x(4xy-x-3y)+(n+1)(6x-1)+(n+11)y)xy$$

8. Asymptotic formulas for Bernstein type quasi-interpolants

We only sketch a study the convergence for polynomials though the results can be extended to smooth functions (this will be developed elsewhere). Given a polynomial $p \in \mathbb{P}$, we are interested in the following limits:

$$\lim n^{r+1}(\mathcal{Q}_n^{(2r)} p(x) - p(x)) \text{ and } \lim n^{r+1}(\mathcal{Q}_n^{(2r+1)} p(x) - p(x))$$

where $\mathcal{Q}_n^{(s)}, s = 2r, 2r+1$ is one of the three types of Bernstein QIs previously defined. For original operators (case $s = 0$), see also [1, 2, 33, 34, 48].

8.1. Bernstein QIs

For beta and alpha polynomials, we define the polynomials

$$\bar{\beta}_{k,\ell} = \lim n^r \beta_{k,\ell} \quad \text{for} \quad k+\ell = 2r-1 \text{ or } 2r$$

$$\bar{\alpha}_{k,\ell} = \lim n^r \alpha_{k,\ell} \quad \text{for} \quad k+\ell = 2r-1 \text{ or } 2r$$

From the recurrence formulas of section 2.2, we immediately deduce the following

Theorem. *The following recurrence relations hold:*

$$(k+1)\bar{\beta}_{k+1,\ell} + (\ell+1)\bar{\beta}_{k,\ell+1} = z\left(x\bar{\beta}_{k-1,\ell} + y\bar{\beta}_{k,\ell-1}\right) \quad \text{for} \quad k+\ell = 2r-1,$$

$$(k+1)\bar{\beta}_{k+1,\ell} + (\ell+1)\bar{\beta}_{k,\ell+1} = z\left(xD^{1,0}\bar{\beta}_{k,\ell} + yD^{0,1}\bar{\beta}_{k,\ell}\right) \quad \text{for} \quad k+\ell = 2r.$$

We have not yet obtained the general formulas for alpha-polynomials. However, for polynomials $\bar{\alpha}_{k,0}$ and $\bar{\alpha}_{0,\ell}$, we deduce from the recurrence formulas of section 2.3.4 :

$$(2r+1)\bar{\alpha}_{2r+1,0} = -2r(1-2x)\bar{\alpha}_{2r,0} - X\bar{\alpha}_{2r-1,0} \qquad (2r+2)\bar{\alpha}_{2r+2,0} = -X\bar{\alpha}_{2r,0},$$

$$(2r+1)\bar{\alpha}_{0,2r+1} = -2r(1-2y)\bar{\alpha}_{0,2r} - Y\bar{\alpha}_{0,2r-1} \qquad (2r+2)\bar{\alpha}_{0,2r+2} = -Y\bar{\alpha}_{0,2r},$$

Here is a table of the first polynomials:

(k,ℓ)	$\beta_{k,\ell}$	$\alpha_{k,\ell}$
$(2,0)$	$X/2$	$-X/2$
$(1,1)$	$-xy$	xy
$(3,0)$	$(1-2x)X/6$	$(1-2x)X/3$
$(2,1)$	$-xy(1-2x)/2$	$-xy(1-2x)$
$(4,0)$	$X^2/8$	$X^2/8$
$(3,1)$	$-xyX/2$	$-xyX/2$
$(2,2)$	$xy(z+3xy)/4$	$xy(z+3xy)/4$

The asymptotic formulas are obtained as follows. For any polynomial f:

$$f - \mathcal{B}_n^{(q)} f = \sum_{p \geq q+1} \sum_{i+j=p} \alpha_{i,j} D^{i,j} f$$

For $q = 2r-1$, we get

$$n^r(f - \mathcal{B}_n^{(2r)} f) = \sum_{p \geq 2r} \sum_{i+j=p} n^r \alpha_{i,j} D^{i,j} \mathcal{B}_n f$$

As $\lim n^r \alpha_{i,j} = \bar{\alpha}_{i,j}$ for $i+j = 2r$, $\lim n^r \alpha_{i,j} = 0$ for $i+j = p > 2r$ and $\lim D^{i,j}\mathcal{B}_n f = D^{i,j} f$, we obtain:

$$\lim n^r(f - \mathcal{B}_n^{(2r)} f) = \sum_{i+j=2r} \bar{\alpha}_{i,j} D^{i,j} f$$

Similarly, for $q = 2r$, we get

$$n^{r+1}(f - \mathcal{B}_n^{(2r+1)} f) = \sum_{p \geq 2r+1} \sum_{i+j=p} n^{r+1} \alpha_{i,j} D^{i,j} \mathcal{B}_n f$$

As $\lim n^{r+1}\alpha_{i,j} = \bar{\alpha}_{i,j}$ for $i+j = 2r+1, 2r+2$, $\lim n^{r+1}\alpha_{i,j} = 0$ for $i+j = p > 2r+2$ and $\lim D^{i,j}\mathcal{B}_n f = D^{i,j} f$, we obtain:

$$\lim n^{r+1}(f - \mathcal{B}_n^{(2r+1)} f) = \sum_{p=2r+1}^{2r+2} \sum_{i+j=p} \bar{\alpha}_{i,j} D^{i,j} f$$

Examples.

$$\lim n(f - \mathcal{B}_n^{(2)} f) = -\frac{1}{2}(XD^{2,0}f - xyD^{1,1}f + YD^{0,2}f)$$

$$\lim n^2(f - \mathcal{B}_n^{(3)} f) = \sum_{|\gamma|=3} \bar{\alpha}_\gamma D^\gamma f + \sum_{|\gamma|=4} \bar{\alpha}_\gamma D^\gamma f$$

$$= \frac{1}{3}(1-2x)XD^{3,0}f - xy(1-2x)D^{2,1}f - xy(1-2xy)D^{1,2}f + (1-2y)YD^{0,3}f$$

$$+\frac{1}{8}X^2D^{4,0}f - \frac{1}{2}xyXD^{3,1}f + \frac{1}{4}xy(z+3xy)D^{2,2}f - \frac{1}{2}xyYD^{1,3}f + \frac{1}{8}Y^2D^{0,4}f$$

8.2. Bernstein-Durrmeyer QIs

For lambda and kappa polynomials, we define

$$\bar{\lambda}_{k,\ell} = \lim n^r \lambda_{k,\ell} \quad \text{for} \quad k+\ell = 2r-1 \text{ or } 2r$$

$$\bar{\kappa}_{k,\ell} = \lim n^r \kappa_{k,\ell} \quad \text{for} \quad k+\ell = 2r-1 \text{ or } 2r$$

Here is a table of the first polynomials $\bar{\kappa}_{k,\ell}$:

(k,ℓ)	$\bar{\kappa}_{k,\ell}$
$(1,0)$	$3x-1$
$(2,0)$	$-X$
$(1,1)$	$2xy$
$(3,0)$	$-X(5x-2)$
$(2,1)$	$x(15xy-x-8y+1)$
$(4,0)$	$X^2/2$
$(3,1)$	$-2xyX$
$(2,2)$	$xy(3xy-(x+y)+1)$

As for Bernstein QIs, we deduce, for any polynomial p :

$$\lim n^r(f - \mathcal{M}_n^{(2r)} f) = \sum_{i+j=2r} \bar{\kappa}_{i,j} D^{i,j} f, \quad q = 2r-1$$

Similarly, for $q = 2r$, we get

$$\lim n^{r+1}(f - \mathcal{M}_n^{(2r+1)} f) = \sum_{p=2r+1}^{2r+2} \sum_{i+j=p} \bar{\kappa}_{i,j} D^{i,j} f, \quad q = 2r$$

Examples.

$$\lim n(f - \mathcal{M}_n^{(2)} f) = -XD^{2,0}f + 2xyD^{1,1}f - YD^{0,2}f$$

$$\lim n^2(f - \mathcal{M}_n^{(3)} f) = \sum_{|\gamma|=3} \bar{\alpha}_\gamma D^\gamma f + \sum_{|\gamma|=4} \bar{\alpha}_\gamma D^\gamma f$$

$$= -X(5x-2)D^{3,0}f - x(15xy-x-8y+1)D^{2,1}f - yx(15xy-8x-y+1)D^{1,2}f$$

$$-(5y-2)YD^{0,3}f+\frac{1}{2}X^2D^{4,0}f-2xyXD^{3,1}f$$

$$+xy(3xy-(x+y)+1)D^{2,2}f-2xyYD^{1,3}f+\frac{1}{2}Y^2D^{0,4}f$$

8.3. Genuine Bernstein-Durrmeyer QIs

For and polynomials, we define

$$\bar{\theta}_{k,\ell}=\lim n^r\theta_{k,\ell}\quad\text{for}\quad k+\ell=2r-1\text{ or }2r$$

$$\bar{\eta}_{k,\ell}=\lim n^r\eta_{k,\ell}\quad\text{for}\quad k+\ell=2r-1\text{ or }2r$$

Here is a table of the first polynomials:

(k,ℓ)	$\bar{\eta}_{k,\ell}$
$(2,0)$	$-X$
$(1,1)$	$2xy$
$(3,0)$	$(1-2x)X$
$(2,1)$	$-3xy(1-2x)$
$(4,0)$	$X^2/2$
$(3,1)$	$-2xyX$
$(2,2)$	$xy(3xy-(x+y)+1)$

9. Numerical experiments on Bernstein quasi-interpolants

We present some numerical tests on the following functions

$$f_1(x,y)=\frac{1}{1+16((x-1/3)^2+(y-1/3)^2)}$$

$$f_2(x,y)=\exp(-x^2-y^2)$$

using classical and genuine Bernstein quasi-interpolants of various degrees and orders.

We denote the uniform errors respectively by $eb_n^{(r)}f:=\|f-\mathcal{B}_n^{(r)}f\|$ for Bernstein QIs and by $eg_n^{(r)}:=\|f-\mathcal{G}_n^{(r)}f\|$ for genuine Bernstein-Durrmeyer QIs.

(n,r)	$eb_n^{(r)}f_1$	$eb_n^{(r)}f_2$	(n,r)	$eg_n^{(r)}f_1$	$eg_n^{(r)}f_2$
$(8,0)$	0.38	3.6(-2)	$(5,1)$	0.6	$8.8(-2)$
$(8,3)$	8.4(-2)	2.3(-3)	$(5,3)$	0.3	$8.8(-3)$
$(8,5)$	2.4(-2)	1.2(-4)	$(5,4)$	0.25	$1.2(-3)$
$(8,8)$	0.12	2.0(-6)	(5,5)	0.14	4.8(-4)
$(15,0)$	0.26	2.0(-2)	(10,0)	0.46	5.2(-2)
$(15,4)$	4.6(-2)	4.4(-5)	(10,2)	0.25	5.2(-3)
$(15,8)$	1.2(-2)	6.0(-8)	(10,4)	0.15	4.0(-4)
$(15,9)$	5.6(-3)	3.0(-8)	(10,6)	8.4(-2)	4.8(-5)
$(15,10)$	9.2(-3)	3.4(-9)	(10,7)	0.12	2.6(-4)
$(15,15)$	1.5(-2)	5.0(-11)	(10,10)		

We see that the behaviours of QIs are quite different for f_1 and f_2.
1) f_1 is a rational function of Runge type : the Lagrange interpolants for $n = 8$ and $n = 15$ both give bad results. However, the errors $eb_n^{(r)} f_1$ seem to have a minimum value for some intermediate QIs, for example for $(n, r) = (8, 5)$ and $(n, r) = (15, 9)$. A similar fact occurs for the errors $eg_n^{(r)} f_1$ where the minimum value is obtained for $(n, r) = (10, 6)$. However the errors are higher than those obtained by Bernstein QIs for $n = 8$.
2) f_1 is a good analytic function with a nice behaviour: the Lagrange interpolant gives the best results. The errors slowly decrease from $r = 0$ to $r = n$. If one does not want a very high precision, the first QIs can be taken as approximants of the given function. For the genuine Durrmeyer operator, the errors for $n = 10$ are higher than those obtained by Bernstein QIs for $n = 8$, except maybe the minimum value for $(n, r) = (10, 6)$.
We also compared the above results with those obtained using the BD operator with Legendre weight (the errors are denoted $ed_n^{(r)} f$). For the two tested functions, the results were worse. We only give them for the exponential functionf_2.

(n, r)	$eb_n^{(r)} f_2$	$eg_n^{(r)} f_2$	$ed_n^{(r)} f_2$
$(5, 0)$	$5.6(-2)$	$8.8(-2)$	0.18
$(5, 3)$	$4.6(-3)$	$8.8(-3)$	4.2(-3)
$(5, 4)$	$6.4(-4)$	$1.2(-3)$	2.3(-3)
(5,5)	6.4(-4)	8.8(-4)	1.6(-3)

As a conclusion of these tests (and of other tests done on various functions), the classical Bernstein QIs seem a priori to be the more efficient. Of course, the values of f on uniform lattices of points of the triangle must be available. If the function is only known by its moments or other mean integral values, then one could consider the approximation by BDQIs with convenient Jacobi weights or by GDQIs.

10. Some applications

In this final section, we briefly present some possible applications of the above quasi-interpolants to various problems in approximation, CAGD and numerical analysis.

- in approximation, the Hausdorff moment problem in T consists in finding a function f having given moments $\mu_\gamma(f) := \int_T f(x, y)x^k y^\ell dxdy$ for some indices $\gamma = (k, \ell) \in \mathbb{N}^2$. Such a function can be approximated by the Bernstein-Durrmeyer quasi-interpolants of Section 5. Indeed, scalar products $\langle f, B_\alpha^n \rangle$ are directly computable from moments, so $\mathcal{M}_n f$ is easily obtained together with its partial derivatives.
- in CAGD, when one is interested in approximating a function defined on a uniform lattice of points in the triangle T, Bernstein quasi-interpolants of Section 3 can sometimes offer an alternative to strict interpolation at

those points since their norms seem to be uniformly bounded in n for a given order r.

- in numerical analysis, it would be perhaps interesting to derive cubature formulas from integration of Bernstein quasi-interpolants. In the same way, approximate formulas for partial derivatives can be obtained by computing derivatives of Bernstein or Bernstein-Durrmeyer type quasi-interpolants.

References

[1] Abel, U., *Asymptotic approximation by Bernstein-Durrmeyer operators and their derivatives*, Approx. Theory & Appl., **16**(2000), no. 2, 1-12.

[2] Abel, U., Ivan, M., *Asymptotic expansion of the multivariate Bernstein polynomials on a simplex*, Approx. Theory & Appl., **16**(2000), no. 3, 85-93.

[3] Berdysheva, E., Jetter, K., *Multivariate Bernstein-Durrmeyer operators with arbitrary weight functions*, J. Approx. Theory, **162**(2010), 576-598.

[4] Berdysheva, E., Jetter, K., *Multivariate Bernstein-Durrmeyer operators with a positive measure*, MAIA Meeting, Edinburgh, 2010.

[5] Berdysheva, E., Jetter, K., Stöckler, J., *New polynomial preserving operators on simplices: Direct results*, J. Approx. Theory, **131**(2004), 59-73.

[6] Berdysheva, E., Jetter, K., Stöckler, J., *Durrmeyer operators and their natural quasi-interpolants*, In: Topics in Multivariate Approximation and Interpolation, K. Jetter et al. (eds.), Elsevier, Amsterdam, 2006, 1-21.

[7] Berens, H., Schmid, H.J., Xu, Y., *Bernstein-Durrmeyer polynomials on a simplex*, J. Approx. Theory, **68**(1992), 247-261.

[8] Berens, H., Xu, Y., *On Bernstein-Durrmeyer polynomials with Jacobi weights*, In: Approximation Theory and Functional Analysis, C.K. Chui (ed.), Academic Press, Boston, 1991, 25-46.

[9] Bos, L., *Bounding the Lebesgue function for Lagrange interpolation on the simplex*, J. Approx. Theory, **38**(1983), 43-59.

[10] Chen, W., Ditzian, Z., Ivanov, K., *Strong converse inequalities for the Bernstein-Durrmeyer operator*, J. Approx. Theory, **75**(1993), 25-43.

[11] Ciarlet, P.G., *The finite element method for elliptic problems*, North-Holland, Amsterdam 1978, reprinted by SIAM, 2002.

[12] Ciesielski, Z., *Biorthogonal systems of polynomials on the standard simplex*, International Series in Numerical Mathematics, Birkhäuser Verlag, Basel, **75**(1985), 116-119.

[13] Derriennic, M.M., *On multivariate approximation by Bernstein-type polynomials*, J. Approx. Theory, **45**(1985), 155-166.

[14] Derriennic, M.M., *Polynômes de Bernstein modifiés sur un simplexe de $\mathbb{R}^\ell$. Problème des moments*, In: Polynômes orthogonaux et Applications, C. Brézinski et al. (eds.), Lecture Notes in Mathematics, Springer Verlag, Berlin **1171**(1985), 296-301.

[15] Derriennic, M.M., *Linear combinations of derivatives of Bernstein type polynomials on a simplex*, In: Coll. Soc. János Bolyai, **58**(1990), 197-220.

[16] Ditzian, Z., *Modified multivariate Bernstein operators*, CMSJB 19, Fourier Analysis and Approximation Theory Budapest, 1976, 291-305.

[17] Ditzian, Z., *A note on Durrmeyer-Bernstein operators in $L^2(S)$*, J. Approx. Theory, **72**(1993), 234-236.

[18] Ditzian, Z., *Multidimensional Jacobi-type Bernstein-Durrmeyer operators*, Acta Sci. Math., Szeged, **60**(1995), 225-243.

[19] Dai, F., Huang, H., Wang, K., *Approximation by the Bernstein-Durrmeyer operator on a simplex*, Constructive Approximation, **31**(2010), 289-308.

[20] Ding, C., Cao, F., *K-functionals and Multivariate Bernstein polynomials*, J. Approx. Theory, **155**(2008), 125-135.

[21] Dunkl, C., Xu, Y., *Orthogonal polynomials of several variables*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, **81**(2001).

[22] Farouki, R., Goodman, T.N.T., Sauer, T., *Construction of orthogonal bases for polynomials in Bernstein form on triangular and simplex domains*, Computer Aided Geometric Design, **20**(2003), no. 4, 209-230.

[23] Gavrea, I., Mache, D.H., *Generalization of Bernstein-type approximation methods*, In: Approximation Theory, IDoMAT 95, M.W. Müller et al. (eds.), Mathematical Research, **86**, Akademie Verlag, Berlin, 1995, 115-126.

[24] Gonska, H., Kacsó, D., Raşa, I., *On Genuine Bernstein-Durrmeyer operators*, Results in Mathematics, **50**(2007), 213-225.

[25] Gonska, H., Maier, J., *A bibliography on approximations of functions by BB type operators*, In: Approximation Theory IV, Chui, Schumaker & Ward (eds.), Academic Press, New-York, 1983, 739-785.

[26] Gonska, H., Maier-Gonska, J., *A bibliography on approximations of functions by BB type operators (supplement)*, In: Approximation Theory V, Chui, Schumaker & Ward (eds.), Academic Press, New-York, 1986, 621-653.

[27] Goodman, T., Sharma, A., *A modified Bernstein-Schoenberg operator*, In: Proc. Conf. on Constructive Theory of Functions, Varna 1987, B. Sendov et al. (eds.), Publ. House Bulg. Acad. Sci., Sofia, 1988, 166-173.

[28] Horst, J., *Multivariate Quasi-Interpolation und gewichtete Polynomapproximation*, Dissertation, Universität Hagen, 2007.

[29] Jetter, K., Stöckler, J., *An identity for multivariate Bernstein polynomials*, Computer Aided Geometric Design, **20**(2003), 563-577.

[30] Kageyama, Y., *Generalization of the left Bernstein quasi-interpolants*, J. Approx. Theory, **94**(1998), no. 2, 306-329.

[31] Kageyama, Y., *A new class of modified Bernstein operators*, J. Approx. Theory, **101**(1999), no. 1, 121-147.

[32] Lai, M.J., *On dual functionals of polynomials in Bernstein form*, J. Approx. Theory, **67**(1991), no. 1, 19-37.

[33] Lai, M.J., *Asymptotic formulae of multivariate Bernstein approximants*, J. Approx. Theory, **70**(1992), no. 2, 229-242.

[34] López-Moreno, A.J., Martínez-Moreno, J., Muñoz-Delgado, F.J., *Asymptotic expressions for multivariate positive linear operators*, in: Approximation Theory X, C.K. Chui et al. (eds.), Vanderbilt University Press, Nashville, 2002, 287-307.

[35] Lorentz, G.G., *List of publications*, J. Approx. Theory, **13**(1975), 8-11.

[36] Lorentz, G.G., *Approximation of functions*, Reprint AMS 2005.

[37] Mache, D.H., *A link between Bernstein polynomials and Durrmeyer polynomials with Jacobi weights*, in: Approximation Theory VIII, Vol. I, C.K. Chui & L.L. Schumaker (eds.), World Scientific Publishers, 1995, 403-410.

[38] Mache, P., Mache, D.H., *Approximation by Bernstein quasi-interpolants*, Numer Funct. Anal. & Optimiz., 22(2001), no. 1-2, 159-171.

[39] Parvanov, P.E., Popov, B.D., *The limit case of Bernstein's operator with Jacobi weights*, Math. Balkanika, **8**(1994), 165-177.

[40] Sablonnière, P., *Bernstein quasi-interpolants on* $[0,1]$, In: Multivariate Approximation IV, C. K. Chui et al. (eds.), International Series in Numerical Mathematics, Birkhäuser Verlag, Basel **90**(1989), 287-294.

[41] Sablonnière, P., *Bernstein quasi-interpolants on a simplex*, Oberwolfach (July 1989), Publ. LANS No 21, Insa de Rennes, Septembre 1989 (unpublished).

[42] Sablonnière, P., *A family of Bernstein quasi-interpolants on* $[0,1]$, Approx. Theory & Appl., **8**(1992), no. 3, 62-76.

[43] Sablonnière, P., *Discrete Bernstein bases and Hahn polynomials*, J. Comput. Appl. Math., **49**(1993), 233-241.

[44] Sablonnière, P., *Bernstein-type quasi-interpolants*, In: Curves and Surfaces, P. J. Laurent et al. (eds.), A. K. Peters, 1991, 421-426.

[45] Sablonnière, P., *Representation of quasi-interpolants as differential operators and applications*, in: New Developments in Approximation Theory, M. W. Müller et al. (eds.), International Series in Numerical Mathematics, Birkhäuser Verlag, Basel, **132**(1999), 233-253.

[46] Sablonnière, P., *Recent progress on univariate and multivariate polynomial and spline quasi interpolants*, in: Trends and Applications in Constructive Approximation, M.G. de Bruin et al. (eds.), International Series in Numerical Mathematics, Birkhäuser Verlag, Basel **151**(2005), 229-245.

[47] Sauer, T., *The genuine Bernstein-Durrmeyer operator on a simplex*, Results in Mathematics, **26**(1994), 99-130.

[48] Waldron, S., *On the Bernstein-Bézier form of Jacobi polynomials on a simplex*, J. Approx. Theory, **140**(2006), 86-99.

[49] Walz, G., *Asymptotics for multivariate Bernstein operators*, J. Comput. Appl. Math., **122**(2000), 317-328.

[50] Zhengchang, W., *Norm of the Bernstein left quasi-interpolant operator*, J. Approx. Theory, **66**(1991), 36-43.

[51] Zhou, X.L., Sun, J., *Dual bases of multivariate Bernstein-Bézier polynomials*, Computer Aided Geometric Design, **5**(1988), 119-125.

[52] Zhou, X.L., *Approximation by multivariate Bernstein operators*, Results in Mathematics, **25**(1994), 166-191.

Paul Sablonnière
INSA de Rennes
20 avenue des Buttes de Coesmes, CS 70839
35708 Rennes Cedex 7
France
e-mail: `psablonn@insa-rennes.fr`

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 587–598

Approximation of the solution of stochastic differential equations driven by multifractional Brownian motion

Anna Soós

Abstract. The aim of this paper is to approximate the solution of a stochastic differential equations

$$dX(t) = F(X(t))dt + G(X(t))dB(t),\ X(0) = X_0,\ \ t \geq 0$$

on $\mathbb{R}^n$. We will use wavelet approximation of multifractional Brownian motion.

Mathematics Subject Classification (2010): Primary 60H10; Secondary: 60H05, 60J65.

Keywords: Stochastic differential equation, fractional Brownian motion.

1. Introduction

The fractional Brownian motion (fBm) with Hurst index $H \in (0, 1)$ is a zero mean Gaussian random process $\Big(B(t)\Big)_{t\geq 0}$ with continuous sample paths and with covariance function

$$E\Big(B(s)B(t)\Big) = \frac{1}{2}\Big(t^{2H} + s^{2H} - |s-t|^{2H}\Big).$$

For $H = \frac{1}{2}$ the fractional Brownian motion is the ordinary standard Brownian motion.

The fractional Brownian motion B has on any finite interval $[0, T]$ Hölder continuous paths with exponent $\gamma \in (0, H)$ (see [5]). Moreover, the quadratic variation on $[a, b] \subseteq [0, T]$ is

$$\lim_{|\Delta_n|\to 0} \sum_{i=1}^{n} \Big(B(t_i^n) - B(t_{i-1}^n)\Big)^2 = \begin{cases} \infty & \text{if } H < \frac{1}{2}, \\ b-a & \text{if } H = \frac{1}{2}, \\ 0 & \text{if } H > \frac{1}{2}, \end{cases} \tag{1.1}$$

where $\Delta_n = (a = t_0^n < \cdots < t_n^n = b)$ is a partition of $[a, b]$ with

$$|\Delta_n| = \max_{1\leq i\leq n} (t_i^n - t_{i-1}^n).$$

If $H \neq \frac{1}{2}$, then the convergence in (1.1) holds with probability 1 uniformly in the set of all partitions of $[a, b]$, while for $H = \frac{1}{2}$ the convergence in (1.1) holds in mean square uniformly in the set of all partitions of $[a, b]$. Note that, if $H \neq \frac{1}{2}$, then B is not a semimartingale, so the classical stochastic integration does not work. But the Hölder continuity of B will ensure the existence of integrals

$$\int_0^T G(u) dB(u),$$

defined in terms of fractional integration as investigated in [15] and [16] for the stochasticc process $(G(t))_{t \in [0,T]}$ with Hölder continuous paths of order $\alpha > 1 - H$. Moreover, the fractional Brownian motion is H-self similar, so for any $c > 0$ the process $\left(c^H B(t/c)\right)_{t \geq 0}$ is again a fractional Brownian motion, has stationary increments. Stochastic differential equations driven by fBm have received considerable attention during the last two decades. Fractional Brownian motion as driving noise is used in electrical engineering ([6]) or biophysics ([11]). Moreover, fBm has established itself also in financial modelling ([4],[8]).

The multifractional Brownian motion (mfBm) is obtained by replacing the constant parameter H of the fractional Brownian motion by a smooth enough functional parameter $H(\cdot)$. We denote by H a function defined on the real line and with values in a fixed interval $[a, b] \subset (0, 1)$. We assume that it is uniformly Hölder continuous of order $\beta > b$ on each compact subset of $\mathbb{R}$.

In this article we study the approximation of the Itô stochastic differential equation

$$dX(t) = F(X(t))dt + G(X(t))dB(t), \; X(0) = X_0, \;\; t \geq 0 \tag{1.2}$$

on $\mathbb{R}^n$. Here $F : \mathbb{R}^n \to \mathbb{R}^n$, $G : \mathbb{R}^n \to \mathbb{R}^n$, $B = \left(B(t)\right)_{t \geq 0}$, $H \in (0, 1)$ is a 1-dimensional multifractional Brownian motion adapted to a filtration $F = (F_t)_{t \geq 0}$ on a probability space (Ω, K, P), and x_0 is a F_0 measurable random variable independent of B.

Suppose with F and G satisfy with probability 1:
$F \in C(\mathbb{R}^n \times [0, T], \mathbb{R}^n), G \in C^1(\mathbb{R}^n \times [0, T], \mathbb{R}^n)$ and
$F(\cdot, t), \dfrac{\partial G(\cdot, t)}{\partial x}, \dfrac{\partial G(\cdot, t)}{\partial t}$ are locally Lipschitz, $\forall t \in [0, T]$.

2. Wavelet approximation for $(B(t))_{t \in [0,1]}$

Let $\{2^{j/2} \Psi(2^j x - k) : (j, k) \in \mathbb{Z}^2\}$ be a Lamarie Meyer wavelet basis of $L^2(\mathbb{R})$ and denote by $\boldsymbol{\Psi}$ the function defined by

$$\boldsymbol{\Psi}(x, \theta) = \int_{\mathbb{R}} e^{ixy} \frac{\overline{\boldsymbol{\Psi}}(y)}{|y|^{\theta + \frac{1}{2}}} dy,$$

where $\overline{\mathbf{\Psi}(y)}$ is the Fourier transform. We use the following wavelet approximation of the multifractional Brownian motion $(B(t))_{t\in[0,1]}$ with Hurst index H investigated in [1].

$$B(t) = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} 2^{-jH(k/2^j)}(\mathbf{\Psi}(2^j t-k, H(k/2^j)) - \mathbf{\Psi}(-k, H(k/2^j)))\epsilon_{j,k}, \tag{2.1}$$

where $\epsilon_{j,k}$ are independent identically distributed $N(0,1)$ random variables. This process was introduced in [3] to model fBm with piecewise constant Hurst index and continuous path.

As in [2] and [12] we consider the following assumptions for Ψ: $\Psi \in C^1$ and there exists a constant $c > 0$ such that

$$|\sup_{\theta\in[a,b]} \mathbf{\Psi}(t,\theta)| \le \frac{c}{(2+|t|)^2} \text{ and } |\sup_{\theta\in[a,b]} \mathbf{\Psi}'(t,\theta)| \le \frac{c}{(2+|t|)^3} \text{ for all } t\in\mathbb{R}. \tag{2.2}$$

We consider the following *high frequency component* of the wavelet representation in (2.1)

$$V_1(t) = \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} 2^{-jH}(\mathbf{\Psi}(2^j t-k, H(k/2^j)) - \mathbf{\Psi}(-k, H(k/2^j)))\epsilon_{j,k}$$

and the *low frequency component*

$$V_2(t) = \sum_{j=-\infty}^{-1}\sum_{k=-\infty}^{\infty} 2^{-jH}(\mathbf{\Psi}(2^j t-k, H(k/2^j)) - \mathbf{\Psi}(-k, H(k/2^j)))\epsilon_{j,k}.$$

Obviously,

$$B(t) = V_1(t) + V_2(t) \text{ for each } t\in[0,1].$$

Let $N\in\mathbb{N}$. In the following we use two approximation components, corresponding to the components V_1, respectively V_2, namely

$$B_1^N(t) = \sum_{j=0}^{N}\sum_{|k|\le\frac{2^{N+4}}{(N-j+1)^2}} 2^{-jH}(\mathbf{\Psi}(2^j t-k, H(k/2^j)) - \mathbf{\Psi}(-k, H(k/2^j)))\epsilon_{j,k}$$

and

$$B_2^N(t) = \sum_{j=-2^{[N/2]}}^{-1}\sum_{|k|\le 2^{[N/2]}} 2^{-jH}(\mathbf{\Psi}(2^j t-k, H(k/2^j)) - \mathbf{\Psi}(-k, H(k/2^j)))\epsilon_{j,k}.$$

We denote

$$B^N(t) = B_1^N(t) + B_2^N(t) \text{ for each } t\in[0,1]. \tag{2.3}$$

Using Theorem 2 and Theorem 3 from [2] we have the following result:

Theorem 2.1. *The sequence $(B^N)_{N\in\mathbb{N}}$ converges to B almost surely in $\omega\in\Omega$ and uniformly in $t\in[0,1]$, i.e.*

$$\mathbb{P}\Big(\lim_{N\to\infty}\sup_{t\in[0,1]}|B^N(t)-B(t)| = 0\Big) = 1.$$

In the sequel we need the following result:

Theorem 2.2. *For all $N \in \mathbb{N}$ the approximating processes $(B^N(t))_{t\in[0,1]}$ are Lipschitz continuous with probability* 1.

Proof. We write

$$|B^N(s) - B^N(t)| \leq |B_1^N(s) - B_1^N(t)| + |B_2^N(s) - B_2^N(t)|$$

$$\leq \sum_{j=0}^{N} \sum_{|k|\leq \frac{2^{N+4}}{(N-j+1)^2}} 2^{-jH} |\mathbf{\Psi}(2^j s - k, H(k/2^j)) - \mathbf{\Psi}(2^j t - k, H(k/2^j)))||\epsilon_{j,k}|$$

$$+ \sum_{j=-2^{[N/2]}}^{-1} \sum_{|k|\leq 2^{[N/2]}} 2^{-jH} |\mathbf{\Psi}(2^j s - k, H(k/2^j)) - \mathbf{\Psi}(2^j t - k, H(k/2^j))||\epsilon_{j,k}|.$$

Using the assumption (2.2) for $\mathbf{\Psi}$ and using that the set of indices of j and k is bounded, it follows that there exists a $c_N > 0$ (depending on ω) such that

$$|B^N(s) - B^N(t)| \leq c_N |s - t| \text{ for all } s, t \in [0, 1] \text{ and all } n \in \mathbb{N}.$$

□

3. Fractional integrals and derivatives

Let $a, b \in \mathbb{R}$, $a < b$ and $f, g : \mathbb{R} \to \mathbb{R}$. We use notions and results about fractional calculus, from [14] and [15]:

$$f(a+) := \lim_{\delta\searrow 0} f(a + \delta), \quad f(b-) := \lim_{\delta\searrow 0} f(b - \delta),$$

$$f_{a+}(x) = \mathbb{I}_{(a,b)}(f(x) - f(a+)), \quad g_{b-}(x) = \mathbb{I}_{(a,b)}(g(x) - g(b-)).$$

Note that for $\alpha > 0$ we have $(-1)^\alpha = e^{i\pi\alpha}$.

For $f \in L_1(a, b)$ and $\alpha > 0$ the **left-** and **right-sided fractional Rieman-Liouville integral of** f **of order** α on (a, b) is given for almost every x by

$$I_{a+}^\alpha f(x) = \frac{1}{\Gamma(\alpha)} \int_a^x (x - y)^{\alpha-1} f(y) dy$$

and

$$I_{b-}^\alpha f(x) = \frac{(-1)^{-\alpha}}{\Gamma(\alpha)} \int_x^b (y - x)^{\alpha-1} f(y) dy.$$

For $p > 1$ let $I_{a+}^\alpha(L_p(a, b))$, be the class of functions f which have the representation $f = I_{a+}^\alpha \Phi$, where $\Phi \in L_p(a, b)$, and let $I_{b-}^\alpha(L_p(a, b))$ be the class of functions g which have the representation $g = I_{b-}^\alpha \varphi$, where $\varphi \in L_p(a, b)$. If $0 < \alpha < 1$, then the function Φ, respectively φ, in the representations above agree almost surely with the **left-sided** and respectively

right-sided fractional derivative of f **of order** α (in the Weyl representation)

$$\Phi(x) = D^{\alpha}_{a+}f(x) = \frac{1}{\Gamma(1-\alpha)}\left(\frac{f(x)}{(x-a)^{\alpha}} + \alpha\int\limits_a^x \frac{f(x)-f(y)}{(x-y)^{\alpha+1}}dy\right)\mathbb{I}_{(a,b)}(x)$$

and

$$\varphi(x) = D^{\alpha}_{b-}g(x) = \frac{(-1)^{\alpha}}{\Gamma(1-\alpha)}\left(\frac{g(x)}{(b-x)^{\alpha}} + \alpha\int\limits_x^b \frac{g(x)-g(y)}{(y-x)^{\alpha+1}}dy\right)\mathbb{I}_{(a,b)}(x).$$

The convergence at the singularity $y = x$ holds in the L_p-sense. Recall that $I^{\alpha}_{a+}(D^{\alpha}_{a+}f) = f$ for $f \in I^{\alpha}_{a+}(L_p(a,b))$, $\quad I^{\alpha}_{b-}(D^{\alpha}_{b-}g) = g$ for $g \in I^{\alpha}_{b-}(L_p(a,b))$ and

$$D^{\alpha}_{a+}(I^{\alpha}_{a+}f) = f, \quad D^{\alpha}_{b-}(I^{\alpha}_{b-}g) = g \text{ for } f, g \in L_1(a,b).$$

For completeness we denote

$$D^0_{a+}f(x) = f(x), D^0_{b-}g(x) = g(x), D^1_{a+}f(x) = f'(x), D^1_{b-}g(x) = g'(x).$$

Let $0 \le \alpha \le 1$. The **fractional integral** of f with respect to g is defined as

$$\begin{aligned}\int\limits_a^b f(x)dg(x) &= (-1)^{\alpha}\int\limits_a^b D^{\alpha}_{a+}f_{a+}(x)D^{1-\alpha}_{b-}g_{b-}(x)dx \\ &\quad + f(a+)(g(b-)-g(a+))\end{aligned} \tag{3.1}$$

if $f_{a+} \in I^{\alpha}_{a+}(L_p(a,b)), g_{b-} \in I^{1-\alpha}_{b-}(L_q(a,b))$ for $\frac{1}{p} + \frac{1}{q} \le 1$.

In our investigations we will take $p = q = 2$. If $0 \le \alpha < \frac{1}{2}$, then the integral in (3.1) can be written as

$$\int\limits_a^b f(x)dg(x) = (-1)^{\alpha}\int\limits_a^b D^{\alpha}_{a+}f(x)D^{1-\alpha}_{b-}g_{b-}(x)dx \tag{3.2}$$

if $f \in I^{\alpha}_{a+}(L_2(a,b))$, $f(a+)$ exists, $g_{b-} \in I^{1-\alpha}_{b-}(L_2(a,b))$ (see [15]).

4. The stochastic integral

Without loss of generality we consider $0 < T \le 1$, because for arbitrary $T > 0$ we can rescale the time variable using the H-self similarity property of the multifractional Brownian motion meaning that $\Big(B(ct)\Big)_{t\ge 0}$ and $\Big(c^H B(t)\Big)_{t\ge 0}$ are equal in distribution for every $c > 0$.

We will define the $\int\limits_0^T G(u)dB(u)$ Itô integral instead of $\int\limits_0^t G(u)dB(u)$ and use

$$\int\limits_0^t G(u)dB(u) = \int\limits_0^T \mathbb{I}_{[0,t]}(u)G(u)dB(u) \text{ for } t \in [0,T]$$

(by Theorem 2.5, p. 345, in [15]).

We consider $\alpha > 1 - H$. It follows by (3.2) that

$$\int_0^T G(u)dB(u) = (-1)^\alpha \int_0^T D_{0+}^\alpha G(u) D_{T-}^{1-\alpha} B_{T-}(u)du \tag{4.1}$$

for $G \in I_{0+}^\alpha(L_2(0,T))$, where $G(0+)$ exists and $B_{T-} \in I_{T-}^{1-\alpha}(L_2(0,T))$.

The condition $G \in I_{0+}^\alpha(L_2(0,T))$ (with probability 1) means that $G \in L_2(0,T)$ and

$$\mathcal{I}_\varepsilon(x) = \int_0^{x-\varepsilon} \frac{G(x) - G(y)}{(x-y)^{\alpha+1}} dy \text{ for } x \in (0,T)$$

converges in $L^2(0,T)$ as $\varepsilon \searrow 0$.

The condition $B_{T-} \in I_{T-}^{1-\alpha}(L_2(0,T))$ means $B_{T-} \in L_2(0,T)$ and

$$\mathcal{J}_\varepsilon(x) = \int_{x+\varepsilon}^{T} \frac{B(x) - B(y)}{(y-x)^{2-\alpha}} dy \text{ for } x \in (0,T)$$

converges in $L_2(0,T)$ as $\varepsilon \searrow 0$ This condition for B is fulfilled for $\alpha > 1 - H$, since the multifractional Brownian motion B is almost surely Hölder continuous with exponent $\gamma \in (0, H)$ (see [5]).

We will use (3.2) for the integrals with respect the approximating processes $\Big(B_N(t)\Big)_{t\in[0,T]}$. Observe that $B_{N,T-} \in I_{T-}^{1-\alpha}(L_2(0,T))$, which follows from the Lipschitz continuity property in Theorem 2.2. We have

$$\int_0^T G(u)dB_N(u) = (-1)^\alpha \int_0^T D_{0+}^\alpha G(u) D_{T-}^{1-\alpha} B_{N,T-}(u)du \tag{4.2}$$

for $G \in I_{0+}^\alpha(L_2(0,T))$, where $G(0+)$ exists.

Let $\Big(Z(t)\Big)_{t\in[0,T]}$ be a cádlág process. Its **generalized quadratic variation process** $\Big([Z](t)\Big)_{t\in[0,T]}$ is defined as

$$[Z](t) = \lim_{\varepsilon \searrow 0} \varepsilon \int_0^1 \int_0^t \frac{1}{u} (Z_{t-}(s+u) - Z_{t-}(s))^2 ds du + (Z(t) - Z(t-))^2,$$

if the limit exists uniformly in probability (see [16]).

In particular, if B is a multifractional Brownian motion with Hurst index $H \in (\frac{1}{2}, 1)$ and B^N is an approximation of B as given in (2.3), it is easy to verify that

$$[B](t) = 0 \quad \text{and} \quad [B^N](t) = 0 \quad \text{for each } \ t \in [0,T], \tag{4.3}$$

because B is locally Hölder continuous and B^N is Lipschitz continuous. The **Itô formula** for change of variable for fractional integrals is given in the next theorem.

Theorem 4.1 ([16], Theorem 5.8, p. 170). *Let* $\Big(Z(t)\Big)_{t\in[0,T]}$ *be a continuous process with generalized quadratic variation* $[Z]$. *Let* $Q : \mathbb{R}\times[0,T]\to\mathbb{R}$ *be a random function such that a.s. we have* $Q \in \mathcal{C}^1(\mathbb{R}\times[0,T])$ *and* $\frac{\partial^2 Q}{\partial x^2} \in \mathcal{C}(\mathbb{R}\times[0,T])$. *Then, for* $t_0, t \in [0,T]$ *we have*

$$\begin{aligned} Q(Z(t),t) - Q(Z(t_0),t_0) &= \int_{t_0}^{t} \frac{\partial Q}{\partial x}(Z(s),s)dZ(s) + \int_{t_0}^{t} \frac{\partial Q}{\partial t}(Z(s),s)ds \\ &\quad + \int_{t_0}^{t} \frac{\partial^2 Q}{\partial^2 x}(Z(s),s)d[Z]s. \end{aligned}$$

Let $1-H<\alpha<\frac{1}{2}$ and let $G \in I^{\alpha}_{0+}(L_2(0,T))$ such that $G(0+)$ exists. We define the processes

$$Z(t) = \int_0^t G(s)dB(s) \text{ and } Z_N(t) = \int_0^t G(s)dB_N(s), \quad t\in(0,T].$$

Then by Theorem 5.6, p. 167 in [16] it follows that

$$[Z](t) = 0 \text{ and } [Z_N](t) = 0.$$

Using Theorem 4.1, it follows that, if $Q : \mathbb{R}\times[0,T]\to\mathbb{R}$ is a random function such that a.s. we have $Q\in\mathcal{C}^1(\mathbb{R}\times[0,T])$ and $\frac{\partial^2 Q}{\partial x^2}\in\mathcal{C}(\mathbb{R}\times[0,T])$, then for $t_0, t\in[0,T]$ we have

$$\begin{aligned} Q(Z(t),t) - Q(Z(t_0),t_0) &= \int_{t_0}^{t} \frac{\partial Q}{\partial x}(Z(s),s)G(s)dB(s) \\ &\quad + \int_{t_0}^{t} \frac{\partial Q}{\partial t}(Z(s),s)ds \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} Q(Z_N(t),t) - Q(Z_N(t_0),t_0) &= \int_{t_0}^{t} \frac{\partial Q}{\partial x}(Z_N(s),s)G(s)dB_N(s) \\ &\quad + \int_{t_0}^{t} \frac{\partial Q}{\partial t}(Z_N(s),s)ds. \end{aligned} \tag{4.5}$$

5. Stochastic differential equations driven by multifractional Brownian motion

Let $\Big(B(t)\Big)_{t\geq 0}$ be a multifractional Brownian motion with Hurst parameter H such that $H>\frac{1}{2}$. We investigate stochastic differential equations of the form

$$\begin{aligned} dX(t) &= F(X(t),t)dt + G(X(t),t)dB(t), \\ X(t_0) &= X_0, \end{aligned} \tag{5.1}$$

where $t_0 \in (0,T]$, X_0 is a random vector in $\mathbb{R}^n$ and the random functions F and G satisfy with probability 1 the following conditions:

(C1) $F \in C(\mathbb{R}^n \times [0,T], \mathbb{R}^n), G \in C^1(\mathbb{R}^n \times [0,T], \mathbb{R}^n)$;

(C2) for each $t \in [0,T]$ the functions $F(\cdot,t), \dfrac{\partial G(\cdot,t)}{\partial x^i}, \dfrac{\partial G(\cdot,t)}{\partial t}$ are locally Lipschitz for each $i \in \{1,\dots,n\}$.

We consider the pathwise auxiliary partial differential equation on $\mathbb{R}^n \times \mathbb{R} \times [0,T]$

$$\begin{aligned} \frac{\partial K}{\partial z}(y,z,t) &= G(K(y,z,t),t), \\ K(Y_0,Z_0,t_0) &= X_0, \end{aligned} \tag{5.2}$$

where Y_0 is an arbitrary random vector in $\mathbb{R}^n$ and Z_0 an arbitrary random variable in $\mathbb{R}$. From the theory of differential equations it follows that with probability 1 there exists a local solution $K \in C^1(\mathbb{R}^n \times [0,T], \mathbb{R}^n)$ in a neighbourhood V of (Y_0, Z_0, t_0) with partial derivatives being Lipschitz in the variable y and

$$\det\left(\frac{K^i}{\partial y^j}(y,z,t)\right)_{1 \le i,j \le n} \neq 0.$$

We have for $(x,y,t) \in V$

$$\frac{\partial^2 K}{\partial z^2}(y,z,t) = \sum_{j=1}^{n} \frac{\partial G}{\partial x^j}(K(y,z,t),t)G^j(K(y,z,t),t).$$

We also consider the pathwise differential equation (in matrix representation) on $[0,T]$

$$\begin{aligned} dY(t) &= \left(\frac{K}{\partial y}(Y(t),B(t),t)\right)^{-1}\Big[F(K(Y(t),B(t),t),t) - \frac{\partial K}{\partial t}(Y(t),B(t),t)\Big]dt \\ Y(t_0) &= Y_0, \end{aligned}$$

which has a unique local solution on a maximal interval $(t_0^1, t_0^2) \subseteq [0,T]$ with $t_0 \in (t_0^1, t_0^2)$ (see [13]).

Applying the Itô formula, see Theorem 4.1 and relation (4.4), to the random function $Q(z,t) = K(Y(t),z,t)$ (in fact, successively for $K^1,\dots,K^n$)

and the fractional Brownian motion B we obtain

$$\begin{aligned}
&K(Y(t),B(t),t)-K(Y(t_0),B(t_0),t_0)\\
&=\sum_{j=1}^{n}\int_{t_0}^{t}\frac{\partial K}{\partial y^j}(Y(s),B(s),s)dY^j(s)+\int_{t_0}^{t}\frac{\partial K}{\partial z}(Y(s),B(s),s)dB(s)\\
&\qquad+\int_{t_0}^{t}\frac{\partial K}{\partial t}(Y(s),B(s),s)ds\\
&=\sum_{j=1}^{n}\int_{t_0}^{t}\frac{\partial K}{\partial y^j}(Y(s),B(s),s)dY^j(s)\\
&\qquad+\int_{t_0}^{t}G(K(Y(s),B(s),s),s)dB(s)+\int_{t_0}^{t}\frac{\partial K}{\partial t}(Y(s),B(s),s)ds\\
&=\int_{t_0}^{t}F(K(Y(s),B(s),s),s)ds+\int_{t_0}^{t}G(K(Y(s),B(s),s),s)dB(s).
\end{aligned}$$

Therefore,

$$X(t):=K(Y(t),B(t),t)$$

satisfies

$$X(t)=X_0+\int_{t_0}^{t}F(X(s),s)ds+\int_{t_0}^{t}G(X(s),s)dB(s).$$

Instead of the process $\Big(B(t)\Big)_{t\in[0,1]}$ we consider its approximations $\Big(B^N(t)\Big)_{t\in[0,1]}$ given in (2.3). For each $N\in\mathbb{N}$ we consider the pathwise differential equation (in matrix representation)

$$\begin{aligned}
dY_N(t) &= \left(\frac{\partial K}{\partial y}(Y_N(t),B^N(t),t)\right)^{-1}\Big[F(K(Y_N(t),B^N(t),t),t)\\
&\qquad -\frac{\partial K}{\partial t}(Y_N(t),B^N(t),t)\Big]dt\\
Y_N(t_0) &= Y_0,
\end{aligned}$$

which has a unique local solution Y_N on a maximal interval $(t^1,t^2)\subset(t_0^1,t_0^2)$ of existence which contains t_0. Applying the Itô formula, see Theorem 4.1 and (4.5), to the random function $Q(z,t)=K(Y_N(t),z,t)$ (in fact, successively for $K^1, \ldots, K^n$) and the process B_N we obtain

$$\begin{aligned}
&K(Y_N(t), B^N(t), t) - K(Y_N(t_0), B^N(t_0), t_0) \\
&= \sum_{j=1}^{n} \int_{t_0}^{t} \frac{\partial K}{\partial y^j}(Y_N(s), B_N(s), s) dY_N^j(s) + \int_{t_0}^{t} \frac{\partial K}{\partial z}(Y_N(s), B^N(s), s) dB^N(s) \\
&\quad + \int_{t_0}^{t} \frac{\partial K}{\partial t}(Y_N(s), B^N(s), s) ds \\
&= \sum_{j=1}^{n} \int_{t_0}^{t} \frac{\partial K}{\partial y^j}(Y_N(s), B^N(s), s) dY_N^j(s) + \int_{t_0}^{t} G(K(Y_N(s), B^N(s), s), s) dB^N(s) \\
&\quad + \int_{t_0}^{t} \frac{\partial K}{\partial t}(Y_N(s), B^N(s), s) ds \\
&= \int_{t_0}^{t} F(K(Y_N(s), B^N(s), s), s) ds + \int_{t_0}^{t} G(K(Y_N(s), B^N(s), s), s) dB^N(s).
\end{aligned}$$

Therefore,

$$X_N(t) := K(Y_N(t), B^N(t), t)$$

satisfies

$$X_N(t) = X_0 + \int_{t_0}^{t} F(X_N(s), s) ds + \int_{t_0}^{t} G(X_N(s), s) dB^N(s), \quad t \in (t_1, t_2).$$

By Theorem 7.2 [13] it follows that we have the following pathwise property

$$\lim_{N \to \infty} \sup_{t \in (t_1, t_2)} \|Y_N(t) - Y(t)\| = 0.$$

Then the continuity properties of K and (2.4) imply that for a.e. $\omega \in \Omega$ it holds

$$\lim_{N \to \infty} \sup_{t \in (t_1, t_2)} \|X_N(t) - X(t)\| = 0.$$

By this we have proved the main result of our paper:

Theorem 5.1. *Let B be a multifractional Brownian motion approximated through the processes B^N given in* (2.1) *and* (2.3). *Let $F, G : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$ be random functions satisfying with probability* 1 *the conditions* (C1) *and* (C2). *Let $t_0 \in (0, T]$ be fixed. Then, each of the stochastic equations*

$$X(t) = X_0 + \int_{t_0}^{t} F(X(s), s) ds + \int_{t_0}^{t} G(X(s), s) dB(s),$$

$$X_N(t) = X_0 + \int_{t_0}^{t} F(X_N(s), s)ds + \int_{t_0}^{t} G(X_N(s), s)dB^N(s), \quad N \in \mathbb{N}$$

admits almost surely a unique local solution on a common interval (t_1, t_2) *(which is independent of* N *and contains* t_0*). Moreover, we have the following approximation result*

$$P\left(\lim_{N\to\infty} \sup_{t\in(t_1,t_2)} \|X_N(t) - X(t)\| = 0\right) = 1.$$

References

[1] Ayache, A., Bertrand, P.R., *A process very similar to multifractional Brownian motion*, submitted.

[2] Ayache, A., Taqqu, M. S., *Rate optimality of wavelet series approximations of fractional Brownian motion*, Jour. Fourier Anal. Appl., **9**(2003), 451-471.

[3] Benassi, A., Bertrand, P.R., Cohen, S., Istas, J., *Identification of the Hurst index of a step fractional Brownian motion*, Stat. Inference Stoch. Proc., **3**(2000), no. 1-2, 101-111.

[4] Bender, C., Sottinen, T., Valkeila, E., *Pricing by Hedging and no Arbitrage beyond Semimartingales,* Finance Stoch., **12**(2008), 441-468.

[5] Decreusefond, L., Üstünel, A.S., *Stochastic Analysis of the Fractional Brownian Motion,* Potential Analysis, **10**(1999), 177-214.

[6] Denk, G., Winkler, R., *Modeling and simulation of transient noise in circuit simulation*, Math. Comput. Model. Dyn. Syst., **13**(2007), no. 4, 383-394.

[7] Dzhaparidze, K., van Zanten, H., *A Series Expansion of Fractional Brownian Motion*, Probab. Theory Relat. Fields, **130**(2004), 39-55.

[8] Guasoni, P., *No arbitrage under tranzaction costs, with fractional Brownian motion and beyond*, Math. Finance, **16**(2006), 569-582.

[9] Klingenhöfer, F., Zähle, M., *Ordinary Differential Equations with Fractal Noise*, Proc. Amer. Math. Soc., **127**(1999), 1021-1028.

[10] Kloeden, P.E., Platen, E., *Numerical Solution of Stochastic Differential equations*, Springer, Berlin, 1999.

[11] Kou, S., *Stochastic modeling in nanoscale Physics: subdiffusion within proteins*, Ann. Appl. Stat., **2**(2008), no. 2, 501-535.

[12] Meyer, Y., Sellan, F., Taqqu, M.R., *Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion*, The Journal of Fourier Analysis and Applications, **5**(1999), 465-494.

[13] Lisei, H., Soós, A, *Approximation of Stochastic Differential Equations Driven by Fractional Brownian Motion*, Seminar on Stochastic Analysis, Random Fields and Applications V, Progress in Probability 59, Birkhäuser Verlag 2007, 229-244.

[14] Samko, S.G., Kilbas, A.A., Marichev, O.I., *Fractional Integrals and Derivatives. Theory and Applications*, Gordon and Breach, 1993.

[15] Zähle, M., *Integration with respect to Fractal Functions and Stochastic Calculus I.*, Probab. Theory Relat. Fields, **111**(1998), 333-374.

[16] Zähle, M., *Integration with respect to Fractal Functions and Stochastic Calculus II.*, Math. Nachr., **225**(2001), 145-183.

Anna Soós
"Babeş-Bolyai" University
Faculty of Mathematics and Computer Sciences
1, Kogălniceanu Street
400084 Cluj-Napoca
Romania
e-mail: asoos@math.ubbcluj.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 599–608

Estimates with optimal constants using Peetre's K-functionals of order 2

Maria Talpău Dimitriu

Abstract. We present estimates of the degree of approximation by positive linear operators which preserve linear function, with the K-functionals K_2^s and $K_{2,\varphi}^s$, $1 \leq s \leq \infty$.

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

Keywords: Positive linear operators, Peetre's K-functional, degree of approximation.

1. Introduction

Estimates with the second order modulus ω_2 given by

$$\omega_2(f,t) = \sup_{|x-y|\leq 2t} \left| f(x) - f\left(\frac{x+y}{2}\right) + f(y) \right|,\ f \in \mathbf{C}[a,b],\ t > 0$$

were established by H. Esser in 1976, G. Freud in 1978, H. Gonska in 1984 and R. Păltănea in 1995.

In [8] is given the following axiomatic definition for the modulus of continuity:

Definition 1.1. *Let X be a linear space of functions $f : I \longrightarrow \mathbb{R}$ ($I \subset \mathbb{R}$ an interval) who include the space of algebric polynomials of degree at most r denoted by Π_r, $r \in \mathbb{N}$. A function $\Omega_r : X \times (0,\infty) \longrightarrow [0,\infty) \cup \{\infty\}$ is called a modulus of continuity of order r on X if and only if the following axioms are satisfied*

1. $\Omega_r(f,t_1) \leq \Omega_r(f,t_2)$ *if* $0 < t_1 < t_2$
2. $\Omega_r(f+p,t) = \Omega_r(f,t)$ *if* $p \in \Pi_{r-1}$
3. $\Omega_r(0,t) = 0.$

Moreover, if there exists a constant $M > 0$ such that $\Omega_r(e_r,t) \leq Mt^r$ for all $t > 0$, then the modulus ω_r is called normalized.

There are established estimates with different second order moduli based on the following general result:

Theorem 1.2. [1] [8, p. 20]*Let* $[c,d]\subset[a,b]$, $L:\mathbf{C}[a,b]\longrightarrow\mathbf{C}[c,d]$ *a positive linear operator such that* $Le_0=e_0$ *and* $Le_1=e_1$, Ω_2 *a second order modulus on* $\mathbf{C}[a,b]$, $f\in\mathbf{C}[a,b]$, $t>0$ *and* $x\in(c,d)$. *Suppose that there exists a function* $\psi:[0,\infty)\longrightarrow[0,\infty)$ *such that* $\psi\left(\dfrac{|e_1-xe_0|}{t}\right)\in\mathbf{C}[a,b]$ *and*

$$\left|\Delta\left(f;x_1,x,x_2\right)\right|\le\delta_t\left(\psi;x_1,x,x_2\right)\Omega_2(f,t),\ a\le x_1<x<x_2\le b.$$

Then

$$|L(f,x)-f(x)|\le L\left(\psi\left(\frac{|e_1-xe_0|}{t}\right),x\right)\Omega_2(f,t).$$

The notations used are:

- e_k for the function $e_k(x)=x^k$, $k\in\mathbb{N}\cup\{0\}$;
- $\Delta(f;x_1,x,x_2)=\dfrac{x_2-x}{x_2-x_1}f\left(x_1\right)+\dfrac{x-x_1}{x_2-x_1}f\left(x_2\right)-f(x)$ for $f:[a,b]\longrightarrow\mathbb{R}$, x_1, x, $x_2\in[a,b]$, $x_1\neq x_2$;
- $\delta_t(\psi;x_1,x,x_2)=\dfrac{x_2-x}{x_2-x_1}\psi\left(\dfrac{x-x_1}{t}\right)+\dfrac{x-x_1}{x_2-x_1}\psi\left(\dfrac{x_2-x}{t}\right)$ for $\psi:[0,\infty)\longrightarrow\mathbb{R}$, x_1, x, $x_2\in[0,\infty)$, $x_1<x<x_2$, $t>0$.

The K-functional $K_r^s(f,t)=K^s\left(f,t^r;\mathbf{C}[a,b],\mathbf{C^r}[a,b]\right),\ t>0,\ 1\le s\le\infty$ defined for the Banach space $(\mathbf{C}[a,b],\|\cdot\|)$ and the semi-Banach subspace $\left(\mathbf{C^r}[a,b],|\cdot|_{C^r}\right)$, $|f|_{C^r}=\left\|f^{(r)}\right\|$ by

$$K_r^s(f,t)=\inf_{g\in\mathbf{C^r}[a,b]}\left\|\left(\|f-g\|,t^r\left\|g^{(r)}\right\|\right)\right\|_s,\ 1\le s\le\infty,$$

where $\|\cdot\|_s$, $1\le s<\infty$, is the Minkowski norm in $\mathbb{R}^2$ and $\|\cdot\|_\infty$ is the Chebychev norm in $\mathbb{R}^2$, respectively, is a modulus of continuity of order r normalized on $\mathbf{C}[a,b]$. An useful relation between the K-functionals is given by:

Lemma 1.3. [13] *Let* $1\le s<\infty$ *and* $r\in\mathbb{N}$. *Then for* $f\in\mathbf{C}[a,b]$ *and* $t>0$

$$K_r^s(f,t)=\inf_{u>0}\left(1+\frac{t^{rs}}{u^{rs}}\right)^{\frac{1}{s}}K_r^\infty(f,u)\ \textit{holds.}\tag{1.1}$$

In the weighted case, for $r\in\mathbb{N}$ and $\varphi(x)=\sqrt{x(1-x)}$, $x\in[0,1]$ we denote by

$$\mathbf{C}_\varphi[0,1]=\left\{f\in\mathbf{C}(0,1)\,|\,(\exists)\lim_{x\to0+}f(x)\varphi(x),\ \lim_{x\to1-}f(x)\varphi(x)\in\mathbb{R}\right\}$$

and

$$\mathbf{W}^{\mathbf{r}}_{\mathbf{C}_{\varphi^{\mathbf{r}}}}[0,1]=\left\{f\in\mathbf{C^{r-1}}[0,1]\,|\,f^{(r)}\in\mathbf{C}_{\varphi^{\mathbf{r}}}[0,1]\right\}.$$

The K-functional $K^s_{r,\varphi}(f,t)=K^s\left(f,t^r;\mathbf{C}[0,1],\mathbf{W}^{\mathbf{r}}_{\mathbf{C}_{\varphi^{\mathbf{r}}}}[0,1]\right)$, $t>0$, $1\le s\le\infty$ defined for the Banach space $(\mathbf{C}[0,1],\|\cdot\|)$ and the semi-Banach subspace

[1] We refer here only the particular case when the operators preserves the linear functions.

$\left(\mathbf{W}^{\mathbf{r}}_{\mathbf{C}_{\varphi^{\mathbf{r}}}}[0,1], |\cdot|_{W^r_{C_{\varphi^r}}}\right)$, $|f|_{W^r_{C_{\varphi^r}}} = \left\|\varphi^r f^{(r)}\right\|$ by

$$K^s_{r,\varphi}(f,t) = \inf_{g\in \mathbf{W}^{\mathbf{r}}_{\mathbf{C}_{\varphi^{\mathbf{r}}}}[0,1]} \left\|\left(\|f-g\|, t^r \left\|\varphi^r g^{(r)}\right\|\right)\right\|_s,\ 1\le s\le\infty,$$

is a modulus of continuity of order r normalized on $\mathbf{C}[0,1]$ and we have

Lemma 1.4. [14] *Let $1\le s<\infty$ and $r\in\mathbb{N}$. Then for $f\in\mathbf{C}[a,b]$ and $t>0$*

$$K^s_{r,\varphi}(f,t) = \inf_{u>0}\left(1+\frac{t^{rs}}{u^{rs}}\right)^{\frac{1}{s}} K^\infty_{r,\varphi}(f,u) \text{ holds.} \tag{1.2}$$

In Section 2 are given estimates with the K-functional K^s_2 and in Section 3 are given estimates with the K-functional $K^s_{2,\varphi}$.

2. General estimates with K^s_2, $1\le s\le\infty$

Theorem 2.1. *Let $[c,d]\subseteq[a,b]$, $L:\mathbf{C}[a,b]\longrightarrow\mathbf{C}[c,d]$ a positive linear operator such that $Le_0=e_0$, $Le_1=e_1$ and $f\in\mathbf{C}[a,b]$. Then for every $x\in(c,d)$ and $t>0$, we have*

$$|L(f,x)-f(x)|\le\left(2+\frac{L\left((e_1-xe_0)^2,x\right)}{2t^2}\right)K^\infty_2(f,t). \tag{2.1}$$

Conversely, if there exist A, $B\ge 0$ such that

$$|L(f,x)-f(x)|\le\left(A+B\frac{L\left((e_1-xe_0)^2,x\right)}{t^2}\right)K^\infty_2(f,t) \tag{2.2}$$

holds for all positive linear operator L, any $f\in\mathbf{C}[a,b]$, any $x\in(c,d)$ and any $t>0$, then $B\ge\dfrac{1}{2}$ and $A\ge 2$.

Proof. Let $g\in\mathbf{C^2}[a,b]$, x_1, x, $x_2\in[a,b]$, $x_1<x<x_2$. We have

$$\begin{aligned}|\Delta(f;x_1,x,x_2)| &\le |\Delta(f-g;x_1,x,x_2)|+|\Delta(g;x_1,x,x_2)|\\ &\le 2\|f-g\|+|\Delta(g;x_1,x,x_2)|\end{aligned}$$

and

$$\begin{aligned}|\Delta(g;x_1,x,x_2)| &= \left|\frac{x_2-x}{x_2-x_1}(g(x_1)-g(x))+\frac{x-x_1}{x_2-x_1}(g(x_2)-g(x))\right|\\ &= \Big|\frac{x_2-x}{x_2-x_1}\left(g'(x)(x_1-x)+\frac{g''(\xi_1)}{2}(x_1-x)^2\right)\\ &\quad+\frac{x-x_1}{x_2-x_1}\left(g'(x)(x_2-x)+\frac{g''(\xi_2)}{2}(x_2-x)^2\right)\Big|\\ &= \frac{(x_2-x)(x-x_1)}{2(x_2-x_1)}\left|g''(\xi_1)(x-x_1)+g''(\xi_2)(x_2-x)\right|\\ &\le \frac{(x_2-x)(x-x_1)}{2}\|g''\|\end{aligned}$$

with ξ_i between x and x_i, $i = 1, 2$. Therefore

$$\begin{aligned}|\Delta(f; x_1, x, x_2)| &\leq 2\|f-g\| + \frac{(x_2-x)(x-x_1)}{2t^2} t^2 \|g''\| \\ &\leq \left(2 + \frac{(x_2-x)(x-x_1)}{2t^2}\right) \max\left\{\|f-g\|, t^2\|g''\|\right\}.\end{aligned}$$

Since g was arbitrary it follows that

$$|\Delta(f; x_1, x, x_2)| \leq \left(2 + \frac{(x_2-x)(x-x_1)}{2t^2}\right) K_2^{\infty}(f, t). \tag{2.3}$$

If we take $\psi(u) = 2 + \dfrac{u^2}{2}$ then (2.3) means

$$|\Delta(f; x_1, x, x_2)| \leq \delta_t(\psi; x_1, x, x_2) K_2^{\infty}(f, t),\ x_1 < x < x_2.$$

By Theorem 1.2 we have

$$|L(f, x) - f(x)| \leq \left(2 + \frac{L\left((e_1 - xe_0)^2, x\right)}{2t^2}\right) K_2^{\infty}(f, t).$$

Now we prove the converse part. We consider the positive linear operator L defined by

$$L(h, x) = (1-x)h(0) + xh(1),\ h \in \mathbf{C}[0, 1].$$

For $f = e_2$ we have $K_2^{\infty}(e_2, t) \leq 2t^2$ and from (2.2) it follows

$$x(1-x) \leq 2At^2 + 2Bx(1-x).$$

Passing to the limit $t \to 0$, we obtain $B \geq \dfrac{1}{2}$.

For $f(x) = \alpha(4x-1)$, $x \in \left[0, \dfrac{1}{2}\right]$, $f(x) = \alpha(3-4x)$, $x \in \left(\dfrac{1}{2}, 1\right]$, $\alpha > 0$, we have $K_2^{\infty}(f, t) \leq \|f\| = \alpha$ and from (2.2) it follows for $x = \dfrac{1}{2}$ that

$$2\alpha \leq \left(A + \frac{B}{4t^2}\right)\alpha.$$

Passing to the limit $t \to \infty$, we obtain $A \geq 2$. □

Corollary 2.2. *Under the conditions of theorem we have*

$$|L(f, x) - f(x)| \leq \max\left\{2, \frac{L\left((e_1 - xe_0)^2, x\right)}{2t^2}\right\} K_2^1(f, t) \tag{2.4}$$

and

$$|L(f, x) - f(x)| \leq \left(2^{s'} + \frac{L\left((e_1 - xe_0)^2, x\right)^{s'}}{2^{s'} t^{2s'}}\right)^{\frac{1}{s'}} K_2^s(f, t) \tag{2.5}$$

where $1 < s < \infty$ *and* $s' = \dfrac{s}{s-1}$.

Conversely

- *if there exist* A, $B \geq 0$ *such that*

$$|L(f,x) - f(x)| \leq \max\left\{A,\ B\frac{L\left((e_1 - xe_0)^2, x\right)}{t^2}\right\} K_2^1(f,t) \tag{2.6}$$

  holds for all positive linear operator L, *any* $f \in \mathbf{C}[a,b]$, *any* $x \in (c,d)$ *and any* $t > 0$, *then* $B \geq \dfrac{1}{2}$ *and* $A \geq 2$.

- *if there exist* A, $B \geq 0$ *such that*

$$|L(f,x) - f(x)| \leq \left(A + B\frac{L\left((e_1 - xe_0)^2, x\right)^{s'}}{t^{2s'}}\right)^{\frac{1}{s'}} K_2^s(f,t) \tag{2.7}$$

  holds for all positive linear operator L, *any* $f \in \mathbf{C}[a,b]$, *any* $x \in (c,d)$ *and any* $t > 0$, *then* $B \geq \dfrac{1}{2^{s'}}$ *and* $A \geq 2^{s'}$.

Proof. Using the estimate (2.1), we obtain

$$\begin{aligned}
|L(f,x) - f(x)| &\leq \left(2 + \frac{L\left((e_1 - xe_0)^2, x\right)}{2u^2}\right) K_2^\infty(f,u) \\
&\leq \max\left\{2, \frac{L\left((e_1 - xe_0)^2, x\right)}{2t^2}\right\}\left(1 + \frac{t^2}{u^2}\right) K_2^\infty(f,u),
\end{aligned}$$

where $u > 0$ is arbitrary. Hence, by Lemma 1.3, we find (2.4). For $1 < s < \infty$, by (2.1) and Hölder's inequality, we have

$$|L(f,x) - f(x)| \leq \left(2^{s'} + \frac{L\left((e_1 - xe_0)^2, x\right)^{s'}}{2^{s'}t^{2s'}}\right)^{\frac{1}{s'}} \left(1 + \frac{t^{2s}}{u^{2s}}\right)^{\frac{1}{s}} K_2^\infty(f,u),$$

where $u > 0$ is arbitrary. Hence, by Lemma 1.3, we find (2.5).

For the converse part we make the same choices like in Theorem 2.1. $\square$

Example 2.3. We consider the Bernstein-type operator $P_{n,m} : \mathbf{C}[0,1] \longrightarrow \mathbf{C}[0,1]$ (see [12], [3])

$$P_{n,m}(f,x) = \sum_{k=0}^{n} b_{n,k,m}(x) \cdot f\left(\frac{k}{n}\right)$$

where

$b_{n,k,m}(x) = \binom{n-m}{k} x^k (1-x)^{n-m-k+1}$ for $0 \leq k < m$,

$b_{n,k,m}(x) = \binom{n-m}{k} x^k (1-x)^{n-m-k+1} + \cdots + \binom{n-m}{k-m} x^{k-m+1} (1-x)^{n-k}$

for $m \leq k \leq n-m$ and

$b_{n,k,m}(x) = \binom{n-m}{k-m} x^{k-m+1}(1-x)^{n-k}$ for $n-m < k \le n$,
with $n \in \mathbb{N}$, $m \in \mathbb{N} \cup \{0\}$, $m < \dfrac{n}{2}$. We have

$$\begin{aligned} P_{n,m}(e_0, x) &= 1 \\ P_{n,m}(e_1, x) &= x \\ P_{n,m}\left((e_1 - xe_0)^2, x\right) &= \left(1 + \frac{m(m-1)}{n}\right) \frac{x(1-x)}{n}. \end{aligned}$$

Theorem 2.1 implies for $f \in \mathbf{C}[0,1]$, $x \in (0,1)$ and $t = \sqrt{\dfrac{x(1-x)}{n}}$

$$|P_{n,m}(f,x) - f(x)| \le \left[2 + \frac{1}{2}\left(1 + \frac{m(m-1)}{n}\right)\right] K_2^\infty\left(f, \sqrt{\frac{x(1-x)}{n}}\right).$$

From Corollary 2.2 we obtain

$$|P_{n,m}(f,x) - f(x)| \le \max\left\{2, \frac{1}{2}\left(1 + \frac{m(m-1)}{n}\right)\right\} K_2^1\left(f, \sqrt{\frac{x(1-x)}{n}}\right)$$

and

$$|P_{n,m}(f,x) - f(x)| \le \left[2^{s'} + \frac{1}{2^{s'}}\left(1 + \frac{m(m-1)}{n}\right)^{s'}\right]^{\frac{1}{s'}} K_2^s\left(f, \sqrt{\frac{x(1-x)}{n}}\right)$$

where $1 < s < \infty$ and $s' = \dfrac{s}{s-1}$. In particular, for $m = 0$ or $m = 1$ we obtain the estimates for the Bernstein operators.

3. General estimates with $K_{2,\varphi}^s$, $1 \le s \le \infty$

Theorem 3.1. *Let $L : \mathbf{C}[0,1] \longrightarrow \mathbf{C}[0,1]$ be a positive linear operator such that $Le_0 = e_0$, $Le_1 = e_1$ and $f \in \mathbf{C}[0,1]$. Then for all $x \in (0,1)$ and $t > 0$ we have*

$$|L(f,x) - f(x)| \le \left(2 + \frac{L\left((e_1 - xe_0)^2, x\right)}{t^2\varphi^2(x)}\right) K_{2,\varphi}^\infty(f,t). \tag{3.1}$$

Conversely, if there exist A, $B \ge 0$ such that

$$|L(f,x) - f(x)| \le \left(A + B\frac{L\left((e_1 - xe_0)^2, x\right)}{t^2\varphi^2(x)}\right) K_{2,\varphi}^\infty(f,t) \tag{3.2}$$

holds for all positive linear operator L, any $f \in \mathbf{C}[0,1]$, any $x \in (0,1)$ and any $t > 0$, then $B \ge 1$ and $A \ge 2$.

Proof. Let $g \in \mathbf{W}^{\mathbf{2}}_{\mathbf{C}_{\varphi^2}}[0,1]$, x_1, x, $x_2 \in [0,1]$, $x_1 < x < x_2$. We have

$$\begin{aligned} |\Delta(f; x_1, x, x_2)| &\le |\Delta(f-g; x_1, x, x_2)| + |\Delta(g; x_1, x, x_2)| \\ &\le 2\|f-g\| + |\Delta(g; x_1, x, x_2)| \end{aligned}$$

and

$$\begin{aligned}
|\Delta(g;x_1,x,x_2)| &= \left|\frac{x_2-x}{x_2-x_1}\left(g(x_1)-g(x)\right)+\frac{x-x_1}{x_2-x_1}\left(g(x_2)-g(x)\right)\right| \\
&= |\frac{x_2-x}{x_2-x_1}\left[g'(x)(x_1-x)+\int_x^{x_1} g''(u)(x_1-u)\,du\right] \\
&\quad +\frac{x-x_1}{x_2-x_1}\left[g'(x)(x_2-x)+\int_x^{x_2} g''(u)(x_2-u)\,du\right]| \\
&= \left|\frac{x_2-x}{x_2-x_1}\int_x^{x_1} g''(u)(x_1-u)\,du-\frac{x_1-x}{x_2-x_1}\int_x^{x_2} g''(u)(x_2-u)\,du\right| \\
&\le \frac{x_2-x}{x_2-x_1}\cdot\left\|\varphi^2 g''\right\|\cdot\int_{x_1}^{x}\frac{u-x_1}{\varphi^2(u)}du+\frac{x-x_1}{x_2-x_1}\cdot\left\|\varphi^2 g''\right\|\cdot\int_x^{x_2}\frac{x_2-u}{\varphi^2(u)}du.
\end{aligned}$$

Let us now make use of the fact that the function $u\mapsto \dfrac{t-u}{u(1-u)}$, $u\in(0,t)$, $t\in(0,1]$ is decreasing [8] and we obtain

$$\begin{aligned}
&|\Delta(g;x_1,x,x_2)| \le \\
&\le \frac{x_2-x}{x_2-x_1}\cdot\left\|\varphi^2 g''\right\|\cdot\int_{1-x}^{1-x_1}\frac{1-u-x_1}{\varphi^2(1-u)}du+\frac{x-x_1}{x_2-x_1}\cdot\left\|\varphi^2 g''\right\|\cdot\int_x^{x_2}\frac{x_2-u}{\varphi^2(u)}du \\
&\le \frac{x_2-x}{x_2-x_1}\cdot\left\|\varphi^2 g''\right\|\cdot\int_{1-x}^{1-x_1}\frac{x-x_1}{\varphi^2(1-x)}du+\frac{x-x_1}{x_2-x_1}\cdot\left\|\varphi^2 g''\right\|\cdot\int_x^{x_2}\frac{x_2-x}{\varphi^2(x)}du \\
&= \frac{(x_2-x)(x-x_1)}{\varphi^2(x)}\left\|\varphi^2 g''\right\|
\end{aligned}$$

therefore

$$\begin{aligned}
|\Delta(f;x_1,x,x_2)| &\le 2\|f-g\|+\frac{(x_2-x)(x-x_1)}{t^2\varphi^2(x)}t^2\left\|\varphi^2 g''\right\| \\
&\le \left(2+\frac{(x_2-x)(x-x_1)}{t^2\varphi^2(x)}\right)\max\left\{\|f-g\|,\,t^2\left\|\varphi^2 g''\right\|\right\}.
\end{aligned}$$

Since g was arbitrary it follows that

$$|\Delta(f;x_1,x,x_2)|\le\left(2+\frac{(x_2-x)(x-x_1)}{t^2\varphi^2(x)}\right)K^{\infty}_{2,\varphi}(f,t). \tag{3.3}$$

If we take $\psi(u)=2+\dfrac{u^2}{\varphi^2(x)}$ then (3.3) means

$$|\Delta(f;x_1,x,x_2)|\le\delta_t(\psi;x_1,x,x_2)\,K^{\infty}_{2,\varphi}(f,t),\ x_1<x<x_2.$$

By Theorem 1.2 we have

$$|L(f,x)-f(x)| \le \left(2+\frac{L\left((e_1-xe_0)^2,x\right)}{t^2\varphi^2(x)}\right) K_{2,\varphi}^{\infty}(f,t).$$

Now we prove the converse part. To show that $A \ge 2$, we consider the positive linear operator L defined by

$$L(h,x)=(1-x)h(0)+xh(1),\ h\in \mathbf{C}[0,1].$$

For $f(x)=\alpha(4x-1)$, $x\in\left[0,\frac{1}{2}\right]$, $f(x)=\alpha(3-4x)$, $x\in\left(\frac{1}{2},1\right]$, $\alpha>0$, we have $K_{2,\varphi}^{\infty}(f,t)\le\|f\|=\alpha$ and from (3.2) it follows for $x=\frac{1}{2}$ that

$$2\alpha \le \left(A+\frac{B}{t^2}\right)\alpha.$$

Passing to the limit $t\to\infty$, we obtain $A\ge 2$.

To show that $B\ge 1$, we choose

$$L(h,x)=\left(1-x^{\beta}\right)h(0)+x^{\beta}h\left(x^{1-\beta}\right),\ \beta\in(0,1),\ h\in\mathbf{C}[0,1]$$

and $f(x)=x^{1+\alpha}$, $\alpha>0$. We have $f\in\mathbf{W}^{\mathbf{2}}_{\mathbf{C}_{\varphi^{\mathbf{2}}}}[0,1]$ and then

$$K_{2,\varphi}^{\infty}(f,t)\le t^2\left\|\varphi^2 f''\right\| = t^2\frac{\alpha^{\alpha+1}}{(\alpha+1)^{\alpha}}.$$

We replace it in (3.2) and passing to the limit $t\to 0$, we obtain

$$x^{1+\alpha}\left(x^{-\alpha\beta}-1\right)\le B\cdot\frac{x\left(x^{-\beta}-1\right)}{1-x}\cdot\frac{\alpha^{\alpha+1}}{(\alpha+1)^{\alpha}}$$

i.e.

$$B\ge x^{\alpha}(1-x)\cdot\frac{\left(x^{-\alpha\beta}-1\right)}{x^{-\beta}-1}\cdot\frac{(\alpha+1)^{\alpha}}{\alpha^{\alpha+1}}.$$

Passing to the limit $\beta\to 0$, we obtain $B\ge x^{\alpha}(1-x)\frac{(\alpha+1)^{\alpha}}{\alpha^{\alpha}}$. Since x is arbitrary, this implies $B\ge\frac{1}{\alpha+1}$. Passing to the limit $\alpha\to 0$, we obtain $B\ge 1$. □

Corollary 3.2. *Under the conditions of theorem we have*

$$|L(f,x)-f(x)|\le \max\left\{2,\ \frac{L\left((e_1-xe_0)^2,x\right)}{t^2\varphi^2(x)}\right\}K_{2,\varphi}^{1}(f,t) \tag{3.4}$$

and

$$|L(f,x)-f(x)|\le\left(2^{s'}+\frac{L\left((e_1-xe_0)^2,x\right)^{s'}}{t^{2s'}\varphi^{2s'}(x)}\right)^{\frac{1}{s'}}K_{2,\varphi}^{s}(f,t) \tag{3.5}$$

where $1 < s < \infty$ *and* $s' = \dfrac{s}{s-1}$.

Conversely

- *if there exist* A, $B \geq 0$ *such that*

$$|L(f,x) - f(x)| \leq \max\left\{A,\ B\frac{L\left((e_1 - xe_0)^2, x\right)}{t^2\varphi^2(x)}\right\} K^1_{2,\varphi}(f,t) \qquad (3.6)$$

  holds for all positive linear operator L, *any* $f \in \mathbf{C}[0,1]$, *any* $x \in (0,1)$ *and any* $t > 0$, *then* $B \geq 1$ *and* $A \geq 2$.

- *if there exist* A, $B \geq 0$ *such that*

$$|L(f,x) - f(x)| \leq \left(A + B\frac{L\left((e_1 - xe_0)^2, x\right)^{s'}}{t^{2s'}\varphi^{2s'}(x)}\right)^{\frac{1}{s'}} K^s_{2,\varphi}(f,t) \qquad (3.7)$$

  holds for all positive linear operator L, *any* $f \in \mathbf{C}[0,1]$, *any* $x \in (0,1)$ *and any* $t > 0$, *then* $B \geq 1$ *and* $A \geq 2^{s'}$.

Proof. We use the estimate (3.1) and Lemma 1.4 (see also the proof of Corollary 2.2). For the converse part we make the same choices like in Theorem 3.1. $\square$

References

[1] Bergh, J., Löfström, J., *Interpolation Spaces*, Springer-Verlag, 1976.

[2] Gonska, H.H., *On approximation in spaces of continuous functions*, Bull. Austral. Math. Soc., **28**(1983), 411-432.

[3] Gonska, H.H., *Quantitative Korovkin type theorems on simultaneous approximation*, Math. Z., **186**(1984), 419-433.

[4] Gonska, H.H., Piţul, P., Raşa, I., *On Peano's form of the Taylor remainder, Voronovskaja's theorem and the commutator of positive linear operators*, Numerical Analysis and Approximation Theory (Proc. Int. Conf. Cluj-Napoca 2006, ed. by O. Agratini and P. Blaga), 55-80.

[5] Mamedov, R.G., *On the order of approximation of functions by sequences of linear positive operators*, (Russian), Dokl. Akad. Nauk SSSR, **128**(1959), 674-676.

[6] Păltănea, R., *Best constants in estimates with second order moduli of continuity*, Approximation Theory (Proc. Int. Dortmund Meeting on Approximation Theory 1995, ed. by M.W. Müller, D. Mache, M. Felten), Berlin, Akad. Verlag, 1995, 251-275.

[7] Păltănea, R., *Optimal estimates with moduli of continuity*, Result. Math., **32**(1997), 318-331.

[8] Păltănea, R., *Approximation theory using positive linear operators*, Birkhäuser, 2004.

[9] Peetre, J., *A theory of interpolation of normed spaces*, Notas de Matematica 39, Rio de Janeiro, Instituto de Matematica Pura e Aplicada, Conselho Nacional de Pesquisas, 1968.

[10] Peetre, J., *On the connection between the theory of interpolation spaces and approximation theory*, Proc. Conf. Const. Theory of Functions, Budapest, Eds. G. Alexits and S. B. Stechkin, Akadémiai Kiadó, Budapest, 1969, 351-363.

[11] Shisha, O., Mond, B., *The degree of convergence of linear positive operators*, Proc. Nat. Acad. Sci. U.S.A., **60**(1968), 1196-1200.

[12] Stancu, D.D., *Quadrature formulas constructed by using certain linear positive operators*, Numerische Integration, 1982, 241-251.

[13] Talpău Dimitriu, M., *Estimates with optimal constants using Peetre's K-functionals*, Carpathian J. Math., **26**(2010), no. 2, 158-169.

[14] Talpău Dimitriu, M., *Estimates with optimal constants for the operator of r-th order*, Bulletin of the Transilvania University of Braşov, **52**(2010), no. 3, 143-154.

Maria Talpău Dimitriu
"Transilvania" University
Department of Mathematics
Eroilor 29, 500 036, Braşov, Romania
e-mail: marias_td@yahoo.com; mdimitriu@unitbv.ro

Stud. Univ. Babeş-Bolyai Math. 56(2011), No. 2, 609–614

On the approximation of the constant of Napier

Andrei Vernescu

Abstract. Starting from some older ideas of [12] and [6], we show new facts concerning the approximation of the constant of Napier.

Mathematics Subject Classification (2010): 26A09, 26D07,40A30, 41A10.

Keywords: The constant of Napier, exponential function, approximation, speed of convergence.

1. Introduction

Consider the two equivalent classical definitions of the real exponential function

$$\mathrm{e}^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots \frac{x^n}{n!} + \dots \tag{1.1}$$

respectively

$$\mathrm{e}^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n, \tag{1.2}$$

both convergences being uniform on compact subsets of $\mathbb{R}$.
Their speed of convergence is different. Concerning the Taylor-Maclaurin approximation (1.1) of the exponential, see D. S. Mitrinović [3], pp. 268-269. For the approximation given by (1.2), also in this classical book are given the following inequalities

$$0 \le \mathrm{e}^x - \left(1 + \frac{x}{n}\right)^n \le \frac{x^2 \mathrm{e}^x}{n}, \quad \text{for } |x| < n \text{ and } n \in \mathbb{N}^*;$$
$$0 \le \mathrm{e}^{-x} - \left(1 - \frac{x}{n}\right)^n \le \frac{x^2 (1+x)\, \mathrm{e}^{-x}}{2n}, \quad \text{for } 0 \le x < n,\ n \in \mathbb{N},\ n \ge 2;$$
$$0 \le \mathrm{e}^{-x} - \left(1 - \frac{x}{n}\right)^n \le \frac{x^2}{2n}, \quad \text{for } 0 \le x \le n \text{ and } n \in \mathbb{N}^*$$

(see [4], [5], [13], [14], [15]).

In [7] we gave some stronger inequalities, namely

i) If $x > 0$, $t > 0$ and $t > \frac{1-x}{2}$ then

$$\frac{x^2 e^x}{2t + x + \max\{x, x^2\}} < e^x - \left(1 + \frac{x}{t}\right)^t < \frac{x^2 e^x}{2t + x}. \tag{1.3}$$

ii) If $x > 0$, $t > 0$ and $t > \frac{x-1}{2}$ then

$$\frac{x^2 e^{-x}}{2t - x + x^2} < e^{-x} - \left(1 - \frac{x}{t}\right)^t < \frac{x^2 e^{-x}}{2t - 2x + \min\{x, x^2\}} \tag{1.4}$$

and we detailed the proof of (1.3) (for the proof of (1.4) see[12], pp. 258-260).

Also, note *en passant*, that the previous inequalities give by the simple particularization $x = 1$, the characterizations of the "speed" of convergence of four standard sequences related to the numbers e and $\frac{1}{e}$, namely [1)]

$$\frac{e}{2n+2} < e - \left(1 + \frac{1}{n}\right)^n < \frac{e}{2n+1} \quad \text{([8], pag. 38, [11])}$$
$$\frac{e}{2n+1} < \left(1 + \frac{1}{n}\right)^{n+1} - e < \frac{e}{2n} \quad \text{([10])}$$
$$\frac{1}{2ne} < \frac{1}{e} - \left(1 - \frac{1}{n}\right)^n < \frac{1}{(2n-1)e} \quad \text{([6], [7])}$$
$$\frac{1}{(2n-1)e} < \left(1 - \frac{1}{n}\right)^{n-1} - \frac{1}{e} < \frac{1}{(2n-2)e} \quad \text{([6], [7]).}$$

2. The main result

Now we will establish the best approximation of e by the family of sequences of general term $\left(1 + \frac{1}{n}\right)^{n+p}$, where p is a real parameter; this may suggest the best approximation of e^x, $x > 0$, by some algebraic functions.

Consider the known limited expansion

$$(1 + x)^{\frac{1}{x}} = e\left(1 - \frac{1}{2}x + \frac{11}{24}x^2 - \frac{7}{16}x^3\right) + O(x^4), \tag{2.1}$$

and also the limited binomial one

$$(1 + x)^p = 1 + \frac{p}{1!}x + \frac{p(p-1)}{2!}x^2 + \frac{p(p-1)(p-2)}{3!}x^3 + O(x^4). \tag{2.2}$$

[1)] Using the notations $e_n = \left(1 + \frac{1}{n}\right)^n$, $f_n = \left(1 + \frac{1}{n}\right)^{n+1}$, $g_n = \left(1 - \frac{1}{n}\right)^n$, $h_n = \left(1 - \frac{1}{n}\right)^{n-1}$ and applying the GM-AM inequality for the numbers $a_1 = a_2 = a_3 = \ldots = a_n = 1 + \frac{1}{n}$, $a_{n+1} = 1$, we obtain that the sequence $(e_n)_n$ is strictly increasing (see [9]). Applying the GM-AM inequality for the numbers $b_1 = b_2 = b_3 = \ldots = b_n = 1 - \frac{1}{n}$, $b_{n+1} = 1$, we obtain analogously that the sequence $(g_n)_n$ is strictly increasing. The identities $f_n = \frac{1}{g_{n+1}}$ and $h_n = \frac{1}{e_{n-1}}$ show us that the sequences $(f_n)_n$ and $(h_n)_n$ are strictly decreasing. Therefore $e_n < e < f_n$ and $g_n < \frac{1}{e} < h_n$.

Remark. The formula (2.1) is can be obtained in a classical way, using the well-known limited expansions $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + O(x^5)$ and $\exp y = 1 + \frac{y}{1!} + \frac{y^2}{2!} + \frac{y^3}{3!} + \frac{y^4}{4!} + O(y^5)$. Then

$$\frac{1}{e}(1+x)^{\frac{1}{x}} = \frac{1}{e}\exp\left(\frac{1}{x}\ln(1+x)\right) =$$

$$= \exp\left(\frac{1}{x}\ln(1+x) - 1\right) = \exp\left(-\frac{x}{2} + \frac{x^2}{2} - \frac{x^3}{4} + O(x^4)\right) =$$

$$= \left(\sum_{k=0}^{3} \frac{1}{k!}\left(-\frac{x}{2} + \frac{x^2}{2} - \frac{x^3}{4} + O(x^4)\right)^k\right) + O(x^4)$$

and some standard calculations give (2.1).

Multiplying (2.1) and (2.2), part by part, performing the usual calculations and replacing x by $\frac{1}{n}$ $(n = 1, 2, 3, \ldots)$, we obtain

$$\begin{aligned}\left(1+\tfrac{1}{n}\right)^{n+p} &= \mathrm{e} + \left(p - \frac{1}{2}\right)\frac{\mathrm{e}}{n} + \frac{12p^2 - 24p + 11}{24} \cdot \frac{\mathrm{e}}{n^2} + \\ &+ \frac{8p^4 - 36p^2 + 50p - 21}{48} \cdot \frac{\mathrm{e}}{n^3} + O\left(\frac{1}{n^4}\right).\end{aligned} \tag{2.3}$$

From (2.3), we see that

$$\lim_{n\to\infty} n\left(\left(1+\frac{1}{n}\right)^{n+p} - \mathrm{e}\right) = \begin{cases} 0, & \text{for } p = \dfrac{1}{2} \\ \left(p - \dfrac{1}{2}\right)\mathrm{e} & \text{for } p \neq \dfrac{1}{2} \end{cases}. \tag{2.4}$$

For $p = \frac{1}{2}$ it results that the term in $\frac{1}{n}$ of (2.3) vanishes and we have

$$\left(1+\frac{1}{n}\right)^{n+1/2} = \mathrm{e} + \frac{\mathrm{e}}{12n^2} - \frac{\mathrm{e}}{12n^3} + O\left(\frac{1}{n^4}\right)$$

and so

$$n^2\left(\left(1+\frac{1}{n}\right)^{n+\frac{1}{2}} - \mathrm{e}\right) = \frac{\mathrm{e}}{12} - \frac{\mathrm{e}}{12n} + O\left(\frac{1}{n^2}\right),$$

which conducts us to the equality

$$\lim_{n\to\infty} n^2\left(\left(1+\frac{1}{n}\right)^{n+\frac{1}{2}} - \mathrm{e}\right) = \frac{\mathrm{e}}{12}. \tag{2.5}$$

Another way to obtain (2.5) consists in a (repeated) use of the *L'Hospital*'s rule, but this gives no idea of the provenance of the result.

So, the best approximation of e by the sequences of general term $\left(1+\frac{1}{n}\right)^{n+p}$ is the one corresponding to $p = \frac{1}{2}$.

3. A two-sided estimate

The equality (2.5) suggests us to search a two sided estimate of the form

$$\frac{\mathrm{e}}{12(n+\alpha)^2} < \left(1+\frac{1}{n}\right)^{n+\frac{1}{2}} - \mathrm{e} < \frac{\mathrm{e}}{12(n+\beta)^2} \tag{3.1}$$

where α and β are two real constants.

Professor Ioan Gavrea communicated me ([1]) a convenient left part of (3.1), namely for $\alpha = \frac{1}{2}$, we have

$$\frac{\mathrm{e}}{12\left(n+\dfrac{1}{2}\right)^2} < \left(1+\frac{1}{n}\right)^{n+\frac{1}{2}} - \mathrm{e}. \tag{3.2}$$

We present here his proof. Let

$$a_n = \frac{\left(1+\dfrac{1}{n}\right)^{n+\frac{1}{2}}}{\mathrm{e}}$$

be and $b_n = \ln a_n$, that is

$$b_n = \left(n+\frac{1}{2}\right)[\ln(n+1) - \ln n] - 1.$$

We have successively

$$b_n = \left(n+\frac{1}{2}\right)\left[\ln\left(n+\frac{1}{2}+\frac{1}{2}\right) - \ln\left(n+\frac{1}{2}-\frac{1}{2}\right)\right] - 1$$

$$= \left(n+\frac{1}{2}\right)\left[\ln\left(n+\frac{1}{2}\right)\left(1+\frac{1}{2\left(n+\dfrac{1}{2}\right)}\right) - \ln\left(n+\frac{1}{2}\right)\left(1-\frac{1}{2\left(n+\dfrac{1}{2}\right)}\right)\right] - 1$$

$$= \left(n+\frac{1}{2}\right)\left[\ln\left(1+\frac{1}{2\left(n+\dfrac{1}{2}\right)}\right) - \ln\left(1-\frac{1}{2\left(n+\dfrac{1}{2}\right)}\right)\right] - 1$$

$$= u\left[\ln\left(1+\frac{1}{2u}\right) - \ln\left(1-\frac{1}{2u}\right)\right] - 1,$$

where we have denoted $n + \frac{1}{2} = u$

Using now the well known expansions

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots \quad |x| < 1$$

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots \quad |x| < 1$$

(uniform convergent in every compact $K \subset (-1,1)$) and performing the usual calculations, we obtain

$$b_n = 2n\left(\frac{1}{2n} + \frac{1}{3}\frac{1}{(2n)^3} + \frac{1}{5}\frac{1}{(2n)^5} + \dots\right) - 1 = \frac{1}{12n^2} + \frac{1}{8n^4} + \dots > \frac{1}{12n^2}$$

(because of $n > 0$). Therefore (using that $\mathrm{e}^x > 1 + x$, for $x > 0$) we have

$$\frac{\left(1+\dfrac{1}{n}\right)^{n+1/2}}{\mathrm{e}} = a_n = \mathrm{e}^{b_n} > \mathrm{e}^{\frac{1}{12u^2}} > 1 + \frac{1}{12u^2}$$

and so

$$\left(1+\frac{1}{n}\right)^{n+1/2} > \mathrm{e}\left(1 + \frac{1}{12\left(n+\dfrac{1}{2}\right)^n}\right),$$

that gives (3.2).

The problem of finding of an adequate constant β in (3.1) remains open.

4. Concluding remarks

The previous results, concerning the approximation of the number e by the sequence $\left(1+\frac{1}{n}\right)^{n+p}$ conduct to the idea to search a similar approximation of the exponential. We mention that an approximation of the exponential using the rational functions was given by J. Karamata (see [2]).

Acknowledgments. I thank Professor Ioan Gavrea for his communication of the inequality (3.2) and the proof.

References

[1] Gavrea, I., *Private communication to the author*, Cluj-Napoca, September 24, 2010.

[2] Karamata, J., *Sur l'approximation de* e^x *par des fonctions rationelles*, Bull. Soc. Math. Phys. Serbie, **1**(1949), no. 1, 7-19.

[3] Mitrinović, D.S., *Analytic Inequalities*, Springer-Verlag, Berlin Heidelberg New York, 1970.

[4] Neville, E.H., *Note 1209: Two inequalities used in the theory of the gamma functions*, Math. Gazette, **20**(1936), 279-280.

[5] Neville, E.H., *Note 1225: Addition to the Note 1209*, Math. Gazette, **21**(1937), 55-56.

[6] Niculescu, C.P., Vernescu, A., *On the order of convergence of the sequence* $\left(1-\frac{1}{n}\right)^n$, (Romanian), Gaz. Mat., **109**(2004), no. 4.

[7] Niculescu, C.P., Vernescu, A., *A two-sided estimate of* $\mathrm{e}^x - (1+\frac{x}{n})^n$, JIPAM, **5**(2004), no. 3, art. 55.

[8] Pólya, G., Szegö, G., *Problems and Theorems in Analysis I*, Springer-Verlag, Berlin Heidelberg New York, 1978.

[9] Sinnadura, J. St.-C., *Applications of the inequality of the means*, Math. Gazette, **45**(1961), no. 354, 330-331.

[10] Vernescu, A., *An inequality related to the number e*, (Romanian), Gaz. Mat., **87**(1982), no. 2-3, 61-62.

[11] Vernescu, A., *An simple proof of an inequality related to the number e*, (Romanian), Gaz. Mat., **93**(1988), no. 5-6, 206-207.

[12] Vernescu, A., *The Number e and the Mathematics of the Exponential*, (Romanian), Publishing House of the University of Bucharest, 2004.

[13] Watson, G.N., *An inequality associated with gamma function*, Messenger Math., **45**(1916), 28-30.

[14] Watson, G.N., *Note 1254: Comments on Note 1225*, Math. Gazette, **21**(1937), 292-295.

[15] Whittaker, E.T., Watson, G.N., *A Course of Modern Analysis*, Cambridge Univ. Press, 1952.

Andrei Vernescu
"Valahia" University of Târgovişte
Department of Sciences
Bd. Unirii 18
130082 Târgovişte
Romania
e-mail: avernescu@clicknet.ro