ISSN 2346-8092

ა. რაზმაძის მათემატიკის ინსტიტუტის შრომები

ივანე ჯავახიშვილის სახელობის თბილისის
სახელმწიფო უნივერსიტეტი

ტომი 172, N 3, ნაწილი I, 2018

Transactions of A. Razmadze Mathematical Institute is a continuation of Travaux de L' Institut Mathematique de Tbilisi, Vol. 1–15 (1937–1947), Trudy Tbilisskogo Matematicheskogo Instituta, Vol. 16–99 (1948–1989), Proceedings of A. Razmadze Mathematical Institute, Vol. 100–169 (1990–2015).

Transactions of A. Razmadze Mathematical Institute
Volume 172, Issue 3 Part A, December 2018

Contents

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 293–298
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Editorial

Professor Gvanji Mania (1918–1985)

Professor Gvanji Mania was born in the village of Etseri, Georgia on May 29, 1918. His father, Mikheil Mania, was a Russian language teacher and his mother, Fedosi was daughter of clergyman. It is noteworthy that Professor Mania's maternal grandfather and two of his uncles served at Saint George's church in the village of Jvari.

From 1932 to 1935 Mania studied at Zugdidi Pedagogical College and immediately after his graduation he entered the Department of Physics and Mathematics of Tbilisi State University. From 1940 to 1945 he worked as Assistant Professor at Zugdidi Pedagogical Institute and at the same time, from 1943 to 1946, as an Assistant Professor at Tbilisi Institute of Railway Engineers. In 1945–1946 he was a higher school inspector at the Ministry of Education of Georgia. In 1945 he was awarded the medal "For labor valor during the Great Patriotic War".

From 1946 till 1949 Gvanji Mania studied at the Moscow Potemkin Pedagogical Institute as a postgraduate student. His research supervisor was a well-known mathematician Professor Smirnov. Smirnov offered him to study problems similar to those Prof. Smirnov was working on together with Academician Kolmogorov. Mania had to compare not just the entire empirical line with the theoretical law of distribution, but only a certain a priori fixed part of this line — to the respective part of the theoretical line. The relevance of the stated problem was due to the fact that empirical data often contain unreliable observations, which, as a rule, are found at the extreme intervals of the distribution line and therefore break fitting on these intervals. Since such observations do not generally define the phenomenon, it is reasonable to omit them when empirical and theoretical distributions are compared. On October 3, 1949 G. Mania defended his thesis *Statistical Estimation of Distribution Law* for Candidate's degree at the Scientific Council of the Department of Physics and Mathematics at Potemkin Institute, his thesis gained a high appreciation of the opponents.

The official opponents of the thesis were Academician Boris Gnedenko from Kiev and Professor Liapunov. After the applicant's speech Academician Gnedenko said: "Glivenko, Kolmogorov and Smirnov always point out to the drawbacks of their theorems. More accurate facts are necessary here – we should perform estimation not on the whole numerical axis, but at points where large deviations can be observed. G. Mania, under Smirnov's guidance, investigated just these particularly interesting and important problems. The results obtained in the thesis are of primary importance (Gnedenko uses the phrase:"первоклассного значения"), and I think that this topic can make a subject of a

Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.09.001

doctoral dissertation. The obtained results – the two beautiful theorems – should be published as soon as possible and included in statistics manuals. It is advisable to prepare this thesis for publication".

The second official opponent, Professor Liapunov, noted: "As a result of calculations, the author obtained boundary laws of distribution both for the first and the second deviation. It is obvious that these two theorems will enter the gold fund of mathematical statistics (эти две теоремы войдут в золотой фонд математической статистики), while in a formal review he wrote: "можно смело сказать, что решения этих задач, прочно войдут в золотой фонд математических методов статистики" — undoubtedly, these theorems will enter the gold fund of mathematical methods in statistics"). Then he added: "I agree with Academician Gnedenko that extension of this topic will make a firm basis for a doctoral dissertation".

We shall present here a fragment from Professor N. Smirnov's, G. Mania's thesis supervisor's, speech: "I remember the year of 1946 when Gvanji Mania first appeared. He was very young then, but creative enthusiasm and love for science characterized each step he took. His Russian was rather poor then. We gave him Hammerstein's memoirs to read and were greatly astonished when this young man, who could hardly arrange Russian words into sentences while speaking, managed to reproduce very precisely heavy German phrases and present not only all basic ideas, but also all details and proofs in a brilliant way. Since the time G. Mania started working independently he introduced a number of different approaches, but some of them led to more difficult problems while others resulted in very long statements, and it was his intuition that made him choose the most convenient and useful method that would become a model for statement and proof of similar problems".

After the Russian period of his activity G. Mania came back to Georgia and from 1949 to 1950 worked as Assistant Professor at Gori Nikoloz Baratashvili Pedagogical Institute. From 1950 to 1953 he was an Assistant Professor at Georgian Polytechnic Institute. From 1955 to 1956 he became a Senior Researcher at Tbilisi Andrea Razmadze Institute of Mathematics.

One cannot overestimate Professor G. Mania's share in the foundation of new scientific centers, such as Computational Centre and Institute of Applied Mathematics. Hence it is quite natural and noteworthy that when during the celebration of the 40th anniversary of the Institute of Applied Mathematics (later – Academician I. Vekua Institute of Applied Mathematics) in 2009 a scientific conference dedicated to this event was held, it was decided that the session devoted to Professor G. Mania's 90th anniversary, prepared on the initiative and under the leadership of Georgian Mathematical Society, would take place just within the walls of this Institute. From 1956 to 1964 Prof. G. Mania was Deputy Director for Science at the Computational Center of Georgian Academy of Sciences (later — Academician Nikoloz Muskhelishvili Institute of Computational Mathematics) and from 1966 to 1972 he worked as Deputy Director for Science at The Institute of Applied Mathematics of TSU.

G. Mania's doctoral dissertation titled *Some Methods of Mathematical Statistics* was as successful as his Candidate's thesis. It was a result of his fruitful ten-year scientific studies. G. Mania defended his doctoral dissertation at A. Razmadze Institute of Mathematics in 1963. His official opponents were Academicians: Khvedelidze, Prokhorov and Sirazhdinov. In 1964 he was elected for the position of TSU Professor.

In 1963 G. Mania organized, basically singlehandedly, very large and important conference — All-Union Conference in Probability Theory and Mathematical Statistics. In Soviet Union, in 1963, the conference was allowed "10 participants from capitalist countries and 15 participants from socialist countries" — an unseen luxury for the times. H. Cramér, was a participant, and Martin Lóff, E. Parzen, J. Wolfowitz and M. Rosenblatt. David Kendall. Not just probabilists and statisticians, but some of the best specialists in the theory of functions and functional analysis, such as Prof. S. Stechkin, also participated. Prof. Yu. Linnik was there with some of his pupils, and young M. Stratonovich, who at that time, worked in approximation methods for PDE in Physics. A. Kolmogorov, along with his pupils and collaborators B. Gnedenko, A. Shityaev, Ya. Sinai, A. Borovkov, and others, formed the scientific 'core', but myriads of problems, small and large, have been laid on shoulders of one young, not yet professor, person — Gvanji Mania.

Some 20–25 years later, and more, colleagues everywhere remembered the Conference as a great and joyful event they experienced.

The foundations of studies in probability theory and mathematical statistics were laid by the first Georgian mathematician, one of the founders of Tbilisi University — Professor Andrea Razmadze (1889–1929). He was a lecturer at the newly established Tbilisi University, while Gvanji Mania (1918–1985) was his successor developing this field of mathematics in Georgia. In 1968 under the direction of Professor G. Mania Probability Theory and Mathematical Statistics Chair was founded at Tbilisi State University, the head of which he remained till the end of his life. This year we celebrate both Professor Mania's centenary and the 100th anniversary of his native university

and the 50th anniversary of the chair he founded (today the Head of the Chair is Professor Elizbar Nadaraya, Member of Georgian Academy of Sciences). At the same time in the period, 1973–1983 Professor G. Mania was Head of the Sector at the Institute of Economics and Law of Georgian Academy of Sciences, while since 1983 he was Head of the Sector of Probability Theory and Mathematical statistics at Tbilisi A. Razmadze Institute of Mathematics.

Professor G. Mania was actively engaged in scientific and pedagogical work. He is an author of more than 50 scientific works. It was G. Mania who laid the foundations for the development of Probability Theory and Mathematical Statistics as a branch of mathematics in Georgia. He is the author of first Georgian manuals and a number of monographs in this field. As a result of his activity since the 50th of the last century teams of scientists were formed studying problems of probability theory and mathematical statistics and solving both theoretical and practical problems using probabilistic and statistical methods. Under the direction of Professor Mania 10 Master's Theses were prepared.

Professor G. Mania took an active part in the work and organization of a number of All-Union and international conferences and symposia. In 1969 he participated in the work of the 37th session of the International Institute of Statistics in London and in 1970 he was delegated to the International Congress of Mathematicians held in Nice, France.

Under G. Mania's leadership All-Union Winter School in Probability Theory and Mathematical Statistics was yearly held in Bakuriani, Georgia, in the course of 20 years. It soon became International since it was regularly attended by famous foreign scientists. In 1982 under Professor G. Mania's direct supervision Tbilisi hosted the VI USSR–Japan Symposium in Probability Theory and Mathematical Statistics. After coming back home Academician Kolmogorov in a letter to Professor G. Mania wrote: "Thank you very much for all your efforts in Tbilisi and Sukhumi. During my visit I was happy to witness and appreciate your major part in the progress of probability theory and mathematical statistics in Georgia".

Professor J. Mania was a member of various scientific societies and councils, including Georgian Mathematical Society, where he was a member of the Presidium, of the International Institute of Statistics, since 1969 — of International Bernoulli Society for Application of Statistics in Probability Theory and Mathematical Statistics, a member of American Mathematical Society, a member of the Editorial Board of the international "Statistics" Journal (published in Berlin). He got two government awards and Academician I. Javakhishvili Medal.

In 1989, to commemorate his 70th anniversary, a book of his works was issued titled *Probability Theory and Mathematical Statistics*, which entered the 92th volume of scientific articles published by A. Razmadze Institute of Mathematics of Georgian Academy of Sciences, where together with a number of other works by outstanding scientist, Academician Korolyuk's paper "Asymptotic Behavior of Mania's Statistics" was also published.

As we have noted earlier, Gvanji Mania's first works were written under Professor Smirnov's guidance, where Mania studied the maximum deviation behavior of the continuous distribution function $F(x)$ from the empirical distribution function $F_n(x)$ taken not on the entire axis, but only on the maximum growth interval of the function $F(x)$. He found the limit distribution of the following statistics:

$$D_n^+(\theta_1, \theta_2) = \sup_{x:\theta_1 \le F(x) \le \theta_2} (F_n(x) - F(x))$$

and

$$D_n(\theta_1, \theta_2) = \sup_{x:\theta_1 \le F(x) \le \theta_2} |F_n(x) - F(x)|,$$

where θ_1 are θ_2 given numbers, $0 \le \theta_1 < \theta_2 \le 1$ (later, he also tabulated the limit distribution of these statistics when $\theta_2 = 1 - \theta_1$). The sharp-witted proof of the above-mentioned results is based on Abel and Tauber type theorems, where errors made in Feller's similar theorems were eliminated. This result brought about immediate attention of specialists and it was often used by other scientists. In scientific literature these statistics are referred to as Mania's statistics (criteria).

In 1961 G. Mania introduced two independent normal sample homogeneity criteria based on the statistics:

$$L = \max_x |\Phi(\frac{x - \overline{x}_1}{s_1}) - \Phi(\frac{x - \overline{x}_2}{s_2})|,$$

where Φ is a standard normal distribution function while $\overline{x}_i$ and s_i are, respectively, empirical mean and variance constructed according to the n_i-size sample, $i = 1,2$. He showed that the limit distribution of the statistics

$$\frac{n_1 \cdot n_2}{n_1 + n_2} \cdot L$$

(when $n_1, n_2 \to \infty$) is independent of normal distribution parameters and is based on the limit distribution of Smirnov's statistics

$$n_1 \max_x |\Phi(\frac{x - \overline{x}_1}{s_1}) - \Phi(x)|.$$

In the same year G. Mania introduced the two-dimensional distribution density

$$f_n(x, y) = \frac{\Delta_{h_1} \Delta_{h_2} F_n(x, y)}{4h_1 h_2}.$$

He found the optimal value of h_1 and h_2 in the sense of integral square deviation and showed that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E(f_n(x, y) - f(x, y))^2 dxdy \approx cn^{-\frac{2}{3}},$$

where c, in a certain sense, depends on the second-order derivative of $f(x, y)$.

After that G. Mania studied the properties of normal distribution density nonparametric estimate. In particular, that of density parametric estimate $n(x|a, C)$ of the k-dimensional normal distribution (by the mean and covariance matrix C) for the square error

$$\Phi_n = \int_{R^k} [n(x|a, C) - n(x|\overline{a}, \overline{C})]^2 dx.$$

He found the limit distribution:

$$P\{n2^{k+3}\pi^{k/2}\sqrt{\det C}\,\Phi_n < u\} \to G(u),$$

where $G(u)$ coincides with a certain type of square distribution of normal values. Thus G. Mania's above-mentioned results imply (the same as the book by L. Devroye and L. Gyofhi) that it is impossible to improve the famous Boyd and Still's Theorem.

Professor G. Mania's above mentioned results and a number of his works and studies in Estimation Theory are published in the monograph *Statistical Estimates of Probability Distribution* issued in 1974, which deserved specialists' appreciation. Before the monograph was published Academician Gnedenko wrote in his review of the manuscript: "В математическом отношении рукопись выполнена безупречно. Она после опубликования, несомненно оживит интерес к тому направлению исследований, которое представляет автор"("As far as mathematics is concerned the manuscript is perfect. After its publication it will undoubtedly enliven the interest in the field of mathematics the Author presents"). After the publication of the monograph on density statistical estimation in *International Statistical Review* in 1979 in Werz and Schneider Reference Book G. Mania's 17 works are mentioned and *International Statistical Review* calls the above-mentioned book "an excellent book in the given field".

The monograph *Some Methods of Mathematical Statistics*, which appeared in Georgian in 1963 together with the book *Mathematical Statistics in Technology*, issued in 1985 was of primary importance for Georgian scientists and engineers enabling them to become aware of certain probabilistic and statistical methods, described in their native language, and apply them for the solution of some practical problems. For years specialists in different fields used to come to Professor G. Mania's Chair to consult on the practical application of probabilistic and statistical methods for the solution of various problems. Among them there were doctors, biologists, engineers, members of the administration of Rustavi Metallurgical Factory and others. His younger colleagues also participated in this activity, and they remember clearly Professor G. Mania's qualified help he rendered to those specialists in different fields.

Professor G. Mania's last studies were devoted to problem of estimation of parameters of stable distributions, and to the investigation of infinitely divisible and stable distributions with random number of summands.

G. Mania's role as a teacher and a tutor cannot be overestimated. His students and younger colleagues always felt his constant support. Caring for them was an important part of his life. The success achieved by contemporary Georgian scientists in the field of probability theory and mathematical statistics largely owes to Professor G. Mania's dedication and help.

Professor G. Mania's wife, Mrs. Irina Nodia, was a scholar specializing in Byzantian Studies. Their son Michael Mania is Head of the Department of Probability Theory and Mathematical Statistics at the A. Razmadze Institute of Mathematics. Their daughter Maia Mania is an Architectural Historian and a Professor at the Tbilisi State Academy of Arts. Both are married.

Professor G. Mania died on March 16, 1985 at the age of 67.

A. Razmadze Institute issued a collection of articles (1989) titled *Theory of Probability and Mathematical Statistics* to celebrate G. Mania's 70th Anniversary.

In 2008 under the auspices of Georgian Mathematical Society G. Mania's 90th Anniversary was celebrated.

In 2013 for his 95th Anniversary lecture hall number 335 of the XI Building of Tbilisi I. Javakhishvili State University was given Professor G. Mania's name.

In the current 2018, to celebrate G. Mania's centenary, Georgian Statistical Association Office decided to establish Professor G. Mania Scholarship at Tbilisi I. Javakhishvili State University.

Main publications

(i) Monographs

1. Some methods of mathematical statistics. (Georgian) Publishing house of Georgian Academy of Sciences, Tbilisi, 1963, pp. 351.
2. Statistical estimation of probability distributions. (Russian) Tbilisi University Press, 1974, pp. 240.
3. Probability theory. (Georgian) Publishing house of Ministry of Education, Tbilisi, 1954, pp. 240.
4. Mathematical statistics in technics. (Georgian), Sabchota Sakartvelo, Tbilisi, 1958, pp. 345.
5. The course of probability theory. (Georgian) Tbilisi University Press, Tbilisi, 1962, pp. 340.
6. Linear programming. (Georgian), "Ganatleba", Tbilisi, 1967, pp. 295.
7. The course of high mathematic. (Georgian) Tbilisi State University, Tbilisi, 1967, pp. 498. (with P.Zeragia)
8. Ilia Vekua. (Georgian) Publishing house of Tbilisi State University, 1967, pp. 75. (with B.Hvedelidze)
9. Probability theory and mathematical statistic. (Georgian) Publishing house of Tbilisi State University, 1976, pp. 350.
10. A book of problems in probability theory and mathematical statistic. (Georgian). Publishing house of Tbilisi State University, 1976, pp. 120. (with A.Ediberidze and N.Anthelava)

(ii) Selected Publications

11. Generalization of A.N. Kolmogorov's criterion for the estimation of distribution laws by empirical data. (Russian) Dokl. Akad. Nauk SSSR, 69(1949), No. 4, 495–497.
12. Statistical estimation of distribution laws. (Russian) Uchenie Zapiski MGPI imeni V.P. Potiomkina 16 (1951), 17–63.
13. Practical applications of an estimation of a maximum of two-sided deviations of empirical distribution in a given interval of growth of a theoretical law. (Russian) Soobshch. AKad. Nayk GSSR,14(1953), No. 9, 521–524.
14. Practical applications of an estimation of a maximum of one-sided deviations of an empirical distribution in a given interval of growth of a theoretical law. (Russian) Proc. of Georgian Polytechnical Institute, 30(1954), No. 9, 89–92.
15. Square estimation of divergence of normal densities by empirical data. (Georgian) Soobshch. AN GSSR, 17 (1956), No. 3, 201–204.
16. Square estimation of normal distribution densities by empirical data (Russian). Trydy Vsesojuznogo Mat. Siezda, IZD. AH SSSR, 1 (1956), 124–125.
17. Quadratic error of an estimation of twodimensional normal density by empirical data. (Russian) Soobshch. AN GSSR, 20 (1958), No. 6, 655–658.
18. Quadratic error of the estimation of normal density by empirical data. (Russian) Tr. Vychisl. Tsentra AH GSSR, 1 (1960), 75–96.
19. On one method of constructing of confidence regions for two samples from general population. (Russian) Soobshch. AN GSSR, 27 (1961), No. 2, 137–142.
20. Remark on non-parametric estimations of twodimensional densities. (Russian) Soobshch. AN GSSR, 27 (1961), No. 4, 385–390.
21. Square estimation of divergence of twodimensional normal distribution densities by empirical data. (Russian) Tr. BC AH GSSR, 2 (1961), 153–211.
22. Square estimation of divergence of twodimensional normal distribution densities by empirical data (Russian) Proc. of 6th Vilnius Conference in Probab. Theory and Math. Stat., (1962), 407–409.
23. Quadratic error of an estimation of densities of normal distributions by two samples (Russian) Trudy. BC AH SSSR, 4 (1963), 213–216.

24. Hypothesis testing of identity of distributions of two independent samples. (Russian) Tr. Vichisl. Tsentra AH GSSR, 7 (1966), 1–34.

25. Square estimation of divergence of densities of multidimensional normal distribution by empirical data (po dannim viborki) (Russian) Proc. of Tbilisi State University, 129 (1962), 373–382.

26. Quadratic error of the estimation of multidimensional normal distribution densities by empirical data. (Russian) Soobshch. AN GSSR, 52 (1968), No. 1, 27–30.

27. Quadratic error of the estimation of multidimensional normal distribution densities by empirical data (Russian) Probability Theory and Appl., 13 (1968), No. 2, 359–362.

28. Quadratic error of an estimation of densities of multidimensional normal distribution by empirical data. (Russian) Probability Theory and Appl., 14 (1969), No. 1, 151–155.

29. Quadratic error of an estimation of densities of multidimensional normal distribution by empirical data (Russian) Proc. of Tbilisi State University, 2 (1969), 223–227.

30. Quadratic error of an estimation of densities of multidimensional normal distribution by empirical data. Congres international des Mathematiciens, Nice, Paris, (1970), Abstracts 260.

31. Quadratic error of an estimation of normal distribution densities by several samples (Russian) Soobshch. AN GSSR, 67 (1972), No. 2, 301–304.

32. One approximation of distributions of positive defined quadratic forms of normal random variables, (Russian).Soobshch. AN GSSR, 107 (1982), No. 2, 241–244.(with E.Khmaladze and V.Felker)

33. On the estimation of parameters of type of stable laws, Proceedings of the first International Tampere Seminar on linear Statistical Models and their Applications (1983) Tampere University, 1985, pp. 202–223. (with L.Klebanov and I.Melamed)

34. One problem of V.M.Zolotarev and analogue of infinitely divisible and stable distributions in the scheme of the sum of random number of random variables. (Russian) Probability Theory and Appl., 29 (1984), 757–760. (with L.Klebanov and I.Melamed)

Further reading

[1] Yu. V. Prokhorov, A.N. Shiryaev, T.L. Shervashidze, Preface (Russian) to Vol. 92 of Proc. Tbilisi A Razmadze Math. Institute dedicated to 70th birthday of GM Mania, 1989, pp. 1-8.

[2] T. Shervashidze, Probability Theory and Mathematical Statistics (Russian) in 50th anniversary of Tbilisi A Razmadze Math. Institute, Metsniereba, Tbilisi, 1985.

E. Nadaraya*
Iv. Javakhishvili Tbilisi State University, Department of Mathematics, Tbilisi, Georgia
E-mail address: elizbar.nadaraya@tsu.ge.

O. Purtukhia
Iv. Javakhishvili Tbilisi State University, A. Razmadze Mathematical Institute, Department of Mathematics, Tbilisi, Georgia
E-mail address: o.purtukhia@gmail.com.

Available online 15 September 2018

* Corresponding editor.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 299–308
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Approximate solution for solving fractional Riccati differential equations via trigonometric basic functions

Bahram Agheli

Department of Mathematics, Qaemshahr Branch, Islamic Azad University, Qaemshahr, Iran

Received 14 June 2018; received in revised form 12 July 2018; accepted 9 August 2018
Available online 22 August 2018

Abstract

In this paper, a method has been proposed to finding a numerical function for the Riccati differential equations of non integer order (FRDEs), in which trigonometric basic functions are used. First, by defining trigonometric basic functions, we define the values of the transformation function in relation to trigonometric basis functions (TBFs). Following that, the numerical function is defined as a linear combination of trigonometric base functions and values of transform function which is named trigonometric transform method (TTM), and the convergence of the method is also presented. To get a numerical solution function with discrete derivatives of the solution function, we have determined the numerical solution function which satisfies the FRDEs. In the end, the algorithm of the method is elaborated with several examples. Numerical results obtained show that the proposed algorithm gives very good numerical solutions. In one example, we have presented an absolute error comparison of some numerical methods.

Keywords: Trigonometric transform; Riccati differential equations; Basic functions; Caputo derivative

1. Introduction

Fractional calculus and arithmetics have a long history which dates back to the birth of the classical calculus. However, during these years, some research works have been carried out but in the last few decades, this new calculus together with dynamic equation has gained more popularity. There are several classes of fractional derivative, but the most prevalent definitions are Riemann–Liouville and Caputo fractional derivatives. The former has an abstraction mathematically but the latter is mostly used by engineers.

Research of fractional differential equations has noticeable flourish in recent years that indicates the significance and station of the fractional calculus in sciences and engineering. On the other hand, because of extensive applications of fractional calculus in natural phenomena such as population growth model [1], fluid flow [2], dynamical problems [3,4], chemistry [5], mathematical physics [6], economics [7], traffic model [8] and medicine [9] and so

E-mail address: b.agheli@qaemiau.ac.ir.
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.08.002

on. There are some books of fractional calculus. All researchers and scholars are demanded to read and investigate books that have been written to take advantage of calculus and arithmetic of non integer order [10–12].

Many researchers have used numerical methods for the purpose of solving the fractional Riccati differential equations (FRDEs) [13–20].

The main objective in this paper is to offer a simple method in which it is possible to apply trigonometric transform method (*TTM*) to tackle with the FRDEs of the following form

$$D_t^{\alpha} u(t) - \sum_{i=0}^{2} p_i(t)\, u^i(t) = 0, \quad 0 < \alpha \leq 1, \quad 0 < t \leq T, \quad u(0) = u_0, \tag{1.1}$$

where $p_i(t)$ are constant functions and $t \in \mathbb{R}$. For time the Caputo differential operator of order α is given by [11] featuring $m - 1 < \alpha \leq m$ and $m \in \mathbb{Z}^+$

$$D_t^{\alpha} u(t) = \frac{1}{\Gamma(m-\alpha)} \int_0^t (t-s)^{\alpha-1} u^{(m)}(s)\, ds. \tag{1.2}$$

There are some more differential operator related to fractional derivative. The interested readers can see [21–23].

Historically by two researchers James Bernoulli (1654–1705) and Count Jacopo Francesco Riccati (1676–1754) a special case of differential equations (1.1) was introduced and evaluated. On the importance and motivation for this differential equation, it should be noted that it has a key role in many of the physical phenomena. Such applications can include control systems, robust stabilization, diffusion problems, network synthesis, optimal filtering, stochastic theory, controls, financial mathematics, optimal control, river flows, robust stabilization, network synthesis and financial mathematics dynamic games, linear systems with Markovian jumps, stochastic control, econometric models and invariant embedding noted that the use of the Riccati differential equations (RDEs) [24–31]. Of other uses, the one dimensional static Schrödinger equation [32] and the traveling wave solutions of a nonlinear partial differential equation [33] are noteworthy with the Riccati differential equations (FRDEs) featuring fractional derivatives.

Many researchers have used numerical methods for the purpose of solving the RDEs and FRDEs. We can refer to a number of familiar methods, including differential transform method [13], Series solutions Adomian's decomposition method [14], Homotopy perturbation method [14], Variational iteration method [15], Homotopy analysis method [16] and etc [17–20].

This paper is organized as what follows: in Section 2, discretization of the fractional derivative is given. In Section 3, we have expressed the trigonometric Basic functions (*TBFs*). In Section 4, a description of the new approach that is named trigonometric transform method (TTM) is presented. Some numerical examples are offered in Section 5. And conclusions are drawn in Section 6.

2. Discretization of the fractional derivative

In this section, we introduce discretization of the fractional derivative. The approximation of derivatives by forward differences is one of the most basic tools in finite difference methods for the numerical solution of differential equations, especially initial value problems. The n-th order forward difference is given by

$$u^{(m)}(t) \approx \frac{1}{h^m} \sum_{i=0}^{m} (-1)^i \binom{n}{i} u\left((n-i)h + t\right), \quad m \in \mathbb{N},$$

Depending on the application, the spacing h may be variable or constant. In this paper, we consider $\tau = t_{j+1} - t_j$ and $t_j = a + j\,\tau$ for $j = 0, 1, 2, \ldots$.

Utilizing the approximation for the Caputo derivative [34] of Eq. (1.2) we have:

$$D^{\alpha} u(t_{k+1}) \approx \frac{1}{\tau^{\alpha}\Gamma(2-\alpha)} \sum_{j=0}^{k} \left(u(t_{j+1}) - u(t_j)\right)\left((k-j+1)^{1-\alpha} - (k-j)^{1-\alpha}\right), \tag{2.1}$$

in which $0 < \alpha \leq 1$ and $u(t_0)$ is known.

3. Trigonometric basic functions (*TBFs*)

In this section, we introduce the trigonometric basis functions and properties that are used in the main sections of the paper to numerical the function of the solution.

Definition 3.1. Presuming that for $n \geq 1$, $a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b$ be specified nodes, we express that basic functions $T_0, T_1, \ldots, T_n$ are defined on $[a, b]$ with their trigonometric functions $T_0, T_1(t), \ldots, T_n(t)$, as follows:

$$T_0(t) = \begin{cases} 0.5\left(1 + cos\frac{\pi}{h_0}(t - t_0)\right), & t_0 \leq t \leq t_1 \\ 0, & \text{otherwise,} \end{cases}$$

$$T_k(t) = \begin{cases} 0.5\left(1 + cos\frac{\pi}{h_{k-1}}(t - t_k)\right), & t_{k-1} \leq t \leq t_k, \\ 0.5\left(1 + cos\frac{\pi}{h_k}(t - t_k)\right), & t_k \leq t \leq t_{k+1}, \quad k = 1, 2, 3, \ldots, n-1, \\ 0, & \text{otherwise,} \end{cases} \tag{3.1}$$

$$T_n(t) = \begin{cases} 0.5\left(1 + cos\frac{\pi}{h_{n-1}}(t - t_n)\right), & t_{n-1} \leq t \leq t_n \\ 0, & \text{otherwise,} \end{cases}$$

in which $h_k = t_{k+1} - t_k$ for $k = 0, 1, \ldots, n-1$.

Remark 3.2. The trigonometric functions introduced in Definition 3.1 are the trigonometric basis functions (*TBFs*) in which the following properties are satisfied.

(1) T_k of $[a, b]$ to $[0, 1]$ is continuous, $\sum_{k=0}^{n} T_k(t) = 1$ for all $t \in [a, b]$ and $T_k(t_k) = 1$, $k = 0, 1, 2, \ldots, n$.
(2) $T_k(t) = 0$ if $t \notin (t_{k-1}, t_{k+1})$, for $k = 1, 2, \ldots, n-1$, $T_0(t) = 0$ if $t \notin (t_0, t_1)$ and $T_n(t) = 0$ if $t \notin (t_{n-1}, t_n)$.
(3) On subinterval $\left[t_{k-1}, t_{k+1}\right]$, for $k = 1, 2, \ldots, n-1$, $T_k(t)$, certainly is an increasing function on $[t_{k-1}, t_k]$ and decreasing function on $\left[t_k, t_{k+1}\right]$.
Basic functions are called uniform as long as $t_{k+1} - t_k = h = \frac{b-a}{n}$ and two additional properties coincide:
(4) $T_k(t_k - t) = T_k(t_k + t)$, for all $t \in [0, h]$, for $k = 1, 2, \ldots, n-1$,
(5) $T_k(t) = T_{k-1}(t - h)$ and $T_{k+1}(t) = T_k(t - h)$, for $k = 1, 2, \ldots, n-1$, and $t \in [t_k, t_{k+1}]$.

Definition 3.3. Let f be a function belonging to $C([a, b])$ and T_i, $i = 0, 1, \ldots, n$, be the *TBFs* which buildup on $[a, b]$. We define the F_k that is the transform of function f on $[a, b]$ with respect to basic functions T_k given by

$$F_k = \frac{\int_a^b f(t)T_k(t)dt}{\int_a^b T_k(t)dt}, \quad k = 0, 1, 2, \ldots, n. \tag{3.2}$$

Definition 3.4. Let f be a function belonging to $C([a, b])$ and T_i, $i = 0, 1, \ldots, n$, be the *TBFs* which buildup on $[a, b]$ and F_k be transform of function f on $[a, b]$ with respect to basic functions T_k. Then,

$$f_n(t) = \sum_{k=0}^{n} F_k T_k(t),$$

is numerical of function f on $[a, b]$ with respect to *TBFs*.

Theorem 3.5 (*Convergence*)**.** *Let f be a uniformly continuous function on $[a, b]$. Thus, for any $\epsilon > 0$, there exists n_ϵ such that for all $n \geq n_\epsilon$:*

$$\left|f(t) - f_{n_\epsilon}(t)\right| < \epsilon. \tag{3.3}$$

Proof. f is a uniformly continuous function on $[a, b]$; therefore,

$$\forall \epsilon > 0, \quad \exists \delta = \delta(\epsilon); \quad |x - t| < \delta \Rightarrow |f(x) - f(t)| < \epsilon \ \ (0 < \delta < \epsilon).$$

For all $\epsilon > 0$, we have

$$|f(t) - f_n(t)| = \left|\sum_{i=0}^{n} T_i(t)f(t) - \sum_{i=0}^{n} F_i T_i(t)\right| \leq \sum_{i=0}^{n} T_i(t)\,|f(t) - F_i| < \epsilon.$$

It is sufficient to show that $|f(t) - F_i| < \epsilon$. Let $x,\ t \in \left[x_{i-1}, x_{i+1}\right]$, $i = 1, 2, \ldots, n-1$, so that we can evaluate

$$|f(x) - F_i| = \left| f(x) - \frac{\int_a^b f(t)T_i(t)dt}{\int_a^b T_i(t)dt} \right| \leq \frac{\int_{x_{i-1}}^{x_{i+1}} T_i(t)\,|f(x) - f(t)|\,dt}{\int_{x_{i-1}}^{x_{i+1}} T_i(t)dt} < \epsilon,$$

iff $\delta < 2h < \epsilon$ or $h < \frac{\epsilon}{2}$.

Regarding $h = \frac{b-a}{n}$, it is sufficient that $n_\epsilon > \frac{2(b-a)}{\epsilon}$. □

For description of fractional derivative, we have the following proposition.

Proposition 3.6. *With substituting* $f_n(t) = \sum_{k=0}^{n} F_k\, T_k(t)$ *in Eq.* (2.1)*, we will have the next equation for* $k = 0, 1, 2, \ldots, n-1$*:*

$$D^\alpha f_n(t_{k+1}) \approx \frac{1}{\tau^\alpha \Gamma(2-\alpha)} \sum_{j=0}^{k} (F_{j+1} - F_j)\left((k-j+1)^{1-\alpha} - (k-j)^{1-\alpha}\right), \quad 0 < \alpha \leq 1. \tag{3.4}$$

4. Description of the new approach

Let solution of (1.1) be continuous on $[0, b]$. To gain numerical solution of $u(x)$, we divide $[0, b]$ to n equal partitions with step length τ:

$$t_0 = 0,\ \ t_i = t_0 + i\tau,\ \ i = 0, 1, \ldots, n,\ \ \tau = \frac{b}{n}. \tag{4.1}$$

Considering the trigonometric functions with regard to Definition 3.1 on $[0, b]$ and Definition 3.4, we can gain numerical function $u(x)$ by $u_n(x) = \sum_{k=0}^{n} U_k\, T_k(t)$. It is evident that for calculating $u_n(t)$, $t \in [0, b]$, we should calculate U_k, $k = 0, 1, 2, \ldots, n$.

In order to gain the numerical solution of Eq. (1.1), $u_n(t)$ for points $t_0, t_1, \ldots, t_n$ must be satisfied in (1.1). Due to the boundary conditions (1.2), $u_n(t_0) := u(t_0) = u_0$ and for other points $t_1, t_2, \ldots, t_n$, we have

$$D^\alpha u_n(t_{k+1}) - \sum_{i=0}^{2} p_i(t_{k+1})\, u_n^i(t_{k+1}) = 0, \quad k = 0, 1, 2, \ldots, n-1. \tag{4.2}$$

in which $m - 1 < \alpha \leq m$ and $m \in \mathbb{Z}^+$.

Using 3.6, Eq. (4.2) converts to the following form for $k = 0, 1, 2 \cdots, n-1$:

$$\frac{1}{\tau^\alpha \Gamma(2-\alpha)} \sum_{j=0}^{k} (U_{j+1} - U_j) \times \left((k-j+1)^{1-\alpha} - (k-j)^{1-\alpha}\right) = p_0(t_{k+1}) + p_1(t_{k+1})U_{k+1} + p_2(t_{k+1})U_{k+1}^2, \tag{4.3}$$

in which $0 < \alpha \leq 1$, $k = 0, 1, 2, \ldots, n-1$ and with boundary conditions $U_0 = u(0)$.

Now, using the boundary condition, we can calculate $U_1, U_2, \ldots, U_n$ by the obtained recursive equation (4.3) and then gain the numerical solution $u(t) \approx u_n(t)$ for Eq. (1.1).

In order to gain numerical solution of FRDEs, an algorithm by this method is offered in the subsequent algorithm.

Algorithm 1. An algorithm for numerical solution of FRDEs.

Step 1. Input $p_0(t)$, $p_1(t)$, $p_2(t)$, n and b.
Step 2. Set $\tau \leftarrow \frac{b}{n}$.
Step 3. Locate $t_k \leftarrow k\,\tau,\ \ k = 0, 1, 2, \ldots, n$.
Step 4. Choose *TBFs* $T_k(t)$ toward $k = 0, 1, 2, \ldots, n$.

Table 1
Absolute error with various n, τ and $\alpha = 1$ in different values of t for Example 5.1.

n	τ	t	TTM	Exact	Absolute error
50	0.02	0.0	0.0	0.0	0.0
		0.2	0.198096	0.197375	0.000721077
		0.4	0.381671	0.379949	0.00172161
		0.6	0.539768	0.53705	0.00271865
		0.8	0.667427	0.664037	0.0033906
		1.0	0.765206	0.761594	0.00361174
500	0.002	0.0	0.0	0.0	0.0
		0.2	0.197773	0.197375	0.000397496
		0.4	0.380422	0.379949	0.000473272
		0.6	0.537449	0.53705	0.000399516
		0.8	0.664285	0.664037	0.000248308
		1.0	0.761671	0.761594	0.0000769757

Step 5. Set recursive equations

$$\frac{1}{\tau^{\alpha}\Gamma(2-\alpha)}\sum_{j=0}^{k}(U_{j+1}-U_j)\times \left((k-j+1)^{1-\alpha}-(k-j)^{1-\alpha}\right)=p_0(t_{k+1})+p_1(t_{k+1})U_{k+1}+p_2(t_{k+1})U_{k+1}^2,$$

where $0<\alpha\leq 1$, $k=0,1,2,\ldots,n-1$ and $U_0=u(0)$.

Step 6. Calculate every U_k, $k=1,2,\ldots,n$ of an equation of degree two.

Step 7. The approximate solution is

$$u_n(t)\approx\sum_{k=0}^{n}U_kT_k(t).$$

5. Examples

Now that it is easier to understand trigonometric transform, a number of examples will be given in this section and then will be calculated. These examples include FRDEs. In all these examples, software *Mathematica11* has been used for calculations and graphs.

Example 5.1. Consider the FRDEs for the first example [18]:

$$D_t^{\alpha}u(t)=1-u^2(t),\quad 0<t<1,\ \ 0<\alpha\leq 1, \tag{5.1}$$

with the primary condition $u_0=u(0)=0$ and the precise solution $u(t)=\frac{\exp(2t)-1}{\exp(2t)+1}$ for $\alpha=1$.

Following the *TTM*, according to what was formulated and presented in Section 4 for Eq. (5.1), we can calculate $U_1,U_2,\ldots,U_n$ and then gain the numerical solution $u_n(t)$ of (5.1). Table 1 and Fig. 1 shows comparison between the numerical solution and the exact of (5.1) with *TBFs* for Example 5.1 with value of $\alpha=1$.

In Table 1, it can be seen that by increasing the amount n and decreasing the amount τ, a more accurate answer can be achieved.

Table 2 shows comparison between the exact and the approximation solution (5.1) with *TTM* of test Example 5.1 for different values of α and t, $n=500$, $\tau=0.002$.

Comparison of the exact and the approximate solution can be seen for test Example 5.1 with different values of α, $n=500$, $\tau=0.002$ and various values of t, in Fig. 3

Example 5.2. Let the FRDEs for the second example [18]:

$$D_t^{\alpha}u(t)=1+2u(t)-u^2(t),\quad 0<t<1,\ \ 0<\alpha\leq 1, \tag{5.2}$$

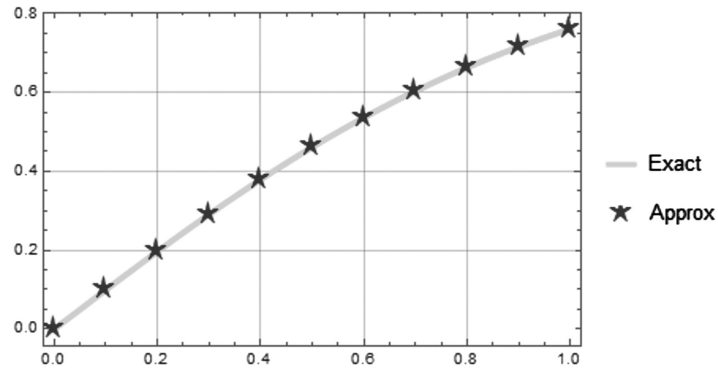


Fig. 1. Comparison between the exact and the numerical solution (5.1) with *TBFs* and $n = 500$ of Example 5.1 for value of $\alpha = 1$ and $\tau = 0.002$.

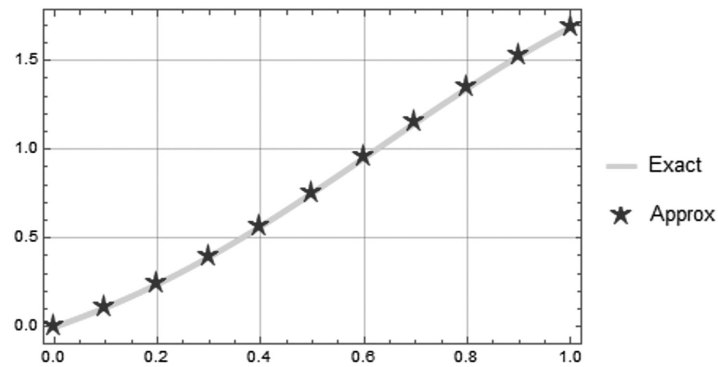


Fig. 2. Comparison between the exact and the numerical solution (5.2) with *TBFs* and $n = 500$ of Example 5.2 for value of $\alpha = 1$ and $\tau = 0.002$.

Table 2
The exact and the approximate result of test Example 5.1 featuring various values of α.

t	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1.0$	Exact
0.0	0.0	0.0	0.0	9
0.2	0.334626	0.260941	0.197773	0.197375
0.4	0.498466	0.442638	0.380422	0.379949
0.6	0.604588	0.577781	0.537449	0.53705
0.8	0.677429	0.67693	0.664285	0.664037
1.0	0.729503	0.749104	0.761671	0.761594

given that the primary condition

$$u_0 = u(0) = 0. \tag{5.3}$$

According to what was formulated and presented in Section 4 for Eqs. (5.2)–(5.3), with the help of the *TTM*, we get $U_1, U_2, \ldots, U_n$. Table 3 represents the present method and the achieved results of particle swarm optimization (PSO), modified homotopy perturbation method (MHPM), Chebyshev wavelets (CW), fractional variational iteration method (FVI), Legendre wavelets method (LWM) and Padé-variational iteration method (PVI) [18]. For $\alpha = 1$ in Fig. 2, we can view the precise and numerical answers via applying *TTM* featuring $n = 500$ and $\tau = 0.002$.

Regard the awareness that $\alpha = 1$, the numerical solution acquired via the offered method corresponds to the precise solution $u(t) = 1 + \sqrt{2}\tanh\left(\sqrt{2}t + \frac{1}{2}\log\left(\frac{\sqrt{2}-1}{\sqrt{2}+1}\right)\right)$.

Noteworthy in the values obtained in the column *TTM* of Table 3 is that by increasing the amount n, a more accurate answer for Eq. (5.4), can be achieved.

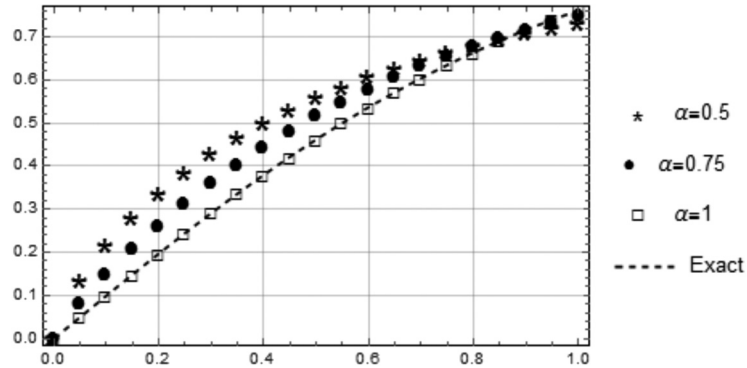


Fig. 3. Comparison betwixt the exact and the approximation solution with *TTM* of test Example 5.1 for value of $n = 500$, $\tau = 0.002$ and different values of α and t.

Table 3
Comparison of the numerical solutions of the equation in Example 5.2 with $\alpha = 1$.

t	SJOM	MHPM	PSO	CW	FVI	PVI	LWM	TTM	Exact
0.6	1.007291	1.370240	1.296320	1.349150	1.331462	1.873658	1.296302	0.953566	0.953653
0.7	1.253674	1.367499	1.416139	1.481449	1.497600	2.112944	1.416311	1.152949	1.15308
0.8	1.467499	1.794879	1.506936	1.599235	1.630234	2.260134	1.506913	1.346364	1.34655
0.9	1.629901	1.962239	1.569252	1.705303	1.724439	2.339134	1.569221	1.526911	1.52715
1.0	1.7872228	2.087384	1.605580	1.801763	1.776542	2.379356	1.605571	1.689498	1.68976
Total errors	7.14559	8.58224	7.39423	7.9369	7.96028	9.74772	7.39432	6.66929	

Table 4
Numerical results of Example 5.3 featuring $\alpha = 1$.

t	TTM	Exact	Absolute error
0.0	0.0	0.0	0.0
0.2	0.450065	0.450166	0.000101334
0.4	0.401178	0.401312	0.000134719
0.6	0.354203	0.354344	0.000140666
0.8	0.309897	0.310026	0.000128611
1.0	0.268837	0.268941	0.000104154

Example 5.3. I offer the FRDEs [18] for the third example:

$$D_t^{\alpha} u(t) - u(t) - u(t)^2 = 0, \quad 0 < t < 1, \ \ 0 < \alpha \leq 1, \tag{5.4}$$

including the primary condition

$$u_0 = u(0) = 0.5. \tag{5.5}$$

The unknown coefficient $U_1, U_2, \ldots, U_n$ with due attention to the *TTM*, according to Section 4 for Eqs. (5.4)–(5.5) is calculated.

Comparison of the exact and the numerical solution can be seen in Table 4 and Fig. 4 for Eq. (5.4) with $n = 500$, $\alpha = 1$, $\tau = 0.002$ and various values of t.

The precise solution $u(t) = \frac{\exp(-t)}{\exp(-t)+1}$ and the numerical solution for $\alpha = 1$, that we have calculated are in agreement.

Example 5.4. Consider the FRDEs for the fourth example:

$$D_t^{\alpha} u(t) = \frac{\alpha^2 t^{1-\alpha}}{\Gamma(3-\alpha)} - \frac{2\alpha t^{1-\alpha}}{\Gamma(3-\alpha)} - \frac{2t^{2-\alpha}}{\Gamma(3-\alpha)} - t^4 - 2\alpha t^3 - \alpha^2 t^2 - u^2(t), \ 0 < t < 1, \ 0 < \alpha \leq 1, \tag{5.6}$$

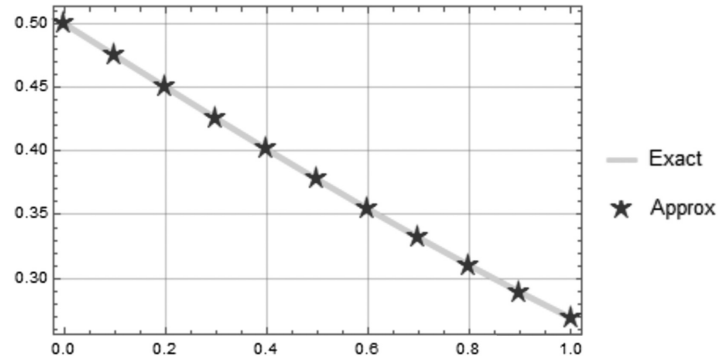


Fig. 4. Comparison of the numerical solution with *TBFs* of Example 5.3 for value of $\alpha = 1$ and various values of t.

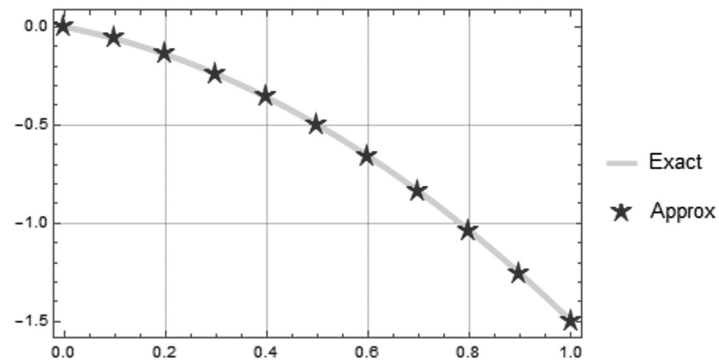


Fig. 5. Comparison of the numerical solution with *TBFs* of Example 5.3 for value of $\alpha = 0.5$ and various values of t.

Table 5
Numerical results of Example 5.4 featuring $\alpha = 0.5$.

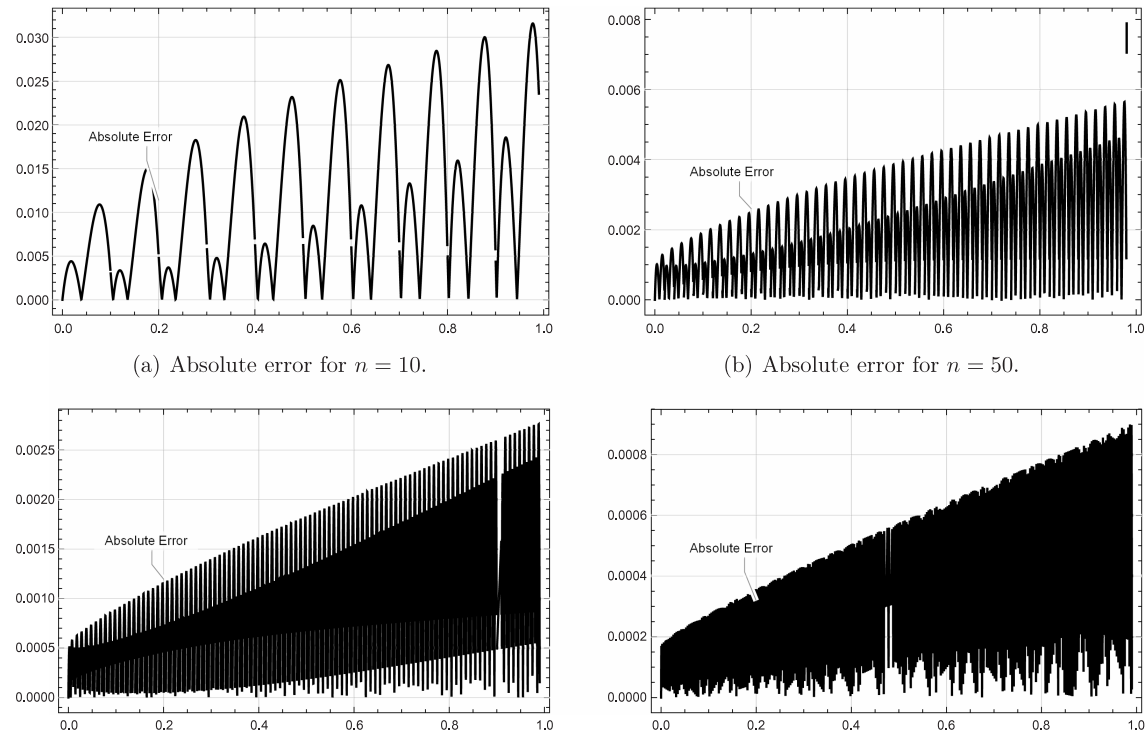
t	TTM	Exact	Absolute error
0.0	0.0	0.0	0.0
0.2	−0.140552	−0.14	0.000551698
0.4	−0.360675	−0.36	0.000675279
0.6	−0.660643	−0.66	0.000642591
0.8	−1.04054	−1.04	0.000541528
1.0	−1.26049	−1.26	0.000485106

with the primary condition $u_0 = u(0) = 0$ and the precise solution $u(t) = -t^2 - \alpha t$. Following the *TTM*, according to what was formulated and presented in Section 4 for Eq. (5.6), we can calculate $U_1, U_2, \ldots, U_n$ and then gain the numerical solution $u_n(t)$ of (5.6). Table 5 and Fig. 5 show comparison between the numerical solution and the exact of (5.6) with *TBFs* for Example 5.4 with value of $\alpha = 0.5$ and $n = 50$.

Fig. 6 shows the absolute error for $n = 10, 50, 100, 300$ with *TTM* for various values of $0 \leq t \leq 1$ and $\alpha = 0.5$.

6. Conclusion

I have proposed a method for finding an approximate function of time fractional Riccati differential equations (FRDEs), in which *TTM* are used. All examples with absolute and relative errors show that we have favorably applied trigonometric transform method *TTM* to obtain approximate solution of the FRDEs. The obtained solutions that are very close analytical solutions indicate that a little iteration of *TTM* will result in some useful solutions. As the result seems necessary from the authors' point of view, the suggested technique has the potentials to be practical in solving other similar ordinary differential equations *ODEs* and partial differential equations *PDEs* of non integer orders.



(a) Absolute error for $n = 10$.

(b) Absolute error for $n = 50$.

(c) Absolute error for $n = 100$.

(d) Absolute error for $n = 300$.

Fig. 6. Absolute errors for test Example 5.4.

References

[1] D. Baleanu, B. Agheli, M.A. Firozja, M.M. Al Qurashi, A method for solving nonlinear Volterra's population growth model of noninteger order, Adv. Difference Equ. 2017 (1) (2017) 368.

[2] C. Ming, F. Liu, L. Zheng, I. Turner, V. Anh, Analytical solutions of multi-term time fractional differential equations and application to unsteady flows of generalized viscoelastic fluid, Comput. Math. Appl. 72 (9) (2016) 2084–2097.

[3] A.S. Deshpande, V. Daftardar-Gejji, Y.V. Sukale, On Hopf bifurcation in fractional dynamical systems, Chaos Solitons Fractals 98 (2017) 189–198.

[4] A. Neamaty, M. Nategh, B. Agheli, Local non-integer order dynamic problems on time scales revisited. International Journal of Dynamics and Control, pp. 1-13.

[5] M.A.Z. Raja, R. Samar, E.S. Alaidarous, E. Shivanian, Bio-inspired computing platform for reliable solution of Bratu-type equations arising in the modeling of electrically conducting solids, Appl. Math. Model. 40 (11) (2016) 5964–5977.

[6] O. Guner, A. Bekir, The Exp-function method for solving nonlinear space–time fractional differential equations in mathematical physics, J. Assoc. Arab Univ. Basic Appl. Sci. (2017).

[7] E. Scalas, The application of continuous-time random walks in finance and economics, Physica A 362 (2) (2006) 225–239.

[8] A. Neamaty, M. Nategh, B. Agheli, Time–space fractional burger's equation on time scales, J. Comput. Nonlinear Dyn. 12 (3) (2017) 031022.

[9] R.L. Magin, O. Abdullah, D. Baleanu, X.J. Zhou, Anomalous diffusion expressed through fractional order differential operators in the Bloch–Torrey equation, J. Magn. Reson. 190 (2) (2008) 255–270.

[10] D. Baleanu, A.C. Luo, in: J.T. Machado (Ed.), Discontinuity and Complexity in Nonlinear Physical Systems, Springer, 2014.

[11] A.A. Kilbas, H.M. Srivastava, J.J. Trujillo, Theory and Application of Fractional Differential Equations, Elsevier B.V, Netherlands, 2006.

[12] I. Podlubny, Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of their Solution and Some of their Applications, Vol. 198, Academic press, 1998.

[13] J. Biazar, M. Eslami, Differential transform method for quadratic Riccati differential equation, Int. J. Nonlinear Sci. 9 (4) (2010) 444–447.

[14] S. Abbasbandy, Homotopy perturbation method for quadratic Riccati differential equation and comparison with Adomian's decomposition method, Appl. Math. Comput. 172 (1) (2006) 485–490.

[15] F. Geng, A modified variational iteration method for solving Riccati differential equations, Comput. Math. Appl. 60 (7) (2010) 1868–1872.

[16] Y. Tan, S. Abbasbandy, Homotopy analysis method for quadratic Riccati differential equation, Commun. Nonlinear Sci. Numer. Simul. 13 (3) (2008) 539–546.

[17] C. Bota, B. Căruntu, Analytical approximate solutions for quadratic Riccati differential equation of fractional order using the Polynomial Least Squares Method, Chaos Solitons Fractals (2017).
[18] A. Neamaty, B. Agheli, R. Darzi, The shifted Jacobi polynomial integral operational matrix for solving Riccati differential equation of fractional order, Appl. Appl. Math. 10 (2) (2015).
[19] K. Maleknejad, L. Torkzadeh, Hybrid Functions Approach for the Fractional Riccati Differential Equation, Filomat 30 (9) (2016).
[20] H. Aminikhah, A.H.R. Sheikhani, H. Rezazadeh, Approximate analytical solutions of distributed order fractional Riccati differential equation, Ain Shams Eng. J. (2016).
[21] A. Atangana, D. Baleanu, New fractional derivatives with nonlocal and non-singular kernel: theory and application to heat transfer model, Therm. Sci. 20 (2) (2016) 763–769.
[22] A. Atangana, Non validity of index law in fractional calculus: A fractional differential operator with Markovian and non-Markovian properties, Physica A 505 (2018) 688–706.
[23] A. Atangana, J.F. Gómez-Aguilar, Decolonisation of fractional calculus rules: Breaking commutativity and associativity to capture more natural phenomena, Eur. Phys. J. Plus 133 (2018) 1–22.
[24] W.T. Reid, Riccati Differ. Equ., in: Mathematics in Science and Engineering, Vol. 86, Academic Press, New York, 1972.
[25] L. Ntogramatzidis, A. Ferrante, On the solution of the Riccati differential equation arising from the LQ optimal control problem, Systems Control Lett. 59 (2) (2010) 114–121.
[26] B.D. Anderson, J.B. Moore, Optimal filtering, Englewood Cliffs, 21, 1979 pp. 22–95.
[27] M. Gerber, B. Hasselblatt, D. Keesing, The Riccati equation: pinching of forcing and solutions, Experiment. Math. 12 (2) (2003) 129–134.
[28] G.A. Einicke, L.B. White, R.R. Bitmead, The use of fake algebraic Riccati equations for co-channel demodulation, IEEE Trans. Signal Process. 51 (9) (2003) 2288–2293.
[29] P.P. Boyle, W. Tian, F. Guan, The Riccati equation in mathematical finance, J. Symbolic Comput. 33 (3) (2002) 343–355.
[30] I. Lasiecka, R. Triggiani, Differential and algebraic Riccati equations with application to boundary/point control problems: continuous theory and approximation theory, Lecture Notes in Control and Inform. Sci. 164 (1991) 1–160.
[31] B.D. Anderson, J.B. Moore, Optimal Control: Linear Quadratic Methods, Prentice-Hall, New Jersey, 2007.
[32] Z. Odibat, A Riccati Equation Approach and Travelling Wave Solutions for Nonlinear Evolution Equations, Int. J. Appl. Comput. Math. 3 (1) (2017) 1–13.
[33] V.V. Kravchenko, Applied Pseudoanalytic Function Theory, Springer Science & Business Media, 2009.
[34] C. Li, Z. Zhao, Y. Chen, Numerical approximation of nonlinear fractional differential equations with subdiffusion and superdiffusion, Comput. Math. Appl. 62 (3) (2011) 855–875.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 309–317
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

On vector valued pseudo metrics and applications

Muhammad Usman Ali[a], Mihai Postolache[b,c,d,*]

[a] *Department of Mathematics, COMSATS University Islamabad, Attock Campus, Attock, Pakistan*
[b] *Center for General Education, China Medical University, Taichung 40402, Taiwan*
[c] *Gh. Mihoc-C. Iacob Institute of Mathematical Statistics and Applied Mathematics of the Romanian Academy, Bucharest 050711, Romania*
[d] *Department of Mathematics and Computer Science, University Politehnica of Bucharest, Bucharest 060042, Romania*

Received 27 April 2018; accepted 22 June 2018
Available online 7 July 2018

Abstract

In this article, we will introduce a new concept of gauge spaces induced by a family of vector valued pseudo metrics. After this, we will also prove some results to ensure the existence of fixed points of self mappings. As an application of our result, we will give an existence theorem to ensure the existence of solutions of n different integral equations.

Keywords: Vector gauge space; Ordered vector space ordered metric space; Order-convergent and order-Cauchy sequences

1. Introduction

The Banach contraction principle is one of the most fundamental results in metric fixed point theory. Its inception has opened new doors of research in mathematical analysis. The applicability of this result in nonlinear analysis and operator theory attracts many researchers to work in this area. As a result, this field has several different dimensions in which many researchers are working all over the world. For example, Frigon [1] and Chis and Precup [2] generalized the Banach contraction principle on gauge spaces. Cevik and Altun [3] introduced the concept of vector metric spaces and extended Banach contraction principle. Li et al. [4] introduced ordered metric spaces, it can be viewed as a slightly modified form of vector metric spaces and few others dimensions are given in [5–13].

In this article, we will introduce the notion of vector gauge spaces by using vector valued pseudo metrics. After this, we will state and prove some fixed point theorems for self mappings by using the structure of vector gauge spaces. We will also construct an example and an application to elaborate our result.

* Corresponding author at: Department of Mathematics and Computer Science, University Politehnica of Bucharest, Bucharest 060042, Romania.
E-mail addresses: muh_usman_ali@yahoo.com (M.U. Ali), emscolar@yahoo.com (M. Postolache).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.06.003

2. Preliminaries

First, we recall some basic concepts of ordered vector spaces from the literature, especially [4]. Throughout the article, by vector space we means a real vector space. A vector space X endowed with a partial order $\succeq^X$ is called a partially ordered vector space, or simply, an ordered vector space, denoted as $(X, \succeq^X)$, if the following properties hold:

(i) $x \succeq^X y$ implies that $x + z \succeq^X y + z$, for each $x, y, z \in X$;
(ii) $x \succeq^X y$ implies that $\alpha x \succeq^X \alpha y$, for each $x, y \in X$ and $\alpha \geq 0$.

A sequence $\{x_n\}$ in an ordered vector space $(X, \succeq^X)$ is said to be order-decreasing, whenever $m > n$ implies that $x_m \preceq^X x_n$. Such a sequence is denoted by $x_n \downarrow$. An order-decreasing sequence $\{x_n\}$ is said to be order-convergent to x, if $x_n \downarrow$ and $\wedge\{x_n\} = \inf\{x_n : n \in \mathbb{N}\}$ exist with $\wedge\{x_n\} = x$, such sequence is denoted by $x_n \downarrow x$. Analogously, we define an order-increasing sequence $\{x_n\}$, denoted as $x_n \uparrow$, further it is order-convergent to x, if $\vee\{x_n\} = \sup\{x_n : n \in \mathbb{N}\}$ exists with $\vee\{x_n\} = x$, denoted as $x_n \uparrow x$.

Lemma 2.1 (*[4]*). *Let $\{x_n\}$, $\{y_n\}$ be two sequences in an ordered vector space $(X, \succeq^X)$. Then, the following properties hold:*

(i) *$x_n \downarrow x$ implies that $\alpha x_n \downarrow \alpha x$, for each $\alpha \geq 0$.*
(ii) *$x_n \downarrow x$ and $y_n \downarrow y$ imply that $(x_n + y_n) \downarrow (x + y)$.*
(iii) *$x_n \downarrow x$ and $y_n \downarrow y$ imply that $(\alpha x_n + \beta y_n) \downarrow (\alpha x + \beta y)$, for any $\alpha, \beta \geq 0$.*

Definition 2.2 (*[4]*). An ordered vector space $(X, \succeq^X)$ is said to be generalized Archimedean, if for any given element $x \succeq^X 0$ and any decreasing sequence of positive real numbers $\{a_n\}$ with limit 0, we have $a_n x \downarrow 0$.

Li et al. [4] introduced the concept of ordered metric spaces in the following way:

Definition 2.3 (*[4]*). Let S be a nonempty set and let $(X, \succeq^X)$ be an ordered vector space. A mapping $d_X : S \times S \to X$ is called an ordered metric on S, with respect to $(X, \succeq^X)$ if, for every x, y, and z in S, it satisfies the following conditions:

(i) $d_X(x, y) \succeq^X 0$ with $d_X(x, y) = 0$ if and only if $x = y$;
(ii) $d_X(x, y) = d_X(y, x)$;
(iii) $d_X(x, z) \preceq^X d_X(x, y) + d_X(y, z)$.

Then (S, d_X) is called an ordered metric space, and $d_X(x, y)$ is called the ordered distance between x and y, with respect to the ordered vector space $(X, \succeq^X)$.

Example 2.4 (*[4]*). Following spaces are ordered metric spaces.

(i) Every metric space is an ordered metric space.
(ii) Every Banach space with the metric induced by its norm is an ordered metric space.
(iii) Every cone metric space is an ordered metric space.

Definition 2.5 (*[4]*). Let $\{s_n\}$ be a sequence in an ordered metric space (S, d_X). Then, the sequence $\{s_n\}$ is:

(i) an order-convergent to s in S, denoted as $s_n \to^o s$, whenever there exists another sequence $\{x_n\}$ in $(X, \succeq^X)$ with $x_n \downarrow 0$ such that $d_X(s_n, s) \preceq^X x_n$ holds, for each n.
(ii) an order-Cauchy sequence, whenever there exists another sequence $\{x_n\}$ in $(X, \succeq^X)$ with $x_n \downarrow 0$ such that $d_X(s_m, s_n) \preceq^X x_n$ holds for each n, and for every $m \geq n$.

Lemma 2.6 (*[4]*). *If a sequence $\{s_n\}$ in an ordered metric space (S, d_X) is order-convergent, then its order-limit is unique.*

Definition 2.7 (*[4]*). An ordered metric space (S, d_X) is said to be order-metric complete, whenever every order-Cauchy sequence in S is order-convergent.

3. Main results

Let S be a nonempty set and let $(X, \succeq^X)$ be an ordered vector space. A mapping $d_X : S \times S \to X$ is called an ordered pseudo b-metric on S, with respect to $(X, \succeq^X)$ if, for every x, y, and z in S, it satisfies the following conditions:

(i) $d_X(x, x) = 0$;
(ii) $d_X(x, y) = d_X(y, x)$;
(iii) $d_X(x, z) \preceq^X c[d_X(x, y) + d_X(y, z)]$

where $c \geq 1$. Then (S, d_X, c) is called an ordered pseudo b-metric with respect to ordered vector space $(X, \succeq^X)$ and $d_X(x, y)$ is called the ordered pseudo b-distance between x and y, with respect to the ordered vector space $(X, \succeq^X)$.

Let S be a nonempty set endowed with ordered pseudo b-metric, with respect to ordered vector space $(X, \succeq^X)$. The d_X-ball of radius $\epsilon \succeq^X 0$ centred at $x \in S$ is the set

$$B(x, d_X, \epsilon) = \{y \in S : d_X(x, y) \preceq^X \epsilon\}.$$

Definition 3.1. A family $\mathcal{F} = \{d_X^\nu \text{ with } c^\nu \geq 1 : \nu \in \mathcal{A}\}$ of ordered pseudo b-metrics on S, with respect to ordered vector space $(X, \succeq^X)$ is said to be separating if for each pair $(x, y) \in S \times S$ with $x \neq y$, there exists $d_X^\nu \in \mathcal{F}$ with $d_X^\nu(x, y) \neq 0$.

Definition 3.2. Let S be a nonempty set and let $\mathcal{F} = \{d_X^\nu \text{ with } c^\nu \geq 1 : \nu \in \mathcal{A}\}$ be a family of ordered pseudo b-metrics on S, with respect to ordered vector space $(X, \succeq^X)$. The topology $\mathcal{T}(\mathcal{F})$ having subbases the family

$$\mathcal{B}(\mathcal{F}) = \{B(x, d_X^\nu, \epsilon) : x \in S, d_X^\nu \in \mathcal{F} \text{ and } \epsilon \succeq^X 0\}$$

of balls is called vector topology induced by the family $\mathcal{F}$ of ordered pseudo b-metrics, with respect to ordered vector space $(X, \succeq^X)$. The pair $(S, \mathcal{T}(\mathcal{F}))$ is called a vector b-gauge space. Note that, the vector b-gauge space $(S, \mathcal{T}(\mathcal{F}))$ induced by the family $\mathcal{F}$ is Hausdorff if $\mathcal{F}$ is separating.

In the following definition, we will discuss some concepts regarding the sequence.

Definition 3.3. Let $(S, \mathcal{T}(\mathcal{F}))$ be a vector b-gauge space with respect to the family $\mathcal{F} = \{d_X^\nu \text{ with } c^\nu \geq 1 : \nu \in \mathcal{A}\}$ of ordered pseudo b-metrics on S, with respect to ordered vector space $(X, \succeq^X)$. Let $\{s_n\}$ be a sequence in S and $s \in S$. Then:

(i) The sequence $\{s_n\}$ order-converges to s, denoted as $s_n \to^o s$, if for each $\nu \in \mathcal{A}$, there exists a sequence $\{x_n^\nu\}$ in $(X, \succeq^X)$ with $x_n^\nu \downarrow 0$ such that $d_X^\nu(s_n, s) \preceq^X x_n^\nu$ holds, for each n.
(ii) The sequence $\{s_n\}$ is an order-Cauchy sequence if for each $\nu \in \mathcal{A}$, there exists a sequence $\{x_n^\nu\}$ in $(X, \succeq^X)$ with $x_n^\nu \downarrow 0$ such that $d_X^\nu(s_m, s_n) \preceq^X x_n^\nu$ holds for each n, and for every $m \geq n$.
(iii) The space $(S, \mathcal{T}(\mathcal{F}))$ is order-complete if each order-Cauchy sequence in S is order-convergent in S.
(iv) A subset of S is said to be order-closed if it contains the order-limit of each order-convergent sequence of its elements.

Remark 3.4. Note that, when $c^\nu = 1$ for each $\nu \in \mathcal{A}$, the vector b-gauge space just called vector gauge space and a family $\mathcal{F} = \{d_X^\nu : \nu \in \mathcal{A}\}$ of ordered pseudo b-metrics on S with respect to ordered vector space $(X, \succeq^X)$, just called family of pseudo metrics on S with respect to ordered vector space $(X, \succeq^X)$.

Subsequently, we consider $\mathcal{A}$ is directed set and S is a nonempty set endowed with order-complete vector b-gauge structure induced by separating family $\mathcal{F} = \{d_X^\nu \text{ with } c^\nu \geq 1 : \nu \in \mathcal{A}\}$ of ordered pseudo b-metrics on S, with respect to a generalized Archimedean ordered vector space $(X, \succeq^X)$. Furthermore, $G = (V, E)$ is a directed graph such that the set of its vertices V is equal to S and set of its edges E contains $\{(s, s) : s \in V\}$, but G has no parallel edges.

Let Ψ denote the family of vector valued functions $\psi : (X, \succeq^X) \to (X, \succeq^X)$ such that $\psi(t) \preceq^X \rho t$ for each $t \in X$, where $\rho \in [0, 1)$.

Theorem 3.5. *Let $T : S \to S$ be a mapping such that for each $s, \overline{s} \in S$ with $(s, \overline{s}) \in E$, we have the following inequality:*

$$c^{\nu} d_X^{\nu}(Ts, T\overline{s}) \preceq^X \psi(d_X^{\nu}(s, \overline{s})) + L^{\nu} d_X^{\nu}(\overline{s}, Ts) \tag{3.1}$$

for each $\nu \in \mathcal{A}$, where $L^{\nu} \geq 0$ for each $\nu \in \mathcal{A}$. Further, assume that the following conditions hold:

(i) *there exists $s_0 \in S$ with $(s_0, Ts_0) \in E$;*
(ii) *$(Ts, T\overline{s}) \in E$, whenever $(s, \overline{s}) \in E$;*
(iii) *for any sequence $\{\overline{s}_n\} \subseteq S$ with $(\overline{s}_n, \overline{s}_{n+1}) \in E$ for each $n \in \mathbb{N}$ and $\overline{s}_n \to^o s$, we have $(\overline{s}_n, s) \in E$ for each $n \in \mathbb{N}$.*

Then T has a fixed point.

Proof. By using hypothesis (i) and (ii), we construct a sequence $\{s_n\}$ such that $s_n = Ts_{n-1} = T^n s_0$ and $(s_n, s_{n+1}) \in E$ for each $n \in \mathbb{N} \cup \{0\}$. From (3.1), for each $n \in \mathbb{N} \cup \{0\}$, we get

$$\begin{aligned} c^{\nu} d_X^{\nu}(s_{n+1}, s_{n+2}) &\preceq^X \psi(d_X^{\nu}(s_n, s_{n+1})) + L^{\nu} d_X^{\nu}(s_{n+1}, Ts_n) \\ &= \psi(d_X^{\nu}(s_n, s_{n+1})) \\ &\preceq^X \rho d_X^{\nu}(s_n, s_{n+1}) \ \forall \nu \in \mathcal{A}. \end{aligned}$$

This yields, the following inequality

$$d_X^{\nu}(s_{n+1}, s_{n+2}) \preceq^X \frac{\rho}{c^{\nu}} d_X^{\nu}(s_n, s_{n+1}) \ \forall \nu \in \mathcal{A}$$

and each $n \in \mathbb{N} \cup \{0\}$. Hence we conclude that

$$d_X^{\nu}(s_{n+1}, s_{n+2}) \preceq^X \left(\frac{\rho}{c^{\nu}}\right)^{n+1} d_X^{\nu}(s_0, s_1) \tag{3.2}$$

for each $\nu \in \mathcal{A}$ and each $n \in \mathbb{N} \cup \{0\}$. We now show that $\{s_n\}$ is an order-Cauchy sequence in S. Take any arbitrary natural numbers m and n, by using the triangle inequality, we get

$$\begin{aligned} d_X^{\nu}(s_{n+m}, s_n) &\preceq^X (c^{\nu})^n d_X^{\nu}(s_n, s_{n+1}) + (c^{\nu})^{n+1} d_X^{\nu}(s_{n+1}, s_{n+2}) + \cdots + (c^{\nu})^{n+m-1} d_X^{\nu}(s_{n+m-1}, s_{n+m}) \\ &\preceq^X (c^{\nu})^n \left(\frac{\rho}{c^{\nu}}\right)^n d_X^{\nu}(s_0, s_1) + (c^{\nu})^{n+1} \left(\frac{\rho}{c^{\nu}}\right)^{n+1} d_X^{\nu}(s_0, s_1) + \cdots + \\ &\quad (c^{\nu})^{n+m-1} \left(\frac{\rho}{c^{\nu}}\right)^{n+m-1} d_X^{\nu}(s_0, s_1) \\ &\preceq^X \frac{\rho^n}{1-\rho} d_X^{\nu}(s_0, s_1) \ \forall \nu \in \mathcal{A}. \end{aligned}$$

For each $\nu \in \mathcal{A}$, we take a sequence $\{x_n^{\nu}\}$ in X as $x_n^{\nu} = \frac{\rho^n}{1-\rho} d_X^{\nu}(s_0, s_1)$ for each $n \in \mathbb{N}$. Since $(X, \succeq^X)$ is generalized Archimedean, by Definition 2.2, we get $x_n^{\nu} \downarrow 0$ for each $\nu \in \mathcal{A}$. Hence, $\{s_n\}$ is an order-Cauchy sequence in $(S, \mathcal{T}(\mathcal{F}))$. Since, $(S, \mathcal{T}(\mathcal{F}))$ is order-complete vector b-gauge space. Then, we have $s^* \in S$ such that $s_n \to^o s^*$. By using hypothesis (iii), we get $(s_n, s^*) \in E$ for each $n \in \mathbb{N}$. Then, from (3.1), for each $n \in \mathbb{N}$, we have

$$\begin{aligned} c^{\nu} d_X^{\nu}(s_{n+1}, Ts^*) &= d_X^{\nu}(Ts_n, Ts^*) \\ &\preceq^X \psi(d_X^{\nu}(s_n, s^*)) + L^{\nu} d_X^{\nu}(s^*, Ts_n) \\ &\preceq^X \rho d_X^{\nu}(s_n, s^*) + L^{\nu} d_X^{\nu}(s^*, s_{n+1}) \ \forall \nu \in \mathcal{A}. \end{aligned}$$

By using the triangle inequality and the above inequality, for each $n \in \mathbb{N}$, we get the following inequality

$$\begin{aligned} d_X^{\nu}(s^*, Ts^*) &\preceq^X c^{\nu} d_X^{\nu}(s^*, s_{n+1}) + c^{\nu} d_X^{\nu}(s_{n+1}, Ts^*) \\ &\preceq^X c^{\nu} d_X^{\nu}(s^*, s_{n+1}) + \rho d_X^{\nu}(s_n, s^*) + L^{\nu} d_X^{\nu}(s^*, s_{n+1}) \ \forall \nu \in \mathcal{A}. \end{aligned} \tag{3.3}$$

As $s_n \to^o s^*$, for each $\nu \in \mathcal{A}$, there is a sequence $\{z_n^{\nu}\} \subseteq X$ such that $z_n^{\nu} \downarrow 0$ and $d_X^{\nu}(s_n, s^*) \preceq^X z_n^{\nu}$ for each $n \in \mathbb{N}$. Thus, by using this fact, Lemma 2.1 and (3.3), we conclude that

$$d_X^{\nu}(s^*, Ts^*) \preceq^X 0 \ \forall \nu \in \mathcal{A}.$$

Hence, we have $d_X^{\nu}(s^*, Ts^*) = 0$ for each $\nu \in \mathcal{A}$. Since the family $\mathcal{F}$ is separating, thus we conclude that $s^* = Ts^*$. $\square$

Example 3.6. Let $S = \mathbb{R}^2$ be the set of all ordered pairs of real numbers. Also, $X = \mathbb{R}^2$ is an ordered vector space with the partial order defined as $(x_1, x_2) \preceq^X (y_1, y_2)$ if $x_1 \leq y_1$ and $x_2 \leq y_2$. Define a mapping

$$T : S \to S, \quad T(a_1, a_2) = \begin{cases} \left(1 + \frac{1}{3}a_1, \frac{2}{3}a_2\right) \text{ if } a_1, a_2 \geq 0 \\ (-a_1, a_2) \text{ otherwise.} \end{cases}$$

Consider the graph G defined as $V = S$ and $E = \{((a_1, a_2), (b_1, b_2)) : a_1, a_2, b_1, b_2 \geq 0\} \cup \{((a, b), (a, b)) : a, b \in \mathbb{R}\}$ and separating family $\mathcal{F} = \{d_X^1, d_X^2, d_X^3\}$ of ordered pseudo b-metrics on S with respect to ordered vector space $(X, \preceq^X)$, defined as

$$d_X^1((a_1, a_2), (b_1, b_2)) = (|a_1 - b_1|, 0)$$

$$d_X^2((a_1, a_2), (b_1, b_2)) = (0, |a_2 - b_2|)$$

and

$$d_X^3((a_1, a_2), (b_1, b_2)) = (|a_1 - b_1|, |a_2 - b_2|).$$

Then, it is easy to see that for each $((x_1, x_2), (y_1, y_2)) \in E$, we have

$$d_X^i(T(x_1, x_2), T(y_1, y_2)) \preceq^X \frac{2}{3} d_X^i(T(x_1, x_2), T(y_1, y_2)), \text{ for each } i = 1, 2, 3.$$

Here we have $\psi(t) = \frac{2}{3}t$ for each $t \in \mathbb{R}^2$ and $c^1 = c^2 = c^3 = 1$. For $s_0 = (0, 0)$, we have $((0, 0), (1, 0)) \in E$. Also for $((a_1, a_2), (b_1, b_2)) \in E$, we have $(T(a_1, a_2), T(b_1, b_2)) \in E$. Moreover, for any sequence $\{\overline{s}_n\} \subseteq S$ with $(\overline{s}_n, \overline{s}_{n+1}) \in E$ for each $n \in \mathbb{N}$ and $\overline{s}_n \to^o s$, we have $(\overline{s}_n, s) \in E$ for each $n \in \mathbb{N}$. Thus, Theorem 3.5 implies that T has a fixed point.

We now move towards the second main result of our article.

Theorem 3.7. *Let $T : S \to S$ be a mapping such that for each $s, \overline{s} \in S$ with $(s, \overline{s}) \in E$, we have the following inequality:*

$$d_X^\nu(Ts, T\overline{s}) \preceq^X \gamma_1^\nu d_X^\nu(s, \overline{s}) + \gamma_2^\nu d_X^\nu(s, Ts) + \gamma_3^\nu d_X^\nu(\overline{s}, T\overline{s}) + \gamma_4^\nu d_X^\nu(s, T\overline{s}) + \gamma_5^\nu d_X^\nu(\overline{s}, Ts) \tag{3.4}$$

for each $\nu \in \mathcal{A}$, where $\gamma_1^\nu, \gamma_2^\nu, \gamma_3^\nu, \gamma_4^\nu, \gamma_5^\nu \geq 0$ and $\gamma_1^\nu + \gamma_2^\nu + \gamma_3^\nu + 2c^\nu\gamma_4^\nu < 1/c^\nu$ for each $\nu \in \mathcal{A}$. Further, assume that the following conditions hold:

(i) *there exists $s_0 \in S$ with $(s_0, Ts_0) \in E$;*
(ii) *$(Ts, T\overline{s}) \in E$, whenever $(s, \overline{s}) \in E$;*
(iii) *for any sequence $\{\overline{s}_n\} \subseteq S$ with $(\overline{s}_n, \overline{s}_{n+1}) \in E$ for each $n \in \mathbb{N}$ and $\overline{s}_n \to^o s$, we have $(\overline{s}_n, s) \in E$ for each $n \in \mathbb{N}$.*

Then T has a fixed point.

Proof. By using hypothesis (i) and (ii), we construct a sequence $\{s_n\}$ such that $s_n = Ts_{n-1} = T^n s_0$ and $(s_n, s_{n+1}) \in E$ for each $n \in \mathbb{N} \cup \{0\}$. From (3.4), for each $n \in \mathbb{N} \cup \{0\}$, we have

$$\begin{aligned} d_X^\nu(s_{n+1}, s_{n+2}) &= d_X^\nu(Ts_n, Ts_{n+1}) \\ &\preceq^X \gamma_1^\nu d_X^\nu(s_n, s_{n+1}) + \gamma_2^\nu d_X^\nu(s_n, Ts_n) + \gamma_3^\nu d_X^\nu(s_{n+1}, Ts_{n+1}) \\ &\quad + \gamma_4^\nu d_X^\nu(s_n, Ts_{n+1}) + \gamma_5^\nu d_X^\nu(s_{n+1}, Ts_n) \\ &= \gamma_1^\nu d_X^\nu(s_n, s_{n+1}) + \gamma_2^\nu d_X^\nu(s_n, s_{n+1}) + \gamma_3^\nu d_X^\nu(s_{n+1}, s_{n+2}) \\ &\quad + \gamma_4^\nu d_X^\nu(s_n, s_{n+2}) + \gamma_5^\nu d_X^\nu(s_{n+1}, s_{n+1}) \\ &\preceq^X \gamma_1^\nu d_X^\nu(s_n, s_{n+1}) + \gamma_2^\nu d_X^\nu(s_n, s_{n+1}) + \gamma_3^\nu d_X^\nu(s_{n+1}, s_{n+2}) \\ &\quad + \gamma_4^\nu c^\nu d_X^\nu(s_n, s_{n+1}) + \gamma_4^\nu c^\nu d_X^\nu(s_{n+1}, s_{n+2}) \ \forall \nu \in \mathcal{A}. \end{aligned}$$

This yields, the following inequality

$$d_X^\nu(s_{n+1}, s_{n+2}) \preceq^X \eta^\nu d_X(s_n, s_{n+1})$$

for each $\nu \in \mathcal{A}$ and $n \in \mathbb{N} \cup \{0\}$, where $\eta^\nu = \frac{\gamma_1^\nu + \gamma_2^\nu + c^\nu \gamma_4}{1 - \gamma_3^\nu - \gamma_4^\nu c^\nu} < \frac{1}{c^\nu}$ for each $\nu \in \mathcal{A}$. Hence we conclude that

$$d_X^\nu(s_{n+1}, s_{n+2}) \preceq^X (\eta^\nu)^{n+1} d_X^\nu(s_0, s_1) \tag{3.5}$$

for each $\nu \in \mathcal{A}$ and $n \in \mathbb{N} \cup \{0\}$. We now show that $\{s_n\}$ is an order-Cauchy sequence in S. Take arbitrary natural numbers m and n, by using the triangle inequality, we have

$$\begin{aligned} d_X^\nu(s_{n+m}, s_n) &\preceq^X (c^\nu)^n d_X^\nu(s_n, s_{n+1}) + (c^\nu)^{n+1} d_X^\nu(s_{n+1}, s_{n+2}) + \cdots + \\ &\quad (c^\nu)^{n+m-1} d_X^\nu(s_{n+m-1}, s_{n+m}) \\ &\preceq^X (c^\nu)^n (\eta^\nu)^n d_X^\nu(s_0, s_1) + (c^\nu)^{n+1} (\eta^\nu)^{n+1} d_X^\nu(s_0, s_1) + \cdots + \\ &\quad (c^\nu)^{n+m-1} (\eta^\nu)^{n+m-1} d_X^\nu(s_0, s_1) \\ &\preceq^X \frac{(c^\nu \eta^\nu)^n}{1 - c^\nu \eta^\nu} d_X^\nu(s_0, s_1) \ \forall \nu \in \mathcal{A}. \end{aligned}$$

For each $\nu \in \mathcal{A}$, we take a sequence $\{x_n^\nu\}$ in X as $x_n^\nu = \frac{(c^\nu \eta^\nu)^n}{1 - c^\nu \eta^\nu} d_X^\nu(s_0, s_1)$ for each $n \in \mathbb{N}$. Since $(X, \succeq^X)$ is generalized Archimedean ordered vector space, by Definition 2.2, we get $x_n^\nu \downarrow 0$ for each $\nu \in \mathcal{A}$. Hence, $\{s_n\}$ is an order-Cauchy sequence in $(S, \mathcal{T}(\mathcal{F}))$. Since, $(S, \mathcal{T}(\mathcal{F}))$ is order-complete vector b-gauge space. Then, we have $s^* \in S$ such that $s_n \to^o s^*$. By hypothesis (iii), we get $(s_n, s^*) \in E$ for each $n \in \mathbb{N} \cup \{0\}$. Then from (3.4), for each $n \in \mathbb{N} \cup \{0\}$, we have

$$\begin{aligned} d_X^\nu(s_{n+1}, Ts^*) &= d_X^\nu(Ts_n, Ts^*) \\ &\preceq^X \gamma_1^\nu d_X^\nu(s_n, s^*) + \gamma_2^\nu d_X^\nu(s_n, Ts_n) + \gamma_3^\nu d_X^\nu(s^*, Ts^*) \\ &\quad + \gamma_4^\nu d_X^\nu(s_n, Ts^*) + \gamma_5^\nu d_X^\nu(s^*, Ts_n) \\ &= \gamma_1^\nu d_X^\nu(s_n, s^*) + \gamma_2^\nu d_X^\nu(s_n, s_{n+1}) + \gamma_3^\nu d_X^\nu(s^*, Ts^*) \\ &\quad + \gamma_4^\nu d_X^\nu(s_n, Ts^*) + \gamma_5^\nu d_X^\nu(s^*, s_{n+1}) \ \forall \nu \in \mathcal{A}. \end{aligned}$$

By using the triangle inequality and the above inequality, we get the following

$$\begin{aligned} d_X^\nu(s^*, Ts^*) &\preceq^X c^\nu d_X^\nu(s^*, s_{n+1}) + c^\nu d_X^\nu(s_{n+1}, Ts^*) \\ &\preceq^X c^\nu d_X^\nu(s^*, s_{n+1}) + c^\nu \gamma_1^\nu d_X^\nu(s_n, s^*) + c^\nu \gamma_2^\nu d_X^\nu(s_n, s_{n+1}) \\ &\quad + c^\nu \gamma_3^\nu d_X^\nu(s^*, Ts^*) + c^\nu \gamma_4^\nu d_X^\nu(s_n, Ts^*) + c^\nu \gamma_5^\nu d_X^\nu(s^*, s_{n+1}) \\ &\preceq^X c^\nu d_X^\nu(s^*, s_{n+1}) + c^\nu \gamma_1^\nu d_X^\nu(s_n, s^*) + c^\nu \gamma_2^\nu d_X^\nu(s_n, s_{n+1}) \\ &\quad + c^\nu \gamma_3^\nu d_X^\nu(s^*, Ts^*) + (c^\nu)^2 \gamma_4^\nu c^\nu d_X^\nu(s_n, s^*) + (c^\nu)^2 \gamma_4^\nu d_X^\nu(s^*, Ts^*) \\ &\quad + c^\nu \gamma_5^\nu d_X^\nu(s^*, s_{n+1}) \ \forall \nu \in \mathcal{A}. \end{aligned} \tag{3.6}$$

As $s_n \to^o s^*$, for each $\nu \in \mathcal{A}$, there is a sequence $\{z_n^\nu\} \subseteq X$ such that $z_n^\nu \downarrow 0$ and $d_X^\nu(s_n, s^*) \preceq^X z_n^\nu$ for each $n \in \mathbb{N}$. Thus, by using this fact, Lemma 2.1 and (3.6), we conclude that

$$(1 - c^\nu \gamma_3^\nu - (c^\nu)^2 \gamma_4^\nu) d_X^\nu(s^*, Ts^*) \preceq^X 0 \ \forall \nu \in \mathcal{A}.$$

Hence, we have $d_X^\nu(s^*, Ts^*) = 0$ for each $\nu \in \mathcal{A}$. Since the family $\mathcal{F}$ is separating, thus we conclude that $s^* = Ts^*$. □

Theorem 3.8. *Let $T : S \to S$ be a mapping such that for each $s, \overline{s} \in S$ with $(s, \overline{s}) \in E$, we have the following inequality:*

$$d_X^\nu(Ts, T\overline{s}) \preceq^X k u(s, \overline{s}) \tag{3.7}$$

$$u(s, \overline{s}) \in \{d_X^\nu(s, \overline{s}), d_X^\nu(s, Ts), d_X^\nu(\overline{s}, T\overline{s}), \frac{d_X^\nu(s, T\overline{s}) + d_X^\nu(\overline{s}, Ts)}{2c^\nu}\}$$

for each $\nu \in \mathcal{A}$, where $0 \le k < 1/c^\nu$ for each $\nu \in \mathcal{A}$. Further, assume that the following conditions hold:

(i) *there exists $s_0 \in S$ with $(s_0, Ts_0) \in E$;*
(ii) *$(Ts, T\overline{s}) \in E$, whenever $(s, \overline{s}) \in E$;*
(iii) *for any sequence $\{\overline{s}_n\} \subseteq S$ with $(\overline{s}_n, \overline{s}_{n+1}) \in E$ for each $n \in \mathbb{N}$ and $\overline{s}_n \to^o s$, we have $(\overline{s}_n, s) \in E$ for each $n \in \mathbb{N}$.*

Then T has a fixed point.

Proof. To prove this theorem we take the help of Theorem 3.7, by considering the following cases:

Case 1: If $u(s,\overline{s}) = d_X^{\nu}(s,\overline{s})$, the conclusion is obtained from Theorem 3.7, by taking $\gamma_1^{\nu} = k$ and $\gamma_2^{\nu} = \gamma_3^{\nu} = \gamma_4^{\nu} = \gamma_5^{\nu} = 0$ for each $\nu \in \mathcal{A}$.

Case 2: If $u(s,\overline{s}) = d_X^{\nu}(s,Ts)$, the conclusion is obtained from Theorem 3.7, by taking $\gamma_2^{\nu} = k$ and $\gamma_1^{\nu} = \gamma_3^{\nu} = \gamma_4^{\nu} = \gamma_5^{\nu} = 0$ for each $\nu \in \mathcal{A}$.

Case 3: If $u(s,\overline{s}) = d_X^{\nu}(\overline{s},T\overline{s})$, the conclusion is obtained from Theorem 3.7, by taking $\gamma_3^{\nu} = k$ and $\gamma_1^{\nu} = \gamma_2^{\nu} = \gamma_4^{\nu} = \gamma_5^{\nu} = 0$ for each $\nu \in \mathcal{A}$.

Case 4: If $u(s,\overline{s}) = \frac{d_X^{\nu}(s,T\overline{s})+d_X^{\nu}(\overline{s},Ts)}{2c^{\nu}}$, the conclusion is obtained from Theorem 3.7, by taking $\gamma_1^{\nu} = \gamma_2^{\nu} = \gamma_3^{\nu} = 0$ and $\gamma_4^{\nu} = \gamma_5^{\nu} = \frac{k}{2c^{\nu}}$ for each $\nu \in \mathcal{A}$. □

4. Consequences and application

Note that the results of this section can be obtained from the above stated results, respectively. When we take $\mathcal{F} = \{d_X^{\nu} : \nu \in \mathcal{A}\} = \{d_X\}$, where d_X is order-metric complete ordered metric on S, with respect to a generalized Archimedean ordered vector space $(X, \preceq^X)$.

Corollary 4.1. *Let (S, d_X) be an order-metric complete ordered metric space, with respect to a generalized Archimedean ordered vector space $(X, \preceq^X)$ and S is endowed with the graph G. Let $T : S \to S$ be a mapping such that for each $s, \overline{s} \in S$ with $(s,\overline{s}) \in E$, we have the following inequality:*

$$d_X(Ts, T\overline{s}) \preceq^X \psi(d_X(s,\overline{s})) + Ld_X(\overline{s}, Ts)$$

where $L \geq 0$ and $\psi \in \Psi$. Further, assume that the following conditions hold:

(i) *there exists $s_0 \in S$ with $(s_0, Ts_0) \in E$;*
(ii) *$(Ts, T\overline{s}) \in E$, whenever $(s,\overline{s}) \in E$;*
(iii) *for any sequence $\{\overline{s}_n\} \subseteq S$ with $(\overline{s}_n, \overline{s}_{n+1}) \in E$ for each $n \in \mathbb{N}$ and $\overline{s}_n \to^o s$, we have $(\overline{s}_n, s) \in E$ for each $n \in \mathbb{N}$.*

Then T has a fixed point.

Corollary 4.2. *Let (S, d_X) be an order-metric complete ordered metric space, with respect to a generalized Archimedean ordered vector space $(X, \preceq^X)$ and S is endowed with the graph G. Let $T : S \to S$ be a mapping such that for each $s, \overline{s} \in S$ with $(s,\overline{s}) \in E$, we have the following inequality:*

$$d_X(Ts, T\overline{s}) \preceq^X \gamma_1 d_X(s,\overline{s}) + \gamma_2 d_X(s, Ts) + \gamma_3 d_X(\overline{s}, T\overline{s}) + \gamma_4 d_X(s, T\overline{s}) + \gamma_5 d_X(\overline{s}, Ts)$$

where $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5 \geq 0$ and $\gamma_1 + \gamma_2 + \gamma_3 + 2\gamma_4 < 1$. Further, assume that the following conditions hold:

(i) *there exists $s_0 \in S$ with $(s_0, Ts_0) \in E$;*
(ii) *$(Ts, T\overline{s}) \in E$, whenever $(s,\overline{s}) \in E$;*
(iii) *for any sequence $\{\overline{s}_n\} \subseteq S$ with $(\overline{s}_n, \overline{s}_{n+1}) \in E$ for each $n \in \mathbb{N}$ and $\overline{s}_n \to^o s$, we have $(\overline{s}_n, s) \in E$ for each $n \in \mathbb{N}$.*

Then T has a fixed point.

Corollary 4.3. *Let (S, d_X) be an order-metric complete ordered metric space, with respect to a generalized Archimedean ordered vector space $(X, \preceq^X)$ and S is endowed with the graph G. Let $T : S \to S$ be a mapping such that for each $s, \overline{s} \in S$ with $(s,\overline{s}) \in E$, we have the following inequality:*

$$d_X(Ts, T\overline{s}) \preceq^X ku(s,\overline{s})$$

$$u(s,\overline{s}) \in \{d_X(s,\overline{s}), d_X(s, Ts), d_X(\overline{s}, T\overline{s}), \frac{d_X(s, T\overline{s}) + d_X(\overline{s}, Ts)}{2}\}$$

where $0 \leq k < 1$. Further, assume that the following conditions hold:

(i) *there exists $s_0 \in S$ with $(s_0, Ts_0) \in E$;*

(ii) $(Ts, T\overline{s}) \in E$, *whenever* $(s, \overline{s}) \in E$;
(iii) *for any sequence* $\{\overline{s}_n\} \subseteq S$ *with* $(\overline{s}_n, \overline{s}_{n+1}) \in E$ *for each* $n \in \mathbb{N}$ *and* $\overline{s}_n \to^o s$, *we have* $(\overline{s}_n, s) \in E$ *for each* $n \in \mathbb{N}$.

Then T *has a fixed point.*

Now, consider an application to the N-Volterra integral equations of the form:

$$h(t)x(t) = \mu \int_0^{f(t)} K_i(t, s, x(s))ds + g(t),\ t \in I = [0, \infty) \tag{4.1}$$

for $i \in \{1, 2, 3, \ldots, N\}$, where $f, g, h : I \to \mathbb{R}$ and $K : I \times I \times \mathbb{R} \to \mathbb{R}$ are continuous functions with $0 \leq f(t)$ and f is nondecreasing for each $t \in I$, and $\mu \in [0, 1)$.

Let $S = C[0, \infty)$ be the space of all realvalued continuous functions defined on $[0, \infty)$. Consider $X = \mathbb{R}^N$ as the set of all N-tuples of real numbers, this set is an ordered vector space with the partial order defined as $x = (x_1, x_2, \ldots, x_N) \preceq^X y = (y_1, y_2, \ldots, y_N)$ if $x_i \leq y_i$ for each $i \in \{1, 2, \ldots, N\}$. We take the family of ordered pseudo metrics on S^N, with respect to a generalized Archimedean ordered vector space $(X = \mathbb{R}^N, \succeq^X)$ such that

$$d_n(x, y) = \Big(\max_{t\in[0,n]} |x_1(t) - y_1(t)|, \max_{t\in[0,n]} |x_2(t) - y_2(t)|, \ldots, \max_{t\in[0,n]} |x_N(t) - y_N(t)|\Big).$$

Clearly, the family $\mathcal{F} = \{d_n : n \in \mathbb{N}\}$ defines a vector gauge structure on S^N, which is complete and separating.

We now state a result which we need in our application. This one is a direct consequence of our main result, by using Remark 3.4 and the graph $G = (V, E)$ defined as $V = S$ and $E = S \times S$.

Corollary 4.4. *Let* S *be a nonempty set endowed with order-complete vector gauge structure induced by separating family* $\mathcal{F} = \{d_X^v : v \in \mathcal{A}\}$ *of ordered pseudo metrics on* S, *with respect to a generalized Archimedean ordered vector space* $(X, \succeq^X)$. *Let* $T : S \to S$ *be a mapping such that for each* $s, \overline{s} \in S$, *we have the following inequality:*

$$d_X^v(Ts, T\overline{s}) \preceq^X \gamma^v d_X^v(s, \overline{s}) \tag{4.2}$$

for each $v \in \mathcal{A}$, *where* $0 \leq \gamma^v < 1$ *for each* $v \in \mathcal{A}$. *Then* T *has a fixed point.*

We now state and prove the theorem which ensures the existence of solutions of integral equations given in (4.1).

Theorem 4.5. *Let* $S = C[0, \infty)$ *and let the operators* $T_i : S \to S$ *be defined by*

$$T_i x(t) = \frac{\mu}{h(t)} \int_0^{f(t)} K_i(t, s, x(s))ds + g(t),\ t \in I = [0, \infty) \tag{4.3}$$

for each $i \in \{1, 2, 3, \ldots, N\}$, *where* $f, g, h : I \to \mathbb{R}$ *and* $K : I \times I \times \mathbb{R} \to \mathbb{R}$ *are continuous functions such that* $0 \leq f(t)$ *and* f *is nondecreasing for each* $t \in I$, *and* $|\mu| < 1$. *Further, assume that for each* $t, s \in I$ *and* $x, y \in S$, *we have*

$$|K_i(t, s, x(s)) - K_i(t, s, y(s))| \leq L_i|x(s) - y(s)|,\ \forall i \in \{1, 2, 3, \ldots, N\}$$

and $h(t) > L^* f(t)$ *for each* $t \in I$, $L^* = \max\{L_i : i \in \{1, 2, 3, \ldots, N\}\}$. *Then the integral equation in* (4.1) *has at least one solution.*

Proof. For any $(x, y) \in S \times S$ and $t \in [0, n]$ for each $n \geq 1$, we have

$$\begin{aligned}
|T_i x(t) - T_i y(t)| &\leq \frac{|\mu|}{h(t)} \int_0^{f(t)} |K_i(t, s, x(s)) - K_i(t, s, y(s))|ds \\
&\leq \frac{|\mu|}{h(t)} \left(\int_0^{f(t)} L_i ds\right) \max_{t\in[0,n]} |x(t) - y(t)| \\
&\leq \frac{|\mu|}{h(t)} \left(\int_0^{f(t)} L^* ds\right) \max_{t\in[0,n]} |x(t) - y(t)|
\end{aligned}$$

for each $i \in \{1, 2, 3, \ldots, N\}$. From the above inequality, we conclude that for every $n \in \mathbb{N}$, we have

$$\max_{t\in[0,n]} |T_i x(t) - T_i y(t)| < |\mu|\big(\max_{t\in[0,n]} |x(t) - y(t)|\big)$$

for each $i \in \{1, 2, 3, \ldots, N\}$. Define an operator $\mathbb{T} : C[0, \infty)^N \to C[0, \infty)^N$ as

$$\mathbb{T}(x_1, x_2, \ldots, x_N) = (T_1x_1, T_2x_2, \ldots, T_Nx_N). \tag{4.4}$$

Then for each $\overline{x} = (x_1, x_2, \ldots, x_N)$ and $\overline{y} = (y_1, y_2, \ldots, y_N) \in S^N$, we have

$$\begin{aligned} d_n(\mathbb{T}\overline{x}, \mathbb{T}\overline{y}) &= \Big(\max_{t\in[0,n]} |T_1x_1(t) - T_1y_1(t)|, \max_{t\in[0,n]} |T_2x_2(t) - T_2y_2(t)|, \ldots, \\ &\qquad \max_{t\in[0,n]} |T_Nx_N(t) - T_Ny_N(t)|\Big) \\ &\leq \Big(|\mu| \max_{t\in[0,n]} |x_1(t) - y_1(t)|, |\mu| \max_{t\in[0,n]} |x_2(t) - y_2(t)|, \ldots, |\mu| \max_{t\in[0,n]} |x_N(t) - y_N(t)|\Big) \\ &= |\mu| d_n(\overline{x}, \overline{y}) \text{ for each } n \in \mathbb{N}. \end{aligned}$$

Therefore, by Corollary 4.4, one can conclude that the operator (4.4) has a fixed point, that is, each integral equation in (4.1) has atleast one solution. □

Remark 4.6. Note that the conclusion of Theorem 4.5 still holds if $|\mu| = 1$.

References

[1] M. Frigon, Fixed point results for generalized contractions in gauge spaces and applications, Proc. Amer. Math. Soc. 128 (10) (2000) 2957–2965.
[2] A. Chis, R. Precup, Continuation theory for general contractions in gauge spaces, Fixed Point Theory Appl. 3 (2004) 173–185.
[3] C. Cevik, I. Altun, Vector metric spaces and some properties, Topol. Methods Nonlinear Anal. 34 (2009) 375–382.
[4] J.L. Li, C.J. Zhang, Q.Q. Chen, Fixed point theorems on ordered vector spaces, Fixed Point Theory Appl. 2014 (2014) 109.
[5] H. Rahimi, M. Abbas, G.S. Rad, Common fixed point results for four mappings on ordered vector metric spaces, Filomat 29 (2015) 865–878.
[6] A. Bucur, L. Guran, A. Petrusel, Fixed points for multivalued operators on a set endowed with vector-valued metrics and applications, Fixed Point Theory 10 (2009) 19–34.
[7] B. Samet, C. Vetro, P. Vetro, Fixed point theorems for α-ψ-contractive type mappings, Nonlinear Anal. 75 (2012) 2154–2165.
[8] C. Cevik, I. Altun, H. Sahin, C.C. Ozeken, Some fixed point theorems for contractive mapping in ordered vector metric spaces, J. Nonlinear Sci. Appl. 10 (2017) 1424–1432.
[9] I. Altun, F. Sola, H. Simsek, H: generalized contractions on partial metric spaces, Topology Appl. 157 (2010) 2778–2785.
[10] J. Jachymski, The contraction principle for mappings on a metric space with a graph, Proc. Amer. Math. Soc. 136 (2008) 1359–1373.
[11] M. Cosentino, P. Vetro, Fixed point results for F-contractive mappings of Hardy-Rogers-type, Filomat 28 (2014) 715–722.
[12] S. Czerwik, Contraction mappings in b-metric spaces, Acta Math. Inf. Univ. Ostrav. 1 (1993) 5–11.
[13] W. Shatanawi, A. Pitea, R. Lazovic, Contraction conditions using comparison functions on b-metric spaces, Fixed Point Theory Appl. (2014) Art. No. 135.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 318–331
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

On the homogeneity test based on the kernel-type estimators of a distribution density

Petre Babilua*, Elizbar Nadaraya

Department of Mathematics, Faculty of Exact and Natural Sciences, Iv. Javakhishvili Tbilisi State University, 3 University Str., Tbilisi 0143, Georgia

Received 29 June 2018; accepted 27 July 2018
Available online 16 August 2018

Abstract

The test of homogeneity is constructed by using kernel-type estimators of a distribution density. The limit power of the constructed test is found for close Pitman-type alternatives. The constructed test is compared with Pearson's Xu-square test.

Keywords: Test for homogeneity; Goodness-of-fit test; Power of test; Test consistency; Kernel-type estimator of density

Let $X^{(i)} = (X_1^{(i)}, \dots, X_{n_i}^{(i)})$, $i = 1, \dots, p$, be independent samples of sizes $n_1, n_2, \dots, n_p$, from $p \geq 2$ general populations with distribution densities $f_1(x), \dots, f_p(x)$. Using the samples $X^{(i)}$, $i = 1, \dots, p$, it is required to check the homogeneity hypothesis

$$H_0 : \ f_1(x) = \cdots = f_p(x) \tag{1}$$

against the class of alternatives $H_1 : \ f_i(x) \neq f_j(x)$ (almost everywhere) for some $i \neq j$, $i, j = 1, \dots, p$. The general distribution density in (1) will be denoted by $f_0(x)$, i.e. $H_0 : \ f_0(x) = f_1(x) = \cdots = f_p(x)$. The homogeneity hypothesis only asserts that the distribution densities $f_i(x)$, $i = 1, \dots, p$, coincide, but it does not fix the form of $f_0(x)$.

In the present paper we construct the criterion for testing the hypothesis H_0 against a sequence of "close" Pitman type alternatives

$$H_1 : \ f_i(x) = f_0(x) + \alpha(n_0)\varphi_i(x),$$
$$\alpha(n_0) \longrightarrow 0, \quad n_0 = \min(n_1, \dots, n_p) \longrightarrow \infty, \quad \int \varphi_i(x)\, dx = 0, \quad i = 1, \dots, p.$$

* Corresponding author.
E-mail addresses: petre.babilua@tsu.ge (P. Babilua), elizbar.nadaraya@tsu.ge (E. Nadaraya).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.07.005

We consider the test of the hypothesis H_0 based on the statistic

$$T(n_1, n_2, \dots, n_p) = \sum_{i=1}^{p} m_i \int \Big[\widehat{f}_i(x) - \frac{1}{m} \sum_{j=1}^{p} m_j \widehat{f}_j(x) \Big]^2 r(x)\, dx, \tag{2}$$

where $\widehat{f}_i(x)$ is the Rosenblatt–Parzen kernel estimator of the distribution density $f_i(x)$:

$$\widehat{f}_i(x) = \frac{a_i}{n_i} \sum_{j=1}^{n_i} K\big(a_i(x - X_j^{(i)})\big), \quad m_i = \frac{n_i}{a_i}, \quad m = m_1 + \cdots + m_p.$$

The particular case $p = 2$ is considered in the works [1] and [2], where the statistic T has the explicit form

$$T(n_1, n_2) = \frac{m_1 m_2}{m_1 + m_2} \int \big(\widehat{f}_1(x) - \widehat{f}_2(x)\big)^2 r(x)\, dx.$$

1. In this section we consider the question about the limit law of distribution of the statistic (2) for the hypothesis H_1 when n_i infinitely increase so that $n_i = nk_i$, where $n \to \infty$, and k_i are constant values. Let $a_1 = a_2 = \cdots = a_p = a_n$, where $a_n \to \infty$ as $n \to \infty$.

To obtain the limit law of distribution of the functional $T_n = T(n_1, \dots, n_p)$ we introduce the following conditions as to the functions $K(x)$, $f_0(x)$, $\varphi_i(x)$, $i = 1, \dots, p$, and $r(x)$:

(i) $K(x) \geq 0$, vanishes outside the finite interval $(-A, A)$ and , together with its derivative, is continuous on this interval; or is absolutely continuous on $(-\infty, \infty)$, $x^2 K(x)$ is integrable and $K^{(1)}(x) \in L_1(-\infty, \infty)$; note that in both cases $\int K(x)\, dx = 1$.

(ii) The function $f_0(x)$ is bounded and positive on $(-\infty, \infty)$ or bounded and positive in some finite interval $[c, d]$. Besides, it has a bounded derivative in the domain of positiveness.

(iii) The functions $\varphi_j(x)$, $j = 1, \dots, p$, are bounded and have bounded derivatives of first order; also $\varphi_i(x)$ and $\varphi_i^{(1)}(x) \in L_1(-\infty, \infty)$.

(iv) The weight function $r(x)$ is piecewise-continuous, bounded and $r(x) \in L_1(-\infty, \infty)$.

The following assertion is true.

Theorem 1. *Let the conditions* (i)–(iv) *be fulfilled. If* $\alpha_n = n^{-1/2} a_n^{1/4}$ ($\alpha_n = \alpha(n_0)$), $n^{-1} a_n^{9/2} \to 0$ *as* $n \to \infty$, *then for the alternative* H_1

$$a_n^{1/2}(T_n - \mu) \xrightarrow{d} N(A(\varphi), \sigma^2),$$

where

$$A(\varphi) = \sum_{i=1}^{p} k_i \int \Big[\varphi_i(x) - \frac{1}{\overline{k}} \sum_{j=1}^{p} k_j \varphi_j(x) \Big]^2 r(x)\, dx,$$

$$\sigma^2 = 2(p-1) \int f_0^2(x) r^2(x)\, dx \cdot R(K_0), \quad K_0 = K * K,$$

$$\mu = (p-1) \int f_0(x) r(x)\, dx \cdot R(K), \quad R(g) = \int g^2(x)\, dx,$$

$$\overline{k} = k_1 + \cdots + k_p, \quad p \geq 2,$$

d denotes convergence in distribution, and $\mathbb{N}(a, b^2)$ *is a random variable having a normal distribution with mean a and dispersion* b^2.

Proof. Let us write T_n as the sum

$$T_n = T_n^{(1)} + A_{1n} + A_{2n},$$

where

$$T_n^{(1)} = \frac{n}{a_n}\sum_{i=1}^{p} k_i \int\left[\widehat{f}_i(x) - E\widehat{f}_i(x) - \frac{1}{\overline{k}}\sum_{j=1}^{p} k_j\big(\widehat{f}_j(x) - E\widehat{f}_j(x)\big)\right]^2 r(x)\,dx,$$

$$A_{1n} = 2\,\frac{n}{a_n}\sum_{i=1}^{p} k_i \int\big[\widehat{f}_i(x) - E\widehat{f}_i(x)\big]\Big[E\widehat{f}_i(x) - \frac{1}{\overline{k}}\sum_{j=1}^{p} k_j E\widehat{f}_j(x)\Big]r(x)\,dx,$$

$$A_{2n} = \frac{n}{a_n}\sum_{i=1}^{p} k_i \int\Big[E\widehat{f}_i(x) - \frac{1}{\overline{k}}\sum_{j=1}^{p} k_j E\widehat{f}_j(x)\Big]^2 r(x)\,dx.$$

Here and in what follows $E(\cdot)$ is a mathematical expectation with respect to the hypothesis H_1.

It is not difficult to see that

$$E\widehat{f}_i(x) = a_n\int K(a_n(x-u))f_0(u)\,du + \alpha_n\varphi_i(x) + \frac{\alpha_n}{a_n}\int tK(t)\int_0^1 \varphi_i^{(1)}\Big(x - \frac{tz}{a_n}\Big)\,dz\,dt.$$

From this relation we find that

$$\begin{aligned} A_{2n} = {} & \frac{n\alpha_n^2}{a_n}\sum_{i=1}^{p} k_i \int\Big[\varphi_i(x) - \frac{1}{\overline{k}}\sum_{j=1}^{p} k_j\varphi_j(x)\Big]^2 r(x)\,dx \\ & + O\Big(\frac{n\alpha_n^2}{a_n^2}\max_i \int\Big(\int tK(t)\int_0^1 \varphi_i^{(1)}\Big(x - \frac{tz}{a_n}\Big)\,dz\,dt\Big)r(x)\,dx\Big) \\ & + O\Big(\frac{n\alpha_n^2}{a_n^3}\max_i \int\Big(\int tK(t)\int_0^1 \varphi_i^{(1)}\Big(x - \frac{tz}{a_n}\Big)\,dz\,dt\Big)^2 r(x)\,dx\Big) = \frac{n\alpha_n^2}{a_n}\,A_n(\varphi) + O\Big(\frac{n\alpha_n^2}{a_n^2}\Big). \end{aligned}$$

Hence, since $\frac{n\alpha_n^2}{\sqrt{a_n}} = 1$, we obtain

$$\sqrt{a_n}\,A_{2n} = A(\varphi) + O\Big(\frac{n\alpha_n^2}{a_n^{3/2}}\Big) = A(\varphi) + O\Big(\frac{1}{a_n}\Big), \tag{3}$$

where

$$A(\varphi) = \sum_{i=1}^{p} k_i \int\Big[\varphi_i(x) - \frac{1}{\overline{k}}\sum_{j=1}^{p} k_j\varphi_j(x)\Big]^2 r(x)\,dx.$$

Now let us show that $a_n^{1/2}A_{1n} \xrightarrow{P} 0$, ($P$ means convergence in probability). For this it suffices to show that $a_n^{1/2}E|A_{1n}| \longrightarrow 0$ as $n \to \infty$. We have

$$E|A_{1n}| \le (EA_{1n}^2)^{1/2} = 2\,\frac{n}{a_n}\left\{\sum_{i=1}^{p} k_i^2 E\Big(\int\big(\widehat{f}_i(x) - E\widehat{f}_i(x)\big)A_i(x)r(x)\,dx\Big)^2\right\}^{1/2},$$

where

$$A_i(x) = E\widehat{f}_i(x) - \frac{1}{\overline{k}}\sum_{j=1}^{p} k_j E\widehat{f}_j(x).$$

Further, it is easy to calculate that

$$\begin{aligned} & E\Big[\int\big(\widehat{f}_i(x) - E\widehat{f}_i(x)\big)A_i(x)r(x)\,dx\Big]^2 = \frac{a_n^2}{k_i n}\,E\Big[\int K\big(a_n(x - X_1^{(i)})\big)A_i(x)r(x)\,dx \\ & \quad - E\int K\big(a_n(x - X_1^{(i)})\big)A_i(x)r(x)\,dx\Big]^2 \le \frac{a_n^2}{k_i n}\,E\Big[\int K\big(a_n(x - X_1^{(i)})\big)A_i(x)r(x)\,dx\Big]^2. \end{aligned}$$

Therefore

$$E|A_{1n}| \le c_1\sqrt{n}\left\{\sum_{i=1}^{p} k_i\int f_i(u)\,du\left[\int K(a_n(x-u))A_i(x)r(x)\,dx\right]^2\right\}^{1/2}. \tag{4}$$

Since $\sup_x |A_i(x)| \le c_2\alpha_n$ for all $i = 1, \dots, p$ and $r(x)$ is bounded, from (4) we obtain

$$a_n^{1/2} E|A_{1n}| \le c_3 \frac{\sqrt{n}\,\alpha_n}{a_n^{1/2}} = O\Big(\frac{1}{a_n^{1/4}}\Big).$$

Thus

$$a_n^{1/2} A_{1n} = o_p(1). \tag{5}$$

Now we proceed to calculating the limit distribution of the functional $T_n^{(1)}$:

$$T_n^{(1)} = \frac{n}{a_n} \sum_{i=1}^{p} k_i \int \Big[\widehat{f}_i(x) - E\widehat{f}_i(x) - \frac{1}{\overline{k}} \sum_{j=1}^{p} k_j \big(\widehat{f}_j(x) - E\widehat{f}_j(x)\big) \Big]^2 r(x)\,dx, \tag{6}$$

where $\overline{k} = k_1 + \cdots + k_p$.

By a simple transformation of (6) we obtain

$$T_n^{(1)} = \int \Bigg[\sum_{i=1}^{p} \Big(\sqrt{\frac{n_i}{a_n}} \big(\widehat{f}_i(x) - E\widehat{f}_i(x)\big) \Big)^2 - \Big(\sum_{j=1}^{p} \alpha_j \sqrt{\frac{n_j}{a_n}} \big(\widehat{f}_j(x) - E\widehat{f}_j(x)\big) \Big)^2 \Bigg] r(x)\,dx,$$

where $\alpha_i^2 = \frac{k_i}{k_1+\cdots+k_p}$.

Let

$$\mathbb{Z}(x) = \big(Z_1(x), \dots, Z_p(x)\big)$$

be the vector with components

$$Z_i(x) = \sqrt{\frac{n_i}{a_n}} \big(\widehat{f}_i(x) - E\widehat{f}_i(x)\big), \quad i = 1, \dots, p.$$

Then

$$T_n^{(1)} = \int \Big[|Z(x)|^2 - \Big(\sum_{j=1}^{p} \alpha_j Z_j(x) \Big)^2 \Big] r(x)\,dx,$$

where $|a|$ is the length of the vector $a = (a_1, \dots, a_p)$.

There exists the orthogonal matrix $\mathbf{C} = \|c_{ij}\|, i, j = 1, \dots, p$ which depends only on $k_1, k_2, \dots, k_p$, and for which

$$c_{pi} = \alpha_i = \sqrt{\frac{k_i}{k_1 + \cdots + k_p}}, \quad i = 1, \dots, p.$$

Since the vector length does not change under orthogonal transformation, we have

$$T_n^{(1)} = \int \Big[|\mathbb{C}\mathbb{Z}|^2 - \Big(\sum_{j=1}^{p} \alpha_j Z_j(x) \Big)^2 \Big] r(x)\,dx = \sum_{i=1}^{p-1} \int \Big(\sum_{j=1}^{p} c_{ij} Z_j(x) \Big)^2 r(x)\,dx. \tag{7}$$

Let $F_i(x)$ be the distribution function of a random variable $X_1^{(i)}$ and $\widehat{F}_{n_i}(x)$ be the empirical distribution function of the sample $X^{(i)} = (X_1^{(i)}, \dots, X_{n_i}^{(i)})$.

By Theorem 3 of [3] we can write that

$$\widehat{F}_{n_i}(x) - F_i(x) = n_i^{-1/2} W_i^0(F_i) + O_p(n^{-1} \ln n) \tag{8}$$

uniformly with respect to $x \in (-\infty, \infty)$, $W_i^0(t)$, $i = 1, \dots, p$, are the independent Brownian bridges which depend only on $X^{(i)}$.

Using (8), we can easily establish [4,5] that

$$Z_i(x) = \sqrt{\frac{n_i}{a_n}} \big(\widehat{f}_i(x) - E\widehat{f}_i(x)\big) = \xi_i(x) + O_p\Big(\frac{\ln n}{\sqrt{n a_n^{-1}}}\Big), \tag{9}$$

uniformly with respect to x, where

$$\xi_i(x) = a_n^{1/2} \int K(a_n(x-u))\,dW_i^0(F_i(u)), \quad i = 1, \dots, p.$$

Then by virtue of (9) we can write

$$\sum_{j=1}^{p} c_{ij} Z_j(x) = \sum_{j=1}^{p} c_{ij}\xi_j(x) + O_p\Big(\frac{\ln n}{\sqrt{n a_n^{-1}}}\Big), \tag{10}$$

uniformly with respect to $x \in (-\infty, \infty)$. Let us estimate the dispersion of the value

$$Y_i = \int \sum_{j=1}^{p} c_{ij}\xi_j(x) r(x)\, dx.$$

From the independence of $W_i^0(t)$, $i = 1, \dots, p$, we get

$$\operatorname{Var} Y_i = \sum_{j=1}^{p} c_{ij}^2 \iint E\xi_j(x)\xi_j(y) r(x) r(y)\, dx\, dy.$$

Furthermore, $\xi_j(x)$ can be represented as

$$\xi_j(x) = a_n^{1/2} \int \Big[K(a_n(x-t)) - \int K(a_n(x-u))\, dF_j(u) \Big] dW_j(F_j),$$

where $W_j(t)$, $j = 1, \dots, p$, are independent Wiener processes on $[0, 1]$.

Hence, after some simple transformations, we obtain

$$\iint E\xi_j(x)\xi_j(y) r(x) r(y)\, dx\, dy = O(a_n^{-1}).$$

Thus

$$\operatorname{Var} Y_j = O(a_n^{-1}), \quad j = 1, \dots, p.$$

Hence it follows that the random value

$$A_n = \sqrt{a_n} \sum_{i=1}^{p-1} \int \sum_{j=1}^{p} c_{ij}\xi_j(x) r(x)\, dx$$

is bounded in probability, i.e. $P\{|A_n| \geq M\} \longrightarrow 0$ as $M \to \infty$ uniformly with respect to n.

Therefore

$$A_n \cdot O_p\Big(\frac{\sqrt{a_n}\ln n}{\sqrt{n}}\Big) = o_p(1), \tag{11}$$

since, by assumption, $a_n^2/n \to 0$.

Thus, from the representations (7) and (10) and also from the relation (11) we find

$$\sqrt{a_n}\,(T_n^{(1)} - T_n^{(2)}) = o_p(1) + O_p\Big(\frac{a_n^{3/2}\ln^2 n}{n}\Big), \tag{12}$$

where

$$T_n^{(2)} = \sum_{i=1}^{p-1} \int \Big(\sum_{j=1}^{p} c_{ij}\xi_j(t) \Big)^2 r(t)\, dt.$$

Denote

$$\eta_i(t) = a_n^{1/2} \int K(a_n(t-u))\, dW_i(F_i(u)), \qquad T_n^{(3)} = \sum_{i=1}^{p-1} \int \Big(\sum_{j=1}^{p} c_{ij}\eta_j(t) \Big)^2 r(t)\, dt,$$

$$\varepsilon_i(t) = a_n^{1/2} W_i(1) \int K(a_n(t-u)) f_i(u)\, du.$$

Then

$$a_n^{1/2}\big(T_n^{(2)} - T_n^{(3)}\big) = o_p(1). \tag{13}$$

Indeed,

$$E\left|T_n^{(2)} - T_n^{(3)}\right| \le 2\sum_{i=1}^{p-1} E\left| \int \sum_{j=1}^{p} c_{ij}\eta_j(t) \sum_{r=1}^{p} c_{ir}\varepsilon_r(t)\right| r(t)\,dt + \sum_{i=1}^{p-1} E\int \Big(\sum_{j=1}^{p} c_{ij}\varepsilon_j(t)\Big)^2 r(t)\,dt$$
$$= B_n^{(1)} + B_n^{(2)}. \tag{14}$$

It is not difficult to see that

$$B_n^{(2)} \le c_4 a_n^{-1}.$$

Let us now estimate $B_n^{(1)}$. We have

$$B_n^{(1)} \le 2\sum_{i=1}^{p-1}\Bigg[\sum_{j,r=1}^{p} |c_{ij}c_{ir}|E|W_r(1)|\left|\int\Big[\int \Psi_r(t)K(a_n(t-u))r(t)\,dt\Big]dW_j(F_j)\right|\Bigg]$$
$$\le 2\sum_{i=1}^{p-1}\Bigg[\sum_{j=1}^{p}\sum_{r=1}^{p}|c_{ij}c_{ir}|E^{1/2}W_r^2(1)E^{1/2}\left\{\int\Big[\int \Psi_r(t)K(a_n(t-u))r(t)\,dt\Big]dW_j(F_j)\right\}^2\Bigg]$$
$$= 2\sum_{i=1}^{p-1}\sum_{j=1}^{p}\sum_{r=1}^{p}|c_{ij}c_{ir}|\left\{\int\Big(\int \Psi_r(t)K(a_n(t-u))r(t)\,dt\Big)^2 dF_j(u)\right\}^{1/2} \le c_5 a_n^{-1},$$

where

$$\Psi_r(t) = \int K(z)f_r(t - za_n^{-1})\,dz.$$

So, after substituting the estimators of $B_n^{(1)}$ and $B_n^{(2)}$ into (14), we find

$$\sqrt{a_n}\,(T_n^{(2)} - T_n^{(3)}) = o_p(1). \tag{15}$$

Denote

$$\eta_i^0(t) = a_n^{1/2}\int K(a_n(t-x))\,dW_i(F_0),$$

where $F_0(x)$ is the distribution function with density $f_0(x)$. Since $F_i(x) = F_0(x) + \alpha_n U_i(x)$, $U_i^{(1)}(x) = \varphi_i(x)$ and, by assumption, $\varphi_i(x)$ is bounded and $K^{(1)}(x) \in L_1(-\infty,\infty)$, we have

$$\begin{aligned} E\big(\eta_j(t) - \eta_j^0(t)\big)^2 &= O(a_n\alpha_n),\\ E(\eta_j^0(t))^2 &= O(1), \end{aligned} \tag{16}$$

uniformly with respect to $t \in (-\infty,\infty)$ j, $j = 1,\dots,p$. Indeed, we have

$$E\big(\eta_j(t) - \eta_j^0(t)\big)^2 = a_n E\left(\int\Big(W_j(F_j(a_n(t-x))) - W_j(F_0(a_n(t-x)))\Big)K^{(1)}(x)\,dx\right)^2$$
$$\le a_n\int E\Big(W_i(F_i(a_n(t-x))) - W_i(F_0(a_n(t-x)))\Big)^2|K^{(1)}(z)|\,dx\int|K^{(1)}(x)|\,dx$$
$$\le a_n\int\Big|F_j(a_n(t-x)) - F_0(a_n(t-x))\Big|\,|K^{(1)}(z)|\,dx\cdot\int|K^{(1)}(x)|\,dx \le c_6 a_n\alpha_n,$$

and also

$$E(\eta_n^0(t))^2 = a_n\int K^2(a_n(t-x))f_0(x)\,dx \le \max_x f_0(x)\cdot\int K^2(u)\,du.$$

Next, using (16) and the Cauchy–Schwarz inequality, we establish that

$$\sqrt{a_n}\,E|T_n^{(3)} - T_n^{(4)}| = O(a_n\sqrt{\alpha_n}) + O(a_n^{3/2}\alpha_n), \tag{17}$$

where

$$T_n^{(4)} = \sum_{i=1}^{p-1} \int \Big(\sum_{j=1}^{p} c_{ij} \eta_j^0(t) \Big)^2 r(t)\, dt.$$

Let us now consider the limit distribution of the functional $T_n^{(4)}$.

The processes $\eta_j^0(t)$, $j = 1, \dots, p$, are independent and Gaussian and thus the new processes $\sum_{j=1}^{p} c_{ij} \eta_j^0(t)$, $i = 1, \dots, p$, are also independent and Gaussian due to the orthogonality of the matrix $\|c_{ij}\|$. Therefore to find the limit distribution of $T_n^{(4)}$ it remains to establish the limit distribution of the functional

$$U_n^{(i)} = \int \Big(\sum_{j=1}^{p} c_{ij} \eta_j^0(t) \Big)^2 r(t)\, dt$$

for each fixed i, $i = 1, \dots, p-1$, $p \geq 2$. The covariational function $R_n^{(i)}(t_1, t_2)$ of the Gaussian process $\sum_{j=1}^{p} c_{ij} \eta_j^0(t)$ is equal to

$$R_n^{(i)}(t_1, t_2) = \sum_{j=1}^{p} c_{ij}^2 E \eta_j^0(t_1) \eta_j^0(t_2).$$

However,

$$E \eta_j^0(t_1) \eta_j^0(t_2) = \int K(u) K\big(a_n(t_1 - t_2) + u\big) f_0(t_1 - a_n^{-1} u)\, du = f_0(t_1) K_0\big(a_n(t_1 - t_2)\big) + O(a_n^{-1}), \tag{18}$$

where the estimator $O(\cdot)$ is uniform with respect to t_1, t_2 and $K_0 = K * K$.

From (18) it follows that

$$R_n^{(i)}(t_1, t_2) = f_0(t_1) K_0\big(a_n(t_1 - t_2)\big) + O(a_n^{-1}). \tag{19}$$

A semi-invariant $\chi_n^{(i)}(s)$ of order s of a random variable $U_n^{(i)}$ is defined by the formula [6]:

$$\chi_n^{(i)}(s) = (s-1)! \cdot 2^{s-1} \int \cdots \int R_n^{(i)}(x_1, x_2) R_n^{(i)}(x_2, x_3) \cdots R_n^{(i)}(x_s, x_1) r(x_1) r(x_2) \cdots r(x_s)\, dx_1\, dx_2 \cdots dx_s. \tag{20}$$

From (19) and (20) it is not difficult to establish that

$$\begin{aligned} E U_n^{(i)} &= \chi_n^{(i)}(1) = R(K) \int f_0(x) r(x)\, dx + O(a_n^{-1}), \\ \operatorname{Var} U_n^{(i)} &= \chi_n^{(i)}(2) = 2R(K_0) a_n^{-1} \int f_0^2(x) r^2(x)\, dx + o(a_n^{-1}), \end{aligned} \tag{21}$$

and the sth semi-invariant $\chi_n^{(i)}(s)$ is equal with an accuracy of terms of higher order smallness to [6]:

$$(s-1)!\, 2^{s-1} (a_n^{-1})^{s-1} [K * K]^{(s)}(0) \int f_0^s(x) r^s(x)\, dx, \tag{22}$$

where $[K * K]^{(s)}(0)$ is the s-multiple convolution of $K_0(x)$ with itself.

From the relations (21) and (22) it follows [6,5] that

$$a_n^{1/2} \left(U_n^{(i)} - R(K) \int f_0(x) r(x)\, dx \right)$$

is distributed in the limit normally with mathematical expectation 0 and dispersion

$$2R(K_0) \int f_0^2(u) r^2(u)\, du, \quad R(g) = \int g^2(x)\, dx,$$

and therefore $\sqrt{a_n}\,(T_n^{(4)} - \mu)$ is distributed in the limit normally with $(0, \sigma^2)$.

Finally, taking into account (3), (5), (12), (13), (17) and the representation

$$a_n^{1/2}(T_n-\mu)=a_n^{1/2}(T_n^{(4)}-\mu)+A(\varphi)+O(a_n^{-1/2})+o_p(1)+O_p\Big(\frac{a_n^{3/2}\ln^2 n}{n}\Big)$$
$$+O_p(a_n\sqrt{\alpha_n})+O(a_n^{3/2}\alpha_n), \tag{23}$$

we conclude that $a_n^{1/2}(T_n-\mu)$ is distributed in the limit normally with $(A(\varphi),\sigma^2)$. □

The conditions of Theorem 1 as to a_n and α_n are fulfilled, for instance, if it is assumed that $a_n=n^{\delta}$, $\alpha_n=n^{-1/2+\delta/4}$ for $0<\delta<\frac{2}{9}$.

Let us introduce the notation

$$f_n^*(x)=\frac{1}{\overline{k}}\sum_{j=1}^{p}k_j\widehat{f}_j(x),$$

$$\overline{\mu}_n=\int f_n^*(x)r(x)\,dx,\quad \Delta_n^2=\frac{1}{\overline{k}}\sum_{i=1}^{p}k_i\Delta_{in}^2,\quad \Delta_{in}^2=\int\widehat{f}_i^{\,2}(x)r^2(x)\,dx.$$

Theorem 2. *Let all conditions of Theorem* 1 *be fulfilled. Then* $a_n^{1/2}(T_n-\mu_n)\sigma_n^{-1}$ *for the hypothesis* H_1 *is distributed in the limit normally with* $(A(\varphi)\sigma^{-1},1)$*, where*

$$\mu_n=(p-1)R(K)\overline{\mu}_n,\quad \sigma_n^2=2(p-1)R(K_0)\Delta_n^2.$$

Proof. It is obvious that

$$a_n^{1/2}(T_n-\mu_n)\sigma_n^{-1}=a_n^{1/2}(T_n-\mu)\sigma^{-1}(\sigma\sigma_n^{-1})+a_n^{1/2}(\mu-\mu_n)\sigma_n^{-1}.$$

Thus it suffices to show that

$$a_n^{1/2}\Big(\overline{\mu}_n-\int f_0(x)r(x)\,dx\Big)=o_p(1) \tag{24}$$

and

$$\Delta_n^2-\int f_0^2(x)r^2(x)\,dx=o_p(1). \tag{25}$$

But (25) directly follows from Theorem 2.1 due to Bhattacharyya G. K., Roussas G. G. [7] (see also [5,8]).

Let us prove (24). We have

$$a_n^{1/2}E\Big|\int f_n^*(x)r(x)\,dx-\int f_0(x)r(x)\,dx\Big|$$
$$\le a_n^{1/2}E\Big|\int\big(f_n^*(x)-Ef_n^*(x)\big)r(x)\,dx\Big|+a_n^{1/2}\int\big|Ef_n^*(x)-f_0(x)\big|r(x)\,dx=A_{1n}+A_{2n}.$$

It is not difficult to verify that

$$A_{2n}\le c_7\big(a_n^{-1/2}+\sqrt{a_n}\,\alpha_n\big).$$

Furthermore, we have

$$A_{1n}\le a_n^{1/2}E^{1/2}\Big(\int\big(f_n^*(x)-Ef_n^*(x)\big)r(x)\,dx\Big)^2\le c_8a_n^{1/2}\max_{1\le j\le p}\Big\{\frac{1}{n}\int f_j(u)\,du\Big(\int K(t)r\Big(u-\frac{t}{a_n}\Big)\,dt\Big)^2\Big\}^{1/2}$$
$$\le c_9\Big(\frac{a_n}{n}\Big)^{1/2}.$$

Therefore

$$A_{1n}+A_{2n}\le c_{10}\Big(a_n^{-1/2}+\sqrt{a_n}\,\alpha_n+\Big(\frac{a_n}{n}\Big)^{1/2}\Big)\longrightarrow 0.\quad\square$$

Theorem 2 gives rise to two corollaries.

Corollary 1. *Let the conditions (i), (ii) and (iv) be fulfilled. If* $n^{-1}a_n^2 \to 0$*, then for the hypothesis* H_0

$$a_n^{1/2}(T_n - \mu_n)\sigma_n^{-1} \xrightarrow{d} N(0,1).$$

This result allows us to construct an asymptotic criterion for testing the hypothesis $H_0: \ f_1(x) = \cdots = f_p(x)$ (homogeneity hypothesis); the critical region is defined by the inequality

$$T_n \geq d_n(\alpha) = \mu_n + a_n^{-1/2}\sigma_n\lambda_\alpha,$$

where λ_α is a quantile of level $1-\alpha$ of the quadratic normal distribution $\Phi(x)$.

Corollary 2. *Under the conditions of Theorem* 2 *the local behavior of the power* $P_{H_1}(T_n \geq \widetilde{d}_n(\alpha))$ *is as follows*

$$P_{H_1}(T_n \geq d_n(\alpha)) \longrightarrow 1 - \Phi\Big(\lambda_\alpha - \frac{A(\varphi)}{\sigma^{-1}}\Big).$$

Let $\inf_{0\leq x\leq 1} f_0(x) > 0$ and $r(x) = f_0^{-1}(x)$, $x \in [0,1]$, and $r(x) = 0$, $x \notin [0,1]$. In this case, for the hypothesis H_1

$$a_n^{1/2}(T_n - \mu_0) \xrightarrow{d} N(A(\varphi), \sigma_0^2),$$

where

$$\mu_0 = (p-1)\int K^2(u)\,du, \quad \sigma_0^2 = 2(p-1)\int K_0^2(u)\,du.$$

Denote

$$\widehat{T}_n = \frac{n}{a_n}\sum_{i=1}^{p} k_i \int_0^1 \Big[\widehat{f}_i(x) - \frac{1}{\overline{k}}\sum_{j=1}^{p} k_j \widehat{f}_j(x)\Big]^2 r_n(x)\,dx, \quad r_n(x) = [f_n^*(x)]^{-1}.$$

Theorem 3. *Let the conditions (i)–(iv) of Theorem* 1 *be fulfilled. Let* $\alpha_n = n^{-1/2}a_n^{1/2}$ *and* $n^{-1}a_n^{9/2}\ln n \to 0$*. Then*

$$a_n^{1/2}(\widehat{T}_n - \mu_0) \xrightarrow{d} N(A(\varphi), \sigma_0^2) \ \textit{for Hypothesis } H_1.$$

Proof. To prove the theorem it suffices to show that $\sqrt{a_n}\,(T_n - \widehat{T}_n) \xrightarrow{P} 0$.

We have

$$\sqrt{a_n}\,|T_n - \widehat{T}_n| \leq L_n^{(1)} \cdot L_n^{(2)}, \tag{26}$$

$$L_n^{(1)} = \sqrt{a_n}\sup_{0\leq x\leq 1}|f_n^*(x) - f_0(x)|\Big(\inf_{0\leq x\leq 1}(f_0(x)f_n^*(x))\Big)^{-1},$$

$$L_n^{(2)} = \frac{n}{a_n}\sum_{i=1}^{p} k_i \int_0^1 \Big[\widehat{f}_i(x) - \frac{1}{\overline{k}}\sum_{j=1}^{p} k_j \widehat{f}_j(x)\Big]^2 dx.$$

Since $E\widehat{f}_j(x) - f_0(x) = O(\frac{1}{a_n}) + O(\alpha_n)$ uniformly with respect to x, we obtain

$$\begin{aligned}\sqrt{a_n}\sup_x|f_n^*(x) - f_0(x)| &\leq \sum_{i=1}^{p} k_i\sqrt{a_n}\sup_x|\widehat{f}_i(x) - E\widehat{f}_j(x)| + \sqrt{a_n}\sum_{j=1}^{p} k_j \sup_x|E\widehat{f}_j(x) - f_0(x)| \\ &\leq \sum_{i=1}^{p} k_i\sqrt{a_n}\sup_x|\widehat{f}_i(x) - E\widehat{f}_i(x)| + O\Big(\frac{1}{\sqrt{a_n}}\Big) + O(\sqrt{a_n}\,\alpha_n).\end{aligned}$$

Next, from the inequality (2) of [9] (see also [5, p. 43]) it follows that

$$\sqrt{a_n}\sup_x|\widehat{f}_i(x) - E\widehat{f}_i(x)| \leq V_0 a_n^{3/2}\sup_x|\widehat{F}_i(x) - F_i(x)| = O(n^{-1}a_n^3\ln n)^{1/2} \ \text{ with probability 1,}$$

where $V_0 = \bigvee_{-\infty}^{\infty}(K)$. From this and the condition $n^{-1}a_n^{9/2}\ln n \to 0$ we have

$$\sqrt{a_n}\sup_x|f_n^*(x) - f_0(x)| \longrightarrow 0 \ \text{ with probability 1.} \tag{27}$$

Furthermore, since

$$\inf_{0\le x\le 1} f_0(x)f_n^*(x) \ge \Delta_0 \inf_{0\le x\le 1} f_n^*(x), \quad \Delta_0 = \inf_{0\le x\le 1} f_0(x) > 0$$

and

$$\inf_{0\le x\le 1} f_n^*(x) \ge \Delta_0 - \sup_{0\le x\le 1} |f_n^*(x) - f_0(x)|,$$

this and (27) imply that $L_n^{(1)} \to 0$ with probability 1.

Further we have

$$EL_n^{(2)} = \frac{n}{a_n}\sum_{i=1}^{p} k_i \int_0^1 E\Big[\widehat{f}_i(x) - E\widehat{f}_i(x) - \frac{1}{\overline{k}}\sum_{j=1}^{p} k_j E(\widehat{f}_j(x) - E\widehat{f}_i(x))\Big]^2 dx$$
$$+ \frac{n}{a_n}\sum_{i=1}^{p} k_i \int_0^1 \Big[E\widehat{f}_i(x) - \frac{1}{\overline{k}}\sum_{j=1}^{p} k_j E\widehat{f}_j(x)\Big]^2 dx. \tag{28}$$

It is not difficult to verify that

$$\int \operatorname{Var}\widehat{f}_i(x)\,dx = O(\frac{a_n}{n}), \quad E\widehat{f}_i(x) = a_n \int k(a_n(x-u))f_0(u)\,du + O(\alpha_n).$$

Therefore from (28) we establish that

$$EL_n^{(2)} \le c_{12} + c_{13}\frac{n}{a_n}\alpha_n^2 = c_{12} + c_{13}\frac{1}{\sqrt{a_n}}.$$

This means that $L_n^{(2)}$ is bounded in probability. Thus $L_n^{(1)} \cdot L_n^{(2)} \xrightarrow{P} 0$. Therefore $\sqrt{a_n}(T_n)\widehat{T}_n \xrightarrow{P} 0$. □

From Theorem 3 we deduce two corollaries.

Corollary 3. *Let the conditions (i), (ii) and (iv) be fulfilled. If $n^{-1}a_n^3 \ln n \to 0$, then for the hypothesis H_0*

$$a_n^{1/2}(\widehat{T}_n - \mu_0) \xrightarrow{d} N(0, \sigma_0^2).$$

The assertion of the corollary allows us to construct a criterion of the asymptotic condition α, $0 < \alpha < 1$, for testing the hypothesis $H_0: f_1(x) = \cdots = f_p(x)$; the critical region is defined by the inequality

$$\widehat{T}_n \ge \widetilde{d}_n(\alpha) = \mu_0 + a_n^{-1/2}\lambda_\alpha\sigma_0, \tag{29}$$
$$\mu_0 = (p-1)\int K^2(u)\,du, \quad \sigma_0^2 = 2(p-1)\int K_0^2(u)\,du.$$

Corollary 4. *By the conditions of Theorem* 3 *the local behavior of the power $P_{H_1}(\widehat{T}_n \ge \widetilde{d}_n(\alpha))$ is as follows*

$$P_{H_1}(\widehat{T}_n \ge \widetilde{d}_n(\alpha)) \longrightarrow 1 - \Phi\Big(\lambda_\alpha - \frac{A(\varphi)}{\sigma_0}\Big). \tag{30}$$

2. Let $f_0(x)$ be the distribution density on $[0, 1]$ and $f_0(x) \ge \mu > 0$, $x \in [0, 1]$. Let further $p = 2$. We give **an example** in which the power of the test $\widehat{T}_n = \widehat{T}(n_1, n_2)$ is compared with the power of Pearson's test $\chi_n^2 = \chi_{n_1 n_2}^2$:

$$\chi_{n_1 n_2}^2 = n_1 n_2 \sum_{i=1}^{s_n} \frac{1}{\nu_i + \mu_i}\Big(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2}\Big)^2,$$

where ν_i and μ_i are respectively the numbers of observations from the first sample $X^{(1)}$ and the second sample $X^{(2)}$, on the interval $\Delta_i = (j-1)s_n^{-1} \le x \le js_n^{-1}$, $|\Delta_j| = \frac{1}{s_n}$, $j = 1, \ldots, s_n = [a_n]$; $[a_n]$ denotes the integer part of a_n.

Let $f_i(x) = f_0(x) + \alpha_n\varphi_i(x)$, $i = 1,2$, $\alpha_n = n^{-1/2+\delta/4}$, $0 < \delta < \frac{2}{9}$, and

$$K(u) = \begin{cases} 1, & |x| \le \frac{1}{2} \\ 0, & |x| > \frac{1}{2}. \end{cases}$$

Then

$$\sigma_0^2 = 2\int K_0^2(x)\,dx = 2\int_{|x|\le 1}(1-|x|)^2\,dx = \frac{4}{3}.$$

From (29) and (30) we have

$$P_{H_1}(\widehat{T}_n \ge \widetilde{d}_n(\alpha)) \longrightarrow 1-\Phi\left(\lambda_\alpha - \frac{\sqrt{3}}{2}\,\frac{k_1k_2}{k_1+k_2}\int_0^1 \varphi^2(x) f_0^{-1}(x)\,dx\right) \quad n\to\infty, \tag{31}$$

where $\varphi(x)=\varphi_1(x)-\varphi_2(x)$, $\widetilde{d}_n(\alpha)=1+\frac{2}{\sqrt{3}}\,a_n^{-1/2}\lambda_\alpha$.

Let, $\widehat{f_i}(x)$, $i=1,2$, be the estimators of histogram type

$$\widehat{f_i}(x) = \frac{s_n}{n_i}\sum_{j=1}^{n_i}\delta_n(x, X_j^{(i)}), \quad i=1,2,$$

where

$$\delta_n(x,y)=\sum_{k=1}^{s_n} I_k(x)I_k(y),$$

$I_k(\cdot)$ is the indicator of the interval Δ_k.

Denote

$$X_n^2 = X_{n_1n_2}^2 = n_1n_2\sum_{i=1}^{s_n}\Big(\frac{\nu_i}{n_1}-\frac{\mu_i}{n_2}\Big)^2\,\frac{1}{n_1p_i+n_2p_i'},$$

where

$$p_i=\int_{\Delta_i} f_1(x)\,dx, \quad p_i'=\int_{\Delta_i} f_2(x)\,dx.$$

We readily see that

$$\overline{T}_n = \overline{T}(n_1,n_2) = \frac{N_1N_2}{N_1+N_2}\int\big[\widehat{f_1}(x)-\widehat{f_2}(x)\big]^2 r_n(x)\,dx = \frac{X_n^2}{s_n}, \tag{32}$$

where

$$r_n(x)=\frac{n_1+n_2}{n_1E\,\widehat{f_1}(x)+n_2E\,\widehat{f_2}(x)}, \quad N_i=\frac{n_i}{s_n}, \quad i=1,2.$$

Since $f_0(x)$ and $\varphi_i(x)$, $i=1,2$, satisfy respectively the conditions (ii) and (iii) of Theorem 1, and $f_0(x)\ge\mu>0$, $x\in[0,1]$, we have

$$r_n(x)=\frac{1}{f_0(x)}+O(\alpha_n)+O(a_n^{-1}) \tag{33}$$

uniformly with respect to $x\in[0,1]$.

Our goal is to establish for the hypothesis H_1 the limit distribution $\overline{T}_n$, after that to define the limit distribution of the statistic X_n^2 and then of χ_n^2.

Let us follow the course of the proof of basic Theorem 1. The analogues of the equalities (3), (5), (12), (13) and (17) hold true also in the case of the histogram $\widehat{f_i}(x)$, $i=1,2$, and

$$A(\varphi)=\frac{k_1k_2}{k_1+k_2}\int(\varphi_1(x)-\varphi_2(x))^2 f_0^{-1}(x)\,dx.$$

Like in the case of the histogram the analogue $T_n^{(4)}$ is defined by

$$\overline{T}_n^{(4)} = \int\Big(\sum_{j=1}^{2} c_{1j}\eta_j^0(t)\Big)^2 r_n(x)\,dx, \quad \eta_j^0(t)=s_n^{1/2}\int\delta_n(t,x)\,dW_j(F_0(x)).$$

But, according to (33), we have

$$\sqrt{s_n}\,E|\overline{T}_n^{(4)}-\overline{T}_n^{(5)}| \le c_{14}\sqrt{s_n}\,\alpha_n\int E(\eta_1^0(t))^2\,dt = c_{15}\sqrt{s_n}\,\alpha_n\int dt\int\delta_n^2(t,x)f_0(x)\,dx \le c_{16}\sqrt{s_n}\,\alpha_n, \tag{34}$$

where

$$\overline{T}_n^{(5)} = \int \Big(\sum_{j=1}^{2} c_{1j}\eta_j^0(t)\Big)^2 f_0^{-1}(t)\,dt.$$

Further, $\overline{T}_n^{(5)}$ can be represented as

$$\overline{T}_n^{(5)} = \sum_{k=1}^{s_n} \xi_k^2,$$

where

$$\xi_k = \sqrt{s_n}\Big(\int I_k(t) f_0^{-1}(t)\,dt\Big)^{1/2}\Big(\sum_{j=1}^{2} c_{1j}\int I_k(x)\,dW_j(F_0)\Big).$$

It is clear that $E\xi_k = 0$ and

$$E(\xi_{k_1}\xi_{k_2}) = \begin{cases} 0, & k_1 \neq k_2, \\ \sigma_k^2 = s_n \int I_k(x) f_0^{-1}(x)\,dx \int I_k(x) f_0(x)\,dx, & k_1 = k_2 = k. \end{cases}$$

Thus the random variables ξ_k are independent and distributed normally with $(0, \sigma_k^2)$, $k = 1, \ldots, s_n$, and by virtue of the condition

$$\min_{0\le x\le 1} f_0(x) \ge \mu > 0 \qquad \max_{0\le x\le 1} |f_0'(x)| \le M < \infty,$$

we have

$$\sum_{k=1}^{s_n} E\xi_k^2 = 1 + O(s_n^{-1}), \quad \sum_{k=1}^{s_n} \operatorname{Var}\xi_k^2 = 2s_n^{-1} + O(s_n^{-2}).$$

Indeed, since $\operatorname{Var}\xi_k^2 = 2(E\xi_k^2)^2$, we have

$$\sum_{k=1}^{s_n} \operatorname{Var}\xi_k^2 = 2s_n^{-1} + L_{1n} + L_{2n},$$

$$|L_{1n}| \le 2s_n^{-2}\sum_{k=1}^{s_n} \frac{|f_0(\theta_k) - f_0(\tau_k)|}{f_0(\theta_k) f_0(\tau_k)} \le 2M\mu^{-2}s_n^{-2}\sum_{k=1}^{s_n}|\theta_k - \tau_k| \le c_{17}s_n^{-2},$$

$$|L_{2n}| \le 2M^2\mu^{-4}s_n^{-2}\sum_{k=1}^{s_n}|\theta_k - \tau_k|^2 \le c_{18}s_n^{-3},$$

where $\theta_k, \tau_k \in \Delta_k$, $k = 1, \ldots, s_n$. In a similar manner one can prove the second relation too.

Using the Lyapunov Central Limit Theorem (Lyapunov fraction $L_n \le c_{19}\frac{1}{\sqrt{s_n}}$), let us establish that $\sqrt{s_n}\,(\overline{T}_n^{(5)} - 1)$ is distributed in the limit normally with $(0, \sqrt{2}\,)$.

Taking (34) into account we can write an analogue of the representation (23) and state that for the hypothesis H_1 $\sqrt{s_n}\,(\overline{T}_n - 1)$ is distributed in the limit normally with $(A(\varphi), \sqrt{2}\,)$.

Further, since $\overline{T}_n = \overline{T}(n_1, n_2) = \frac{X_n^2}{s_n}$ (see (32)), we have

$$\frac{\sqrt{s_n}\,(\overline{T}_n - 1)}{\sqrt{2}} = \frac{X_n^2 - s_n}{\sqrt{2s_n}}.$$

Hence it follows that

$$s_n^{-1/2}(X_n^2 - s_n)\cdot 2^{-1/2} \xrightarrow{d} N\Big(\frac{A(\varphi)}{\sqrt{2}}, 1\Big)$$

for the hypothesis H_1.

Let us now proceed to proving the fact that $s_n^{-1/2}(\chi_n^2 - X_n^2) \xrightarrow{P} 0$. For this it suffices to show that

$$s_n^{-1/2} E|\chi_n^2 - X_n^2| \longrightarrow 0 \quad n \to \infty.$$

By assumption, ν_i μ_i are independent random variables distributed according to the binomial laws (n_1, p_i) and (n_2, p_i'), respectively. If the value of the random variable $\nu_i + \mu_i$ is equal to zero, then we will assume that the corresponding summands in χ_n^2 X_n^2 are also equal to zero. By this assumption, in what follows it will be assumed that $\nu_i + \mu_i \geq 1$.

By the above argumentation we obtain

$$s_n^{-1/2} E|\chi_n^2 - X_n^2| \leq E\Big|\sum_{i=1}^{s_n} R_i Z(\nu_i, \mu_i)\Big(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2}\Big)^2\big(\nu_i - n_1 p_i - (\mu_i - n_2 p_i')\big)\Big|,$$

where

$$Z(\nu_i, \mu_i) = \begin{cases} 0 \ \text{ if } \ \nu_i + \mu_i = 0, \\ \dfrac{1}{\nu_i + \nu_i} \ \text{ if } \ \nu_i + \mu_i \geq 1, \end{cases} \qquad R_i = \frac{n_1 n_2}{n_1 p_i + n_2 p_i'} \cdot s_n^{-1/2}.$$

From the Cauchy–Schwarz inequality we can write

$$s_n^{-1/2} E|\chi_n^2 - X_n^2| \leq \sum_{i=1}^{s_n} R_i E^{1/2} Z^2(\nu_i, \mu_i) E^{1/4}\big(\nu_i - n_1 p_i - (\mu_i - n_2 p_i')\big)^4 \cdot E^{1/4}\Big(\frac{\nu_i}{n_1} - \frac{\mu_i}{n_2}\Big)^8. \tag{35}$$

Further, we have

$$EZ^2(\nu_i, \mu_i) = \sum_{k_1 + k_2 \geq 1} \Big(\frac{1}{k_1 + k_2}\Big)^2 p_1(k_1) p_2(k_2),$$
$$p_1(k_1) = C_{n_1}^{k_1} p_i^{k_1} q_i^{n_1 - k_1}, \quad p_2(k_2) = C_{n_2}^{k_2} (p_i')^{k_2} (q_i')^{n_2 - k_2}.$$

It is easy to see that

$$\begin{aligned} EZ^2(\nu_i, \mu_i) &\leq \frac{1}{2}\Big[\sum_{k_1 \geq 1,\, k_2 \geq 0} \Big(\frac{1}{k_1}\Big)^2 p_1(k_1) p_2(k_2) + \sum_{k_1 \geq 0,\, k_2 \geq 1} \Big(\frac{1}{k_2}\Big)^2 p_2(k_2) p_1(k_1)\Big] \\ &= \frac{1}{2}\Big[\sum_{k_1=1}^{n_1} \Big(\frac{1}{k_1}\Big)^2 p_1(k_1) + \sum_{k_2=1}^{n_2} \Big(\frac{1}{k_2}\Big)^2 p_2(k_2)\Big]. \end{aligned} \tag{36}$$

In the proof of Lemma 1 in [10, p. 1184] it is shown that

$$\sum_{k=1}^{N} \frac{1}{k^m} C_N^k p^k q^{N-k} \leq \Big(\frac{m+1}{Np}\Big)^m.$$

From this and (36) we have that

$$EZ^2(\nu_i, \mu_i) \leq \frac{1}{2}\Big[\Big(\frac{3}{n_1 p_i}\Big)^2 + \Big(\frac{3}{n_2 p_i'}\Big)^2\Big]. \tag{37}$$

For H_1 and by the condition $f_0(x) \geq \mu > 0$, $x \in [0, 1]$, the following inequalities take place for large values of n:

$$c_{20}\frac{n_1}{s_n} \leq n_1 p_i \leq c_{21}\frac{n_1}{s_n}, \quad c_{22}\frac{n_2}{s_n} \leq n_2 p_i' \leq c_{23}\frac{n_2}{s_n}. \tag{38}$$

Therefore (37) implies

$$R_i E^{1/2} Z^2(\nu_i, \mu_i) = O(s_n^{3/2}). \tag{39}$$

Next, applying (38) and the expressions from [11] for $E(\nu - Np)^4$ and $E(\frac{\nu}{N} - p)^8$ (ν has the binomial distribution (N, p)), we obtain

$$\begin{aligned} &E(\nu_i - n_1 p_i)^4 = O(n^2 s_n^{-2}), \quad E(\mu_i - n_2 p_i')^4 = O(n^2 s_n^{-2}), \\ &E\Big(\frac{\nu_i}{n_1} - p_i\Big)^8 = O((n s_n)^{-4}), \quad E\Big(\frac{\mu_i}{n_2} - p_i'\Big)^8 = O((n s_n)^{-4}). \end{aligned} \tag{40}$$

Finally, using $|p_i - p_i'| = O(\frac{\alpha_n}{s_n})$ and substituting (39) and (40) into (35), we find

$$s_n^{-1/2} E|\chi_n^2 - X_n^2| \le c_{24} \frac{s_n}{\sqrt{n}} + c_{25}\sqrt{n}\,\alpha_n^2 \longrightarrow 0.$$

Therefore

$$\frac{\chi_{n_1 n_2}^2 - s_n}{\sqrt{2s_n}} \xrightarrow{d} N\Big(\frac{A(\varphi)}{\sqrt{2}}, 1\Big)$$

for the hypothesis H_1 (the authors presume that the result for $A(\varphi) = 0$ is known with high probability). From the proven facts it also follows that the local behavior of the power of the test χ_n^2 is defined as follows

$$P_{H_1}\big(\chi_n^2 \ge s_n + \lambda_\alpha\sqrt{2s_n}\big) \longrightarrow 1 - \Phi\left(\lambda_\alpha - \frac{k_1 k_2}{k_1 + k_2}\,\frac{1}{\sqrt{2}}\int_0^1 \varphi^2(u) f_0^{-1}(x)\,ds u\right) \quad n \to \infty \tag{41}$$

where $\varphi(u) = \varphi_1(u) - \varphi_2(u)$. Comparing (31) and (41), we conclude that the asymptotic test $\widehat{T}(n_1 n_2)$ is more powerful than the test $\chi_{n_1 n_2}^2$ as regards the alternative hypothesis H_1.

References

[1] N.H. Anderson, P. Hall, D.M. Titterington, Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates, J. Multivariate Anal. 50 (1) (1994) 41–54.

[2] E.A. Nadaraya, Limit distribution of the quadratic deviation of two nonparametric estimators of the density of a distribution. (Russian), Soobshch. Akad. Nauk Gruz. SSR 78 (1975) 25–28.

[3] J. Komlós, P. Major, G. Tusnády, An approximation of partial sums of independent RV's and the sample DF, I. Z. Wahrscheinlichkeitstheorie Verwandte. Geb. 32 (1975) 111–131.

[4] P. Hall, Limit theorems for stochastic measures of the accuracy of density estimators, Stochastic Process. Appl. 13 (1) (1982) 11–25.

[5] E.A. Nadaraya, Nonparametric Estimation of Probability Densities and Regression Curves, in: Mathematics and its Applications (Soviet Series), vol. 20, Kluwer Academic Publishers Group, Dordrecht, 1989 Translated from the Russian by Samuel Kotz.

[6] P.J. Bickel, M. Rosenblatt, On some global measures of the deviations of density function estimates, Ann. Statist. 1 (1973) 1071–1095.

[7] G.K. Bhattacharyya, G.G. Roussas, Estimation of a certain functional of a probability density function, Skand. Aktuarietidskr 1969 (1969) 201–206 (1970).

[8] D.M. Mason, E.A. Nadaraya, G.A. Sokhadze, Integral functionals of the density, in: Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Honor of Professor Jana Jurečková, in: Inst. Math. Stat. Collect., vol. 7, Inst. Math. Statist., Beachwood, OH, 2010, pp. 153–168.

[9] E.A. Nadaraya, On non-parametric estimates of density functions and regression. (Russian), Teor. Verojatnost. i Primenen. 10 (1965) 199–203.

[10] I.W. McKeague, K.J. Utikal, Inference for a nonlinear counting process regression model, Ann. Statist. 18 (3) (1990) 1172–1187.

[11] G. Kramer, Mathematical methods of statistics. (Russian) Translated from the English by A. S. Monin and A. A. Petrov. Edited by A. N. Kolmogorov. Second, unrevised edition. With a supplement to the second edition by A. V. Prohorov. Izdat. "Mir", Moscow, 1975.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 332–336
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

On the optimal stopping with incomplete data

Petre Babilua, Besarion Dochviri, Zaza Khechinashvili*

Ivane Javakhishvili Tbilisi State University, Georgia

Received 10 July 2018; received in revised form 25 July 2018; accepted 28 July 2018
Available online 13 August 2018

Abstract

The Kalman–Bucy continuous model of partially observable stochastic processes is considered. The problem of optimal stopping of a stochastic process with incomplete data is reduced to the problem of optimal stopping with complete data. The convergence of payoffs is proved when $\varepsilon_1 \to 0$, $\varepsilon_2 \to 0$, where ε_1, and ε_2 are small perturbation parameters of the non observable and observable processes respectively.

Keywords: Partially observable process; Gain function; Payoff; Stopping time; Optimal stopping

1. Introduction

On the probability space $(\Omega, \mathcal{F}, P)$ we consider a partially observable stochastic process (θ_t, ξ_t), $0 \le t \le T$, of Kalman–Bucy model

$$d\theta_t = [a_0(t) + a_1(t)\theta_t]dt + \varepsilon_1 dw_1(t), \tag{1}$$

$$d\xi_t = d\theta_t + \varepsilon_2 dw_2(t), \tag{2}$$

where $\varepsilon_1 > 0$, $\varepsilon_2 > 0$ are constants, the coefficients $a_i(t)$, $i = 0, 1$, non random measurable functions and $w_1(t)$, $w_2(t)$ are independent Wiener processes. It is assumed that in model (1), (2) θ_t is the non observable process and ξ_t is the observable process [1].

Consider a linear gain function of such form

$$g(x, t) = f_1(t) + f_2(t)x, \tag{3}$$

* Corresponding author.
E-mail address: khechinashvili@gmail.com (Z. Khechinashvili).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.07.006

where $f_i(t)$, $i = 1, 2$, is non random measurable function, $x \in R$, and introduce the value

$$S_T^0 = \sup_{\tau \in \Re_T^\theta} Eg(\tau, \theta_\tau), \quad S_T^{\varepsilon_1,\varepsilon_2} = \sup_{\tau \in \Re_T^\xi} Eg(\tau, \theta_\tau), \tag{4}$$

where as usual we denote a class of all stopping times relative to a family of σ-algebras $F^X = (\mathcal{F}_t^X)$ with $\mathcal{F}_t^X = \sigma\{X_s, 0 \le s \le t\}$ as $\Re_T^X$ [1,2].

The payoff S_T^0 corresponds to an optimal stopping problem with complete data for the process θ_t, while the payoff $S_T^{\varepsilon_1,\varepsilon_2}$ corresponds to the process θ_t with incomplete data. The first problem (reduction problem) consists in reducing the optimal stopping problem with incomplete data of the process θ_t to the optimal stopping problem of some completely observable process. The second problem (convergence of payoffs problem) is a proof of the convergence $S_T^{\varepsilon_1,\varepsilon_2} \to S_T^0$ as $\varepsilon_1 \to 0, \varepsilon_2 \to 0$ [3–5].

Consider the example which show that from the smallness of coefficients ε_1 and ε_2 does not necessarily follows the closeness of the payoffs. We suppose that $\theta_t = \varepsilon_1 w_1(t)$, $g(x, t) = g(x) = a$, when $x = x_0$ and $g(t, x) = 0$, when $x \ne x_0$, $x_0 \ne 0$. Then it is possible to show that $S_T^{\varepsilon_1,\varepsilon_2} \to 0 \ne S_T^0 = a$, when $\varepsilon_1 \to 0, \varepsilon_2 \to 0$.

In this paper the problem of reduction and convergence of payoff are investigated for model (1), (2).

2. The reduction problem

Let us introduce the following notations

$$m_t = E(\theta_t/\mathcal{F}_t^\xi), \quad \gamma_t = E(\theta_t - m_t)^2. \tag{5}$$

Theorem 1. *The payoff $S_T^{\varepsilon_1,\varepsilon_2}$ can be represented in the following form*

$$S_T^{\varepsilon_1,\varepsilon_2} = \sup_{\tau \in \Re_T^\xi} Eg(\tau, m_\tau). \tag{6}$$

Proof. Note that for arbitrary $\tau \in \Re_T^\xi$ and $A \in \mathcal{F}$ we have $A \cap \{\tau \le t\} \in \mathcal{F}_T^\xi$ for all $t \le T$. Because we have

$$S_T^{\varepsilon_1,\varepsilon_2} = \sup_{\tau \in \Re_T^\xi} E\{f_1(\tau) + f_2(\tau)\theta_\tau\} = S_T^{\varepsilon_1,\varepsilon_2} = \sup_{\tau \in \Re_T^\xi} E\{E[f_1(\tau) + f_2(\tau)\theta_\tau]/\mathcal{F}_\tau^\xi\}$$
$$= S_T^{\varepsilon_1,\varepsilon_2} = \sup_{\tau \in \Re_T^\xi} E\{f_1(\tau) + f_2(\tau)E(\theta_\tau/\mathcal{F}_\tau^\xi)\}.$$

Next we can write

$$I_{\{\tau=t\}}E(\theta_\tau/\mathcal{F}_\tau^\xi) = E(I_{\{\tau=t\}}\theta_\tau/\mathcal{F}_\tau^\xi) = E(I_{\{\tau=t\}}\theta_t/\mathcal{F}_\tau^\xi) = I_{\{\tau=t\}}E(\theta_t/\mathcal{F}_\tau^\xi),$$

where I_A is the indicator of set A. According to Lemma 1.9 [1], on the set $\{\tau = t\}$, we have $E(\theta_t/\mathcal{F}_\tau^\xi) = E(\theta_t/\mathcal{F}_t^\xi)$, i.e.

$$I_{\{\tau=t\}}E(\theta_\tau/\mathcal{F}_\tau^\xi) = I_{\{\tau=t\}}E(\theta_t/\mathcal{F}_t^\xi). \quad \text{a.}s.$$

Thus we get the proof of (6).

Theorem 2. *The payoff $S_T^{\varepsilon_1,\varepsilon_2}$ can be represented in the following form*

$$S_T^{\varepsilon_1,\varepsilon_2} = \sup_{\tau \in \Re_T^\theta} Eg(\tau, \tilde{\theta}_\tau), \tag{7}$$

where the stochastic process $\tilde{\theta}_t$ is defined by the following stochastic differential equation

$$d\tilde{\theta}_t = [a_0(t) + a_1(t)\tilde{\theta}_t]dt + \frac{a_1(t)\gamma_t}{\sqrt{\varepsilon_1^2 + \varepsilon_2^2}}dw_1(t). \tag{8}$$

Proof. It follows from (1) and (2) that

$$d\xi_t = [a_0 + a_1\theta]dt + \varepsilon_1 dW_1 + \varepsilon_2 dW_2 =$$
$$d\xi_t = [a_0(t) + a_1(t)\theta_t]dt + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}d\tilde{w}(t),$$

where

$$\tilde{w}(t) = \frac{\varepsilon_1}{\sqrt{\varepsilon_1^2+\varepsilon_2^2}} W_1 + \frac{\varepsilon_2}{\sqrt{\varepsilon_1^2+\varepsilon_2^2}} W_2$$

is a new Wiener process.

According to the Theorem 10.3 [1] and Theorem 7.12 [1] we have

$$dm_t = [a_0(t) + a_1(t)m_t]dt + (\varepsilon_1^2 + a_1(t)\gamma_t)(d\xi_t - [a_0(t) + a_1(t)m_t]dt),$$

$$dm_t = [a_0(t) + a_1(t)m_t]dt + \frac{\varepsilon_1^2 + a_1(t)\gamma_t}{\sqrt{\varepsilon_1^2+\varepsilon_2^2}} d\overline{w}(t), \tag{9}$$

where

$$\overline{w}(t) = \frac{d\xi_t - [a_0(t) + a_1(t)m_t]dt}{\sqrt{\varepsilon_1^2+\varepsilon_2^2}}$$

is so called innovation Wiener process, which has such property that the σ-algebras $\mathcal{F}_t^{\xi}$ and $\mathcal{F}_t^{\overline{w}}$ coincide. From (8) and (10) we have

$$d\tilde{\theta}_t = \Phi_t[\int_0^t \Phi_s^{-1} a_0(s)ds + \int_0^t \Phi_s^{-1} \frac{\varepsilon_1^2 + a_1(s)\gamma_s}{\sqrt{\varepsilon_1^2+\varepsilon_2^2}} dw_1(s)], \tag{10}$$

$$dm_t = \Phi_t[\int_0^t \Phi_s^{-1} a_0(s)ds + \int_0^t \Phi_s^{-1} \frac{\varepsilon_1^2 + a_1(s)\gamma_s}{\sqrt{\varepsilon_1^2+\varepsilon_2^2}} d\overline{w}(s)], \tag{11}$$

where the deterministic function Φ_t is defined by the following relation

$$\Phi_t = \exp\{\int_0^t a_1(s)ds\}. \tag{12}$$

From (11), (12) we can write

$$\sup_{\tau\in\Re^{\tilde{\theta}}} Eg(\tau, \tilde{\theta}_\tau) = \sup_{\tau\in\Re^{\xi}} Eg(\tau, m_\tau), \tag{13}$$

where $\Re^{\tilde{\theta}} = \Re^{\theta}$. Thus

$$\sup_{\tau\in\Re^{\tilde{\theta}}} Eg(\tau, \tilde{\theta}_\tau) = \sup_{\tau\in\Re^{\theta}} Eg(\tau, \tilde{\theta}_\tau).$$

According to Theorem 1 $\sup_{\tau\in\Re^{\xi}} Eg(\tau, m_\tau) = S_T^{\varepsilon_1,\varepsilon_2}$ and we get (7).

3. Convergence of payoffs

In proving the payoffs convergence rate, an estimation of the conditional variance γ_t by means of small parameters $\varepsilon_1, \varepsilon_2$ plays an essential role. We recall that for γ_t we have the ordinary differential equation

$$\gamma_t' = 2a_1(t)\gamma_t - \frac{a_1^2(t)\gamma_t^2}{\varepsilon_1^2+\varepsilon_2^2} + \varepsilon_1^2, \quad \gamma_0 = 0. \tag{14}$$

Let $\rho(t)$ denote a continuous increasing majorant of the function

$$\phi(t) = \frac{\varepsilon_1}{a_1(t)} \Phi_t^{-2},$$

where the function Φ_t is defined by (13).

Theorem 3. *Let $\rho(t) \geq \phi(t)$. Then the following estimate holds for all $0 \leq t \leq T$:*

$$\gamma_t \leq \sqrt{\varepsilon_1^2+\varepsilon_2^2}\, \Phi_t^2 \rho(t). \tag{15}$$

Proof. We introduce a function u_t by using the following transformation

$$\gamma_t = \sqrt{\varepsilon_1^2 + \varepsilon_2^2}\,\Phi_t^2 u_t, \quad u_0 = 0. \tag{16}$$

It is not difficult to see that the function u_t satisfies the ordinary differential equation

$$u_t' = \frac{a_1^2(t)\Phi_t^2}{\sqrt{\varepsilon_1^2 + \varepsilon_2^2}}\left[\frac{\varepsilon_1^2\Phi_t^{-4}}{a_1^2(t)} - u_t^2\right], \quad u_0 = 0. \tag{17}$$

Let us show that $u_t \le \rho(t), 0 \le t \le T$. Assume the opposite. Then there exist points t_0 and t_1 with $t_0 < t_1$ such that $u_{t_0} = \rho(t_0)$ and $u_t > \rho(t)$ for $t_0 < t \le t_1$. For $t \in [t_0, t_1]$ we have

$$u_t' = \frac{a_1^2(t)\Phi_t^2}{\sqrt{\varepsilon_1^2 + \varepsilon_2^2}}\left[\rho^2(t) - u_t^2\right] < 0.$$

Therefore $u_t < u_{t_0} = \rho(t_0) \le \rho(t)$ and we have obtained $u_t < \rho(t)$, which contradicts our assumption. Thus $u_t \le \rho(t), 0 \le t \le T$, and we obtain the estimate (16).

We introduce the notations [5]:

$$h(t) = \varepsilon_1^2 \int_0^t \Phi_s^{-2} ds, \quad \tilde{h}(t) = \int_0^t \Phi_s^{-2} \frac{a_1^2(s)\gamma_s^2}{\sqrt{\varepsilon_1^2 + \varepsilon_2^2}} ds, \tag{18}$$

$$l = \exp\{2\int_0^T a_1(s)ds\}\rho(T), \tag{19}$$

$$Lg(t,x) = f_1'(t) + f_2'(t)x + f_2(t)[a_0(t) + a_1(t)x]. \tag{20}$$

Theorem 4. *Let the following condition hold:*

$$E(\sup_{t\le T} g(t,\theta_t)) < \infty. \tag{21}$$

Then the estimate is true

$$0 \le S_T^0 - S_T^{\varepsilon_1,\varepsilon_2} \le (\varepsilon_1 + \varepsilon_2) l \sup_{t\le T} E(Lg(t,\theta_t)). \tag{22}$$

Proof. First we show that $S_T^0 \ge S_T^{\varepsilon_1,\varepsilon_2}$. From Theorem 3 [5] and the identity of the σ-algebras $\mathcal{F}_t^{\overline{w}}$ and $\mathcal{F}_t^{\xi}$ it follows that

$$S_T^{\varepsilon_1,\varepsilon_2} = \sup_{\tau\in\Re^{\overline{w}}} Eg(\tau, m_\tau + \eta\sqrt{\gamma}_\tau), \tag{23}$$

where η is standard normal random variable. The process $m_t, 0 \le t \le T$, is Markovian with respect to the family $F^{\overline{w}} = (\mathcal{F}_t^{\overline{w}})$ and in that case as it is well known, the class of stopping times $\mathcal{F}_T^m$ is sufficient [2], i.e. we have

$$S_T^{\varepsilon_1,\varepsilon_2} = \sup_{\tau\in\Re_T^m} Eg(\tau, m_\tau + \eta\sqrt{\gamma}_\tau). \tag{24}$$

Let us now introduce an auxiliary payoff for stopping times $\tau \in \Re_T^\theta$:

$$\tilde{S}_T^{\varepsilon_1,\varepsilon_2} = \sup_{\tau\le T_{\varepsilon_1,\varepsilon_2}} Eg(\tau,\theta_\tau), \tag{25}$$

where $T_{\varepsilon_1,\varepsilon_2}$ is defined by the relation $\tilde{h}(T) = h(T_{\varepsilon_1,\varepsilon_2})$. It is easy to see that for $\tau \in \mathcal{F}_t^\theta$:

$$0 \le S_T^0 - \tilde{S}_T^{\varepsilon_1,\varepsilon_2} \le \sup_{\tau\le T} E[g(\tau,\theta_\tau) - g(\tau \wedge T_{\varepsilon_1,\varepsilon_2}, \theta_{\tau\wedge T\varepsilon_1,\varepsilon_2})],$$

where $s \wedge t := \min(s,t)$.

Further, by Ito formula we can write

$$E[g(\tau,\theta_\tau)-g(\tau\wedge T_{\varepsilon_1,\varepsilon_2},\theta_{\tau\wedge T\varepsilon_1,\varepsilon_2})]=E\int_{\tau\wedge T_{\varepsilon_1,\varepsilon_2}}^{\tau}Lg(t,\theta_t)dt\le\int_{T_{\varepsilon_1,\varepsilon_2}}^{T}E[Lg(t,\theta_t)]dt$$
$$\le(T-T_{\varepsilon_1,\varepsilon_2})\sup_{t\le T}E[g(t,\theta_t)]\le(\varepsilon_1+\varepsilon_2)l\sup_{t\le T}E[g(t,\theta_t)].$$

Therefore we have

$$S_T^0-\tilde{S}_T^{\varepsilon_1,\varepsilon_2}\le(\varepsilon_1+\varepsilon_2)l\sup_{t\le T}E[g(t,\theta_t)]. \tag{26}$$

From (26), by Theorem 4 [5], we obtain the estimate (23).

References

[1] R.Sh. Liptser, A.N. Shiryaev, Statistics of Random Processes. Vol. 1-2, Springer-Verlag, Berlin, New York, 1977 1978.
[2] A.N. Shiryaev, Optimal Stopping Rules, Springer-Verlag, New York, 1978.
[3] B. Dochviri, On optimal stopping with incomplete data, in: Probability Theory and Mathematical Statistics (Kyoto, 1986), in: Lecture Notes in Math, vol. 1299, Springer-Verlag, Berlin, 1988, pp. 64–68.
[4] P. Babilua, I. Bokuchava, B. Dochviri, M. Shashiashvili, Convergence of costs in an optimal stopping problem for a partially observable model, Appl. Math. Inform. Mech. 11 (2006) 6–11.
[5] P. Babilua, I. Bokuchava, B. Dochviri, The optimal stopping problem for the Kalman-Bucy scheme, Theory Probab. Appl. 55 (1) (2010) 133–142.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 337–353
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Forks, noodles and the Burau representation for $n = 4$

A. Beridze[a,*], P. Traczyk[b]

[a] *Department of Mathematics, Batumi ShotaRustaveli State University, 35, Ninoshvili St., Batumi 6010, Georgia*
[b] *Institute of Mathematics, University of Warsaw, Banacha 2, 02-097 Warszawa, Poland*

Received 16 January 2018; received in revised form 6 May 2018; accepted 15 May 2018
Available online 30 May 2018

Abstract

The reduced Burau representation is a natural action of the braid group B_n on the first homology group $H_1(\tilde{D}_n; \mathbb{Z})$ of a suitable infinite cyclic covering space $\tilde{D}_n$ of the n-punctured disc D_n. It is known that the Burau representation is faithful for $n \leq 3$ and that it is not faithful for $n \geq 5$. We use forks and noodles homological techniques and Bokut–Vesnin generators to analyze the problem for $n = 4$. We present a Conjecture implying faithfulness and a Lemma explaining the implication. We give some arguments suggesting why we expect the Conjecture to be true. Also, we give some geometrically calculated examples and information about data gathered using a C++ program.

1. Introduction

Let us recall the definition of the reduced Burau representation in terms of the first homology group $H_1(\tilde{D}_4; \mathbb{Z})$ of a suitable infinite cyclic covering space $\tilde{D}_4$ of the 4-punctured disc D_4. Let D_4 be the unit closed disc on the plane with center $(0, 0)$ and four punctures at: $p_1 = \left(-\frac{1}{2}, \frac{1}{2}\right)$, $p_2 = \left(\frac{1}{2}, \frac{1}{2}\right)$, $p_3 = \left(\frac{1}{2}, -\frac{1}{2}\right)$, $p_4 = \left(-\frac{1}{2}, -\frac{1}{2}\right)$ (see Fig. 1).

The braid group B_4 is the group of all equivalence classes of orientation preserving homeomorphisms $\varphi : D_4 \to D_4$ which fix the boundary ∂D_4 pointwise, where equivalence relation is isotopy relative to ∂D_4. Let $\pi_1(D_4)$ be the fundamental group of the 4-punctured disc D_4 with respect to the basepoint $p_0 = (-1, 0)$. Consider the map $\varepsilon : \pi_1(D_4) \to \langle t \rangle$ which sends a loop $\gamma \in \pi_1(D_4)$ to $t^{[\gamma]}$, where $[\gamma]$ is the winding number of γ around punctured points p_1, p_2, p_3, p_4 (meaning: the sum of the four winding numbers for individual points). Let $\pi : \tilde{D}_4 \to D_4$ be the infinite cyclic covering space corresponding to the kernel $\ker(\varepsilon)$ of the map $\varepsilon : \pi_1(D_4) \to \langle t \rangle$. Let $\tilde{p}_0$ be any fixed basepoint which is a lift of the basepoint p_0. In this case $H_1(\tilde{D}_4; \mathbb{Z})$ is free $\mathbb{Z}\left[t, t^{-1}\right]$-module of rank 3 (see [1]). Let $\varphi : D_4 \to D_4$ be a homeomorphism representing an element $\sigma \in B_4$. It can be lifted to a map $\tilde{\varphi} : \tilde{D}_4 \to \tilde{D}_4$ which

* Corresponding author.
E-mail addresses: a.beridze@bsu.edu.ge (A. Beridze), traczyk@mimuw.edu.pl (P. Traczyk).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.05.001

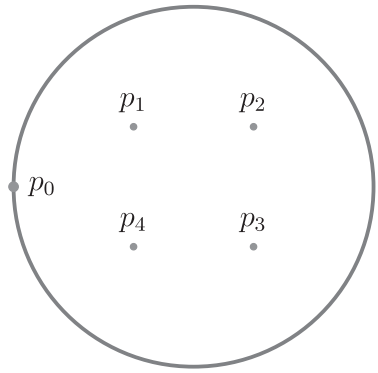


Fig. 1. The 4-punctured disc with basepoint p_0 and puncture points: p_1, p_2, p_3, p_4.

fixes the fiber over p_0. Therefore it induces a $\mathbb{Z}\left[t, t^{-1}\right]$-module automorphism $\widetilde{\varphi}_* : H_1\left(\tilde{D}_4; \mathbb{Z}\right) \to H_1\left(\tilde{D}_4; \mathbb{Z}\right)$. Consequently, the reduced Burau representation

$$\rho : B_4 \to Aut\left(H_1\left(\tilde{D}_4; \mathbb{Z}\right)\right) \tag{1.1}$$

is given [1] by

$$\rho(\sigma) = \widetilde{\varphi}_*, \quad \forall \sigma \in B_4. \tag{1.2}$$

It is known that the Burau representation is faithful for $n \leq 3$ [2,3] and it is not faithful for $n \geq 5$ [3–5]. Therefore, the problem is open for $n = 4$. In this paper, we use the Bokut–Vesnin generators a, a^{-1}, b, b^{-1} of a certain free subgroup of B_4 (see [6]) and a technique developed in [1], to prove the crucial lemma, which gives the opportunity to decompose entries $\rho_{11}(a^n\sigma)$ and $\rho_{13}(a^n\sigma)$ of the Burau matrix $\rho(a^n\sigma)$ as a sum of three uniquely determined polynomials and the formula to calculate $\rho_{13}(a^{n+m}\sigma)$ and $\rho(a^{n+m}\sigma)$ polynomials using the given decomposition. Besides, we formulate Conjecture 4.2, which implies that if a non-trivial braid $\sigma \in \ker\rho$ has a certain additional property, then there exists a sufficiently large l_0 with respect to the length of σ (to be explained in Section 3, Corollary 3.2) and a sufficiently large m_0 such that for each $m > m_0$ and $l > l_0$ the difference of lowest degrees of polynomials $\rho_{13}(a^{n+m}\sigma)$ and $\rho(a^{n+m}\sigma)$ is -1. We will present arguments and experimental data showing why we expect the conjecture to be true. Also, we will consider several examples calculated geometrically. We will show that the conjecture implies faithfulness of the Burau representation for $n = 4$.

2. The Burau representation, forks and noodles

The Burau representation for $n = 4$ was defined by (1.1) and (1.2). On the other hand $H_1(\tilde{D}_4; \mathbb{Z})$ is a free $\mathbb{Z}\left[t, t^{-1}\right]$-module of rank 3 and if we take a basis of it, then $Aut\left(H_1\left(\tilde{D}_4; \mathbb{Z}\right)\right)$ can be identified with $GL\left(3, \mathbb{Z}\left[t, t^{-1}\right]\right)$. For this reason we will review the definition of the forks.

Definition 2.1. A fork is an embedded oriented tree F in the disc D with four vertices p_0, p_i, p_j and z, where $i \neq j, i, j \in \{1, 2, 3, 4\}$ such that (see [1]):

1. F meets the puncture points only at p_i and p_j;
2. F meets the boundary ∂D_4 only at p_0;
3. All three edges of F have z as a common vertex.

The edge of F which contains p_0 is called the handle. The union of the other two edges is denoted by $T(F)$ and it is called tine of F. Orient $T(F)$ so that the handle of F lies to the right of $T(F)$ (see Fig. 2) [1].

For a given fork F, let $h : I \to D_4$ be the handle of F, viewed as a path in D_4 and take a lift $\tilde{h} : I \to \tilde{D}_4$ of h so that $\tilde{h}(0) = \tilde{p}_0$. Let $\tilde{T}(F)$ be the connected component of $\pi^{-1}(T(F))$ which contains the point $\tilde{h}(1)$. In this case any element of $H_1\left(\tilde{D}_4; \mathbb{Z}\right)$ can be viewed as a homology class of $\tilde{T}(F)$ and it is denoted by F [1].

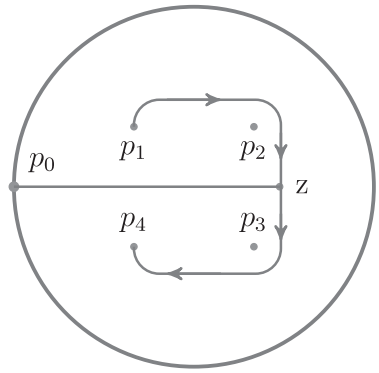


Fig. 2. The line from p_0 to z is the handle and the curve from p_1 to p_4 is the tine $T(F)$ of the fork F.

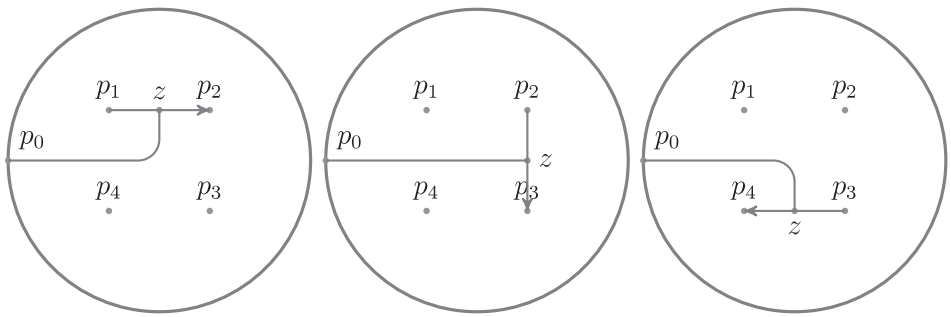


Fig. 3. Standard forks: $F_1,\ \ F_2,\ \ F_3$.

Standard fork $F_i,\ \ i = 1, 2, 3$ is the fork whose tine edge is the straight arc connecting the ith and the $(i + 1)$-st punctured points and whose handle has the form as in Fig. 3. It is known that if $F_1,\ F_2$ and F_3 are the corresponding homology classes, then they form a basis of $H_1\left(\tilde{D}_4; \mathbb{Z}\right)$ (see [1]).

Using the basis derived from F_1, F_2, F_3, any automorphism $\widetilde{\varphi}_* : H_1\left(\tilde{D}_4; \mathbb{Z}\right) \to H_1\left(\tilde{D}_4; \mathbb{Z}\right)$ can be viewed as a 3×3 matrix with elements in the free $\mathbb{Z}\left[t, t^{-1}\right]$-module [1]. If $\varphi : D_4 \to D_4$ is representing an element $\sigma \in B_4$, then we need to write the matrix $\rho(\sigma) = \widetilde{\varphi}_*$ in terms of homology (algebraic) intersection pairing

$$\langle -, - \rangle : H_1\left(\tilde{D}_4; \mathbb{Z}\right) \times H_1\left(\tilde{D}_4, \partial\tilde{D}_4; \mathbb{Z}\right) \to \mathbb{Z}\left[t, t^{-1}\right].$$

For this aim we need to define the noodles which represent relative homology classes in $H_1\left(\tilde{D}_4, \partial\tilde{D}_4; \mathbb{Z}\right)$.

Definition 2.2. A noodle is an embedded oriented arc in D_4, which begins at the base point p_0 and ends at some point of the boundary ∂D_4 [1].

For each $a \in H_1\left(\tilde{D}_4; \mathbb{Z}\right)$ and $b \in H_1\left(\tilde{D}_4, \partial\tilde{D}_4; \mathbb{Z}\right)$ we should take the corresponding fork F and noodle N and define the polynomial $\langle F, N\rangle \in \mathbb{Z}\left[t, t^{-1}\right]$. It does not depend on the choice of representatives of homology classes and so

$$\langle -, - \rangle : H_1\left(\tilde{D}_4; \mathbb{Z}\right) \times H_1\left(\tilde{D}_4, \partial\tilde{D}_4; \mathbb{Z}\right) \to \mathbb{Z}\left[t, t^{-1}\right]$$

is well-defined [1]. The map defined by the above formula is called the noodle–fork pairing. Note that geometrically it can be computed in the following way: Let F be a fork and N be a noodle, such that $T(F)$ intersects N transversely. Let $z_1, z_2,\ \ldots, z_n$ be the intersection points. For each point z_i let ε_i be the sign of the intersection between $T(F)$

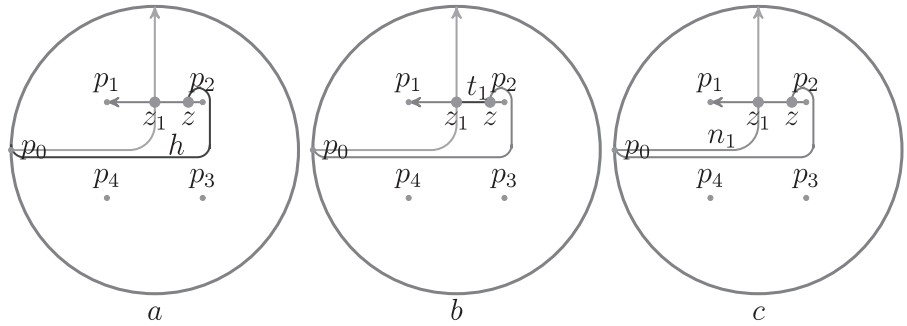


Fig. 4. h — a path from p_0 to z; t_1 — a path from z to z_1; n_1 — a path from z_1 to p_0.

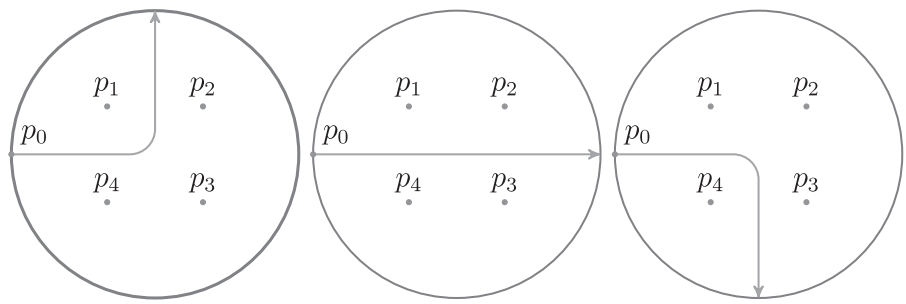


Fig. 5. Standard noodles: N_1, N_2, N_3.

and N at z_i (the intersection is positive if going from tine to noodle according to the chosen directions means turning left) and $e_i = [\gamma_i]$ be the winding number of the loop γ_i around the puncture points p_1, p_2, p_3, p_4, where γ_i is the composition of three paths h, t_i and n_i:

1. h is a path from p_0 to z along the handle of F (see Fig. 4(a));
2. t_i is a path from z to z_i along the tine $T(F)$ (see Fig. 4(b));
3. n_i is a path from z_i to p_0 along the noodle N (see Fig. 4(c)).

In such case the noodle–fork pairing of F and N is given by (see [1]):

$$\langle F, N\rangle = \sum_{1\le i\le n} \varepsilon_i t^{e_i} \in \mathbb{Z}\left[t, t^{-1}\right]. \tag{2.1}$$

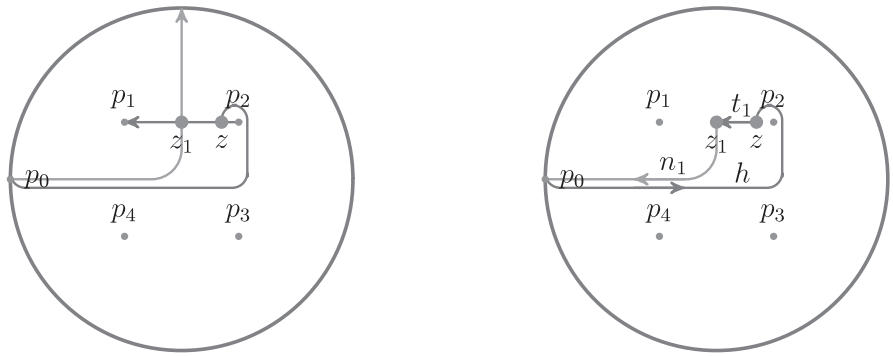
Let N_1, N_2 and N_3 be the noodles given in Fig. 5. These are called standard noodles. For each braid $\sigma \in B_4$, the corresponding Burau matrix $\rho(\sigma)$ can be computed using noodle–fork pairing of standard noodles and standard forks. In particular the following is true.

Lemma 2.3 (*See [7]*)**.** *Let $\sigma \in B_n$. Then for $1 \le i, j \le n-1$, the entry $\rho_{ij}(\sigma)$ of its Burau matrix $\rho(\sigma)$ is given by*

$$\rho_{ij}(\sigma) = \left\langle F_i\sigma, N_j\right\rangle.$$

Note that under the convention adopted here we have

$$\rho(\sigma_1) = \begin{pmatrix} -t^{-1} & 0 & 0 \\ t^{-1} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \rho(\sigma_2) = \begin{pmatrix} 1 & 1 & 0 \\ 0 & -t^{-1} & 0 \\ 0 & t^{-1} & 1 \end{pmatrix}$$



(a) $e_1 = -1$. (b) $\gamma_1 = n_1 * t_1 * h$; $\varepsilon_1 = 1$.

Fig. 6. $\langle F_1\sigma_1, N_1\rangle = -t^{-1}$ is monomial, because the intersection of $F_1\sigma_1$ and the noodle N_1 is just one point z_1.

and

$$\rho(\sigma_3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & -t^{-1} \end{pmatrix}.$$

For example, to calculate $\rho_{1,1}(\sigma_1)$ entry of the matrix $\rho(\sigma_1)$ see the corresponding Fig. 6. Note that intersection of the tine $T(F_1\sigma_1)$ of the fork $F_1\sigma_1$ and the noodle N_1 at point z_1 is negative which means that $\varepsilon_i = -1$ (see Fig. 6(a)). On the other hand the winding number e_1 of the loop γ_1 (see Fig. 6(b)) around puncture points equals -1 because the considered loop misses p_1, p_3 and p_4 and it goes around p_2 once in anti-clockwise direction. Therefore

$$\rho_{11}(\sigma_1) = \langle F_1\sigma_1, N_1\rangle = -t^{-1}.$$

3. The Bokut–Vesnin generators and kernel elements of the Burau representation

The braid groups B_4 and B_3 are defined by the following standard presentations [2]:

$$B_4 = \langle \sigma_1,\ \sigma_2, \sigma_3 \,|\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2,\quad \sigma_3\sigma_2\sigma_3 = \sigma_2\sigma_3\sigma_2,\quad \sigma_3\sigma_1 = \sigma_1\sigma_3 \rangle,$$

$$B_3 = \langle \sigma_1,\ \sigma_2 \,|\sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2 \rangle.$$

Let $\varphi : B_4 \to B_3$ be the homomorphism defined by

$$\varphi(\sigma_1) = \sigma_1,\ \ \varphi(\sigma_2) = \sigma_2,\ \ \varphi(\sigma_3) = \sigma_1.$$

The kernel of φ is known to be a free group $F(a, b)$ of two generators [6];

$$a = \sigma_1\sigma_2\sigma_1^{-1}\sigma_3\sigma_2^{-1}\sigma_1^{-1},\ \ b = \sigma_3\sigma_1^{-1}.$$

This was proved by L. Bokut and A. Vesnin [6]. We will refer to a and b as the Bokut–Vesnin generators. The generators a and b are in fact much more similar than they look at the first glance. This becomes obvious when we interpret B_4 as the mapping class group of the 4-punctured disc. In this well-known approach a braid is an isotopy class of homeomorphisms of the punctured disc fixing the boundary. Fig. 7 shows a and b as homeomorphisms of the punctured disc. The punctures are arranged to make the similarity more visible. Another advantage of this approach is that it gives natural interpretation to various actions of B_4 to be considered later in this paper.

The following Proposition is crucial to our considerations.

Proposition 3.1. $\ker\rho_4 \subset \ker\varphi$.

Proof. Let us make a slight detour into the realm of the Temperley–Lieb algebras TL_3 and TL_4. The Temperley–Lieb algebra TL_n is defined as an algebra over $\mathbb{Z}[t, t^{-1}]$. It has $n-1$ generators $\{U_i^i\}_{i=1}^{n-1}$, and the following relations:

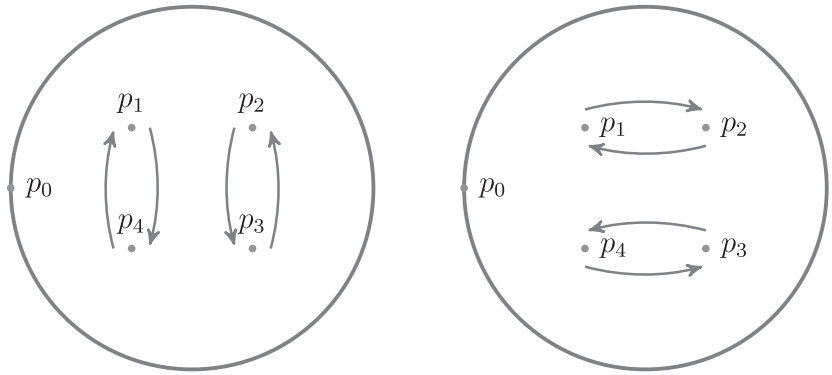


Fig. 7.

(TL1) $U_i^i U_i^i = (-t^{-2} - t^2)U_i^i$,
(TL2) $U_i^i U_j^j U_i^i = U_i^i$, for $|i - j| = 1$,
(TL3) $U_i^i U_j^j = U_j^j U_i^i$, for $|i - j| > 1$.
Let us consider the homomorphism $\psi : TL_4 \to TL_3$ defined by

$$U_1^1 \to U_1^1,\ U_2^2 \to U_2^2,\ U_3^3 \to U_1^1.$$

Also, we need to use the Jones' representation $\theta : B_n \to TL_n$ defined by sending σ_i to $A + A^{-1}U_i^i$. It is known (see [1], Proposition 1.5) that for $n = 3, 4$ we have $\ker\theta_n = \ker\rho_n$. Moreover, the following diagram is obviously commutative:

$$\begin{array}{ccc} B_4 & \xrightarrow{\theta_4} & TL_4 \\ \downarrow{\scriptstyle\varphi} & & \downarrow{\scriptstyle\psi} \\ B_3 & \xrightarrow[\theta_3]{} & TL_3 \end{array}$$

On the other hand the representation θ_3 is faithful and therefore $\ker\rho_4 = \ker\theta_4 \subset \ker\varphi$. □

Corollary 3.2. *All kernel elements of the Burau representation may be written as words in the Bokut–Vesnin generators a, b, a^{-1}, b^{-1}. Moreover, all possible nontrivial elements in the kernel may be written as reduced words of positive length.*

We will use this fact in the next section.

We present for future use the images of a, b, a^{-1} and b^{-1} under the Burau representation:

$$\rho(a) = \begin{pmatrix} -t^{-1}+1 & -t^{-1}+t & -t^{-1} \\ 0 & -t & 0 \\ -1 & 0 & 0 \end{pmatrix},$$

$$\rho(b) = \begin{pmatrix} -t & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & -t^{-1} \end{pmatrix}.$$

$$\rho\left(a^{-1}\right) = \begin{pmatrix} 0 & 0 & -1 \\ 0 & -t^{-1} & 0 \\ -t & t^{-1}-t & 1-t \end{pmatrix},$$

$$\rho\left(b^{-1}\right) = \begin{pmatrix} -t^{-1} & 0 & 0 \\ t^{-1} & 1 & t \\ 0 & 0 & -t \end{pmatrix}.$$

4. Faithfulness problem of the Burau representation

Let us outline the strategy for analyzing $\ker \rho_4$ in general terms. Consider a braid σ that is a candidate for a non-trivial kernel element of the Burau representation. Of course we can exclude from our considerations all those non-trivial braids for which we *know* for whatever reason that they do not belong to the kernel. Also, we can adjust the remaining candidates in some ways — like replacing σ with a suitably chosen conjugate of σ. For such a suitably chosen braid σ we need to give some argument which shows that $\rho_{11}(\sigma)$ and $\rho_{31}(\sigma)$ should be non-zero and that $\deg_{min}(\rho_{11}) - \deg_{min}(\rho_{31}) = -1$, where $\deg_{min}$ denotes the exponent of the lowest degree term in the considered Laurent polynomial.

To simplify notation we will denote by $S_i(t^{\pm 1})$ the ith partial sum of the geometric series with initial term 1 and quotient $-t$ or $-t^{-1}$ (e.g. $S_2(t^{-1}) = 1 - t^{-1} + t^{-2}$).

Lemma 4.1. *For each braid $\sigma \in B_4$ there exists $n \in \mathbb{N}$, such that*

(1) the $\rho_{11}(a^n\sigma)$ and $\rho_{31}(a^n\sigma)$ entries of the Burau matrix $\rho\left(a^n\sigma\right)$ can be decomposed as a sum of three uniquely determined polynomials

$$\rho_{11}\left(a^n\sigma\right) = P\left(t, t^{-1}\right)\left(1 - t^{-1}\right) + Q\left(t, t^{-1}\right) + R\left(t, t^{-1}\right)(1 - t),$$

$$\rho_{31}\left(a^n\sigma\right) = -P\left(t, t^{-1}\right) - Q\left(t, t^{-1}\right) - R\left(t, t^{-1}\right),$$

such that

(2) for each $m \in \mathbb{N}$ we have

$$\rho_{11}\left(a^{m+n}\sigma\right) = P\left(t, t^{-1}\right)\left(S_{m+1}(t^{-1})\right) + Q\left(t, t^{-1}\right) + R\left(t, t^{-1}\right)\left(S_{m+1}(t)\right),$$

$$\rho_{31}\left(a^{m+n}\sigma\right) = -P\left(t, t^{-1}\right)\left(S_m(t^{-1})\right) - Q\left(t, t^{-1}\right) - R\left(t, t^{-1}\right)\left(S_m(t)\right).$$

(3) Moreover, if σ is a pure braid, then the polynomial P is non-zero.

Proof. First of all let us observe that uniqueness of P, Q and R follows from properties (1) and (2) and general algebra. This means that we only need to prove existence and property (3). While it is possible to give specific algebraic formulas for P, Q and R we prefer to prove existence using forks and noodles. We will always assume that the fork/noodle configuration considered is irreducible.

Let $\sigma \in B_4$ be any braid. By Lemma 2.3 $\rho_{11}\left(a^n\sigma\right) = \langle F_1 a^n\sigma, N_1\rangle$ and $\rho_{31}\left(a^n\sigma\right) = \langle F_3 a^n\sigma, N_1\rangle$. On the other hand $\langle -, -\rangle$ is a bilinear form, so $\langle F_1 a^n\sigma, N_1\rangle = \left\langle F_1 a^n, N_1\sigma^{-1}\right\rangle$ and $\langle F_3 a^n\sigma, N_1\rangle = \left\langle F_3 a^n, N_1\sigma^{-1}\right\rangle$. It follows that

$$\rho_{11}\left(a^n\sigma\right) = \left\langle F_1 a^n, N_1\sigma^{-1}\right\rangle,$$

$$\rho_{31}\left(a^n\sigma\right) = \left\langle F_3 a^n, N_1\sigma^{-1}\right\rangle.$$

Let us consider $N_1\sigma^{-1}$, the image of the standard noodle N_1 under the action of σ^{-1}. $N_1\sigma^{-1}$ is a path in D_4 that begins at the base point p_0 and ends at the point $(0, 1) \in \partial D_4$. By the definition of the standard noodle N_1 it is clear that $N_1\sigma^{-1}$ divides D_4 into two components, such that there is one puncture point in one component and three puncture points in the other. Let us assume that the single point is p_1. For example see Fig. 8.

We intend to define P and R by grouping some terms in the sum originally used to define the representation in terms of fork/noodle pairing. The pairing is defined as a certain sum (2.1) of terms corresponding to crossings between forks and noodles. We will choose some of the crossings to define P and some other to define R. In order to do this we will need some preparations.

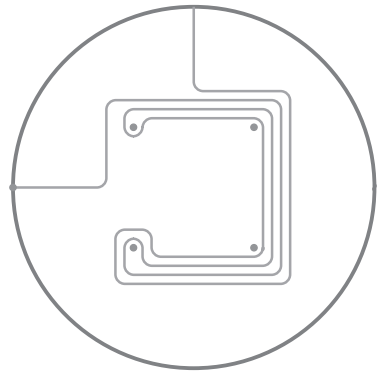


Fig. 8. p_1 is in one component and $p_2,\ p_2,\ p_3$ are in the other.

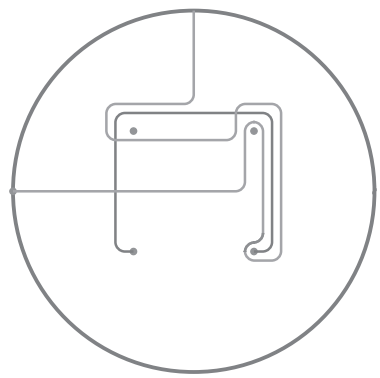


Fig. 9. The tine $T(F_1)a$ (shown as the black curve) does not intersect the blue segment at which the noodle N (the union of blue and red segments) intersects T_2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Let T be the boundary of the square whose vertices are the puncture points. We denote the sides with $T_1, \ldots, T_4$, where T_i connects P_i with the next crossing (clockwise). We would like to work with a fork/noodle arrangement that has certain special properties. We need the pair (of a fork and a noodle) to be irreducible. We need the fork to be drawn in the standard way. We need the noodle to intersect T transversally with minimum possible number of intersection points. And finally we need the tine of the fork to intersect all segments at which the noodle intersects T_4 and T_2. While general position arguments show that we can take care of the first three conditions, there is no possibility of the fourth being satisfied without some further adjustments. Fig. 9 shows an example.

However it is automatically corrected if we increase the exponent n. The effect is just that we add a number of turns around two pairs of punctures. They do not affect the three properties already dealt with and with sufficient increase of n we obtain the fourth property. So we are interested in strings between puncture points which has transversal intersection with T (see Fig. 10).

Note that it is possible to be no such string between p_2 and p_3 or p_3 and p_4, but by our assumption (p_1 is in the first component) there is an odd number of strings between p_1 and p_2 and an odd number between p_1 and p_4 (this guaranties that P and Q are not zero).

The pictures of $T\,(F_1)\,a^n$ and $T\,(F_3)\,a^n$ are as given in Fig. 11. Therefore, they differ from each other by just one string and by the direction. The number of strings around $p_1,\ p_4$ and $p_2,\ p_3$ is n for $T\,(F_1)\,a^n$ and $n-1$ for $T\,(F_3)\,a^n$.

If we take curves $T\,(F_1)\,a$ and $N_1\sigma^{-1}$ in the same D_4 and assume that their intersection is transversal, then it is possible that $T\,(F_1)\,a$ does not intersect all strings between p_1 and p_4 or p_2 and p_3. For example see Fig. 9.

In this case we must take more numbers of a's and finally we will obtain the curves $T\,(F_1)\,a^n$ and $N_1\sigma^{-1}$ such that we can find neighborhoods U_1 and U_2 of T_4 and T_2 respectively, with the following picture, illustrated in Fig. 12(a).

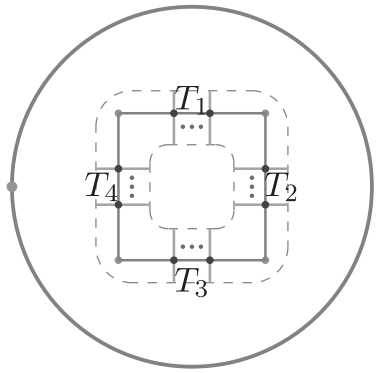


Fig. 10.

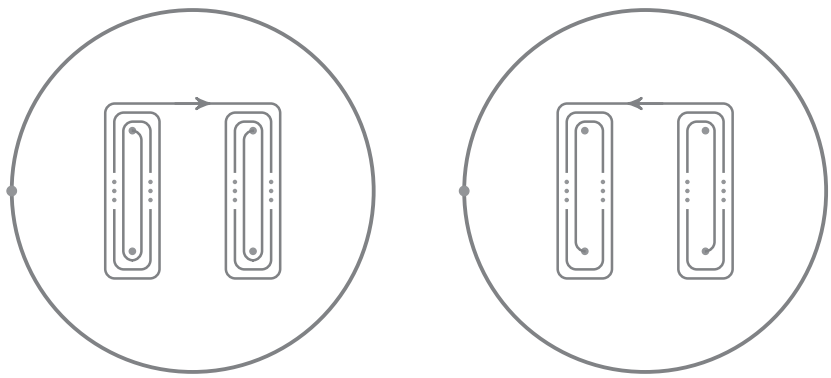


Fig. 11.

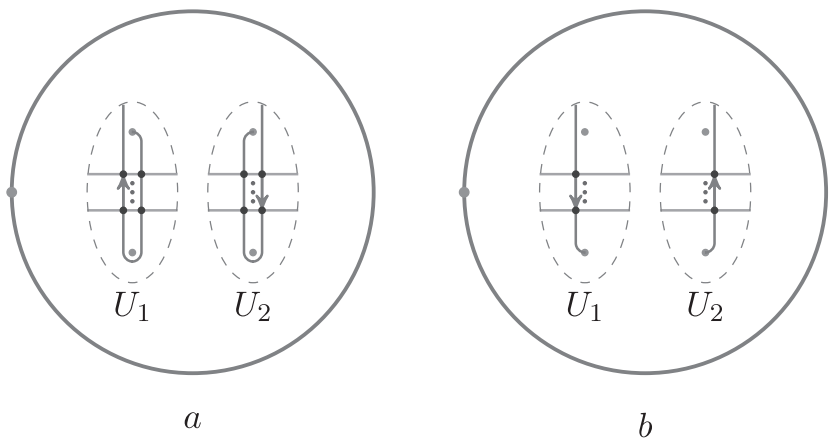


Fig. 12.

In this case for Fig. 12(a) the polynomial corresponding to intersections inside U_1 and U_2 can be written as $P(t,t^{-1})(1-t^{-1})$ and $R(t,t^{-1})(1-t)$ respectively. Let $Q\left(t,t^{-1}\right)=\rho_{11}\left(a^n\sigma\right)-P\left(t,t^{-1}\right)\left(1-t^{-1}\right)-R(t,t^{-1})(1-t)$, then we have

$$\rho_{11}\left(a^n\sigma\right)=P\left(t,t^{-1}\right)\left(1-t^{-1}\right)+Q\left(t,t^{-1}\right)+R\left(t,t^{-1}\right)(1-t).$$

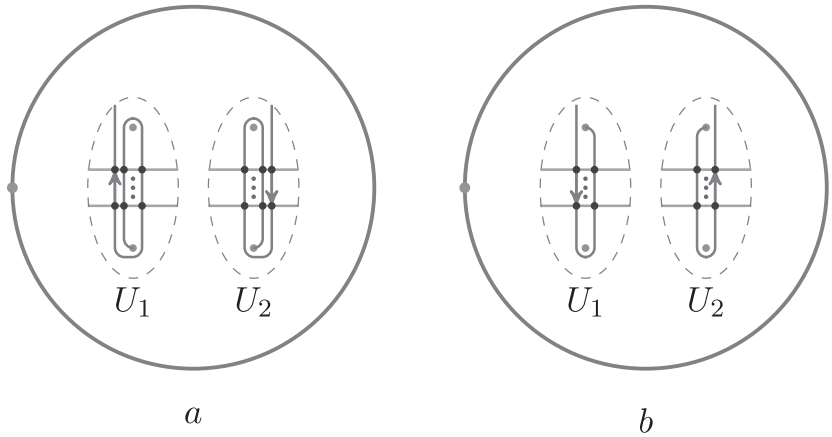


Fig. 13.

On the other hand if we look at Fig. 12(b) and keep in mind that directions of $T(F_1)\,a^n$ and $T(F_3)\,a^n$ are different we can say that

$$\rho_{31}\left(a^n\sigma\right) = -P\left(t,t^{-1}\right) - Q\left(t,t^{-1}\right) - R\left(t,t^{-1}\right).$$

After that if we multiply the braid $a^n\sigma$ by a on the left side then we obtain the following picture, illustrated in Fig. 13.

Therefore we will have

$$\rho_{11}\left(a^{n+1}\sigma\right) = P\left(t,t^{-1}\right)\left(1-t^{-1}+t^{-2}\right) +$$

$$Q\left(t,t^{-1}\right) + R\left(t,t^{-1}\right)\left(1-t^{1}+t^{2}\right),$$

$$\rho_{31}\left(a^{n+1}\sigma\right) = -P\left(t,t^{-1}\right)\left(1-t^{-1}\right) - Q\left(t,t^{-1}\right) - R\left(t,t^{-1}\right)\left(1-t^{1}\right).$$

Note that the same argument is sufficient to complete the proof. □

Example 1. Let $\sigma = b^{-1}a^{-1}b$, then for $n = 2$ we have

$$\rho_{11}\left(a^2b^{-1}a^{-1}b\right) = -t^{-3}+t^{-2}-1+2t-t^2-t^3+2t^4-t^5 =$$

$$t^{-2}\left(1-t^{-1}\right) + \left(-1+t-t^3+t^4\right)(1-t),$$

$$\rho_{31}\left(a^2b^{-1}a^{-1}b\right) = -t^{-2}+1-t+t^3-t^4 = -t^{-2} - \left(-1+t-t^3+t^4\right).$$

See Fig. 14. Let $m = 3$, then we can see that

$$\rho_{11}\left(a^5b^{-1}a^{-1}b\right) = t^{-2}\left(1-t^{-1}+t^{-2}-t^{-3}+t^{-4}\right) +$$

$$\left(-1+t-t^3+t^4\right)\left(1-t+t^2-t^3+t^4\right) =$$

$$t^{-6}-t^{-5}+t^{-4}-t^{-3}+t^{-2}-$$

$$1+2t-2t^2+t^3-t^5+2t^6-2t^7+t^8,$$

$$\rho_{31}\left(a^5b^{-1}a^{-1}b\right) = -t^{-2}\left(1-t^{-1}+t^{-2}-t^{-3}\right) -$$

$$\left(-1+t-t^3+t^4\right)\left(1-t+t^2-t^3\right) =$$

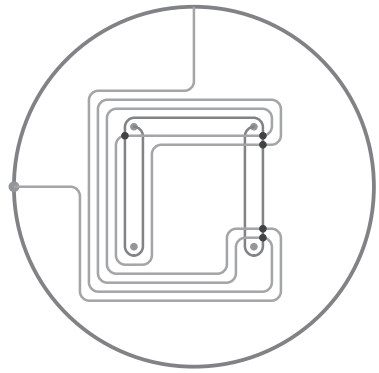


Fig. 14. The polynomial corresponding to the marked intersection point on the left side is $P(t, t^{-1}) = t^2$ and the polynomial corresponding to the marked intersection points on the right side is $R(t, t^{-1}) = -1 + t - t^3 + t^4$.

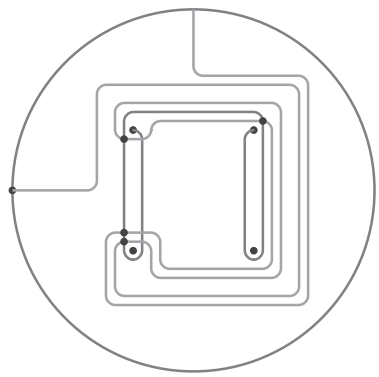


Fig. 15. The polynomial corresponding to the three marked intersection points on the left side is $P(t, t^{-1}) = -t^{-5} + t^{-4} - t^{-2}$ and the polynomial corresponding to the single marked intersection point on the right side is the monomial $Q(t, t^{-1}) = t^{-1}$.

$$t^{-5} - t^{-4} + t^{-3} - t^{-2} +$$

$$1 - 2t + 2t^2 - t^3 - t^4 + 2t^5 - 2t^6 + t^7.$$

Example 2. Let $\sigma = ba^{-2}b^{-1}$, then for $n = 2$ we have

$$\rho_{11}\left(a^2ba^{-2}b^{-1}\right) = t^{-6} - 2t^{-5} + t^{-4} + t^{-3} - t^{-2} + t^{-1} =$$

$$\left(-t^{-5} + t^{-4} - t^{-2}\right)\left(1 - t^{-1}\right) + t^{-1},$$

$$\rho_{31}\left(a^2ba^{-2}b^{-1}\right) = t^{-5} - t^{-4} + t^{-2} - t^{-1} = -\left(-t^{-5} + t^{-4} - t^{-2}\right) - t^{-1}.$$

See Fig. 15 Let $m = 4$ then we can see that

$$\rho_{11}\left(a^6ba^{-2}b^{-1}\right) = \left(-t^{-5} + t^{-4} - t^{-2}\right)\left(1 - t^{-1} + t^{-2} - t^{-3} + t^{-4} - t^{-5}\right) + t^{-1} =$$

$$t^{-10} - 2t^{-9} + 2t^{-8} - t^{-7} + t^{-6} - t^{-5} + t^{-3} - t^{-2} + t^{-1},$$

$$\rho_{13}\left(a^6ba^{-2}b^{-1}\right) = \left(-t^{-5} + t^{-4} - t^{-2}\right)\left(1 - t^{-1} + t^{-2} - t^{-3} + t^{-4}\right) + t^{-1} =$$

$$t^{-9} - 2t^{-8} + 2t^{-7} - t^{-6} + t^{-5} - t^{-3} + t^{-2} - t^{-1}.$$

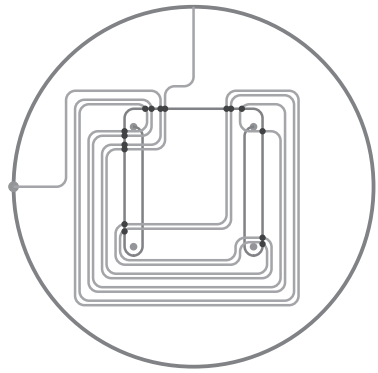


Fig. 16. The polynomial corresponding to the intersection points marked on the left side is $P(t,t^{-1}) = t^{-5} - t^{-4} + t^{-2} - 2t^{-1} + 1$. The polynomial corresponding to the marked intersection points above puncture points is $Q(t,t^{-1}) = -t^{-4} + t^{-3} - 2t^{-1} + 2 - t$ and the polynomial corresponding to the marked intersection points on the right side is $R(t,t^{-1}) = 1 - t^2 + t^3$.

Example 3. Let $\sigma = ab^2ab^{-1}$, then for $n = 2$ we have

$$\rho_{11}\left(a^2ab^2ab^{-1}\right) = -t^{-6} + 2t^{-5} - t^{-4} + t^{-3} + 3t^{-2} - 3t^{-1} + 2 - t - t^2 + 2t^3 - t^4 =$$

$$\left(t^{-5} - t^{-4} + t^{-2} - 2t^{-1} + 1\right)\left(1 - t^{-1}\right)$$

$$\left(-t^{-4} + t^{-3} - 2t^{-1} + 2 - t\right) + \left(1 - t^2 + t^3\right)(1 - t)\,,$$

$$\rho_{11}\left(a^2ab^2ab^{-1}\right) = -t^{-5} + 2t^{-4} - t^{-3} - t^{-2} + 4t^{-1} - 4 + t + t^2 - t^3 =$$

$$-\left(t^{-5} - t^{-4} + t^{-2} - 2t^{-1} + 1\right) -$$

$$\left(-t^{-4} + t^{-3} - 2t^{-1} + 2 - t\right) - \left(1 - t^2 + t^3\right).$$

See Fig. 16 Let $m = 2$ then we can see that

$$\rho_{11}\left(a^4ab^2ab^{-1}\right) = \left(t^{-5} - t^{-4} + t^{-2} - 2t^{-1} + 1\right)\left(1 - t^{-1} + t^{-2} - t^{-3}\right)$$

$$\left(-t^{-4} + t^{-3} - 2t^{-1} + 2 - t\right) + \left(1 - t^2 + t^3\right)\left(1 - t + t^2 - t^3\right) =$$

$$-t^{-8} + 2t^{-7} - 2t^{-6} + t^{-5} + t^{-4} - 3t^{-3} + 4t^{-2} - 5t^{-1} + 4$$

$$-2t + t^3 - 2t^4 + 2t^5 - t^6,$$

$$\rho_{13}\left(a^4ab^2ab^{-1}\right) = -\left(t^{-5} - t^{-4} + t^{-2} - 2t^{-1} + 1\right)\left(1 - t^{-1} + t^{-2}\right) -$$

$$\left(-t^{-4} + t^{-3} - 2t^{-1} + 2 - t\right) - \left(1 - t^2 + t^3\right)\left(1 - t + t^2\right) =$$

$$-t^{-7} + 2t^{-6} - 2t^{-5} + t^{-4} + 2t^{-3} - 4t^{-2} + 5t^{-1} - 4 + 2t - 2t^3 + 2t^4 - t^5.$$

We formulate a conjecture that describes a certain regularity, experimentally observed for images (matrices) of braids of a special form. For future reference let us state clearly that what we mean by *regularity* is that the (1, 1) and (3, 1) entries in the considered matrix are non-zero Laurent polynomials and that the difference of the degrees of lowest degree terms is equal to -1.

Conjecture 4.2. *Let $\sigma \in B_4$ be any non-trivial pure braid which is not equivalent to Δ^m, for some $m \in \mathbb{N}$. We assume that σ acts non-trivially on T_4. Then there exists a sufficiently large $l_0 \in \mathbb{N}$ with respect to the length of σ and a sufficiently large $m_0 \in \mathbb{N}$ such that for each $m > m_0, l \geq l_0$ the difference of the lowest degrees of the polynomials $\rho_{11}(a^m\sigma a^{-l})$ and $\rho_{31}(a^m\sigma a^{-l})$ is equal to -1 and the polynomials are non-zero.*

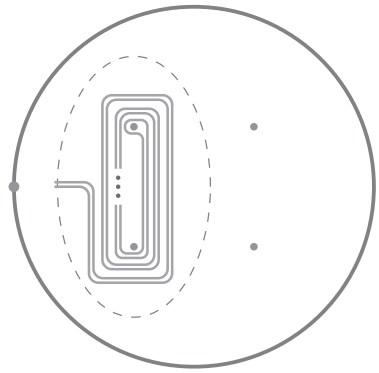


Fig. 17.

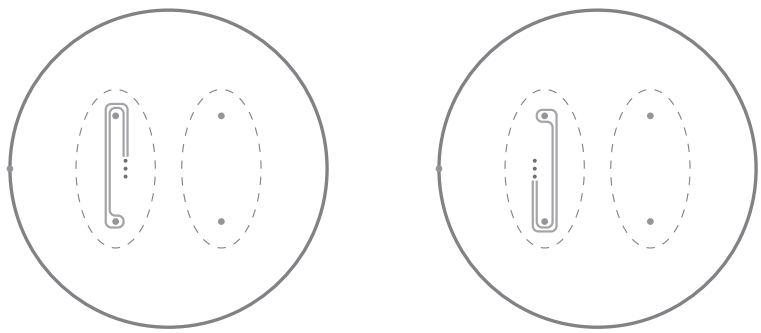


Fig. 18.

While experimental data suggest that the Conjecture is true as formulated, we are really interested in the situation when σ is a product of Bokut–Vesnin generators. Therefore we may refer to the length of σ, meaning the length of σ as a reduced word in a, b, a^{-1}, b^{-1}. Now, we give some arguments showing why we expect the Conjecture to be true. Take a sufficiently large $l_0 \in \mathbb{N}$ with respect to the length of σ. Consider the curves $N_1 a^{l_0}$ and neighborhood U_1 of T_4 inside of which it looks as in Fig. 17:

Apply to the curve $N_1 a^{l_0}$ the transformation corresponding to the braid σ^{-1}. Note that l_0 is sufficiently large with respect to the length of σ^{-1} and so in the curve $N_1 a^{l_0}\sigma^{-1}$ almost all parallel lines to line T_4 are followed by the curve $T_4\sigma^{-1}$. On the other hand the transformation corresponding to the braid σ^{-1} acts non-trivially on the line T_4. Therefore the final image $N_1 a^{l_0}\sigma^{-1}$ does not have problematic strings around T_4, as in Fig. 18:

In general, if we take any braid σ, then $F_1\sigma$ may have the strings in the form illustrated in Fig. 18. For example if $\sigma = b^{-1}ab^{-1}$ or $\sigma = a^3$ then the corresponding curves are shown in Fig. 19.

Because the curve $N_1 a^{l_0}\sigma^{-1}$ does not have any problematic strings for $n = 3$ the intersection of curves $F_1 a^3$ and $N_1 a^{l_0}\sigma^{-1}$ inside the neighborhoods U_1 and U_2 of T_4 and T_2 looks as in Fig. 13. So the $\rho_{11}(a^3\sigma a^{-l_o})$ and $\rho_{31}(a^3\sigma a^{-l_0})$ entries of the Burau matrix $\rho\left(a^n\sigma a^{-l_0}\right)$ can be written as

$$\rho_{11}\left(a^3\sigma a^{-l_0}\right) = P\left(t,t^{-1}\right)\left(1-t^{-1}\right) + Q\left(t,t^{-1}\right) + R\left(t,t^{-1}\right)(1-t),$$

$$\rho_{31}\left(a^3\sigma a^{-l_0}\right) = -P\left(t,t^{-1}\right) - Q\left(t,t^{-1}\right) - R\left(t,t^{-1}\right),$$

where polynomial $P\left(t,t^{-1}\right)$ is not zero and for each $m' \in \mathbb{N}$ we have

$$\begin{aligned}\rho_{11}\left(a^{3+m'}\sigma a^{-l_0}\right) \;=\; & P\left(t,t^{-1}\right)\left(S_{m'+1}(t^{-1})\right) \quad + Q\left(t,t^{-1}\right) \\ & + R\left(t,t^{-1}\right)\left(S_{m'+1}(t)\right),\end{aligned}$$

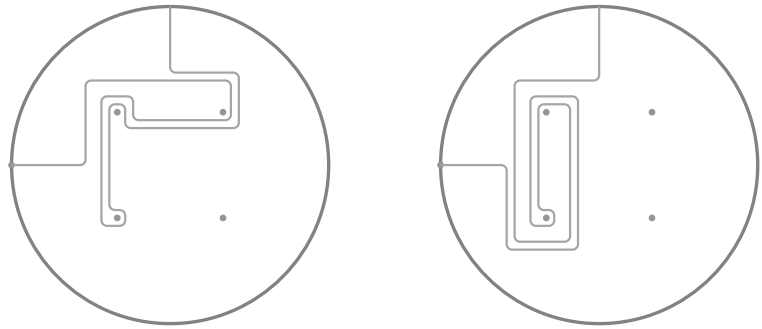


Fig. 19.

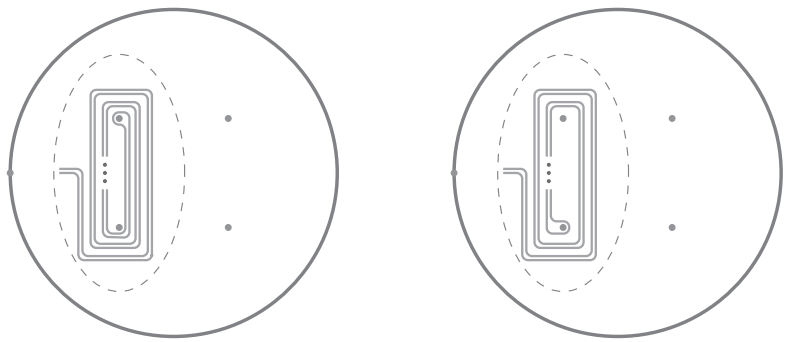


Fig. 20.

$$\rho_{31}\left(a^{3+m'}\sigma a^{-l_0}\right) = -P\left(t,t^{-1}\right)\left(S_{m'}(t^{-1})\right) \quad -Q\left(t,t^{-1}\right) \\ -R\left(t,t^{-1}\right)\left(S_{m'}(t)\right).$$

On the other hand if we compare the curves $N_1a^{l_0+1}$ and $N_1a^{l_0}$ (see Fig. 20), it is clear that they differ only by strings around T_4.

Moreover, if we consider the intersections of curves $N_1a^{l_0+1}$ and $N_1a^{l_0}$ with the strings between the puncture points p_1 and p_4 as in Fig. 21, then corresponding polynomials up to sign ϵ and multiplications t^α have the forms:

$$\left(\epsilon t^\alpha + S(t,t^{-1})(1-t^{-1})\right)\left(1-t+\cdots+(-1)^l t^{l-1}\right), \tag{*}$$

$$\left(\epsilon t^\alpha + S(t,t^{-1})(1-t^{-1})\right)\left(1-t+\cdots+(-1)^{l+1} t^{l-2}\right). \tag{**}$$

Therefore their lowest degrees are equal. Note that, l_0 is sufficiently large with respect to the length of σ and so the pictures of curves $N_1a^{l_0+1}\sigma^{-1}$ and $N_1a^{l_0}\sigma^{-1}$ 'globally' are the same. That means that in some 'local' pictures there are just different numbers of strings. Now we must look at pictures $N_1a^{l_0+1}\sigma^{-1}$ and $N_1a^{l_0}\sigma^{-1}$ inside a neighborhood of T as it was done in the previous proof (see Fig. 10). Note that by the arguments in the proof of Lemma 4.1 and same 'global' picture of the curves $N_1a^{l_0+1}\sigma^{-1}$ and $N_1a^{l_0}\sigma^{-1}$ we have

$$\rho_{11}\left(a^3\sigma a^{-l_0-1}\right) = P'\left(t,t^{-1}\right)\left(1-t^{-1}\right) + Q'\left(t,t^{-1}\right) + R'\left(t,t^{-1}\right)(1-t),$$

$$\rho_{31}\left(a^3\sigma a^{-l_0-1}\right) = -P'\left(t,t^{-1}\right) - Q'\left(t,t^{-1}\right) - R'\left(t,t^{-1}\right),$$

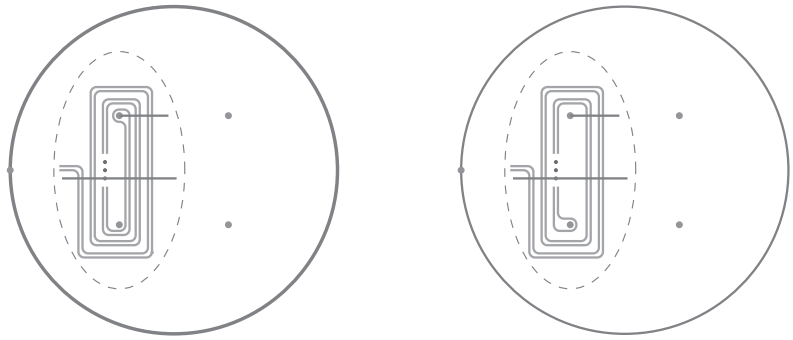


Fig. 21.

and for each $m' \in \mathbb{N}$ we have:

$$\begin{aligned}
\rho_{11}\left(a^{3+m'}\sigma a^{-l_0-1}\right) &= P'\left(t,t^{-1}\right)\left(S_{m'+1}(t^{-1})\right) + Q'\left(t,t^{-1}\right) \\
&\quad + R'\left(t,t^{-1}\right)\left(S_{m'+}(t)\right), \\
\rho_{31}\left(a^{3+m'}\sigma a^{-l_0-1}\right) &= -P'\left(t,t^{-1}\right)\left(S_{m'}(t^{-1})\right) - Q'\left(t,t^{-1}\right) \\
&\quad - R'\left(t,t^{-1}\right)\left(S_{m'+}(t)\right).
\end{aligned}$$

Take a large $m_0 = 3 + m'$ such that the difference of lowest degrees of polynomials $\rho_{11}\left(a^{m_0}\sigma a^{-l_0}\right)$ and $\rho_{31}\left(a^{m_0}\sigma a^{-l_0}\right)$ is equal to -1 and these lowest degrees come from the lowest degree of the polynomial $P(t,t^{-1})$. By (*) and (**) the polynomials $P\left(t,t^{-1}\right)$ and $P'\left(t,t^{-1}\right)$, $Q\left(t,t^{-1}\right)$ and $Q'\left(t,t^{-1}\right)$ and also $R\left(t,t^{-1}\right)$ and $R'\left(t,t^{-1}\right)$ have the same lowest degrees and so the same regularity will be true for the polynomials $\rho_{11}\left(a^{m}\sigma b a^{-l_0-1}\right)$ and $\rho_{31}\left(a^{m}\sigma a^{-l_0-1}\right)$. By induction on the length of σ it will be true for the braid $a^m\sigma a^{-l}$, $l > l_0$ as well.

Example 4. Let $\sigma = b^6a\,b^{-1}a^{-1}b^{-6}a^{-6}$. Our aim is to find n which satisfies the conditions of Lemma 4.1 and to calculate the corresponding polynomials P, Q and R. Then we will take any $l > 6$ (in our case we consider $l = 9$) and will show that lowest degrees of the corresponding polynomials do not change. For the given braid it is difficult to see the picture and write down the polynomials P, Q and R. Therefore we will use the following method: If n (in our case $n = 2$) is a number as in Lemma 4.1, then we have

$$\rho_{11}(\sigma) + \rho_{31}(\sigma) = -t^{-1}P(t,t^{-1}) - tR(t,t^{-1}),$$

$$\rho_{11}(a\sigma) + \rho_{31}(a\sigma) = t^{-2}P(t,t^{-1}) + t^{2}R(t,t^{-1}).$$

Therefore

$$P(t,t^{-1}) = \frac{t(\rho_{11}(\sigma)+\rho_{31}(\sigma)) + (\rho_{11}(a\sigma)+\rho_{31}(a\sigma))}{t^{-2}-1},$$

$$R(t,t^{-1}) = \frac{t^{-1}(\rho_{11}(\sigma)+\rho_{31}(\sigma)) + (\rho_{11}(a\sigma)+\rho_{31}(a\sigma))}{t^{2}-1}.$$

In this way we can see that for the braid $\sigma = a^2b^6ab^{-1}a^{-1}b^{-6}a^{-6}$ we have

$$\begin{aligned}
P\left(t,t^{-1}\right) &= t^{-8} - 3t^{-7} + 6t^{-6} - 9t^{-5} + 11t^{-4} - 11t^{-3} + 8t^{-2} - 2t^{-1} - \\
&\quad - 7 + 16t^{1} - 22t^{2} + 23t^{3} - 20t^{4} + 14t^{5} - 5t^{6} - 4t^{7} + \\
&\quad + 10t^{8} - 12t^{9} + 11t^{10} - 9t^{11} + 6t^{12} - 3t^{13} + t^{14}
\end{aligned}$$

$$Q\left(t,t^{-1}\right) = -t^{-1} + 2 - 2t^{1} + t^{2} - 2t^{4} + 4t^{5} - 6t^{6} + 6t^{7} - 4t^{8} + t^{9} +$$

$$+ t^{10} - 2t^{11} + 3t^{12} - 3t^{13} + 2t^{14} - t^{15},$$

$$R\left(t, t^{-1}\right) = -t^{-6} + 3t^{-5} - 6t^{-4} + 9t^{-3} - 11t^{-2} + 11t^{-1} -$$

$$- 7 - t^{1} + 10t^{2} - 18t^{3} + 23t^{4} - 22t^{5} + 17t^{6} - 8t^{7} - t^{8} + 8t^{9} - 11t^{10} +$$

$$+ 11t^{11} - 9t^{12} + 6t^{13} - 3t^{14} + t^{15}.$$

Similarly, for the braid $\sigma' = a^2b^6ab^{-1}a^{-1}b^{-6}a^{-9}$ we obtain

$$P'\left(t, t^{-1}\right) = t^{-8} - 3t^{-7} + 6t^{-6} - 9t^{-5} + 11t^{-4} - 12t^{-3} + 11t^{-2} - 8t^{-1} +$$

$$+ 1 + 8t^{1} - 16t^{2} + 21t^{3} - 23t^{4} + 23t^{5} - 19t^{6} + 12t^{7} -$$

$$- 4t^{8} - 3t^{9} + 8t^{10} - 11t^{11} + 12t^{12} - 11t^{13} + 9t^{14} - 6t^{15} + 3t^{16} - t^{17}$$

$$Q'\left(t, t^{-1}\right) = -t^{-1} + 2 - 2t^{1} + t^{2} - t^{4} + 2t^{5} - 4t^{6} + 6t^{7} - 6t^{8} + 4t^{9} -$$

$$- 2t^{10} + t^{11} - t^{13} + 2t^{14} - 3t^{15} + 3t^{16} - 2t^{17} + t^{18},$$

$$R'\left(t, t^{-1}\right) = -t^{-6} + 3t^{-5} - 6t^{-4} + 9t^{-3} - 11t^{-2} + 12t^{-1} -$$

$$- 10 + 5t^{1} + 2t^{2} - 10t^{3} + 17t^{4} - 21t^{5} + 22t^{6} - 20t^{7} + 14t^{8} - 7t^{9} +$$

$$+ 5t^{11} - 9t^{12} + 11t^{13} - 11t^{14} + 9t^{15} - 6t^{16} + 3t^{17} - t^{18}.$$

Therefore the lowest degrees of polynomials $P\left(t, t^{-1}\right)$ and $P'\left(t, t^{-1}\right)$ (same situation is with the polynomials $Q\left(t, t^{-1}\right)$ and $Q'\left(t, t^{-1}\right)$ or $R\left(t, t^{-1}\right)$ and $R'\left(t, t^{-1}\right)$) are equal.

Theorem 4.3. Conjecture 4.2 *implies faithfulness of the Burau representation for* $n = 4$.

Proof. Let us consider a nontrivial braid $\sigma \in B_4$ written as a reduced word in the Bokut–Vesnin generators. We may assume that it begins and ends with a or a^{-1} (otherwise we will conjugate by a suitable power of a).

If we interpret the braid group as the mapping class group, then there is a natural induced action on the set of isotopy classes of forks. Let σ act non-trivially on T_4. Then by Lemma 4.1 it is possible to find sufficiently large l_0 with respect to the length of σ and sufficiently large m_0, such that for each $m > m_0$ and $l > l_0$ the difference of lowest degrees of the polynomials $\rho_{11}(a^m\sigma a^{-l})$ and $\rho_{31}(a^m\sigma a^{-l})$ is equal to -1 and the polynomials are both non-zero. In particular, we can assume that $m = l$ and so $\rho_{11}(a^m\sigma a^{-m})$ and $\rho_{31}(a^m\sigma a^{-m})$ are both non-zero which contradicts the assumption that $a^m\sigma a^{-m} \in \ker\rho$.

The general case (when we do not assume that σ acts non-trivially on T_4) is easily reduced to the one discussed above. The reason is that if σ acts trivially on all four segments, then σ is a power of Δ which is not possible if σ is a product of the Bokut–Vesnin generators. And if σ acts non-trivially on at least one of the four segments, then we can rotate the whole disc to make the action non-trivial for T_4. □

Remark 4.4. We have a C++ program checking whether our regularity works or not for randomly generated examples. We calculated millions of examples and the regularity was always confirmed. In fact we considered examples of type $a^3b^3wb^{-3}a^{-3}$, where $a^3b^3wb^{-3}a^{-3}$ is a reduced word in the Bokut–Vesnin generators. Such a version of Proposition 3.1 is sufficient for the Burau representation faithfulness problem.

Acknowledgment

Most of this research was conducted while the first author was a postdoc at the University of Warsaw during the Spring 2014 semester, with support of Erasmus Mundus Project (WEBB).

References

[1] Stephen Bigelow, Does the Jones polynomial detect the unknot?, J. Knot Theory Ramifications 11 (4) (2002) 493–505.
[2] Joan S. Birman, Braids, links, and mapping class groups, in: Annals of Mathematics Studies, vol. 82, Princeton University Press, Princeton, NJ, 1974.
[3] Stephen Bigelow, The Burau representation is not faithful for $n = 5$, Geom. Topol. 3 (1999) 397–404.
[4] D.D. Long, M. Paton, The Burau representation is not faithful for $n \geq 6$, Topology 32 (2) (1993) 439–447.
[5] John Atwell Moody, The Burau representation of the braid group B_n is unfaithful for large n, Bull. Amer. Math. Soc. (N.S.). 25 (2) (1991) 379–384.
[6] Leonid Bokut, Andrei Vesnin, New rewriting system for the braid group B_4, in: Proceedings of Symposium in Honor of Bruno Buchberger'S 60th Birthday "Logic, Mathematics and Computer Sciences: Intersections", Research Institute for Symbolic Computations, Linz, Austia, 2002, pp. 48–60. Report Series No. 02-60.
[7] Matthieu Calvez, Tetsuya Ito, Garside-theoretic analysis of Burau representations. arXiv:1401.2677v2.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 354–360
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Besov continuity for global operators on compact Lie groups: The critical case $p = q = \infty$.

Duván Cardona

Department of Mathematics, Pontificia Universidad Javeriana, Bogotá, Colombia

Received 8 April 2018; received in revised form 2 July 2018; accepted 1 August 2018
Available online 11 August 2018

Abstract

In this note, we study the mapping properties of global pseudo-differential operators with symbols in Ruzhansky–Turunen classes on Besov spaces $B^s_{\infty,\infty}(G)$. The considered classes satisfy Fefferman type conditions of limited regularity.

Keywords: Pseudo-differential operator; Compact Lie groups; Ruzhansky–Turunen calculus; Global analysis

1. Introduction

This note on the Besov boundedness of pseudo-differential operators on compact Lie groups in $B^s_{\infty,\infty}$ is based on the matrix-valued quantization procedure developed by M. Ruzhansky and V. Turunen in [1].

The Besov spaces $B^s_{p,q}$ arose from attempts to unify the various definitions of several fractional-order Sobolev spaces. By following the historical note of Grafakos [2, p. 113], we recall that Taibleson studied the generalized Hölder–Lipschitz spaces $\Lambda^s_{p,q}$ on $\mathbb{R}^n$, and these spaces were named after Besov spaces in honor to O. V. Besov who obtained a trace theorem and important embedding properties for them (see Besov [3,4]). Dyadic decompositions for Besov spaces on $\mathbb{R}^n$ were introduced by J. Peetre as well as other embedding properties (see Peetre [5,6]). We will use the formulation of Besov spaces $B^s_{p,q}(G)$, through the representation theory of compact Lie groups G introduced and consistently developed by E. Nursultanov, M. Ruzhansky, and S. Tikhonov in [7,8].

The present paper is a continuation of a series of our previous papers [9–12], in which were investigated the mapping properties of global operators (i.e., global pseudo-differential operators on compact Lie groups) on Besov spaces $B^s_{p,q}$, $-\infty < s < \infty$, $1 < p < \infty$ and $0 < q \leq \infty$. So, in this note we study the mapping properties for global operators in the case of Besov spaces $B^s_{\infty,\infty}$, $-\infty < s < \infty$.

The main tool in the proof of the Besov boundedness results presented in [9–12], is the L^p-multipliers theorems proved in Ruzhansky and Wirth [13,14], Delgado and Ruzhansky [15], Fischer [16] and Akylzhanov and

E-mail address: cardonaduvan@javeriana.edu.co.
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.08.001

Ruzhansky [17]. In general, such L^p-estimates, $1 < p < \infty$, cannot be extended to $L^\infty(G)$ and consequently we need to consider other techniques for the formulation of a boundedness result on $B^s_{\infty,\infty}$. Let us recall some L^p and Besov estimates for global operators, in order to announce our main theorem. First of all, we recall some notions on the global analysis of pseudo-differential operators.

If G is a compact Lie group and $\widehat{G}$ is its unitary dual, that is the set of equivalence classes of continuous irreducible unitary representations of G, Ruzhansky–Turunen's approach associates to every bounded linear operator A on $C^\infty(G)$, a matrix valued symbol $\sigma_A(x,\xi)$ given by $\sigma_A(x,\xi) := \xi(x)^*(A\xi)(x)$, $x \in G$ and $\xi \in [\xi] \in \widehat{G}$. This allows us to write the operator A in terms of representations in G as

$$Af(x) = \sum_{[\xi]\in\widehat{G}} d_\xi \mathrm{Tr}(\xi(x)\sigma_A(x,\xi)\widehat{f}(\xi)), \tag{1.1}$$

for all $f \in C^\infty(G)$, where $\mathscr{F}(f) := \widehat{f}$ is the Fourier transform on the group G.

The Hörmander classes $\Psi^m_{\rho,\delta}(G)$, $m \in \mathbb{R}$, $\rho > \max\{\delta, 1-\delta\}$, where characterized in [1,18] by the following condition: $A \in \Psi^m_{\rho,\delta}(G)$ if only if its matrix-valued symbol $\sigma_A(x,\xi)$ satisfies the inequalities

$$\|\partial_x^\alpha \Delta^\beta \sigma_A(x,\xi)\|_{op} \leq C_{\alpha,\beta}\langle\xi\rangle^{m-\rho|\beta|+\delta|\alpha|}, \tag{1.2}$$

for every $\alpha, \beta \in \mathbb{N}^n$. The discrete differential operator Δ^β (called difference operator of first order) is the main tool in this theory (see [1,18]).

The L^p-mapping properties for global operators on compact Lie groups can be summarized as follows: if G is a compact Lie group and n is its dimension, $\varkappa$ is the less integer larger that $\frac{n}{2}$ and $l := [n/p]+1$, under one of the following conditions

- $\|\partial_x^\beta \mathbb{D}_\xi^\alpha \sigma_A(x,\xi)\|_{op} \leq C_\alpha \langle\xi\rangle^{-|\alpha|}$, for all $|\alpha| \leq \varkappa$, $|\beta| \leq l$ y $[\xi] \in \widehat{G}$, (Ruzhansky and Wirth [13,14]),
- $\|\mathbb{D}_\xi^\alpha \partial_x^\beta \sigma_A(x,\xi)\|_{op} \leq C_{\alpha,\beta}\langle\xi\rangle^{-m-\rho|\alpha|+\delta|\beta|}$, $|\alpha| \leq \varkappa, |\beta| \leq l$, $m \geq \varkappa(1-\rho)|\frac{1}{p} - \frac{1}{2}| + \delta l$, (Delgado and Ruzhansky [15]),
- $\|\mathbb{D}_\xi^\alpha \partial_x^\beta \sigma_A(x,\xi)\|_{op} \leq C_{\alpha,\beta}\langle\xi\rangle^{-\nu-\rho|\alpha|+\delta|\beta|}$, $\alpha \in \mathbb{N}^n, |\beta| \leq l, 0 \leq \nu < \frac{n}{2}(1-\rho)$, $|\frac{1}{p}-\frac{1}{2}| \leq \frac{\nu}{n}(1-\rho)^{-1}$, (Delgado and Ruzhansky [15]),
- $\|\sigma_A\|_{\Sigma_s} := \sup_{[\xi]\in\widehat{G}}[\|\sigma_A(\xi)\|_{op} + \|\sigma_A(\xi)\eta(r^{-2}\mathcal{L}_G)\|_{\dot{H}^s(\widehat{G})}] < \infty$, (Fischer [16]),

the global operator $A \equiv T_a$ is bounded on $L^p(G)$. On the other hand, if $A : C^\infty(G) \to C^\infty(G)$ is a linear and bounded operator, then under any one of the following conditions

- $\|\mathbb{D}_\xi^\alpha \partial_x^\beta \sigma_A(x,\xi)\|_{op} \leq C_\alpha \langle\xi\rangle^{-|\alpha|}$, for all $|\alpha| \leq \varkappa$, $|\beta| \leq l$ and $[\xi] \in \widehat{G}$,
- $\|\mathbb{D}_\xi^\alpha \partial_x^\beta \sigma_A(x,\xi)\|_{op} \leq C_{\alpha,\beta}\langle\xi\rangle^{-\nu-\rho|\alpha|}$, $\alpha \in \mathbb{N}^n, |\beta| \leq l$, $|\frac{1}{p}-\frac{1}{2}| \leq \frac{\nu}{n}(1-\rho)^{-1}$, $0 \leq \nu < \frac{n}{2}(1-\rho)$,
- $\|\mathbb{D}_\xi^\alpha \partial_x^\beta \sigma_A(x,\xi)\|_{op} \leq C_{\alpha,\beta}\langle\xi\rangle^{-m-\rho|\alpha|+\delta|\beta|}$, $|\alpha| \leq \varkappa, |\beta| \leq l$, $m \geq \varkappa(1-\rho)|\frac{1}{p}-\frac{1}{2}| + \delta l$,

the corresponding pseudo-differential operator A extends to a bounded operator from $B^r_{p,q}(G)$ into $B^r_{p,q}(G)$ for all $r \in \mathbb{R}$, $1 < p < \infty$ and $0 < q \leq \infty$, (see Cardona [9–11]). For $p, q = \infty$ we have the following theorem which is our main result in this paper.

Theorem 1.1. *Let G be a compact Lie group of dimension n. Let $0 < \rho \leq 1$, $k := [\frac{n}{2}] + 1$, and let $A : C^\infty(G) \to C^\infty(G)$ be a pseudo-differential operator with symbol σ satisfying*

$$\|\mathbb{D}_\xi^\alpha \sigma(x,\xi)\|_{op} \leq C_\alpha \langle\xi\rangle^{-\frac{n}{2}(1-\rho)-\rho|\alpha|} \tag{1.3}$$

for all $|\alpha| \leq k$. Then $A : B^s_{\infty,\infty}(G) \to B^s_{\infty,\infty}(G)$ extends to a bounded linear operator for all $-\infty < s < \infty$. Moreover,

$$\|A\|_{\mathscr{B}(B^s_{\infty,\infty})} \leq C \sup\{C_\alpha : |\alpha| \leq k\}. \tag{1.4}$$

Theorem 1.1 implies the following result.

Corollary 1.2. *Let G be a compact Lie group of dimension n. Let $0 < \rho \leq 1$, $0 \leq \delta \leq 1$, $\ell \in \mathbb{N}$, $k := [\frac{n}{2}] + 1$, and let $A : C^\infty(G) \to C^\infty(G)$ be a pseudo-differential operator with symbol σ satisfying*

$$\|\partial_x^\beta \mathbb{D}_\xi^\alpha \sigma(x,\xi)\|_{op} \leq C_\alpha \langle\xi\rangle^{-m-\rho|\alpha|+\delta|\beta|} \tag{1.5}$$

for all $|\alpha| \leq k$, $|\beta| \leq \ell$. *Then* $A : B^{s}_{\infty,\infty}(G) \to B^{s}_{\infty,\infty}(G)$ *extends to a bounded linear operator for all* $-\infty < s < \infty$ *provided that* $m \geq \delta\ell + \frac{n}{2}(1-\rho)$.

Besov spaces on graded Lie groups, as well as the action of Fourier multipliers and spectral multipliers on these spaces can be found in Cardona and Ruzhansky [19,20]. We refer the reader to the Refs. [21,22] for boundedness properties of pseudo-differential operators in Besov spaces on $\mathbb{R}^n$. This paper is organized as follows. In the next section we present some basics on the calculus of global operators. Finally, in Section 3 we prove our Besov estimates.

2. Pseudo-differential operators on compact Lie groups

2.1. Fourier analysis and Sobolev spaces on compact Lie groups

In this section we will introduce some preliminaries on pseudo-differential operators on compact Lie groups and some of its properties on L^p-spaces. There are two notions of pseudo-differential operators on compact Lie groups. The first notion is the case of general manifolds (based on the idea of *local symbols* as in Hörmander [23]) and, in a much more recent context, the one of global pseudo-differential operators on compact Lie groups as defined by Ruzhansky and Turunen [1]. We adopt this last notion for our work, because we will use a description of the Besov spaces B^{s}_{∞}, ∞ through representation theory. We will always equip a compact Lie group with the Haar measure μ_G. For simplicity, we will write $L^{\infty}(G)$ for $L^{\infty}(G, \mu_G)$. The following assumptions are respectively the Fourier transform and the Fourier inversion formula for smooth functions,

$$\widehat{\varphi}(\xi) = \int_G \varphi(x)\xi(x)^* dx, \qquad \varphi(x) = \sum_{[\xi]\in\widehat{G}} d_\xi \mathrm{Tr}(\xi(x)\widehat{\varphi}(\xi)).$$

The Peter–Weyl Theorem on G implies the Plancherel identity on $L^2(G)$,

$$\|\varphi\|_{L^2(G)} = \left(\sum_{[\xi]\in\widehat{G}} d_\xi \mathrm{Tr}(\widehat{\varphi}(\xi)\widehat{\varphi}(\xi)^*)\right)^{\frac{1}{2}} = \|\widehat{\varphi}\|_{L^2(\widehat{G})}.$$

Here

$$\|A\|^2_{HS} = \mathrm{Tr}(AA^*),$$

denotes the Hilbert–Schmidt norm of matrices. Now, we introduce global pseudo-differential operators in the sense of Ruzhansky and Turunen. Any continuous linear operator A on G mapping $C^{\infty}(G)$ into $\mathcal{D}'(G)$ gives rise to a *matrix-valued global (or full) symbol* $\sigma_A(x,\xi) \in \mathbb{C}^{d_\xi \times d_\xi}$ given by

$$\sigma_A(x,\xi) = \xi(x)^*(A\xi)(x), \tag{2.1}$$

which can be understood from the distributional viewpoint. Then it can be shown that the operator A can be expressed in terms of such a symbol as [1]

$$Af(x) = \sum_{[\xi]\in\widehat{G}} d_\xi \mathrm{Tr}[\xi(x)\sigma_A(x,\xi)\widehat{f}(\xi)]. \tag{2.2}$$

In this paper we use the notation $\mathrm{Op}(\sigma_A) = A$. $L^p(\widehat{G})$ spaces on the unitary dual can be well defined. If $p = 2$, $L^2(\widehat{G})$ is defined by the norm

$$\|\Gamma\|^2_{L^2(\widehat{G})} = \sum_{[\xi]\in\widehat{G}} d_\xi \|\Gamma(\xi)\|^2_{HS}.$$

Now, we want to introduce Sobolev spaces and, for this, we give some basic tools. Let $\xi \in \mathrm{Rep}(G) := \cup\widehat{G} = \{\xi : [\xi] \in \widehat{G}\}$, if $x \in G$ is fixed, $\xi(x) : H_\xi \to H_\xi$ is a unitary operator and $d_\xi := \dim H_\xi < \infty$. There exists a non-negative real number $\lambda_{[\xi]}$ depending only on the equivalence class $[\xi] \in \hat{G}$, but not on the representation ξ, such that $-\mathcal{L}_G\xi(x) = \lambda_{[\xi]}\xi(x)$; here $\mathcal{L}_G$ is the Laplacian on the group G (in this case, defined as the Casimir element on G). Let $\langle\xi\rangle$ denote the function $\langle\xi\rangle = (1+\lambda_{[\xi]})^{\frac{1}{2}}$.

Definition 2.1. For every $s \in \mathbb{R}$, the *Sobolev space* $H^s(G)$ on the Lie group G is defined by the condition: $f \in H^s(G)$ if only if $\langle \xi \rangle^s \widehat{f} \in L^2(\widehat{G})$. The Sobolev space $H^s(G)$ is a Hilbert space endowed with the inner product $\langle f, g \rangle_s = \langle \Lambda_s f, \Lambda_s g \rangle_{L^2(G)}$, where, for every $r \in \mathbb{R}$, $\Lambda_s : H^r \to H^{r-s}$ is the bounded pseudo-differential operator with symbol $\langle \xi \rangle^s I_\xi$. In L^p spaces, the p-Sobolev space of order s, $H^{s,p}(G)$, is defined by functions satisfying

$$\|f\|_{H^{s,p}(G)} := \|\Lambda_s f\|_{L^p(G)} < \infty. \tag{2.3}$$

2.2. Differential and difference operators

In order to classify symbols by its regularity we present the usual definition of differential operators and the difference operators used introduced in [1].

Definition 2.2. Let $(Y_j)_{j=1}^{\dim(G)}$ be a basis for the Lie algebra $\mathfrak{g}$ of G, and let ∂_j be the left-invariant vector fields corresponding to Y_j. We define the differential operator associated to such a basis by $D_{Y_j} = \partial_j$ and, for every $\alpha \in \mathbb{N}^n$, the *differential operator* ∂_x^α is the one given by $\partial_x^\alpha = \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}$. Now, if ξ_0 is a fixed irreducible representation, the matrix-valued *difference operator* is given by $\mathbb{D}_{\xi_0} = (\mathbb{D}_{\xi_0,i,j})_{i,j=1}^{d_{\xi_0}} = \xi_0(\cdot) - I_{d_{\xi_0}}$. If the representation is fixed we omit the index ξ_0 so that, from a sequence $\mathbb{D}_1 = \mathbb{D}_{\xi_0,j_1,i_1}, \ldots, \mathbb{D}_n = \mathbb{D}_{\xi_0,j_n,i_n}$ of operators of this type we define $\mathbb{D}_\xi^\alpha = \mathbb{D}_1^{\alpha_1} \cdots \mathbb{D}_n^{\alpha_n}$, where $\alpha \in \mathbb{N}^n$.

2.3. Besov spaces

We introduce the Besov spaces on compact Lie groups using the Fourier transform on the group G as follows.

Definition 2.3. Let $r \in \mathbb{R}$, $0 \leq q < \infty$ and $0 < p \leq \infty$. If f is a measurable function on G, we say that $f \in B^r_{p,q}(G)$ if f satisfies

$$\|f\|_{B^r_{p,q}} := \left(\sum_{m=0}^{\infty} 2^{mrq} \| \sum_{2^m \leq \langle \xi \rangle < 2^{m+1}} d_\xi \operatorname{Tr}[\xi(x) \widehat{f}(\xi)] \|^q_{L^p(G)} \right)^{\frac{1}{q}} < \infty. \tag{2.4}$$

If $q = \infty$, $B^r_{p,\infty}(G)$ consists of those functions f satisfying

$$\|f\|_{B^r_{p,\infty}} := \sup_{m \in \mathbb{N}} 2^{mr} \| \sum_{2^m \leq \langle \xi \rangle < 2^{m+1}} d_\xi \operatorname{Tr}[\xi(x) \widehat{f}(\xi)] \|_{L^p(G)} < \infty. \tag{2.5}$$

If we denote by $\operatorname{Op}(\chi_m)$ the Fourier multiplier associated to the symbol

$$\chi_m(\eta) = 1_{\{[\xi]: 2^m \leq \langle \xi \rangle < 2^{m+1}\}}(\eta),$$

we also write,

$$\|f\|_{B^r_{p,q}} = \|\{2^{mr} \|\operatorname{Op}(\chi_m) f\|_{L^p(G)}\}_{m=0}^{\infty}\|_{l^q(\mathbb{N})}, \ 0 < p, q \leq \infty, \ r \in \mathbb{R}. \tag{2.6}$$

Remark 2.4. For every $s \in \mathbb{R}$, $H^{s,2}(G) = H^s(G) = B^s_{2,2}(G)$. Besov spaces according to Definition 2.3 were introduced in [7] on compact homogeneous manifolds where, in particular, the authors obtained its embedding properties. On compact Lie groups such spaces were characterized, via representation theory, in [8].

3. Global operators on $B^s_{\infty,\infty}(G)$-spaces

In this section we prove our Besov estimate for global pseudo-differential operators. Our starting point is the following lemma which is slight variation of one due to J. Delgado and M. Ruzhansky (see Lemma 4.11 of [15]) and whose proof is verbatim to Delgado–Ruzhansky's proof.

Lemma 3.1. *Let G be a compact Lie group of dimension n. Let $0 < \rho \leq 1$, $k := [\frac{n}{2}] + 1$, and let $A : C^\infty(G) \to C^\infty(G)$ be a pseudo-differential operator with symbol σ satisfying*

$$\|\mathbb{D}_\xi^\alpha \sigma(x, \xi)\|_{op} \leq C_\alpha \langle \xi \rangle^{-\frac{n}{2}(1-\rho)-\rho|\alpha|} \tag{3.1}$$

for all $|\alpha| \leq k$. Let us assume that σ is supported in $\{\xi : R \leq \langle \xi \rangle \leq 2R\}$ for some $R > 0$. Then $A : L^\infty(G) \to L^\infty(G)$ extends to a bounded linear operator with norm operator independent of R. Moreover,

$$\|A\|_{\mathscr{B}(L^\infty)} \leq C \sup\{C_\alpha : |\alpha| \leq k\}. \tag{3.2}$$

So, we are ready for the proof of our main result.

Proof of Theorem 1.1. Our proof consists of two steps. In the first one, we prove the statement of the theorem for Fourier multipliers, i.e., pseudo-differential operators depending only on the Fourier variables ξ. Later, we extend the result for general global operators.

Step 1. Let us consider $\mathcal{R} := (I - \mathcal{L}_G)^{\frac{1}{2}}$ where $\mathcal{L}_G$ is the Laplacian on G. Let us denote by ψ the characteristic function of the interval $I = [1/2, 1]$. Denote by ψ_l the function $\psi_l(t) = \psi(2^{-l}t)$, $t \in \mathbb{R}$. We will use the following characterization for $B^s_{\infty,\infty}(G)$: $f \in B^s_{\infty,\infty}(G)$, if and only if,

$$\|f\|_{B^s_{\infty,\infty}(G)} := \sup_{l \geq 0} 2^{ls} \|\psi_l(\mathcal{R}) f\|_{L^\infty(G)} \tag{3.3}$$

where $\psi_l(\mathcal{R})$ is defined by the functional calculus associated to the self-adjoint operator $\mathcal{R}$. If $A \equiv \sigma(D_x)$ has a symbol depending only on the Fourier variables ξ, then

$$\|\sigma(D_x) f\|_{B^s_{\infty,\infty}(G)} := \sup_{l \geq 0} 2^{ls} \|\psi_l(\mathcal{R})\sigma(D_x) f\|_{L^\infty(G)}. \tag{3.4}$$

Taking into account that the operator $\sigma(D_x)$ commutes with $\psi_l(\mathcal{R})$ for every l, that

$$\psi_l(\mathcal{R})\sigma(D_x) = \sigma(D_x)\psi_l(\mathcal{R}) = \sigma_l(D_x)\psi_l(\mathcal{R}) \tag{3.5}$$

where $\sigma_l(D_x)$ is the pseudo-differential operator with matrix-valued symbol

$$\sigma_l(\xi) = \sigma(\xi) \cdot 1_{\{\xi : 2^{l-1} \leq \langle \xi \rangle \leq 2^{l+1}\}},$$

and that $\sigma_l(D_x)$ has a symbol supported in $\{\xi : 2^{l-1} \leq \langle \xi \rangle \leq 2^{l+1}\}$, by Lemma 3.1 we deduce that $\sigma_l(D_x)$ is a bounded operator on $L^\infty(G)$ with operator norm independent on l. In fact, σ_l satisfies the symbol inequalities

$$\|\mathbb{D}^\alpha_\xi \sigma_l(\xi)\|_{op} \leq C_\alpha \langle \xi \rangle^{-\frac{n}{2}(1-\rho)-\varepsilon|\alpha|},$$

for all $|\alpha| \leq k$, and consequently

$$\|\sigma_l(D_x)\|_{\mathscr{B}(L^\infty)} \leq C \sup\{C_\alpha : |\alpha| \leq k\}. \tag{3.6}$$

So, we have

$$\begin{aligned} &\|\psi_l(\mathcal{R})\sigma(D_x) f\|_{L^\infty(G)} \\ &\qquad = \|\sigma_l(D_x)\psi_l(\mathcal{R}) f\|_{L^\infty(G)} \leq \|\sigma_l(D_x)\|_{\mathscr{B}(L^\infty)} \|\psi_l(\mathcal{R}) f\|_{L^\infty(G)} \\ &\qquad \lesssim \sup\{C_\alpha : |\alpha| \leq k\} \|\psi_l(\mathcal{R}) f\|_{L^\infty(G)}. \end{aligned}$$

As a consequence, we obtain

$$\begin{aligned} \|\sigma(D_x) f\|_{B^s_{\infty,\infty}(G)} &= \sup_{l \geq 0} 2^{ls} \|\psi_l(\mathcal{R})\sigma(D_x) f\|_{L^\infty(G)} \\ &\lesssim \sup\{C_\alpha : |\alpha| \leq k\} \sup_{l \geq 0} 2^{ls} \|\psi_l(\mathcal{R}) f\|_{L^\infty(G)} \\ &\asymp \sup\{C_\alpha : |\alpha| \leq k\} \|f\|_{B^s_{\infty,\infty}(G)}. \end{aligned}$$

Step 2. Now, we extend the estimate from multipliers to pseudo-differential operators. So, let us define for every $z \in G$, the multiplier

$$\sigma_z(D_x) f(x) = \sum_{[\xi] \in \widehat{G}} d_\xi \mathrm{Tr}[\xi(x)\sigma(z, \xi)\widehat{f}(\xi)].$$

For every $x \in G$ we have the equality,

$$\sigma_x(D_x) f(x) = Af(x),$$

and we can estimate the Besov norm of the function $\sigma(x, D_x)f$, as follows

$$\begin{aligned}
\|\sigma_x(D_x)f(x)\|_{B^s_{\infty,\infty}} &\asymp \sup_{l\geq 0} 2^{ls} \operatorname*{ess\,sup}_{x\in G} |\psi_l(\mathcal{R})\sigma(x, D_x)f(x)| \\
&= \sup_{l\geq 0} 2^{ls} \operatorname*{ess\,sup}_{x\in G} |\sum_{[\xi]\in\widehat{G}} d_\xi \operatorname{Tr}[\xi(x)\psi_l(\xi)\sigma(x,\xi)\widehat{f}(\xi)]| \\
&\leq \sup_{l\geq 0} 2^{ls} \operatorname*{ess\,sup}_{x\in G} \sup_{z\in G} |\sum_{[\xi]\in\widehat{G}} d_\xi \operatorname{Tr}[\xi(x)\psi_l(\xi)\sigma(z,\xi)\widehat{f}(\xi)]| \\
&= \sup_{l\geq 0} 2^{ls} \operatorname*{ess\,sup}_{x\in G} \sup_{z\in G} |\psi_l(\mathcal{R})\sigma_z(D_x)f(x)| \\
&= \sup_{l\geq 0} 2^{ls} \operatorname*{ess\,sup}_{x\in G} \sup_{z\in G} |\sigma_z(D_x)\psi_l(\mathcal{R})f(x)| \\
&\leq \sup_{l\geq 0} 2^{ls} \operatorname*{ess\,sup}_{x\in G} \sup_{z\in G} \operatorname*{ess\,sup}_{\varkappa\in G} |\sigma_z(D_\varkappa)\psi_l(\mathcal{R})f(\varkappa)| \\
&= \sup_{l\geq 0} 2^{ls} \sup_{z\in G} \|\sigma_z(D_\varkappa)\psi_l(\mathcal{R})f(\varkappa)\|_{L^\infty(G)}.
\end{aligned}$$

From the estimate for the operator norm of multipliers proved in the first step, we deduce

$$\sup_{z\in G} \|\sigma_z(D_x)\psi_l(\mathcal{R})f\|_{L^\infty(G)} \lesssim \sup\{C_\alpha : |\alpha| \leq k\} \|\psi_l(\mathcal{R})f\|_{L^\infty(G)}.$$

So, we have

$$\|\sigma_x(D_x)f(x)\|_{B^s_{\infty,\infty}} \lesssim \sup\{C_\alpha : |\alpha| \leq k\} \|f\|_{B^s_{\infty,\infty}(G)}. \tag{3.7}$$

Thus, we finish the proof. □

Now, we present the following result for symbols admitting differentiability in the spatial variables.

Corollary 3.2. *Let G be a compact Lie group of dimension n. Let $0 < \rho \leq 1$, $0 \leq \delta \leq 1$, $\ell \in \mathbb{N}$, $k := [\frac{n}{2}] + 1$, and let $A : C^\infty(G) \to C^\infty(G)$ be a pseudo-differential operator with symbol σ satisfying*

$$\|\partial_x^\beta \mathbb{D}_\xi^\alpha \sigma(x,\xi)\|_{op} \leq C_\alpha \langle\xi\rangle^{-m-\rho|\alpha|+\delta|\beta|} \tag{3.8}$$

for all $|\alpha| \leq k$, $|\beta| \leq \ell$. Then $A : B^s_{\infty,\infty}(G) \to B^s_{\infty,\infty}(G)$ extends to a bounded linear operator for all $-\infty < s < \infty$ provided that $m \geq \delta\ell + \frac{n}{2}(1-\rho)$.

Proof. Let us observe that

$$\langle\xi\rangle^{-m-\rho|\alpha|+\delta|\beta|} \leq \langle\xi\rangle^{-\frac{n}{2}(1-\rho)+\rho|\alpha|} \tag{3.9}$$

when $m \geq \delta\ell + \frac{n}{2}(1-\rho)$. So, we finish the proof if we apply Theorem 1.1. □

Acknowledgment

This project was partially supported by the Department of Mathematics of the Pontificia Universidad Javeriana, Bogotá-Colombia.

References

[1] M. Ruzhansky, V. Turunen, Pseudo-Differential Operators and Symmetries: Background Analysis and Advanced Topics, Birkhäuser-Verlag, Basel, 2010.
[2] L. Grafakos, Classical Fourier Analysis Grad, in: Texts in Math., vol. 249, Springer-Verlag, New York, 2008.
[3] O.V. Besov, On a family of function spaces. Embeddings theorems and applications [in Russian], Dokl. Akad. Nauk SSSR 126 (1959) 1163–1165.
[4] O.V. Besov, On a family of function spaces in connection with embeddings and extensions, [in Russian], Trudy Mat. Inst. Steklov 60 (1961) 42–81.
[5] J. Peetre, Sur les espaces de Besov, C. R. Acad. Sci. Paris 264 (1967) 281–283.
[6] J. Peetre, Remarques sur les espaces de Besov. Le case $0 < p < 1$, C. R. Acad. Sci. Paris 277 (1973) 947–950.

[7] E. Nursultanov, M. Ruzhansky, S. Tikhonov, Nikolskii inequality and Besov, Triebel-Lizorkin, Wiener and Beurling spaces on compact homogeneous manifolds, Ann. Sc. Norm. Super. Pisa Cl. Sci. XVI (2016) 981–1017.
[8] E. Nursultanov, M. Ruzhansky, S. Tikhonov, Nikolskii inequality and functional classes on compact Lie groups, Funct. Anal. Appl. 49 (2015) 226–229.
[9] D. Cardona, Besov continuity for Multipliers defined on compact Lie groups, Palest. J. Math. 5 (2) (2016) 35–44.
[10] D. Cardona, Besov continuity of pseudo-differential operators on compact Lie groups revisited, C. R. Math. Acad. Sci. Paris 355 (5) (2017) 533–537.
[11] D. Cardona, Besov continuity for pseudo-differential operators on compact homogeneous manifolds, J. Pseudo-Diff. Oper. Appl. (2018) (in press).
[12] D. Cardona, Nuclear pseudo-differential operators in Besov spaces on compact Lie groups, J. Fourier Anal. Appl. 23 (5) (2017) 1238–1262.
[13] M. Ruzhansky, J. Wirth, On multipliers on compact Lie groups, Funct. Anal. Appl. 47 (2013) 72–75.
[14] M. Ruzhansky, J. Wirth, L^p Fourier multipliers on compact Lie groups, Math. Z. 280 (2015) 3–4, 621–642.
[15] J. Delgado, M. Ruzhansky, L^p-bounds for pseudo-differential operators on compact Lie groups, J. Inst. Math. Jussieu (2018) (in press).
[16] V. Fischer, Hörmander condition for Fourier multipliers on compact Lie groups. arXiv:1610.06348.
[17] R. Akylzhanov, E. Nursultanov, M. Ruzhansky, Hardy-Littlewood-Paley type inequalities on compact Lie groups, Math. Notes 100 (2016) 287–290.
[18] V. Fischer, Intrinsic pseudo-differential calculi on any compact Lie group, J. Funct. Anal. 268 (11) (2015) 3404–3477.
[19] D. Cardona, M. Ruzhansky, Littlewood-Paley theorem, Nikolskii inequality, Besov spaces, Fourier and spectral multipliers on graded Lie groups. arXiv:1610.04701.
[20] D. Cardona, M. Ruzhansky, Multipliers for Besov spaces on graded Lie groups, C. R. Math. Acad. Sci. Paris 355 (4) (2017) 400–405.
[21] R. Duduchava, F.O. Speck, Pseudodiferential operators on compact manifolds with Lipschitz boundary, Math. Nachr. 160 (1993) 149–191.
[22] E. Shargorodsky, Some remarks on the boundedness of pseudodiferential operators, Math. Nachr. 183 (1) (1997) 229–273.
[23] L. Hörmander, The Analysis of the Linear Partial Differential Operators, Vol. III, Springer-Verlag, 1985.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 361–371
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Invex programming problems with equality and inequality constraints

A.K. Das[a], R. Jana[b,*], Deepmala[a]

[a] *SQC & OR Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata, 700 108, India*
[b] *Department of Mathematics, Jadavpur University , Kolkata, 700 032, India*

Received 9 October 2017; accepted 14 April 2018
Available online 3 May 2018

Abstract

The class of functions is known as invex function (invariant convex) in the literature and the name derives from the fact that the convex like property of such functions remains invariant under all diffeomorphisms of R^n into R^n. A noteworthy result here is that the class of invex functions is precisely the class of differentiable functions whose stationary points are global minimizers. We revisit some of the important results obtained by Hanson and Martin and extend them to constrained minimization problems with equality constraints in addition to inequality constraints. We address some conditions by which a function is invex. We propose a result to solve pseudo-invex programming problem with the help of an equivalent programming problem.

Keywords: Invex function; Invex set; KT-invex; WD-invex; Pseudo-invex function

1. Introduction

We say that f is said to be invex [1] if there exists a vector function $\eta : R^n \times R^n \to R^n$ such that for any $x, \bar{x} \in R^n$

$$f(x) - f(\bar{x}) \geq \eta^t(x, \bar{x}) \nabla f(\bar{x}) \tag{1.1}$$

where $f : R^n \to R$ is differentiable function and $\nabla f(\bar{x})$ is the gradient vector of f at $\bar{x}$. We call a set $A \subseteq R^n$ invex ([2], Definition 2.1) with respect to a given $\eta : R^n \times R^n \to R^n$ if

$$x, \bar{x} \in A,\ 0 \leq \lambda \leq 1 \Rightarrow \bar{x} + \lambda \eta(x, \bar{x}) \in A.$$

The importance of convex functions is well known in optimization problems. The property of convexity is invariant with respect to certain operations and transformations. However for many problems encountered in applications

* Corresponding author.
E-mail addresses: akdas@isical.ac.in (A.K. Das), dmrai23@gmail.com (R. Jana), rwitamjanaju@gmail.com (Deepmala).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.04.001

mainly economics, engineering etc., the notion of convexity does no longer suffice. To meet this demand and the convexity requirement to prove sufficient optimality conditions for differentiable mathematical programming problem, the notion of invexity was introduced by Hanson by substituting the linear term $(x - \bar{x})$. In this paper, we revisit a class of generalized convex functions introduced by Hanson [3] in 1981 and other related functions introduced later in [4–6] and [7]. Hanson [3] considers the following problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0,\ 1 \leq i \leq m \end{array} \tag{1.2}$$

where $f,\ g_i : R^n \to R$ are differentiable functions.

We say that a function f is said to be incave function ([8], page 63) if there exists a function $\eta : R^n \times R^n \to R^n$ such that for any $x, \bar{x} \in R^n$

$$f(x) - f(\bar{x}) \leq \eta^t(x, \bar{x}) \nabla f(\bar{x})$$

where $f : R^n \to R$ is differentiable function and $\nabla f(\bar{x})$ is the gradient vector of f at $\bar{x}$. The Karush–Kuhn–Tucker(KKT) conditions are said to be satisfied at an $\bar{x} \in S = \{x \mid g_i(x) \leq 0,\ 1 \leq i \leq m\}$ if there exists a vector $u \in R^m$

$$\left.\begin{array}{l} \nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) = 0 \\ u_i g_i(\bar{x}) = 0,\ 1 \leq i \leq m \\ u \geq 0 \end{array}\right\}. \tag{1.3}$$

For details, see [3,9] and [10]. Hanson proved the following theorem in [3].

Theorem 1.1. *Consider the problem in* (1.1)*. Suppose* f *and* g_i*'s* $1 \leq i \leq m$ *satisfy* (1.2) *with the same* $\eta : R^n \times R^n \to R^n$*. If the KKT conditions hold at an* $\bar{x} \in S$*, then* $\bar{x}$ *is a solution to the problem in* (1.1)*.*

The name invex (invariant convex) has been given to a function satisfying (1.1) with the function η by Craven [1]. Clearly, a differentiable convex function on an open convex subset A is invex with respect to $\eta(x, \bar{x}) = x - \bar{x}$. But the converse is not true. For invex functions every stationary points are global minimizers. Invexity is considered as a generalization of convexity. Invexity requires the differentiability assumption and all the needed functions to be invex with respect to the same η. It is also possible that a function can be invex with respect to $\eta(x, \bar{x}) = 0$ since if one can take $f : R^n \to R$ such that $f(x) = c\ \forall\ x \in R^n$, where c is a constant. Consider $f(x) = x^n\ \forall\ x \in R^n$, where n is odd and $n > 1$. In this case $f(x)$ is not invex function as $x = 0$ is a stationary point for the function, which is not minimum point.

Certain generalizations of invex functions, such as pseudo-invex and quasi-invex functions have been considered in [11] similar to pseudo-convex and quasi-convex functions. Recently, Noor et al. [12] introduced harmonic pre-invex functions which include harmonic functions as a special case. Ivanov [13] introduced the notion of a second-order KT invex with inequality constraints and established optimality conditions for global minimum. Giorgi and Guerraggio [14] considered generalized invexity of locally Lipschitz vector-valued functions to formulate sufficient Karush–Kuhn–Tucker conditions for a multiobjective programming problem. Galewski [15] showed that invex functions are convex transformable under some assumptions. Craven [16] showed that pseudo-invexity property of a vector function coincides with invexity in a restricted set of directions. Niezgoda and Peari [17] attempted to obtain an extension from convex functions to invex functions and from majorized vectors to separable vectors. Li and Zhang [18] considered multiple-objective semi-infinite programming involving these invex functions to obtain generalized saddle point necessary and sufficient conditions. Presently a large number of computational methods are reported in the literature for solving invex programming problems and other mathematical programming problems using invex functions. Ruiz-Garzón et al. [19] considered generalized invex monotone functions to characterize the solutions of the variational inequality problem and mathematical programming problem. Antczak [20] considered the exact penalty function method to establish optimality conditions in case of a constrained optimization problem involving continuously differentiable invex functions. Li et al. [21] proposed penalty function method for solving constrained non-smooth invex problems.

The paper is organized as follows. In Section 2, we show that Theorem 1.1 and the weak duality result go thorough for a minimization problem with equality constraints in addition to the inequality constraints. We show that

the characterization results proved by Martin [4] can be extended for a problem in which equality constraints are permitted. We prove a sufficient condition to show that a class of functions are invex. In Section 3, we propose an equivalent mathematical programming problem involving pseudo-invex functions for a class of mathematical programming problem. A numerical example for solving pseudo-invex programming problem is considered.

2. Invex and related functions

Hanson ([3], Theorem 2.1) considered the sufficient conditions for the minimizing problems with inequality constraints. We generalize this sufficient condition with equality constraints in addition to inequality constraints.

Theorem 2.1. *Let $f,\ g_i : R^n \to R$ for $1 \le i \le m$ and $h_j : R^n \to R$ for $1 \le j \le r$ be differentiable functions. Consider the following problem P:*

$$\begin{aligned} &\textit{minimize} && f(x) \\ &\textit{subject to} && \\ & && g_i(x) \le 0,\ 1 \le i \le m \\ & && h_j(x) = 0,\ 1 \le j \le r. \end{aligned}$$

Let $\bar{x}$ be a point in R^n such that

$$\bar{x} \in S = \{x \mid g_i(x) \le 0,\ 1 \le i \le m \text{ and } h_j(x) = 0,\ 1 \le j \le r\}.$$

If there exist $u \in R^m$ and $v \in R^r$ with

$$\left.\begin{aligned} &\nabla f(\bar{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) + \sum_{j \in J} v_j \nabla h_j(\bar{x}) + \sum_{j \in K} v_j \nabla h_j(\bar{x}) = 0 \\ &u_i g_i(\bar{x}) = 0,\ 1 \le i \le m \\ &u \ge 0 \end{aligned}\right\}. \tag{2.1}$$

Further suppose the following condition holds for $\bar{x}$ such that there exists a function $\eta : R^n \times R^n \to R^n$ such that $f,\ g_i$ for $1 \le i \le m$, and h_j are invex function for $j \in J$ and h_j are incave function for $j \in K$ with respect to η, where $J = \{j : v_j > 0\}$ and $K = \{j : v_j < 0\}$ and $J \cup K = \{1, 2, \ldots, r\}$. Then $\bar{x}$ is a solution to the problem P.

Proof. Suppose that there exist $u \in R^m$ and $v \in R^r$ such that the inequalities and (2.1) hold. Then note that for any $x \in S$ such that

$$g_i(x) \le 0,\ 1 \le i \le m \text{ and } h_j(x) = 0,\ 1 \le j \le r$$

Now as $f,\ g_i$ for $1 \le i \le m$ are invex function with respect to η, then

$$\begin{aligned} &f(x) - f(\bar{x}) \ge \eta^t(x, \bar{x}) \nabla f(\bar{x}), \\ &g_i(x) - g_i(\bar{x}) \ge \eta^t(x, \bar{x}) \nabla g_i(\bar{x}),\ 1 \le i \le m. \end{aligned}$$

Also h_j are invex for $j \in J$ and h_j are incave for $j \in K$ with respect to η, so we can write

$$\begin{aligned} &h_j(x) - h_j(\bar{x}) \ge \eta^t(x, \bar{x}) \nabla h_j(\bar{x}) \ for\ j \in J, \\ &-h_j(x) + h_j(\bar{x}) \ge -\eta^t(x, \bar{x}) \nabla h_j(\bar{x}) \ for\ j \in K. \end{aligned}$$

We have

$$\eta^t(x, \bar{x}) \nabla f(\bar{x}) = \sum_{i=1}^{m} u_i(-\eta^t(x, \bar{x}) \nabla g_i(\bar{x})) + \sum_{j \in J} v_j(-\eta^t(x, \bar{x}) \nabla h_j(\bar{x})) + \sum_{j \in K} v_j(-\eta^t(x, \bar{x}) \nabla h_j(\bar{x})).$$

Noting that

$$u_i(-\eta^t(x, \bar{x})) \nabla g_i(\bar{x}) \ \ge\ -u_i g_i(x) + u_i g_i(\bar{x}) \ge 0$$

and

$$v_j(-\eta^t(x,\bar{x})\nabla h_j(\bar{x})) \geq v_j\{-h_j(x)+h_j(\bar{x})\} = 0 \; for \; j \in J.$$

$$v_j(-\eta^t(x,\bar{x})\nabla h_j(\bar{x})) \geq v_j\{-h_j(x)+h_j(\bar{x})\} = 0 \; for \; j \in K.$$

Therefore we say that $\eta^t(x,\bar{x})\nabla f(\bar{x}) \geq 0$. Since f is invex function, we write

$$f(x) - f(\bar{x}) \geq \eta^t(x,\bar{x})\nabla f(\bar{x}) \geq 0.$$

Hence $f(x) \geq f(\bar{x})$. ■

Example 2.1. We consider the modified version of example given in [3] to illustrate our result.

$$\begin{array}{ll} \text{minimize} & x_1 - \sin x_2 \\ \text{subject to} & \sin x_1 - 4\,\sin x_2 \leq 0 \\ & 2\,\sin x_1 + 7\,\sin x_2 + x_1 \leq 6 \\ & 2\,x_1 + 2\,x_2 \leq 3 \\ & 4\,x_1^2 + 4\,x_2^2 \leq 9 \\ & \text{-}\sin x_2 \leq 0 \\ & \sin x_1 = 0 \end{array}$$

Note that f, g_i for $i = 1, 2, \ldots, 5$ and h_1 are invex with respect to $\eta(x,\bar{x}) = \begin{pmatrix} \frac{\sin x_1 - \sin \bar{x}_1}{\cos \bar{x}_1} \\ \frac{\sin x_2 - \sin \bar{x}_2}{\cos \bar{x}_2} \end{pmatrix}$ The KKT points of this problem is as follows:

$$u = [0, 1/7, 0, 0, 0]^t, \qquad v = [10/7] \qquad and \qquad x = \begin{pmatrix} 0 \\ sin^{-1}(\frac{6}{7}) \end{pmatrix} \approx \begin{pmatrix} 0 \\ 1.0296 \end{pmatrix}.$$

It is easy to verify that $u_i g_i(\bar{x}) = 0 \; \forall \; i$. By using our result we say that $\bar{x} = \begin{pmatrix} 0 \\ 1.0296 \end{pmatrix}$ is the solution of this problem.

The following theorem was observed by Martin [4], Ben-Israel and Mond [22]. It shows that the class of invex functions has an interesting characterization property.

Theorem 2.2.

$$\{f | f : R^n \to R, \; f \textit{is invex with respect to some } \eta : R^n \times R^n \to R^n\}$$

$$= \{f | f : R^n \to R, \; f \textit{ is differentiable, } \nabla f(x) = 0 \Rightarrow x \textit{ is a global minimizer }\}.$$

In other words, the class of invex functions is the same as the class of differentiable functions for which every stationary point is a global minimizer.

We call $\bar{x}$ a stationary point if $g_i(\bar{x}) \leq 0$, $1 \leq i \leq m$ and $h_j(\bar{x}) = 0$, $1 \leq j \leq r$ and there exist $u \in R^m$ and $v \in R^r$ satisfying the KKT conditions at $\bar{x}$ given by (2.1) with the assumptions. In order to consider this issue, Martin in his paper [4] introduced the notion of KT-invexity (Kuhn–Tucker invexity) for a programming problem. However, the problem considered by Hanson in [3] does not admit any equality constraint. In what follows, we consider the notion of KT-invexity for a problem of the type P with equality constraints in addition to inequality constraints and extend the theorem of Martin [4] for this problem.

Definition 2.1. Consider the problem P

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \\ & g_i(x) \leq 0, 1 \leq i \leq m \\ & h_j(x) = 0, 1 \leq j \leq r \end{array}$$

where f, g_i and $h_j : R^n \to R$ are also differentiable. Also h_j are invex for $j \in J$ and h_j are incave for $j \in K$ with respect to η where $J = \{j : v_j > 0\}$ and $K = \{j : v_j < 0\}$. We say that a problem P is KT-invex if there exists a

function $\eta : R^n \times R^n \to R^n$ such that for any $x, z \in R^n$

$$\left.\begin{array}{l} g_i(x) \le 0,\ 1 \le i \le m \\ g_i(z) \le 0,\ 1 \le i \le m \\ h_j(x) = h_j(z) = 0,\ 1 \le j \le r \end{array}\right\} \Rightarrow \begin{cases} f(x) - f(z) \ge \eta^t(x,z)\nabla f(z) \\ \text{for } i \in I(z) = \{i \mid g_i(z) = 0\} \\ -\eta^t(x,z)\nabla g_i(z) \ge 0,\ 1 \le i \le m \text{ and} \\ -\eta^t(x,z)\nabla h_j(z) \ge 0,\ \ j \in J \\ \eta^t(x,z)\nabla h_j(z) \ge 0,\ \ j \in K \end{cases} \tag{2.2}$$

where $J \cup K = \{1, 2, \ldots, r\}$.

We show the following result :

Theorem 2.3.

$$\{P \mid P \textit{ is a KT-invex }\} = \{\, P \mid \textit{Every stationary point is a global minimizer.}\}$$

Proof. If P is KT-invex then every stationary point is a global minimizer for P which follows from Theorem 2.1. Therefore to complete the proof, we need to show that if P is such that every stationary point is a global minimizer then P is KT-invex.

To show this, suppose x and z are two points such that

$$\left.\begin{array}{l} g_i(x) \le 0,\ 1 \le i \le m, h_j(x) = 0,\ 1 \le j \le r \\ g_i(z) \le 0,\ 1 \le i \le m, h_j(z) = 0,\ 1 \le j \le r \end{array}\right\} \tag{2.3}$$

Also suppose that $f(x) < f(z)$. Let $J = \{j : v_j > 0\}$ and $K = \{j : v_j < 0\}$. Then z is not a KKT point (i.e. a stationary point) by our hypothesis. Hence, the following system does not have a solution.

$$\sum_{i \in I(z)} u_i \nabla g_i(z) + \sum_{j \in J} v_j \nabla h_j(z) + \sum_{j \in K} v_j \nabla h_j(z) = -\nabla f(z), u_i \ge 0,\ i \in I(z),\ v_j \in R,\ 1 \le j \le r$$

where $I(z) = \{i \mid g_i(z) = 0\}$.

Therefore, by Farkas theorem of alternative $\exists$ a $y(z)$ such that

$$y^t(z)\nabla g_i(z) \ge 0, i \in I(z)$$

$$y^t(z)\nabla h_j(z) \ge 0,\ \ j \in J$$

$$y^t(z)(-\nabla h_j(z)) \ge 0,\ \ j \in K$$

$$y^t(z)\nabla f(z) > 0.$$

We now define

$$\eta(x,z) = [f(x) - f(z)][\nabla f(z) \cdot y^t(z)]^{-1} y^t(z).$$

If x and z satisfying (2.3) but $f(x) \ge f(z)$ then we set $\eta(x,z) = 0$. Also if either x or z does not satisfy (2.3), we set $\eta(x,z) = 0$. With $\eta : R^n \times R^n \to R^n$ as defined above, we see that P is KT-invex. ■

In the literature, we also come across the definition of pseudo-invexity and quasi-invexity for a differential function.

Definition 2.2 (*[5]*)**.** Let $f : R^n \to R$. We say that f is pseudo-invex with respect to η where $\eta : R^n \times R^n \to R^n$ if

$$\eta^t(x,\bar{x})\nabla f(\bar{x}) \ge 0 \Rightarrow f(x) \ge f(\bar{x}).$$

Remark 2.1. As noted by Pini [5], the class of pseudo-invex function is same as the class of invex function in view of Theorem 2.2.

Definition 2.3 (*[5]*)**.** Let $f : R^n \to R$ be differentiable. We say that f is quasi-invex with respect to η where $\eta : R^n \times R^n \to R^n$ if

$$f(x) \le f(\bar{x}) \Rightarrow \eta^t(x,\bar{x})\nabla f(\bar{x}) \le 0.$$

Remark 2.2. Definition of KT-invexity of a problem imposes the condition that g_i for $i = 1, 2, \ldots, m$ are quasi-invex at a feasible point z such that $g_i(z) = 0$ and h_j are both quasi-invex and quasi-incave at points which are feasible with respect to the same $\eta : R^n \times R^n \to R^n$ over a feasible set to which f is also invex at the feasible z. So when h_j for $j = 1, 2, \ldots, r$ are quasi-invex then $-\eta^t(x, z)\nabla h_j(z) \geq 0$ and when it is quasi-incave then $\eta^t(x, z)\nabla h_j(z) \geq 0$.

Remark 2.3. The implication of Theorem 2.3 is that a point $\bar{x}$ feasible to problem P which is also stationary in the sense of (2.1) is a global minimizer if and only if P is KT-invex. However, if we want to claim that every global minimizer of P is a stationary point of P in addition to a constraint qualification we may also require to modify the definition of a stationary point by allowing v to be unrestricted in (2.1). This means that to get a necessary and sufficient condition in addition to a constraint qualification, we may have to impose a stronger condition on h_j, i.e. h_j are invex for $j \in J$ and h_j are incave for $j \in K$, where $J = \{j : v_j > 0\}$ and $K = \{j : v_j < 0\}$ and $J \cup K = \{1, 2, \ldots, r\}$.

Now again consider the problem P,

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \\ & g_i(x) \leq 0, 1 \leq i \leq m \\ & h_j(x) = 0, 1 \leq j \leq r \end{array}$$

where f, g_i and $h_j : R^n \to R$ are also differentiable.

We now look at the following problem which is a formal dual to problem P. We call this the problem D.

$$\underset{(u,\mu,v)}{\text{maximize}} f(u) + \sum_{i=1}^{m} \mu_i g_i(u) + \sum_{j=1}^{r} v_j h_j(u)$$

subject to

$$\nabla f(u) + \sum_{i=1}^{m} \mu_i \nabla g_i(u) + \sum_{j=1}^{r} v_j \nabla h_j(u) = 0$$

$$\mu \geq 0.$$

We observe the following results.

Theorem 2.4. *Suppose $f : R^n \to R$, $g_i : R^n \to R$, $1 \leq i \leq m$ are invex with respect to $\eta : R^n \times R^n \to R^n$. Let $J = \{j : v_j > 0\}$ and $K = \{j : v_j < 0\}$. Suppose h_j are invex for $j \in J$ and h_j are incave for $j \in K$ with respect to same η. If x is feasible to P and (u, μ, v) is feasible to D then*

$$f(x) \geq f(u) + \sum_{i=1}^{m} \mu_i\, g_i(u) + \sum_{j \in J} v_j\, h_j(u) + \sum_{j \in K} v_j\, h_j(u)$$

Proof. As $f : R^n \to R$, $g_i : R^n \to R$, $1 \leq i \leq m$ are invex with respect to $\eta : R^n \times R^n \to R^n$ then

$$f(x) - f(u) \geq \eta^t(x, u)\nabla f(u)$$
$$g_i(x) - g_i(u) \geq \eta^t(x, u)\nabla g_i(u),\ 1 \leq i \leq m$$

Also h_j are invex for $j \in J$ and h_j are incave for $j \in K$ with respect to same η, where $J = \{j : v_j > 0\}$ and $K = \{j : v_j < 0\}$. Then we can write,

$$h_j(x) - h_j(u) \geq \eta^t(x, u)\nabla h_j(u) \text{ for } j \in J.$$

$$h_j(x) - h_j(u) \leq \eta^t(x, u)\nabla h_j(u) \text{ for } j \in K.$$

Now we have,

$$f(x) - f(u) - \eta^t(x,u)\nabla f(u) + \sum_{i=1}^{m} \mu_i[g_i(x) - g_i(u) - \eta^t(x,u)\nabla g_i(u)]$$
$$+ \sum_{j\in J} v_j[h_j(x) - h_j(u) - \eta^t(x,u)\nabla h_j(u)] + \sum_{j\in K} v_j[h_j(x) - h_j(u) - \eta^t(x,u)\nabla h_j(u)] \geq 0.$$

So,

$$f(x) - f(u) - [\nabla f(u) + \sum_{i=1}^{m} \mu_i \nabla g_i(u) + \sum_{j\in J} v_j \nabla h_j(u) + \sum_{j\in K} v_j \nabla h_j(u)]\eta^t(x,u)$$
$$+ \sum_{i=1}^{m} \mu_i g_i(x) - \sum_{i=1}^{m} \mu_i g_i(u) + \sum_{j\in J} v_j h_j(x) - \sum_{j\in J} v_j h_j(u) + \sum_{j\in K} v_j h_j(x) - \sum_{j\in K} v_j h_j(u) \geq 0.$$

Now we say,

$$[\nabla f(u) + \sum_{i=1}^{m} \mu_i \nabla g_i(u) + \sum_{j\in J} v_j \nabla h_j(u) + \sum_{j\in K} v_j \nabla h_j(u)]\eta^t(x,u) = 0,$$

$$g_i(x) \leq 0, \quad 1 \leq i \leq m,$$

$$h_j(x) = 0, \quad \forall \ j.$$

Hence,

$$f(x) \geq f(u) + \sum_{i=1}^{m} \mu_i g_i(u) + \sum_{j\in J} v_j h_j(u) + \sum_{j\in K} v_j h_j(u).$$

This completes the proof. ■

Definition 2.4. Problem P is said to be weak duality invex (WD-invex) if there exists a function $\eta : R^n \times R^n \to R^n$ such that for any $x, u \in R^n$

$$\left.\begin{array}{l} g_i(x) \leq 0, 1 \leq i \leq m \\ h_j(x) = 0, 1 \leq j \leq r \end{array}\right\} \Rightarrow \begin{cases} either \\ f(x) - f(u) \geq \eta^t(x,u)\nabla f(u), \\ -g_i(u) - \eta^t(x,u)\nabla g_i(u) \geq 0 \ for \ 1 \leq i \leq m \text{ and} \\ -h_j(u) - \eta^t(x,u)\nabla h_j(u) \geq 0 \ \ for \ j \in J, \\ h_j(u) + \eta^t(x,u)\nabla h_j(u) \geq 0 \ \ for \ j \in K. \\ or \\ -\eta^t(x,u)\nabla f(u) > 0, \\ -\eta^t(x,u)\nabla g_i(u) \geq 0, \\ -\eta^t(x,u)\nabla h_j(u) \geq 0 \ for \ j \ \in J, \\ \eta^t(x,u)\nabla h_j(u) \geq 0 \ for \ j \ \in K. \end{cases} \tag{2.4}$$

where $J = \{j : v_j > 0\}$ and $K = \{j : v_j < 0\}$ and $J \cup K = \{1, 2, \ldots, r\}$.

Theorem 2.5.

$$\{P \mid P \text{ is WD-invex}\} = \{P \mid \text{weak duality holds for } P\}$$

Proof. If P is WD invex then weak duality holds. For converse without loss of generality let us consider $v_j = -v'_j$ for $j \in K$ where $v'_j > 0$. Suppose weak duality holds, then for any $g_i(x) \leq 0, \quad 1 \leq i \leq m$ and $h_j(x) = 0$ for $1 \leq j \leq r$ the system

$$\nabla f(u) + \sum_{i=1}^{m} \mu_i \nabla g_i(u) + \sum_{j\in J} v_j \nabla h_j(u) + \sum_{j\in K} -v'_j \nabla h_j(u)$$

$\mu \geq 0$ and

$$f(x) < f(u) + \sum_{i=1}^{m} \mu_i g_i(u) + \sum_{j\in J} v_j h_j(u) + \sum_{j\in K} -v'_j h_j(u)$$

is inconsistent in μ and $v_j \in J$ or $v_j \in K$. Then the homogeneous system,

$$[\alpha : \beta]\begin{bmatrix} 0 & 1 \\ \nabla f(u) & f(x) - f(u) \end{bmatrix} + \sum_{i=1}^{m} \mu_i [\nabla g_i(u) \ \vdots \ -g_i(u)]$$

$$+ \sum_{j\in J} v_j [\nabla h_j(u) \ \vdots \ -h_j(u)] + \sum_{j\in K} v'_j [-\nabla h_j(u) \ \vdots \ h_j(u)] = 0$$

$[\alpha : \beta] > 0$ is inconsistent in $(\mu_i, v_j, \alpha, \beta)$. Then by Motzkin's theorem of alternative,

$$\begin{bmatrix} 0 & 1 \\ \nabla f(u) & f(x) - f(u) \end{bmatrix} \begin{bmatrix} \eta \\ \varepsilon \end{bmatrix} \leq 0$$

$$[\nabla g_i(u) \ \vdots \ -g_i(u)] \begin{bmatrix} \eta \\ \varepsilon \end{bmatrix} \leq 0$$

$$[\nabla h_j(u) \ \vdots \ -h_j(u)] \begin{bmatrix} \eta \\ \varepsilon \end{bmatrix} \leq 0$$

$$[-\nabla h_j(u) \ \vdots \ h_j(u)] \begin{bmatrix} \eta \\ \varepsilon \end{bmatrix} \leq 0$$

Consider $\varepsilon = -1$ in the above, we say

$$\eta^t(x, u)\nabla f(u) - f(x) + f(u) \leq 0$$

$$f(x) - f(u) \geq \eta^t(x, u)\nabla f(u)$$

$$-g_i(u) - \nabla g_i(u)\eta \geq 0, \quad \text{for } 1 \leq i \leq m$$

Again for $j \in J$

$$-h_j(u) - \nabla h_j(u)\eta \geq 0,$$

for $j \in K$

$$-h_j(u) - \eta\nabla h_j(u) \leq 0.$$

Again, if $\varepsilon = 0$, we say

$$\begin{aligned} &-\eta\nabla f(u) \geq 0 \\ &-\eta\nabla g_i(u) \geq 0, \ 1 \leq i \leq m \\ &-\eta\nabla h_j(u) \geq 0, \quad for\ j \in J \\ &\ \ \eta\nabla h_j(u) \geq 0, \quad for\ j \in K. \end{aligned}$$

This shows the existence of η and completes the proof. ■

Remark 2.4. We define the notion weak duality invexity (WD-invexity) for a problem P as Martin [4] did for his primal problem which did not admit any equalities. The definition is similar to that given by Martin [4] with the equality constraints handled as in Definition 2.2.

Now we state a sufficient condition to show that a class of functions are invex. In this context, we prove the following theorem to show some conditions under which a type of function is invex with respect to an appropriate η. The proof of this result shows the constructions of η.

Theorem 2.6. *Let $f : R^n \to R$ be differentiable function with respect to $\eta : R^n \times R^n \to R^n$, $b : A \times A \times [0, 1] \to R_+$ on a invex set $A \subset R^n$ with respect to η such that*

$$f(y + \lambda\eta(x, y)) \leq b(x, y, \lambda)f(x) + (1 - b(x, y, \lambda))f(y), \quad x, y \in A, \ \lambda \in [0, 1].$$

Further suppose b is continuous at $\lambda = 0$. Then f is invex with respect to some $\eta_1 : R^n \times R^n \to R^n$.

Proof. Note that we have

$$f(y + \lambda\eta(x, y)) \leq b(x, y, \lambda)f(x) + (1 - b(x, y, \lambda))f(y), \quad x, y \in A, \ \lambda \in [0, 1].$$

Using the expansion of f at y, we say that

$$f(y) + \lambda\eta^t(x, y)\nabla f(y) + \lambda\|\eta(x, y)\|\beta(y, \lambda\eta(x, y)) \leq f(y) + b(x, y, \lambda)[f(x) - f(y)]$$

$$\Longrightarrow b(x, y, \lambda)[f(x) - f(y)] \geq \lambda\eta^t(x, y)\nabla f(y) + \lambda\|\eta(x, y)\|\beta(y, \lambda\eta(x, y)).$$

Taking the limit $\lambda \to 0^+$ we get,

$$[f(x) - f(y)] \geq \frac{\eta^t(x, y)}{k(x, y)}\nabla f(y)$$

where $k(x, y) = \limsup_{\lambda \to 0^+} b(x, y, \lambda)$ and $\beta(y, \lambda\eta(x, y)) \to 0$ as $\lambda \to 0^+$. Let $\eta_1^t(x, y) = \frac{\eta^t(x,y)}{k(x,y)}$. Consequently f is invex with respect to $\eta_1 : R^n \times R^n \to R^n$. ■

Example 2.2. We consider the example given in [2] to illustrate the theorem given above. Let us consider $A = (0, \pi/2)$, and $f : A \to R$ such that $f(x) = sin(x)$ and

$$\eta(x, \bar{x}) = \begin{cases} \dfrac{sin(x) - sin(\bar{x})}{cos(\bar{x})}, & \text{if } x \geq \bar{x} \\ 0, & \text{if } x \leq \bar{x}. \end{cases}$$

Now $f(x)$ satisfies

$$f(\bar{x} + \lambda\eta(x, \bar{x})) \leq b(x, \bar{x}, \lambda)f(x) + (1 - b(x, \bar{x}, \lambda))f(\bar{x}), \quad x, \bar{x} \in A, \ \lambda \in [0, 1],$$

where for $0 < \lambda \leq 1$,

$$b(x, \bar{x}, \lambda) = \begin{cases} 1 - \lambda, & \text{if } x \geq \bar{x} \\ 1, & \text{if } x < \bar{x}. \end{cases}$$

If $\lambda = 0$, then $b(x, \bar{x}, \lambda) = 1$. So $f(x)$ satisfies all the condition of Theorem 2.6. Hence $f(x)$ is invex function.

3. Pseudo-invex function

If f is a pseudo-invex function then for a class of nonlinear programs, we can construct another nonlinear program for which optimal solutions of two nonlinear programs are same. The following theorem establishes a link between two nonlinear programs involving f and η. This is a generalization of the results by Kortanek and Evans [23].

Theorem 3.1. *Let $f : R^n \to R$ be a differentiable pseudo-invex function with respect to $\eta : R^n \times R^n \to R^n$ on an invex set $S \subseteq R^n$. Suppose $x^* \in S$ and assume that $\eta(x, x^*) = \bar{\eta}(x, x^*) - \bar{\eta}(x^*, x^*)$, $x \in S$, $\bar{\eta} : R^n \times R^n \to R^n$. Consider the following two programs:*

P_1. *minimize* $f(x)$ *subject to* $x \in S$

P_2. *minimize* $\bar{\eta}^t(x, x^*)\nabla f(x^*)$ *subject to* $x \in S$.

Then x^* is an optimal solution of P_1 if and only if x^* is an optimal solution for P_2.

Proof. x^* is an optimal solution of P_1. Then for any $x \in S$, we have

$$f(x^* + \lambda\eta(x, x^*)) - f(x^*) \geq 0.$$
$$\Rightarrow \frac{f(x^* + \lambda\eta(x, x^*)) - f(x^*)}{\lambda} \geq 0$$

Taking the limit $\lambda \to 0^+$ we get,

$$\Rightarrow \eta^t(x, x^*)\nabla f(x^*) \geq 0.$$

According to Hanson [3] and Hanson and Mond [24], we have $\eta(x, x^*) = \bar{\eta}(x, x^*) - \bar{\eta}(x^*, x^*)$. Therefore, $\bar{\eta}^t(x, x^*)\nabla f(x^*) - \bar{\eta}^t(x^*, x^*)\nabla f(x^*) \geq 0$, which implies x^* is an optimal solution for P_2.

Conversely, let x^* be an optimal solution for P_2. Therefore

$$\bar{\eta}^t(x, x^*)\nabla f(x^*) \geq \bar{\eta}^t(x^*, x^*)\nabla f(x^*)$$
$$\Rightarrow \eta^t(x, x^*)\nabla f(x^*) \geq 0$$
$$\Rightarrow f(x) \geq f(x^*) \text{ by pseudo-invexity of } f$$
$$\Rightarrow x^* \text{ is an optimal solution for } P_1. \quad \blacksquare$$

Theorem 3.1 is useful to find the solutions of problem P_1 by using the proposed alternative method. Note that problem P_2 is an equivalent mathematical programming problem of P_1. Hence to solve P_1 one can either solve P_1 directly or solve the proposed equivalent method P_2. We consider the example given in Example 2.1 with $\eta(x, \bar{x}) = \begin{pmatrix} \frac{\sin x_1 - sin\bar{x}_1}{\cos \bar{x}_1} \\ \frac{\sin x_2 - sin\bar{x}_2}{\cos \bar{x}_2} \end{pmatrix}$ for numerical illustrations. To solve this problem we consider problem P_2. We start the iteration with $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$. We find $\bar{x} = \begin{pmatrix} 0 \\ 1.0296 \end{pmatrix}$ as the solution of the problem P_2. By using our result we say that $\bar{x} = \begin{pmatrix} 0 \\ 1.0296 \end{pmatrix}$ is the solution of problem P_1.

Acknowledgments

The second author R. Jana is thankful to the Department of Science and Technology, Govt. of India, INSPIRE Fellowship Scheme for financial support. The research work of the third author Deepmala is supported by the Science and Engineering Research Board (SERB), Government of India under SERB N-PDF scheme, File Number: PDF/2015/000799.

References

[1] B.D. Craven, Invex functions and constrained local minima, Bull. Aust. Math. Soc. 24 (03) (1981) 357–366.
[2] S.K. Suneja, C. Singh, C.R. Bector, Generalization of preinvex and B-vex functions, J. Optim. Theory Appl. 76 (3) (1993) 577–587.
[3] M.A. Hanson, On sufficiency of the Kuhn–Tucker conditions, J. Math. Anal. Appl. 80 (2) (1981) 545–550.
[4] D.H. Martin, The essence of invexity, J. Optim. Theory Appl. 47 (1) (1985) 65–76.
[5] R. Pini, Invexity and generalized convexity, Optimization 22 (4) (1991) 513–525.
[6] V. Jeyakumar, Strong and Weak Invexity in Mathematical Programming, University of Melbourne, Department of Mathematics, 1984.
[7] T. Weir, B. Mond, Pre-invex functions in multiple objective optimization, J. Math. Anal. Appl. 136 (1) (1988) 29–38.
[8] Ramík. Jaroslav, Milan. Vlach, Generalized Concavity in Fuzzy Optimization & Decision Analysis, Vol. 41, Springer Science and Business Media, 2012.
[9] M.S. Bazaraa, H.D. Sherali, C.M. Shetty, Nonlinear Programming: Theory and Algorithms, John Wiley & Sons, 2013.
[10] H.W. Kuhn, A.W. Tucker, Nonlinear programming, in: J. Neyman (Ed.), Proceedings 2nd Berkley Symposium on Mathematical Statistics and Probablity, University of California Press, 1951.
[11] O.L. Mangasarian, Pseudo-convex functions, J. Soc. Ind. Appl. Math., A 3 (2) (1965) 281–290.
[12] M.A. Noor, K.I. Noor, S. Iftikhar, Hermite-Hadamard inequalities for harmonic preinvex functions, Saussurea 6 (2) (2016) 34–53.
[13] V.I. Ivanov, Second-order Kuhn–Tucker invex constrained problems, J. Global Optim. 50 (3) (2011) 519–529.
[14] G. Giorgi, A. Guerraggio, Nonsmooth vector-valued invex functions and applications, J. Inf. Optim. Sci. 21 (2) (2000) 243–255.
[15] M. Galewski, A note on invex problems with nonnegative variable, European J. Oper. Res. 163 (2) (2005) 565–568.
[16] B.D. Craven, Characterizing invex and related properties. Generalized convexity, generalized monotonicity and applications, Nonconvex Optim. Appl. 77 (2005) 183–191.
[17] M. Niezgoda, J. Peari, Hardy-Littlewood-Pòlya-type theorems for invex functions, Comput. Math. Appl. 64 (4) (2012) 518–526.
[18] X.Y. Li, Q.X. Zhang, Saddle point condition for multiple-objective semi-infinite programming with generalized invex functions, Commun. Appl. Math. Comput. 27 (4) (2013) 501–507.

[19] G. Ruiz-Garzón, R. Osuna-Gómez, A. Rufián-Lizana, Generalized invex monotonicity, European J. Oper. Res. 144 (3) (2003) 501–512.
[20] T. Antczak, Exact penalty functions method for mathematical programming problems involving invex functions, European J. Oper. Res. 198 (1) (2009) 29–36.
[21] G. Li, Z. Yan, J. Wang, A one-layer recurrent neural network for constrained nonsmooth invex optimization, Neural Netw. 50 (2014) 79–89.
[22] A. Ben-Israel, B. Mond, What is invexity, J. Aust. Math. Soc. Ser. B 28 (1) (1986) 1–9; Math. Soc., B 28 (1986) 1–9.
[23] K.O. Kortanek, J.P. Evans, Pseudo-concave programming and lagrange regularity, Oper. Res. 15 (5) (1967) 882–891.
[24] M.A. Hanson, B. Mond, Convex transformable programming problems and invexity, J. Inf. Optim. Sci. 8 (2) (1987) 201–207.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 372–377
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original Article

A new collection which contains the topology via ideals☆

Erdal Ekici

Department of Mathematics, Canakkale Onsekiz Mart University, Terzioglu Campus, 17020 Canakkale, Turkey

Received 15 July 2017; received in revised form 15 November 2017; accepted 21 November 2017
Available online 12 December 2017

Abstract

In 1999, Dontchev studied the concept of pre-I-open sets (Dontchev, 1999). In 2002, Hatir and Noiri studied the concept of semi-I-open sets (Hatir and Noiri, 2002). In 2013, Ekici studied the concept of $\mathcal{C}_I^*$-sets (Ekici, 2013). In this paper, the concept of ζ-I-open sets is introduced. The concept of ζ-I-open sets is related to topology, pre-I-open sets, semi-I-open sets and $\mathcal{C}_I^*$-sets. Main properties of ζ-I-open sets are presented.

Keywords: $\mathcal{C}_I^*$-set; Pre-I-open; Semi-I-open; Topology; Strongly β-I-open; Semi*-I-open; ζ-I-open; ζ-I-closed; $\star$-nowhere dense; $\star$-hyperconnected; Pre-I-regular

1. Introduction

Many collections of the generalized open and closed sets in the literature have been studied for various point-set topological problems (for example [1–7]). After the advent of the concept of ideals, several research papers appeared in the context of ideal topological spaces. Various new problems in topology have been introduced via ideal, for example [8–13]. In 1999, Dontchev studied the concept of pre-I-open sets [8]. In 2002, Hatir and Noiri studied the concept of semi-I-open sets [14]. In 2013, Ekici studied the concept of $\mathcal{C}_I^*$-sets [15]. The main goal of the present paper is to introduce the concept of ζ-I-open sets. The concept of ζ-I-open sets is related to topology, pre-I-open sets, semi-I-open sets and $\mathcal{C}_I^*$-sets. Main properties of ζ-I-open sets are discussed.

In this paper, a topological space will be denoted by (X, ϑ). For a subset S of X, the closure and the interior of S will be denoted by $\hat{\mathbf{c}}(S)$ and $\hat{\imath}(S)$, respectively.

An ideal $\mathcal{L}$ is a nonempty collection of sets in a topological space X such that $\mathcal{L}$ is closed under heredity and finite additivity [16], i.e. $\mathcal{L}$ satisfies the following properties [16]:

☆ This work was supported by Çanakkale Onsekiz Mart University, The Scientific Research Coordination Unit, Project number: FHD-2017-1171.
E-mail address: eekici@comu.edu.tr.
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2017.11.001

(a) $T \in \mathcal{L}$ and $S \subseteq T$ implies $S \in \mathcal{L}$,
(b) $S \in \mathcal{L}$ and $T \in \mathcal{L}$ implies $S \cup T \in \mathcal{L}$.

For a topological space (X, ϑ) with an ideal $\mathcal{L}$, $(.)^\star : P(X) \longrightarrow P(X)$, $S^\star = \{a \in X \mid S \cap T \notin \mathcal{L}$ for every $T \in \vartheta$ with $a \in T\}$ is called the local function of S with respect to $\mathcal{L}$ and ϑ [16].

$\hat{\mathbf{c}}^\star(S) = S \cup S^\star$ defines a Kuratowski closure operator for a topology $\vartheta^\star$ and it will be called $\star$-topology [13].

Definition 1. Let S be a subset of a topological space (X, ϑ) with an ideal $\mathcal{L}$. S is said to be
(a) strongly β-I-open [17] if $S \subseteq \hat{\mathbf{c}}^\star(\hat{\imath}(\hat{\mathbf{c}}^\star(S)))$.
(b) semi-I-open [14] if $S \subseteq \hat{\mathbf{c}}^\star(\hat{\imath}(S))$.
(c) pre-I-open [8] if $S \subseteq \hat{\imath}(\hat{\mathbf{c}}^\star(S))$.
(d) pre-I-closed [8] if $X \setminus S$ is pre-I-open.

Definition 2. Let S be a subset of a topological space (X, ϑ) with an ideal $\mathcal{L}$. S is said to be
(a) semi*-I-open [18] if $S \subseteq \hat{\mathbf{c}}(\hat{\imath}^\star(S))$.
(b) semi*-I-closed [18] if $X \setminus S$ is semi*-I-open.

2. A new collection which contains the topology: ζ-I-open sets

In this section, the collection of ζ-I-open sets in a topological space (X, ϑ) with an ideal $\mathcal{L}$ is introduced. Main properties for the collection of ζ-I-open sets are discussed.

Definition 3. Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. Then S is called ζ-I-open if $S \in \{T \neq \varnothing \mid$ There exists a nonempty pre-I-open subset V such that $V \setminus \hat{\mathbf{c}}^\star(T) \in \mathcal{L}\} \cup \{\varnothing\} \subseteq P(X)$.

Theorem 4. *Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. If S is a strongly β-I-open subset of X, then S is a ζ-I-open subset of X.*

Proof. Let S be a strongly β-I-open set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. We have $S = \varnothing$ or $S \neq \varnothing$. Let S be a nonempty subset of X. Since S is strongly β-I-open, then

$$S \subseteq \hat{\mathbf{c}}^\star(\hat{\imath}(\hat{\mathbf{c}}^\star(S))).$$

By Theorem 4.1 of [19], there exists a pre-I-open subset $U = \hat{\imath}(\hat{\mathbf{c}}^\star(S))$ such that $U \subseteq \hat{\mathbf{c}}^\star(S)$. Also, $U = \hat{\imath}(\hat{\mathbf{c}}^\star(S))$ is a nonempty subset of X. We have $U \setminus \hat{\mathbf{c}}^\star(S) \in \mathcal{L}$. Hence, S is a ζ-I-open subset of X. ■

Definition 5. ([15]) Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. Then S is said to be
(a) pre-I-regular if S is pre-I-open and pre-I-closed.
(b) a $\mathcal{C}_I^*$-set if $S = T \cap V$ where T is an open set and V is a pre-I-regular set.

Corollary 6. *Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. If S is a $\mathcal{C}_I^*$-set, then S is a ζ-I-open subset of X. So, every pre-I-regular subset of X is a ζ-I-open subset of X. Also, every open subset of X is a ζ-I-open subset of X.*

Proof. Let S be a $\mathcal{C}_I^*$-set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. By Theorem 12 of [15], each $\mathcal{C}_I^*$-set is a pre-I-open subset of X. So, S is pre-I-open. By Theorem 4, S is a ζ-I-open subset of X.

Let S be a pre-I-regular subset of X. By Remark 15 of [15], each pre-I-regular subset of X is a $\mathcal{C}_I^*$-set. So, S is a ζ-I-open subset of X.

Let S be an open subset of X. By Remark 15 of [15], every open subset of X is a $\mathcal{C}_I^*$-set. So, S is a ζ-I-open subset of X. ■

Remark 7. Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. The following implications hold by Theorem 4 and Corollary 6:

$$
\begin{array}{ccc}
 & & \zeta\text{-}I\text{-open} \\
 & & \uparrow \\
\text{semi-}I\text{-open} & \longrightarrow & \text{strongly } \beta\text{-}I\text{-open} \\
 & & \uparrow \\
\uparrow & & \text{pre-}I\text{-open} \\
 & & \uparrow \\
\text{open} & \longrightarrow & \mathcal{C}_I^*\text{-set} \\
 & & \uparrow \\
 & & \text{pre-}I\text{-regular.}
\end{array}
$$

Remark 8. The reverses of the implications in the above diagram are not true in general as shown in the below example and in [15] and [17].

Example 9. Take $X = \{s, t, u, v\}$ with a topology $\vartheta = \{\{s\}, \{s, t\}, \{u, v\}, \{s, u, v\}, \varnothing, X\}$ and the ideal $\mathcal{L} = \{\{s\}, \{v\}, \{s, v\}, \varnothing\}$. Put $S = \{v\} \subseteq X$. Then S is a ζ-I-open subset of X and S is not a strongly β-I-open subset of X.

Definition 10 (*[20]*)**.** Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. Then S is said to be $\star$-dense if $\hat{\mathbf{c}}^{\star}(S) = X$.

Theorem 11. *Let (X, ϑ) be a topological space with an ideal $\mathcal{L}$. Each $\star$-dense subset of X is ζ-I-open.*

Proof. Let S be a $\star$-dense subset of X. Since $\hat{\mathbf{c}}^{\star}(S) = X$, S is strongly β-I-open. By Theorem 4, S is ζ-I-open. ■

Definition 12 (*[18]*)**.** Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. Then S is called $\star$-nowhere dense if $\hat{\imath}(\hat{\mathbf{c}}^{\star}(S)) = \varnothing$.

Theorem 13. *Let $S \neq \varnothing$ be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. Suppose that S is not a $\star$-nowhere dense set. Then S is a ζ-I-open set in X.*

Proof. For a set $S \neq \varnothing$ in X, suppose that S is not a $\star$-nowhere dense set. This implies $\hat{\imath}(\hat{\mathbf{c}}^{\star}(S)) \neq \varnothing$. Also,

$$U = \hat{\imath}(\hat{\mathbf{c}}^{\star}(S)) \neq \varnothing$$

is a pre-I-open subset of X. Furthermore, $U \setminus \hat{\mathbf{c}}^{\star}(S) \in \mathcal{L}$. Consequently, S is a ζ-I-open subset of X. ■

Remark 14. The following example shows that we have a nonempty, ζ-I-open and $\star$-nowhere dense set.

Example 15. Take $X = \{s, t, u, v\}$ with a topology $\vartheta = \{X, \{s\}, \{t, u\}, \{s, t, u\}, \varnothing\}$ and an ideal $\mathcal{L} = \{\varnothing, \{s\}, \{v\}, \{s, v\}\}$. Put $S = \{v\}$. Then S is ζ-I-open $\star$-nowhere dense in X.

Definition 16. Let (X, ϑ) and (Y, γ) be topological spaces with the ideals $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively. A function $h : X \longrightarrow Y$ is called pre-I-open if the image of each pre-I-open subset of X is pre-I-open in Y.

Theorem 17. *Let (X, ϑ) and (Y, γ) be topological spaces with the ideals $\mathcal{L}$ and $h(\mathcal{L})$, respectively, where $h : X \longrightarrow Y$ is a bijective, pre-I-open and $\star$-continuous function and $h(\mathcal{L}) = \{h(I) \mid I \in \mathcal{L}\}$. Then $h(S)$ is ζ-I-open for each ζ-I-open subset S of X.*

Proof. Suppose that $h : X \longrightarrow Y$ is bijective, pre-I-open, $\star$-continuous. Let S be ζ-I-open in X. Then $S \neq \varnothing$ or $S = \varnothing$. Take $S \neq \varnothing$. Since S be ζ-I-open in X, then there exists a nonempty pre-I-open set U such that $U \setminus \hat{\mathbf{c}}^{\star}(S) \in \mathcal{L}$. Since $h : X \longrightarrow Y$ is bijective, pre-I-open, $\star$-continuous, we have

$$
\begin{aligned}
& h(U) \setminus h(\hat{\mathbf{c}}^{\star}(S)) \\
&= h(U) \cap (X \setminus h(\hat{\mathbf{c}}^{\star}(S))) \in h(\mathcal{L})
\end{aligned}
$$

and this implies

$$h(U) \setminus \hat{\mathbf{c}}^{\star}(h(S)) \in h(\mathcal{L}).$$

Consequently, $h(S)$ is a ζ-I-open subset of Y. ■

Theorem 18. *Let (X, ϑ) be a topological space with an ideal $\mathcal{L}$ on X and $S \subseteq X$. Then S is a ζ-I-open set in X if and only if $S = \varnothing$ or there exist a set U in $\mathcal{L}$ and a nonempty pre-I-open set N such that $N \setminus U \subseteq \hat{\mathbf{c}}^{\star}(S)$.*

Proof. ($\Longrightarrow$): Let S be a ζ-I-open set in X. Then $S = \varnothing$ or $S \neq \varnothing$. Suppose $S \neq \varnothing$. There exists a nonempty pre-I-open set N such that $N \setminus \hat{\mathbf{c}}^{\star}(S) \in \mathcal{L}$. Put $U = N \setminus \hat{\mathbf{c}}^{\star}(S)$. This implies

$$N \setminus U \subseteq \hat{\mathbf{c}}^{\star}(S).$$

($\Longleftarrow$): Suppose that $S = \varnothing$ or there exist a set U in $\mathcal{L}$ and a nonempty pre-I-open set N such that $N \setminus U \subseteq \hat{\mathbf{c}}^{\star}(S)$. If $S = \varnothing$, then S is a ζ-I-open set in X. Assume that there exist a set U in $\mathcal{L}$ and a nonempty pre-I-open set N such that

$$N \setminus U \subseteq \hat{\mathbf{c}}^{\star}(S).$$

We have $N \setminus \hat{\mathbf{c}}^{\star}(S) \subseteq U$. This implies $N \setminus \hat{\mathbf{c}}^{\star}(S) \in \mathcal{L}$. Consequently, S is a ζ-I-open set in X. ■

Theorem 19. *Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. Then S is a ζ-I-open subset of X if and only if $S \in \{\varnothing \neq T \mid$ there exist a nonempty pre-I-open subset U and an element V of $\mathcal{L}$ such that $U \subseteq \hat{\mathbf{c}}^{\star}(T) \cup V\} \cup \{\varnothing\} \subseteq P(X)$.*

Proof. It follows by Theorem 18. ■

3. ζ-I-closed sets and other properties

In this section, the collection of ζ-I-closed sets in a topological space (X, ϑ) with an ideal $\mathcal{L}$ is introduced. Properties for the collection of ζ-I-closed sets and other properties for the collection of ζ-I-open sets are studied.

Definition 20. Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. Then S is called ζ-I-closed set if $X \setminus S$ is a ζ-I-open subset of X.

Theorem 21. *Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. Then S is ζ-I-closed if and only if there exist $U \in \mathcal{L}$ and a pre-I-closed subset $V \neq X$ such that $\hat{\imath}^{\star}(S) \setminus U \subseteq V$ or $S = X$.*

Proof. ($\Longrightarrow$): Let S be a ζ-I-closed subset of X. This implies $X \setminus S$ is ζ-I-open. Since $X \setminus S$ is ζ-I-open, $X \setminus S = \varnothing$ or there exists a nonempty pre-I-open subset N such that $N \setminus \hat{\mathbf{c}}^{\star}(X \setminus S) \in \mathcal{L}$. This implies that $S = X$ or there exists a nonempty pre-I-open set N such that $N \setminus \hat{\mathbf{c}}^{\star}(X \setminus S) \in \mathcal{L}$. Take

$$U = N \setminus \hat{\mathbf{c}}^{\star}(X \setminus S).$$

We have $N \subseteq (X \setminus \hat{\imath}^{\star}(S)) \cup U$. It follows that

$$\begin{aligned} &X \setminus (X \setminus \hat{\imath}^{\star}(S)) \cap (X \setminus U) \\ &\subseteq X \setminus N. \end{aligned}$$

Since N is a nonempty pre-I-open subset, then $X \setminus N$ is a pre-I-closed subset and $X \setminus N \neq X$.
Also, we have

$$\begin{aligned} &\hat{\imath}^{\star}(S) \cap (X \setminus U) \\ &\subseteq X \setminus N. \end{aligned}$$

Thus, $\hat{\imath}^{\star}(S) \setminus U \subseteq X \setminus N$.
($\Longleftarrow$): Let S be a subset of X. If $S = X$, then $X \setminus S = \varnothing$. This implies that $X \setminus S$ is ζ-I-open and so S is ζ-I-closed. Suppose that there exist $U \in \mathcal{L}$ and a pre-I-closed subset $V \neq X$ such that $\hat{\imath}^{\star}(S) \setminus U \subseteq V$.

We have

$$\hat{\imath}^{\star}(S) \cap (X \setminus U) \subseteq V$$

and

$$X \setminus V \subseteq (X \setminus \hat{\imath}^{\star}(S)) \cup U.$$

Since V is a pre-I-closed subset and $V \neq X$, then $X \setminus V$ is a nonempty pre-I-open subset of X. Since $X \setminus V \subseteq (X \setminus \hat{\imath}^{\star}(S)) \cup U$, then

$$X \setminus V \subseteq \hat{\mathbf{c}}^{\star}(X \setminus S) \cup U.$$

It follows by Theorem 19 that $X \setminus S$ is a ζ-I-open subset of X. Consequently, S is ζ-I-closed. ■

Theorem 22. *Let S be a set in a topological space (X, ϑ) with an ideal $\mathcal{L}$. Then S is ζ-I-closed if and only if there exists a pre-I-closed subset $V \neq X$ such that $\hat{\imath}^{\star}(S) \setminus V \in \mathcal{L}$ or $S = X$.*

Proof. ($\Longrightarrow$): Let S be a ζ-I-closed subset of X. By Theorem 21, there exist $U \in \mathcal{L}$ and a pre-I-closed subset $V \neq X$ such that

$$\hat{\imath}^{\star}(S) \setminus U \subseteq V \text{ or } S = X.$$

This implies $\hat{\imath}^{\star}(S) \setminus V \subseteq U$. Since $U \in \mathcal{L}$, then $\hat{\imath}^{\star}(S) \setminus V \in \mathcal{L}$.

($\Longleftarrow$): Let $V \neq X$ be a pre-I-closed subset with $\hat{\imath}^{\star}(S) \setminus V \in \mathcal{L}$. We have $\hat{\imath}^{\star}(S) \setminus V \in \mathcal{L}$ and also $\hat{\imath}^{\star}(S) \setminus (\hat{\imath}^{\star}(S) \setminus V) \subseteq V$. Thus, by Theorem 21, S is a ζ-I-closed subset of X. ■

Theorem 23. *Let (X, ϑ) be a topological space with an ideal $\mathcal{L}$. For ζ-I-open subsets S_α in X for each $\alpha \in I$, $\cup\{S_\alpha \mid \alpha \in I\}$ is a ζ-I-open subset of X .*

Proof. Let $\cup\{S_\alpha \mid \alpha \in I\} = \varnothing$. Then $\cup\{S_\alpha \mid \alpha \in I\}$ is a ζ-I-open subset of X.

Let $\cup\{S_\alpha \mid \alpha \in I\} \neq \varnothing$. This implies that there exists an $\alpha_i \in I$ such that $S_{\alpha_i} \neq \varnothing$. Since $S_{\alpha_i} \neq \varnothing$ is a ζ-I-open subset, then there exists a nonempty pre-I-open subset U such that $U \setminus \hat{\mathbf{c}}^{\star}(S_{\alpha_i}) \in \mathcal{L}$. Since

$$\hat{\mathbf{c}}^{\star}(S_{\alpha_i}) \subseteq \hat{\mathbf{c}}^{\star}(\cup\{S_\alpha \mid \alpha \in I\}),$$

we have

$$U \setminus \hat{\mathbf{c}}^{\star}(\cup\{S_\alpha \mid \alpha \in I\}) \subseteq U \setminus \hat{\mathbf{c}}^{\star}(S_{\alpha_i}).$$

Since $U \setminus \hat{\mathbf{c}}^{\star}(S_{\alpha_i}) \in \mathcal{L}$, then $U \setminus \hat{\mathbf{c}}^{\star}(\cup\{S_\alpha \mid \alpha \in I\}) \in \mathcal{L}$. Thus, $\cup\{S_\alpha \mid \alpha \in I\}$ is ζ-I-open. ■

Theorem 24. *Let (X, ϑ) be a topological space with an ideal $\mathcal{L}$. Assume that every $\star$-open subset of X is pre-I-closed. Then each subset of X is ζ-I-open.*

Proof. Let $S \neq \varnothing$ be a subset of X. Since every $\star$-open subset of X is pre-I-closed, then $\hat{\mathbf{c}}^{\star}(S)$ is pre-I-open. Furthermore, $\hat{\mathbf{c}}^{\star}(S) \neq \varnothing$ and $\hat{\mathbf{c}}^{\star}(S) \setminus \hat{\mathbf{c}}^{\star}(S) \in \mathcal{L}$. Consequently, S is ζ-I-open. ■

Remark 25. The following example shows that there are two ζ-I-open sets but the intersection of these two sets is not ζ-I-open for any topological space (X, ϑ) with an ideal $\mathcal{L}$.

Example 26. Take $X = \{s, t, u, v\}$ with a topology $\vartheta = \{X, \{s\}, \{u\}, \{s, t\}, \{s, u\}, \{s, t, u\}, \{s, u, v\}, \varnothing\}$ and an ideal $\mathcal{L} = \{\varnothing, \{t\}\}$. Put $S = \{s, t\}$ and $T = \{t, u\}$. Then S and T are ζ-I-open subsets but $S \cap T$ is not a ζ-I-open subset of X.

Theorem 27. *Let (X, ϑ) be a topological space with an ideal $\mathcal{L}$. Assume that the ideal has a pre-I-open element $\{a\}$. Then each subset of X is ζ-I-open.*

Proof. Let $\varnothing \neq S \subseteq X$ and $\{a\} \in \mathcal{L}$ be pre-I-open. Then

$$\{a\} \setminus \hat{\mathbf{c}}^{\star}(S) = \{a\}$$

or

$$\{a\} \setminus \hat{\mathbf{c}}^{\star}(S) = \varnothing.$$

Since $\{a\} \in \mathcal{L}$ and $\{a\}$ is a pre-I-open subset of X, we have $\{a\} = \{a\} \setminus \hat{\mathbf{c}}^{\star}(S) \in \mathcal{L}$ or $\varnothing = \{a\} \setminus \hat{\mathbf{c}}^{\star}(S) \in \mathcal{L}$. Thus, S is ζ-I-open. ■

Theorem 28. *Let (X, ϑ) be a topological space with an ideal $\mathcal{L}$ and $\varnothing \neq S \subseteq T \subseteq X$ and S be ζ-I-open. Then T is ζ-I-open.*

Proof. Let $\varnothing \neq S \subseteq T \subseteq X$ and S be ζ-I-open. This implies there exists a nonempty pre-I-open subset U such that $U \setminus \hat{\mathbf{c}}^{\star}(S) \in \mathcal{L}$. Since $\varnothing \neq S \subseteq T \subseteq X$, we have

$$U \setminus \hat{\mathbf{c}}^{\star}(T) \subseteq U \setminus \hat{\mathbf{c}}^{\star}(S).$$

Since $U \setminus \hat{\mathbf{c}}^{\star}(S) \in \mathcal{L}$, then $U \setminus \hat{\mathbf{c}}^{\star}(T) \in \mathcal{L}$. Consequently, T is ζ-I-open. ■

Theorem 29. *Let (X, ϑ) be a topological space with an ideal $\mathcal{L}$. Then $\{a\}$ is semi*-I-closed or $\{a\}$ is a ζ-I-open subset of X for each $a \in X$.*

Proof. Let $\{a\}$ be not semi*-I-closed. Since $\{a\}$ is not semi*-I-closed, then $\hat{\imath}(\hat{\mathbf{c}}^{\star}(\{a\})) \not\subseteq \{a\}$. Furthermore, we have $\hat{\imath}(\hat{\mathbf{c}}^{\star}(\{a\})) \neq \varnothing$ and $\hat{\imath}(\hat{\mathbf{c}}^{\star}(\{a\}))$ is a pre-I-open subset of X. On the other hand,

$$\hat{\imath}(\hat{\mathbf{c}}^{\star}(\{a\})) \setminus \hat{\mathbf{c}}^{\star}(\{a\}) \in \mathcal{L}.$$

Thus, $\{a\}$ is ζ-I-open. ■

Acknowledgments

I would like to thank the editor and the referees.

References

[1] M. Caldas, S. Jafari, S.P. Moshokoa, On some new maximal and minimal sets via θ-open sets, Commun. Korean Math. Soc. 25 (4) (2010) 623–628.
[2] I. Dochviri, On some properties of pairwise extremally disconnected bitopological spaces, Proc. A. Razmadze Math. Inst. 142 (2006) 1–7.
[3] E. Ekici, On an openness which is placed between topology and Levine's openness, Jordan J. Math. Stat. 9 (4) (2016) 303–313.
[4] E. Ekici, S. Jafari, On $\mathcal{DS}^*$-sets and decompositions of continuous functions, Filomat 22 (2) (2008) 65–73.
[5] E. Ekici, A note on a-open sets and e^*-open sets, Filomat 22 (1) (2008) 89–96.
[6] E. Ekici, T. Noiri, Decompositions of continuity, α-continuity and $\mathcal{AB}$-continuity, Chaos Solitons Fractals 41 (2009) 2055–2061.
[7] E. Ekici, On e-open sets, $\mathcal{DP}^*$-sets and $\mathcal{DPE}^*$-sets and decompositions of continuity, Arab. J. Sci. Eng. 33 (2A) (2008) 269–282.
[8] J. Dontchev, Idealization of Ganster-Reilly decomposition theorems, 1999, arxiv:math.GN/9901017v1.
[9] E. Ekici, Ö. Elmalı, On decompositions via generalized closedness in ideal spaces, Filomat 29 (4) (2015) 879–886.
[10] E. Ekici, On I-Alexandroff and $I_{\mathbf{g}}$-Alexandroff ideal topological spaces, Filomat 25 (4) (2011) 99–108.
[11] E. Ekici, T. Noiri, Properties of I-submaximal ideal topological spaces, Filomat 24 (4) (2010) 87–94.
[12] E. Ekici, T. Noiri, On subsets and decompositions of continuity in ideal topological spaces, Arab. J. Sci. Eng. 34 (1A) (2009) 165–177.
[13] D. Janković, T.R. Hamlett, New topologies from old via ideals, Amer. Math. Monthly 97 (1990) 295–310.
[14] E. Hatir, T. Noiri, On decompositions of continuity via idealization, Acta Math. Hungar. 96 (2002) 341–349.
[15] E. Ekici, On $\mathcal{A}_I^*$-sets, $\mathcal{C}_I$-sets, $\mathcal{C}_I^*$-sets and decompositions of continuity in ideal topological spaces, An. Stiint. Univ. Al. I. Cuza Iasi. (S.N.) Mat. Tomul LIX (2013) 173–184 f. 1.
[16] K. Kuratowski, Topology, Vol. 1, Academic Press, New York, 1966.
[17] E. Hatir, A. Keskin, T. Noiri, On a new decomposition of continuity via idealization, JP J. Geom. Topol. 3 (1) (2003) 53–64.
[18] E. Ekici, T. Noiri, $\star$-hyperconnected ideal topological spaces, An. Stiint. Univ. Al. I. Cuza Iasi. (S.N.) Mat. Tomul LVIII (1) (2012) 121–129.
[19] E. Ekici, T. Noiri, Certain subsets in ideal topological spaces, An. Univ. Oradea Fasc. Mat. Tom XVII (2) (2010) 125–132.
[20] J. Dontchev, M. Ganster, D. Rose, Ideal resolvability, Topol. Appl. 93 (1999) 1–16.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 378–387
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Generated sets of the complete semigroup binary relations defined by semilattices of the finite chains

Omari Givradze, Yasha Diasamidze, Nino Tsinaridze*

Department of Mathematics, Faculty of Physics, Mathematics and Computer Sciences, Shota Rustaveli Batumi State University, 35 Ninoshvili St., Batumi 6010, Georgia

Received 20 April 2018; received in revised form 30 July 2018; accepted 18 August 2018
Available online 31 August 2018

Abstract

In this article, we study generated sets of the complete semigroups of binary relations defined by X-semilattices unions of the finite chains. We found uniquely irreducible generating set for the given semigroups.

Keywords: Semigroup; Semilattice; Binary relation

1. Introduction

Let X be an arbitrary nonempty set, D be an X-semilattice of unions which is closed with respect to the set-theoretic union of elements from D, f be an arbitrary mapping of the set X in the set D. To each mapping f we put into correspondence a binary relation α_f on the set X that satisfies the condition

$$\alpha_f = \bigcup_{x\in X}\big(\{x\} \times f(x)\big).$$

The set of all such α_f $(f : X \to D)$ is denoted by $B_X(D)$. It is easy to prove that $B_X(D)$ is a semigroup with respect to the operation of multiplication of binary relations, which is called a complete semigroup of binary relations defined by an X-semilattice of unions D. We denote by $\varnothing$ an empty binary relation or an empty subset of the set X. The condition $(x, y) \in \alpha$ will be written in the form $x\alpha y$. Further, let $x, y \in X$, $Y \subseteq X$, $\alpha \in B_X(D)$, $\breve{D} = \bigcup_{Y\in D} Y$ and

* Corresponding author.
E-mail address: n.tsinaridze@bsu.edu.ge (N. Tsinaridze).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.08.004

$T \in D$. We denote by the symbols $y\alpha$, $Y\alpha$, $V(D,\alpha)$, X^* and $V(X^*,\alpha)$ the following sets:

$$y\alpha = \{x \in X \mid yx\alpha x\}, \quad Y\alpha = \bigcup_{y\in Y} y\alpha, \quad V(D,\alpha) = \{Y\alpha \mid Y \in D\}, \quad X^* = \{Y \mid \varnothing \neq Y \subseteq X\},$$
$$V(X^*,\alpha) = \{Y\alpha \mid \varnothing \neq Y \subseteq X\}, \quad D_T = \{Z \in D \mid T \subseteq Z\}, \quad Y_T^\alpha = \{y \in X \mid y\alpha = T\}.$$

It is well known the following statements:

Theorem 1.1. *Let $D = \{\check{D}, Z_1, Z_2, \ldots, Z_{m-1}\}$ be some finite X-semilattice of unions and $C(D) = \{P_0, P_1, P_2, \ldots, P_{m-1}\}$ be the family of sets of pairwise nonintersecting subsets of the set X(the set $\varnothing$ can be repeated several times). If φ is a mapping of the semilattice D on the family of sets $C(D)$ which satisfies the conditions*

$$\varphi = \begin{pmatrix} \check{D} & Z_1 & Z_2 & \cdots & Z_{m-1} \\ P_0 & P_1 & P_2 & \cdots & P_{m-1} \end{pmatrix}$$

and $\widehat{D}_Z = D \setminus D_Z$, then the following equalities are valid:

$$\check{D} = P_0 \cup P_1 \cup P_2 \cup \cdots \cup P_{m-1},$$
$$Z_i = P_0 \cup \bigcup_{T\in\widehat{D}_{Z_i}} \varphi(T). \tag{1.1}$$

In the sequel these equalities will be called formal.

It is proved that if the elements of the semilattice D are represented in the form (1.1), then among the parameters P_i $(0 < i \leq m-1)$ there exist such parameters that cannot be empty sets for D. Such sets P_i are called bases sources, whereas sets P_j $(0 \leq j \leq m-1)$ which can be empty sets too are called completeness sources.

It is proved that under the mapping φ the number of covering elements of the pre-image of a bases source is always equal to one, while under the mapping φ the number of covering elements of the pre-image of a completeness source either does not exist or is always greater than one (see [1, Chapter 11]).

Definition 1.1. We say that an element α of the semigroup $B_X(D)$ is external if $\alpha \neq \delta \circ \beta$ for all $\delta, \beta \in B_X(D) \setminus \{\alpha\}$ (see [1, Definition 1.15.1]).

It is well known, that if B is all external elements of the semigroup $B_X(D)$ and B' is any generated set for the $B_X(D)$, then $B \subseteq B'$ (see [1, Lemma 1.15.1]).

Definition 1.2. The representation $\alpha = \bigcup_{T\in D}(Y_T^\alpha \times T)$ of binary relation α is called quasinormal, if

$$\bigcup_{T\in D} Y_T^\alpha = X \text{ and } Y_T^\alpha \cap Y_{T'}^\alpha = \varnothing \text{ for any } T, T' \in D, \quad T \neq T'.$$

Definition 1.3. Let $\alpha, \beta \subseteq X \times X$. Their product $\delta = \alpha \circ \beta$ is defined as follows: $x\delta y$ $(x, y \in X)$ if there exists an element $z \in X$ such that $x\alpha z\beta y$ (see [1, Chapter 1.3]).

2.

Let $\Sigma_m(X, m)$ be a class of all X-semilattices of unions whose every element is isomorphic to an X-semilattice of unions $D = \{Z_1, Z_2, Z_3, \ldots, Z_{m-2}, Z_{m-1}, Z_m = \check{D}\}$, which satisfies the condition

$$Z_1 \subset Z_2 \subset Z_3 \subset \cdots \subset Z_{m-2} \subset Z_{m-1} \subset Z_m = \check{D} \tag{2.1}$$

(see Fig. 1).

Let $C(D) = \{P_0, P_1, P_2, P_3, \ldots, P_{m-2}, P_{m-1}\}$ be a family of sets, where $P_0, P_1, P_2, P_3, \ldots, P_{m-2}, P_{m-1}$ are pairwise disjoint subsets of the set X and $\varphi = \begin{pmatrix} \check{D} & Z_1 & Z_2 & Z_3 & \cdots & Z_{m-2} & Z_{m-1} \\ P_0 & P_1 & P_2 & P_3 & \cdots & P_{m-2} & P_{m-1} \end{pmatrix}$ is a mapping of the semilattice

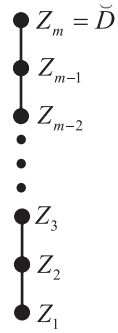


Fig. 1.

D onto the family of sets $C(D)$. Then the formal equalities of the semilattice D have a form:

$$\begin{aligned}
\breve{D} = Z_m &= P_0 \cup P_1 \cup P_2 \cup P_3 \cup \cdots \cup P_{m-1},\\
Z_m &= P_0 \cup P_1 \cup P_2 \cup P_3 \cup \cdots \cup P_{m-2},\\
Z_{m-1} &= P_0 \cup P_1 \cup P_2 \cup P_3 \cup \cdots \cup P_{m-3},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
Z_3 &= P_0 \cup P_1 \cup P_2,\\
Z_2 &= P_0 \cup P_1,\\
Z_1 &= P_0,
\end{aligned} \tag{2.2}$$

Here the elements $P_1, P_2, P_3, \ldots, P_{m-2}, P_{m-1}$ are bases sources, the element P_0 is sources of completeness of the semilattice D. Therefore $|X| \geq m-1$, since $|P_i| \geq 1$, $i = 1, 2, \ldots, m-1$ (see Theorem 1.1).

In this paper we are learning irreducible generating sets of the semigroup $B_X(D)$ defined by semilattices of the class $\Sigma_m(X, m)$.

Definition 2.1. In the sequel, by symbol $\Sigma_{m.0}(X, m)$ we denote all semilattices $D = \{Z_1, Z_2, Z_3, \ldots, Z_{m-2}, Z_{m-1}, Z_m\}$ of the class $\Sigma_m(X, m)$ for which Z_1 is not empty. From the last inequality and from the formal equalities (2.2) it follows that $Z_1 = P_0 \neq \varnothing$, i.e. in this case $|X| \geq m$.

Note that, if $D \in \Sigma_{m,0}(X, m)$ and by symbol $|X|$ is denoted the power of a set X, then the inequality $|X| \geq |D|$ always is true.

It is easy to see that $\widetilde{D} = \{Z_1, Z_2, Z_3, \ldots, Z_{m-2}, Z_{m-1}, Z_m\}$ is irreducible generating set for the semilattice D.

Lemma 2.1. *If $D \in \Sigma_m(X, m)$, then the following statements are true:*

$$\begin{aligned}
&\text{(a)} \quad P_0 = Z_1;\\
&\text{(b)} \quad P_{i-1} = Z_i \setminus Z_{i-1}, \quad i = 2, 3, \ldots, m-2, m-1, m.
\end{aligned} \tag{2.3}$$

Proof. From the formal equalities of the semilattice D immediately follows the following statements:

$$P_0 = Z_1, \quad P_{i-1} = Z_i \setminus Z_{i-1}, \quad i = 2, 3, \ldots, m-2, m-1, m.$$

The statements (a) and (b) of Lemma 2.1 are proved. □

It is well known, that the binary relation α, representation, which has a form

$$\varepsilon = (Z_1 \times Z_1) \cup ((Z_2 \setminus Z_1) \times Z_2) \cup \cdots \cup ((Z_{m-1} \setminus Z_{m-2}) \times Z_{m-1}) \cup ((Z_m \setminus Z_{m-1}) \times Z_m) \cup ((X \setminus \breve{D}) \times \breve{D})$$

is largest right unit of the semigroup $B_X(D)$ (see [1, Theorem 15.1.1]) and if by $E_X^{(r)}(D)$ is denoted all right units of the semigroup $B_X(D)$, then all right units of the semigroup are its external elements.

Definition 2.2. We denote the following sets by symbol $\mathfrak{A}_k$ and $B(\mathfrak{A}_k)$:

$$\mathfrak{A}_k = \big\{V(X^*, \alpha) \mid \ |V(X^*, \alpha)| = k \text{ for any} \alpha \in B_X(D)\big\},$$
$$B(\mathfrak{A}_k) = \big\{\alpha \in B_X(D) \mid \ V(X^*, \alpha) \in \mathfrak{A}_k\big\},$$

where $k = m, m-1, m-2, \ldots, 3, 2, 1$.

Lemma 2.2. *Let $\alpha \in B(\mathfrak{A}_m)$, then α is external element for the semigroup $B_X(D)$.*

Proof. Let $\alpha \in B(\mathfrak{A}_m)$ and $\alpha = \delta \circ \beta$ for some $\delta, \beta \in B_X(D) \setminus \{\alpha\}$. If quasinormal representation of binary relation δ has a form

$$\delta = \bigcup_{i=1}^{m}(Y_i^{\delta} \times Z_i),$$

then

$$\alpha = \delta \circ \beta = \Big(\bigcup_{i=1}^{m}(Y_i^{\delta} \times Z_i)\Big) \circ \beta = \bigcup_{i=1}^{m}(Y_i^{\delta} \times Z_i\beta). \tag{2.4}$$

It is easy to see that

$$Z_1\beta \subseteq Z_2\beta \subseteq Z_3\beta \subseteq \cdots \subseteq Z_{m-2}\beta \subseteq Z_{m-1}\beta \subseteq Z_m\beta = \breve{D}\beta \tag{2.5}$$

since $Z_1 \subset Z_2 \subset Z_3 \subset \cdots \subset Z_{m-2} \subset Z_{m-1} \subset Z_m = \breve{D}$ by definition of the semilattice D. By preposition

$$\big\{Z_1\beta, Z_2\beta, Z_3\beta, \ldots, Z_{m-2}\beta, Z_{m-1}\beta, Z_m\beta\big\} = \big\{Z_1, Z_2, Z_3, \ldots, Z_{m-2}, Z_{m-1}, Z_m\big\}$$
$$\big(|V(D, \alpha)| = |V(D, \delta \circ \beta)| = m\big),$$

i.e. there exists one to one mapping φ of the set $\{Z_1\beta, Z_2\beta, Z_3\beta, \ldots, Z_{m-2}\beta, Z_{m-1}\beta, Z_m\beta\}$ on the set $\{Z_1, Z_2, Z_3, \ldots, Z_{m-2}, Z_{m-1}, Z_m\}$ which satisfies the condition (2.1) and (2.5). Now, let $\varphi(Z_i\beta) = Z_j$ for some $Z_i \subseteq Z_j (i \leq j)$. If $Z_i \neq Z_j$, then there exist such p and q $(1 \leq p \neq q \leq m)$, that $Z_p\beta = Z_q$ for which $(p > q)$, since D is finite. But equality $Z_p\beta = Z_q$ and inclusion $p > q$ contradicts the inclusions (2.5), i.e. $i = j$. Of this follows that $Z_i\beta = Z_i$ for all $i = 1, 2, 3, \ldots, m$. We have that

$$\alpha = \delta \circ \beta = \bigcup_{i=1}^{m}(Y_i^{\delta} \times Z_i\beta) = \bigcup_{i=1}^{m}(Y_i^{\delta} \times Z_i) = \delta.$$

But equality $\alpha = \delta$ contradicts the preposition $\delta \in B_X(D) \setminus \{\alpha\}$. □

Lemma 2.3. *Let $D \in \Sigma_{m.0}(X, m)$. If $|X \setminus \breve{D}| \geq 1$, then any element of set $B(\mathfrak{A}_k)$ is generated by elements of the set $B(\mathfrak{A}_{k+1})$ $(k = m-1, \ldots, 4, 3, 2)$.*

Proof. Let $D \in \Sigma_{m.0}(X, m)$, then $|X| \geq |D| = m$. By Lemma 2.2 it follows that if $\alpha \in B(\mathfrak{A}_m)$, then α is external element for the semigroup $B_X(D)$. Now, let $k < m$ and $\alpha \in B(\mathfrak{A}_k)$, then quasinormal representation of a binary relation has a form $\alpha = \bigcup_{i=1}^{k}(Y_i^{\alpha} \times T_i)$, where $Y_1^{\alpha}, Y_2^{\alpha}, \ldots, Y_k^{\alpha} \notin \{\varnothing\}$, $T_1, T_2, \ldots, T_k \in D$ and $T_1 \subset T_2 \subset \cdots \subset T_k$. It is easy to see, that $V(X^*, \alpha) = \{T_1, T_2, \ldots, T_k\}$, i.e. $|V(X^*, \alpha)| = k$. From the sets $Y_1^{\alpha}, Y_2^{\alpha}, \ldots, Y_k^{\alpha}$ there exists such set Y_q^{α}, that satisfies the condition $|Y_q^{\alpha}| \geq 2$ since $k < m$. Therefore, we may suppose, that $Y_q^{\alpha} = Y_{q1}^{\alpha} \cup Y_{q2}^{\alpha}$, where $Y_{q1}^{\alpha}, Y_{q2}^{\alpha} \notin \{\varnothing\}$ and $Y_{q1}^{\alpha} \cap Y_{q2}^{\alpha} = \varnothing$. Further, there exist such element $T \in D$ $(T \notin \{T_1, T_2, \ldots, T_k\})$. For the element T we consider the following cases:

(1) $T \subset T_1$;
(2) $T_k \subset T$;
(3) $T_1 \subset T \subset T_q$;
(4) $T_q \subset T \subset T_k$.

(1). Let $T \subset T_1$ and $T = T_0$. If quasinormal representations of the binary relations δ and β have the forms

$$\begin{aligned}\delta &= (Y_1^\alpha \times T_0) \cup (Y_2^\alpha \times T_1) \cup (Y_3^\alpha \times T_2) \cup \cdots \\ &\quad \cup (Y_{q-1}^\alpha \times T_{k-2}) \cup (Y_{q1}^\alpha \times T_{q-1}) \cup (Y_{q2}^\alpha \times T_q) \cup (Y_{q+1}^\alpha \times T_{q+1}) \cup \cdots \cup (Y_k^\alpha \times T_k);\\ \beta &= (T_0 \times T_1) \cup ((T_1 \setminus T_0) \times T_2) \cup ((T_2 \setminus T_1) \times T_3) \cup \cdots \\ &\quad \cup ((T_{q-2} \setminus T_{q-3}) \times T_{q-1}) \cup ((T_q \setminus T_{q-2}) \times T_q) \cup ((T_{q+1} \setminus T_q) \times T_{q+1}) \cup \cdots \\ &\quad \cup ((T_k \setminus T_{k-1}) \times T_k) \cup ((X \setminus T_k) \times T_0),\end{aligned}$$

where $Y_1^\alpha, Y_2^\alpha, Y_3^\alpha, \ldots, Y_{q-1}^\alpha Y_{q1}^\alpha Y_{q2}^\alpha Y_{q+1}^\alpha, \ldots, Y_k^\alpha \notin \{\varnothing\}$, then it is easy to see, that $|V(X^*, \delta)| = |V(X^*, \beta)| = k+1$, i.e. $V(X^*, \delta), V(X^*, \beta) \in \mathfrak{A}_{k+1}$ and

$$T_0\beta = T_1, T_1\beta = T_2, \ldots, T_{q-2}\beta = T_{q-1}, T_{q-1}\beta = T_q\beta = T_q, T_{q+1}\beta = T_{q+1}, T_{q+2}\beta = T_{q+2}, \ldots, T_k\beta = T_k.$$

From the last equalities we obtain, that:

$$\begin{aligned}\delta \circ \beta &= (Y_1^\alpha \times T_0\beta) \cup (Y_2^\alpha \times T_1\beta) \cup (Y_3^\alpha \times T_2\beta) \cup \cdots \\ &\cup (Y_{q-1}^\alpha \times T_{q-2}\beta) \cup (Y_{q1}^\alpha \times T_{q-1}\beta) \cup (Y_{q2}^\alpha \times T_q\beta) \cup (Y_{q+1}^\alpha \times T_{q+1}\beta) \cup \cdots \cup (Y_k^\alpha \times T_k\beta) \\ &= (Y_1^\alpha \times T_1) \cup (Y_2^\alpha \times T_2) \cup (Y_3^\alpha \times T_3) \cup v \cup (Y_{q-1}^\alpha \times T_{q-1}) \cup \\ &\quad ((Y_{q1}^\alpha \cup Y_{q2}^\alpha) \times T_q) \cup (Y_{q+1}^\alpha \times T_{q+1}) \cup \cdots \cup (Y_k^\alpha \times T_k) = \alpha,\end{aligned}$$

since $Y_q^\alpha = Y_{q1}^\alpha \cup Y_{q2}^\alpha$, i.e. $\alpha = \delta \circ \beta$, $V(X^*, \alpha) = \{T_1, T_2, \ldots, T_k\}$ and $|V(X^*, \alpha)| = k$. Therefore, $V(X^*, \alpha) \in \mathfrak{A}_k$, i.e. element α is generated by elements of the set $B(\mathfrak{A}_{k+1})$.

(2). Let $T_k \subset T$ and $T_{k+1} = T$. If quasinormal representations of the binary relations δ and β have the forms

$$\begin{aligned}\delta &= (Y_1^\alpha \times T_1) \cup (Y_2^\alpha \times T_2) \cup \cdots \cup (Y_{q-1}^\alpha \times T_{q-1}) \cup (Y_{q1}^\alpha \times T_q) \cup \\ &\quad (Y_{q2}^\alpha \times T_{q+1}) \cup (Y_{q+1}^\alpha \times T_{q+2}) \cup \cdots \cup (Y_k^\alpha \times T_{k+1});\\ \beta &= (T_1 \times T_1) \cup ((T_2 \setminus T_1) \times T_2) \cup ((T_3 \setminus T_2) \times T_3) \cup \cdots \\ &\quad \cup ((T_{q-1} \setminus T_{q-2}) \times T_{q-1}) \cup ((T_{q+1} \setminus T_{q-1}) \times T_q) \cup ((T_{q+2} \setminus T_{q+1}) \times T_{q+1}) \\ &\quad \cup ((T_{q+3} \setminus T_{q+2}) \times T_{q+2}) \cup \cdots \cup ((T_{k+1} \setminus T_k) \times T_k) \cup ((X \setminus T_{k+1}) \times T_{k+1}),\end{aligned}$$

where $Y_1^\alpha, Y_2^\alpha, \ldots, Y_{q-1}^\alpha Y_{q1}^\alpha Y_{q2}^\alpha Y_{q+1}^\alpha, \ldots, Y_k^\alpha \notin \{\varnothing\}$. It is easy to see, that $|V(X^*, \delta)| = |V(X^*, \beta)| = k+1$, i.e. $V(X^*, \delta), V(X^*, \beta) \in \mathfrak{A}_{k+1}$ and

$$T_1\beta = T_1, T_2\beta = T_2, \ldots, T_{q-1}\beta = T_{q-1}, T_q\beta = T_{q+1}\beta = T_q, T_{q+2}\beta = T_{q+1}, T_{q+3}\beta = T_{q+2}, \ldots, T_{k+1}\beta = T_k.$$

From the last equalities we obtain, that:

$$\begin{aligned}\delta \circ \beta &= \delta = (Y_1^\alpha \times T\beta_1) \cup (Y_2^\alpha \times T\beta_2) \cup \cdots \cup (Y_{q-1}^\alpha \times T_{q-1}\beta) \\ &\cup (Y_{q1}^\alpha \times T_q\beta) \cup (Y_{q2}^\alpha \times T_{q+1}\beta) \cup (Y_{q+1}^\alpha \times T_{q+2}\beta) \cup \cdots \cup (Y_k^\alpha \times T_{k+1}\beta) \\ &= (Y_1^\alpha \times T_1) \cup (Y_2^\alpha \times T_2) \cup \cdots \cup (Y_{q1}^\alpha \times T_q) \cup (Y_{q2}^\alpha \times T_q) \cup \cdots \cup (Y_k^\alpha \times T_k) \\ &= (Y_1^\alpha \times T_1) \cup (Y_2^\alpha \times T_2) \cup \cdots \cup ((Y_{q1}^\alpha \cup Y_{q2}^\alpha) \times T_q) \cup \cdots \cup (Y_k^\alpha \times T_k) = \alpha,\end{aligned}$$

since $Y_q^\alpha = Y_{q1}^\alpha \cup Y_{q2}^\alpha$, i.e. $\alpha = \delta \circ \beta$, $V(X^*, \alpha) = \{T_1, T_2, \ldots, T_k\}$ and $|V(X^*, \alpha)| = k$.

Therefore, $V(X^*, \alpha) \in \mathfrak{A}_k$, i.e. element α is generated by elements of the set $B(\mathfrak{A}_{k+1})$.

(3). Let $T_1 \subset T \subset T_q$. Then there exists an element T which is covered by element T_l, where $2 \le l \le q$, i.e. $T_{l-1} \subset T \subset T_l$. So, we have the following chain:

$$T_1 \subset T_2 \subset T_{l-1} \subset T \subset T_l \subset T_{l+1} \subset T_{l+2} \subset \cdots \subset T_q \subset T_{q+1} \subset T_{q+2} \subset \cdots \subset T_k.$$

If quasinormal representations of the binary relations δ and β have the forms

$$\begin{aligned}\delta &= (Y_1^\alpha \times T_1) \cup (Y_2^\alpha \times T_2) \cup \cdots \cup (Y_{l-1}^\alpha \times T_{l-1}) \cup (Y_l^\alpha \times T) \cup (Y_{l+1}^\alpha \times T_l) \cup (Y_{l+2}^\alpha \times T_{l+1}) \cup \cdots \\ &\quad \cup (Y_{q-1}^\alpha \times T_{q-2}) \cup (Y_{q1}^\alpha \times T_{q-1}) \cup (Y_{q2}^\alpha \times T_q) \cup (Y_{q+1}^\alpha \times T_{q+1}) \cup (Y_{q+2}^\alpha \times T_{q+2}) \cup \cdots \cup (Y_k^\alpha \times T_k);\\ \beta &= (T_1 \times T_1) \cup ((T_2 \setminus T_1) \times T_2) \cup ((T_3 \setminus T_2) \times T_3) \cup \cdots \cup ((T_{l-1} \setminus T_{l-2}) \times T_{l-1}) \cup ((T \setminus T_{l-1}) \times T_l) \\ &\quad \cup ((T_l \setminus T) \times T_{l+1}) \cup ((T_{l+1} \setminus T_l) \times T_{l+2}) \cup \cdots \cup ((T_{q-2} \setminus T_{q-3}) \times T_{q-1}) \cup ((T_q \setminus T_{q-2}) \times T_q) \\ &\quad \cup ((T_{q+1} \setminus T_q) \times T_{q+1}) \cup \cdots \cup ((T_k \setminus T_{k-1}) \times T_k) \cup ((X \setminus T_k) \times T),\end{aligned}$$

where $Y_1^\alpha, \dots, Y_{l-1}^\alpha, Y_l^\alpha, Y_{l+1}^\alpha, \dots, Y_{q-1}^\alpha, Y_{q1}^\alpha, Y_{q2}^\alpha, Y_{q+1}^\alpha, Y_{q+2}^\alpha, \dots, Y_k^\alpha \notin \{\varnothing\}$. It is easy to see, that $|V(X^*, \delta)| = |V(X^*, \beta)| = k+1$, i.e. $V(X^*, \delta), V(X^*, \beta) \in \mathfrak{A}_{k+1}$ and

$$T_1\beta = T_1, T_2\beta = T_2, \dots, T_{l-1}\beta = T_{l-1}, T\beta = T_l, T_l\beta = T_{l+1}, T_{l+1}\beta = T_{l+2}, \dots,$$
$$T_{q-2}\beta = T_{q-1}, T_{q-1}\beta = T_q\beta = T_q, T_{q+1}\beta = T_{q+1}, T_{q+2}\beta = T_{q+2}, \dots, T_k\beta = T_k.$$

From the last equalities we obtain, that:

$$\begin{aligned}
\delta\circ\beta &= (Y_1^\alpha \times T_1\beta)\cup(Y_2^\alpha \times T_2\beta)\cup\dots\cup(Y_{l-1}^\alpha \times T_{l-1}\beta)\cup(Y_l^\alpha \times T\beta)\\
&\quad\cup(Y_{l+1}^\alpha \times T_l\beta)\cup(Y_{l+2}^\alpha \times T_{l+1}\beta)\cup\dots\cup(Y_{q-1}^\alpha \times T_{q-2}\beta)\cup(Y_{q1}^\alpha \times T_{q-1}\beta)\cup(Y_{q2}^\alpha \times T_q\beta)\\
&\quad\cup(Y_{q+1}^\alpha \times T_{q+1}\beta)\cup(Y_{q+2}^\alpha \times T_{q+2}\beta)\cup\dots\cup(Y_k^\alpha \times T_k\beta)\\
&= (Y_1^\alpha \times T_1)\cup(Y_2^\alpha \times T_2)\cup\dots\cup(Y_{l-1}^\alpha \times T_{l-1})\cup(Y_l^\alpha \times T_l)\cup(Y_{l+1}^\alpha \times T_{l+1})\cup(Y_{l+2}^\alpha \times T_{l+2})\cup\dots\\
&\quad\cup(Y_{q1}^\alpha \times T_q)\cup(Y_{q2}^\alpha \times T_q)\cup(Y_{q+1}^\alpha \times T_{q+1})\cup(Y_{q+2}^\alpha \times T_{q+2})\cup\dots\cup(Y_k^\alpha \times T_k)\\
&= (Y_1^\alpha \times T_1)\cup(Y_2^\alpha \times T_2)\cup\dots\cup(Y_{l-1}^\alpha \times T_{l-1})\cup(Y_l^\alpha \times T_l)\cup(Y_{l+1}^\alpha \times T_{l+1})\cup(Y_{l+2}^\alpha \times T_{l+2})\cup\dots\\
&\quad\cup((Y_{q1}^\alpha \cup Y_{q2}^\alpha)\times T_q)\cup(Y_{q+1}^\alpha \times T_{q+1})\cup(Y_{q+2}^\alpha \times T_{q+2})\cup\dots\cup(Y_k^\alpha \times T_k) = \alpha,
\end{aligned}$$

since $Y_q^\alpha = Y_{q1}^\alpha \cup Y_{q2}^\alpha$, i.e. $\alpha = \delta\circ\beta$, $V(X^*, \alpha) = \{T_1, T_2, \dots, T_k\}$ and $|V(X^*, \alpha)| = k$.

Therefore, $V(X^*, \alpha) \in \mathfrak{A}_k$, i.e. element α is generated by elements of the set $B(\mathfrak{A}_{k+1})$.

(4). Let $T_q \subset T \subset T_k$. Then there exists an element T which is covered by element T_r, where $q+1 \le r \le k$, i.e. $T_{r-1} \subset T \subset T_r$. So, we have the following chain:

$$T_1 \subset T_2 \subset \dots \subset T_q \subset T_{q+1} \subset \dots \subset T_{r-1} \subset T \subset T_r \subset T_{r+1} \subset \dots \subset T_k.$$

If quasinormal representations of the binary relations δ and β have the forms

$$\begin{aligned}
\delta &= (Y_1^\alpha \times T_1)\cup(Y_2^\alpha \times T_2)\cup\dots\cup(Y_{q-1}^\alpha \times T_{q-1})\cup\dots\cup(Y_{q1}^\alpha \times T_q)\\
&\quad\cup(Y_{q2}^\alpha \times T_{q+1})\cup(Y_{q+1}^\alpha \times T_{q+2})\cup(Y_{q+2}^\alpha \times T_{q+3})\cup\dots\cup(Y_{r-2}^\alpha \times T_{r-1})\\
&\quad\cup(Y_{r-1}^\alpha \times T)\cup(Y_r^\alpha \times T_r)\cup(Y_{r+1}^\alpha \times T_{r+1})\cup\dots\cup(Y_k^\alpha \times T_k);\\
\beta &= (T_1 \times T_1)\cup((T_2 \setminus T_1)\times T_2)\cup((T_3 \setminus T_2)\times T_3)\cup\dots\cup((T_{q-1} \setminus T_{q-2})\times T_{q-1})\\
&\quad\cup((T_{q+1} \setminus T_{q-1})\times T_q)\cup((T_{q+2} \setminus T_{q+1})\times T_{q+1})\cup\dots\cup((T_{r-1} \setminus T_{r-2})\times T_{r-2})\\
&\quad\cup((T \setminus T_{r-1})\times T_{r-1})\cup((T_r \setminus T)\times T_r)\cup((T_{r+1} \setminus T_r)\times T_{r+1})\cup\dots\\
&\quad\cup((T_k \setminus T_{k-1})\times T_k)\cup((X \setminus T_k)\times T),
\end{aligned}$$

where $Y_1^\alpha, Y_2^\alpha, \dots, Y_{q-1}^\alpha, \dots, Y_{q1}^\alpha, Y_{q2}^\alpha, Y_{q+1}^\alpha, Y_{q+2}^\alpha, \dots, Y_{r-2}^\alpha, Y_{r-1}^\alpha, Y_r^\alpha, Y_{r+1}^\alpha, \dots, Y_k^\alpha \notin \{\varnothing\}$. It is easy to see, that $|V(X^*, \delta)| = |V(X^*, \beta)| = k+1$, i.e. $V(X^*, \delta), V(X^*, \beta) \in \mathfrak{A}_{k+1}$ and

$$T_1\beta = T_1, T_2\beta = T_2, \dots, T_{q-1}\beta = T_{q-1}, T_q\beta = T_{q+1}\beta = T_q, T_{q+2}\beta = T_{q+1}, T_{q+3}\beta = T_{q+2}, \dots,$$
$$T_{r-1}\beta = T_{r-2}, T\beta = T_{r-1}, T_r\beta = T_r, T_{r+1}\beta = T_{r+1}, \dots, T_k\beta = T_k.$$

From the last equalities we obtain, that:

$$\begin{aligned}
\delta\circ\beta &= (Y_1^\alpha \times T_1\beta)\cup(Y_2^\alpha \times T_2\beta)\cup\dots\cup(Y_{q-1}^\alpha \times T_{q-1}\beta)\cup\dots\cup(Y_{q1}^\alpha \times T_q\beta)\cup(Y_{q2}^\alpha \times T_{q+1}\beta)\\
&\quad\cup(Y_{q+1}^\alpha \times T_{q+2}\beta)\cup(Y_{q+2}^\alpha \times T_{q+3}\beta)\cup\dots\cup(Y_{r-2}^\alpha \times T_{r-1}\beta)\\
&\quad\cup(Y_{r-1}^\alpha \times T\beta)\cup(Y_r^\alpha \times T_r\beta)\cup(Y_{r+1}^\alpha \times T_{r+1}\beta)\cup\dots\cup(Y_k^\alpha \times T_k\beta)\\
&= (Y_1^\alpha \times T_1)\cup(Y_2^\alpha \times T_2)\cup\dots\cup(Y_{q-1}^\alpha \times T_{q-1})\cup\dots\cup(Y_{q1}^\alpha \times T_q)\cup(Y_{q2}^\alpha \times T_q)\\
&\quad\cup(Y_{q+1}^\alpha \times T_{q+1})\cup(Y_{q+2}^\alpha \times T_{q+2})\cup\dots\cup(Y_{r-2}^\alpha \times T_{r-2})\\
&\quad\cup(Y_{r-1}^\alpha \times T_{r-1})\cup(Y_r^\alpha \times T_r)\cup(Y_{r+1}^\alpha \times T_{r+1})\cup\dots\cup(Y_k^\alpha \times T_k)\\
&= (Y_1^\alpha \times T_1)\cup(Y_2^\alpha \times T_2)\cup\dots\cup(Y_{q-1}^\alpha \times T_{q-1})\cup\dots\cup((Y_{q1}^\alpha \cup Y_{q2}^\alpha)\times T_q)\\
&\quad\cup(Y_{q+1}^\alpha \times T_{q+1})\cup(Y_{q+2}^\alpha \times T_{q+2})\cup\dots\cup(Y_{r-2}^\alpha \times T_{r-2})\\
&\quad\cup(Y_{r-1}^\alpha \times T_{r-1})\cup(Y_r^\alpha \times T_r)\cup(Y_{r+1}^\alpha \times T_{r+1})\cup\dots\cup(Y_k^\alpha \times T_k) = \alpha,
\end{aligned}$$

since $Y_q^\alpha = Y_{q1}^\alpha \cup Y_{q2}^\alpha$, i.e. $\alpha = \delta \circ \beta$, since

$$Y_q^\alpha = Y_{q1}^\alpha \cup Y_{q2}^\alpha, \quad V(X^*, \alpha) = \{T_1, T_2, \ldots, T_k\} \text{ and } |V(X^*, \alpha)| = k.$$

Therefore, $V(X^*, \alpha) \in \mathfrak{A}_k$, i.e. element α is generated by elements of the set $B(\mathfrak{A}_{k+1})$.

So, any element of set $B(\mathfrak{A}_k)$ is generated by elements of the set $B(\mathfrak{A}_{k+1})$, where $k = m-1, \ldots, 4, 3, 2$. □

Lemma 2.4. *Let $D \in \Sigma_{m.0}(X, m)$. If $|X \setminus \breve{D}| \geq 1$, then any element of the set $B(\mathfrak{A}_1)$ is generated by elements of the set $B(\mathfrak{A}_2)$.*

Proof. Let T is any element of the semilattice D and $\alpha = X \times T \in B(\mathfrak{A}_1)$. If quasinormal representations of the binary relations δ and β ($\delta, \beta \in B_X(D)$) have forms

$$\delta = (Y_1^\delta \times T) \cup (Y_2^\delta \times T'), \quad \beta = (\breve{D} \times T) \cup ((X \setminus \breve{D}) \times T'),$$

where $Y_1^\delta, Y_2^\delta \notin \{\varnothing\}$ ($T' \in D,\ T \neq T'$), then $|V(X^*, \delta)| = |V(X^*, \beta)| = 2$, i.e. $\delta, \beta \in B(\mathfrak{A}_2)$ and

$$T\beta = T'\beta = T,$$
$$\delta \circ \beta = (Y_1^\delta \times T\beta) \cup (Y_2^\delta \times T'\beta) = (Y_1^\delta \times T) \cup (Y_2^\delta \times T) = X \times T = \alpha,$$

since the representation of a binary relation δ is quasinormal.

Therefore, if $|X \setminus \breve{D}| \geq 1$, then any element of the set $B(\mathfrak{A}_1)$ is generated by elements of the set $B(\mathfrak{A}_2)$. □

Theorem 2.1. *Let $D \in \Sigma_{m.0}(X, m)$ and D is a finite set. If $|X \setminus \breve{D}| \geq 1$ and*

$$\mathfrak{A}_m = \left\{V(X^*, \alpha) \mid\ |V(X^*, \alpha)| = m \text{ for any } \alpha \in B_X(D)\right\}, \quad B(\mathfrak{A}_m) = \left\{\alpha \in B_X(D) \mid\ V(X^*, \alpha) \in \mathfrak{A}_m\right\},$$

then the set $B(\mathfrak{A}_m)$ is irreducible generating set for the semigroup $B_X(D)$.

Proof. Let $|X \setminus \breve{D}| \geq 1$. First, we proved that every element of the semigroup $B_X(D)$ is generated by elements of the set $B(\mathfrak{A}_m)$. Indeed, let α be an arbitrary element of the semigroup $B_X(D)$. Then quasinormal representation of a binary relation α has a form $\alpha = \bigcup_{i=1}^m (Y_i^\alpha \times Z_i)$, where $\bigcup_{i=1}^m Y_i^\alpha = X$ and $Y_i^\alpha \cap Y_j^\alpha = \varnothing$ for any $1 \leq i \neq j \leq m$. For the number $|V(X^*, \alpha)|$ we consider the following cases:

(1) If $k = m-1, \ldots, 4, 3, 2$, then from Lemma 2.3 immediately follows that any element of the set $B(\mathfrak{A}_k)$ is generated by elements of the set $B(\mathfrak{A}_{k+1})$.
(2) If $k = 1$, then from Lemma 2.4 immediately follows that any element of the set $B(\mathfrak{A}_1)$ is generated by elements of the set $B(\mathfrak{A}_2)$.

So, if $|X \setminus \breve{D}| \geq 1$, then every element of the semigroup $B_X(D)$ are generated by elements of the set $B(\mathfrak{A}_m)$. From Lemma 2.2 it follows that, every element of the set $B(\mathfrak{A}_m)$ is external element for the semigroup $B_X(D)$. Therefore, we have that $B(\mathfrak{A}_m)$ is irreducible generating set for the semigroup $B_X(D)$. □

Theorem 2.2. *Let $D \in \Sigma_{m.0}(X, m)$ and D is a finite set. If $|X \setminus \breve{D}| \geq 1$ and*

$$\mathfrak{A}_m = \left\{V(X^*, \alpha) \mid\ |V(X^*, \alpha)| = m \text{ for any } \alpha \in B_X(D)\right\}, \quad B(\mathfrak{A}_m) = \left\{\alpha \in B_X(D) \mid\ V(X^*, \alpha) \in \mathfrak{A}_m\right\}.$$

Then the paver of a set $B(\mathfrak{A}_m)$ is equal to the paver surjection of the set X on the set D.

Proof. Let X and D be finite sets. If $\alpha \in B(\mathfrak{A}_m)$, then by definition of a semigroup $B_X(D)$ there exists such a mapping f of a set X on the set $D = \{Z_1, Z_2, \ldots, Z_m\}$, since $|X| \geq |D|$ and $\alpha = \alpha_f$, i.e. mapping f is some surjection of the set X on the set D. □

Corollary 2.1. *Let $D \in \Sigma_{m.0}(X, m)$. If $|X \setminus \breve{D}| \geq 1$ and*

$$\mathfrak{A}_m = \left\{V(X^*, \alpha) \mid\ |V(X^*, \alpha)| = m \text{ for any } \alpha \in B_X(D)\right\}, \quad B(\mathfrak{A}_m) = \left\{\alpha \in B_X(D) \mid\ V(X^*, \alpha) \in \mathfrak{A}_m\right\}.$$

If X and D are finite sets $|X| = n$ and $|D| = m$, then the number $|B(\mathfrak{A}_m)|$ elements of the set $B(\mathfrak{A}_m)$ is equal to

$$|B(\mathfrak{A}_m)| = n^m + \sum_{i=1}^{n-1} (-1)^i \cdot C_n^i \cdot (n-i)^m.$$

Proof. Let X and D be finite sets, $|X| = n$ and $|D| = m$. In our case we have $|X| \geq |D|$. It is well known that, the number surjection of the set X on the set D is equal to the number $n^m + \sum_{i=1}^{n-1}(-1)^i \cdot C_n^i \cdot (n-i)^m$. Now, Corollary 2.1 immediately follows from Theorem 2.2. □

Example 2.1. $X = \{1, 2, 3\}$ and $D = \{\{1\}, \{1, 2\}\}$. Then $3 = |X| > |D| = 2$ and

$$\alpha_1 = \big\{(1,1),(2,1),(3,1)\big\}, \quad \alpha_2 = \big\{(1,1),(2,1),(3,1),(3,2)\big\}, \quad \alpha_3 = \big\{(1,1),(2,1),(2,2)(3,1)\big\},$$
$$\alpha_4 = \big\{(1,1),(1,2)(2,1),(3,1)\big\}, \quad \alpha_5 = \big\{(1,1),(1,2),(2,1),(2,2)(3,1)\big\},$$
$$\alpha_6 = \big\{(1,1),(1,2),(2,1),(3,1),(3,2)\big\},$$
$$\alpha_7 = \big\{(1,1),(2,1),(2,2),(3,1),(3,2)\big\}, \quad \alpha_8 = \big\{(1,1),(1,2),(2,1),(2,2),(3,1),(3,2)\big\}.$$

In this case we have

$$B_X(D) = \{\alpha_1, \alpha_2, \ldots, \alpha_7, \alpha_8\}; \quad B(\mathfrak{A}_2) = \{\alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7\}, \quad B(\mathfrak{A}_1) = \{\alpha_1, \alpha_8\}$$

and

$$\alpha_3 \circ \alpha_2 = \big\{(1,1),(2,1),(2,2)(3,1)\big\} \circ \big\{(1,1),(2,1),(3,1),(3,2)\big\} = \big\{(1,1),(2,1),(3,1)\big\} = \alpha_1,$$
$$\alpha_2 \circ \alpha_4 = \big\{(1,1),(2,1),(3,1),(3,2)\big\} \circ \big\{(1,1),(1,2)(2,1),(3,1)\big\}$$
$$= \big\{(1,1),(1,2),(2,1),(2,2),(3,1),(3,2)\big\} = \alpha_8,$$

i.e. the set $B(\mathfrak{A}_2)$ is irreducible generating set for the semigroup $B_X(D)$.

Example 2.2. $X = \{1, 2, 3, 4\}$ and $D = \{\{1\}, \{1, 2\}, \{1, 2, 3\}\}$. Then $4 = |X| > |D| = 3$ and

$$\alpha_1 = \big\{(1,1),(2,1),(3,1),(4,1)\big\}, \quad \alpha_2 = \big\{(1,1),(1,2),(2,1),(2,2),(3,1),(3,2),(4,1),(4,2)\big\},$$
$$\alpha_3 = \big\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),(4,1),(4,2),(4,3)\big\},$$
$$\alpha_4 = \big\{(1,1),(1,2),(2,1),(3,1),(4,1)\big\}, \quad \alpha_5 = \big\{(1,1),(2,1),(2,2),(3,1),(4,1)\big\},$$
$$\alpha_6 = \big\{(1,1),(2,1),(3,1),(3,2),(4,1)\big\}, \quad \alpha_7 = \big\{(1,1),(2,1),(3,1),(4,1),(4,2)\big\},$$
$$\alpha_8 = \big\{(1,1),(2,1),(3,1),(3,2),(4,1),(4,2)\big\}, \quad \alpha_9 = \big\{(1,1),(1,2),(2,1),(2,2),(3,1),(4,1)\big\},$$
$$\alpha_{10} = \big\{(1,1),(2,1),(2,2),(3,1),(4,1),(4,2)\big\}, \quad \alpha_{11} = \big\{(1,1),(1,2),(2,1),(3,1),(3,2),(4,1)\big\},$$
$$\alpha_{12} = \big\{(1,1),(2,1),(2,2),(3,1),(3,2),(4,1)\big\}, \quad \alpha_{13} = \big\{(1,1),(1,2),(2,1),(3,1),(4,1),(4,2)\big\},$$
$$\alpha_{14} = \big\{(1,1),(2,1),(2,2),(3,1),(3,2),(4,1),(4,2)\big\},$$
$$\alpha_{15} = \big\{(1,1),(1,2),(2,1),(3,1),(3,2),(4,1),(4,2)\big\},$$
$$\alpha_{16} = \big\{(1,1),(1,2),(2,1),(2,2),(3,1),(4,1),(4,2)\big\},$$
$$\alpha_{17} = \big\{(1,1),(1,2),(2,1),(2,2),(3,1),(3,2),(4,1)\big\},$$
$$\alpha_{18} = \big\{(1,1),(1,2),(1,3),(2,1),(3,1),(4,1)\big\}, \quad \alpha_{19} = \big\{(1,1),(2,1),(2,2),(2,3),(3,1),(4,1)\big\},$$
$$\alpha_{20} = \big\{(1,1),(2,1),(3,1),(3,2),(3,3),(4,1)\big\}, \ \alpha_{21} = \big\{(1,1),(2,1),(3,1),(4,1),(4,2),(4,3)\big\},$$
$$\alpha_{22} = \big\{(1,1),(2,1),(3,1),(3,2),(3,3),(4,1),(4,2),(4,3)\big\},$$
$$\alpha_{23} = \big\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(4,1)\big\},$$
$$\alpha_{24} = \big\{(1,1),(2,1),(2,2),(2,3),(3,1),(4,1),(4,2),(4,3)\big\},$$
$$\alpha_{25} = \big\{(1,1),(1,2),(1,3),(2,1),(3,1),(3,2),(3,3),(4,1)\big\},$$
$$\alpha_{26} = \big\{(1,1),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),(4,1)\big\},$$
$$\alpha_{27} = \big\{(1,1),(1,2),(1,3),(2,1),(3,1),(4,1),(4,2),(4,3)\big\},$$
$$\alpha_{28} = \big\{(1,1),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),(4,1),(4,2),(4,3)\big\},$$
$$\alpha_{29} = \big\{(1,1),(1,2),(1,3),(2,1),(3,1),(3,2),(3,3),(4,1),(4,2),(4,3)\big\},$$
$$\alpha_{30} = \big\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(4,1),(4,2),(4,3)\big\},$$
$$\alpha_{31} = \big\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),(4,1)\big\},$$
$$\alpha_{32} = \big\{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(3,2),(4,1),(4,2)\big\},$$

$$\alpha_{33} = \{(1,1),(1,2),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1),(4,2)\},$$
$$\alpha_{34} = \{(1,1),(1,2),(2,1),(2,2),(3,1),(3,2),(3,3),(4,1),(4,2)\},$$
$$\alpha_{35} = \{(1,1),(1,2),(2,1),(2,2),(3,1),(3,2),(4,1),(4,2),(4,3)\},$$
$$\alpha_{36} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1),(4,2)\},$$
$$\alpha_{37} = \{(1,1),(1,2),(2,1),(2,2),(3,1),(3,2),(3,3),(4,1),(4,2),(4,3)\},$$
$$\alpha_{38} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(3,2),(3,3),(4,1),(4,2)\},$$
$$\alpha_{39} = \{(1,1),(1,2),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1),(4,2),(4,3)\},$$
$$\alpha_{40} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(3,2),(4,1),(4,2),(4,3)\},$$
$$\alpha_{41} = \{(1,1),(1,2),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),(4,1),(4,2)\},$$
$$\alpha_{42} = \{(1,1),(1,2),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),(4,1),(4,2),(4,3)\},$$
$$\alpha_{43} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(3,2),(3,3),(4,1),(4,2),(4,3)\},$$
$$\alpha_{44} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1),(4,2),(4,3)\},$$
$$\alpha_{45} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),(4,1),(4,2)\},$$
$$\alpha_{46} = \{(1,1),(2,1),(2,2),(3,1),(3,2),(3,3),(4,1)\},$$
$$\alpha_{47} = \{(1,1),(2,1),(2,2),(3,1),(3,2),(3,3),(4,1),(4,2)\},$$
$$\alpha_{48} = \{(1,1),(2,1),(2,2),(3,1),(3,2),(3,3),(4,1),(4,2),(4,3)\},$$
$$\alpha_{49} = \{(1,1),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1)\},$$
$$\alpha_{50} = \{(1,1),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1),(4,2)\},$$
$$\alpha_{51} = \{(1,1),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1),(4,2),(4,3)\},$$
$$\alpha_{52} = \{(1,1),(1,2),(2,1),(3,1),(3,2),(3,3),(4,1)\},$$
$$\alpha_{53} = \{(1,1),(1,2),(2,1),(3,1),(3,2),(3,3),(4,1),(4,2)\},$$
$$\alpha_{54} = \{(1,1),(1,2),(2,1),(3,1),(3,2),(3,3),(4,1),(4,2),(4,3)\},$$
$$\alpha_{55} = \{(1,1),(1,2),(2,1),(2,2),(2,3),(3,1),(4,1)\},$$
$$\alpha_{56} = \{(1,1),(1,2),(2,1),(2,2),(2,3),(3,1),(4,1),(4,2)\},$$
$$\alpha_{57} = \{(1,1),(1,2),(2,1),(2,2),(2,3),(3,1),(4,1),(4,2),(4,3)\},$$
$$\alpha_{58} = \{(1,1),(1,2),(1,3),(2,1),(3,1),(3,2),(4,1)\},$$
$$\alpha_{59} = \{(1,1),(1,2),(1,3),(2,1),(3,1),(3,2),(4,1),(4,2)\},$$
$$\alpha_{60} = \{(1,1),(1,2),(1,3),(2,1),(3,1),(3,2),(4,1),(4,2),(4,3)\},$$
$$\alpha_{61} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(4,1)\},$$
$$\alpha_{62} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(4,1),(4,2)\},$$
$$\alpha_{63} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(4,1),(4,2),(4,3)\},$$
$$\alpha_{64} = \{(1,1),(2,1),(2,2),(3,1),(4,1),(4,2),(4,3)\},$$
$$\alpha_{65} = \{(1,1),(2,1),(2,2),(3,1),(3,2),(4,1),(4,2),(4,3)v\},$$
$$\alpha_{66} = \{(1,1),(2,1),(2,2),(2,3),(3,1),(4,1),(4,2)\},$$
$$\alpha_{67} = \{(1,1),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),(4,1),(4,2)\},$$
$$\alpha_{68} = \{(1,1),(1,2),(2,1),(3,1),(4,1),(4,2),(4,3)\},$$
$$\alpha_{69} = \{(1,1),(1,2),(2,1),(3,1),(3,2),(4,1),(4,2),(4,3)\},$$
$$\alpha_{70} = \{(1,1),(1,2),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1)\},$$
$$\alpha_{71} = \{(1,1),(1,2),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3),(4,1)\},$$
$$\alpha_{72} = \{(1,1),(1,2),(1,3),(2,1),(3,1),(4,1),(4,2)\},$$
$$\alpha_{73} = \{(1,1),(1,2),(1,3),(2,1),(3,1),(3,2),(3,3),(4,1),(4,2)\},$$

$$\alpha_{74} = \big\{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(3,2),(4,1)\big\},$$
$$\alpha_{75} = \big\{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1),(3,2),(3,3),(4,1)\big\},$$
$$\alpha_{76} = \big\{(1,1),(2,1),(3,1),(3,2),(4,1),(4,2),(4,3)\big\},$$
$$\alpha_{77} = \big\{(1,1),(2,1),(3,1),(3,2),(3,3),(4,1),(4,2)\big\},$$
$$\alpha_{78} = \big\{(1,1),(1,2),(2,1),(2,2),(3,1),(4,1),(4,2),(4,3)\big\},$$
$$\alpha_{79} = \big\{(1,1),(1,2),(2,1),(2,2),(3,1),(3,2),(3,3),(4,1)\big\},$$
$$\alpha_{80} = \big\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(4,1),(4,2)\big\},$$
$$\alpha_{81} = \big\{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1)\big\}.$$

In this case

$$B_X(D) = \{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_{80}, \alpha_{81}\}, \quad B(\mathfrak{A}_3) = \{\alpha_{46}, \alpha_{47}, \ldots, \alpha_{80}, \alpha_{81}\},$$
$$B(\mathfrak{A}_2) = \{\alpha_4, \alpha_5, \ldots, \alpha_{44}, \alpha_{45}\}, \quad B(\mathfrak{A}_1) = \{\alpha_1, \alpha_2, \alpha_3\}$$

and

$$\alpha_4 = \alpha_{72} \circ \alpha_{76}, \quad \alpha_5 = \alpha_{49} \circ \alpha_{76}, \quad \alpha_6 = \alpha_{46} \circ \alpha_{76}, \quad \alpha_7 = \alpha_{76} \circ \alpha_{76},$$
$$\alpha_8 = \alpha_{54} \circ \alpha_{76}, \quad \alpha_9 = \alpha_{80} \circ \alpha_{76}, \quad \alpha_{10} = \alpha_{51} \circ \alpha_{76}, \quad \alpha_{11} = \alpha_{52} \circ \alpha_{64},$$
$$\alpha_{12} = \alpha_{46} \circ \alpha_{64}, \quad \alpha_{13} = \alpha_{60} \circ \alpha_{76}, \quad \alpha_{14} = \alpha_{47} \circ \alpha_{64}, \quad \alpha_{15} = \alpha_{53} \circ \alpha_{64},$$
$$\alpha_{16} = \alpha_{56} \circ \alpha_{64}, \quad \alpha_{17} = \alpha_{79} \circ \alpha_{65}, \quad \alpha_{18} = \alpha_{72} \circ \alpha_{77}, \quad \alpha_{19} = \alpha_{49} \circ \alpha_{77}, \quad \alpha_{20} = \alpha_{46} \circ \alpha_{77},$$
$$\alpha_{21} = \alpha_{76} \circ \alpha_{77}, \quad \alpha_{22} = \alpha_{48} \circ \alpha_{77}, \quad \alpha_{23} = \alpha_{80} \circ \alpha_{77}, \quad \alpha_{24} = \alpha_{51} \circ \alpha_{77}, \quad \alpha_{25} = \alpha_{52} \circ \alpha_{49},$$
$$\alpha_{26} = \alpha_{46} \circ \alpha_{49}, \quad \alpha_{27} = \alpha_{60} \circ \alpha_{77}, \quad \alpha_{28} = \alpha_{47} \circ \alpha_{49}, \quad \alpha_{29} = \alpha_{53} \circ \alpha_{49}, \quad \alpha_{30} = \alpha_{78} \circ \alpha_{67},$$
$$\alpha_{31} = \alpha_{79} \circ \alpha_{67}, \quad \alpha_{32} = \alpha_{72} \circ \alpha_{79}, \quad \alpha_{33} = \alpha_{49} \circ \alpha_{53}, \quad \alpha_{34} = \alpha_{46} \circ \alpha_{52}, \quad \alpha_{35} = \alpha_{76} \circ \alpha_{79},$$
$$\alpha_{36} = \alpha_{80} \circ \alpha_{79}, \quad \alpha_{37} = \alpha_{48} \circ \alpha_{52}, \quad \alpha_{38} = \alpha_{52} \circ \alpha_{55}, \quad \alpha_{39} = \alpha_{51} \circ \alpha_{79}, \quad \alpha_{40} = \alpha_{60} \circ \alpha_{79},$$
$$\alpha_{41} = \alpha_{67} \circ \alpha_{79}, \quad \alpha_{42} = \alpha_{65} \circ \alpha_{71}, \quad \alpha_{43} = \alpha_{53} \circ \alpha_{55}, \quad \alpha_{44} = \alpha_{78} \circ \alpha_{71}, \quad \alpha_{45} = \alpha_{79} \circ \alpha_{71},$$
$$\alpha_1 = \alpha_7 \circ \alpha_7, \quad \alpha_2 = \alpha_7 \circ \alpha_{17}, \quad \alpha_3 = \alpha_{31} \circ \alpha_{31}.$$

References

[1] Ya. Diasamidze, Sh. Makharadze, Complete Semigroups of Binary Relations, Kriter, Turkey, 2013.

Further Reading

[1] Ya. Diasamidze, N. Aydin, A. Erdoğan, Generating set of the complete semigroups of binary relations, Appl. Math. Irvin 7 (2016) 98–107.
[2] Ya.I. Diasamidze, Sh.I. Makharadze, Complete semigroups of binary relations defined by elementary and nodal X-semilattices of unions. Algebra, 19, J. Math. Sci. (New York) 111 (2002), no. 1, 3171–3226.
[3] Ya.I. Diasamidze, T.T. Sirabidze, Complete semigroups of binary relations determined by three-element X-chains. Semigroups of binary relations, J. Math. Sci. (N.Y.) 117 (4) (2003) 4320–4350.
[4] O. Givradze, Irreducible generating sets of complete semigroups of unions $B_x(D)$ defined by semilattices of the class $\Sigma_2(X, 4)$. (Russian), Sovrem. Mat. Prilozh. (2011) 74 translation in J. Math. Sci. (N.Y.) 186 (2012) no. 5, 745–750.
[5] O. Givradze, Irreducible generating sets of complete semigroups of unions. (Russian), Sovrem. Mat. Prilozh. (2012) 84 translation in Part 3. J. Math. Sci. (N.Y.) 197 (2014) no. 6, 755–760.
[6] A. Bakuridze, Generated sets of the complete semigroup binari relations defined by semilattices of the class $\Sigma_1(X, 2)$, Int. J. Eng. Sci. Innovative Technol. (IJESIT) 5 (6) (2016) 17–26.
[7] Ya. Diasamidze, O. Givradze, A. Bakuridze, Generated sets of the complete semigroup binari relations defined by semilattices of the class $\Sigma_1(X, 3)$, Int. J. Eng. Sci. Innovative Technology (IJESIT) 5 (6) (2016) 52–69.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 388–403
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Bayesian inverse problems with partial observations

Shota Gugushvili*, Aad W. van der Vaart, Dong Yan

Mathematical Institute, Faculty of Science, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands

Received 2 March 2018; received in revised form 9 July 2018; accepted 8 September 2018
Available online 22 October 2018

Abstract

We study a nonparametric Bayesian approach to linear inverse problems under discrete observations. We use the discrete Fourier transform to convert our model into a truncated Gaussian sequence model, that is closely related to the classical Gaussian sequence model. Upon placing the truncated series prior on the unknown parameter, we show that as the number of observations $n \to \infty$, the corresponding posterior distribution contracts around the true parameter at a rate depending on the smoothness of the true parameter and the prior, and the ill-posedness degree of the problem. Correct combinations of these values lead to optimal posterior contraction rates (up to logarithmic factors). Similarly, the frequentist coverage of Bayesian credible sets is shown to be dependent on a combination of smoothness of the true parameter and the prior, and the ill-posedness of the problem. Oversmoothing priors lead to zero coverage, while undersmoothing priors produce highly conservative results. Finally, we illustrate our theoretical results by numerical examples.

Keywords: Credible set; Frequentist coverage; Gaussian prior; Gaussian sequence model; Heat equation; Inverse problem; Nonparametric Bayesian estimation; Posterior contraction rate; Singular value decomposition; Volterra operator

1. Introduction

Linear inverse problems have been studied since long in the statistical and numerical analysis literature; see, e.g., [1–9], and references therein. Emphasis in these works has been on the signal-in-white noise model,

$$Y = Af + \varepsilon W, \tag{1}$$

where the parameter of interest f lies in some infinite-dimensional function space, A is a linear operator with values in a possibly different space, W is white noise, and ε is the noise level. Applications of linear inverse problems include, e.g., computerized tomography, see [10], partial differential equations, see [11], and scattering theory, see [12].

* Corresponding author.
E-mail addresses: shota.gugushvili@math.leidenuniv.nl (S. Gugushvili), avdvaart@math.leidenuniv.nl (A.W. van der Vaart), d.yan@math.leidenuniv.nl (D. Yan).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.09.002

Arguably, in practice one does not have access to a full record of observations on the unknown function f as in the idealized model (1), but rather one indirectly observes it at a finite number of points. This statistical setting can be conveniently formalized as follows: let the signal of interest f be an element in a Hilbert space H_1 of functions defined on a compact interval $[0, 1]$. The forward operator A maps f to another Hilbert space H_2. We assume that H_1, H_2 are subspaces of $L^2([0, 1])$, typically collections of functions of certain smoothness as specified in the later sections, and that the design points are chosen deterministically,

$$\left\{x_i = \frac{i}{n}\right\}_{i=1,\dots,n}. \tag{2}$$

Assuming continuity of Af and defining

$$Y_i = Af(x_i) + \xi_i, \quad i = 1, \dots, n, \tag{3}$$

with ξ_i i.i.d. standard Gaussian random variables, our observations are the pairs $(x_i, Y_i)_{i\leq n}$, and we are interested in estimating f. A prototype example we think of is the case when A is the solution operator in the Dirichlet problem for the heat equation acting on the initial condition f; see Example 2.8 for details.

Model (3) is related to the inverse regression model studied e.g. in [13] and [14]. Although the setting we consider is somewhat special, our contribution is arguably the first one to study from a theoretical point of view a nonparametric Bayesian approach to estimation of f in the inverse problem setting with partial observations (see [15] for a monographic treatment of modern Bayesian nonparametrics). In the context of the signal-in-white noise model (1), a nonparametric Bayesian approach has been studied thoroughly in [16] and [17], and techniques from these works will turn out to be useful in our context as well. Our results will deal with derivation of posterior contraction rates and study of asymptotic frequentist coverage of Bayesian credible sets. A posterior contraction rate can be thought of as a Bayesian analogue of a convergence rate of a frequentist estimator, cf. [18] and [15]. Specifically, we will show that as the sample size $n \to \infty$, the posterior distribution concentrates around the 'true' parameter value, under which data have been generated, and hence our Bayesian approach is consistent and asymptotically recovers the unknown 'true' f. The rate at which this occurs will depend on the smoothness of the true parameter and the prior and the ill-posedness degree of the problem. Correct combinations of these values lead to optimal posterior contraction rates (up to logarithmic factors). Furthermore, a Bayesian approach automatically provides uncertainty quantification in parameter estimation through the spread of the posterior distribution, specifically by means of posterior credible sets. We will give an asymptotic frequentist interpretation of these sets in our context. In particular, we will see that the frequentist coverage will depend on a combination of smoothness of the true parameter and the prior, and the ill-posedness of the problem. Oversmoothing priors lead to zero coverage, while undersmoothing priors produce highly conservative results.

The article is organized as follows: in Section 2, we give a detailed description of the problem, introduce the singular value decomposition and convert the model (3) into an equivalent truncated sequence model that is better amenable to our theoretical analysis. We show how a Gaussian prior in this sequence model leads to a Gaussian posterior and give an explicit characterization of the latter. Our main results on posterior contraction rates and Bayesian credible sets are given in Section 3, followed by simulation examples in Section 4 that illustrate our theoretical results. Section 5 contains the proofs of the main theorems, while the technical lemmas used in the proofs are collected in Section 6.

1.1. Notation

The notational conventions we use in this work are the following: definitions are marked by the $:=$ symbol; $|\cdot|$ denotes the absolute value and $\|\cdot\|_H$ indicates the norm related to the space H; $\langle\cdot,\cdot\rangle_H$ is understood as the canonical inner product in the inner product space H; subscripts are omitted when there is no danger of confusion; $\mathcal{N}(\mu, \Sigma)$ denotes the Gaussian distribution with mean μ and covariance operator Σ; subscripts $\mathcal{N}_n$ and $\mathcal{N}_H$ may be used to emphasize the fact that the distribution is defined on the space $\mathbb{R}^n$ or on the abstract space H; $\mathrm{Cov}(\cdot,\cdot)$ denotes the covariance or the covariance operator, depending on the context; for positive sequences $\{a_n\}$, $\{b_n\}$ of real numbers, the notation $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ mean respectively that there exist positive constants C_1, C_2 independent of n, such that $a_n \leq C_1 b_n$ or $a_n \geq C_2 b_n$ hold for all n; finally, $a_n \asymp b_n$ indicates that the ratio a_n/b_n is asymptotically bounded from zero and infinity, while $a_n \sim b_n$ means $a_n/b_n \to 1$ as $n \to \infty$.

2. Sequence model

2.1. Singular value decomposition

We impose a common assumption on the forward operator A from the literature on inverse problems, see, e.g., [1,2] and [3].

Assumption 2.1. Operator A is injective and compact.

It follows that A^*A is also compact and in addition self-adjoint. Hence, by the spectral theorem for self-adjoint compact operators, see [19], we have a representation $A^*Af = \sum_{k\in\mathbb{N}} a_k^2 f_k \varphi_k$, where $\{\varphi_k\}$ and $\{a_k\}$ are the eigenbasis on H_1 and eigenvalues, respectively, (corresponding to the operator A^*A), and $f_k = \langle f, \varphi_k\rangle$ are the Fourier coefficients of f. This decomposition of A^*A is known as the singular value decomposition (SVD), and $\{a_k\}$ are also called singular values.

It is easy to show that the conjugate basis $\psi_k := A\varphi_k/a_k$ of the orthonormal basis $\{\varphi_k\}_k$ is again an orthonormal system in H_2 and gives a convenient basis for $Range(A)$, the range of A in H_2. Furthermore, the following relations hold (see [1]),

$$A\varphi_k = a_k\psi_k,\ A^*\psi_k = a_k\varphi_k. \tag{4}$$

Recall a standard result (see, e.g., [20]): a Hilbert space H is isometric to ℓ^2, and Parseval's identity $\|f\|_{\ell^2}^2 := \sum_k |f_k|^2 = \|f\|_H^2$ holds; here f_k are the Fourier coefficients with respect to some known and fixed orthonormal basis.

We will employ the eigenbasis $\{\varphi_k\}$ of A^*A to define the Sobolev space of functions. This will define the space in which the unknown function f resides.

Definition 2.2. We say f is in the Sobolev space S^β with smoothness parameter $\beta \geq 0$, if it can be written as $f = \sum_{k=1}^\infty f_k\varphi_k$ with $f_k = \langle f, \varphi_k\rangle$, and if its norm $\|f\|_\beta := \left(\sum_{k=1}^\infty f_k^2 k^{2\beta}\right)^{1/2}$ is finite.

Remark 2.3. The above definition agrees with the classical definition of the Sobolev space if the eigenbasis is the trigonometric basis, see, e.g., [21]. With a fixed basis, which is always the case in this article, one can identify the function f and its Fourier coefficients $\{f_k\}$. Thus, we use S^β to denote both the function space and the sequence space. For example, it is easy to verify that $S^0 = \ell^2$ (correspondingly $S^0 = L^2$), $S^\beta \subset \ell^2$ for any nonnegative β, and $S^\beta \subset \ell^1$ when $\beta > 1/2$.

Recall that $Af = \sum a_i f_i \psi_i$. Then we have $Af \in S^{\beta+p}$ if $a_k \asymp k^{-p}$, and $Af \in S^\infty := \cap_{k\in\mathbb{N}} S^k$, if a_k decays exponentially fast. Such a lifting property is beneficial in the forward problem, since it helps to obtain a smooth solution. However, in the context of inverse problems it leads to a difficulty in recovery of the original signal f, since information on it is washed out by smoothing. Hence, in the case of inverse problems one does not talk of the lifting property, but of ill-posedness, see [3].

Definition 2.4. An inverse problem is called mildly ill-posed, if $a_k \asymp k^{-p}$ as $k \to \infty$, and extremely ill-posed, if $a_k \asymp e^{-k^s p}$ with $s \geq 1$ as $k \to \infty$, where p is strictly positive in both cases.

In the rest of the article, we will confine ourselves to the following setting.

Assumption 2.5. The unknown true signal f in (3) satisfies $f \in S^\beta \subset H_1$ for $\beta > 0$. Furthermore, the ill-posedness is of one of the two types in Definition 2.4.

Remark 2.6. As an immediate consequence of the lifting property, we have $H_2 \subset H_1$.

We conclude this section with two canonical examples of the operator A.

Example 2.7 (*Mildly Ill-Posed Case: Volterra Operator [16]*)**.** The classical Volterra operator $A : L^2[0,1] \to L^2[0,1]$ and its adjoint A^* are

$$Af(x) = \int_0^x f(s)\,ds, \quad A^*f(x) = \int_x^1 f(s)\,ds.$$

The eigenvalues, eigenfunctions of A^*A and the conjugate basis are given by

$$\begin{aligned} a_i^2 &= \frac{1}{(i-1/2)^2\pi^2}, \\ \varphi_i(x) &= \sqrt{2}\cos((i-1/2)\pi x), \\ \psi_i(x) &= \sqrt{2}\sin((i-1/2)\pi x), \end{aligned}$$

for $i \geq 1$.

Example 2.8 (*Extremely Ill-Posed Case: Heat Equation [17]*)**.** Consider the Dirichlet problem for the heat equation:

$$\begin{aligned} &\frac{\partial}{\partial t}u(x,t) = \frac{\partial^2}{\partial x^2}u(x,t), \quad u(x,0) = f(x), \\ &\quad u(0,t) = u(1,t) = 0, \quad t \in [0,T], \end{aligned} \tag{5}$$

where $u(x,t)$ is defined on $[0,1]\times[0,T]$ and $f(x) \in L^2[0,1]$ satisfies $f(0) = f(1) = 0$. The solution of (5) is given by

$$u(x,t) = \sqrt{2}\sum_{k=1}^{\infty} f_k e^{-k^2\pi^2 t}\sin(k\pi x) =: Af(x),$$

where $\{f_k\}$ are the coordinates of f in the basis $\{\sqrt{2}\sin(k\pi x)\}_{k\geq 1}$.

For the solution map A, the eigenvalues of A^*A are $e^{-k^2\pi^2 t}$, the eigenbasis and conjugate basis coincide and $\varphi_k(x) = \psi_k(x) = \sqrt{2}\sin(k\pi x)$.

2.2. Equivalent formulation

In this subsection we develop a sequence formulation of the model (3), which is very suitable for asymptotic Bayesian analysis. First, we briefly discuss the relevant results that provide motivation for our reformulation of the problem.

In Examples 2.7 and 2.8, the sine and cosine bases form the eigenbasis. In fact, the Fourier basis (trigonometric polynomials) frequently arises as an eigenbasis for various operators, e.g. in the case of differentiation, see [22], or circular deconvolution, see [4]. For simplicity, we will use Fourier basis as a primary example in the rest of the article. Possible generalization to other bases is discussed in Remark 2.10.

Restriction of our attention to the Fourier basis is motivated by its special property: discrete orthogonality. The next lemma illustrates this property for the sine basis (Example 2.8).

Lemma 2.9 (*Discrete Orthogonality*)**.** *Let $\{\psi_k\}_{k\in\mathbb{N}}$ be the sine basis, i.e.*

$$\psi_k(x) = \sqrt{2}\sin(k\pi x), \quad k = 1,2,3,\ldots.$$

Then:

(i.) *Discrete orthogonality holds:*

$$\langle \psi_j, \psi_k\rangle_d := \frac{1}{n}\sum_{i=1}^{n}\psi_j(i/n)\psi_k(i/n) = \delta_{jk}, \quad j,k = 1,\ldots,n-1. \tag{6}$$

Here δ_{jk} is the Kronecker delta.

(ii.) *Fix $l \in \mathbb{N}$. For any fixed $1 \leq k \leq n-1$ and all $j \in \{ln, ln+1,\ldots,(l+1)n-1\}$, there exists only one $\bar{k} \in \{1,2,\ldots,n-1\}$ depending only on the parity of l, such that for $\tilde{j} = ln + \bar{k}$, the equality*

$$|\langle \psi_{\tilde{j}}, \psi_k\rangle_d| = 1 \tag{7}$$

holds, while $\langle \psi_{\tilde{j}}, \psi_k\rangle_d = 0$ for all $\tilde{j} = ln + \tilde{k}$ such that $\tilde{k} \neq \bar{k}, \tilde{k} \in \{1,2,\ldots,n-1\}$.

Remark 2.10. For other trigonometric bases, discrete orthogonality can also be attained. Thus, the conjugate eigenbasis in Example 2.7 is discretely orthogonal with design points $\{(i-1/2)/n\}_{i=1,\dots,n}$. We refer to [23] and references therein for details. With some changes in the arguments, our asymptotic statistical results still remain valid with such modifications of design points compared to (2). We would like to stress the fact that restricting attention to bases with discrete orthogonality property does constitute a loss of generality. However, there exist classical bases other than trigonometric bases that are discretely orthogonal (possibly after a suitable modification of design points). See, for instance, [24] for an example of Lagrange polynomials.

Motivated by the observations above, we introduce our central assumption on the basis functions.

Assumption 2.11. Given the design points $\{x_i\}_{i=1,\dots,n}$ in (2), we assume the conjugate basis $\{\psi_k\}_{k\in\mathbb{N}}$ of the operator A in (3) possesses the following properties:

(i.) for $1 \le j,k \le n-1$,

$$\langle \psi_j(x), \psi_k(x)\rangle_d := \frac{1}{n}\sum_{i=1}^{n} \psi_j(x_i)\psi_k(x_i). = \delta_{jk}$$

(ii.) For $1 \le k \le n-1$ and $j \in \{ln,\dots,(l+1)n-1\}$ with fixed $l \in \mathbb{N}$, there exists only one $\tilde{j} = ln + \bar{k}$, such that $0 < |\langle \psi_{\tilde{j}}, \psi_k\rangle_d| < M$, where M is a fixed constant, and $\bar{k}$ depends on the parity of l only. For other $j \neq \tilde{j}$, $|\langle \psi_j, \psi_k\rangle_d| = 0$.

Using the shorthand notation

$$f = \sum_j f_j\varphi_j = \sum_{j=1}^{n-1} f_j\varphi_j + \sum_{j\ge n} f_j\varphi_j =: f^n + f^r,$$

we obtain for $k = 1,\dots,n-1$ that

$$\begin{aligned} U_k &= \frac{1}{n}\sum_{i=1}^{n} Y_i\psi_k(x_i) = \langle Af^n, \psi_k\rangle_d + \langle Af^r, \psi_k\rangle_d + \frac{1}{n}\sum_{i=1}^{n}\xi_i\psi_k(x_i) \\ &= a_k f_k + R_k + \frac{1}{\sqrt{n}}\zeta_k, \end{aligned} \tag{8}$$

where

$$R_k := R_k(f) = \langle Af^r, \psi_k\rangle_d, \quad \zeta_k := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\psi_k(x_i).$$

By Assumption 2.11, we have

$$|R_k| = |\langle Af^r, \psi_k\rangle_d| \le \sum_{j\ge n} a_j|f_j||\langle \psi_j, \psi_k\rangle_d| = \sum_{l=1}^{\infty} a_{ln+\bar{k}}|f_{ln+\bar{k}}|, \tag{9}$$

which leads to (via Cauchy–Schwarz)

$$R_k^2(f) \le \Big(\sum_{l=1}^{\infty} a_{ln+\bar{k}}^2 (ln+\bar{k})^{-2\beta}\Big)\|f\|_\beta^2.$$

Hence, for a mildly ill-posed problem, i.e. $a_k \asymp k^{-p}$, the following bound holds, uniformly in the ellipsoid $\{f : \|f\|_\beta \le K\}$,

$$\begin{aligned} \sup_{f:\|f\|_\beta\le K} R_k^2(f) &\lesssim \sum_{l=1}^{\infty}(ln)^{-2\beta-2p} = n^{-2(\beta+p)}\sum_{l=1}^{\infty} l^{-2(\beta+p)} \\ &\asymp n^{-2(\beta+p)} = o(1/n), \end{aligned} \tag{10}$$

for any $1 \le k \le n-1$ when $\beta + p > 1/2$.

If the problem is extremely ill-posed, i.e. $a_k \asymp e^{-k^s p}$, we use the inequality

$$R_k^2(f) \leq \left(\sum_{j \geq n} a_j |f_j| \right)^2 \leq (\sum_{j \geq n} a_j^2) \| f^r \|^2.$$

Since $a_j \asymp \exp(-pj^s) \leq \exp(-pj)$, it follows that $\sum_{j \geq n} a_j^2$ is up to a constant bounded from above by $\exp(-2pn)$. Hence

$$\sup_{f : \|f\|_\beta \leq K} R_k^2(f) \lesssim \exp(-2pn) \ll o(1/n). \tag{11}$$

In [16,17], the Gaussian prior $\Pi = \otimes_{i \in \mathbb{N}} \mathcal{N}(0, \lambda_i)$ is employed on the coordinates of the eigenbasis expansion of f. If $\lambda_i = \rho_n^2 i^{-1-2\alpha}$, the sum $\sum_{i \in \mathbb{N}} \lambda_i = \rho_n^2 \sum_{i \in \mathbb{N}} i^{-1-2\alpha}$ is convergent, and hence this prior is the law of a Gaussian element in H_1.

In our case, we consider the same type of the prior with an additional constraint that only the first $n-1$ components of the prior are non-degenerate, i.e. $\Pi = (\otimes_{i<n} \mathcal{N}(0, \lambda_i)) \times \left(\otimes_{i \geq n} \mathcal{N}(0, 0) \right)$, where λ_i is as above. In addition, we assume the prior on f is independent of the noise ζ_k, $k = 1, \ldots, n-1$, in (8). With these assumptions in force, we see $\Pi(R_k = 0) = 1$, for $k = 1, \ldots, n-1$. Furthermore, the posterior can be obtained from the product structure of the model and the prior via the normal conjugacy,

$$\begin{aligned} &\Pi(f|U^n) = \otimes_{k \in \mathbb{N}} \mathcal{N}(\hat{f}_k, \sigma_k^2), \\ &\text{with} \quad \hat{f}_k = \frac{n a_k \lambda_k \mathbb{1}_{\{k<n\}}}{n a_k^2 \lambda_k + 1} U_k, \quad \sigma_k^2 = \frac{\lambda_k \mathbb{1}_{\{k<n\}}}{n a_k^2 \lambda_k + 1}. \end{aligned} \tag{12}$$

We also introduce

$$\hat{f} = \mathbb{E}(f|U^n) = (\mathbb{E}(f_k|U_k)) = (\hat{f}_k)_{k \in \mathbb{N}} = (b_k U_k)_{k \in \mathbb{N}}, \tag{13}$$

where $b_k = \frac{n a_k \lambda_k \mathbb{1}_{\{k<n\}}}{n a_k^2 \lambda_k + 1}$. We conclude this section with a useful fact that will be applied in later sections:

$$\hat{f}_k = b_k U_k = b_k \left(a_k f_k + R_k + \frac{\zeta_k}{\sqrt{n}} \right) = \mathbb{E} \hat{f}_k + \tau_k \zeta_k, \tag{14}$$

where $\mathbb{E} \hat{f}_k = a_k b_k f_k + b_k R_k$ and $\tau_k = b_k / \sqrt{n}$.

3. Main results

3.1. Contraction rates

In this section, we determine the rate at which the posterior distribution concentrates on shrinking neighbourhoods of the 'true' parameter f_0 as the sample size n grows to infinity.

Assume the observations in (3) have been collected under the parameter value $f_0 = \sum_{k \in \mathbb{N}} f_{0,k} \varphi_k$. Thus our observations $(U_k)_{k<n}$ given in (8) have the law $\otimes_{k<n} \mathcal{N}(a_k f_{0,k} + R_k, 1/n)$. We will use the notation $\Pi_n(\cdot|U)$ to denote the posterior distribution given in (12).

Theorem 3.1 (*Posterior Contraction: Mildly Ill-Posed Problem*)**.** *If the problem is mildly ill-posed as $a_k \asymp k^{-p}$ with $p > 0$, the true parameter $f_0 \in S^\beta$ with $\beta > 0$, and furthermore $\beta + p > 1/2$, by letting $\lambda_k = \rho_n^2 k^{-1-2\alpha}$ with $\alpha > 0$ and any positive ρ_n satisfying $\rho_n^2 n \to \infty$, we have, for any $K > 0$ and $M_n \to \infty$,*

$$\sup_{\|f_0\|_\beta \leq K} \mathbb{E}_{f_0} \Pi_n \left(f : \|f - f_0\|_{H_1} \geq M_n \varepsilon_n | U^n \right) \to 0,$$

where

$$\varepsilon_n = \varepsilon_{n,1} \vee \varepsilon_{n,2} = (\rho_n^2 n)^{-\beta/(2\alpha+2p+1) \wedge 1} \vee \rho_n (\rho_n^2 n)^{-\alpha/(2\alpha+2p+1)}. \tag{15}$$

In particular,

(i.) *if* $\rho_n = 1$, *then* $\varepsilon_n = n^{-(\alpha\wedge\beta)/(2\alpha+2p+1)}$;
(ii.) *if* $\beta \leq 2\alpha + 2p + 1$ *and* $\rho_n \asymp n^{(\alpha-\beta)/(2\beta+2p+1)}$, *then* $\varepsilon_n = n^{-\beta/(2\beta+2p+1)}$;
(iii.) *if* $\beta > 2\alpha + 2p + 1$, *then for every scaling* ρ_n, $\varepsilon_n \gg n^{-\beta/(2\beta+2p+1)}$.

Thus we recover the same posterior contraction rates as obtained in [16], at the cost of an extra constraint $\beta + p > 1/2$. The frequentist minimax convergence rate for mildly ill-posed problems in the white noise setting with $\varepsilon = n^{-1/2}$ is $n^{-\beta/(2\beta+2p+1)}$, see [3]. We will compare our result to this rate. Our theorem states that in case (i.) the posterior contraction rate reaches the frequentist optimal rate if the regularity of the prior matches the truth ($\beta = \alpha$) and the scaling factor ρ_n is fixed. Alternatively, as in case (ii.), the optimal rate can also be attained by proper scaling, provided a sufficiently regular prior is used. In all other cases the contraction rate is slower than the minimax rate. Our results are similar to those in [16] in the white noise setting. The extra constraint $\beta + p > 1/2$ that we have in comparison to that work demands an explanation. As (10) shows, the size of negligible terms $R_k(f_0)$ in (8) decreases as the smoothness $\beta + p$ of the transformed signal Af_0 increases. In order to control R_k, a minimal smoothness of Af_0 is required. The latter is guaranteed if $p + \beta \geq 1/2$, for it is known that in that case Af_0 will be at least continuous, while it may fail to be so if $p + \beta < 1/2$, see [21].

Remark 3.2. The control on $R_k(f_0)$ from (9) depends on the fact that the eigenbasis possesses the properties in Assumption 2.11. If instead of Assumption 2.11(ii.) one only assumes $|\langle\psi_j, \psi_k\rangle| \leq 1$ for any $k \leq n-1$ and $j \geq n$, the constraint on the smoothness of Af_0 has to be strengthened to $\beta + p \geq 1$ in order to obtain the same results as in Theorem 3.1, because the condition $\beta + p \geq 1$ guarantees that the control on $R_k(f_0)$ in (10) remains valid.

Now we consider the extremely ill-posed problem. The following result holds.

Theorem 3.3 (*Posterior Contraction: Extremely Ill-Posed Problem*)**.** *Let the problem be extremely ill-posed as* $a_k \asymp e^{-pk^s}$ *with* $s \geq 1$, *and let the true parameter* $f_0 \in S^\beta$ *with* $\beta > 0$. *Let* $\lambda_k = \rho_n^2 k^{-1-2\alpha}$ *with* $\alpha > 0$ *and any positive* ρ_n *satisfying* $\rho_n^2 n \to \infty$. *Then*

$$\sup_{\|f_0\|_\beta \leq K} \mathbb{E}_{f_0} \Pi_n\left(f : \|f - f_0\|_{H_1} \geq M_n \varepsilon_n | U^n\right) \to 0,$$

for any $K > 0$ *and* $M_n \to \infty$, *where*

$$\varepsilon_n = \varepsilon_{n,1} \vee \varepsilon_{n,2} = \left(\log(\rho_n^2 n)\right)^{-\beta/s} \vee \rho_n \left(\log(\rho_n^2 n)\right)^{-\alpha/s}. \tag{16}$$

In particular,

(i.) *if* $\rho_n = 1$, *then* $\varepsilon_n = (\log n)^{-(\alpha\wedge\beta)/s}$,
(ii.) *if* $n^{-1/2+\delta} \lesssim \rho_n \lesssim (\log n)^{(\alpha-\beta)/s}$ *for some* $\delta > 0$, *then* $\varepsilon_n = (\log n)^{-\beta/s}$.

Furthermore, if $\lambda_k = \exp(-\alpha k^s)$ *with* $\alpha > 0$, *the following contraction rate is obtained:* $\varepsilon_n = (\log n)^{-\beta/s}$.

Since the frequentist minimax estimation rate in extremely ill-posed problems in the white noise setting is $(\log n)^{-\beta/s}$ (see [3]), Theorem 3.3 shows that the optimal contraction rates can be reached by suitable choice of the regularity of the prior, or by using an appropriate scaling. In contrast to the mildly ill-posed case, we have no extra requirement on the smoothness of Af_0. The reason is obvious: because the signal is lifted to S^∞ by the forward operator A, the term (11) converges to zero exponentially fast, implying that $R_k(f_0)$ in (8) is always negligible.

3.2. Credible sets

In the Bayesian paradigm, the spread of the posterior distribution is a common measure of uncertainty in parameter estimates. In this section we study the frequentist coverage of Bayesian credible sets in our problem.

When the posterior is Gaussian, it is customary to consider credible sets centred at the posterior mean, which is what we will also do. In addition, because in our case the covariance operator of the posterior distribution does not depend on the data, the radius of the credible ball is determined by the credibility level $1 - \gamma$ and the sample size n.

A credible ball centred at the posterior mean $\hat{f}$ from (13) is given by

$$\hat{f} + B(r_{n,\gamma}) := \{f \in H_1 : \|f - \hat{f}\|_{H_1} \leq r_{n,\gamma}\}, \tag{17}$$

where the radius $r_{n,\gamma}$ is determined by the requirement that

$$\Pi_n(\hat{f} + B(r_{n,\gamma})|U^n) = 1 - \gamma. \tag{18}$$

By definition, the frequentist coverage or confidence of the set (17) is

$$\mathbb{P}_{f_0}(f_0 \in \hat{f} + B(r_{n,\gamma})), \tag{19}$$

where the probability measure is the one induced by the law of U^n given in (8) with $f = f_0$. We are interested in the asymptotic behaviour of the coverage (19) as $n \to \infty$ for a fixed f_0 uniformly in Sobolev balls, and also along a sequence f_0^n changing with n.

The following two theorems hold.

Theorem 3.4 (*Credible Sets: Mildly Ill-Posed Problem*)**.** *Assume the same assumptions as in Theorem* 3.1 *hold, and let* $\tilde{\beta} = \beta \wedge (2\alpha + 2p + 1)$. *The asymptotic coverage of the credible set* (17) *is*

(i.) *1, uniformly in* $\{f_0 : \|f_0\|_\beta \leq 1\}$, *if* $\rho_n \gg n^{(\alpha-\tilde{\beta})/(2\tilde{\beta}+2p+1)}$;
(ii.) *1, for every fixed* $f_0 \in S^\beta$, *if* $\beta < 2\alpha+2p+1$ *and* $\rho_n \asymp n^{(\alpha-\tilde{\beta})/(2\tilde{\beta}+2p+1)}$; *c, along some* f_0^n *with* $\sup_n \|f_0^n\|_\beta < \infty$, *if* $\rho_n \asymp n^{(\alpha-\tilde{\beta})/(2\tilde{\beta}+2p+1)}$ *(any* $c \in [0, 1)$*).*
(iii.) *0, along some* f_0^n *with* $\sup_n \|f_0^n\|_\beta < \infty$, *if* $\rho_n \ll n^{(\alpha-\tilde{\beta})/(2\tilde{\beta}+2p+1)}$.

Theorem 3.5 (*Credible Sets: Extremely Ill-Posed Problem*)**.** *Assume the setup of Theorem* 3.3. *Then if* $\lambda_k = \rho_n^2 k^{-1-2\alpha}$ *with* $\alpha > 0$ *and any positive* ρ_n *satisfying* $\rho_n^2 n \to \infty$, *the asymptotic coverage of the credible set* (17) *is*

(i.) *1, uniformly in* $\{f_0 : \|f_0\|_{S^\beta} \leq 1\}$, *if* $\rho_n \gg (\log n)^{(\alpha-\beta)/2}$;
(ii.) *1, uniformly in* f_0 *with* $\|f_0\|_\beta \leq r$ *with r small enough;*
1, for any fixed $f_0 \in S^\beta$,
provided the condition $\rho_n \asymp (\log n)^{(\alpha-\beta)/s}$ *holds;*
(iii.) *0, along some* f_0^n *with* $\sup_n \|f_0^n\|_\beta < \infty$, *if* $\rho_n \lesssim (\log n)^{(\alpha-\beta)/s}$.

Moreover, if $\lambda_k = e^{-\alpha^s}$ *with* $\alpha > 0$ *and any positive* ρ_n *satisfying* $\rho_n^2 n \to \infty$, *the asymptotic coverage of the credible set* (17) *is*

(iv.) *0, for every* f_0 *such that* $|f_{0,i}| \gtrsim e^{-ci^s/2}$ *for some* $c < \alpha$.

For the two theorems in this section, the most intuitive explanation is offered by the case $\rho_n \equiv 1$. The situations (i.), (ii.) and (iii.) correspond to $\alpha < \beta$, $\alpha = \beta$ and $\alpha > \beta$, respectively. The message is that the oversmoothing prior ((iii.) in Theorem 3.4 and (iii.), (iv.) in Theorem 3.5) leads to disastrous frequentist coverage of credible sets, while the undersmoothing prior ((i.) in both theorems) delivers very conservative frequentist results (coverage 1). With the right regularity of the prior (case (ii.)), the outcome depends on the norm of the true parameter f_0. Our results are thus similar to those obtained in the white noise setting in [16] and [17].

4. Simulation examples

In this section we carry out a small-scale simulation study illustrating our theoretical results. Examples we use to that end are those given in Section 2.1. These were also used in simulations in [16] and [17].

In the setting of Example 2.7, we use the following true signal,

$$f_0(x) = \sum_{i=1}^{\infty} f_{0,i}\varphi_i(x) \text{ with } f_{0,k} = k^{-3/2}\sin(k). \tag{20}$$

It is easy to check that $f_0 \in S^1$.

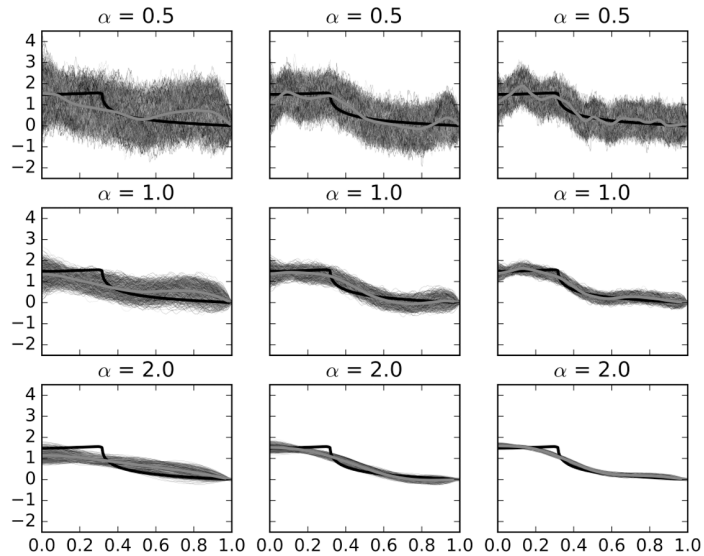


Fig. 1. Realizations of the posterior mean (red) and 950 of 1000 draws from the posterior (coloured thin lines) with smallest L^2 distance to the posterior mean. From left to right columns, the posterior is computed based on sample size 10^3, 10^4 and 10^5 respectively. The true parameter (black) is of smoothness $\beta = 1$ and given by coefficients $f_{0,k} = k^{-3/2}\sin(k)$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In the setup of Example 2.8, the initial condition is assumed to be

$$f_0(x) = 4x(x-1)(8x-5). \tag{21}$$

One can verify that in this case

$$f_{0,k} = \frac{8\sqrt{2}(13 + 11(-1)^k)}{\pi^3 k^3},$$

and $f_0 \in S^\beta$ for any $\beta < 5/2$.

First, we generate noisy observations $\{Y_i\}_{i=1,\dots,n}$ from our observation scheme (3) at design points $x_i = \frac{i-1/2}{n}$ in the case of Volterra operator, and $x_i = i/n$ in the case of the heat equation. Next, we apply the transform described in (8) and obtain transformed observations $\{U_i\}_{i=1,\dots,n-1}$. Then, by (12), the posterior of the coefficients with the eigenbasis φ_i is given by

$$f_k|U^n \sim \mathcal{N}\left(\frac{na_k\lambda_k \mathbb{1}_{\{k<n\}}}{na_k^2\lambda_k + 1}U_k, \frac{\lambda_k \mathbb{1}_{\{k<n\}}}{na_k^2\lambda_k + 1}\right).$$

Figs. 1 and 2 display plots of 95% L_2-credible bands for different sample sizes and different priors. For all priors we assume $\rho_n \equiv 1$, and use different smoothness degrees α, as shown in the titles of the subplots. In addition, the columns from left to right correspond to 10^3, 10^4 and 10^5 observations. The (estimated) credible bands are obtained by generating 1000 realizations from the posterior and retaining 95% of them that are closest in the L^2-distance to the posterior mean.

Two simulations reflect several similar facts. First, because of the difficulty due to the inverse nature of the problem, the recovery of the true signal is relatively slow, as the posteriors for the sample size 10^3 are still rather diffuse around the true parameter value. Second, it is evident that undersmoothing priors (the top rows in the figures) deliver conservative credible bands, but still capture the truth. On the other hand, oversmoothing priors lead to overconfident, narrow bands, failing to actually express the truth (bottom rows in the figures). As already anticipated due to a greater degree of ill-posedness, recovery of the initial condition in the heat equation case is more difficult than recovery of the true function in the case of the Volterra operator. Finally, we remark that qualitative behaviour of the posterior in our examples is similar to the one observed in [16] and [17]; for larger samples sizes n, discreteness of the observation scheme does not appear to have a noticeably adversary effect compared to the fully observed case in [16] and [17].

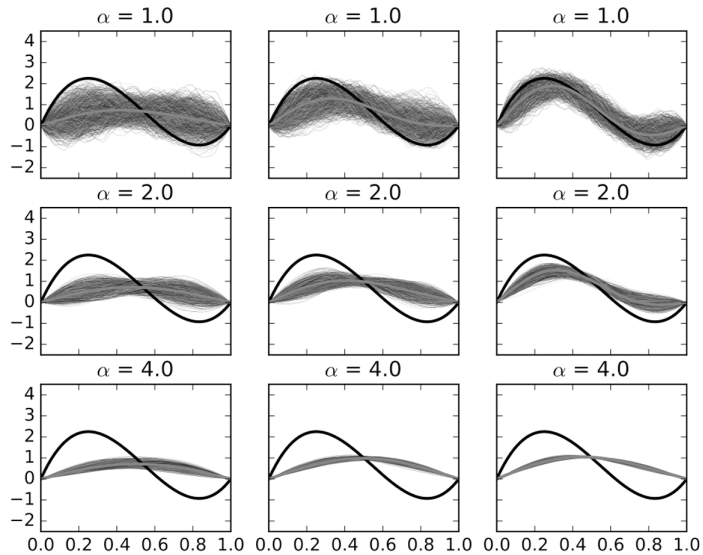


Fig. 2. Realizations of the posterior mean (red) and 950 of 1000 draws from the posterior (coloured thin lines) with smallest L^2 distance to the posterior mean. From left to right columns, the posterior is computed based on sample size 10^3, 10^4 and 10^5 respectively. The true parameter (black) is of smoothness β for any $\beta < 5/2$ and given by (21). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Proofs

5.1. Proof of Lemma 2.9

This proof is a modification of the one of Lemma 1.7 in [21]. With the following temporary definitions $a := e^{i\pi\frac{j}{n}}$ and $b := e^{i\pi\frac{k}{n}}$, using Euler's formula, we have

$$\begin{aligned}\langle \psi_j, \psi_k\rangle_d &= -\frac{1}{2n}\sum_{s=1}^{n}(a^s - a^{-s})(b^s - b^{-s})\\ &= -\frac{1}{2n}\sum_{s=1}^{n}\left[(ab)^s - (a/b)^s - (a/b)^{-s} + (ab)^{-s}\right],\\ &= -\frac{1}{2n}\left[\underbrace{\sum_{s=1}^{n}(ab)^s}_{A} - \underbrace{\sum_{s=1}^{n}(a/b)^s}_{B} - \underbrace{\sum_{s=1}^{n}(a/b)^{-s}}_{C} + \underbrace{\sum_{s=1}^{n}(ab)^{-s}}_{D}\right].\end{aligned} \tag{22}$$

Furthermore,

$$ab = e^{i\pi\frac{j+k}{n}},\ \frac{a}{b} = e^{i\pi\frac{j-k}{n}}.$$

Observe that when $ab \neq 1$, we have

$$A = \frac{ab(1-(ab)^n)}{1-ab},\quad D = \frac{1-(ab)^{-n}}{ab-1},\quad A + D = \frac{ab(1-(ab)^n) - (1-(ab)^{-n})}{1-ab}.$$

Similarly, if $a/b \neq 1$,

$$B + C = \frac{(a/b)(1-(a/b)^n) - (1-(a/b)^{-n})}{1-(a/b)}.$$

We fix $1 \leq k \leq n-1$ and discuss different situations depending on j.

(I.) $1 \leq j \leq n-1$ and $j+k \neq n$.
Since $n \neq j+k < 2n$, we always have $ab = e^{i\pi \frac{j+k}{n}} \neq 1$, and the terms A and D can be calculated as above. Similarly, since $-n < j-k < n$, $a/b = 1$ only when $j = k$. Moreover, $j+k$ and $j-k$ have the same parity, and so $j = k$ is only possible if $j+k$ is even.

(i.) $j+k$ is even.
In this case, $(ab)^n = 1$. This leads to $A = D = 0$.
Further, if $j = k$, we have $a/b = b/a = 1$ and $B = C = n$. Otherwise, if $j \neq k$, we have $a/b \neq 1$ and $(a/b)^n = 1 = (b/a)^n$ (since $j-k$ is even), and so

$$B = \frac{a/b(1-(a/b)^n)}{1-a/b} = 0, \quad C = 0,$$

which implies (22) equals 1.

(ii.) $j+k$ is odd. We have $(ab)^n = (a/b)^n = -1$, which results in $A + D = B + C = -2$, and so (22) equals 0.

(II.) $1 \leq j < n$ and $j+k = n$. We have $ab = -1$. Arguing as above, if n is odd, $A + D = -2$ and $B + C = -2$. If n is even, $A = D = 0$ and $B = C = n\delta_{jk}$.

The remaining cases follow the same arguments, and hence we omit the (lengthy and elementary) calculations.

(III.) $j = ln$ with $l \in \mathbb{N}$.
It can be shown that $A + D = B + C$ always holds.

(IV.) $j \in \{ln+1, \ldots, (l+1)n-1\}$.
When l is even, one obtains $\langle \psi_j, \psi_k \rangle_d = \delta_{\tilde{j}k}$, where $\tilde{j} = j - ln$. Otherwise, for odd l, $\langle \psi_j, \psi_k \rangle_d = -\delta_{\tilde{j}k}$ where $\tilde{j} = (l+1)n - j$.

5.2. *Proof of Theorem 3.1*

In this proof we use the notation $\|\cdot\| = \|\cdot\|_{H_1} = \|\cdot\|_{\ell^2}$. To show

$$\sup_{\|f_0\|_\beta \leq K} \mathbb{E}_{f_0} \Pi_n \left(f : \|f - f_0\| \geq M_n \varepsilon_n | U^n\right) \to 0,$$

we first apply Markov's inequality,

$$M_n^2 \varepsilon_n^2 \Pi_n \left(f : \|f - f_0\|^2 \geq M_n^2 \varepsilon_n^2 | U^n\right) \leq \int \|f - f_0\|^2 \, \mathrm{d}\Pi_n(f|U^n).$$

From (12) and the bias–variance decomposition,

$$\int \|f - f_0\|^2 \, \mathrm{d}\Pi_n(f|U^n) = \|\hat{f} - f_0\|^2 + \|\sigma\|^2,$$

where $\sigma = (\sigma_k)_k$ is given in (12). Because σ is deterministic,

$$\mathbb{E}_{f_0} \left[\Pi_n \left(f : \|f - f_0\| \geq M_n \varepsilon_n | U^n\right)\right] \leq \frac{1}{M_n^2 \varepsilon_n^2} \left(\mathbb{E}_{f_0} \|\hat{f} - f_0\|^2 + \|\sigma\|^2\right).$$

Since $M_n \to \infty$ is assumed, it suffices to show that the terms in brackets are bounded by a constant multiple of ε_n^2 uniformly in f_0 in the Sobolev ellipsoid.

Using (14), we obtain

$$\mathbb{E}_{f_0} \|\hat{f} - f_0\|^2 = \|\mathbb{E}_{f_0} \hat{f} - f_0\|^2 + \|\tau\|^2 = \|\mathbb{E}_{f_0} \hat{f} - f_0^n\|^2 + \|f_0^r\|^2 + \|\tau\|^2,$$

where $\tau = (\tau_k)_k$ given in (14) and

$$f_0^n = (f_{0,1}, \ldots, f_{0,n-1}, 0, \ldots),$$
$$f_0^r = (0, \ldots, 0, f_{0,n}, f_{0,n+2}, \ldots).$$

We need to obtain a uniform upper bound over the ellipsoid $\{f_0 : \|f_0\|_\beta \leq K\}$ for

$$\|\mathbb{E}_{f_0}\hat{f} - f_0^n\|^2 + \|f_0^r\|^2 + \|\tau\|^2 + \|\sigma\|^2. \tag{23}$$

We have

$$\begin{aligned}\|\mathbb{E}_{f_0}\hat{f} - f_0^n\|^2 &= \sum_{k=1}^{n-1}\left(\frac{na_k^2\lambda_k}{na_k^2\lambda_k+1}f_{0,k} + \frac{na_k\lambda_k}{na_k^2\lambda_k+1}R_k - f_{0,k}\right)^2 \\ &\lesssim \underbrace{\sum_{k=1}^{n-1}\frac{1}{\left(na_k^2\lambda_k+1\right)^2}f_{0,k}^2}_{A_1} + n\sup_{k<n}R_k^2\underbrace{\sum_{k=1}^{n-1}\frac{na_k^2\lambda_k^2}{\left(na_k^2\lambda_k+1\right)^2}}_{A_2},\end{aligned} \tag{24}$$

and

$$\|f_0^r\|^2 = \sum_{k\geq n} f_{0,k}^2, \quad \|\tau\|^2 = \sum_{k=1}^{n-1}\frac{na_k^2\lambda_k^2}{\left(na_k^2\lambda_k+1\right)^2} = A_2, \quad \|\sigma\|^2 = \sum_{k=1}^{n-1}\frac{\lambda_k}{na_k^2\lambda_k+1}.$$

Recall that we write (15) as $\varepsilon_n = \varepsilon_{n,1} \vee \varepsilon_{n,2}$. The statements (i.)–(iii.) follow by elementary calculations. Specifically, in (ii.) the given ρ_n is the best scaling, as it gives the fastest rate. From [16] (see the argument below (7.3) on page 21), A_1 is bounded by a fixed multiple of $(\varepsilon_{n,1})^2$, and $\|\tau\|^2$, $\|\sigma\|^2$ are bounded by multiples of $(\varepsilon_{n,2})^2$. Hence, to show that the rate is indeed (15), it suffices to show that $n\sup_{k\leq n}R_k^2A_2$ and $\|f_0^r\|^2$ can be bounded by a multiple of $(\varepsilon_n)^2$ uniformly in the ellipsoid $\{f_0 : \|f_0\|_\beta \leq K\}$. Since $A_2 = \|\tau\|^2$, to that end it is sufficient to show that $\sup_{k<n}nR_k^2 = O(1)$, and that $\|f_0^r\|^2 = O(\varepsilon_n)^2$.

Since $f_0 \in S^\beta$, we have the following straightforward bound,

$$\|f_0^r\|^2 \leq n^{-2\beta}\sum_{k\geq n} f_{0,k}^2k^{2\beta} \leq n^{-2\beta}\|f_0\|_\beta^2 \lesssim n^{-2\beta},$$

which is uniform in $\{f_0 : \|f_0\|_\beta \leq K\}$. By comparing to the rates in the statements (ii.)–(iii.), it is easy to see that $n^{-2\beta}$ is always negligible with respect to ε_n^2.

Proving $\sup_{k\leq n}nR_k^2 = O(1)$ is equivalent to showing $\sup_{k\leq n}R_k^2 = O(1/n)$; but the latter has been already proved in (10). Notice that we actually obtained a sharper bound $\sup_{k\leq n}nR_k^2 = o(1)$ than the one necessary for our purposes in this proof. However, this sharper bound will be used in the proof of Theorem 3.4. By taking supremum over f_0, we thus have

$$\sup_{\|f_0\|_{S^\beta}\leq K}\left(\|\mathbb{E}_{f_0}\hat{f} - f_0^n\|^2 + \|f_0^r\|^2\right) \lesssim \varepsilon_n^2 + n^{-2\beta} \lesssim \varepsilon_n^2, \tag{25}$$

with which we conclude that up to a multiplicative constant, (23) is bounded by ε_n^2 uniformly over the ellipsoid $\sup_{\|f_0\|_\beta\leq K}$. This completes the proof.

5.3. Proof of Theorem 3.3

We start by generalizing Theorem 3.1 in [17]. Following the same lines as in the proof of that theorem and using Lemmas 6.1, 6.2, 6.3, 6.4 in Section 6 of the present paper instead of analogous technical results in [17], the statement of Theorem 3.1 in [17] can be extended from $s = 2$ to a general $s \geq 1$, for which the posterior rate is given by (16), or $\varepsilon_n = \varepsilon_{n,1} \vee \varepsilon_{n,2}$ in short.

In our model, we again obtain (23) and also that a fixed multiple of $(\varepsilon_{n,1})^2$ is an upper bound of A_1, and that $\|\tau\|^2$, $\|\sigma\|^2$ can be bounded from above by fixed multiples of $(\varepsilon_{n,1})^2$.

Now as in the proof of Theorem 3.1 in Section 5.2, we will show that $\sup_{\|f_0\|_\beta\leq K}(\|\mathbb{E}_{f_0}\hat{f} - f_0^n\|^2 + \|f_0^r\|^2)$ can be bounded by a fixed multiple of $(\varepsilon_n)^2$ by proving that $\sup_{k\leq n}nR_k^2 = O(1)$. By (11), $n(R_k)^2 \lesssim \exp(-2pn)n$, and the right hand side converges to zero. Therefore,

$$\sup_{\|f_0\|_\beta\leq K}\left(\|\mathbb{E}_{f_0}\hat{f} - f_0^n\|^2 + \|f_0^r\|^2\right) \lesssim \varepsilon_n.$$

Parts (i.) and (ii.) of the statement of the theorem are obtained by direct substitutions, using the fact that $\log n \ll n$. Notice that if $\rho_n \gtrsim (\log n)^{(\alpha-\beta)/s}$, the rate ε_n deteriorates and is dominated by the second term in (16).

For the case $\lambda_k = \exp(-\alpha k^s)$, the argument follows the same lines as in Section 5.1 in [17], and our arguments above.

5.4. Proof of Theorem 3.4

The proof runs along the same lines as the proof of Theorem 4.2 in [16]. We will only show the main steps here.

In Section 2.2, we have shown that the posterior distribution is $\otimes_{k\in\mathbb{N}}\mathcal{N}(\hat{f}_k, \sigma_k^2)$, the radius $r_{n,\gamma}$ in (17) satisfies $\mathbb{P}_{X_n}(X_n < r_{n,\gamma}^2) = 1-\gamma$, where X_n is a random variable distributed as the square norm of an $\otimes_{k\in\mathbb{N}}\mathcal{N}(\hat{f}_k, \sigma_k^2)$ variable. Let $T = (\tau_k^2)_{k\in\mathbb{N}}$. Under (8), the variable $\hat{f}$ is distributed as $\mathcal{N}_{H_1}(\mathbb{E}_{f_0}\hat{f}, T) \coloneqq \otimes_{k\in\mathbb{N}}\mathcal{N}(\mathbb{E}_{f_0}\hat{f}_k, \tau_k^2)$. Hence the coverage (19) can be rewritten as

$$\mathbb{P}_{W_n}(\|W_n + \mathbb{E}_{f_0}\hat{f} - f_0\|_{H_1} \le r_{n,\gamma}), \tag{26}$$

where $W_n \sim \mathcal{N}_{H_1}(0, T)$. Denote $V_n = \|W_n\|_{H_1}^2$ and observe that one has in distribution

$$X_n = \sum_{1\le i<n} \sigma_i^2 Z_i^2, \quad V_n = \sum_{1\le i<n} \tau_i^2 Z_i^2$$

for $\{Z_i\}$ independent standard Gaussian random variables with

$$\sigma_i^2 = \frac{\lambda_i}{na_i^2\lambda_i + 1}, \quad \tau_i^2 = \frac{na_i^2\lambda_i^2}{(na_i^2\lambda_i+1)^2}.$$

By the same argument as in [16], one can show that the standard deviations of X_n and V_n are negligible with respect to their means,

$$\mathbb{E}X_n \asymp \rho_n^2(\rho_n^2 n)^{-2\alpha/(2\alpha+2p+1)}, \quad \mathbb{E}V_n \asymp \rho_n^2(\rho_n^2 n)^{-2\alpha/(2\alpha+2p+1)}, \tag{27}$$

and the difference of their means,

$$\mathbb{E}(X_n - V_n) \asymp \rho_n^2(\rho_n^2 n)^{-2\alpha/(2\alpha+2p+1)}.$$

Since $X_n \ge V_n$, the distributions of X_n and V_n are asymptotically separated, i.e. $\mathbb{P}(V_n \le v_n \le X_n) \to 1$ for some v_n, e.g. $v_n = \mathbb{E}(V_n + X_n)/2$. Since $r_{n,\gamma}^2$ are $1-\gamma$ quantiles of X_n, we also have $\mathbb{P}(V_n \le r_{n,\gamma}^2(1+o(1))) \to 1$. In addition, by (27),

$$r_{n,\gamma}^2 \asymp \rho_n^2(\rho_n^2 n)^{-2\alpha/(2\alpha+2p+1)}.$$

Introduce

$$B_n \coloneqq \sup_{\|f_0\|_\beta \lesssim 1} \|\mathbb{E}_{f_0}\hat{f} - f_0\|_{H_1} = \sup_{\|f_0\|_\beta\lesssim 1} \left(\|\mathbb{E}_{f_0}\hat{f} - f_0^n\|_{H_1} + \|f_0^r\|_{H_1}\right). \tag{28}$$

It follows from the arguments for (10) in the proof of Theorem 3.1 that

$$B_n \lesssim \varepsilon_{n,1} \vee \left(\sqrt{n}R\varepsilon_{n,2}\right),$$

where $R = \sup_{k<n} R_k \lesssim n^{-(p+\beta)}$. Now apply the argument on the lower bound from Lemma 8.1 in [16] (with $q=\beta, t=0, u=2\alpha+2p+1, v=2, N=\rho_n^2 n$) to obtain that $B_n \gtrsim \varepsilon_{n,1}$. Thus we have

$$\varepsilon_{n,1} \lesssim B_n \lesssim \varepsilon_{n,1} \vee \left(\sqrt{n}R\varepsilon_{n,2}\right).$$

We consider separate cases. In case (i.), substituting the corresponding ρ_n into the expression of $\varepsilon_{n,1}$ and $\varepsilon_{n,2}$, we have $\varepsilon_{n,1} \ll \varepsilon_{n,2}$. By (10), $B_n \lesssim \varepsilon_{n,1} \vee \left(\sqrt{n}R\varepsilon_{n,2}\right) \ll \varepsilon_{n,2} \asymp r_{n,\gamma}$. This leads to

$$\begin{aligned}\mathbb{P}(\|W_n + \mathbb{E}_{f_0}\hat{f} - f_0\|_{H_1} \le r_{n,\gamma}) &\ge \mathbb{P}(\|W_n\|_{H_1} \le r_{n,\gamma} - B_n)\\ &= \mathbb{P}(V_n \le r_{n,\gamma}^2(1+o(1))) \to 1\end{aligned} \tag{29}$$

uniformly in the set $\{f_0 : \|f_0\|_\beta \lesssim 1\}$.

In case (iii.), the given ρ_n leads to $\varepsilon_{n,1} \gg \varepsilon_{n,2}$ and consequently $B_n \gg r_{n,\gamma}$. Hence,

$$\mathbb{P}(\|W_n + \mathbb{E}_{f_0}\hat{f}^n - f_0^n\|_{H_1} \leq r_{n,\gamma}) \leq \mathbb{P}(\|W_n\|_{H_1} \geq B_n - r_{n,\gamma}) \to 0,$$

for any f_0^n (nearly) attaining the supremum.

In case (ii.), we have $B_n \asymp r_{n,\gamma}$. If $\beta < 2\alpha + 2p + 1$, by Lemma 8.1 in [16] the bias $\mathbb{E}_{f_0}\hat{f} - f_0$ at a fixed f_0 is of strictly smaller order than B_n. Following the argument of case (i.), the asymptotic coverage can be shown to converge to 1.

For existence of a sequence along which the coverage is $c \in [0, 1)$, we only give a sketch of the proof here; the details can be filled in as in [16].

The coverage (26) with f_0 replaced by f_0^n tends to c, if for $b_n = \mathbb{E}_{f_0}\hat{f}^n - f_0^n$ and z_c a standard normal quantile,

$$\frac{\|W_n + b_n\|_{H_1}^2 - \mathbb{E}\|W_n + b_n\|_{H_1}^2}{\operatorname{sd}\|W_n + b_n\|_{H_1}^2} \rightsquigarrow \mathcal{N}(0, 1), \tag{30}$$

$$\frac{r_{n,\gamma}^2 - \mathbb{E}\|W_n + b_n\|_{H_1}^2}{\operatorname{sd}\|W_n + b_n\|_{H_1}^2} \to z_c, \tag{31}$$

Since W_n is centred Gaussian $\mathcal{N}_{H_1}(0, T)$, (31) can be expressed as

$$\frac{r_{n,\gamma}^2 - \mathbb{E}V_n - \sum_{i=1}^{n-1} b_{n,i}^2}{\sqrt{\operatorname{var} V_n + 4\sum_{i=1}^{n-1} \tau_{i,n}^2 b_{n,i}^2}} \to z_c. \tag{32}$$

Here $\{b_{n,i}\}$ has exactly one nonzero entry depending on the smoothness cases $\beta \leq 2\alpha + 2p + 1$ and $\beta > 2\alpha + 2p + 1$. The nonzero entry, which we call b_{n,i_n}, has the following representation, with d_n to be yet determined,

$$b_{n,i_n}^2 = r_{n,\gamma}^2 - \mathbb{E}V_n - d_n \operatorname{sd} V_n.$$

Since $r_{n,\gamma}^2$, $\mathbb{E}V_n$ and $r_{n,\gamma}^2 - \mathbb{E}V_n$ have the same order and $\operatorname{sd} V_n$ is of strictly smaller order, one can show that the left hand side of (32) is equivalent to

$$\frac{d_n \operatorname{sd} V_n}{\sqrt{\operatorname{var} V_n + 4\tau_{i_n,n}^2 (r_{n\gamma}^2 - \mathbb{E}V_n)(1 + o(1))}},$$

for bounded or slowly diverging d_n. Then (32) can be obtained by discussing different smoothness cases separately, by a suitable choice of i_n, d_n.

To prove the asymptotic normality in (30), the numerator can be written as

$$\|W_n + b_n\|_{H_1}^2 - \mathbb{E}\|W_n + b_n\|_{H_1}^2 = \sum_i \tau_{i,n}^2 (Z_i^2 - 1) + 2b_{n,i_n}\tau_{i_n,n} Z_{i_n}.$$

Next one applies the arguments as in [16].

5.5. *Proof of Theorem 3.5*

This proof is almost identical to the proof of Theorem 2.2 in [17]. We supply the main steps.

Following the same arguments as in the proof of Theorem 3.4, we obtain

$$\mathbb{E}X_n \asymp \rho_n^2 (\log(\rho_n^2 n))^{-2\alpha/s} \gg \operatorname{sd} X_n \asymp \rho_n^2 (\log(\rho_n^2 n))^{-1/(2s) - 2\alpha/s},$$
$$\mathbb{E}V_n \asymp \rho_n^2 (\log(\rho_n^2 n))^{-1/s - 2\alpha/s} \asymp \operatorname{sd} V_n,$$

as in the proof of Theorem 2.2 in [17]. This leads to

$$r_{n,\gamma}^2 \asymp \rho_n^2 (\log(\rho_n^2 n))^{-2\alpha/s},$$

and furthermore,

$$\mathbb{P}(V_n \leq \delta r_{n,\gamma}^2) = \mathbb{P}\left(\frac{V_n - \mathbb{E}V_n}{\operatorname{sd} V_n} \leq \frac{\delta r_{n,\gamma}^2 - \mathbb{E}V_n}{\operatorname{sd} V_n}\right) \to 1,$$

for every $\delta > 0$.

Similar to Theorem 3.4, the bounds on the square norm B_n (defined in (28)) of the bias are known: upper bound from the proof of Theorem 3.3, and lower bound from Lemma 6.1,

$$\varepsilon_{n,1} \lesssim B_n \lesssim \varepsilon_{n,1} \vee \left(\sqrt{n}R\varepsilon_{n,2}\right),$$

where $\varepsilon_{n,1}$, $\varepsilon_{n,2}$ are given in (16), and $\sqrt{n}R$ satisfies the bound (11).

In case (i.), $B_n \ll r_{n,\gamma}$, and hence (29) applies. The rest of the results can be obtained in a similar manner.

6. Auxiliary lemmas

The following lemmas are direct generalizations of the case $s = 2$ in the Appendix of [17] to a general s. They can be easily proved by simple adjustments of the original proofs in [17], and we only state the results.

Lemma 6.1 (*Lemma 6.1 in [17]*). *For $q \in \mathbb{R}$, $u \geq 0$, $v > 0$, $t + 2q \geq 0$, $p > 0$, $0 \leq r < pv$ and $s \geq 1$,*

$$\sup_{\|f\|_{S^q} \leq 1} \sum_{i=1}^{\infty} \frac{f_i^2 i^{-t} e^{-ri^s}}{(1 + Ni^{-u}e^{-pi^s})^v} \asymp N^{-r/p}(\log N)^{-t/s-2q/s+ru/ps},$$

as $N \to \infty$.

In addition, for any fixed $f \in S^q$,

$$N^{r/p}(\log N)^{t/s+2q/s-ru/ps} \sum_{i=1}^{\infty} \frac{f_i^2 i^{-t} e^{-ri^s}}{(1 + Ni^{-u}e^{-pi^s})^v} \to 0,$$

as $N \to \infty$.

Lemma 6.2 (*Lemma 6.2 in [17]*). *For $t, u \geq 0$, $v > 0$, $p > 0$, $0 < r < vp$ and $s \geq 1$, as $N \to \infty$,*

$$\sum_{i=1}^{\infty} \frac{i^{-t} e^{-ri^s}}{(1 + Ni^{-u}e^{-pi^s})^v} \asymp N^{-r/p}(\log N)^{-t/s+ru/ps}.$$

If $r = 0$ and $t > 1$, while other assumptions remain unchanged,

$$\sum_{i=1}^{\infty} \frac{i^{-t} e^{-ri^s}}{(1 + Ni^{-u}e^{-pi^s})^v} \asymp (\log N)^{(-t+1)/s}.$$

Lemma 6.3 (*Lemma 6.4 in [17]*). *Assume $s \geq 1$. Let I_N be the solution in i to $Ni^{-u}e^{-pi^s} = 1$, for $u \geq 0$ and $p > 0$. Then*

$$I_N \sim \left(\frac{1}{p}\log N\right)^{1/s}$$

Lemma 6.4 (*Lemma 6.5 in [17]*). *Let $s \geq 1$. As $K \to \infty$, we have*

(i.) *for $a > 0$ and $b \in \mathbb{R}$,*

$$\int_1^K e^{ax^s} x^b \, dx \sim \frac{1}{as} e^{aK^s} K^{b-s+1};$$

(ii.) *for $a, b, K > 0$,*

$$\int_K^{\infty} e^{-ax^s} x^{-b} \, dx \leq \frac{1}{as} e^{-aK^s} K^{-b-s+1}.$$

Acknowledgement

The research leading to the results in this paper has received funding from the European Research Council under ERC Grant Agreement 320637.

References

[1] P. Alquier, E. Gautier, G. Stoltz, Inverse Problems and High-Dimensional Estimation: Stats in the Château Summer School, August 31–September 4, 2009, in: Lecture Notes in Statistics, Springer, 2011.
[2] N. Bissantz, T. Hohage, A. Munk, F. Ruymgaart, Convergence rates of general regularization methods for statistical inverse problems and applications, SIAM J. Numer. Anal. 45 (6) (2007) 2610–2636.
[3] L. Cavalier, Nonparametric statistical inverse problems, Inverse Problems 24 (3) (2008) 034004.
[4] L. Cavalier, A. Tsybakov, Sharp adaptation for inverse problems with random noise, Probab. Theory Related Fields 123 (3) (2002) 323–354.
[5] A. Cohen, M. Hoffmann, M. Reiß, Adaptive wavelet Galerkin methods for linear inverse problems, SIAM J. Numer. Anal. 42 (4) (2004) 1479–1501.
[6] D.L. Donoho, Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition, Appl. Comput. Harmon. Anal. 2 (2) (1995) 101–126.
[7] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, in: Applied Mathematical Sciences, Springer New York, 2006.
[8] A. Kirsch, An Introduction to the Mathematical Theory of Inverse Problems, in: Applied Mathematical Sciences, Springer, 2011.
[9] G. Wahba, Practical approximate solutions to linear operator equations when the data are noisy, SIAM J. Numer. Anal. 14 (4) (1977) 651–667.
[10] F. Natterer, The Mathematics of Computerized Tomography, in: Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, 2001.
[11] V. Isakov, Inverse Problems for Partial Differential Equations, in: Applied Mathematical Sciences, Springer New York, 2013.
[12] D. Colton, R. Kress, Inverse Acoustic and Electromagnetic Scattering Theory, in: Applied Mathematical Sciences, Springer New York, 2012.
[13] M. Birke, N. Bissantz, H. Holzmann, Confidence bands for inverse regression models, Inverse Problems 26 (11) (2010) 115020.
[14] N. Bissantz, H. Dette, K. Proksch, Model checks in inverse regression models with convolution-type operators, Scand. J. Stat. 39 (2) (2012) 305–322.
[15] S. Ghosal, A. van der Vaart, Fundamentals of Nonparametric Bayesian Inference, in: Cambridge Series in Statistical and Probabilistic Mathematics, vol. 44, Cambridge University Press, Cambridge, 2017, p. xxiv+646.
[16] B.T. Knapik, A.W. van der Vaart, J.H. van Zanten, Bayesian inverse problems with Gaussian priors, Ann. Statist. 39 (5) (2011) 2626–2657.
[17] B.T. Knapik, A.W. van der Vaart, J.H. van Zanten, Bayesian recovery of the initial condition for the heat equation, Comm. Statist. Theory Methods 42 (7) (2013) 1294–1313.
[18] S. Ghosal, J.K. Ghosh, A.W. van der Vaart, Convergence rates of posterior distributions, Ann. Statist. 28 (2) (2000) 500–531.
[19] J. Conway, A Course in Functional Analysis, in: Graduate Texts in Mathematics, Springer, 1990.
[20] M. Haase, Functional Analysis: An Elementary Introduction, in: Graduate Studies in Mathematics, Amer. Mathematical Society, 2014.
[21] A. Tsybakov, Introduction to Nonparametric Estimation, in: Springer Series in Statistics, Springer, 2008.
[22] S. Efromovich, Simultaneous sharp estimation of functions and their derivatives, Ann. Statist. 26 (1) (1998) 273–278.
[23] A. Akansu, H. Agirman-Tosun, Generalized discrete Fourier transform with nonlinear phase, IEEE Trans. Signal Process. 58 (9) (2010) 4547–4556.
[24] A. Quarteroni, R. Sacco, F. Saleri, Numerical Mathematics, in: Texts in Applied Mathematics, Springer, 2010.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 404–408
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Numerical computation of charge carriers optical phonon scattering mobility in III–V semiconductor compounds

Revaz Kobaidze*, Elza Khutsishvili, Nodar Kekelidze

Laboratory of Semiconductor Materials, Ferdinand Tavadze Institute of Metallurgy and Materials Science, 10, Mindeli str. Tbilisi 0186, Georgia

Received 1 April 2018; accepted 22 June 2018
Available online 31 July 2018

Abstract

Optical phonon scattering mobility has been calculated using numerical methods, and a general program was developed in Matlab to calculate mobility due to scattering on optical phonons. Calculations were done for InAs material that was irradiated by fast neutrons.

1. Introduction

The semiconductor compounds of III–V type differ from monoatomic semiconductors mainly in that in the chemical bond of the atoms that make up their lattice there is a certain fraction of the ionic bond. Therefore, scattering of charge carriers by optical phonons can play a significant role in transport phenomena [1–3]. The presence of charge carriers scattering by optical phonons makes the task of interpreting transport phenomena in semiconductor compounds of III–V extremely complex. A quantitative analysis of transport phenomena is considered by a quantum mechanical method. Qualitative evaluation of these phenomena for individual limited temperature intervals, impurity concentrations, and for each of the crystals of semiconductor compounds of III–V was carried out in a large number of original papers systematized in [4–6]. However, a detailed analysis of the kinetic effects taking into account charge carriers scattering by optical phonon in crystals of semiconductor compounds of III–V type is scanty. III–V compounds are semiconductors materials where scattering by optical phonons is the principal mechanism at temperatures $T > 200$ K. The conduction electrons (or holes) in semiconductors with an ionic bond interact much more strongly with optical vibrations than with acoustic ones. This is related with the fact that in crystals with a partially ionic bond, such as compounds of III–V semiconductor, at optical vibrations in each crystal cell, an electric dipole moment arises with which the electrons (or holes) of conduction interact strongly. The scattering of charge

* Corresponding author.
E-mail address: kobaidzerezo@yandex.ru (R. Kobaidze).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.06.002

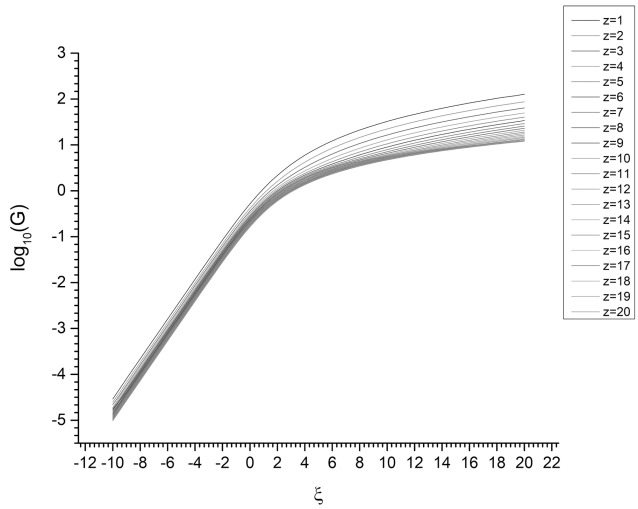


Fig. 1. $G(\xi, z)$ values for different z and ξ parameters.

carriers by thermal lattice vibrations is complicated and can be carried out using complex kinetic integrals [1–3]. For the basic understanding of properties and the nature of the material being studied, theoretical treatment of experimental data is essential. Comparison of experimental results with theoretical prediction is a necessary stage at investigation of transport phenomena. Usually, such complex calculations are performed analytically in rough approximation. However, detailed discussion of experimental data requires accurate calculation. This paper presents the possibilities of such numerical calculations using computer software. The role of scattering of charge carriers by optical lattice vibrations in current carriers mobility as a function of temperature for n-InAs by using Matlab program has been calculated.

2. The mobility for the electron–optic phonons interaction

The transport equation for electric conduction connected with the scattering of the conduction electrons (holes) by the optical lattice vibrations has been developed in detail for III–V materials in [7]. Obtained appropriate expression of conduction arbitrary degeneracy of electrons gas is given by the following formula:

$$\sigma = \frac{16a^3 M \nu_0 (kT)^2 (e^z - 1) G(\xi, z)}{3e^2 h^2} \tag{1}$$

where a—lattice parameter or interionic distance, M—reduced mass, e—the charge of free electron, k—Boltzmann's constant, h—Planck's constant, ν_0—fixed vibrational frequency of optical waves and

$$G(\xi, z) = \frac{25F_{3/2}^2(\xi)D_{00}(\xi, z) + 9F_{1/2}^2(\xi)D_{11}(\xi, z) - 30F_{1/2}(\xi)F_{3/2}(\xi)D_{01}(\xi, z)}{4(D_{01}(\xi, z)D_{11}(\xi, z) - D_{01}(\xi, z)^2)} \tag{2}$$

$$D_{00} = \int_0^\infty \frac{\sqrt{y(y+z)}dy}{(e^{-y+\xi}+1)(e^{y-\xi}+e^{-z})} \tag{3}$$

$$D_{01} = \int_0^\infty z^2 \sinh^{-1}(\sqrt{y/z}) + (2y+z)\sqrt{y(y+z)}\frac{dy}{(e^{-y+\xi}+1)(e^{y-\xi}+e^{-z})} \tag{4}$$

$$D_{11} = 2\int_0^\infty z^2(2y+z)\sinh^{-1}(\sqrt{y/z}) + y^{3/2}(y+z)^{3/2}\frac{dy}{(e^{-y+\xi}+1)(e^{y-\xi}+e^{-z})} \tag{5}$$

G (ξ, z) is a complicated function of two parameters ξ and z. "Degeneracy parameter" $\xi = \zeta/kT$ (ζ-Fermi level) appropriate is equal to $-\infty$ in the limiting non-degenerate case of the distribution of electrons over energies and ∞ in the limiting highly degenerate case of electrons gas. ξ-Fermi parameter is determined by experimentally known

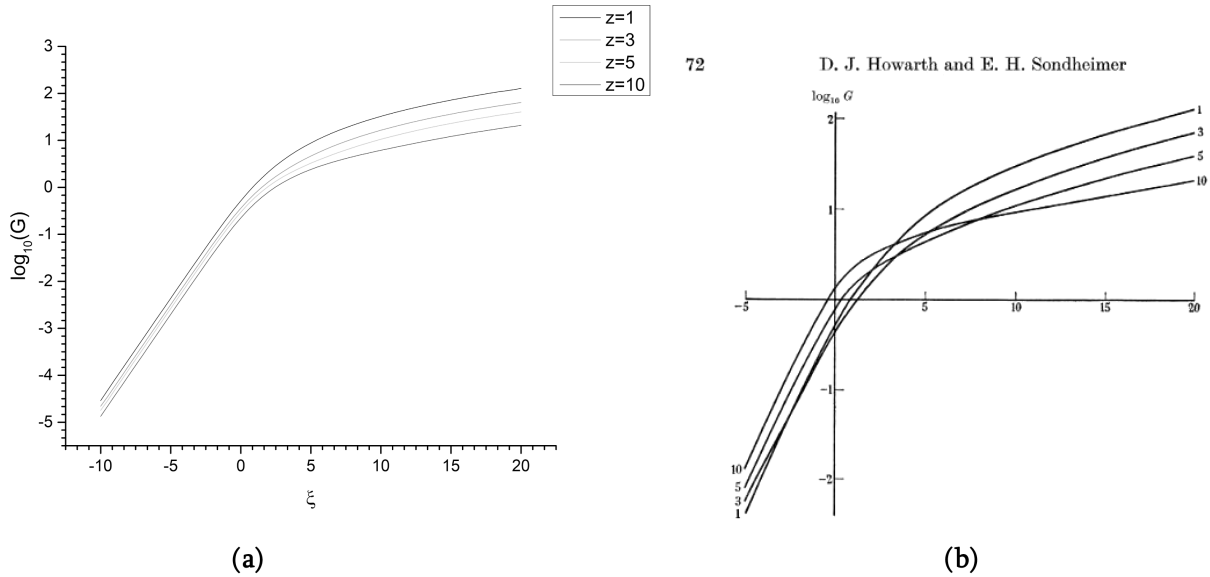


Fig. 2. $G(\xi, z)$ values for different z and ξ parameters: (a) calculated using (2) formula; (b) figure from [7] article calculated from (xx) formula.

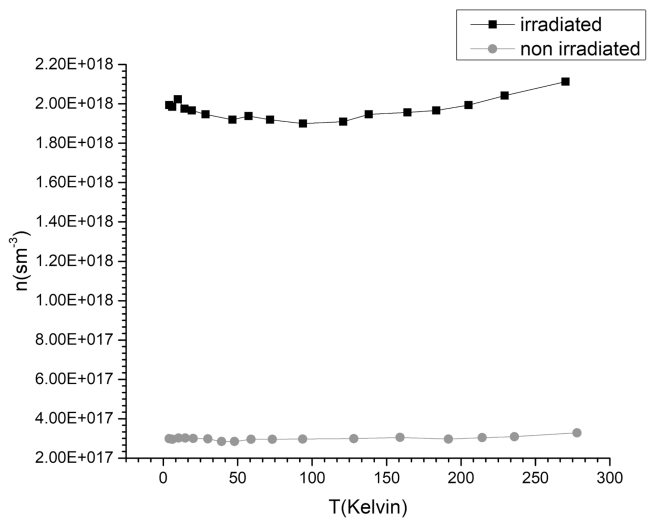


Fig. 3. InAs concentration for neutron irradiated and non irradiated samples.

charge carriers concentration (n) and temperature (T). So ξ-parameter can be found from the expression of

$$n = \frac{4\pi(2m^*kT)^{3/2}F_{1/2}(\xi)}{h^3} \tag{6}$$

$F_{3/2}(\xi)$, $F_{1/2}(\xi)$-Fermi-Dirac integrals are functions of ξ parameter. $z = \upsilon/T$, υ is the characteristic temperature of lattice being determined as $\upsilon = h\nu_0/k$ and it slowly depends on temperature. D_{00}, D_{01}, and D_{11} are bordered determinants.

If we take into account the electrical conductivity:

$$\sigma = \mu e n \tag{7}$$

Table 1
Optical phonon scattering mobility of n-type InAs.

T	ξ	$F_{1/2}(\xi)$	$F_{3/2}(\xi)$	D_{00}	D_{01}	D_{11}	$G(\xi, z)$	μ
4.06248	181.531161	1695.396378	195425.492764	3.027854e+04	6.271358e+06	1.255030e+16	213.594508	8.470445e+37
6.21244	123.523921	886.380834	65880.764539	1.170367e+04	1.591092e+06	1.245318e+16	151.043222	5.151503e+25
10.31336	69.645792	422.488246	18789.486715	5.056470e+03	3.968494e+05	1.238813e+16	79.426308	3.298882e+16
14.75334	52.855164	246.917260	7712.175285	2.185628e+03	1.234310e+05	1.236030e+16	62.763805	2.924787e+12
20.17784	36.867708	153.759509	3481.232822	1.261049e+03	5.081857e+04	1.234338e+16	42.182728	8.061397e+09
29.95897	25.164336	84.312867	1284.097324	5.503960e+02	1.501411e+04	1.232862e+16	29.059955	5.353658e+07
39.17465	18.642098	53.847480	610.843501	3.111710e+02	6.353801e+03	1.232152e+16	20.965935	4.933346e+06
47.91849	15.222281	39.805670	371.240702	2.078345e+02	3.483155e+03	1.231734e+16	17.153577	1.260388e+06
58.99126	12.668222	30.291719	237.244482	1.410030e+02	1.969214e+03	1.231383e+16	14.642055	4.192296e+05
73.27798	10.169297	21.879683	139.757549	9.168692e+01	1.040799e+03	1.231087e+16	11.747818	1.696643e+05
93.6311	7.938594	15.208278	77.943343	5.695641e+01	5.157420e+02	1.230822e+16	9.136924	7.775258e+04
128.05729	5.763024	9.577542	37.755960	3.163384e+01	2.178885e+02	1.230566e+16	6.524372	3.745210e+04
159.07065	4.635006	7.054126	23.727165	2.161227e+01	1.254035e+02	1.230430e+16	5.180463	2.556716e+04
191.60213	3.696233	5.195016	15.130335	1.522574e+01	7.579614e+01	1.230335e+16	3.988210	1.927054e+04
214.23307	3.307313	4.496559	12.305561	1.278206e+01	5.912492e+01	1.230286e+16	3.559116	1.671729e+04
235.87664	2.985448	3.953098	10.267121	1.098536e+01	4.775217e+01	1.230248e+16	3.200688	1.489737e+04
277.9847	2.562255	3.289184	7.971589	8.785215e+00	3.492184e+01	1.230191e+16	2.770808	1.253008e+04

Table 2
Optical phonon scattering mobility of n-type InAs irradiated by neutrons.

T	ξ	$F_{1/2}(\xi)$	$F_{3/2}(\xi)$	D_{00}	D_{01}	D_{11}	$G(\xi, z)$	μ
4.05207	540.731102	11349.983993	4564285.988143	3.242599e+05	1.887945e+08	1.255107e+16	893.881230	6.540358e+37
6.05502	538.051873	6187.887793	1777729.335024	6.512774e+04	3.698062e+07	1.245765e+16	1322.821919	2.605510e+26
9.82251	290.144235	3052.698575	533487.959131	3.022075e+04	9.314985e+06	1.239286e+16	693.817355	1.984829e+17
14.36107	181.493908	1685.837562	193690.050403	1.425860e+04	2.758133e+06	1.236203e+16	448.473920	5.628583e+12
19.2175	125.332922	1083.583257	90344.531206	8.914264e+03	1.185276e+06	1.234566e+16	296.361386	1.803556e+10
28.49651	90.307742	594.137529	33522.257109	2.778758e+03	2.670864e+05	1.233017e+16	285.828704	1.296352e+08
46.58459	54.428726	280.299333	9507.550447	9.340186e+02	5.347843e+04	1.231787e+16	189.265359	2.380839e+06
57.42277	44.468790	206.858431	5710.569135	6.322596e+02	2.910435e+04	1.231424e+16	152.276732	7.363762e+05
71.78876	36.039203	146.515797	3222.113237	3.495332e+02	1.294795e+04	1.231112e+16	138.185644	3.247308e+05
93.87171	27.724355	97.016111	1616.937574	2.012546e+02	5.578935e+03	1.230820e+16	105.226320	1.392293e+05
120.87187	21.523982	66.734740	871.163390	1.272624e+02	2.768254e+03	1.230607e+16	78.738337	7.451209e+04
138.12927	19.057840	55.672799	645.407348	1.016609e+02	1.960558e+03	1.230515e+16	68.598488	5.714584e+04
164.04597	16.089529	43.229063	425.215979	7.553481e+01	1.233874e+03	1.230413e+16	55.665618	4.230763e+04
183.42213	14.426559	36.747610	325.562890	6.283041e+01	9.230570e+02	1.230356e+16	48.358274	3.550450e+04
205.08689	13.012439	31.522043	253.202571	5.283416e+01	7.025730e+02	1.230305e+16	42.315203	3.027137e+04
229.31056	11.812896	27.307929	200.306112	4.494541e+01	5.447375e+02	1.230259e+16	37.331433	2.616340e+04
270.24647	10.233372	22.083484	141.870247	3.570429e+01	3.777084e+02	1.230200e+16	30.732462	2.142882e+04

The final expression for mobility (μ) connected with scattering by optical phonons is of the following form:

$$\mu = \frac{16a^3 M\omega(kT)^2(e^z - 1)G(\xi, z)}{3e^3h^2n} \tag{8}$$

3. Results and discussion

Gauss–Legendre numerical method with 100 points of weights was used to calculate Fermi-integrals in (2) and (6) formulas. Also, Gauss–Legendre method was used to calculate D_{00}, D_{01} and D_{11} values. Bisection root finding method was used to extract ξ parameter from (6) formula. The program was written in Matlab and is uploaded on repository [8]. In Fig. 1, the results of $G(\xi,z)$-function for different ξ and z values are given. Those results can be used for calculation mobility (8) if n, z are known.

It should be mentioned that in article [7] where $G(\xi, z)$ was first introduced, values of this function were calculated approximately by introducing new parameter and expanding integrand in power of $1/\xi$. We calculated $G(\xi, z)$ using numerical method compression of results from article [7] and our calculations are given in Fig. 2.

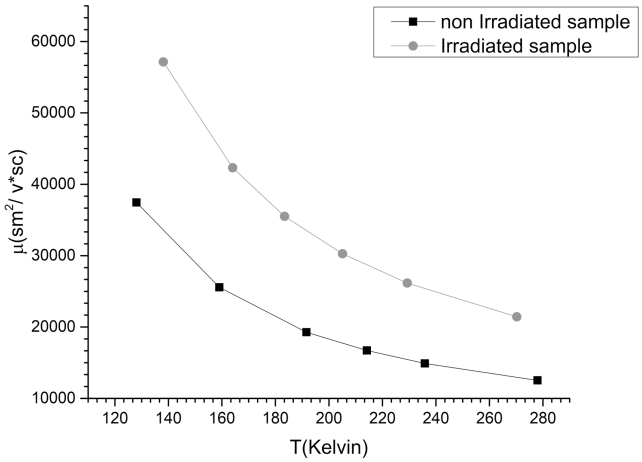


Fig. 4. Calculated Optical phonon scattering mobility in n-type InAs irradiated with neutrons and non irradiated sample.

N-type InAs was taken as an example. One sample is irradiated by neutrons. Fig. 3shows the electron concentrations for different temperature measured by Hall effect. We calculated Optical phonon scattering mobility using (7) formula for given concentrations and Debye temperature equal to 336.23. Results for both samples are given in Tables 1 and 2. In irradiated sample, calculated Optical phonon scattering mobility is higher than in non-irradiated sample. This may be due to higher electron concentrations that may be caused by creation of defects after irradiation (see Fig. 4).

References

[1] L.D. Landau, E.M. Lifshitz, Quantum Mechanics, in: A Course of Theoretical Physics, vol. 3, Pergamon Press, 1965.
[2] M. Born, Ken Huang, Dynamical Theory of Crystal Lattices, Oxford University Press, ISBN: 9780198503699, 1968.
[3] A.I. Anselm, Introduction to the Theory of Semiconductors, Nauka, Moscow, 1978.
[4] O. Madelung, Physics of III-V Compounds, J. Wiley, New York, 1964.
[5] M.P. Mikhailova, in: M. Levinshtein, S. Rumyantsev, M. Shur (Eds.), HandBook Series on Semiconductor Parameters, vol.1, World Scientific, London, 1996, pp. 147–168.
[6] M.A. Alzamil, Digest J. Nanomater. Biostruct. 5 (2011) 725–729.
[7] D.J. Howarth, E.H. Sondheimer, The theory of electronic conduction in polar semi-conductors, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 219 (1953) 53–74. http://dx.doi.org/10.1098/rspa.1953.0130.
[8] https://github.com/scienceman001/optical-phonon-scattering-#optical-phonon-scattering-.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 409–419
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

New coupled fixed point theorems in cone metric spaces with applications to integral equations and Markov process

D. Ramesh Kumar*, M. Pitchaimani

Ramanujan Institute for Advanced Study in Mathematics, University of Madras, Chepauk, Chennai 600 005, Tamil Nadu, India

Received 10 March 2017; received in revised form 22 November 2017; accepted 28 January 2018
Available online 6 February 2018

Abstract

In this paper, we define a generalized T-contraction and derive some new coupled fixed point theorems in cone metric spaces with total ordering condition. An illustrative example is provided to support our results. As an application, we utilize the results obtained to study the existence of common solution to a system of integral equations. We also present an application to Markov process.

MSC: 47H10; 54H25
Keywords: T-contraction; Coupled fixed point; Sequentially convergent; Integral equations; Markov process; Cone metric space

1. Introduction and preliminaries

In 1922, the Banach contraction principle [1] was introduced and it remains a powerful tool in nonlinear analysis which incites many authors to extend it, for instance [2–13]. In 2007, Huang and Zhang [14] generalized the notion of metric space by replacing the set of real numbers by ordered normed spaces, defined a cone metric space and extended the Banach contraction principle on these spaces over a normal solid cone. There are many fixed point results for generalized contractive conditions in metric spaces which were extended to cone metric spaces when the underlying cone is normal or not normal. Fixed point theory in cone metric spaces has been studied recently by many authors [2,14–26]. Bhaskar and Lakshmikantham [27] introduced the concept of coupled fixed point and applied their results to the study of existence and uniqueness of solution for a periodic boundary value problem in partially ordered metric spaces. Recently, Rahimi et al. [28] defined the concept of T-contraction in coupled fixed-point theory and obtained some coupled fixed point results on cone metric spaces without normality condition. For the detailed study on coupled fixed point results in ordered metric spaces and ordered cone metric spaces, we refer the reader to [29–34].

* Corresponding author.
E-mail address: rameshkumard14@gmail.com (D. Ramesh Kumar).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.01.006

After the study of T-contractions in a metric space by Beiranvand et al. [35], in [19], Filipović et al. obtained some fixed and periodic points satisfying T-Hardy-Rogers contraction in a cone metric space. Recently, Rahimi et al. [36] proved new fixed and periodic point results under T-contractions of two maps in cone metric spaces.

Motivated by the above work, we define generalized T-contraction and establish the existence and uniqueness of a coupled fixed point in cone metric spaces with a total ordering cone and dropping the normality condition which in turn will extend and generalize the results of [27,28,37]. We state some illustrative example to justify the obtained results. Also, we prove the existence of common solution to a system of integral equations. Further, we present an application to Markov process. The presented results improve and generalize many known results in cone metric spaces.

Now, we recall the definition of cone metric spaces and some of their properties.

Definition 1.1 (*[14]*)**.** Let E be a real Banach space. A subset P of E is called a cone if the following conditions are satisfied:

(i) P is closed, nonempty and $P \neq \{\theta\}$;
(ii) $a, b \in \mathbb{R}, a, b \geq 0$ and $x, y \in P$ imply that $ax + by \in P$.
(iii) $P \cap (-P) = \{\theta\}$.

Given a cone P of E, we define a partial ordering $\preceq$ with respect to P by $x \preceq y$ if and only if $y - x \in P$. We shall write $x \prec y$ to indicate that $x \preceq y$ but $x \neq y$, while $x \ll y$ will stand for $y - x \in int\, P$ (interior of P).

A cone P is called normal if there is a number $K > 0$ such that for all $x, y \in E$,

$$\theta \preceq x \preceq y \ \text{ implies } \ \|x\| \leq K\|y\|. \tag{1}$$

or equivalently, if, for any n, $x_n \preceq y_n \preceq z_n$ and

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} z_n = x \ \text{ imply } \ \lim_{n\to\infty} y_n = x.$$

The least positive number K satisfying the inequality (1) is called the normal constant of P.

Recently, Küçük et al. studied the characterization of total ordering cones with some properties and optimality conditions in [38].

Proposition 1.2 (*[38]*)**.** *Let E be a vector space and P be a partial ordering cone with partial order "$\preceq$" defined by $x \preceq y$ if and only if $y - x \in P$. Then "$\preceq$" is a total order on X if and only if $P \cup (-P) = E$.*

Definition 1.3 (*[14]*)**.** Let X be a nonempty set and $d : X \times X \to E$ be a mapping such that the following conditions hold:

(i) $\theta \preceq d(x, y)$ for all $x, y \in X$ and $d(x, y) = \theta$ if and only if $x = y$;
(ii) $d(x, y) = d(y, x)$ for all $x, y \in X$;
(iii) $d(x, y) \preceq d(x, z) + d(z, y)$ for all $x, y, z \in X$.

Then d is called a cone metric on X and (X, d) is called a cone metric space.

Example 1.4 (*[14]*)**.** Let $X = R, E = R^2, P = \{(x, y) \in E : x, y \geq 0\} \subset R^2$ and $d : X \times X \to E$ such that $d(x, y) = (|x - y|, \delta|x - y|)$, where $\delta \geq 0$ is a constant. Then (X, d) is a cone metric space.

Definition 1.5 (*[14]*)**.** Let (X, d) be a cone metric space. We say that $\{x_n\}$ is;

(i) a Cauchy sequence if for every $c \in E$ with $\theta \ll c$, there is N such that for all $m, n > N$, $d(x_n, x_m) \ll c$;
(ii) a convergent sequence if for every $c \in E$ with $\theta \ll c$, there is N such that for all $n > N$, $d(x_n, x) \ll c$, for some $x \in X$. We denote it by $\lim_{n\to\infty} x_n = x$ or $x_n \to x$.

A cone metric space X is said to be complete if every Cauchy sequence in X is convergent in X.

Proposition 1.6 (*[19]*)**.** *Let (X, d) be a cone metric space. Then the following properties are often used, particularly when dealing with cone metric spaces in which the cone need not be normal.*

($P1$) *If $u \preceq v$ and $v \ll w$, then $u \ll w$.*
($P2$) *If $\theta \preceq u \ll c$ for each $c \in int\, P$, then $u = \theta$.*
($P3$) *If E is a real Banach space with a cone P and if $a \preceq \lambda a$ where $a \in P$ and $0 \leq \lambda < 1$, then $a = \theta$.*
($P4$) *If $c \in int\, P$, $a_n \in E$ and $a_n \to \theta$, then there exists n_0 such that for all $n > n_0$, we have $a_n \ll c$.*

Definition 1.7 (*[19]*)**.** Let (X, d) be a cone metric space, P be a solid cone and $f : X \to X$. Then

(i) f is said to be continuous if $\lim_{n\to\infty} x_n = x$ implies that $\lim_{n\to\infty} fx_n = fx$, for all $\{x_n\}$ in X.
(ii) f is said to be sequentially convergent if, for every sequence $\{x_n\}$, such that $\{fx_n\}$ is convergent, then $\{x_n\}$ also is convergent.
(iii) f is said to be subsequentially convergent if, for every sequence $\{x_n\}$, such that $\{fx_n\}$ is convergent, then $\{x_n\}$ has a convergent subsequence.

Definition 1.8 (*[19]*)**.** Let (X, d) be a cone metric space and T, $f : X \to X$ two mappings. A mapping f is said to be a T-Hardy-Rogers contraction, if there exist $a_i \geq 0$, $i = 1, 2, \ldots, 5$ with $\sum_{i=1}^{5} a_i < 1$ such that for $x, y \in X$,

$$d(Tfx, Tfy) \preceq a_1 d(Tx, Ty) + a_2 d(Tx, Tfx) + a_3 d(Ty, Tfy) + a_4 d(Tx, Tfy) + a_5 d(Ty, Tfx). \tag{2}$$

Taking $a_1 = a_4 = a_5 = 0, a_2 = a_3 \neq 0$ (respectively $a_1 = a_2 = a_3 = 0, a_4 = a_5 \neq 0$) in (2), we obtain T-Kannan (respectively T-Chatterjea) contraction.

Definition 1.9 (*[28]*)**.** Let (X, d) be a cone metric space and $T : X \to X$ be a mapping. A mapping $S : X \times X \to X$ is called a T-Sabetghadam-contraction if there exist $a, b \geq 0$ with $a + b < 1$ such that for all $x, y \in X$,

$$d(TS(x, y), TS(\tilde{x}, \tilde{y})) \preceq ad(Tx, T\tilde{x}) + bd(Ty, T\tilde{y}).$$

Definition 1.10 (*[31]*)**.** Let (X, d) be a cone metric space. An element $(x, y) \in X \times X$ is called a coupled fixed point of the mapping $F : X \times X \to X$ if $F(x, y) = x$ and $F(y, x) = y$.

Note that if (x, y) is a coupled fixed point of F, then also (y, x) is a coupled fixed point of F.

2. Main results

Initially we define the following contraction condition which generalizes T-Sabetghadam-contraction.

Definition 2.1. Let (X, d) be a cone metric space with $P \cup (-P) = E$, (i.e. P is a total ordering cone) and $T : X \to X$ be a mapping. A mapping $S : X \times X \to X$ is called a generalized T-contraction if there exists λ with $0 \leq \lambda < 1$ such that

$$d(TS(x, y), TS(\tilde{x}, \tilde{y})) \preceq \lambda \max\{d(Tx, T\tilde{x}), d(Ty, T\tilde{y})\}. \tag{3}$$

for all $x, y, \tilde{x}, \tilde{y} \in X$.

The first main result in this paper is the following coupled fixed point result which generalizes Theorem 3 of Rahimi et al. [28].

Theorem 2.2. *Let (X, d) be a complete cone metric space, P be a solid cone with $P \cup (-P) = E$ and $T : X \to X$ be a continuous, one-to-one mapping and $S : X \times X \to X$ be a mapping such that* (3) *holds for all $x, y, \tilde{x}, \tilde{y} \in X$. Then*

(i) *there exist $z_{x_0}, z_{y_0} \in X$ such that*

$$\lim_{n\to\infty} TS^n(x_0, y_0) = z_{x_0} \quad \text{and} \quad \lim_{n\to\infty} TS^n(y_0, x_0) = z_{y_0};$$

where $S^n(x_0, y_0) = x_n$ and $S^n(y_0, x_0) = y_n$ are the iterative sequences.
(ii) *if T is subsequentially convergent, then $\{S^n(x_0, y_0)\}$ and $\{S^n(y_0, x_0)\}$ have a convergent subsequence;*
(iii) *there exist unique $w_{x_0}, w_{y_0} \in X$ such that $S(w_{x_0}, w_{y_0}) = w_{x_0}$ and $S(w_{y_0}, w_{x_0}) = w_{y_0}$;*

(iv) *if T is sequentially convergent, then for every $x_0, y_0 \in X$, the sequence $\{S^n(x_0, y_0)\}$ converges to $w_{x_0} \in X$ and the sequence $\{S^n(y_0, x_0)\}$ converges to $w_{y_0} \in X$.*

Proof. For $x_0, y_0 \in X$, we define a sequence as follows:

$$x_{n+1} = S(x_n, y_n) = S^{n+1}(x_0, y_0) \text{ and } y_{n+1} = S(y_n, x_n) = S^{n+1}(y_0, x_0), \ \forall n = 0, 1, 2, \ldots.$$

Now, using (3), we have

$$\begin{aligned} d(Tx_n, Tx_{n+1}) &= d(TS(x_{n-1}, y_{n-1}), TS(x_n, y_n)) \\ &\preceq \lambda \max\{d(Tx_{n-1}, Tx_n), d(Ty_{n-1}, Ty_n)\}, \end{aligned} \tag{4}$$

and

$$\begin{aligned} d(Ty_n, Ty_{n+1}) &= d(TS(y_{n-1}, x_{n-1}), TS(y_n, x_n)) \\ &\preceq \lambda \max\{d(Ty_{n-1}, Ty_n), d(Tx_{n-1}, Tx_n)\}. \end{aligned} \tag{5}$$

Let $D_n = \max\{d(Tx_n, Tx_{n+1}), d(Ty_n, Ty_{n+1})\}$. Applying (4) and (5), we get

$$D_n \preceq \lambda \max\{d(Ty_{n-1}, Ty_n), d(Tx_{n-1}, Tx_n)\} = \lambda D_{n-1},$$

where $0 \leq \lambda < 1$. Continuing in this fashion, we obtain

$$\theta \preceq D_n \preceq \lambda D_{n-1} \preceq \cdots \preceq \lambda^n D_0.$$

If we take $D_0 = \theta$, then (x_0, y_0) is a coupled fixed point of S. Suppose that $D_0 > \theta$ and for $n > m$, we have

$$d(Tx_m, Tx_n) \preceq d(Tx_m, Tx_{m+1}) + d(Tx_{m+1}, Tx_{m+2}) + \cdots + d(Tx_{n-1}, Tx_n) \tag{6}$$

and

$$d(Ty_m, Ty_n) \preceq d(Ty_m, Ty_{m+1}) + d(Ty_{m+1}, Ty_{m+2}) + \cdots + d(Ty_{n-1}, Ty_n). \tag{7}$$

From (6) and (7), we get

$$\begin{aligned} \max\{d(Tx_m, Tx_n), d(Ty_m, Ty_n)\} &\preceq \max\{d(Tx_m, Tx_{m+1}), d(Ty_m, Ty_{m+1})\} + \cdots \\ &\quad \max\{d(Tx_{n-1}, Tx_n), d(Ty_{n-1}, Ty_n)\} \\ &= D_m + D_{m+1} + \cdots + D_{n-1} \\ &\preceq (\lambda^m + \lambda^{m+1} + \cdots + \lambda^{n-1}) D_0 \\ &\preceq \frac{\lambda^m}{1-\lambda} D_0. \end{aligned}$$

Now applying (P1) and (P4), we have for every $c \in intP$, there exists a positive integer N such that $\max\{d(Tx_m, Tx_n), d(Ty_m, Ty_n)\} \ll c$ for every $n > m > N$ which implies that $\{Tx_n\}$ and $\{Ty_n\}$ are Cauchy sequences in X. By the completeness of X, we can find $z_{x_0}, z_{y_0} \in X$ such that

$$\lim_{n\to\infty} TS^n(x_0, y_0) = z_{x_0} \text{ and } \lim_{n\to\infty} TS^n(y_0, x_0) = z_{y_0}. \tag{8}$$

If T is subsequentially convergent, then $S^n(x_0, y_0)$ and $S^n(y_0, x_0)$ have convergent subsequences. Thus, there exist w_{x_0}, w_{y_0} in X and sequences $\{x_{n_j}\}$ and $\{y_{n_j}\}$ such that

$$\lim_{j\to\infty} S^{n_j}(x_0, y_0) = w_{x_0} \text{ and } \lim_{j\to\infty} S^{n_j}(y_0, x_0) = w_{y_0}.$$

Now, since T is continuous, we obtain

$$\lim_{j\to\infty} TS^{n_j}(x_0, y_0) = Tw_{x_0} \text{ and } \lim_{j\to\infty} TS^{n_j}(y_0, x_0) = Tw_{y_0}. \tag{9}$$

Hence, from (8) and (9), we have

$$Tw_{x_0} = z_{x_0}, Tw_{y_0} = z_{y_0}.$$

On the other hand, using (3) we get

$$d(TS(w_{x_0}, w_{y_0}), Tw_{x_0}) \preceq \lambda \max\{d(Tw_{x_0}, Tx_{n_j}), d(Tw_{y_0}, Ty_{n_j})\} + d(Tx_{n_j+1}, Tw_{x_0}).$$

Applying Proposition 1.6, we obtain $d(TS(w_{x_0}, w_{y_0}), Tw_{x_0}) = \theta$, that is, $TS(w_{x_0}, w_{y_0}) = Tw_{x_0}$. As T is one-to-one, we have $S(w_{x_0}, w_{y_0}) = w_{x_0}$. Similarly, $S(w_{y_0}, w_{x_0}) = w_{y_0}$. Therefore, (w_{x_0}, w_{y_0}) is a coupled fixed point of S. Suppose that (v_{x_0}, v_{y_0}) is another coupled fixed point of S, then

$$d(Tw_{x_0}, Tv_{x_0}) = d(TS(w_{x_0}, w_{y_0}), TF(v_{x_0}, v_{y_0})) \preceq \lambda \max\{d(Tw_{x_0}, Tv_{x_0}), d(Tw_{y_0}, Tv_{y_0})\}$$

and

$$d(Tw_{y_0}, Tv_{y_0}) = d(TS(w_{y_0}, w_{x_0}), TF(v_{y_0}, v_{x_0})) \preceq \lambda \max\{d(Tw_{y_0}, Tv_{y_0}), d(Tw_{x_0}, Tv_{x_0})\},$$

which implies that

$$\max\{d(Tw_{y_0}, Tv_{y_0}), d(Tw_{x_0}, Tv_{x_0})\} \leq \lambda \max\{d(Tw_{y_0}, Tv_{y_0}), d(Tw_{x_0}, Tv_{x_0})\}$$

which yields

$$d(Tw_{x_0}, Tv_{x_0}) = d(Tw_{y_0}, Tv_{y_0}) = \theta,$$

as $\lambda < 1$. Thus, $Tw_{x_0} = Tv_{x_0}$, $Tw_{y_0} = Tv_{y_0}$. Since T is one-to-one, we have $(w_{x_0}, w_{y_0}) = (v_{x_0}, v_{y_0})$. Further, if T is sequentially convergent, by replacing n by n_j, we obtain

$$\lim_{n\to\infty} S^n(x_0, y_0) = w_{x_0} \quad \text{and} \quad \lim_{n\to\infty} S^n(y_0, x_0) = w_{y_0}.$$

This completes the proof. □

Example 2.3. Let $X = [0, 1]$, $E = \mathbb{R}$ with $P = \{x \in E : x \geq 0\}$ and define $d(x, y) = |x - y|$. Then (X, d) is a cone metric space. Consider the mappings $T : X \to X$ defined by $Tx = \frac{x^2}{2}$ and $S : X \times X \to X$ defined by $S(x, y) = \frac{\sqrt{x^8+y^8}}{5}$. Clearly, T is one-to-one, continuous and (3) holds for all $x, y, u, v \in X$ and $\lambda > \frac{1}{4}$. Further, all the conditions of Theorem 2.2 are satisfied. Therefore, S has a unique coupled fixed point $(0, 0)$.

Theorem 2.4. *Let (X, d) be a complete cone metric space, P be a solid cone with $P \cup (-P) = E$ and $T : X \to X$ be a continuous, one-to-one mapping and $S : X \times X \to X$ be a mapping such that*

$$d(TS(x, y), TS(\tilde{x}, \tilde{y})) \preceq \lambda \max\{d(TS(x, y), Tx), d(TS(\tilde{x}, \tilde{y}), T\tilde{x})\} \tag{10}$$

for all $x, y, \tilde{x}, \tilde{y} \in X$ where $0 \leq \lambda < 1$. Then the conclusions of Theorem 2.2 hold.

Proof. The proof is similar to that of Theorem 2.2. □

Theorem 2.5. *Let (X, d) be a complete cone metric space, P be a solid cone with $P \cup (-P) = E$ and $T : X \to X$ be a continuous, one-to-one mapping and $S : X \times X \to X$ be a mapping such that*

$$d(TS(x, y), TS(\tilde{x}, \tilde{y})) \preceq \lambda \max\{d(TS(x, y), T\tilde{x}), d(TS(\tilde{x}, \tilde{y}), Tx)\} \tag{11}$$

for all $x, y, \tilde{x}, \tilde{y} \in X$ where $0 \leq \lambda < 1$. Then the conclusions of Theorem 2.2 hold.

Proof. We omit the proof as it is immediate from Theorem 2.2. □

Remarks 2.6. Theorems 2.2, 2.4 and 2.5 generalize the following results:

(i) Theorems 3, 4 and 5 of Rahimi et al. [28].
(ii) Theorems 2.2, 2.5 and 2.6 of Sabetghadam et al. [37].

Remarks 2.7. Note that the main results of Shatanawi [39] can be proved in a total ordering cone P under the weaker contractive condition of the type (3).

Now we obtain the following corollaries as consequences of Theorems 2.2, 2.4 and 2.5.

Corollary 2.8. *Let (X, d) be a complete cone metric space, P be a solid cone and $T : X \to X$ be a continuous, one-to-one mapping and $S : X \times X \to X$ be a mapping such that*

$$d(TS(x, y), TS(\tilde{x}, \tilde{y})) \preceq \lambda \max\left\{d(Tx, T\tilde{x}), d(Ty, T\tilde{y}), \frac{d(Tx, T\tilde{x}) + d(Ty, T\tilde{y})}{2}\right\}$$

for all $x, y, \tilde{x}, \tilde{y} \in X$ where $0 \leq \lambda < 1$. Then the conclusions of Theorem 2.2 hold.

Corollary 2.9. *Let (X, d) be a complete cone metric space, P be a solid cone and $T : X \to X$ be a continuous, one-to-one mapping and $S : X \times X \to X$ be a mapping such that*

$$d(TS(x, y), TS(\tilde{x}, \tilde{y})) \preceq \lambda \max\left\{d(TS(x, y), Tx), d(TS(\tilde{x}, \tilde{y}), T\tilde{x}), \frac{d(TS(x, y), Tx) + d(TS(\tilde{x}, \tilde{y}), T\tilde{x})}{2}\right\}$$

for all $x, y, \tilde{x}, \tilde{y} \in X$ where $0 \leq \lambda < 1$. Then the conclusions of Theorem 2.2 hold.

Corollary 2.10. *Let (X, d) be a complete cone metric space, P be a solid cone and $T : X \to X$ be a continuous, one-to-one mapping and $S : X \times X \to X$ be a mapping such that*

$$d(TS(x, y), TS(\tilde{x}, \tilde{y})) \preceq \lambda \max\left\{d(TS(x, y), T\tilde{x}), d(TS(\tilde{x}, \tilde{y}), Tx), \frac{d(TS(x, y), T\tilde{x}) + d(TS(\tilde{x}, \tilde{y}), Tx)}{2}\right\}$$

for all $x, y, \tilde{x}, \tilde{y} \in X$ where $0 \leq \lambda < 1$. Then the conclusions of Theorem 2.2 hold.

Next, we explain a general approach to our previous results.

Lemma 2.11. *Let (X, d) be a cone metric space. Then we have the following:*

(i) *$(X \times X, d_1)$ is a cone metric space with*

$$d_1((x, y), (u, v)) = \max\{d(x, u), d(y, v)\}.$$

In addition, (X, d) is complete if and only if $(X \times X, d_1)$ is complete.

(ii) *The mapping $S : X \times X \to X$ has a coupled fixed point if and only if the mapping $F_S : X \times X \to X \times X$ defined by $F_S(x, y) = (S(x, y), S(y, x))$ has a fixed point in $X \times X$.*

Proof.

(i) Notice that (i) and (ii) of Definition 1.3 are satisfied. Now it suffices to prove the triangle inequality. Since (X, d) is a cone metric space, we obtain

$$\begin{aligned} d_1((x, y), (u, v)) &= \max\{d(x, u), d(y, v)\} \\ &\preceq \max\{d(x, s) + d(s, u), d(y, t) + d(t, v)\} \\ &\preceq \max\{d(x, s), d(y, t)\} + \max\{d(s, u), d(t, v)\} \\ &= d_1((x, y), (s, t)) + d_1((s, t), (u, v)) \end{aligned}$$

for all $(x, y), (u, v), (s, t) \in X \times X$. Hence, $(X \times X, d_1)$ is a cone metric space. The completeness part can be easily proved.

(ii) Suppose that (x, y) is a coupled fixed point of S, that is, $S(x, y) = x$ and $S(y, x) = y$. Then

$$F_S(x, y) = (S(x, y), S(y, x)) = (x, y)$$

which shows that $(x, y) \in X \times X$ is a fixed point of F_S. Conversely, assume that $(x, y) \in X \times X$ is a fixed point of F_S , then $F_S(x, y) = (x, y)$ which implies that $S(x, y) = x$ and $S(y, x) = y$. □

Theorem 2.12. *Let (X, d) be a complete cone metric space, P be a total ordering solid cone and $T : X \to X$ be a continuous and one-to-one mapping. Moreover, let $S : X \times X \to X$ be a mapping satisfying*

$$\max\{d(TS(x, y), TS(\tilde{x}, \tilde{y})), d(TS(y, x), TS(\tilde{y}, \tilde{x}))\} \preceq \lambda \max\big\{d(Tx, T\tilde{x}), d(Ty, T\tilde{y})\big\} \tag{12}$$

for all $x, y, \tilde{x}, \tilde{y} \in X$, where $0 \leq \lambda < 1$. Then the conclusions of Theorem 2.2 hold.

Proof. Let us define $T_1 : X \times X \to X \times X$ by $T_1(x, y) = (Tx, Ty)$. Note that T_1 is continuous and one-to-one. Now, applying $Y = (x, y)$, $V = (u, v) \in X \times \times X$ and (ii) of Lemma 2.11, (12) becomes

$$d_1(T_1F_S(Y), T_1F_S(V)) \preceq \lambda d_1(T_1Y, T_1V).$$

It can be viewed that the conclusions follow by setting $a_1 = \lambda$ and $a_2 = a_3 = a_4 = a_5 = 0$ in Theorem 2.1 of [19] as $\lambda < 1$. □

Remarks 2.13. The cone metric defined in Theorem 2.2 is the generalized form of the cone metric defined in [28]. Further, Theorem 2.12 generalizes Theorem 6 of Rahimi et al. [28].

Example 2.14. Let $X = [0, 1]$ and $E = C^1_{\mathbb{R}}[0, 1]$ with $d(x, y) = |x - y|e^t$ where $e^t \in E$ on $P = \{\varphi \in E : \varphi \geq 0\}$. Then (X, d) is a cone metric space. Suppose that $T : X \to X$ defined by $Tx = \frac{x}{2}$ and $S : X \times X \to X$ defined by $S(x, y) = \frac{x-y}{10}$. Note that

$$d(TS(x, y), TS(u, v)) = \frac{e^t}{20}|(x - u) - (y - v)| \tag{13}$$

and

$$d(Tx, Tu) = \frac{e^t}{2}|x - u| \text{ and } d(Ty, Tv) = \frac{e^t}{2}|y - v|. \tag{14}$$

From (13) and (14), it can be easily seen that the condition (12) holds for all $x, y, u, v \in X$ and $\lambda > \frac{1}{10}$. Further, all the conditions of Theorem 2.12 are satisfied. Hence, $(0, 0)$ is a unique coupled fixed point of S.

3. An application to integral equations

The purpose of this section is to study the existence of solution of a system of nonlinear integral equations using the results we obtained.

Let $X = C\big([0, T], \mathbb{R}\big)$ (the set of continuous functions defined on $[0, T]$ and taking values in $\mathbb{R}$) be together with the metric given by

$$d(x, y) = \sup_{t\in[0,T]} |x(t) - y(t)|, \ \forall x, y \in X.$$

Consider the following system of integral equations, for $t \in [0, T]$, $T > 0$,

$$F(x, y)(t) = \int_0^T G(t, s)f(t, x(s), y(s))ds + g(t), \tag{15}$$

$$F(y, x)(t) = \int_0^T G(t, s)f(t, y(s), x(s))ds + g(t). \tag{16}$$

Theorem 3.1. *Suppose that the following hold:*

(i) $G : [0, T] \times [0, T] \to \mathbb{R}$ *is a continuous function.*
(ii) $g \in C([0, T], \mathbb{R})$.
(iii) $f : [0, T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *is a continuous function.*

(iv) *For all* $x, y, u, v \in X$ *and* $t \in [0, T]$*, we have*

$$|f(t, x(t), y(t)) - f(t, u(t), v(t))| \leq \lambda \max\{|x(t) - u(t)|, |y(t) - v(t)|\},$$

where $0 \leq \lambda < 1$.

(v) $\int_0^T |G(t, s)| \leq 1$.

Then the system (15)–(16) *has at least one solution in* $C\big([0, T], \mathbb{R}\big)$.

Proof. It is easy to see that (x, y) is a solution to (15)–(16) if and only if (x, y) is a coupled fixed point of F. Existence of such a point follows from Theorem 2.2, by taking T as identity mapping. So we have to check that all the conditions of Theorem 2.2 hold. For all $x, y, u, v \in X$ and $t \in [0, T]$, we have

$$\begin{aligned}
|F(x, y)(t) - F(u, v)(t)| &\leq \int_0^T \big|G(t, s)\big|\big|f(t, x(s), y(s)) - f(t, u(s), v(s))\big|ds, \\
&\leq \int_0^T \big|G(t, s)\big|\lambda \max\big\{|x(t) - u(t)|, |y(t) - v(t)|\big\}ds, \\
&\leq \left(\int_0^T \big|G(t, s)\big|ds\right)\lambda \max\big\{d(x, u), d(y, v)\big\},
\end{aligned}$$

which yields that

$$d(F(x, y), F(u, v)) \leq \lambda \max\big\{d(x, u), d(y, v)\big\}, \quad \forall x, y, u, v \in X.$$

This shows that the contractive condition of Theorem 2.2 holds. Therefore, F has a unique coupled fixed point $(\tilde{x}, \tilde{y}) \in C\big([0, T], \mathbb{R}\big) \times C\big([0, T], \mathbb{R}\big)$ which is the unique solution of (15)–(16). □

4. An application to Markov process

Let $\mathbb{R}^n_+ = \{x = (x_1, x_2, \ldots, x_n) : x_i > 0, i = 1, 2, \ldots, n\}$ and $\Delta^2_{n-1} = \{z = (x, y) \in \mathbb{R}^n_+ \times \mathbb{R}^n_+ : \sum_{i=1}^n z_i = \sum_{i=1}^n (x_i + y_i) = 1\}$ denote the $2(n-1)$ dimensional unit simplex. Note that any $z \in \Delta^2_{n-1}$ may be regarded as a probability over the $2n$ possible states. A random process in which one of the $2n$ states is realized in each period $t = 1, 2, \ldots$ with the probability conditioned on the current realized state is called Markov Process. Let a_{ij} denote the conditional probability that state i is reached in succeeding period starting in state j. Then, given the prior probability vector z^t in period t, the posterior probability in period $t + 1$ is given by $z_i^{t+1} = \sum_j a_{ij} z_j^t$ for each $i = 1, 2, \ldots, n$. To express this in matrix notation, we let z^t denote a column vector. Then, $z^{t+1} = Az^t$. Observe that the properties of conditional probability require each $a_{ij} \geq 0$ and $\sum_{i=1}^n a_{ij} = 1$ for each j. If for any period t, $z^{t+1} = z^t$ then z^t is a stationary distribution of the Markov process. Thus, the problem of finding a stationary distribution is equivalent to the fixed point problem $Az^t = z^t$.

For each i, let $\varepsilon_i = \min_j a_{ij}$ and define $\varepsilon = \sum_{i=1}^n \varepsilon_i$.

Theorem 4.1. *Under the assumption* $a_{ij} > 0$*, a unique stationary distribution exists for the Markov process.*

Proof. Let $d : \Delta^2_{n-1} \times \Delta^2_{n-1} \to \mathbb{R}^2$ be given by

$$d(s, t) = d((x, y), (u, v)) = \left(\sum_{i=1}^n \big(|x_i - u_i| + |y_i - v_i|\big), \alpha \sum_{i=1}^n \big(|x_i - y_i| + |y_i - v_i|\big)\right)$$

for all $s, t \in \Delta^2_{n-1}$ and some $\alpha \geq 0$.

Note that $d(s, t) \geq (0, 0)$ for all $s, t \in \Delta^2_{n-1}$ and $d(s, t) = (0, 0) \Rightarrow \Big(\sum_{i=1}^n \big(|x_i - u_i| + |y_i - v_i|\big), \alpha \sum_{i=1}^n \big(|x_i - u_i| + |y_i - v_i|\big)\Big) = (0, 0) \Rightarrow \big(|x_i - u_i| + |y_i - v_i|\big) = 0$ for all i, which implies that $s = t$. Assume $s = t$ then $x_i = u_i$

and $y_i = v_i$ for all i which implies that $|x_i - y_i| = |y_i - v_i| = 0 \Rightarrow \sum_{i=1}^{n}\big(|x_i - u_i| + |y_i - v_i|\big) \Rightarrow d(s,t) = (0,0)$.

$$\begin{aligned} d(s,t) &= \left(\sum_{i=1}^{n}\big(|x_i - u_i| + |y_i - v_i|\big), \alpha\sum_{i=1}^{n}\big(|x_i - u_i| + |y_i - v_i|\big)\right) \\ &= \left(\sum_{i=1}^{n}\big(|u_i - x_i| + |v_i - y_i|\big), \alpha\sum_{i=1}^{n}\big(|u_i - x_i| + |v_i - y_i|\big)\right) = d(t,s). \end{aligned}$$

Now

$$\begin{aligned} d(s,t) &= \left(\sum_{i=1}^{n}\big(|x_i - u_i| + |y_i - v_i|\big), \alpha\sum_{i=1}^{n}\big(|x_i - u_i| + |y_i - v_i|\big)\right) \\ &= \Bigg(\sum_{i=1}^{n}\big(|(x_i - p_i) + (p_i - u_i)| + |(y_i - q_i) + (q_i - v_i)|\big), \\ &\quad \alpha\sum_{i=1}^{n}\big(|(x_i - p_i) + (p_i - u_i)| + |(y_i - q_i) + (q_i - v_i)|\big)\Bigg) \\ &\preceq \Bigg(\sum_{i=1}^{n}\big(|(x_i - p_i)| + |(p_i - u_i)| + |(y_i - q_i)| + |(q_i - v_i)|\big), \\ &\quad \alpha\sum_{i=1}^{n}\big(|(x_i - p_i)| + |(p_i - u_i)| + |(y_i - q_i)| + |(q_i - v_i)|\big)\Bigg) \\ &= \left(\sum_{i=1}^{n}\big(|x_i - p_i| + |y_i - q_i|\big), \alpha\sum_{i=1}^{n}\big(|x_i - p_i| + |y_i - q_i|\big)\right) \\ &\quad + \left(\sum_{i=1}^{n}\big(|p_i - u_i| + |q_i - v_i|\big), \alpha\sum_{i=1}^{n}\big(|p_i - u_i| + |q_i - v_i|\big)\right) \\ &= d(s,r) + d(r,t) \ \text{ for } \ s = (x,y), r = (p,q), t = (u,v) \in \Delta_{n-1}^2. \end{aligned}$$

Thus (Δ_{n-1}^2, d) is a cone metric space with $P = \{(x_1, x_2, \ldots, x_n) : x_i \geq 0, \forall i = 1, 2, \ldots, n\}$. For $z \in \Delta_{n-1}$, let $t = Az$. Then each $\beta_i = \sum_{j=1}^{n} a_{ij} z_j \geq 0$. Further more, since each $\sum_{j=1}^{n} a_{ij} = 1$, we have

$$\sum_{i=1}^{n}\beta_i = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} z_j = \beta_i = \sum_{j=1}^{n} a_{ij}\sum_{j=1}^{n}(x_j + y_j) = \sum_{j=1}^{n}(x_j + y_j) = 1$$

which shows that $t \in \Delta_{n-1}^2$. Thus, we see that $A : \Delta_{n-1}^2 \to \Delta_{n-1}^2$. We shall show that A is a contraction. Let A_i denote the ith row of A. Then for any $(x, y), (u, v) \in \Delta_{n-1}$, we have

$$\begin{aligned} d(A(x,y), A(u,v)) &= \Bigg(\sum_{i=1}^{n}\Big|\sum_{j=1}^{n}(a_{ij}(x_j + y_j) - a_{ij}(u_j + v_j))\Big|, \\ &\qquad \alpha\sum_{i=1}^{n}\Big|\sum_{j=1}^{n}(a_{ij}(x_j + y_j) - a_{ij}(u_j + v_j))\Big|\Bigg) \\ &= \Bigg(\sum_{i=1}^{n}\Big|\sum_{j=1}^{n}(a_{ij} - \epsilon_i)((x_j + y_j) - (u_j + v_j)) + \epsilon_i((x_j + y_j) - (u_j + v_j))\Big|, \\ &\qquad \alpha\sum_{i=1}^{n}\Big|\sum_{j=1}^{n}(a_{ij} - \epsilon_i)((x_j + y_j) - (u_j + v_j)) + \epsilon_i((x_j + y_j) - (u_j + v_j))\Big|\Bigg) \end{aligned}$$

$$
\begin{aligned}
&\leq \Bigg(\sum_{i=1}^{n} |\sum_{j=1}^{n} (a_{ij} - \epsilon_i)((x_j + y_j) - (u_j + v_j))| + |\sum_{j=1}^{n} \epsilon_i((x_j + y_j) - (u_j + v_j))|, \\
&\quad \alpha \sum_{i=1}^{n} |\sum_{j=1}^{n} (a_{ij} - \epsilon_i)((x_j + y_j) - (u_j + v_j))| + |\sum_{j=1}^{n} \epsilon_i((x_j + y_j) - (u_j + v_j))| \Bigg) \\
&\leq \Bigg(\sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij} - \epsilon_i)(|x_j - u_j| + |y_j - v_j|), \\
&\quad \alpha \sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij} - \epsilon_i)(|x_j - u_j| + |y_j - v_j|) \Bigg) \\
&= \Bigg(\sum_{j=1}^{n} (|x_j - u_j| + |y_j - v_j|) \sum_{i=1}^{n} (a_{ij} - \epsilon_i), \\
&\quad \alpha \sum_{j=1}^{n} (|x_j - u_j| + |y_j - v_j|) \sum_{i=1}^{n} (a_{ij} - \epsilon_i) \Bigg) \\
&= \Bigg(\sum_{j=1}^{n} (|x_j - u_j| + |y_j - v_j|)(1 - \epsilon), \alpha \sum_{j=1}^{n} (|x_j - u_j| + |y_j - v_j|)(1 - \epsilon) \Bigg) \\
&= (1 - \epsilon) d((x, y), (u, v))
\end{aligned}
$$

which establishes that A is a contraction mapping. Thus, Theorem 2.2 with T as identity mapping ensures a unique stationary distribution for the Markov Process. Moreover, for any $z^* \in \Delta_{n-1}$, the sequence $\langle A^n z^* \rangle$ converges to the unique stationary distribution. □

References

[1] S. Banach, Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales, Fund Math. 3 (1922) 133–181.
[2] I. Beg, A.R. Butt, Fixed point for set-valued mappings satisfying an implicit relation in partially ordered metric spaces, Nonlinear Anal. 71 (9) (2009) 3699–3704.
[3] W.A. Kirk, Some recent results in metric fixed point theory, J. Fixed Point Theory Appl. 2 (2007) 195–207.
[4] J.J. Nieto, R.L. Pouso, R. Rodríguez-López, Fixed point theorems in ordered abstract spaces, Proc. Amer. Math. Soc. 135 (2007) 2505–2517.
[5] J.J. Nieto, R. Rodríguez-López, Existence and uniqueness of fixed point in partially ordered sets and applications to ordinary differential equations, Acta Math. Sinica, Engl. Ser. Mar. 23 (2007) 2205–2212.
[6] M. Pitchaimani, D. Ramesh Kumar, Some common fixed point theorems using implicit relation in 2-Banach spaces, Surv. Math. Appl. 10 (2015) 159–168.
[7] M. Pitchaimani, D. Ramesh Kumar, Common and coincidence fixed point theorems for asymptotically regular mappings in 2-Banach Spaces, Nonlinear Funct. Anal. Appl. 21 (1) (2016) 131–144.
[8] M. Pitchaimani, D. Ramesh Kumar, On construction of fixed point theory under implicit relation in Hilbert spaces, Nonlinear Funct. Anal. Appl. 21 (3) (2016) 513–522.
[9] M. Pitchaimani, D. Ramesh Kumar, On Nadler type results in ultrametric spaces with application to well-posedness, Asian-Eur. J. Math. 10 (4) (2017) 1750073(1–15). http://dx.doi.org/10.1142/S1793557117500735.
[10] M. Pitchaimani, D. Ramesh Kumar, Generalized Nadler type results in ultrametric spaces with application to well-posedness, Afr. Mat. 28 (2017) 957–970.
[11] D. Ramesh Kumar, M. Pitchaimani, Set-valued contraction mappings of Prešić-Reich type in ultrametric spaces, Asian-Eur. J. Math. 10 (4) (2017) 1750065(1–15). http://dx.doi.org/10.1142/S1793557117500656.
[12] D. Ramesh Kumar, M. Pitchaimani, A generalization of set-valued Prešić-Reich type contractions in ultrametric spaces with applications, J. Fixed Point Theory Appl. 19 (3) (2017) 1871–1887.
[13] T. Suzuki, A new type of fixed point theorem in metric spaces, Nonlinear Anal. 71 (2009) 5313–5317.
[14] L.G. Huang, X. Zhang, Cone metric spaces and fixed point theorems of contractive mappings, J. Math. Anal. Appl. 332 (2) (2007) 1468–1476.
[15] M. Abbas, B.E. Rhoades, Fixed and periodic point results in cone metric spaces, Appl. Math. Lett. 22 (4) (2009) 511–515.
[16] M. Abbas, B.E. Rhoades, T. Nazir, Common fixed points for four maps in cone metric spaces, Appl. Math. Comput. 216 (2010) 80–86.
[17] A. Azam, M. Arshad, I. Beg, Common fixed points of two maps in cone metric spaces, Rend. Circ. Mat. Palermo 57 (2008) 433–441.
[18] A.P. Farajzadeh, A. Amini-Harandi, D. Baleanu, Fixed point theory for generalized contractions in cone metric spaces, Commun. Nonlinear Sci. Numer. Simul. 17 (2) (2012) 708–712.

[19] M. Filipović, L. Paunović, S. Radenović, M. Rajović, Remarks on Cone metric spaces and fixed point theorems of T-Kannan and T-Chatterjea contractive mappings, Math. Comput. Modelling 54 (2011) 1467–1472.
[20] D. Ilić, V. Rakočević, Common fixed points for maps on cone metric space, J. Math. Anal. Appl. 341 (2008) 876–882.
[21] Z. Kadelburg, M. Pavlović, S. Radenović, Common fixed point theorems for ordered contractions and quasicontractions in ordered cone metric spaces, Comput. Math. Appl. 59 (2010) 3148–3159.
[22] Z. Kadelburg, S. Radenović, V. Rakočević, Remarks on quasi-contraction on a cone metric space, Appl. Math. Lett. 22 (2009) 1674–1679.
[23] S. Rezapour, R. Hamlbarani, Some notes on the paper: Cone metric spaces and fixed point theorems of contractive mappings, J. Math. Anal. Appl. 345 (2008) 719–724.
[24] G. Song, X. Sun, Y. Zhao, G. Wang, New common fixed point theorems for maps on cone metric spaces, Appl. Math. Lett. 23 (2010) 1033–1037.
[25] M. Turkoglu, M. Abuloha, Cone metric spaces and fixed point theorems in diametrically contractive mappings, Acta Math. Appl. Sin. 26 (2010) 489–496.
[26] P. Vetro, Common fixed points in cone metric spaces, Rend. Circ. Mat. Palermo 56 (2007) 464–468.
[27] T. Bhaskar, V. Lakshmikantham, Fixed point theorems in partially ordered metric spaces and applications, Nonlinear Anal. 65 (2006) 1379–1393.
[28] H. Rahimi, P. Vetro, G. Soleimani Rad, Coupled fixed-point results for T-contractions on cone metric spaces with applications, Math. Notes 98 (1) (2015) 158–167.
[29] M. Abbas, M. Ali Khan, S. Radenović, Common coupled fixed point theorems in cone metric spaces for ω-compatible mappings, Appl. Math. Comput. 217 (2010) 195–202.
[30] V. Berinde, Generalized coupled fixed-point theorems for mixed monotone mappings in partially ordered metric spaces, Nonlinear Anal. 74 (2011) 7347–7355.
[31] V. Lakshmikantham, L. Ćirić, Coupled fixed point theorems for nonlinear contractions in partially ordered metric spaces, Nonlinear Anal. 70 (2009) 4341–4349.
[32] H. Nashine, W. Shatanawi, Coupled common fixed-point theorems for a pair of commuting mappings in partially ordered complete metric spaces, Comput. Math. Appl. 62 (2011) 1984–1993.
[33] B. Samet, C. Vetro, Coupled fixed-point theorems for multi-valued nonlinear contraction mappings in partially ordered metric spaces, Nonlinear Anal. 74 (2011) 4260–4268.
[34] P. Semwal, R.C. Dimri, A Suzuki type coupled fixed point theorem for generalized multivalued mapping, Abstr. Appl. Anal. 2014 (2014) 1–8.
[35] A. Beiranvand, S. Moradi, M. Omid, H. Pazandeh, Two fixed-point theorems for special mappings, 2009. arXiv:0903.1504v1 [math.FA].
[36] H. Rahimi, B.E. Rhoades, S. Radenović, G. Soleimani Rad, Fixed and periodic point theorems for T-contractions on cone metric spaces, Filomat 27 (5) (2013) 881–888.
[37] F. Sabetghadam, H.P. Masiha, A.H. Sanatpour, Some coupled fixed-point theorems in cone metric space, Fixed Point Theory Appl. 2009 (2009) Article ID 125426.
[38] M. Küçük, M. Soyertem, Y. Küçük, On constructing total orders and solving vector optimization problems with total orders, J. Global Optim. 50 (2) (2011) 235–247.
[39] W. Shatanawi, Partially ordered cone metric spaces and coupled fixed point results, Comput. Math. Appl. 60 (2010) 2508–2515.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 420–428
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

On functionals of the Wiener process in a Banach space

Badri Mamporia[a,*], Omar Purtukhia[b]

[a] *Georgian Technical University N. Muskhelishvili Institute of Computational Mathematics, 4 Grigol Peradze st., Tbilisi 0131, Georgia*
[b] *Iv. Javakhishvili Tbilisi State University, A.Razmadze Mathematical Institute, Faculty of Exact and Natural Sciences, Department of Mathematics, Tbilisi, Georgia*

Received 1 June 2018; received in revised form 27 July 2018; accepted 29 July 2018
Available online 14 August 2018

Abstract

In development of stochastic analysis in a Banach space one of the main problem is to establish the existence of the stochastic integral from predictable Banach space valued (operator valued) random process. In the problem of representation of the Wiener functional as a stochastic integral we are faced with an inverse problem: we have the stochastic integral as a Banach space valued random element and we are looking for a suitable predictable integrand process. There are positive results only for a narrow class of Banach spaces with special geometry (UMD Banach spaces). We consider this problem in a general Banach space for a Gaussian functional.

Keywords: Wiener process; Functional of the Wiener process; Ito stochastic integral; Covariance operator in Banach space

1. Introduction and preliminaries

The problem of construction of the Ito stochastic integral in a Banach space is developing in three directions. In the first (relatively) direction the integrand is Banach space valued predictable random process and the stochastic integral is taken by the one dimensional Wiener process. In the second direction the integrand is operator valued (from Banach space to Banach space) predictable random process and stochastic integral is taken from Wiener process in a Banach space. In the third direction the integrand is operator-valued (from Hilbert space to Banach space) predictable process and stochastic integral is taken from cylindrical Wiener process in a Hilbert space. In all of these cases difficulties are the same. Therefore, for simplicity, in this article we consider the first case (Wiener process is one dimensional).

Using traditional methods, to find the suitable conditions that guarantee the construction of the stochastic integral is possible only in a very narrow class of Banach spaces. This class is so called UMD Banach spaces class (see survey in [1]). We consider the generalized stochastic integral for a wide class of predictable random processes and

* Corresponding author.
E-mail addresses: badrimamporia@yahoo.com (B. Mamporia), o.purtukhia@gmail.com (O. Purtukhia).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.07.007

the problem of existence of the stochastic integral we reduced to the problem of decomposability of the generalized random element (see [2]).

In this article we consider the problem of representation of the Wiener functional by the stochastic integral in an arbitrary separable Banach space. This problem is, in some sense, opposite to the problem of existence of the stochastic integral: here we have the stochastic integral as a random element and the problem is to find the integrand as a Banach space valued predictable process. In this direction there exists the following result in UMD Banach space case: under special condition every Wiener functional is represented by the stochastic integral and is generalized the Clark–Ocone formula of representation of the functional of the Wiener process by Malliavin derivative (see [3]).

Let X be a real separable Banach space. X^* – its conjugate, (Ω, B, P) – a probability space. Let $(W_t)_{t\in[0,1]}$ -be a real valued Wiener process. Denote by F_t^W the minimal σ-algebra generated by the random variables $(W_s)_{s\le t}$ ($F_t^W = \sigma(W_s, s \le t)$). The random element ξ is a weak second order if for all x^*, $E\langle \xi, x^*\rangle^2 < \infty$. Suppose that ξ is F_1^W- measurable i.e., ξ is the functional of the Wiener process. Our main aim is to represent the random element ξ by the Ito stochastic integral

$$\xi = E\xi + \int_0^1 f(t,\omega)dW_t,$$

where $f(t,\omega)$ is Banach space valued predictable random process. In the development of this difficult problem firstly, in this article, we consider the case when ξ is a Gaussian random element which with the Wiener process generates mutually Gaussian system. In this case the integrand (if it exists) will be nonrandom function. Remember, that the continuous linear operator $T : X^* \to L_2(\Omega, B, P)$ is called the generalized random element (GRE).[1]

Denote by $\mathcal{M}_1 := L(X^*, L_2(\Omega, B, P))$ the Banach space of GRE with the norm

$$\|T\|^2 = \sup_{\|x^*\|\le 1} E(Tx^*)^2.$$

We can realize the weak second order random element ξ as an element of $\mathcal{M}_1$, $T_\xi x^* = \langle \xi, x^*\rangle$, but not conversely: in infinite dimensional Banach space for all $T : X^* \to L_2(\Omega, B, P)$, there does not always exist the random element $\xi : \Omega \to X$ such that $Tx^* = \langle \xi, x^*\rangle$ for all $x^* \in X^*$. The problem of existence of such random element is the well known problem of decomposability of the GRE. Denote by $\mathcal{M}_2$ the linear normed space of all random elements of the weak second order with the norm

$$\|\xi\|^2 = \sup_{\|x^*\|\le 1} E\langle \xi, x^*\rangle^2.$$

Thus, we have $M_2 \subset M_1$. The family of random processes $(T_t)_{t\in[0,1]}$ is called the generalized random processes (GRP). In this paper we will consider the linear bounded operators $T : X^* \to L_2[0,1]$. In this special case instead of $L_2(\Omega, B, P)$, we have $L_2([0,1], B([0,1]), \lambda)$, Nevertheless we use the term GRE in this special case too. The decomposability problem is: for the GRE $T : X^* \to L_2[0,1]$ existence of the weak second order function $f : [0,1] \to X$ such that for all $x^* \in X^*$, $Tx^* = \langle f, x^*\rangle$ λ-a.e. Denote by $\mathcal{M}_1^\lambda$ the linear space of GRE $T : X^* \to L_2[0,1]$. $\mathcal{M}_1^\lambda$ is a Banach space with the norm

$$\|T\|^2 = \sup_{\|x^*\|\le 1} \int_0^1 [Tx^*(t)]^2 dt = \sup_{\|x^*\|\le 1} \|Tx^*\|_{L_2}^2.$$

Denote by $\mathcal{M}_2^\lambda$ the linear space of functions $f : [0,1] \to X$, such that $\int_0^1 \langle f(t), x^*\rangle^2 < \infty$. $\mathcal{M}_2^\lambda \subset \mathcal{M}_1^\lambda$.

2. Integral representation of functionals

For simplicity assume that $E\xi = 0$.

Proposition 2.1. *Let ξ be a F_1^W-measurable Gaussian random element. There exists a GRE $T : X^* \to L_2[0,1]$, such that for all $x^* \in X^*$,*

$$\langle \xi, x^*\rangle = \int_0^1 Tx^*(t)dW_t. \tag{2.1}$$

[1] sometimes it is used the terms: random linear function or cylindrical random element.

Proof. We use the technique developed in one dimensional case (see [4]). Denote

$$F_n^W = \sigma\{W_{\frac{1}{2^n}}, W_{\frac{2}{2^n}}, \dots, W_1\} = \sigma\{W_{\frac{1}{2^n}}, (W_{\frac{2}{2^n}} - W_{\frac{1}{2^n}}), \dots, (W_1 - W_{\frac{2^n-1}{2^n}})\}$$
$$= \sigma\{2^{\frac{n}{2}} g_1, 2^{\frac{n}{2}} g_2, \dots, 2^{\frac{n}{2}} g_{2^n}\} = \sigma\{g_1, g_2, \dots, g_{2^n}\},$$

where

$$g_1, g_2, \dots, g_{2^n}, \quad g_i = 2^{\frac{n}{2}}(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})$$

are independent, standard Gaussian random variables. Denote $\xi_n \equiv E(\xi|F_n^W)$ — the conditional mathematical expectation.

It is obvious that

$$\xi_n = E(\xi|F_n^W) = \sum_{i=0}^{2^n-1} E(\xi g_i) g_i$$
$$= \sum_{i=0}^{2^n-1} 2^n E(\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})) = \int_0^1 f_n(t) dW_t,$$

where

$$f_n(t) = \sum_{i=0}^{2^n-1} 2^n E\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}}) I_{(\frac{i}{2^n}, \frac{i+1}{2^n}]}(t).$$

As $(\xi_n)_{n\in N}$ is Gaussian martingale, we have

$$E\|\xi_n - \xi_m\|^2 = \left\| \int_0^1 f_n(t) dW_t - \int_0^1 f_m(t) dW_t \right\|^2 \to 0.$$

That is

$$\xi = \lim_{n\to\infty} \int_0^1 f_n(t) dW_t.$$

For all $x^* \in X^*$ denote $Tx^*(t) = \lim_{n\to\infty} \langle f_n(t), x^* \rangle$. We have

$$\sup_{\|x^*\|\le 1} \int_0^1 (Tx^*(t))^2 dt = \sup_{\|x^*\|\le 1} E\langle \xi, x^* \rangle^2 \le E\|\xi\|^2 < \infty.$$

Therefore, $T : X^* \to L_2[0, 1]$ is GRE and

$$\langle \xi, x^* \rangle = \int_0^1 Tx^*(t) dW_t. \quad \square$$

Remark 2.1. Note that if we have two representations of the F_1^W measurable random element ξ, by the stochastic integral

$$\langle \xi, x^* \rangle = \int_0^1 T_1 x^*(t) dW_t = \int_0^1 T_2 x^*(t) dW_t,$$

then

$$0 = \sup_{\|x^*\|\le 1} E\left(\int_0^1 (T_1 x^*(t) - T_2 x^*(t)) dW(t) \right)^2 = \sup_{\|x^*\|\le 1} \int_0^1 ((T_1 - T_2)x^*)^2 dt.$$

Hence, the representation (2.1) is unique.

The main problem is to find the function $f : [0, 1] \to X$ such that $Tx^*(t) = \langle f(t), x^* \rangle$ a. e. for all $x^* \in X^*$. In this case we will have $\xi = \int_0^1 f(t) dW_t$. The following example shows that, in general, such function (even $f : [0, 1] \to X^{**}$) does not exist.

Example. Let $X = c_0$ and denote $e_{k,n}(t) = a_n I_{(\frac{k-1}{n}, \frac{k}{n}]}(t)$, $n \in N, k = 1, 2, \ldots, n$, the sequence $(a_n)_{n\in N}$ we will choose later. Suppose $f(t) \equiv (e_{kn}(t))_{n\in N, k\le n}$, $t \in [0, 1]$. $f : [0, 1] \to R^N$. Consider the map $T(t) : l_1 \to R^1$, $t \in [0, 1]$.

$$T(t)\vec{\lambda} = \sum_{n=1}^{\infty}\sum_{k=1}^{n} \lambda_{nk} e_{nk}(t),$$

where $\vec{\lambda} = (\lambda_{nk})_{n\in N, k\le n} \in l_1$ ($\sum_{n=1}^{\infty}\sum_{k=1}^{n}\|\lambda_{nk}\| < \infty$).

We will show that T is the linear bounded operator from l_1 to $L_2[0, 1]$. We have

$$\int_0^1 (T(t)(\vec{\lambda}))^2 dt = \int_0^1 \sum_{n=1}^{\infty}\sum_{k=1}^{n}\sum_{m=1}^{\infty}\sum_{l=1}^{m} \lambda_{nk}\lambda_{ml} a_n a_m I_{(\frac{k-1}{n}, \frac{k}{n}]}(t) I_{(\frac{l-1}{m}, \frac{l}{m}]}(t).$$

If we demand that $|a_n| \le n^{\frac{1}{2}}$, we receive

$$\begin{aligned}\int_0^1 (T(t)(\vec{\lambda}))^2 dt &\le \int_0^1 \sum_{n=1}^{\infty}\sum_{k=1}^{n}\sum_{m=1}^{\infty}\sum_{l=1}^{m} \lambda_{nk}\lambda_{ml}(mn)^{\frac{1}{2}} \min\Big(\frac{1}{n}, \frac{1}{m}\Big)\\ &\le \sum_{n=1}^{\infty}\sum_{k=1}^{n}\sum_{m=1}^{\infty}\sum_{l=1}^{m}|\lambda_{nk}||\lambda_{ml}| = \|\vec{\lambda}\|_{l_1}^2.\end{aligned}$$

Therefore, $T : c_0^* \to L_2[0, 1]$ is bounded linear operator, $\|T\| \le 1$.

On the other hand,

$$\int_0^1 T(t)\vec{\lambda} dw_t = \sum_{n=1}^{\infty}\sum_{k=1}^{n} \lambda_{nk} a_n (W_{\frac{k}{n}} - W_{\frac{k-1}{n}}) = \langle\vec{\lambda}, \xi\rangle,$$

where $\xi := ((W_{\frac{k}{n}} - W_{\frac{k-1}{n}})a_n)_{n\in N, k\le n}$.

Let us check the sufficient condition of N. Vakhania (see [5], prop. 5.5.8) on the belonging of the random element ξ to the Banach space c_0. If we denote $\sigma_{k,n} = E[a_n(W_{\frac{k}{n}} - W_{\frac{k-1}{n}})]^2 = \frac{a_n^2}{n}$, then

$$\sum_{n=1}^{\infty}\sum_{k=1}^{n} \exp\Big(-\frac{\mu}{\sigma_{kn}}\Big) = \sum_{n=1}^{\infty}\sum_{k=1}^{n} \exp\Big(-\frac{n\mu}{a_n^2}\Big).$$

For example, if $a_n = n^{\frac{1}{3}}$, then the series $\sum_{n=1}^{\infty} n exp(-n^{\frac{1}{3}}\mu)$ converges for all fixed μ. Therefore, by the theorem N. Vakhania $\xi \in c_0$.

In this example we have $\langle\xi, x^*\rangle = \int_0^1 T(t)x^* dW_t$, where $T(t) : X^* \to L_2[0, 1]$. There does not exist $f : [0, 1] \to X$ or $f : [0, 1] \to X^{**}$ such that $T(t)x^* = \langle f(t), x^*\rangle$ (if $a_n \to 0$, then $f \in X$; if $a_n = 1, n = 1, 2, \ldots$, then $f \in X^{**}$; if $a_n \to \infty$, $a_n \le n^{\frac{1}{3}}$, then $f \bar{\in} X^{**}$).

Proposition 2.2. *Let $\xi : \Omega \to X$ be Gaussian F_1^W measurable random element. $T_\xi : X^* \to L_2[0, 1]$ be such that $\langle\xi, x^*\rangle = \int_0^1 Tx^*(t)dW(t)$, then $T_\xi \in \bar{\mathcal{M}}_2^\lambda \subseteq \mathcal{M}_1^\lambda$, $T^*T : X^* \to X \subset X^{**}$ is Gaussian covariance, there exists $a_\xi \in X$ such that*

$$\int_0^1 T_\xi x^*(t)dt = \langle a, x^*\rangle, \quad for\ \ all\ \ x^* \in X^*.$$

Proof. From the proof of Proposition 2.1, we have

$$f_n(t) = \sum_{i=0}^{2^n-1} 2^n E\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}}) I_{(\frac{i}{2^n}, \frac{i+1}{2^n}]}(t), \quad f_n \in \mathcal{M}_2^\lambda$$

and

$$\|f_n - T\|^2_{\mathcal{M}_1^\lambda} = \sup_{\|x^*\|\le 1} \int_0^1 T_\xi x^*(t) - \langle f_n(t), x^*\rangle^2 dt \to 0.$$

Therefore, $T_\xi \in \bar{\mathcal{M}}_2^\lambda$.

As $T_\xi \in \bar{\mathcal{M}}_2^\lambda$, by the proposition 1 of [2] $T_\xi^* T_\xi : X^* \to X$. This statement follows also from the equality

$$\langle T_\xi^* T_\xi x^*, y^*\rangle = \int_0^1 T_\xi x^*(t) T_\xi y^*(t) dt = E\langle \xi, x^*\rangle\langle \xi, y^*\rangle = \langle R_\xi x^*, y^*\rangle$$

and as $R_\xi : X^* \to X$, then $T^*T : X^* \to X \subset X^{**}$ too.

Further, we have

$$\int_0^1 f_n(t)dt = \sum_{i=0}^{2^n-1} 2^n E\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})2^{-n} = \sum_{i=0}^{2^n-1} E\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})$$
$$= E\xi\left(\sum_{i=0}^{2^n-1}(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})\right) = E\xi W_1.$$

Denote $a := E(\xi W_1)$, then we have

$$\int_0^1 T_\xi x^*(t)dt = \langle a, x^*\rangle, \quad for\ \ all\ \ x^* \in X^*.$$

Whereas that T_ξ is not X-valued function, the integral $\int_0^1 T_\xi(.)(t)dt$ is X-valued, that is, the Pettis integral from T_ξ exists. More clearly, if there exists $f : [0, 1] \to X$ such that $T_\xi x^* = \langle f(t), x^*\rangle$ for all $x^* \in X^*$, then a is the Pettis integral from f. □

Proposition 2.3. *For any X-valued F_1^W-measurable Gaussian random element ξ and $g \in L_2[0, 1]$, there exists $a_g \in X$ such that*

$$\int_0^1 T_\xi x^*(t)g(t)dt = \langle a_g, x^*\rangle, \ for\ \ all\ \ x^* \in X^*.$$

Proof. Consider the family of σ-algebras

$$F_n = \sigma\left\{\left(\frac{i}{2^n}, \frac{i+1}{2^n}\right], \ \ i = 0, 1, \ldots, 2^n\right\}.$$

It is evident that $F_n \subset F_{n+1}$, $n = 1, 2, \ldots$; $E(g|F_n) \to g \ \ in \ \ L_2[0, 1]$;

$$E(g|F_n)(t) = \sum_{i=0}^{2^n-1} 2^n \int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} g(s)ds I_{(\frac{i}{2^n}, \frac{i+1}{2^n}]}(t);$$

$$\int_0^1 T_\xi x^*(t)g(t)dt = \lim_{n\to\infty} \int_0^1 \sum_{i=0}^{2^n-1} 2^n E(\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}}))I_{(\frac{i}{2^n}, \frac{i+1}{2^n}]}(t) \sum_{i=0}^{2^n-1} g_i I_{(\frac{i}{2^n}, \frac{i+1}{2^n}]}(t)dt$$

$$= \lim_{n\to\infty} \int_0^1 \sum_{i=0}^{2^n-1} [2^n E(\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})g_i)]I_{(\frac{i}{2^n}, \frac{i+1}{2^n}]}(t)dt$$

$$= \lim_{n\to\infty} \sum_{i=0}^{2^n-1} [E[\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})g_i]]$$

$$= \lim_{n\to\infty} E\xi(\sum_{i=0}^{2^n-1}(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})g_i) = E\xi \int_0^1 g(t)dW_t,$$

where

$$g_i \equiv 2^n \int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} g(t)dt.$$

As

$$E\|\xi \int_0^1 g(t)dW_t\| \leq (E\|\xi\|^2)^{\frac{1}{2}} + (\int_0^1 g(t)^2dt)^{\frac{1}{2}} < \infty,$$

the Bochner integral $\int_0^1 T_\xi x^*(t)g(t)dt \equiv a_g$ exists. □

3. Representation of functionals in the form of series

Consider now the minimal subspace of $L_2(\Omega, B, P)$, consisting only random variables $(W_t, t \in [0, 1])$. Denote $G_0 = L(W_t, t \in [0, 1])$. G_0 consists only Gaussian Random variables. Denote $G := \bar{G}_0$. $G \subset L_2(\Omega, B, P)$ is a Hilbert space. As limit of Gaussian random variables is also Gaussian, G-contains only Gaussian random variables. Denote $\xi_t = E(\xi|F_t^W)$, $t \in [0, 1]$. $\xi_t : \Omega \to X$. $(\xi_t)_{t\in[0,1]}$ is a Gaussian process of independent increments. Firstly we consider one dimensional case and give the representation of the functional of the Wiener process by the sum of independent Gaussian random variable.

Theorem 3.1. *Let $\xi_t : \Omega \to R^1$ be F_t^W -measurable Gaussian random process, $f : [0, 1] \to R^1$ be such, that*

$$\xi(t) = \int_0^t f(\tau)dW(\tau).$$

For any orthonormal Basis $(e_n)_{n\in N}$ of $L_2[0, 1]$, there exists the sequence of independent, identically distributed, standard Gaussian random variables $(g_n)_{n\in N}$, such that

$$\xi_t = \sum_{k=1}^{\infty} \int_0^t f(\tau)e_k(\tau)d\tau g_k.$$

The convergence of the sum is a.s. uniformly in t.

Proof. As $(W_t)_{t\in[0,1]}$ has a.s. continuous sample paths, we can consider the corresponding $C[0, 1]$ valued random element $W : \Omega \to C[0, 1]$. The covariance operator $R_W : C[0, 1]^* \to C[0, 1]$,

$$R_W\varphi(t) = \int_0^1 \min(t, s)d\varphi(s)$$

admits the factorization (see [4] factorization lemma 3.1.1) through the Hilbert space $L_2[0, 1]$, $R_W = AA^*$:

$$A : L_2[0, 1] \to C[0, 1], \quad Ah(t) = \int_0^t h(\tau)d\tau, \quad A^*(t) : C[0, 1]^* \to L_2[0, 1],$$
$$A^*\delta_t = \chi_{[0,t]}(\tau), \quad \delta_t \in C[0, 1]^*, \quad \langle\delta_t, \psi\rangle = \psi(t), \quad \psi \in C[0, 1], \quad t \in [0, 1].$$

We have also another factorization of R_W: $R_W = T^*T$, $T : C[0, 1]^* \to G$, $T\delta_t = \langle W, \delta_t\rangle = W_t$. By the factorization lemma, there exists the isometric operator $I : G \to L_2[0, 1]$, such that $IT = A^*$. Therefore, $I(T\delta_t) = I(W_t) = A^*\delta_t = \chi_{[0,t]}(\tau)$. Thus,

$$EW_t g_k = \langle T\delta_t, g_k\rangle = \langle IT\delta_t, e_k\rangle = \langle\chi_{[0,t]}, e_k\rangle = \int_0^t e_k(\tau)d\tau.$$

Accordingly, we have $W_t = \sum_{k=1}^{\infty} E(W_t g_k)g_k$. Therefore

$$W_t = \sum_{k=1}^{\infty} \int_0^t e_k(\tau)d\tau g_k. \tag{3.1}$$

We have convergence of sum (3.1) in $C[0, 1]$. Therefore, this formula gives representation of the Wiener process by the a.s. uniformly in t convergent sum of independent Gaussian random variables.

Return now to the functional of the Wiener process $\xi = \int_0^1 f(t)dW_t$. $\xi_t = E(\xi|F_t^W) = \int_0^t f(\tau)dW_\tau$. $(\xi_t)_{t\in[0,1]}$ is a Gaussian process of independent increments. Dispersion of the random variable ξ_t is $\int_0^t f^2(\tau)d\tau$. As $(\xi_t)_{t\in[0,1]}$ is the process with continuous sample paths, we can consider corresponding random element $\xi : \Omega \to C[0,1]$. The covariance operator of this random element is $R_\xi : C[0,1]^* \to C[0,1]$,

$$R_\xi\varphi(t) = \int_0^1 \min(t,s)f(s)d\varphi(s), \quad R_\xi = AA^*, \quad A : L_2[0,1] \to C[0,1],$$
$$Ah(t) = \int_0^t h(\tau)f(\tau)d(\tau), \quad A^* : C[0,1]^* \to L_2[0,1], \quad A^*\delta_t = \chi_{[0,t]}(\tau)f(\tau).$$

It is clear that

$$\langle R\delta_t, \delta_s\rangle = \langle AA^*\delta_t, \delta_s\rangle = \langle T\delta_t, T\delta_s\rangle$$
$$= \int_0^1 \chi_{[0,t]}(\tau)f(\tau)\chi_{[0,s]}(\tau)f(\tau)d\tau = \int_0^{\min(t,s)} f^2(\tau)d\tau.$$

Further, we have:

$$\xi = \int_0^1 f(t)dW_t = \lim_{n\to\infty}\int_0^1 f_n(t)dW_t,$$

where

$$f_n(t) = \sum_{i=0}^{2^n-1} 2^n E\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(t).$$

Hence,

$$\xi = \lim_{n\to\infty}\sum_{i=0}^{2^n-1} 2^n E\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})$$
$$= \lim_{n\to\infty}\sum_{i=0}^{2^n-1} 2^n E\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})\sum_{k=1}^{\infty}\int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} e_k(\tau)d(\tau)g_k$$
$$= \lim_{n\to\infty}\sum_{k=1}^{\infty}\int_0^1 \sum_{i=0}^{2^n-1} 2^n E\xi(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})I_{(\frac{i}{2^n},\frac{i+1}{2^n}]}(\tau)e_k(\tau)d(\tau)g_k$$
$$= \lim_{n\to\infty}\sum_{k=1}^{\infty}\int_0^1 f_n(\tau)e_k(\tau)d(\tau)g_k = \sum_{k=1}^{\infty}\int_0^1 f(\tau)e_k(\tau)d(\tau)g_k,$$

as

$$\lim_{n\to\infty} E\left(\sum_{k=1}^{\infty}\int_0^1 f(\tau)e_k(\tau)d(\tau)g_k - \sum_{k=1}^{\infty}\int_0^1 f_n(\tau)e_k(\tau)d(\tau)g_k\right)^2$$
$$= \lim_{n\to\infty} E\left(\sum_{k=1}^{\infty}\int_0^1 (f_n(\tau) - f(\tau))e_k(\tau)d(\tau)g_k\right)^2$$
$$= \lim_{n\to\infty}\sum_{k=1}^{\infty}\left(\int_0^1 (f_n(\tau) - f(\tau))e_k(\tau)d(\tau)\right)^2$$
$$= \lim_{n\to\infty}\sum_{k=1}^{\infty}\langle(f(\tau) - f_n(\tau)), e_k\rangle^2 = \lim_{n\to\infty}\| f(\tau) - f_n(\tau)\|^2_{L_{[0,1]}} \to 0.$$

We received that

$$\xi = \int_0^1 f(t)dW_t = \sum_{k=1}^{\infty}\int_0^1 f(\tau)e_k(\tau)d(\tau)g_k.$$

Therefore

$$\xi_t = \sum_{k=1}^{\infty} \int_0^t f(\tau) e_k(\tau) d(\tau) g_k. \tag{3.2}$$

Consider the partial sum

$$(\xi_t)_n = \sum_{k=1}^{n} \int_0^t f(\tau) e_k(\tau) d(\tau) g_k$$

of independent, $C[0, 1]$-valued random elements. By the Ito–Nisio theorem [6], we have convergence of the last sum in $C[0, 1]$. Therefore, the last sum converges a.s. uniformly in t. □

Remark 3.1. N.Wiener [7] shows that the series

$$W_t \equiv g_0 t + \sum_{k=1}^{\infty} \sum_{n=2^{k-1}}^{2^k-1} n^{-1} g_n \sqrt{2} \sin \pi n t$$

converges (along an appropriate subsequence) uniformly in t to the Wiener process. Paul Levy [8] simplified Wiener's construction using Haar functions. Z.Ciesielsky [9] proved a.s. uniformly in t convergence of the series (3.1) in case, when $(e_k)_{k\in N}$ are Haar functions. K.Ito and M.Nisio [6] proved convergence of the sum (3.1) uniformly in t for an arbitrary orthonormal Basis $(e_k)_{k\in N}$ of $L_2[0, 1]$. Representation any fixed Wiener process by the sum (3.1) requires an additional effort, for example, to use the factorization lemma.

Consider now the corresponding problem for Banach space valued Wiener functional. Let $\xi : \Omega \to X$ be F_1^W measurable Gaussian random element. By Proposition 2.1, there exists GRE $T : X^* \to L_2[0, 1]$ such that

$$\langle \xi, x^* \rangle = \int_0^1 T x^*(t) dW_t,$$

for all $x^* \in X^*$. Denote $\xi_t = E(\xi | F_t^W)$. $(\xi_t)_{t\in[0,1]}$ is the Gaussian process with independent increments in a Banach space X. By continuity of the family $(F_t^W)_{t\in[0,1]}$ follows stochastically continuity of the process $(\xi_t)_{t\in[0,1]}$ and, as it is a stochastically continuous Gaussian process with independent increments, it has continuous sample paths (see [10]). Therefore, we can consider the corresponding random element in a Banach space $C([0, 1], X)$. The following theorem is similar to Theorem 3.1 for the case of Banach space X.

Theorem 3.2. *Let $\xi_t : \Omega \to X$ be F_t^W-measurable Gaussian random process, $T_\xi : X^* \to L_2[0, 1]$ be corresponding GRE, $\xi_t = \int_0^t T_\xi x^*(\tau) dW_\tau$. For any orthonormal basis $(e_n)_{n\in N}$ of $L_2[0, 1]$ there exists the sequence of independent, identically distributed, standard Gaussian random variables $(g_n)_{n\in N}$ such that*

$$\xi_t = \sum_{k=1}^{\infty} \int_0^t T_\xi(\tau) e_k(\tau) d\tau g_k. \tag{3.3}$$

The elements of the sum are X-valued and convergence of the sum is a.s. uniformly in t in X.

Proof. For any fixed orthonormal basis $(e_n)_{n\in N}$ of $L_2[0, 1]$, there exists the sequence of independent, identically distributed, standard Gaussian random variables $(g_n)_{n\in N}$ such that $W_t = \int_0^t e_k(\tau) d\tau g_k$. By Theorem 3.1, for any $x^* \in X^*$, we have

$$\langle \xi_t, x^* \rangle = \sum_{k=1}^{\infty} \int_0^t T_\xi x^*(\tau) e_k(\tau) d\tau g_k.$$

By Proposition 2.3 for an arbitrary $k \in N$, and $t \in [0, 1]$ $\int_0^t T_\xi(\tau) e_k(\tau) d\tau \equiv a_k(t)$ belongs to X. As $(\xi_t)_{t\in[0,1]}$ is X-valued Gaussian process of independent increments with continuous sample paths, we can consider corresponding random element with values in Banach space valued continuous functions: $\bar{\xi} : \Omega \to C([0, 1], X)$. In the process of proof of Proposition 2.1 we have considered the sequence

$$f_n(t) = \sum_{i=0}^{2^n-1} 2^n E\xi (W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}}) I_{(\frac{i}{2^n}, \frac{i+1}{2^n}]}(t),$$

for which we have $\|f_n - T_\xi\|_{M_1^\lambda} \to 0$ and

$$\xi = \int_0^1 T_\xi x^*(t) dW_t = \lim_{n\to\infty} \int_0^1 f_n(t) dW_t.$$

Analogously to the proof of Theorem 3.1 we obtain

$$\begin{aligned}\xi &= \lim_{n\to\infty} \int_0^1 f_n(t) dW_t = \lim_{n\to\infty} \sum_{i=0}^{2^n-1} 2^n E\xi (W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}})(W_{\frac{i+1}{2^n}} - W_{\frac{i}{2^n}}) \\ &= \lim_{n\to\infty} \sum_{k=1}^{\infty} \int_0^1 f_n(\tau) e_k(\tau) d(\tau) g_k = \sum_{k=1}^{\infty} \int_0^1 T_\xi(\tau) e_k(\tau) d(\tau) g_k\end{aligned}$$

According to Proposition 2.3 we have

$$\int_0^1 T_\xi(\tau) e_k(\tau) d(\tau) = \int_0^1 a_k(\tau) d(\tau) \in X.$$

Therefore, the last sum converges in X to ξ.

Consider the partial sum

$$\sum_{k=1}^{n} \int_0^t T_\xi(\tau) e_k(\tau) d(\tau) g_k \tag{3.4}$$

of independent $C([0, 1], X)$ valued random elements. By the Ito–Nisio Theorem this sum converges in $C([0, 1], X)$ to $(\xi_t)_{t\in[0,1]}$, therefore, we have a.s. uniformly in t convergence of the partial sum (3.4) to the sum (3.3). □

References

[1] J.M.A.M. van Neerven, M. Veraar, L. Weis, Stochastic Integration in UMD Banach Spaces –A Survey, in: Stochastic Analysis: A Series of Lectures, 2015, pp. 297–332.

[2] B. Mamporia, Stochastic differential equation for generalized stochastic processes in a Banach space, Teor. Veroyatn. Primen. 56 (4) (2011) 704—725; Theory of Probability and its Applications, SIAM, 2012, 602–620 (4).

[3] J. Maas, Jan, J. van Neerven, A Clark-Ocone formula in UMD banach spaces, Electron. Comm. Probab. 13 (2008) 151–164.

[4] R.S. Liptser, A.N. Shiriaev, Statistics of Random Processes, Nauka, Springer, Moscow, 1974 (in Russian), (2001).

[5] N.N. Vakhania, V.I. Tarieladze, S.A. Chobanyan, Probability Distributions on Banach Spaces, Nauka, Moscow, 1985. English translation: Reidel, Dordrecht, the Netherlands, 1987.

[6] K. Ito, M. Nisio, On the convergence of sums of independent Banach space valued random variables, Osaka J. Math. 5 (1968) 35–48.

[7] N. Wiener, Differential space, J. Math. Phys. 58 (1923) 131–174.

[8] P. Levy, Processus Stochastiques et Mouvement Brownien. Suivi d'une Note de M. Loéve, Gauthier-Villars, Paris, 1948 (French).

[9] Z. Ciesielski, Hölder conditions for realizations of Gaussian processes, Trans. Amer. Math. Soc. 99 (1961) 403–413.

[10] I.I. Gikhman, A.V. Skorokhod, The Theory of Stochastic Processes. II, Nauka, Moscow, 1973. Translated from the Russian by S Kotz. Reprint of the 1975 edition. Classics in Mathematics. Springer-Verlag, Berlin, 2004.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 429–439
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Connections between a system of forward–backward SDEs and backward stochastic PDEs related to the utility maximization problem

Michael Mania[a,b,*], Revaz Tevzadze[c,b]

[a] *A. Razmadze Mathematical Institute of Tbilisi State University, 6 Tamarashvili Str., Tbilisi 0176, Georgia*
[b] *Business School, Georgian American University, 8 Aleksidze Str., Tbilisi 0193, Georgia*
[c] *Institute of Cybernetics, 5 S. Euli Str., Tbilisi 0186, Georgia*

Received 21 May 2018; received in revised form 13 July 2018; accepted 12 August 2018
Available online 29 August 2018

Abstract

Connections between a system of Forward–Backward SDEs derived in Horst et al., (2014) and Backward Stochastic PDEs (Mania and Tevzadze, 2010) related to the utility maximization problem are established. Besides, we derive another version of Forward–Backward SDE of the same problem and prove the existence of solution.

Keywords: Utility maximization problem; Backward stochastic partial differential equation; Forward–backward stochastic differential equation

1. Introduction

We consider a financial market model, where the dynamics of asset prices is described by the continuous R^d-valued continuous semimartingale S defined on a complete probability space $(\Omega, \mathcal{F}, P)$ with filtration $F = (F_t, t \in [0, T])$ satisfying the usual conditions, where $\mathcal{F} = F_T$ and $T < \infty$. We work with discounted terms, i.e. the bond is assumed to be constant.

Let $U = U(x) : R \to R$ be a utility function taking finite values at all points of real line R such that U is continuously differentiable, increasing, strictly concave and satisfies the Inada conditions

$$U'(\infty) = \lim_{x\to\infty} U'(x) = 0, \quad U'(-\infty) = \lim_{x\to-\infty} U'(x) = \infty. \tag{1}$$

* Corresponding author.
E-mail addresses: misha.mania@gmail.com (M. Mania), tevzadze@cybernet.ge (R. Tevzadze).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.08.003

We also assume that U satisfies the condition of reasonable asymptotic elasticity (see [1] and [2] for a detailed discussion of these conditions), i.e.

$$\limsup_{x\to\infty}\frac{xU'(x)}{U(x)}<1,\quad \liminf_{x\to-\infty}\frac{xU'(x)}{U(x)}>1. \tag{2}$$

For the utility function U we denote by $\widetilde{U}$ its convex conjugate

$$\widetilde{U}(y)=\sup_x(U(x)-xy),\quad y>0. \tag{3}$$

Denote by $\mathcal{M}^e$ (resp. $\mathcal{M}^a$) the set of probability measures Q equivalent (resp. absolutely continuous) with respect to P such that S is a local martingale under Q.

Let $\mathcal{M}^a_U$ (resp. $\mathcal{M}^e_U$) be the convex set of probability measures $Q\in\mathcal{M}^a$ (resp. $\mathcal{M}^e$) such that

$$E\widetilde{U}\big(\frac{dQ_T}{dP_T}\big)<\infty. \tag{4}$$

It follows from proposition 4.1 of [3] that (4) implies $E\widetilde{U}\big(y\frac{dQ_T}{dP_T}\big)<\infty$ for any $y>0$.

Throughout the paper we assume that

$$\mathcal{M}^e_U\neq\emptyset. \tag{5}$$

The wealth process, determined by a self-financing trading strategy π and initial capital x, is defined as a stochastic integral

$$X^{x,\pi}_t=x+\int_0^t\pi_u dS_u,\quad 0\le t\le T.$$

We consider the utility maximization problem with random endowment H, where H is a liability that the agent must deliver at the terminal time T. H is an F_T-measurable random variable which for simplicity is assumed to be bounded (one can use also weaker assumption 1.6 from [4]). The value function $V(x)$ associated to the problem is defined by

$$V(x)=\sup_{\pi\in\Pi_x}E\bigg[U\bigg(x+\int_0^T\pi_u\,dS_u+H\bigg)\bigg], \tag{6}$$

where Π_x is a class of strategies which (following [2] and [4]) we define as the class of predictable S-integrable processes π such that $U(x+(\pi\cdot S)_T+H)\in L^1(P)$ and $\pi\cdot S$ is a supermartingale under each $Q\in\mathcal{M}^a_U$.

The dual problem to (6) is

$$\widetilde{V}(y)=\inf_{Q\in\mathcal{M}^e_U}E[\widetilde{U}(y\rho^Q_T)+y\rho^Q_T H],\quad y>0, \tag{7}$$

where $\rho^Q_t=dQ_t/dP_t$ is the density process of the measure $Q\in\mathcal{M}^e$ relative to the basic measure P.

It was shown in [4] that under assumptions (2) and (5) an optimal strategy $\pi(x)$ in the class Π_x exists. There exists also an optimal martingale measure $Q(y)$ to the problem (7), called the minimax martingale measure and by $\rho^*=(\rho^*_t(y),t\in[0,T])$ we denote the density process of this measure relative to the measure P.

It follows also from [4] that under assumptions (2) and (5) optimal solutions $\pi^*(x)\in\Pi_x$ and $Q(y)\in\mathcal{M}^e_U$ are related as

$$U'\bigg(x+\int_0^T\pi^*_u(x)dS_u+H\bigg)=y\rho^*_T(y),\quad P\text{-}a.s. \tag{8}$$

The continuity of S and the existence of an equivalent martingale measure imply that the structure condition is satisfied, i.e. S admits the decomposition

$$S_t=M_t+\int_0^t d\langle M\rangle_s\lambda_s,\quad \int_0^t\lambda^T_s\,d\langle M\rangle_s\lambda_s<\infty$$

for all t P-a.s., where M is a continuous local martingale and λ is a predictable process. The sign ^{T} here denotes the transposition.

Let us introduce the dynamic value function of problem (6) defined as

$$V(t,x) = \operatorname*{ess\,sup}_{\pi\in\Pi_x} E\left(U\left(x+\int_t^T \pi_u\, dS_u + H\right)\,\Big|\, F_t\right). \tag{9}$$

It is well known that for any $x \in R$ the process $(V(t,x), t \in [0,T])$ is a supermartingale admitting an RCLL (right-continuous with left limits) modification.

Therefore, using the Galtchouk–Kunita–Watanabe (GKW) decomposition, the value function is represented as

$$V(t,x) = V(0,x) - A(t,x) + \int_0^t \psi(s,x)\, dM_s + L(t,x),$$

where for any $x \in R$ the process $A(t,x)$ is increasing and $L(t,x)$ is a local martingale orthogonal to M.

Definition 1. We shall say that $(V(t,x), t \in [0,T])$ is a regular family of semimartingales if

(a) $V(t,x)$ is two-times continuously differentiable at x P- a.s. for any $t \in [0,T]$,

(b) for any $x \in R$ the process $V(t,x)$ is a special semimartingale with bounded variation part absolutely continuous with respect to an increasing predictable process $(K_t, t \in [0,T])$, i.e.

$$A(t,x) = \int_0^t a(s,x)\, dK_s,$$

for some real-valued function $a(s,x)$ which is predictable and K-integrable for any $x \in R$,

(c) for any $x \in R$ the process $V'(t,x)$ is a special semimartingale with the decomposition

$$V'(t,x) = V'(0,x) - \int_0^t a'(s,x)\, dK_s + \int_0^t \psi'(s,x)\, dM_s + L'(t,x).$$

where a', φ' and L' are partial derivatives of a, φ and L respectively.

If $F(t,x)$ is a family of semimartingales then $\int_0^T F(ds,\xi_s)$ denotes a generalized stochastic integral, or a stochastic line integral (see [5], or [6]). If $F(t,x) = xG_t$, where G_t is a semimartingale then the stochastic line integral coincides with the usual stochastic integral denoted by $\int_0^T \xi_s dG_s$ or $(\xi\cdot G)_T$.

It was shown in [7–9] (see, e.g., Theorem 3.1 from [9]) that if the value function satisfies conditions (a)–(c) then it solves the following BSPDE

$$\begin{aligned} V(t,x) &= V(0,x) \\ &+ \frac{1}{2}\int_0^t \frac{1}{V''(s,x)}(\varphi'(s,x)+\lambda(s)V'(s,x))^T\, d\langle M\rangle_s(\varphi'(s,x)+\lambda(s)V'(s,x)) \\ &+ \int_0^t \varphi(s,x)\, dM_s + L(t,x), \quad V(T,x) = U(x) \end{aligned} \tag{10}$$

and optimal wealth satisfies the SDE

$$X_t(x) = x - \int_0^t \frac{\varphi'(s,X_s(x)) + \lambda(s)V'(s,X_s(x))}{V''(s,X_s(x))} dS_s. \tag{11}$$

This assertion is a verification theorem since conditions are required directly on the value function $V(t,x)$ and not on the basic objects (on the asset price model and on the objective function U) only. In the case of complete markets [10] conditions on utility functions are given to ensure properties (a)–(c) and thus existence of a solution to the BSPDE (10), (11) is established. Note that the BSPDE (10), (11) is of the same form for random utility functions $U(\omega,x)$, for utility functions defined on half real line and properties (a)–(c) are also satisfied for standard (exponential, power and logarithmic) utility functions.

In the paper [11] a new approach was developed, where a characterization of optimal strategies to the problem (6) in terms of a system of Forward–Backward Stochastic Differential Equations (FBSDE) in the Brownian framework was given. The key observation was an existence of a stochastic process Y with $Y_T = H$ such that $U'(X_t + Y_t)$ is a martingale. The same approach was used in [12], where these results were generalized in semimartingale setting with

continuous filtration rejecting also some technical conditions imposed in [11]. The FBSDE for the pair (X, Y) (where X is the optimal wealth and Y the process mentioned above) is of the form (see, [12])

$$Y_t = Y_0 + \int_0^t \Big[\lambda_s^T \frac{U'(X_s+Y_s)}{U''(X_s+Y_s)} - \frac{1}{2}\lambda_s^T \frac{U'''(X_s+Y_s)U'(X_s+Y_s)^2}{U''(X_s+Y_s)^3} \tag{12}$$

$$+ Z_s^T\Big]d\langle M\rangle_s\lambda_s - \frac{1}{2}\int_0^t \frac{U'''(X_s+Y_s)}{U''(X_s+Y_s)}d\langle N\rangle_s + \int_0^t Z_s dM_s + N_t, \quad Y_T = H.$$

$$X_t = x - \int_0^t \big(\lambda_s \frac{U'(X_s+Y_s)}{U''(X_s+Y_s)} + Z_s\big)dS_s, \tag{13}$$

where N is a local martingale orthogonal to M.

Note that in [11] and [12] an existence of a solution of FBSDE (12), (13) is not proved, since not all conditions of corresponding theorems are formulated in terms of basic objects. E.g., in both papers it is imposed that $E(U'(X_T^* + H))^2 < \infty$ and it is not clear if an optimal strategy satisfying this condition exists. Note that in [11] in the case of complete markets an existence of a solution of FBSDE (12), (13) is proved under certain regularity assumptions on the objective function U.

One of our goal is to derive another version of FBSDE (12), (13) and to prove the existence of a solution which will imply the existence of a solution of the system (12), (13) also.

The second goal is to establish relations between equations BSPDE (10), (11) and FBSDE (12), (13). Solutions of these equations give constructions of the optimal strategy of the same problem. BSPDE (29), (30) can be considered as a generalization of Hamilton–Jacobi–Bellman equation to the non Markovian case and FBSDE (12), (13) is linked with the stochastic maximum principle (see [11]), although Eqs. (12)–(13) is not obtained directly from the maximum principle. It is well known that the relation between Bellman's dynamic programming and the Pontryagin's maximum principle in optimal control is of the form $\psi_t = V'(t, X_t)$, where V is the value function, X an optimal solution and ψ is an adjoint process (see, e.g. [13,14]). Therefore, somewhat similar relation between above mentioned equations should be expected. In particular, it is shown in Theorem 2, that the first components of solutions of these equations are related by the equality

$$Y_t = -\widetilde{U}'(V'(t, X_t)) - X_t.$$

In addition, conditions are given when the existence of a solution of BSPDE (29), (30) imply the existence of a solution of the system (12)–(13) and vice versa.

2. Another version of the forward–backward system (12)–(13)

In this section we derive another version of the Forward–Backward system (12), (13) in which the backward component P_t is a process, such that $P_t + U'(X_t)$ is a martingale.

Theorem 1. *Let utility function U be three-times continuously differentiable and let the filtration F be continuous. Assume that conditions* (2) *and* (5) *are satisfied. Then there exists a quadruple (P, ψ, L, X), where P and X are continuous semimartingales, ψ is a predictable M-integrable process and L is a local martingale orthogonal to M, that satisfies the FBSDE*

$$X_t = x - \int_0^t \frac{\lambda_s P_s + \lambda_s U'(X_s) + \psi_s}{U''(X_s)} dS_s, \tag{14}$$

$$P_t = P_0 + \int_0^t \left[\lambda_s - \frac{1}{2}U'''(X_s)\frac{\big(\lambda_s P_s + \lambda_s U'(X_s) + \psi_s\big)}{U''(X_s)^2}\right]^T d\langle M\rangle_s\big(\lambda_s P_s + \lambda_s U'(X_s) + \psi_s\big)$$

$$+ \int_0^t \psi_s dM_s + L_t, \qquad P_T = U'(X_T + H) - U'(X_T). \tag{15}$$

In addition the optimal strategy is expressed as

$$\pi_t^* = -\frac{\lambda_t P_t + \lambda_t U'(X_t) + \psi_t}{U''(X_t)} \tag{16}$$

and the optimal wealth X^* coincides with X.

Proof. Define the process

$$P_t = E(U'(X_T^* + H)/F_t) - U'(X_t^*). \tag{17}$$

Note that the integrability of $U'(X_T^* + H)$ follows from the duality relation (8). It is evident that $P_T = U'(X_T^* + H) - U'(X_T^*)$.

Since U is three-times differentiable, $U'(X_t^*)$ is a continuous semimartingale and P_t admits the decomposition

$$P_t = P_0 + A_t + \int_0^t \psi_u dM_u + L_t, \tag{18}$$

where A is a predictable process of finite variations and L is a local martingale orthogonal to M.

Since ρ_t^* is the density of a martingale measure, it is of the form $\rho_t^* = \mathcal{E}_t(-\lambda \cdot M + R)$, $R \perp M$. Therefore, (8) and (17) imply that

$$E(U'(X_T^* + H)/F_t) = y\rho_t^* = y - \int_0^t \lambda_s y \rho_s^* dM_s + \tilde{R}_t$$

$$= y - \int_0^t \big(P_s + U'(X_s^*)\big)\lambda_s dM_s + \tilde{R}_t, \tag{19}$$

where $y = EU'(X_T^* + H)$ and $\tilde{R}$ is a local martingale orthogonal to M.

By definition of the process P_t, using the Itô formula for $U'(X_t^*)$ and taking decompositions (18), (19) in mind, we obtain

$$P_0 + A_t + \int_0^t \psi_s dM_s + L_t = y - \int_0^t \big(P_s + U'(X_s^*)\big)\lambda_s dM_s + \tilde{R}_t -$$

$$- U'(x) - \int_0^t U''(X_s^*)\pi_s^{*T} d\langle M\rangle_s \lambda_s - \frac{1}{2}\int_0^t U'''(X_s^*)\pi_s^{*T} d\langle M\rangle_s \pi_s^*$$

$$- \int_0^t U''(X_s^*)\pi_s^* dM_s. \tag{20}$$

Equalizing the integrands of stochastic integrals with respect to dM we have that $\mu^{\langle M\rangle}$-a.e.

$$\pi_t^* = -\frac{\lambda_t P_t + \lambda_t U'(X_t^*) + \psi_t}{U''(X_t^*)} \tag{21}$$

Equalizing the parts of finite variations in (20) we get

$$A_t = -\int_0^t \Big(U''(X_s^*)\lambda_s + \frac{1}{2}U'''(X_s^*)\pi_s^*\Big)^T d\langle M\rangle_s \pi_s^* \tag{22}$$

and from (21), substituting the expression for π^* in (22) we obtain that

$$A_t = \int_0^t \left[\lambda_s - \frac{1}{2}U'''(X_s)\frac{\big(\lambda_s P_s + \lambda_s U'(X_s) + \psi_s\big)}{U''(X_s)^2}\right]^T d\langle M\rangle_s \big(\lambda_s P_s + \lambda_s U'(X_s) + \psi_s\big) \tag{23}$$

Therefore, (23) and (18) imply that P_t satisfies Eq. (15). Integrating both parts of equality (21) with respect to dS and adding the initial capital we obtain Eq. (14) for the optimal wealth. □

Corollary. Let conditions of Theorem 1 be satisfied. Then there exists a solution of FBSDE (12), (13). In particular, if the pair (X, P) is a solution of (14), (15), then the pair (X, Y), where

$$Y_t = -\tilde{U}'(P_t + U'(X_t)) - X_t,$$

satisfies the FBSDE (12), (13).

Conversely, if the pair (X, Y) solves the FBSDE (12), (13), then $(X_t, P_t = U'(X_t + Y_t) - U'(X_t))$ satisfies (14), (15).

3. Relations between BSPDE (10)–(11) and FBSDE (12)–(13)

To establish relations between equations BSPDE (10), (11) and FBSDE (12), (13) we need the following

Definition 2 *([15])*. The function $u(t,x)$ is called a decoupling field of the FBSDE (12), (13) if

$$u(T,x)=H,\quad a.s. \tag{24}$$

and for any $x\in R$, $s,\tau\in R_+$ such that $0\le s<\tau\le T$ the FBSDE

$$\begin{aligned} Y_t &= u(s,x) \\ &+\int_s^t\Big(\lambda_r^T\frac{U'(X_r+Y_r)}{U''(X_r+Y_r)}-\frac{1}{2}\lambda_r^T\frac{U'''(X_r+Y_r)U'(X_r+Y_r)^2}{U''(X_r+Y_r)^3}+Z_r^T\Big)d\langle M\rangle_r\lambda_r \\ &-\frac{1}{2}\int_s^t\frac{U'''(X_r+Y_r)}{U''(X_r+Y_r)}d\langle N\rangle_r+\int_s^t Z_r dM_r+N_t-N_s,\quad Y_\tau=u(\tau,X_\tau), \end{aligned} \tag{25}$$

$$X_t=x-\int_s^t\Big(\lambda_r\frac{U'(X_r+Y_r)}{U''(X_r+Y_r)}+Z_r\Big)dS_r, \tag{26}$$

has a solution (Y,Z,N,X) satisfying

$$Y_t=u(t,X_t),\quad a.s. \tag{27}$$

for all $t\in[s,\tau]$. We mean that all integrals are well defined.

We shall say that $u(t,x)$ is a regular decoupling field if it is a regular family of semimartingales (in the sense of Definition 1).

If we differentiate equation BSPDE (10) at x (assuming that all derivatives involved exist), we obtain the BSPDE

$$\begin{aligned} V'(t,x) &= V'(0,x) \\ &+\frac{1}{2}\int_0^t\left(\frac{(\varphi'(s,x)+\lambda_s V'(s,x))^T}{V''(s,x)}\,d\langle M\rangle_s(\varphi'(s,x)+\lambda_s V'(s,x))\right)' \\ &+\int_0^t\varphi'(s,x)\,dM_s+L'(t,x),\quad V'(T,x)=U'(x+H). \end{aligned} \tag{28}$$

Thus, we consider the following BSPDE

$$\begin{aligned} V'(t,x) &= V'(0,x)+\int_0^t\Bigg[\frac{(V''(s,x)\lambda_s+\varphi''(s,x))^T}{V''(s,x)} \\ &-\frac{1}{2}V'''(s,x)\frac{(V'(s,x)\lambda_s+\varphi'(s,x))^T}{V''(s,x)}\Bigg]d\langle M\rangle_s(V'(s,x)\lambda_s+\varphi'(s,x)) \\ &+\int_0^t\varphi'(s,x)\,dM_s+L'(t,x),\quad V'(T,x)=U'(x+H), \end{aligned} \tag{29}$$

where the optimal wealth satisfies the same SDE

$$X_t(x)=x-\int_0^t\frac{\varphi'(s,X_s(x))+\lambda(s)V'(s,X_s(x))}{V''(s,X_s(x))}dS_s. \tag{30}$$

The FBSDE (12), (13) is equivalent, in some sense, to BSPDE (29), (30) and the following statement establishes a relation between these equations.

Theorem 2. *Let the utility function $U(x)$ be three-times continuously differentiable and let the filtration F be continuous.*

(a) If $V'(t,x)$ is a regular family of semimartingales and $(V'(t,x), \varphi'(t,x), L'(t,x), X_t)$ is a solution of BSPDE (29), (30)*, then the quadruple*
(Y_t, Z_t, N_t, X_t), where

$$Y_t = -\widetilde{U}'(V'(t,X_t)) - X_t, \tag{31}$$

$$Z_t = \lambda_t \widetilde{U}'(V'(t,X_t)) + \frac{\varphi'(t,X_t) + \lambda_t V'(t,X_t)}{V''(t,X_t)}, \tag{32}$$

$$N_t = -\int_0^t \widetilde{U}''(V'(s,X_s))d\Big(\int_0^s L'(dr,X_r)\Big), \tag{33}$$

will satisfy the FBSDE (12), (13)*. Moreover, the function $u(t,x) = -\widetilde{U}'(V'(t,x)) - x$ will be the decoupling field of this FBSDE.*

(b) Let $u(t,x)$ be a regular decoupling field of FBSDE (12), (13) *and let $(U'(X_t + Y_t), s \le t \le T)$ be a true martingale for every $s \in [0,T]$. Then $(V'(t,x), \varphi'(t,x), L'(t,x), X)$ will be a solution of BSPDE* (29), (30) *and following relations hold*

$$V'(t,x) = U'(x + u(t,x)), \quad \textit{hence} \quad V'(t,X_t) = U'(X_t + Y_t), \tag{34}$$

$$\varphi'(t,X_t) = (Z_t + \lambda_s \frac{U'(X_t+Y_t)}{U''(X_t+Y_t)})V''(t,X_t) - \lambda_t U'(X_t+Y_t), \tag{35}$$

$$\int_0^t L'(ds,X_s) = \int_0^t U''(X_s+Y_s)dN_s, \tag{36}$$

where $\int_0^t L'(ds,X_s)$ is a stochastic line integral with respect to the family $(L'(t,x), x \in R)$ along the process X.

Proof. (a) It follows from BSPDE (29), (30) and from the Itô – Ventzel formula that $V'(t,X_t)$ is a local martingale with the decomposition

$$V'(t,X_t) = V'(0,x) - \int_0^t \lambda_s V'(s,X_s)dM_s + \int_0^t L'(ds,X_s). \tag{37}$$

Let $Y_t = -\widetilde{U}'(V'(t,X_t)) - X_t$. Since U is three-times differentiable (hence so is $\tilde{U}$), Y_t will be a special semimartingale and by GKW decomposition

$$Y_t = Y_0 + A_t + \int_0^t Z_u dM_u + N_t, \tag{38}$$

where A is a predictable process of finite variations and N is a local martingale orthogonal to M.

The definition of the process Y, decompositions (37), (38) and the Itô formula for $\widetilde{U}'(V'(t,X_t))$ imply that

$$A_t + \int_0^t Z_s dM_s + N_t = \tag{39}$$

$$= \int_0^t \widetilde{U}''(V'(s,X_s))V'(s,X_s)\lambda_s dM_s - \int_0^t \widetilde{U}''(V'(s,X_s))d\Big(\int_0^s L'(dr,X_r)\Big)$$

$$- \frac{1}{2}\int_0^t \widetilde{U}'''(V'(s,X_s))V'(s,X_s)^2 \lambda_s^T d\langle M\rangle_s \lambda_s - \frac{1}{2}\int_0^t \widetilde{U}'''(V'(s,X_s))d\langle \int_0^{\cdot} L'(dr,X_r)\rangle_s$$

$$+ \int_0^t \frac{\lambda_s V'(s,X_s) + \varphi'(s,X_s)}{V''(s,X_s)} dM_s + \int_0^t \frac{\lambda_s^T V'(s,X_s) + \varphi'(s,X_s)^T}{V''(s,X_s)} d\langle M\rangle_s \lambda_s.$$

Equalizing the integrands of stochastic integrals with respect to dM in (39) we have that $\mu^{\langle M\rangle}$-a.e.

$$Z_s = \frac{\lambda_s V'(s, X_s) + \varphi'(s, X_s)}{V''(s, X_s)} + \widetilde{U}''(V'(s, X_s))V'(s, X_s)\lambda_s. \tag{40}$$

Equalizing the orthogonal martingale parts we get P-a.s.

$$N_t = -\int_0^t \widetilde{U}''(V'(s, X_s))d\Big(\int_0^s L'(dr, X_r)\Big). \tag{41}$$

Equalizing the parts of finite variations in (39) we have

$$A_t = \int_0^t \frac{\lambda_s^T V'(s, X_s) + \varphi'(s, X_s)^T}{V''(s, X_s)} d\langle M\rangle_s \lambda_s \tag{42}$$

$$-\frac{1}{2}\int_0^t \widetilde{U}'''(V'(s, X_s))V'(s, X_s)^2\lambda_s^T d\langle M\rangle_s\lambda_s - \frac{1}{2}\int_0^t \widetilde{U}''''(V'(s, X_s))d\langle \int_0^{\cdot} L'(dr, X_r)\rangle_s$$

and by equalities (40), (41) we obtain from (42) that

$$A_t = \int_0^t \left(Z_s - \widetilde{U}''(V'(s, X_s))V'(s, X_s)\lambda_s - \frac{1}{2}\widetilde{U}'''(V'(s, X_s))V'(s, X_s)^2\lambda_s\right)^T d\langle M\rangle_s\lambda_s$$

$$-\frac{1}{2}\int_0^t \frac{\widetilde{U}''''(V'(s, X_s))}{\widetilde{U}''(V'(s, X_s))^2} d\langle N\rangle_s. \tag{43}$$

Therefore, using the duality relations

$$V'(t, X_t) = U'(X_t + Y_t),$$

$$\widetilde{U}''(V'(t, X_t)) = -\frac{1}{U''(X_t + Y_t)},$$

$$\widetilde{U}'''(V'(t, X_t)) = -\frac{U'''(X_t + Y_t)}{(U''(X_t + Y_t))^3},$$

we obtain from (43) that

$$A_t = \int_0^t \left(\lambda_s \frac{U'(X_s + Y_s)}{U''(X_s + Y_s)} - \frac{1}{2}\lambda_s \frac{U'''(X_s + Y_s)U'(X_s + Y_s)^2}{U''(X_s + Y_s)^3} + Z_s\right)^T d\langle M\rangle_s\lambda_s$$

$$-\frac{1}{2}\int_0^t \frac{U'''(X_s + Y_s)}{U''(X_s + Y_s)} d\langle N\rangle_s \tag{44}$$

Thus, (38) and (44) imply that Y satisfies Eq. (12).

Since

$$\widetilde{U}''(V'(s, X_s))V'(s, X_s) = -\frac{1}{U''(X_s + Y_s)},$$

from (30) and (40) we obtain Eq. (13) for the optimal wealth.

The proof that the function $u(t, x) = -\widetilde{U}'(V'(t, x)) - x$ is the decoupling field of the FBSDE (12) is similar. One should take integrals from s to t and use the same arguments.

(b) Since the quadruple $(Y^{s,x}, Z^{s,x}, N^{s,x}, X^{s,x})$ satisfies the FBSDE (25), (26), it follows from the Itô formula that for any $t \geq s$

$$U'(X_t^{s,x} + Y_t^{s,x}) = U'(x + u(s, x)) - \int_s^t \lambda_r U'(X_r^{s,x} + Y_r^{s,x})dM_r \tag{45}$$

$$+ \int_s^t U''(X_r^{s,x} + Y_r^{s,x})dN_r.$$

Thus $U'(X_t^{s,x}+Y_t^{s,x})$, $t\geq s$, is a local martingale and a true martingale by assumption. Therefore, it follows from (24) and (27) that

$$U'(X_t^{s,x}+Y_t^{s,x})=E(U'(X_T^{s,x}+H)/F_t)=V'(t,X_t^{s,x}), \tag{46}$$

where the last equality is proved similarly to [3]. For $t=s$ we obtain that

$$U'(x+u(s,x))=V'(s,x), \tag{47}$$

hence

$$u(t,x)=-\tilde{U}'(V'(t,x))-x. \tag{48}$$

Since $U(x)$ is three-times differentiable and $u(t,x)$ is a regular decoupling field, equality (47) implies that $V'(t,x)$ will be a regular family of semimartingales. Therefore, using the Itô – Ventzel formula for $V'(t,X_t^{s,x})$ and equalities (45), (46) we have

$$\begin{aligned}
&\int_s^t\big[\varphi'(r,X_r^{s,x})-V''(r,X_r^{s,x})(\lambda_s\frac{U'(X_r^{s,x}+Y_r^{s,x})}{U''(X_r^{s,x}+Y_r^{s,x})}+Z_r^{s,x})\big]dM_r \qquad (49)\\
&+\int_s^t L'(dr,X_r)+\int_s^t a'(r,X_r^{s,x})dK_r\\
&-\int_s^t\big(\lambda_r\frac{U'(X_r^{s,x}+Y_r^{s,x})}{U''(X_r^{s,x}+Y_r^{s,x})}+Z_r^{s,x}\big)^T d\langle M\rangle_r(V''(r,X_r^{s,x})\lambda_r+\varphi''(r,X_r^{s,x}))\\
&-\frac{1}{2}\int_s^t(V'''(r,X_r^{s,x}))\big(\lambda_r\frac{U'(X_r^{s,x}+Y_r^{s,x})}{U''(X_r^{s,x}+Y_r^{s,x})}+Z_r^{s,x}\big)^T d\langle M\rangle_r\big(\lambda_r\frac{U'(X_r^{s,x}+Y_r^{s,x})}{U''(X_r^{s,x}+Y_r^{s,x})}+Z_r^{s,x}\big)\\
&=-\int_s^t\lambda_r U'(X_r^{s,x}+Y_r^{s,x})dM_r+\int_s^t U''(X_r^{s,x}+Y_r^{s,x})dN_r.
\end{aligned}$$

Equalizing the integrands of stochastic integrals with respect to dM in (49) we have that μ^K-a.e.

$$Z_r^{s,x}=\frac{\lambda_r V'(r,X_r^{s,x})+\varphi'(r,X_r^{s,x})}{V''(r,X_r^{s,x})}-\lambda_r\frac{U'(X_r^{s,x}+Y_r^{s,x})}{U''(X_r^{s,x}+Y_r^{s,x})}. \tag{50}$$

Equalizing the parts of finite variations in (49), taking (50) in mind, we get that for any $t>s$

$$\begin{aligned}
\int_s^t a'(r,X_r^{s,x})dK_r=\int_s^t\big[&\frac{(V''(r,X_r^{s,x})\lambda_r+\varphi''(r,X_r^{s,x}))}{V''(r,X_r^{s,x})} \qquad (51)\\
&-\frac{1}{2}V'''(r,X_r^{s,x})\frac{(V'(r,X_r^{s,x})\lambda_r+\varphi'(r,X_r^{s,x}))}{V''(r,X_r^{s,x})^2}\big]^T d\langle M\rangle_r(V'(r,X_r^{s,x})\lambda_r+\varphi'(r,X_r^{s,x})).
\end{aligned}$$

Let $\tau_s(\varepsilon)=\inf\{t\geq s: K_t-K_s\geq\varepsilon\}$. Since $\langle M^i,M^j\rangle\ll\tilde{K}$ for any $1\leq i,j\leq d$, where $\tilde{K}=\sum_{i=1}^d\langle M^i\rangle$, taking an increasing process $K+\tilde{K}$ (which we denote again by K), without loss of generality we can assume that $\langle M\rangle\ll K$ and denote by C_t the matrix of Radon–Nikodym derivatives $C_t=\frac{d\langle M\rangle_t}{dK_t}$. Then from (51)

$$\begin{aligned}
\int_s^{\tau_s(\varepsilon)}\Big[&\frac{(V''(r,X_r^{s,x})\lambda_r+\varphi''(r,X_r^{s,x}))^T C_r(V'(r,X_r^{s,x})\lambda_r+\varphi'(r,X_r^{s,x}))}{V''(r,X_r^{s,x})} \qquad (52)\\
&-\frac{1}{2}V'''(r,X_r^{s,x})\frac{(V'(r,X_r^{s,x})\lambda_r+\varphi'(r,X_r^{s,x}))^T C_r(V'(r,X_r^{s,x})\lambda_r+\varphi'(r,X_r^{s,x}))}{V''(r,X_r^{s,x})^2}\\
&-a'(r,X_r^{s,x})\Big]dK_r=0.
\end{aligned}$$

Since for any $x \in R$ the process $X_r^{s,x}$ is a continuous function on $\{(r,s), r \geq s\}$ with $X_s^{s,x} = x$ (as a solution of Eq. (26)) and $V'(t,x)$ is a regular family of semimartingales, dividing equality (52) by ε and passing to the limit as $\varepsilon \to 0$ from [7] (Proposition B1) we obtain that for each x

$$a'(s,x) = \frac{(V''(s,x)\lambda_s + \varphi''(s,x))^T C_s (V'(s,x)\lambda_s + \varphi'(s,x))}{V''(s,x)} \tag{53}$$

$$- \frac{1}{2} V'''(s,x) \frac{(V'(s,x)\lambda_s + \varphi'(s,x))^T C_s (V'(s,x)\lambda_s + \varphi'(s,x))}{V''(s,x)^2}$$

$$= \frac{1}{2} \left(\frac{(V'(s,x)\lambda_s + \varphi'(s,x))^T C_s (V'(s,x)\lambda_s + \varphi'(s,x))}{V''(s,x)} \right)', \ \mu^K - a.e.,$$

which implies that $V'(t,x)$ satisfies the BSPDE

$$V'(t,x) = V'(0,x) + \frac{1}{2} \int_0^t \left(\frac{(V'(s,x)\lambda_s + \varphi'(s,x))^T C_s (V'(s,x)\lambda_s + \varphi'(s,x))}{V''(s,x)} \right)' dK_s$$

$$+ \int_0^t \varphi'(s,x)\, dM_s + L'(t,x), \quad V'(T,x) = U'(x+H). \quad \square \tag{54}$$

Remark 1. In the proof of the part (a) of the theorem we need the condition that $V'(t,x)$ is a regular family of semimartingales only to show equality (37) and to obtain representation (33). Equality (37) can be proved without this assumption (replacing the stochastic line integral by a local martingale orthogonal to M) from the duality relation

$$V'(t, X_t(x)) = \rho_t(y), \quad y = V'(x),$$

where $\rho_t(y)/y$ is the density of the minimax martingale measure (see [2] and [4] for the version with random endowment). Since $\rho_t(y)/y$ is representable in the form $\mathcal{E}(-\lambda \cdot M + D)$, for a local martingale D orthogonal to M, using the Dolean Dade equation we have

$$V'(t, X_t) = \rho_t = y - \int_0^t \lambda_s \rho_s dM_s + \int_0^t \rho_s dD_s =$$

$$= 1 - \int_0^t \lambda_s V'(s, X_s) dM_s + R_t,$$

where $R_t \equiv (Z \cdot D)_t$ is a local martingale orthogonal to M. Further the proof will be the same if we always use a local martingale R_t instead of the stochastic line integral $\int_0^t (L'(ds, X_s)$. Hence the representation (33) will be of the form

$$N_t = - \int_0^t \widetilde{U}''(V'(s, X_s)) dR_t.$$

Remark 2. It follows from the proof of Theorem 2, that if a regular decoupling field for the FBSDE (12), (13) exists, then the second component of the solution Z is also of the form $Z_t = g(\omega, t, X_t)$ for some measurable function g and if we assume that any orthogonal to M local martingale L is represented as a stochastic integral with respect to the given continuous local martingale $M^\perp$, then the third component N of the solution will take the same form $N_t = \int_0^t g^\perp(s, X_s) dM_s^\perp$, for some measurable function $g^\perp$.

Remark 3. Similarly to Theorem 2(b) one can show that $u(t,x) = V'(t,x) - U'(x)$ is the decoupling field of (14), (15).

Acknowledgments

We would like to thank the referees for careful reading.

References

[1] D. Kramkov, W. Schachermayer, The asymptotic elasticity of utility functions and optimal investment in incomplete markets, Ann. Appl. Probab. 9 (9) (1999) 904–950.

[2] W. Schachermayer, A super-martingale property of the optimal portfolio process, Finance Stoch. 7 (4) (2003) 433–456.

[3] W. Schachermayer, Optimal investment in incomplete markets when wealth may be negative, Ann. Appl. Probab. 11 (3) (2001) 694–734.

[4] M.P. Owen, G. Zitkovich, Optimal investment with an unbounded random endowment and utility-based pricing, Math. Finance 19 (1) (2009) 129–159.

[5] H. Kunita, Stochastic Flows and Stochastic Differential Equations, Cambridge University Press, 1990.

[6] R. Chitashvili, Martingale ideology in the theory of controlled stochastic processes, in: Probability Theory and Mathematical Statistics, Proc. 4th USSR-Jap. Symp., Tbilisi, 1982, in: Lecture Notes in Mathematics, N. 1021, Springer, Berlin etc, 1983, pp. 73–92.

[7] M. Mania, R. Tevzadze, Backward stochastic PDE and imperfect hedging, Int. J. Theor. Appl. Finance 6 (7) (2003) 663–692.

[8] M. Mania, R. Tevzadze, Backward stochastic partial differential equations related to utility maximization and hedging, J. Math. Sci. 153 (3) (2008) 292–376.

[9] M. Mania, R. Tevzadze, Backward stochastic PDEs related to utility maximization problem, Georgian Math. J. 17 (4) (2010) 705–741.

[10] M. Mania, R. Tevzadze, On regularity of primal and dual dynamic value functions related to investment problem and their representations as BSPDE solutions, SIAM J. Financial Math. 8 (2017) 483–503.

[11] U. Horst, Y. Hu, P. Imkeller, A. Reveillac, J. Zhang, Forward-backward systems for expected utility maximization, Stochastic Process. Appl. 124 (5) (2014) 1813–1848.

[12] M. Santacroce, B. Trivellato, Forward backward semimartingale systems for utility maximization, SIAM J. Control Optim. 52 (6) (2014) 3517–3537.

[13] J.M. Bismut, Conjugate convex functions in optimal stochastic control, J. Math. Anal. Appl. 44 (1973) 384–404.

[14] X.Y. Zhou, The connection between the maximum principle and dynamic programming in stochastic control, Stoch. Stoch. Rep. 31 (1–4) (1990) 1–15.

[15] A. Fromm, P. Imkeller, Existence, Uniqueness and regularity of decoupling fields to multidimensional fully coupled FBSDEs, arXiv:1310.04 99v2, 2013.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 440–447
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Nonparametric density estimation based on the scaled Laplace transform inversion

Fairouz Elmagbri, Robert M. Mnatsakanov*

Department of Statistics, P.O. Box 6330, West Virginia University, Morgantown, WV 26506, USA

Received 17 May 2018; received in revised form 15 July 2018; accepted 7 September 2018
Available online 23 October 2018

Abstract

New nonparametric procedure for estimating the probability density function of a positive random variable is suggested. Asymptotic expressions of the bias term and Mean Squared Error are derived. By means of graphical illustrations and evaluating the Average of L_2-errors we conducted comparisons of the finite sample performance of proposed estimate with the one based on kernel density method.

Keywords: Laplace transform; Nonparametric estimation; Mean squared error; Kernel density estimation

1. Introduction

In this paper we propose new nonparametric procedure to estimate the density function of a positive random variable that is based on the moment-recovered approach proposed in Mnatsakanov et al. [1].

There are many articles devoted to investigation of the properties of nonparametric density estimators based on the kernel-smoothing technique. A common method for estimating a density function is obtained by using a fixed symmetric kernel density estimate (KDE) with bounded (unbounded) support and choosing the optimal bandwidth (Silverman [2]). The properties of such estimator have been studied by Mielniczuk [3], Zhang [4], among many others. There are series of papers where the asymmetric kernels are used (Bouezmarni and Rolin [5], Chen [6], Mnatsakanov and Ruymgaart [7], and Mnatsakanov and Sarkisian [8]).

Also there are several well-known techniques, e.g. Lejeune and Sarda [9] as well as Nielsen at al. [10] worked on the local linear estimation. Jones [11] and Jones and Forster [12] investigated the performance of the estimates based on boundary kernels, Müller [13] studied the smoothed optimum kernels, and Marron and Ruppert [14] used transformation approach.

* Corresponding author.
E-mail address: robert.mnatsakanov@mail.wvu.edu (R.M. Mnatsakanov).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.09.003

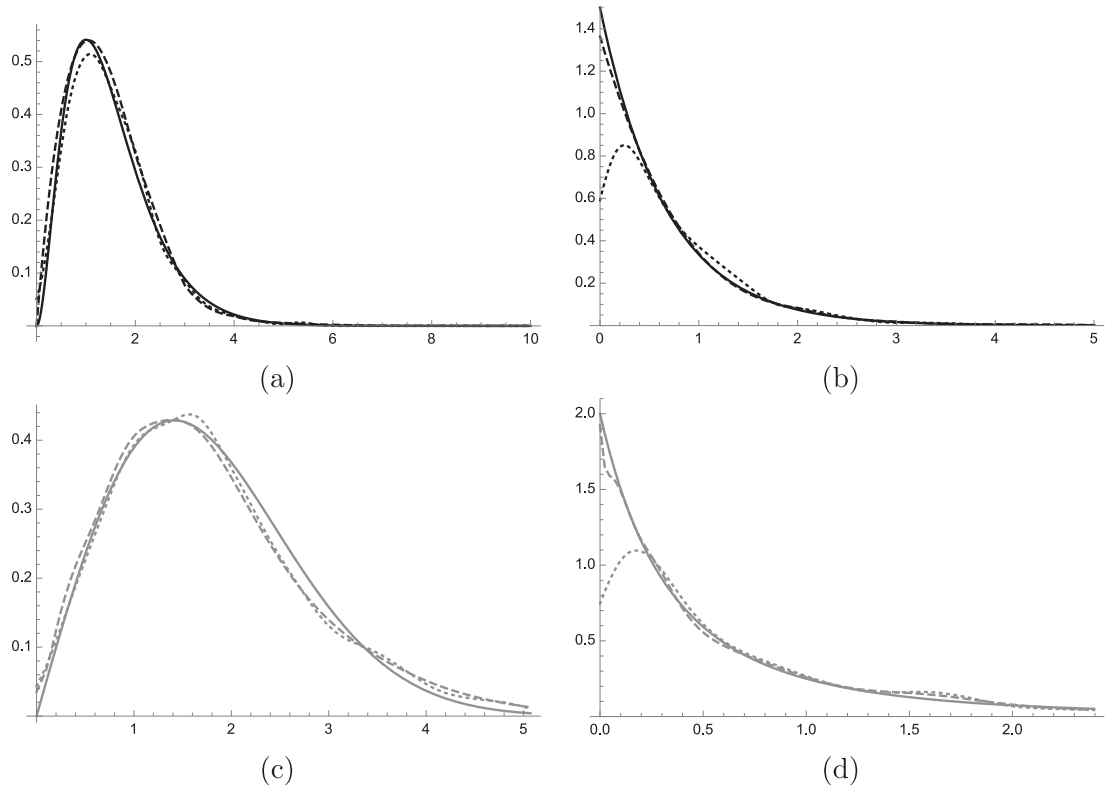


Fig. 1. Estimates $\hat{f}_{\alpha,b}$ (dashed curve) and $\hat{f}_h$ (dotted curve) (a): when $X_i \sim$ Gamma$(3, 1/2)$ with $\alpha = 120, b = 1.9, n = 500$; (b): when $X_i \sim$ Exp$(3/2)$ with $\alpha = 120, b = 1.15, n = 500$; (c): when $X_i \sim$ Weibull$(2, 2)$ with $\alpha = 100, b = 1.23, n = 800$; (d): when $X_i \sim$ Pareto$(1, 2)$ with $\alpha = 100, b = 1.55, n = 800$.

To reduce the boundary effect for nonnegative data, Chen [6] proposed the estimator based on asymmetric gamma kernels. Jones [11] and Chen [15] showed that the local linear estimator achieves better results than the boundary kernel estimator of Müller [13]. Bouezmarni and Rolin [5] derived the exact asymptotic constants for uniform and L_1-errors of the kernel density estimator defined by asymmetric beta kernels.

Bouezmarni et al. [16] studied another estimator based on the gamma kernels. It was shown that such construction is free of boundary effect and achieves the optimal rate of convergence in terms of integrated mean squared error. Similar properties of the so-called varying kernel density estimates have been established in [7] and [8] as well.

The main goal of the current work is to introduce totally different approach for estimating the density function f of a positive random variable via the values of its scaled Laplace transform. It is based on the empirical counterpart of approximate $f_{\alpha,b}$ suggested in Mnatsakanov et al. [1] (see also (2) in Section 2). The proposed approach provides a unified estimation method that could be applied for estimating the density functions in several indirect models as well, including the right-censored one. The properties of corresponding estimates will be investigated in the forthcoming paper. Simulation study justifies that suggested estimate does not have the edge effect in contrary to those based on symmetric KDE. In particular, when $f(0) > 0$, we recommend the use the estimate defined in (4) instead of traditional KDE (see, for example, plots (b) and (d) in Fig. 1). Besides, it is shown that proposed estimate (in the case of direct model) is reduced to the one based on asymmetric beta kernel density construction.

The rest of the paper is organized as follows. In Section 2, we describe the construction of new nonparametric estimate $\hat{f}_{\alpha,b}$ defined in (4). In Section 3, the finite sample properties of $\hat{f}_{\alpha,b}$ are investigated. In particular, the bias and the Mean Squared Error (MSE) of $\hat{f}_{\alpha,b}$ are derived. Section 4 is devoted to simulation study. Here we evaluated the average L_2-errors of proposed estimate and compared it to the one based on KDE. Several graphs of the estimate are displayed as well. The advantages of the proposed estimates are discussed in Section 5.

2. Some preliminaries and notations

Assume F is an absolutely continuous distribution supported by $\mathbf{R}_+ = [0, \infty)$. Let f be its probability density function (pdf) with respect to the Lebesgue measure on $\mathbf{R}_+$. In this section we introduce the estimate based on approximation of the Laplace transform inversion that recovers f. Denote by X a random variable distributed according to F.

Assume that we are given the sequence $\mathscr{L}(F) = \{\mathscr{L}_t(F), t \in \mathbb{N}_\alpha\}$ of the values of the scaled Laplace transform of F:

$$\mathscr{L}_t(F) = \int_0^\infty e^{-ctx}\, dF(x), \qquad t \in \mathbb{N}_\alpha = \{0, 1, 2, \ldots, \alpha\}. \tag{1}$$

To simplify the notations, assume in (1) that the scaling parameter $c = \ln b$, for some $1 < b \le \exp(1)$. Besides, denote the pdf of a Beta (p, q) distribution by

$$\beta(u, p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} u^{p-1}(1-u)^{q-1}, \quad 0 < u < 1.$$

Here the shape parameters $p, q > 0$, are defined as follows: $p = [\alpha b^{-x}] + 1$ and $q = \alpha - [\alpha b^{-x}] + 1$. Denote also $\beta^*_{\alpha,x}(\cdot) := \beta(\cdot, p, q)$. Consider the following approximation of f introduced in Mnatsakanov et al. [1]. Namely, for each $x \in \mathbf{R}_+$, define $f_{\alpha,b}(x) := (\mathscr{B}_\alpha^{-1}\mathscr{L}(F))(x)$, with

$$(\mathscr{B}_\alpha^{-1}\mathscr{L}(F))(x) = \frac{\ln b\,[\alpha b^{-x}]\,\Gamma(\alpha+2)}{\alpha\,\Gamma([\alpha b^{-x}]+1)} \sum_{m=0}^{\alpha-[\alpha b^{-x}]} \frac{(-1)^m \mathscr{L}_{m+[\alpha b^{-x}]}(F)}{\Gamma(m+1)\,\Gamma(\alpha-[\alpha b^{-x}]-m+1)}. \tag{2}$$

In a general setting, the proposed construction (2) can be applied when the $\sqrt{n}$-consistent estimate of the Laplace transform of F, say, $\tilde{\mathscr{L}}(F) = \{\tilde{\mathscr{L}}_t(F), t \in \mathbb{N}_\alpha\}$ is available, while F is not observed directly. Hence, using $\tilde{\mathscr{L}}(F)$ instead of $\mathscr{L}(F)$ in (2), we arrive at the estimate of f:

$$\tilde{f}_{\alpha,b}(x) = (\mathscr{B}_\alpha^{-1}\tilde{\mathscr{L}}(F))(x), \quad x \in \mathbb{R}_+. \tag{3}$$

In the current work, we apply (2) in direct model, i.e., when the sample of $i.i.d.$ random variables $X_1, \ldots, X_n$ from F is given. In this case, one can consider the empirical Laplace transform $\widehat{\mathscr{L}}(F)$ instead of $\tilde{\mathscr{L}}(F)$ in (3). As a result, the following estimate of f (after multiplying by $\alpha/(\alpha+1)$) is derived:

$$\hat{f}_{\alpha,b}(x) = \frac{\alpha}{\alpha+1}(\mathscr{B}_\alpha^{-1}\widehat{\mathscr{L}}(F))(x), \quad x \in \mathbb{R}_+. \tag{4}$$

Recall that

$$\widehat{\mathscr{L}}_t(F) = \int_{\mathbf{R}_+} e^{-ctx} d\widehat{F}_n(x), \tag{5}$$

and $\widehat{F}_n$ is the empirical cdf of the sample $X_1, \ldots, X_n$.

Remark 1. In order to simplify the evaluations of the bias and variance of estimate $\hat{f}_{\alpha,b}$, the normalizing factor in $\alpha/(\alpha+1)$ is used (4). Also note that

$$\widehat{\mathscr{L}}_t(F) = \int_{\mathbf{R}_+} b^{-tx} d\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n b^{-tX_i}. \tag{6}$$

Finally, note that given the observations $X_1, \ldots, X_n$ drawn from F, the estimate $\hat{f}_{\alpha,b}$ is reduced to:

$$\begin{aligned}\hat{f}_{\alpha,b}(x) &= \Big(\frac{\alpha}{\alpha+1}\Big)\frac{[\alpha b^{-x}]\ln(b)\Gamma(\alpha+2)}{\alpha\,\Gamma([\alpha b^{-x}]+1)}\frac{1}{n}\sum_{i=1}^n \sum_{m=0}^{\alpha-[\alpha b^{-x}]} \frac{(-1)^m b^{-(m+[\alpha b^{-x}])X_i}}{m!(\alpha-[\alpha b^{-x}]-m)!} \\ &= \frac{[\alpha b^{-x}]\ln(b)\Gamma(\alpha+2)}{(\alpha+1)\Gamma([\alpha b^{-x}]+1)}\frac{1}{n}\sum_{i=1}^n \sum_{m=0}^{\alpha-[\alpha b^{-x}]} \frac{(-b^{-X_i})^m (b^{-X_i})^{[\alpha b^{-x}]}}{m!(\alpha-[\alpha b^{-x}]-m)!}\end{aligned}$$

$$= \frac{[\alpha b^{-x}]\ln b}{\alpha+1}\,\frac{1}{n}\sum_{i=1}^{n}\beta(b^{-X_i},[\alpha b^{-x}]+1,\alpha-[\alpha b^{-x}]+1)$$

$$= \frac{[\alpha b^{-x}]\ln b}{\alpha+1}\,\frac{1}{n}\sum_{i=1}^{n}\beta^{*}_{\alpha,x}(b^{-X_i}),\quad x\in\mathbb{R}_+, \tag{7}$$

with $\beta^{*}_{\alpha,x}(\cdot)$ defined in Section 2.

It is worth mentioning that construction (3) has a unified form, and can be applied as in direct as well as in indirect models. As a special case, it reduces to the asymmetric beta density kernel estimator (see last line in (7)) when underlying density f is observed directly via the data-sets $\{X_i\}_{i=1}^{n}$ (cf. with [5,16,15]).

In Section 3 it will be assumed that the following conditions are satisfied:

$$f\in C^{(2)}(\mathbf{R}_+),\quad \sup_{x\in[0,\infty)}|b^{kx}f'(x)|<\infty,\quad k=1,2,\quad\text{and}\quad \sup_{x\in[0,\infty)}|b^{x}f''(x)|<\infty. \tag{8}$$

3. Mean squared errors of $\hat{f}_{\alpha,b}$

In this section we investigate the asymptotic behavior of the estimate $\hat{f}_{\alpha,b}(x)$ defined in (4). In particular, to derive the upper bound for MSE of $\hat{f}_{\alpha,b}(x)$. Recall that

$$\mathrm{MSE}\{\hat{f}_{\alpha,b}(x)\}:=E|\hat{f}_{\alpha,b}(x)-f(x)|^2=var\{\hat{f}_{\alpha,b}(x)\}+\big(\mathrm{Bias}\{\hat{f}_{\alpha,b}(x)\}\big)^2,$$

where $\mathrm{Bias}\{\hat{f}_{\alpha,b}(x)\}=f_{\alpha,b}(x)-f(x)$ and $f_{\alpha,b}(x):=E(\hat{f}_{\alpha,b}(x))$, for each $x\in\mathbb{R}_+$.

Theorem 1. *If conditions* (8) *are satisfied, then for MSE of* $\hat{f}_{\alpha,b}(x)$ *we have:*

$$\mathrm{MSE}\{\hat{f}_{\alpha,b}(x)\}\le n^{-4/5}\left[\left(\frac{2\,b^x|f'(x)|}{\ln b}+\frac{b^{2x}|f'(x)|}{2\ln^2 b}+\frac{b^x|f''(x)|}{2\,\ln b}\right)^2+\frac{f(x)\ln b}{\sqrt{\pi\,(b^x-1)}}\right]+o(n^{-4/5}) \tag{9}$$

provided that we choose $\alpha=\alpha(n)\sim n^{2/5}$, $n\to\infty$.

The proof of Theorem 1 is based on investigating the asymptotic behavior of the bias and variance terms of $\mathrm{MSE}\{\hat{f}_{\alpha,b}(x)\}$.

Lemma 2. *(i):* $f_{\alpha,b}(x)$ *converges uniformly to* $f(x)$ *as* $\alpha\to\infty$, *and for each* $x>0$, *the absolute value of the bias term of* $\hat{f}_{\alpha,b}(x)$ *is estimated from above as follows:*

$$|\mathrm{Bias}\{\hat{f}_{\alpha,b}(x)\}|\le\frac{1}{\alpha+1}\left\{\frac{2\,b^x|f'(x)|}{\ln b}+\frac{b^{2x}|f'(x)|}{2\ln^2 b}+\frac{b^x|f''(x)|}{2\,\ln b}\right\}+o\Big(\frac{1}{\alpha}\Big),$$

as $\alpha\to\infty$.

(ii): For each $x>0$, *the asymptotic expression for variance of* $\hat{f}_{\alpha,b}(x)$ *we have*

$$var\{\hat{f}_{\alpha,b}(x)\}=\frac{\sqrt{\alpha}}{n}\,\frac{f(x)\,\ln b}{\sqrt{\pi(b^x-1)}}+o\Big(\frac{\sqrt{\alpha}}{n}\Big),$$

as $\sqrt{\alpha}/n\to 0$, $\alpha,n\to\infty$.

Proof of Lemma 2. Because $\hat{f}_{\alpha,b}=(\alpha/(\alpha+1))\mathscr{B}_{\alpha}^{-1}\widehat{\mathscr{L}}(F)$ has representation (7), we have

$$E\hat{f}_{\alpha,b}(x)=\frac{[\alpha b^{-x}]\ln b}{\alpha+1}\,\frac{1}{n}\sum_{i=1}^{n}E\beta(b^{-X_i},[\alpha b^{-x}]+1,\alpha-[\alpha b^{-x}]+1)$$

$$=\frac{\ln b\;\Gamma(\alpha+1)}{\Gamma([\alpha b^{-x}])\Gamma(\alpha+[\alpha b^{-x}]+1)}\int_0^{\infty}(b^{-u})^{[\alpha b^{-x}]}(1-b^{-u})^{\alpha-[\alpha b^{-x}]}f(u)du. \tag{10}$$

The change of variable under integral in (10) with $\tau = b^{-u}$ gives

$$\begin{aligned} E\,\hat{f}_{\alpha,b}(x) &= \frac{\ln b\,\Gamma(\alpha+1)}{\Gamma([\alpha b^{-x}])\Gamma(\alpha+[\alpha b^{-x}]+1)}\int_0^1 \tau^{[\alpha b^{-x}]}(1-\tau)^{\alpha-[\alpha b^{-x}]}\frac{f(-\log_b\tau)}{\tau\ln b}d\tau \\ &= \frac{\Gamma(\alpha+1)}{\Gamma([\alpha b^{-x}])\Gamma(\alpha+[\alpha b^{-x}]+1)}\int_0^1 \tau^{[\alpha b^{-x}]-1}(1-\tau)^{\alpha-[\alpha b^{-x}]}f(-\log_b\tau)d\tau \\ &= \int_0^1 \beta(\tau,[\alpha b^{-x}],\alpha-[\alpha b^{-x}]+1)q(\tau)d\tau, \end{aligned} \tag{11}$$

where $q = f\circ\phi$ denotes the composition of functions f and $\phi(\tau) = -\log_b(\tau)$. Therefore for the expected value of $\hat{f}_{\alpha,b}$ we have

$$f_{\alpha,b}(x) = E\,\hat{f}_{\alpha,b}(x) = \int_0^1 \beta_{\alpha,x}(\tau)q(\tau)d\tau, \quad q(\tau) = f(-\log_b\tau). \tag{12}$$

Here, $\beta_{\alpha,x}(\cdot) := \beta(\cdot,[\alpha b^{-x}],\alpha-[\alpha b^{-x}]+1)$. To complete the proof of Lemma 2(i) one can proceed in a similar way as it is done in Mnatsakanov et al. [1].

Namely, let us mention that the first and second derivatives of q with respect to τ can be written as

$$q'(\tau) = (f'\circ\phi)(\tau)\phi'(\tau) \tag{13}$$

$$q''(\tau) = (f''\circ\phi)(\tau)\phi''(\tau) + (f'\circ\phi)(\tau)\phi''(\tau). \tag{14}$$

Evaluation of q' and q'' at $\tau = b^{-x}$ gives: $q'(b^{-x}) = \frac{b^x f'(x)}{\ln b}$ and $q''(b^{-x}) = \frac{b^{2x}f'(x)}{\ln^2 b} + \frac{b^x f''(x)}{\ln b}$. Now, note that the sequence $\{\beta_{\alpha,x}(\cdot), \alpha = 1,2,\dots\}$ represents the sequence of δ-functions with the mean and variance specified as follows:

$$\eta_\alpha := \int_0^1 \tau\beta_{\alpha,x}(\tau)d\tau = \frac{[\alpha b^{-x}]}{\alpha+1} \tag{15}$$

$$\sigma_\alpha^2 := \int_0^1 (\tau-\eta_\alpha)^2\beta_{\alpha,x}(\tau)d\tau = \frac{[\alpha b^{-x}](\alpha-[\alpha b^{-x}]+1)}{(\alpha+1)^2(\alpha+2)} < \frac{1}{\alpha+1}, \tag{16}$$

and

$$|\eta_\alpha - b^{-x}| \le \frac{2}{\alpha+1}. \tag{17}$$

To derive the asymptotic form of the Bias$\{\hat{f}_{\alpha,b}(x)\} = f_{\alpha,b}(x) - f(x)$ let us write

$$f_{\alpha,b}(x) - f(x) = \int_0^1 \beta_{\alpha,x}(\tau)\{q(\tau) - q(b^{-x})\}d\tau. \tag{18}$$

Applying the Taylor expansion of $q(\tau)$ around $\tau = b^{-x}$ we get:

$$\text{Bias}\{\hat{f}_{\alpha,b}(x)\} = \int_0^1 \beta_{\alpha,x}(\tau)\left\{(\tau-b^{-x})q'(b^{-x}) + \frac{1}{2}(\tau-b^{-x})^2 q''(\tilde{\tau})\right\}d\tau. \tag{19}$$

Now adding and subtracting η_α from the first and second terms in the right hand side of (19) we obtain the upper bound for the absolute value of Bias$\{\hat{f}_{\alpha,b}(x)\}$:

$$\begin{aligned} &\left|\int_0^1 \beta_{\alpha,x}(\tau)(\tau-\eta_\alpha)q'(b^{-x})d\tau + (\eta_\alpha - b^{-x})q'(b^{-x})\int_0^1 \beta_{\alpha,x}(\tau)d\tau\right. \\ &\quad\left. + \frac{1}{2}\int_0^1 \beta_{\alpha,x}(\tau)(\tau-\eta_\alpha)^2 q''(\tilde{\tau})d\tau + \frac{1}{2}(\eta_\alpha - b^{-x})^2\int_0^1 \beta_{\alpha,x}(\tau)q''(\tilde{\tau})d\tau\right| \\ &\le |\eta_\alpha - b^{-x}||q'(b^{-x})| + \frac{1}{2}\sigma_\alpha^2\,\|q''\| + \frac{1}{2}(\eta_\alpha - b^{-x})^2\,\|q''\|. \end{aligned} \tag{20}$$

Now taking into account the bounds for σ_α^2 and $|\eta_\alpha - b^{-x}|$ mentioned in (16) and (17), respectively, we obtain for the bias of $\hat{f}_{\alpha,b}$:

$$|\text{Bias}\{\hat{f}_{\alpha,b}(x)\}| \le \frac{1}{\alpha+1}\left\{\frac{2\,b^x|f'(x)|}{\ln b} + \frac{b^{2x}|f'(x)|}{2\ln^2 b} + \frac{b^x|f''(x)|}{2\,\ln b}\right\} + o(\frac{1}{\alpha}) \tag{21}$$

as $\alpha\to\infty$. The uniform convergence of $f_{\alpha,b}$ to f follows from (21) and the conditions (8).

Now, consider the variance of $\hat{f}_{\alpha,b}$. Taking into account (7) and (12), we obtain:

$$var\{\hat{f}_{\alpha,b}(x)\} = \frac{1}{n}\Big\{\frac{[\alpha b^{-x}]\ln b}{\alpha+1}\Big\}^2 var\{\beta^*_{\alpha,x}(b^{-X_1})\}$$
$$= \frac{1}{n}\Big\{\frac{[\alpha b^{-x}]\ln b}{\alpha+1}\Big\}^2 E\beta^{*2}_{\alpha,x}(b^{-X_i}) - \frac{1}{n}\Big\{\frac{[\alpha b^{-x}]\ln b}{\alpha+1}\Big\}^2 (E\beta^*_{\alpha,x}(b^{-X}))^2. \tag{22}$$

At first let us investigate the asymptotic behavior of the first term in (22). We have

$$\frac{1}{n}\Big\{\frac{[\alpha b^{-x}]\ln b}{\alpha+1}\Big\}^2 E\beta^{*2}_{\alpha,x}(b^{-X_i}) = \frac{1}{n}\Big\{\frac{[\alpha b^{-x}]\ln b}{\alpha+1}\Big\}^2 \int_0^\infty \beta^{*2}_{\alpha,x}(b^{-u}) f(u)du$$
$$= \frac{1}{n}\Big(\frac{\ln b\, \Gamma(\alpha+1)}{\Gamma([\alpha b^{-x}])\Gamma(\alpha-[\alpha b^{-x}]+1)}\Big)^2 \int_0^\infty (b^{-u})^{2[\alpha b^{-x}]}(1-b^{-u})^{2\alpha-2[\alpha b^{-x}]} f(u)du$$
$$= \frac{1}{n}\Big(\frac{\ln b\, \Gamma(\alpha+1)}{\Gamma([\alpha b^{-x}])\Gamma(\alpha-[\alpha b^{-x}]+1)}\Big)^2 \int_0^1 \tau^{2[\alpha b^{-x}]-1}(1-\tau)^{2\alpha-2[\alpha b^{-x}]} q(\tau)\frac{d\tau}{\ln b}$$
$$= \xi_\alpha(x)\int_0^1 \beta(\tau, 2[\alpha b^{-x}], 2\alpha-2[\alpha b^{-x}]+1)q(\tau)d\tau. \tag{23}$$

Here

$$\xi_\alpha(x) = \frac{\ln b}{n}\left(\frac{\Gamma(\alpha+1)}{\Gamma([\alpha b^{-x}])\Gamma(\alpha-[\alpha b^{-x}]+1)}\right)^2 \frac{\Gamma(2[\alpha b^{-x}])\Gamma(2\alpha-2[\alpha b^{-x}]+1)}{\Gamma(2\alpha+1)}.$$

and according to Lemma 2(i):

$$\int_0^1 \beta(\tau, 2[\alpha b^{-x}], 2\alpha-2[\alpha b^{-x}]+1)q(\tau)d\tau \to f(x) \quad \text{as} \quad \alpha\to\infty,$$

uniformly with the rate of $1/\alpha$. The order of magnitude of $\xi_\alpha(x)$ in (23) is specified as follows: for each $x > 0$ we have

$$\xi_\alpha(x) \sim \frac{\ln b}{\sqrt{\pi}}\frac{\alpha^{1/2}}{n}\left(\frac{b^{-x}}{1-b^{-x}}\right)^{1/2}, \quad \alpha, n\to\infty. \tag{24}$$

Hence, for the first term in the right-hand side of (22) we can write:

$$\frac{1}{n}\Big\{\frac{[\alpha b^{-x}]\ln b}{\alpha+1}\Big\}^2 E\beta^{*2}_{\alpha,x}(b^{-X_1}) = \frac{\sqrt{\alpha}}{n}\frac{f(x)\ln b}{\sqrt{\pi(b^x-1)}} + o(\frac{\sqrt{\alpha}}{n}), \tag{25}$$

as $\alpha, n\to\infty$. Consider the second term of (22):

$$\frac{1}{n}\{\frac{[\alpha b^{-x}]\ln b}{\alpha+1}\}^2 (E\beta^*_{\alpha,x}(b^{-X}))^2 = \frac{1}{n}\{\frac{[\alpha b^{-x}]\ln b}{\alpha+1}\}^2 \left(\frac{\Gamma(\alpha+2)}{\Gamma([\alpha b^{-x}]+1)\,\Gamma(\alpha-[\alpha b^{-x}]+1)}\right)^2$$
$$\times \left(\int_0^\infty (b^{-u})^{[\alpha b^{-x}]}(1-b^{-u})^{\alpha-[\alpha b^{-x}]} f(u)du\right)^2$$
$$= \frac{1}{n}\left(\frac{\ln b\, \Gamma(\alpha+1)}{\Gamma([\alpha b^{-x}])\,\Gamma(\alpha-[\alpha b^{-x}]+1)}\right)^2 \left(\int_0^1 \tau^{[\alpha b^{-x}]-1}(1-\tau)^{\alpha-[\alpha b^{-x}]} q(\tau)\frac{d\tau}{\ln(b)}\right)^2$$
$$= \frac{1}{n}\left(\int_0^1 \beta_{\alpha,x}(\tau)q(\tau)d\tau\right)^2 = \frac{1}{n}\Big(f(x)+O(\frac{1}{\alpha})\Big)^2, \quad \text{as} \quad \alpha\to\infty. \tag{26}$$

Finally, from (22), (25), and (26) we obtain

$$var\{\hat{f}_{\alpha,b}(x)\} = \frac{\sqrt{\alpha}}{n}\frac{f(x)\ln b}{\sqrt{\pi(b^x-1)}} + o(\frac{\sqrt{\alpha}}{n}), \tag{27}$$

as $\sqrt{\alpha}/n\to 0, \alpha, n\to\infty$. □

Table 1
Records of the average L_2-errors ($\hat{d}_n$), when $X \sim$ Gamma $(3, 1/2)$ and $R = 50$.

b	$n = 300$			$n = 500$			$n = 800$		
	α=80	α=100	α=120	α=80	α=100	α=120	α=80	α=100	α=120
1.05	0.154	0.1153	0.0969	0.1492	0.1213	0.0998	0.1507	0.1209	0.0948
1.15	0.0404	0.0294	0.0227	0.0397	0.0257	0.0179	0.0404	0.0269	0.0196
1.21	0.0235	0.0158	0.0125	0.0219	0.0154	0.0084	0.0244	0.0154	0.0119
1.23	0.0219	0.0140	0.0104	0.0217	0.0134	0.0090	0.0232	0.0147	0.0081
1.25	0.0167	0.0111	0.0091	0.0166	0.0105	0.0092	0.0177	0.0109	0.0127
1.26	0.0147	0.0104	0.0087	0.0175	0.0103	0.0120	0.0158	0.0111	0.0101
1.27	0.0156	00103	0.0082	0.0145	0.0103	0.0104	0.0168	0.0096	0.0119
1.28	0.0144	0.0099	0.0093	0.0144	0.0099	0.0102	0.015	0.0094	0.0117
1.29	0.0139	0.0101	0.0136	0.0159	0.0101	0.0122	0.0158	0.0269	0.0095
1.30	0.0130	0.0984	0.0107	0.0123	0.0098	0.0106	0.0137	0.0095	0.0099
1.50	0.0097	0.0063	0.0061	0.0089	0.0063	0.0064	0.0106	0.0075	0.0059
1.80	0.0079	0.0039	0.0055	0.0073	0.0039	0.0058	0.0068	0.0042	0.0025
1.90	0.0079	0.0049	0.0069	0.0057	0.0049	0.0021	0.0053	0.0037	0.0017
2.0	0.0058	0.0071	0.0052	0.0051	0.0071	0.0039	0.0051	0.0039	0.0030

Proof of Theorem 1. Combining the statements (i) and (ii) of Lemma 2, we obtain:

$$\mathrm{MSE}\{\hat{f}_{\alpha,b}(x)\} \le \frac{1}{(\alpha+1)^2}\left\{\frac{2\,b^x|f'(x)|}{\ln b} + \frac{b^{2x}|f'(x)|}{2\ln^2 b} + \frac{b^x|f''(x)|}{2\,\ln b}\right\}^2 + \frac{\sqrt{\alpha}}{n}\frac{f(x)\,\ln b}{\sqrt{\pi(b^x-1)}} + o(\frac{\sqrt{\alpha}}{n}). \tag{28}$$

Finally, taking $\alpha \sim n^{2/5}$ in (28), we arrive at (9). □

4. Simulation study

In this section we evaluated the average errors using L_2-norm of estimate (4) when the normalizing factor $\alpha/(\alpha+1)$ is removed. The following notation is used:

$$\hat{d}_n = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{m}\sum_{j=1}^{m}(\hat{f}(x_j) - f(x_j))^2\right)^{1/2}. \tag{29}$$

Here by R we denote the number of replications, and $\{x_j,\ j = 1, \ldots, m\}$ represents the partition of support of f. In this section, simulations from three different models (Gamma, Weibull, and shifted Pareto) with three different values of the sample size $n = 300$,500, and 800, combined with the number of replications $R = 50$ are conducted.

Now, let us simulate a random sample from $X \sim$ Gamma(a, β). In addition, the Average L_2-errors are computed for different values of α, b, and n, as it is shown in Table 1.

The records from this table specify the optimal values for parameters (α, b) as follows: (100, 1.8), (120, 2.0), and (120, 1.9) for $n = 300$,500, and 800, respectively. Corresponding average errors $\hat{d}_n$ are equal to 0.0039, 0.0021, and 0.0017, respectively. Hence, the Average L_2-error is decreasing as a function of sample size n.

Also, we compared $\hat{f}_{\alpha,b}(x)$ with the KDE $\hat{f}_h$ when $X \sim$ Gamma$(3, 1/2)$ and $X \sim$ Exp$(2/3)$. Assume the kernel density function K is specified as the Gaussian one and the bandwidth $h = n^{-1/5}\,1.06\,\hat{\sigma}$ (cf. with Silverman [2]). Here $\hat{\sigma}$ represents the standard deviation of the sample $X_1, \ldots, X_n$. In plots (a) and (b) of Fig. 1, the performances of $\hat{f}_{\alpha,b}$ and $\hat{f}_h$ are compared graphically. Two cases when $X_i \sim$ Gamma$(3, 1/2)$ and $X_i \sim$ Exp(rate $= 2/3$) are considered when $\alpha = 120$, $b = 1.9$, and $\alpha = 120$, $b = 1.15$, respectively. In both plots the sample size $n = 500$. In plots (c) and (d) of Fig. 1, the Weibull (2, 2) and shifted Pareto (1, 2), respectively, are considered when $\alpha = 100$ and $n = 800$. We can say that $\hat{f}_{\alpha,b}$ performs better if compared to KDE when $f(0) > 0$ (see plots (b) and (d) in Fig. 1).

Table 2 displays the Average L_2-errors of $\hat{f}_{\alpha,b}$ evaluated for several values of parameter α when $b = 1.23$ and of $\hat{f}_h$ when $X \sim$ Gamma (3, 1/2) and $n = 300$,500. Here the number of replications $R = 50$. From this table we conclude that performances of $\hat{f}_{\alpha,b}$ and $\hat{f}_h$ are similar to each other when $X \sim$ Gamma (3, 1/2).

Table 2
Comparison of Average L_2-errors of $\hat{f}_{\alpha,b}$, when $\alpha = 80, 100, 120$, and $b = 1.23$, and of KDE $\hat{f}_h$ when $h = n^{-1/5}\, 1.06\, \hat{\sigma}$. Here $X \sim$ Gamma (3, 1/2) and $R = 50$.

n	$\hat{f}_{80,b}$	$\hat{f}_{100,b}$	$\hat{f}_{120,b}$	$\hat{f}_h$
300	0.01004	0.0097	0.01071	0.01108
500	0.0084	0.0088	0.0083	0.0088
800	0.0083	0.0080	0.0077	0.0081

5. Conclusion

The paper deals with investigation of asymptotic properties of nonparametric density estimate $\hat{f}_{\alpha,b}$ defined in (4). The main advantages of the proposed construction are: (a) it can be used in models where the only available information about the underlying distribution F represents the finite values of the scaled Laplace transform of F; (b) the estimate (3) has a unified form and can be easily implemented in different incomplete models as soon as there exists a consistent estimate of the Laplace transform of f. In the forthcoming paper we are planning to study the properties of corresponding estimates in the models with right-censored and the length-biased observations. Note also that in the case of direct model, the proposed estimate $\hat{f}_{\alpha,b}$ is reduced to the one based on asymmetric beta kernel density construction (see (7)).

It is worth mentioning that the values of $\hat{f}_{\alpha,b}(x)$ became constant for $x > \ln\alpha/\ln b$. That is why it is recommended to choose the values of b closer to 1 (from the right), when one is dealing with distribution having a long tail.

We compared the finite sample performances of $\hat{f}_{\alpha,b}$ with its counterpart based on the kernel density construction $\hat{f}_h$. We found out that it behaves a little bit better in terms of average L_2-errors and is free from the edge effect.

References

[1] R.M. Mnatsakanov, K. Sarkisian, A. Hakobyan, Approximation of the ruin probability using the scaled Laplace transform inversion, Appl. Math. Comput. 268 (2015) 717–727.
[2] B. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, New York, 1986.
[3] J. Mielniczuk, Some asymptotic properties of kernel estimators of a density function in case of censored data, Ann. Statist. 14 (1986) 766–773.
[4] B. Zhang, Some asymptotic result for kernel density estimation under random censorship, Bernoulli 2 (1996) 183–198.
[5] T. Bouezmarni, J. Rolin, Consistency of beta kernel density function estimator, Canad. J. Statist. 31 (2003) 89–98.
[6] S.X. Chen, Probability density function estimation using gamma kernels, Ann. Inst. Statist. Math. 52 (2000) 471–480.
[7] R.M. Mnatsakanov, F.H. Ruymgaart, Moment-density estimation for positive random variables, Statistics 46 (2012) 215–230.
[8] R.M. Mantsakanov, K. Sarkisian, Varying kernel density estimation on R_+, Statist. Probab. Lett. 82 (2012) 1337–1345.
[9] M. Lejeune, P. Sarda, Smooth estimators of distribution and density functions, Comput. Statist. Data Anal. 14 (1992) 457–471.
[10] J.P. Nielsen, C. Tanggaard, M.C. Jones, Local linear density estimation for filtered survival data, with bias correction, Statistics 43 (2) (2009) 167–186.
[11] M.C. Jones, Simple boundary correction for kernel density estimation, Stat. Comput. 3 (1993) 135–146.
[12] M.C. Jones, P.J. Foster, A simple nonnegative boundary correction method for kernel density estimation, Statist. Sinica 6 (1996) 1005–1013.
[13] H.G. Müller, Smooth optimum kernel estimators near endpoints, Biometrika 78 (1991) 521–530.
[14] J.S. Marron, D. Ruppert, Transformation to reduce the boundary bias in kernel estimation, J. Roy. Statist. Soc. 56 (1994) 653–671.
[15] S.X. Chen, Beta kernel estimators for density functions, Comput. Statist. Data Anal. 31 (1999) 131–145.
[16] T. Bouezmarni, A. El Ghouch, M. Mesfioui, Gamma kernel estimators for density and hazard rate of right censored data, J. Probab. Stat. (2011).

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 448–452
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

On the integral relationship between the early exercise boundary and the value function of the American put option

Malkhaz Shashiashvili

Ivane Javakhishvili Tbilisi State University, Faculty of Exact and Natural Sciences, Department of Mathematics, 13 University St., Tbilisi 0186, Georgia

Received 17 May 2018; accepted 24 July 2018
Available online 14 August 2018

Abstract

We prove in this paper a new integral relationship between the American put option early exercise boundary and its value function in the generalized Black–Scholes model. Based on this relationship we show that it is possible to construct the L^2-approximation to the unknown early exercise boundary provided that we have at hand any uniform approximation of the American put option value function.

Keywords: American put option; Early exercise boundary; Energy inequality for the difference of two convex functions

1. Introduction

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $W = (W_t, \mathcal{F}_t)_{0\le t\le T}$ a standard Wiener process on it, where we assume that the time horizon T is finite. We will consider on $(\Omega, \mathcal{F}, W_t, \mathcal{F}_t, P)$, $0 \le t \le T$, the financial market with two assets (B_t, X_t), $0 \le t \le T$, where B_t denotes the value of the unit bank account at time t and X_t is the stock price at time t. They evolve according to the following ordinary and stochastic differential equations

$$dB_t = rB_t\,dt, \quad B_0 = 1, \quad 0 \le t \le T, \tag{1.1}$$

$$dX_u(x,t) = rX_u(x,t)\,du + \sigma(X_u(x,t),u)X_u(x,t)\,dW_u, \quad X_t(x,t) = x, \quad t \le u \le T, \tag{1.2}$$

where $r > 0$ is the bank interest rate and $\sigma(x,t)$, $x \ge 0$, $0 \le t \le T$, — called the local volatility function that satisfies the following conditions

$$0 < \underline{\sigma} \le \sigma(x,t) \le \overline{\sigma}, \tag{1.3}$$

E-mail address: malkhaz.shashiashvili@tsu.ge.
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.07.003

$\sigma(x,t)$ is a twice continuously differentiable function with respect to the argument x in $(0,\infty)\times[0,T]$ and

$$\big|x\sigma(x,t)-y\sigma(y,t)\big|\le c|x-y|. \tag{1.4}$$

The latter condition guarantees the existence and the uniqueness of the strong solution of the stochastic differential equation (1.2).

Consider the American put option on the stock with price X_t at time t and with a payoff function

$$g(x)=(K-x)^+,\quad x\ge 0. \tag{1.5}$$

Denote $V(x,t)$, $x\ge 0$, $0\le t\le T$, the American put option value function and define the continuation region and the exercise region of the American option in the following manner

$$\begin{aligned} C&=\big\{(x,t):\ x\ge 0,\ 0\le t\le T,\ V(x,t)>(K-x)^+\big\},\\ E&=\big\{(x,t):\ x\ge 0,\ 0\le t\le T,\ V(x,t)=(K-x)^+\big\}. \end{aligned} \tag{1.6}$$

The boundary between C and E forms the graph of the function $x=b(t)$, $0\le t\le T$, and is called the early exercise boundary of the American put option (here the early exercise means to exercise before the maturity time T). It is optimal to exercise the put option if the stock price at time t is below $b(t)$ and it is optimal not to exercise if the stock price at time t is above $b(t)$. For arbitrary t, $0\le t\le T$, $b(t)$ is defined in the following way

$$b(t)=\inf\big\{x\ge 0:\ V(x,t)>(K-x)^+\big\},\quad \text{where}\ \ b(t)\le K,\ \ b(T)=K. \tag{1.7}$$

It is well known (see, e.g. Achdou [1]) that the American put option value function $V(X,t)$, $x\ge 0$, $0\le t\le T$, is the unique strong solution (in the class of functions belonging to the parabolic Sobolev space $W^{2,1}_{2,loc}((0,\infty)\times(0,T))$) of the following parabolic obstacle problem

$$\begin{aligned} &LV(x,t)\le 0,\quad V(x,t)\ge(K-x)^+,\quad V(x,T)=(K-x)^+,\\ &LV(x,t)(V(x,t)-(K-x)^+)=0\ \ \text{a.e. } dx\times dt \text{ in } (0,\infty)\times(0,T), \end{aligned} \tag{1.8}$$

where LV is a parabolic partial differential operator of the form

$$LV(x,t)=\frac{\partial V(x,t)}{\partial t}+\frac{\sigma^2(x,t)}{2}x^2\frac{\partial^2 V(x,t)}{\partial x^2}+rx\frac{\partial V(x,t)}{\partial x}-rV(x,t),\quad 0<x<\infty,\quad 0<t<T, \tag{1.9}$$

with its adjoint operator $L^*h(x)$ for function $h(x)$ of the single argument x

$$\begin{aligned} L^*h(x)=&\frac{\sigma^2(x,t)}{2}x^2\frac{d^2h(x)}{dx^2}\\ &+\Big(-rx+\frac{d}{dx}\Big(\frac{\sigma^2(x,t)}{2}x^2\Big)\Big)\frac{dh(x)}{dx}+\Big(\frac{d^2}{dx^2}\Big(\frac{\sigma^2(x,t)}{2}x^2\Big)-2r\Big)h(x),\\ &0<x<\infty,\quad 0<t<T, \end{aligned} \tag{1.10}$$

Moreover, $V(x,t)$ satisfies the following partial differential equation

$$\begin{aligned} LV(x,t)=&-rKI(x\le b(t))\ \ \text{a.e. } dx\times dt \text{ in } (0,\infty)\times(0,T), \text{ with final condition}\\ &V(x,T)=(K-x)^+. \end{aligned} \tag{1.11}$$

We have from Achdou [1] that

$$\begin{aligned} &V(x,t)\in C([0,\infty)\times[0,T]),\quad -1\le\frac{\partial V(x,t)}{\partial x}\le 0,\\ &0<\gamma\le b(t)\le K,\quad 0\le t\le T,\\ &V(x,t)\in W^{2,1}_2((0,R)\times(s,t))\ \ \text{for arbitrary}\ \ R>0,\quad 0<s<t<T. \end{aligned} \tag{1.12}$$

Consider now continuously differentiable function $\rho(x)$, $-\infty<x<\infty$, such that

$$\rho(x)=\begin{cases}0 & \text{if } x\le 0 \text{ or } x\ge 1,\\ \text{is nonnegative} & \text{if } 0<x<1,\end{cases}\qquad \int_0^1\rho(x)\,dx=1, \tag{1.13}$$

and define the following twice continuously differentiable function $h(x)$, $-\infty<x<\infty$,

$$h(x)=\int_x^\infty\rho(y)\,dy,\quad -\infty<x<\infty. \tag{1.14}$$

Then it is clear that

$$h(x)=\begin{cases}1 & \text{if } x\le 0,\\ 0 & \text{if } x\ge 1,\\ \text{belongs to } [0,1] & \text{if } 0\le x\le 1.\end{cases} \tag{1.15}$$

2. The formulation and the proof of the main results

The principal objective of this paper is to establish the following proposition.

Theorem 2.1. *Let the local volatility function $\sigma(x,t)$ satisfy conditions* (1.3)*,* (1.4)*. Then the following integral relationship is valid between the early exercise boundary $b(t)$, $0\le t\le T$, and the value function $V(x,t)$ of the American put option*

$$rK\int_s^t b(u)\,du=-\int_s^t\left(\int_0^{K+1}V(x,u)\,L^*h(x-K)\,dx\right)du + \int_0^{K+1}(V(x,s)-V(x,t))\,h(x-K)\,dx,\quad 0\le s\le t\le T. \tag{2.1}$$

Proof. We recall the Green's classical identity for functions $V(x,u)\in C^{2,1}((0,\infty)\times(s,t))$ and $h(x)\in C^2(-\infty,+\infty)$

$$\begin{aligned}&h(x-K)\,LV(x,u)-V(x,u)\,L^*h(x-K)\\ &\quad=\frac{\partial}{\partial x}\Big[h(x-K)\,\frac{\sigma^2(x,u)}{2}\,x^2\,\frac{\partial V(x,u)}{\partial x}-V(x,u)\,\frac{\sigma^2(x,u)}{2}\,x^2\,\frac{\partial h(x-K)}{\partial x}\\ &\qquad-V(x,u)\,h(x-K)\,\frac{\partial}{\partial x}\Big(\frac{\sigma^2(x,u)}{2}\,x^2\Big)+rxV(x,u)h(x-K)\Big]+\frac{\partial}{\partial u}\,(V(x,u)h(x-K)).\end{aligned} \tag{2.2}$$

Integrate the Green's identity (2.2) over the set $(0,K+1)\times(s,t)$ taking into account that

$$h(1)=0,\quad h'(1)=0, \tag{2.3}$$

then we get

$$\begin{aligned}&\int_s^t\int_0^{K+1}\Big(h(x-K)\cdot LV(x,u)-V(x,u)\cdot L^*h(x-K)\Big)\,dx\,du\\ &\quad=\int_0^{K+1}\int_s^t\frac{\partial}{\partial u}\,(V(x,u)\cdot h(x-K))\,du\,dx=\int_0^{K+1}(V(x,t)-V(x,s))\cdot h(x-K)\,dx.\end{aligned} \tag{2.4}$$

Since the coefficients of the differential operator LV are bounded on the set $(0,K+1)\times(s,t)$, using standard approximation arguments of the functions $V(x,u)$ belonging to parabolic Sobolev space $W_2^{2,1}((0,K+1)\times(s,t))$ by the sequence of smooth functions $V_m(x,u)$, $V_m(x,u)\in C^{2,1}((0,K+1)\times(s,t))$ we see, that the equality (2.4) remains valid for functions $V(x,u)$ belonging to the space $W_2^{2,1}((0,K+1)\times(s,t))$ and hence for the value function $V(x,u)$ of the American put, that is we get

$$\begin{aligned}&\int_s^t\int_0^{K+1}LV(x,u)\cdot h(x-K)\,dx\,du\\ &\quad=\int_s^t\int_0^{K+1}V(x,u)\cdot L^*h(x-K)\,dx\,du+\int_0^{K+1}(V(x,t)-V(x,s))\cdot h(x-K)\,dx.\end{aligned} \tag{2.5}$$

Now, we have the following obvious identity

$$rKb(u)=\int_0^{K+1}rK\,I_{(x\le b(u))}\cdot h(x-K)\,dx,\quad 0\le u\le T. \tag{2.6}$$

Integrating the last identity over the time interval $[s,t]$, we get

$$rK\int_s^t b(u)\,du=\int_s^t\int_0^{K+1}rK\,I_{(x\le b(u))}\cdot h(x-K)\,dx\,du. \tag{2.7}$$

From here and the partial differential equation (1.11) we obtain

$$rK\int_s^t b(u)\,du = \int_s^t\int_0^{K+1}(-LV(x,u)\cdot h(x-K))\,dx\,du. \tag{2.8}$$

Comparing the equalities (2.5) and (2.8), ultimately we come to the desired relationship (2.1). □

Remark 2.1. Note that in both sides of the equality (2.1) we have the linear integral operators from functions $b(t)$ and $V(x,t)$, whereas by definition (1.7) $b(t)$ is obtained as a result of the nonlinear transform of $V(x,t)$.

From now on we will consider the case when the local volatility depends only on the argument x, that is

$$\sigma(x,t)=\sigma(x),\quad x\geq 0,\quad 0\leq t\leq T. \tag{2.9}$$

From Shashiashvili [2] we know that $b(t)$ is a nondecreasing continuous function on the time interval $[0,T]$.

Introduce the function

$$B(t)=\int_0^t b(u)\,du,\quad 0\leq t\leq T. \tag{2.10}$$

It is clear, that $B(t)$, $0\leq t\leq T$, is a convex function with continuous derivative $b(t)$, $0\leq t\leq T$.

We have from the relationship (2.1)

$$\begin{aligned}B(t) = &-\frac{1}{rK}\int_0^t\int_0^{K+1} V(x,u)\cdot L^*h(x-K)\,dx\,du\\ &+\frac{1}{rK}\int_0^{K+1}(V(x,0)-V(x,t))\cdot h(x-K)\,dx,\quad 0\leq t\leq T.\end{aligned} \tag{2.11}$$

Several authors, among them Bally, Saussereau [3], Hu, Liang and Jiang [4], constructed the uniform approximations to American put option value function $V(x,t)$, $x\geq 0$, $0\leq t\leq T$. Having at hand any uniform approximation $V_\varepsilon(x,t)$, $x\geq 0$, $0\leq t\leq T$, to the latter function, our next objective is to approximate (in some reasonable sense) the unknown early exercise boundary $b(t)$, $0\leq t\leq T$. Let us assume that for some small parameter $\varepsilon>0$ we have at hand the function $V_\varepsilon(x,t)$ such that

$$\sup_{\substack{0\leq x\leq K+1\\ 0\leq t\leq T}}|V_\varepsilon(x,t)-V(x,t)|\leq c_0\varepsilon, \tag{2.12}$$

where c_0 is a positive constant which depends on the parameters of our financial model r, $\underline{\sigma}$, $\overline{\sigma}$, K, T.

Denote

$$\begin{aligned}B_\varepsilon(t) = &-\frac{1}{rK}\int_0^t\int_0^{K+1} V_\varepsilon(x,u)\cdot L^*h(x-K)\,dx\,du\\ &+\frac{1}{rK}\int_0^{K+1}(V_\varepsilon(x,0)-V_\varepsilon(x,t))\cdot h(x-K)\,dx,\quad 0\leq t\leq T.\end{aligned} \tag{2.13}$$

Consider now the lower convex envelope $\breve{B}_\varepsilon(t)$, $0\leq t\leq T$, of the function $B_\varepsilon(t)$, $0\leq t\leq T$, which is the maximal convex function dominated by the function $B_\varepsilon(t)$, $0\leq t\leq T$, and then consider its left-hand derivative

$$b_\varepsilon(t)=\breve{B}'_\varepsilon(t-),\ 0<t\leq T. \tag{2.14}$$

We will prove that $b_\varepsilon(t)$ approximates $b(t)$ in the weighted L^2 norm.

Theorem 2.2. *Suppose that the local volatility $\sigma(x)$ satisfy conditions* (1.3), (1.4). *Then the following weighted L^2-estimate is valid for the unknown early exercise boundary $b(t)$, $0\leq t\leq T$, through the function $b_\varepsilon(t)$, $0\leq t\leq T$*

$$\int_0^T (b_\varepsilon(t)-b(t))^2\,t\,(T-t)\,dt\leq 5Tc(2KT+c)\varepsilon, \tag{2.15}$$

where

$$c=\frac{c_0}{rK}\left(T\int_0^{K+1}|L^*h(x-K)|\,dx+2(K+1)\right). \tag{2.16}$$

Proof. Easy to check that

$$\sup_{0\le t\le T} |B_\varepsilon(t) - B(t)| \le c\varepsilon. \tag{2.17}$$

From Lemma 3.1 in K. Shashiashvili and M. Shashiashvili [5] we get

$$\sup_{0\le t\le T} |\breve{B}_\varepsilon(t) - B(t)| \le c\varepsilon. \tag{2.18}$$

Consider the weight function

$$\varphi(t) = \frac{t(T-t)}{2}, \quad 0 \le t \le T,$$

and note

$$\varphi''(t) = -1, \quad 0 < t < T.$$

Let us apply now the energy inequality for the difference of two convex functions from the paper [5] by K. Shashiashvili and M. Shashiashvili (see therein Theorem 2.2 and the energy estimate (44)), then we shall get the desired result (2.15). □

References

[1] Y. Achdou, An inverse problem for a parabolic variational inequality arising in volatility calibration with American options, SIAM J. Control Optim. 43 (5) (2005) 1583–1615.
[2] M. Shashiashvili, Mathematical analysis of the early exercise boundary for the American put option, Rep. Enlarged Sess. Semin. I. Vekua Appl. Math. 31 (2017) 123–126.
[3] V. Bally, B. Saussereau, Approximation of the Snell envelope and American options prices in dimension one, ESAIM Probab. Stat. 6 (2002) 1–19.
[4] B. Hu, J. Liang, L. Jiang, Optimal convergence rate of the explicit finite difference scheme for American option valuation, J. Comput. Appl. Math. 230 (2) (2009) 583–599.
[5] K. Shashiashvili, M. Shashiashvili, From the uniform approximation of a solution of the PDE to the L^2-approximation of the gradient of the solution, J. Convex Anal. 21 (1) (2014) 237–252.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 453–465
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

The method of probabilistic solution for 3D Dirichlet ordinary and generalized harmonic problems in finite domains bounded with one surface

Mamuli Zakradze*, Badri Mamporia, Murman Kublashvili, Nana Koblishvili

Georgian Technical University, N. Muskhelishvili Institute of Computational Mathematics, 4 Grigol Peradze st., Tbilisi 0131, Georgia

Received 1 June 2018; received in revised form 7 August 2018; accepted 18 August 2018
Available online 24 September 2018

Abstract

The Dirichlet ordinary and generalized harmonic problems for some 3D finite domains are considered. The term "generalized" indicates that a boundary function has a finite number of first kind discontinuity curves. An algorithm of numerical solution by the method of probabilistic solution (MPS) is given, which in its turn is based on a computer simulation of the Wiener process. Since, in the case of 3D generalized problems there are none exact test problems, therefore, for such problems, the way of testing of our method is suggested. For examining and to illustrate the effectiveness and simplicity of the proposed method five numerical examples are considered on finding the electric field. In the role of domains are taken ellipsoidal, spherical and cylindrical domains and both the potential and strength of the field are calculated. Numerical results are presented.

Keywords: Dirichlet ordinary and generalized problems; Harmonic function; Discontinuity curve; Probabilistic solution; Wiener process

1. Introduction

Let D be a finite domain in the Euclidian space R^3, bounded by one closed piecewise smooth surface $S(i.e., S = \bigcup_{j=1}^{p} S^j)$, where each part S^j is a smooth surface. Besides, we assume: (1) equations of the parts S^j are given; (2) for the surface S it is easy to show that a point $x = (x_1, x_2, x_3) \in R^3$ lies in $\overline{D}$ or not.

It is known (see, e.g., [1–8]) that in practical stationary problems (for example, for the determination of the temperature of the thermal field or the potential of the electric field, and so on) there are cases when it is necessary to consider the Dirichlet ordinary (or generalized) harmonic problems: A (or B).

* Corresponding author.
E-mail addresses: mamuliz@yahoo.com (M. Zakradze), badrimamporia@yahoo.com (B. Mamporia), mkublashvili@mail.ru (M. Kublashvili), nanakoblishvili@yahoo.com (N. Koblishvili).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.08.005

Problem A. Find a function $u(x) \equiv u(x_1, x_2, x_3) \in C^2(D) \bigcap C(\overline{D})$ satisfying the conditions:

$$\Delta u(x) = 0, \quad x \in D,$$
$$u(y) = g(y), \quad y \in S,$$

where $\Delta = \sum_{i=1}^{3} \frac{\partial^2}{\partial x_i^2}$ is the Laplace operator and $g(y) \equiv g(y_1, y_2, y_3)$ is a continuous function on S.

It is known (see, e.g., [2–4]) that Problem A is correct, i.e., its solution exists, is unique and depends on data continuously. It should be noted that in general the difficulties and respectively the laboriousness of solving problems sharply increase along with the dimension of the problems considered. Therefore, as a rule, one fails to develop standard methods for solving a wide class of multidimensional problems with the same high accuracy as in the one-dimensional case. For example, the exact solution of Problem A for a disk is written by one-dimensional Poisson's integral and in the case of kernel by two-dimensional Poisson's integral

$$u(x) = \frac{1}{4\pi r} \int_{|y|} \int_{=r} \frac{r^2 - |x|^2}{|x - y|^3} g(y) dS_y, \quad |x| < r, \tag{1.1}$$

where r is the radius of the kernel with the center at the origin O ($r = |y| = |Oy|, \quad y \in S$). Integral (1.1) loses sense, when $|x| = r$. However, it is proved that $u(x) \to g(y)$ when $x \to y$ ($x \in D$). A simple example given by us, shows the difficulty in determining of the solution with the high accuracy of the Dirichlet ordinary harmonic problem when the dimension increases.

Problem B. Function $g(y)$ is given on the boundary S of the domain D and is continuous everywhere, except a finite number of curves $l_1, l_2, \ldots, l_n$ which represent discontinuity curves of the first kind for the function $g(y)$. It is required to find a function $u(x) \equiv u(x_1, x_2, x_3) \in C^2(D) \bigcap C(\overline{D} \setminus \bigcup_{k=1}^{n} l_k)$ satisfying the conditions:

$$\Delta u(x) = 0, \qquad x \in D, \tag{1.2}$$
$$u(y) = g(y), \quad y \in S, \quad y \overline{\in} l_k \ \ (k = \overline{1, n}), \tag{1.3}$$
$$|u(y)| < c, \qquad y \in \overline{D}, \tag{1.4}$$

where c is a real constant.

It is shown (see [9,10]) that Problem (1.2),(1.3),(1.4) has a unique solution depending continuously on the data, and for a generalized solution $u(x)$ the generalized extremum principal is valid:

$$\min_{x \in S} u(x) < \begin{matrix} u(x) \\ x \in D \end{matrix} < \max_{x \in S} u(x), \tag{1.5}$$

where for $x \in S$ it is assumed that $x \overline{\in} l_k (k = \overline{1, n})$.

It is evident that actually, the surface S is divided into parts $S_i (i = \overline{1, m})$ by curves $l_k (k = \overline{1, n})$ or $S = \bigcup_{i=1}^{m} S_i$. On the basis of noted, the boundary function $g(y)$ has the following form

$$g(y) = \begin{cases} g_1(y), & y \in S_1, \\ g_2(y), & y \in S_2, \\ \ldots\ldots\ldots\ldots\ldots \\ g_m(y), & y \in S_m, \end{cases} \tag{1.6}$$

where the functions $g_i(y)$, $y \in S_i$ are continuous on the parts S_i of S, respectively.

Note that the additional requirement (1.4) of boundedness concerns actually only the neighborhoods of discontinuity curves of the function $g(y)$ and it plays an important role in the extremum principle (1.5).

On the basis of (1.4), in general, the values of $u(y)$ are not defined on the curves l_k. For example, if Problem B concerns the investigation of the thermal (or the electric) field, then $u(y) = 0$ when $y \in l_k$, respectively, in this case, in physical sense the curves l_k are non-conductors (or dielectrics).

Remark 1. If inside the surface S there is a vacuum then we have the ordinary and generalized problems with respect to closed shells.

In general, it is known (see [5,10,11]) that the methods used to obtain an approximate solution to ordinary boundary problems are less suitable (or not suitable at all) for solving boundary problems of type B. In particular, the convergence is very slow in the neighborhood of boundary singularities and, consequently, the accuracy of the approximate solution of the generalized problem is very low.

The choice and construction of computational schemes (algorithms) mainly depend on problem class, its dimension, geometry and location of singularities on the boundary. e.g., plane Dirichlet generalized problems for harmonic functions with concrete location of discontinuity points in the cases of simply connected domains are considered in [5–8,12], and general cases for finite and infinite domains are studied in [1,13–17].

In the case of spatial (3D) harmonic generalized problems, due to their higher dimension, the difficulties become more significant. In particular, there does not exist a standard scheme which can be applied to a wide class of domains. In the classical literature, simplified, or so called "solvable" generalized problems (problems whose "exact" solutions can be constructed by series, whose terms are represented by special functions) are considered, and for their solution the classical method of separation of variables is mainly applied and therefore the accuracy of the solution is rather low. In the mentioned problems, the boundary functions (conditions) are mainly constants, and in the general case, the analytic form of the "exact" solution is so difficult in the sense of numerical implementation, that it only has theoretical significance (see e.g., [4–8]).

As a consequence of the above, from our viewpoint, the construction of high accuracy and effectively realizable computational schemes for approximate solution of 3D Dirichlet generalized harmonic problems (whose application is possible to a wide class of domains) have both theoretical and practical importance.

It should be noted that in classical literature (see e.g.,[4–8]), while solving Dirichlet generalized harmonic problems, the existence of discontinuity curves is neglected. This fact and application of classical methods to solving problems of type B are reasons of the inaccuracies. Therefore, for numerical solution of generalized harmonic problems we should apply such methods which do not require approximation of a boundary function and in which the existence of discontinuity curves is not ignored. The algorithm suggested by us is one of such methods.

2. The method of probabilistic solution

Let (Ω, F, P) be a probability space. The family of random values $(w(t), t \geq 0)$ is called the standard real valued Wiener process, if:

1. $w(0) = 0$ almost surely (a.s.); 2. for any $0 = t_0 < t_1 < \cdots < t_n$, the random variables $w(t_{i+1}) - w(t_i), i = \overline{0, n-1}$, are independents; 3. for all s, t, the random variable $w(t+s) - w(t)$ is normally distributed (Gaussian) random variable with mean 0 and variance s; 4. for almost all $\omega \in \Omega$, the function $w(t) = w(t, \omega)$ is continuous. (It is known, that the condition 4 follows from the conditions 1–3).

Let $x \in R$, denote $w(x, t) = x + w(t)$—the Wiener process starting at the point x, $w(t) = w(0, t)$. If we have independent standard Wiener processes $w_1(t)$, $w_2(t)$, $w_3(t)$, we can consider the standard Wiener process $W(t) = (w_1(t), w_2(t), w_3(t))$ in R^3 and, correspondingly, for any $x = (x_1, x_2, x_3) \in R^3$ consider the Wiener process $x(t, \omega) \equiv W(x, t, \omega) = x + W(t, \omega)$ starting at the point x.

Denote by X the space of continuous functions $x(t), t \geq 0$ with values in R^3. These functions are interpreting as trajectories of the Wiener process. Let on X is given the family of probabilistic distributions (measures) P_x. The measure P_x is the distribution of the random trajectory of the Wiener process starting at the point x. The mathematical expectation by the measure P_x is denoted by E_x. If $\Gamma \in R^3$ is a measurable set, then

$$P_x\{x(t) \in \Gamma\} = \frac{1}{(2\pi t)^{3/2}} \int_\Gamma e^{-\frac{(y_1-x_1)^2+(y_2-x_2)^2+(y_3-x_3)^2}{2t}} dy_1 dy_2 dy_3 \equiv P(t, x, \Gamma),$$

where $P(t, x, \Gamma)$—the transition function of the Markov process, that is, this is a probability of the event that the Wiener trajectory starting at the point x, after the time interval t arrives at Γ.

The main theorem in realization of the MPS is the following one (see e.g., [9])

Theorem 1. *If a finite domain $D \in R^3$ is bounded by piecewise smooth surface S and $g(y)$ is continuous (or discontinuous) bounded function on S, then the solution of the Dirichlet ordinary (or generalized) boundary problem for the Laplace equation at the fixed point $x \in D$ has the form*

$$u(x) = Eg(W(x, \tau(\omega), \omega)) = E_x g(x(\tau)), \tag{2.1}$$

where $\tau(\omega)$ is the first exit moment of the trajectory $W(x, \tau(\omega), \omega)$ from the domain D.

For any fixed point $x \in D$ we generate the realization of the Wiener process $x(t) = (x_1(t), x_2(t), x_3(t))$ starting at the point x. After, we check the first exit moment of the trajectory $W(x, \tau(\omega), \omega)$ from the domain D and fix the real value $g(W(x, \tau(\omega), \omega)) \equiv g(y^1)$. We repeat this procedure and receive $g(y^i)$, $i = \overline{1, N}$. If the number N is sufficiently large, then by the law of large numbers, from (2.1) we have

$$u(x) = Eg(W(x, \tau(\omega), \omega)) \approx u_N(x) = \frac{1}{N} \sum_{i=1}^{N} g(y^i), \tag{2.2}$$

where $y^i = (y_1^i, y_2^i, y_3^i)$ are the intersection points of the Wiener process and the surface S and $u(x) = \lim u_N(x)$ a. s. for $N \to \infty$. Thus, in the presence of the Wiener process the approximate values of the solutions to Problems A and B at a point $x \in D$ are calculated by formula (2.2).

On the basis of Theorem 1, the existence of solutions of the Dirichlet ordinary and generalized problems in the case of Laplace's equation for a sufficiently wide class of domains is shown. Besides, we have also an explicit formula giving such solutions.

Consider now the general case . That is the elliptic differential operator

$$L = \sum_{i,j=1}^{3} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{3} b_i(x) \frac{\partial}{\partial x_i},$$

where the matrix $A(x) = (a_{ij}(x))$ is symmetric and positive defined (i.e. all principal minors of $det(a_{ij})$ are positive at any point of the domain D) and $a_{ij}(x)$ and $b_i(x)$ are sufficiently smooth functions.

Consider the Dirichlet problem for equation $Lu(x) = 0$, in a finite domain $D \in R^3$, bounded by sufficiently smooth surface S, and boundary function $g(y)$ is continuous (or discontinuous) bounded function on S. In this case the solution of the Dirichlet problem for above mentioned equation at the fixed point $x \in D$ has the form (see e.g.,[18])

$$u(x) = Eg(\xi(x, \tau(\omega), \omega)) = E_x g(x(\tau)),$$

where ξ is the solution of the Ito's stochastic differential equation

$$d\xi_t = B(\xi_t)dt + C(\xi_t)dW_t, \tag{2.3}$$

with initial condition $\xi_0 = x$.

In (2.3) the matrix $C(x) = (c_{ij}(x))$ $(i, j = \overline{1, 3})$ is such that $A = C \times C^T$, $B(x) = (b_1(x), b_2(x), b_3(x))$ and $W_t = (w_1(t), w_2(t), w_3(t))$ is standard Wiener process in R^3. The integral form of (2.3) is

$$\xi_t = x + \int_0^t B(\xi_s)ds + \int_0^t C(\xi_s)dW_s,$$

where the last integral is the Ito's stochastic integral. Investigation and numerical solution of problems of types A and B for equation $Lu(x) = 0$ are problems of future development.

Since, in this article, we consider the problem of numerical solution of Dirichlet problems of type A and B for the Laplace equation, then it is clear that $b_i = 0$ $(i = 1, 2, 3)$, $a_{ij} = 1$ when $i = j$ and $a_{ij} = 0$ when $i \neq j$, respectively. In this case, in the role of the matrix C identity matrix is taken. $(c_{ij} = 1)$ when $i = j$ and $c_{ij} = 0$ when $i \neq j$.

According of definition of the Wiener process we have

$$\begin{aligned} x_1(t) &= x_1 + w_1(t) - w_1(0), \\ x_2(t) &= x_2 + w_2(t) - w_2(0), \\ x_3(t) &= x_3 + w_3(t) - w_3(0), \end{aligned} \tag{2.4}$$

where $w_1(0)$, $w_2(0)$, $w_3(0)$ equal to zero. In order, to computer simulate of the Wiener process based on (2.4), we use the following recursion relations:

$$\begin{aligned} &x_1(t_k) = x_1(t_{k-1}) + \gamma_1(t_k)/nq, \\ &x_2(t_k) = x_2(t_{k-1}) + \gamma_2(t_k)/nq, \\ &x_3(t_k) = x_3(t_{k-1}) + \gamma_3(t_k)/nq, \\ &(k = 1, 2, \dots), \quad x(t_0) = x, \end{aligned} \tag{2.5}$$

with the help of which the coordinates of the point $x(t_k) = (x_1(t_k), x_2(t_k), x_3(t_k))$ are being determined. In (2.5): $\gamma_1(t_k)$, $\gamma_2(t_k)$, $\gamma_3(t_k)$ are three normally distributed independent random numbers for the kth step, with zero means and variances one; nq is a number of quantification (nq) such that $1/nq = \sqrt{t_k - t_{k-1}}$ and when $nq \to \infty$, then the discrete process approaches to the continuous Wiener process. In the implementation, the random process is simulated at each step of the walk and continues until it crosses the boundary.

In our case computations and generation of random numbers are done in MATLAB.

3. Numerical examples

In this section, problems of type A and B considered by us are solved for one and the same domain. The reason of this is the following: since there exist exact test problems for problems of type A, and there are none for problems of type B, therefore, besides the fact that numerical solution of problems of type A is interesting and important (see e.g.,[19–21]), it has an additional role in this paper. Namely, verification of a scheme needed for numerical solution of Problem B and corresponding calculating program is carried out with the help of Problem A, which consists in following.

Function

$$u(x^0, x) = \frac{1}{|x - x^0|}, \quad x \in D, \quad x^0 = (x_1^0, x_2^0, x_3^0) \overline{\in} \overline{D}, \tag{3.1}$$

is taken in the role of the exact test solution for Problem A under boundary condition $g(y) = 1/|y - x^0|$, $y \in S$, where $|x - x^0|$ denotes the distance between the points x and x^0. After this function $g(y)$ is taken in the role of functions $g_i(y)$ ($i = \overline{1, m}$) in Problem B and consequently in calculating program. Evidently, in this case curves l_k represent removable discontinuity curves for functions $g_i(y)$, therefore instead of problem of type B we have problem of type A. Just for the obtained problem, verification of the scheme needed for numerical solution of Problem B and corresponding calculating program (comparison of the obtained results with exact solution) is carried out first of all, and then Problem B is being solved under boundary conditions (1.6).

In the case when Problems A and B concern electrostatic field, for full investigation of the field it is necessary to find both potential and strength of the field. It is known [5,6] that the strength $E(x) = (E_1(x), E_2(x), E_3(x))$ of electrostatic field is defined as follows:

$$E(x) = -gra\partial u(x) \equiv -\left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}, \frac{\partial u}{\partial x_3}\right), \ x \in D, \tag{3.2}$$

where $u(x)$ is potential of electrostatic field. It is known that the vector $E(x)$ is directed where the potential of the electric field is less.

Since in our case Problems A and B are solved by numerical method, for test problem, coordinates of vector $E(x)$ are defined by formula (3.2), and in the case of numerical solution by central difference formula

$$f'(t) = \frac{f(t + h) - f(t - h)}{2h} \tag{3.3}$$

is used, whose accuracy with respect to h is $O(h^2)$.

Thus on the basis of (3.2) and (3.3) for definition components of the vectors $E(x)$ and $E^N(x)$ we have:

$$E_k(x) = -\frac{\partial u(x)}{\partial x_k} = \frac{x_k - x_k^0}{|x - x^0|^3}, \quad (k = 1, 2, 3); \tag{3.4}$$

$$E_k^N(x) = -\frac{\partial u_N(x)}{\partial x_k} \approx -[u_N((x_1 + h)\delta_{1k}, (x_2 + h)\delta_{2k}, (x_3 + h)\delta_{3k}) - u_N((x_1 - h)\delta_{1k}, (x_2 - h)\delta_{2k}, (x_3 - h)\delta_{3k})]/(2h), \tag{3.5}$$

where δ_{ik} is Kronecker delta.

In the present paper we examine the application of the MPS to five examples. In tables, N is the number of the implementation of the Wiener process for the given points $x^i = (x_1^i, x_2^i, x_3^i) \in D$, and nq is the number of the quantification. For simplicity, in the considered examples the values of nq and N are one and the same. In tables for problems of type A we present absolute errors Δ^i at the points $x^i \in D$ of $u_N(x)$, in the MPS approximation, for $nq = 200$ and various values of N, and under notations of type $(E \pm k)$, $10^{\pm k}$ are meant. In particular,

$\Delta^i = |u_N(x^i) - u(x^0, x^i)|$, where $u_N(x^i)$ is the approximate solution of Problem A at the point x^i, which is defined by formula (2.2), and the exact solution $u(x^0, x^i)$ of the test problem is given by (3.1). In tables, for problems of type B, the probabilistic solution $u_N(x)$ is presented at the points x^i, defined by (2.2).

Example 3.1. In the first example the domain D is the interior of the triaxial ellipsoid S:

$$\left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2 + \left(\frac{x_3}{c}\right)^2 = 1, \tag{3.6}$$

where a, b, c are semi-axes of the ellipsoid, and $x(x_1, x_2, x_3)$ is a current point of the surface S.

In numerical experiments for the considered example, we took: (1) $a = 1$, $b = 2$, $c = 0.5$; (2) in the test problem (3.1) $x^0 = (0, 0, 5)$; (3) in Problem B the boundary function $g(y)$ has the form

$$g(y) \equiv g(y_1, y_2, y_3) = \begin{cases} 1, & y \in S_1 = \{y \in S \mid -2 \le y_2 < -1\}, \\ 2, & y \in S_2 = \{y \in S \mid -1 < y_2 < 1\}, \\ 1, & y \in S_3 = \{y \in S \mid 1 < y_2 \le 2\}, \\ 0, & y \in l_k \quad (k = 1, 2). \end{cases} \tag{3.7}$$

It is evident that in the considered case l_1 and l_2 are curves, which are obtained by intersection of the planes $x_2 = -1$, $x_2 = 1$ and the surface S (in the physical sense the curves l_1 and l_2 are non-conductors).

In order to determine the intersection points $y^i = (y_1^i, y_2^i, y_3^i)$ $(i = \overline{1, N})$ of the Wiener process and the surface S, we operate in the following way. During the implementation of the Wiener process, for each current point $x(t_k)$, defined from (2.5), its location with respect to S is checked, i,e., for the point $x(t_k)$ the value

$$d = \left(\frac{x_1(t_k)}{a}\right)^2 + \left(\frac{x_2(t_k)}{b}\right)^2 + \left(\frac{x_3(t_k)}{c}\right)^2$$

is calculated and the conditions $d = 1$, $d < 1$ or $d > 1$ are checked. If $d = 1$ then $x(t_k) \in S$ and $y^i = x(t_k)$. If $d < 1$ then $x(t_k) \in D$ and the process continues until it crosses the boundary, and if $d > 1$ then $x(t_k) \overline{\in} \overline{D}$.

Let $x(t_k) \in D$ for the moment $t = t_{k-1}$ and $x(t_k) \overline{\in} \overline{D}$ for the moment $t = t_k$. In this case, for approximate determination of the point y^i, a parametric equation of a line L passing through the points $x(t_{k-1})$ and $x(t_k)$ is firstly obtained, which has the following form

$$\begin{cases} x_1 = x_1^{k-1} + (x_1^k - x_1^{k-1})\theta, \\ x_2 = x_2^{k-1} + (x_2^k - x_2^{k-1})\theta, \\ x_3 = x_3^{k-1} + (x_3^k - x_3^{k-1})\theta, \end{cases} \tag{3.8}$$

where (x_1, x_2, x_3) is the current point of L and θ is a parameter $(\infty < \theta < \infty)$. After this, for definition the intersection points x^* and x^{**} of the line L and the surface S Eq. (3.6) is solved with respect to θ.

It is easy to see that for parameter θ we obtain an equation

$$A\theta^2 + 2B\theta + C = 0 \tag{3.9}$$

where $A > 0$ and $C < 0$.

Since the discriminant of (3.9) is positive, the points x^* and x^{**} are defined respectively on the basis of (3.8) for solutions of (3.9) θ_1 and θ_2. In the role of the points y^i we choose the one (from x^* and x^{**}) for which $|x(t_k) - x|$ is minimal.

In Table 3.1A the absolute errors Δ^i of the approximate solution $u_N(x)$ to the test problem at the points $x^i \in D$ $(i = \overline{1, 5})$ are presented. In the considered case $x^0 \in Ox_3$, therefore, the exact solution $u(x^0, x)$ is symmetric with respect to the planes Ox_1x_3 and Ox_2x_3, respectively, for control, in the role of the points x^i we took points which are symmetric with respect to the plane Ox_1x_3 and $x^i \in Ox_2$. On the basis of (3.4) and (3.5) we calculated exact and approximate electric field (or $E_2(x)$ and $E_2^N(x)$) on the axis Ox_2 at the points x^i.

For points x^i we examined the case when $N = 10^6, nq = 200, h = 0.03$ and by (3.4) and (3.5) we obtained the following results: $E_2(0, -1.8, 0) = -0.0119943$; $E_2(0, -1, 0) = -0.007543$; $E_2(0, 0, 0) = 0$; $E_2(0, 1, 0) = -E_2(0, -1, 0)$; $E_2(0, 1.8, 0) = -E_2(0, -1.8, 0)$; $E_2^N(0, -1.8, 0) = -0.011567$; $E_2^N(0, -1, 0) = -0.0078483$; $E_2^N(0, 0, 0) = 0.54E - 3$; $E_2^N(0, 1, 0) = 0.0078483$; $E_2^N(0, 1.8, 0) = 0.0119996$;

It is evident that the results obtained for $E_2^N(x^i)$ are in good agreement with the values of $E_2(x^i)$.

Table 3.1A
Results for Problem A (in Example 3.1).

x^i	$(0, -1.8, 0)$	$(0, -1, 0)$	$(0, 0, 0)$	$(0, 1, 0)$	$(0, 1.8, 0)$
N	$nq = 200$ Δ^1	$nq = 200$ Δ^2	$nq = 200$ Δ^3	$nq = 200$ Δ^4	$nq = 200$ Δ^5
$5E{+}3$	$0.42E{-}4$	$0.19E{-}3$	$0.68E{-}4$	$0.46E{-}4$	$0.17E{-}3$
$1E{+}4$	$0.63E{-}4$	$0.26E{-}3$	$0.86E{-}4$	$0.67E{-}4$	$0.32E{-}4$
$5E{+}4$	$0.12E{-}4$	$0.62E{-}4$	$0.56E{-}5$	$0.14E{-}4$	$0.91E{-}6$
$1E{+}5$	$0.57E{-}5$	$0.94E{-}4$	$0.77E{-}4$	$0.11E{-}4$	$0.88E{-}5$
$4E{+}5$	$0.22E{-}4$	$0.42E{-}5$	$0.13E{-}4$	$0.73E{-}5$	$0.13E{-}4$
$1E{+}6$	$0.14E{-}4$	$0.37E{-}4$	$0.21E{-}4$	$0.54E{-}5$	$0.96E{-}5$

Table 3.1B
Results for Problem B (in Example 3.1).

x^i	$(0, -1.8, 0)$	$(0, -1, 0)$	$(0, 0, 0)$	$(0, 1, 0)$	$(0, 1.8, 0)$
N	$nq = 200$	$nq = 200$	$nq = 200$	$nq = 200$	$nq = 200$
$5E{+}3$	1.01000	1.46480	1.96540	1.47460	1.00560
$1E{+}4$	1.00610	1.46190	1.96740	1.48200	1.00550
$5E{+}4$	1.00758	1.46084	1.96610	1.46296	1.00682
$1E{+}5$	1.00706	1.46610	1.96758	1.46413	1.00734
$4E{+}5$	1.00762	1.46517	1.96705	1.46528	1.00734
$1E{+}6$	1.00745	1.46499	1.96757	1.46524	1.00748

In Table 3.1B the values of the approximate solution $u_N(x)$ to Problem B at the same points $x^i (i = \overline{1,5})$ are given. The boundary function (3.7) (or Problem B) is symmetric with respect to the coordinate planes, respectively, the obtained results are symmetric with respect to the plane Ox_1x_3 and have sufficient accuracy for many practical problems.

For illustration, we calculated the electrostatic field strength by (3.5) on the axis Ox_2 at the same points x^i $(i = \overline{1,5})$ for $N = 1E + 6$, $nq = 200$, $h = 0.03$. We obtained the following results: $E_2^N(0, -1.8, 0) = -0.0670$; $E_2^N(0, -1, 0) = -1.25241$; $E_2^N(0, 0, 0) = 0.56E{-}3$; $E_2^N(0, 1, 0) = 1.25581$; $E_2^N(0, 1.8, 0) = 0.06731$. The obtained results are in good agreement with the real physical picture.

Example 3.2. In this example the interior of the unit sphere $S : x_1^2+x_2^2+x_3^2 = 1$ with the center at the origin $O(0, 0, 0)$ is taken in the role of the domain D, where $x(x_1, x_2, x_3)$ is a current point of the surface S. In the considered case in the role of x^0 (see (3.1)) $x^0 = (0, 0, 5)$ is taken, and the boundary function $g(y)$ has the form

$$g(y) = \begin{cases} 1, & y \in S_1 = \{y \in S \,|y_1 > 0, y_2 > 0, y_3 > 0\}, \\ 0, & y \in S_2 = \{y \in S \,|y_1 < 0, y_2 > 0, y_3 > 0\}, \\ 1, & y \in S_3 = \{y \in S \,|y_1 < 0, y_2 < 0, y_3 > 0\}, \\ 0, & y \in S_4 = \{y \in S \,|y_1 > 0, y_2 < 0, y_3 > 0\}, \\ 2, & y \in S_5 = \{y \in S \,|y_1 > 0, y_2 > 0, y_3 < 0\}, \\ 1, & y \in S_6 = \{y \in S \,|y_1 < 0, y_2 > 0, y_3 < 0\}, \\ 2, & y \in S_7 = \{y \in S \,|y_1 < 0, y_2 < 0, y_3 < 0\}, \\ 1, & y \in S_8 = \{y \in S \,|y_1 > 0, y_2 < 0, y_3 < 0\}, \\ 0, & y \in l_k \quad (k = 1, 2, 3). \end{cases} \tag{3.10}$$

It is evident that in this case the discontinuity curves l_k $(k = 1, 2, 3)$ are the circles, obtained by intersection of the coordinate planes and the sphere S. Actually, the sphere S is divided into equal parts S_i $(i = \overline{1,8})$ by curves l_k $(k = 1, 2, 3)$. Besides, in the considered case, l_k, S_2, S_4 are non-conductors. Since the sphere is special case of the triaxial ellipsoid, the algorithm which is given in Example 3.1 for approximate determination of the intersection points of the Wiener process and the ellipsoid is applied to this case.

In Table 3.2A the absolute errors Δ^i of the approximate solution $u_N(x)$ of the test problem are presented at the points $x^i \in D$ $(i = 1, 2, 3)$ which are symmetric with respect to the plane Ox_1x_3 and $x^i \in Ox_2$. By use of (3.4) and (3.5) we calculated $E_2(x)$ and $E_2^N(x)$ on the axis Ox_2 at the points x^i for $N = 10^6$, $nq = 200$, $h = 0.03$ and we

Table 3.2A
Results for Problem A (in Example 3.2).

x^i	$(0, 0, -0.8)$	$(0, 0, 0)$	$(0, 0, 0.8)$
N	$nq = 200$ Δ^1	$nq = 200$ Δ^2	$nq = 200$ Δ^3
$5E{+}3$	$0.31E{-}3$	$0.41E{-}3$	$0.26E{-}3$
$1E{+}4$	$0.12E{-}3$	$0.14E{-}3$	$0.62E{-}4$
$5E{+}4$	$0.58E{-}4$	$0.76E{-}4$	$0.23E{-}4$
$1E{+}5$	$0.60E{-}4$	$0.79E{-}4$	$0.15E{-}4$
$4E{+}5$	$0.24E{-}4$	$0.18E{-}4$	$0.20E{-}4$
$1E{+}6$	$0.16E{-}4$	$0.57E{-}4$	$0.23E{-}5$

Table 3.2B
Results for Problem B (in Example 3.2).

x^i	$(0, 0, -0.8)$	$(0, 0, 0)$	$(0, 0, 0.8)$	(b, b, b)	$(-b, -b, b)$
N	$nq = 200$	$nq = 200$	$nq = 200$	$nq = 200$	$nq = 200$
$5E{+}3$	1.45260	1.00500	0.54440	0.96180	0.97000
$1E{+}4$	1.45020	0.99940	0.55340	0.96870	0.96000
$5E{+}4$	1.44740	0.99816	0.55102	0.97130	0.96890
$1E{+}5$	1.44879	1.00006	0.55123	0.97089	0.96920
$4E{+}5$	1.44945	0.99793	0.55193	0.96967	0.97022
$1E{+}6$	1.44952	1.00089	0.55095	0.97003	0.97008

obtained the following results: $E_2(0, -0.8, 0) = -0.006161876$; $E_2(0, 0, 0) = 0$; $E_2(0, 0.8, 0) = -E_2(0, -0.8, 0)$; $E_2^N(0, -0.8, 0) = -0.006192$; $E_2^N(0, 0, 0) = 0.46E - 4$; $E_2^N(0, 0.8, 0) = 0.005729$.

The approximate results obtained for $E_2^N(x^i)$ are in good agreement with the values of $E_2(x^i)$.

In Table 3.2B, the values of the approximate solution $u_N(x)$ of Problem B at the points $x^i \in D$ ($i = \overline{1, 5}$) are given and $b = 0.5$. In the role of the points x^i ($i = 1, 2, 3$), points which are symmetric with respect to the plane Ox_1x_2 and situated on the axis Ox_3 are taken. Since the boundary function (3.10) (or Problem B) is symmetric with respect to the axis Ox_3, for control, in the role of x^i ($i = 4, 5$) we took symmetric points with respect to Ox_3. The obtained results have sufficient accuracy for many practical problems (see Table 3.2B). By use of (3.5) we calculated the strength $E^N(x)$ on the axis Ox_3 at the points x^i ($i = 1, 2, 3$) for $N = 10^6$, $nq = 200$, $h = 0.03$ and we obtained the following results: $E_3^N(0, 0, -0.8) = 0.2287$; $E_3^N(0, 0, 0,) = 0.8356$; $E_3^N(0, 0, 0.8) = 0.256$. The obtained results are in good agreement with the real physical picture.

Example 3.3. Here we consider a simplified generalized problem of type B, in which the kernel with the center at the origin O(0,0,0) and with radius a is taken in the role of the domain D. In particular, we solve Problem B when on the sphere $S(O; a)$, the boundary function $g(y)$ has the following form

$$g(y) = \begin{cases} V_1, & y \in S_1 = \{y \in S \,|0 < y_3 \le a\}, \\ V_2, & y \in S_2 = \{y \in S \mid -a \le y_3 < 0\}, \\ 0, & y \in l. \end{cases} \tag{3.11}$$

In (3.11): V_1 and V_2 are constants; l is the discontinuity curve (circle) which is obtained by intersection of the plane $x_3 = 0$ and the surface S (it is non-conductor). In numerical experiments we took: $a = 1$, $V_1 = 2$, $V_2 = 1$ and $x^0 = (0, 0, 5)$. Analogously to Example 3.2, in this case, for determination of the intersection points y^i ($i = \overline{1, N}$) the same algorithm is applied , which is described in Example 3.1.

In Table 3.3A the absolute errors Δ^i of the approximate solution $u_N(x)$ of the test problem are presented at the points $x^i \in D$ ($i = 1, 2, 3$), which are symmetric with respect to the plane Ox_1x_2 and $x^i \in Ox_3$. By use of (3.4) and (3.5) we calculated $E_3(x)$ and $E_3^N(x)$ on the axis Ox_3 at the points x^i for $N = 2 \times 10^5$, $nq = 200$, $h = 0.03$ and we obtained the following results: $E_3(0, 0, -0.8) = -0.0297265$; $E_3(0, 0, 0) = -0.040$; $E_3(0, 0, 0.8) = -0.056689$; $E_3^N(0, 0, -0.8) = -0.030010$; $E_3^N(0, 0, 0) = -0.04130$; $E_3^N(0, 0, 0.8) = -0.057218$.

The approximate results obtained for $E_3^N(x^i)$ are in good agreement with the values of $E_3(x^i)$.

Table 3.3A
Results for Problem A (in Example 3.3).

x^i	$(0, 0, -0.8)$	$(0, 0, 0)$	$(0, 0, 0.8)$
N	$nq = 200$ Δ^1	$nq = 200$ Δ^2	$nq = 200$ Δ^3
$5E{+}3$	$0.93E{-}4$	$0.26E{-}3$	$0.18E{-}3$
$1E{+}4$	$0.20E{-}3$	$0.17E{-}3$	$0.35E{-}3$
$5E{+}4$	$0.58E{-}4$	$0.97E{-}4$	$0.75E{-}4$
$1E{+}5$	$0.27E{-}4$	$0.99E{-}5$	$0.20E{-}3$
$2E{+}5$	$0.92E{-}7$	$0.39E{-}4$	$0.11E{-}3$
$4E{+}5$	$0.82E{-}4$	$0.61E{-}4$	$0.15E{-}3$
$1E{+}6$	$0.62E{-}4$	$0.24E{-}4$	$0.15E{-}3$

Table 3.3B
Results for Problem B (in Example 3.3).

x^i	$(0, 0, -0.8)$	$(0, 0, 0)$	$(0, 0, 0.8)$	$(0, -0.8, 0)$	$(0, 0.8, 0)$
N	$nq = 200$	$nq = 200$	$nq = 200$	$nq = 200$	$nq = 200$
$5E{+}3$	1.05140	1.50260	1.94860	1.50400	1.48380
$1E{+}4$	1.05380	1.50350	1.95040	1.50290	1.50630
$5E{+}4$	1.05122	1.50030	1.94808	1.49706	1.49812
$1E{+}5$	1.05039	1.50085	1.94840	1.49797	1.50033
$4E{+}5$	1.05132	1.50018	1.94893	1.49830	1.50017
$1E{+}6$	1.05170	1.50016	1.94847	1.49929	1.49971

The values of the approximate solution $u_N(x)$ of Problem B at the points $x^i \in D(i = \overline{1, 5})$ are given in Table 3.3B. Points symmetric with respect to the planes Ox_1x_2, Ox_1x_3 and situated on the axes Ox_3 and Ox_2, respectively, are taken in the role of the points x^i. The obtained results have the sufficient accuracy for many practical problems.

By use of (3.5) we calculated $E_3^N(x)$ at the points x^i $(i = 1, 2, 3)$ for $N = 10^6$, $nq = 200$, $h = 0.03$ and we obtained the following results: $E_3^N(0, 0, -0.8) = -0.3013$; $E_3^N(0, 0, 0) = -0.75931$; $E_3^N(0, 0, 0.8) = -0.30104$. The obtained results are in good agreement with the real physical picture.

It should be noted that Example 3.3 is considered in [4,8], where it is solved by the method of separation of variables (under solving to Problem B an existence of the dielectric l is neglected). It is shown that in conditions (3.11) the "exact" analytical solution to Problem B has the following form (in spherical coordinates)

$$u(r, \theta) = \frac{V_1 + V_2}{2} + \frac{V_1 - V_2}{2} w(r, \theta), \tag{3.12}$$

$$w(r, \theta) = \sum_{n=1}^{\infty} (r/a)^n \frac{2n + 1}{n + 1} P_{n-1}(0) P_n(\cos\theta), \tag{3.13}$$

where $0 \le r \le a$ $(r^2 = x_1^2 + x_2^2 + x_3^2)$, $0 \le \theta \le \pi$, and P_n is Legendre's polynomial of order n. Namely, $P_0(\lambda) = 1$, $P_1(\lambda) = \lambda$, $|P_n(\lambda)| \le 1$, where $-1 \le \lambda \le 1$, and $P_n(0) = 0$ for $n = 2k + 1$.

It is evident that the series (3.13) converges rapidly for all points $(r, \theta) \in D$, i.e., when $0 \le r < a$. If $r = a$, then the rate of convergence becomes worse on S, especially in the neighborhood of curve l. In particular, if $r = a$ and $\theta \to \pi/2$ then the convergence is very slow and consequently, the accuracy in the satisfaction of boundary condition on S is very low (see Section 1). This is caused by the fact that, when $\theta \to \pi/2$, all terms of the series (3.13) tend to zero. Indeed, if $\theta \to \pi/2$ then $P_n(cos\theta) \to P_n(0)$ and in (3.13) $P_{n-1}(0) = 0$ or $P_n(0) = 0$ for given n. Besides, from (3.12) $u(1, \pi/2) = (V_1 + V_2)/2 = V \ne 0$, consequently, the discontinuity curve l is conductor. Actually, Problem B for Example 3.3 is solved for that case, when in the boundary condition (3.11) $g(y) = V$, $y \in l$, respectively, the initial physical model is changed.

Since, boundary function (3.11) is independent on the angle of rotation with respect to axis Ox_3, therefore, for illustration in Table 3.3C the results of the calculations for the sum of the first $m + 1$ terms of the series (3.12) are presented. In numerical experiments, in the role of a, V_1, V_2 were taken the same values. Because of the above-mentioned $u(r, \theta)$ is calculated at the points (r, θ) $(r = a$ and $0 \le \theta \le \pi$, $\theta \ne \pi/2)$ which represent a certain interest.

Table 3.3C
Results for series (3.12).

m	$\theta = 0.0$	$\theta = 0.7854$	$\theta = 1.521$	$\theta = 1.5707$	$\theta = 1.57079$
$1E{+}2$	1.9602	2.0049	1.9918	1.5031	1.5001
$1E{+}3$	1.9874	1.9995	1.9939	1.5307	1.5010
$1E{+}4$	1.9960	1.99995	2.0006	1.7913	1.5106
$5E{+}4$	1.9982	1.99998	1.9999	2.0050	1.5529
m	$\theta = 1.5708$	$\theta = 1.5718$	$\theta = 1.671$	$\theta = 3.1415$	$\theta = 3.14159$
$1E{+}2$	1.4999	1.4986	0.9729	1.0398	1.0398
$1E{+}3$	1.4988	1.1977	1.0028	1.0126	1.0126
$1E{+}4$	1.4883	0.9728	1.0002	1.0032	1.0040
$5E{+}4$	1.4416	1.0063	1.0000	0.9995	1.0018

In (3.13) for calculation of Legendre's polynomial $P_n(x)$ we used the following recursion formula (see [2])

$$P_n(\lambda) = \frac{2n-1}{n}\ \lambda\ P_{n-1}(\lambda) - \frac{n-1}{n} P_{n-2}(\lambda), \quad n \geq 2.$$

From Table 3.3C it is clear that accuracy of the "exact" solution $u(r, \theta)$ is very low in the neighborhood of the circle l, that was possible. As we above mentioned when $r < a$, then the series (3.13) converges rapidly, indeed, the calculations have shown that practically $u(0.8, 0) = 1.9493$, $u(0.8, \pi) = 1.0507$ and $u(0, \theta) = 1.5$ when $m = 100$, 1000, 10,000, 50,000. These results are sufficiently close to results of $u(0, 0, 0.8)$, $u(0, 0, -0.8)$ and $u(0, 0, 0)$ which are presented in Table 3.3B. This fact was expected, since on the basis of the extremum principle and condition (3.11) $|u(x) - u(r, \theta)|$ is minimal on the axis Ox_3, where $u(x)$ is the exact solution of Problem 3.3B.

Remark 2. If V_1 or V_2 is not constant then the analytic form of the "exact" solution is so difficult in the sense of numerical implementation, that it has only theoretical significance (see [4]).

Example 3.4. Here in the role of the domain D we took the kernel layer, which is bounded by the planes $x_3 = h_1$, $x_3 = h_2$ and the unit sphere $x_1^2 + x_2^2 + x_3^2 = 1$, where $-1 < h_1 < h_2 < 1$. Actually, the boundary S of D is $S = S_1 \bigcup S_2 \bigcup S_3 \bigcup l_1 \bigcup l_2$ where $S_1 : x_1^2 + x_2^2 < r_1^2,\ x_3 = h_1;\ S_2 : x_1^2 + x_2^2 + x_3^2 = 1,\ h_1 < x_3 < h_2;\ S_3 : x_1^2 + x_2^2 < r_2^2,\ x_3 = h_2$; l_1 and l_2 are the circles of the disks S_1 and S_2 with the radii $r_1 = \sqrt{1 - h_1^2},\ r_2 = \sqrt{1 - h_2^2}$, respectively. If $h_1 = 0$ and $h_2 = 1$, then the domain D will be hemikernel.

We solved Problems A and B when $h_1 = -0.5$, $h_2 = 0.5$, $x^0 = (5, 0, 0)$ and on the boundary S the function $g(y)$ has the form

$$\begin{cases} 1, & y \in S_1, \\ 2, & y \in S_2, \\ 1, & y \in S_3, \\ 0, & y \in l_k (k = 1, 2), \end{cases} \tag{3.14}$$

where the discontinuity curves l_k $(k = 1, 2)$ are non-conductors.

In the considered case, for determination of the intersection points y^i $(i = \overline{1, N})$ of the Wiener process and the spherical surface S_2 the same algorithm, described in Example 3.1 is applied.

In Table 3.4A, the absolute errors Δ^i of the approximate solution $u_N(x)$ of the test problem at the points $x^i \in D$ $(i = \overline{1, 5})$ are given. Since in the considered case $x^0 \in Ox_1$, the exact solution $u(x^0, x)$ (see (3.1)) is symmetric with respect to the planes Ox_1x_2 and Ox_1x_3. For control, in the role of the points x^i we took points symmetric with respect to the mentioned planes and are situated on Ox_2 and Ox_3. By use of (3.4) and (3.5) we calculated $E_2(x)$, $E_2^N(x)$, $E_3(x)$, $E_3^N(x)$ for $N = 10^6, nq = 200, h = 0.03$ and we obtained the following results: $E_2^N(0, -0.8, 0) = -0.0061745$; $E_2^N(0, 0, 0) = 0.83E - 3$; $E_2^N(0, 0.8, 0) = 0.005921$; $E_3(0, 0, -0.4) = -0.00316952$; $E_3(0, 0, 0) = 0$; $E_3(0, 0, 0.4) = -E_3(0, 0, -0.4)$; $E_3^N(0, 0, -0.4) = -0.003004$; $E_3^N(0, 0, 0) = 0.79E - 4$; $E_3^N(0, 0, 0.4) = 0.003206$.

The approximate results obtained for $E_2^N(x^i)$ and $E_2^N(x^i)$ are in good agreement with the values of $E_2(x^i)$ and $E_3(x^i)$ at the corresponding points.

Table 3.4A
Results for Problem A (in Example 3.4).

x^i	$(0, -0.8, 0)$	$(0, 0, 0)$	$(0, 0.8, 0)$	$(0, 0, -0.4)$	$(0, 0, 0.4)$
N	$nq = 200$ Δ^1	$nq = 200$ Δ^2	$nq = 200$ Δ^3	$nq = 200$ Δ^4	$nq = 200$ Δ^5
$5E{+}3$	$0.14E{-}4$	$0.17E{-}3$	$0.14E{-}4$	$0.17E{-}3$	$0.41E{-}3$
$1E{+}4$	$0.17E{-}4$	$0.14E{-}3$	$0.21E{-}4$	$0.34E{-}3$	$0.11E{-}3$
$5E{+}4$	$0.10E{-}3$	$0.42E{-}4$	$0.64E{-}4$	$0.84E{-}4$	$0.85E{-}4$
$1E{+}5$	$0.57E{-}4$	$0.56E{-}4$	$0.17E{-}4$	$0.22E{-}4$	$0.51E{-}4$
$4E{+}5$	$0.37E{-}4$	$0.13E{-}4$	$0.22E{-}4$	$0.25E{-}4$	$0.24E{-}4$
$1E{+}6$	$0.58E{-}4$	$0.16E{-}4$	$0.52E{-}4$	$0.17E{-}4$	$0.88E{-}6$

Table 3.4B
Results for Problem B (in Example 3.4).

x^i	$(0, -0.8, 0)$	$(0, 0, 0)$	$(0, 0.8, 0)$	$(0, 0, -0.4)$	$(0, 0, 0.4)$
N	$nq = 200$	$nq = 200$	$nq = 200$	$nq = 200$	$nq = 200$
$5E{+}3$	1.77700	1.27280	1.76940	1.08900	1.08700
$1E{+}4$	1.77150	1.27940	1.77190	1.09020	1.08600
$5E{+}4$	1.76886	1.27582	1.76780	1.08770	1.08660
$1E{+}5$	1.76898	1.27285	1.76781	1.08640	1.08620
$4E{+}5$	1.76829	1.27251	1.76782	1.08722	1.08725
$1E{+}6$	1.76900	1.27342	1.76883	1.08680	1.08703

The values of the approximate solution $u_N(x)$ of Problem B at the same points $x^i \in D$ $(i = \overline{1,5})$ are given in Table 3.4B. It is evident that the boundary function (3.14) is symmetric with respect to the coordinate planes, respectively. The obtained results have sufficient accuracy for many practical problems.

On the basis of (3.4) and (3.5) we calculated $E_2^N(x)$ and $E_3^N(x)$ at the points x^i for $N = 10^6$, $nq = 200$, $h = 0.03$ and we obtained the following results: $E_2^N(0, -0.8, 0) = 1.20433$; $E_2^N(0, 0, 0) = 0.52E - 2$; $E_2^N(0, 0.8, 0) = -1.1932$; $E_3^N(0, 0, -0.4) = -0.8075$; $E_3^N(0, 0, 0) = 0.75E - 2$; $E_3^N(0, 0, 0.4) = 0.8205$. The obtained results are in good agreement with the real physical picture.

Example 3.5. In the role of the domain D we took the finite right circular cylinder $D(0 \le r \le a, 0 \le x_3 \le h)$, where $r = \sqrt{x_1^2 + x_2^2}$, h is height of the cylinder, and a is its radius. We solved Problems A and B for $a = 0.5$, $h = 2$, $x^0 = (5, 0, 1)$ and the boundary function $g(y)$ with

$$g(y) = \begin{cases} 1, & y \in S_1 = \{y \in S \,|0 \le r < 0.5, y_3 = 0\}, \\ 0.5, & y \in S_2 = \{y \in S \,|0 \le r < 0.5, y_3 = 2\}, \\ 1.5, & y \in S_3 = \{y \in S \,|0 < \varphi < \pi/2\}, \\ 0, & y \in S_4 = \{y \in S \,|\pi/2 < \varphi < \pi\}, \\ 1.5, & y \in S_5 = \{y \in S \,|\pi < \varphi < 3\pi/2\}, \\ 0, & y \in S_6 = \{y \in S \,|3\pi/2 < \varphi < 2\pi\}, \\ 0, & y \in l_k \ (k = \overline{1,6}). \end{cases} \tag{3.15}$$

on the cylindrical surface S.

In the expression of $g(y)$, φ is the polar angle; l_1 and l_2 are the circles of the bases of the cylinder; l_3, l_4, l_5, l_6 are the generatrices of the cylinder D for $\varphi = 0$, $\varphi = \pi/2$, $\varphi = \pi$, $\varphi = 3 * \pi/2$, respectively. It should be noted that in [5,6,10] the simple case when only l_1 and l_2 are the discontinuity curves is considered. In this case l_k $(k = \overline{1,6})$, S_4 and S_6 are non-conductors.

In the considered case, for determination of the intersection points y^i $(i = \overline{1,N})$ of the Wiener process and the lateral surface of the cylinder the same algorithm described in Example 3.1 is applied, where instead of Eq. (3.1) is taken the equation $x_1^2 + x_2^2 = a^2$.

Table 3.5A
Results for Problem A (in Example 3.5).

x^i	$(0, 0, 0.2)$	$(0, 0, 1)$	$(0, 0, 1.8)$
N	$nq = 200$ Δ^1	$nq = 200$ Δ^2	$nq = 200$ Δ^3
$5E{+}3$	$0.22E{-}4$	$0.29E{-}4$	$0.73E{-}4$
$1E{+}4$	$0.11E{-}4$	$0.20E{-}4$	$0.50E{-}4$
$5E{+}4$	$0.31E{-}4$	$0.34E{-}4$	$0.34E{-}4$
$1E{+}5$	$0.24E{-}4$	$0.38E{-}4$	$0.12E{-}4$
$4E{+}5$	$0.18E{-}6$	$0.66E{-}5$	$0.18E{-}4$
$1E{+}6$	$0.83E{-}6$	$0.21E{-}4$	$0.21E{-}5$

Table 3.5B
Results for Problem B (in the cylinder).

x^i	$(0, 0, 0.2)$	$(0, 0, 1)$	$(0, 0, 1.8)$	$(b, b, 1)$	$-b, -b, 1$
N	$nq = 200$	$nq = 200$	$nq = 200$	$nq = 200$	$nq = 200$
$5E{+}3$	0.89750	0.74150	0.64850	1.18200	1.16800
$1E{+}4$	0.88520	0.75950	0.64225	1.18330	1.17265
$1E{+}4$	0.88036	0.75391	0.63800	1.19005	1.18498
$1E{+}5$	0.87881	0.74886	0.64000	1.18904	1.18904
$4E{+}5$	0.87801	0.75263	0.63992	1.18706	1.18934
$1E{+}6$	0.87939	0.75054	0.63987	1.18785	1.18802

The absolute errors Δ^i of the approximate solution $u_N(x)$ of the test problem at the points $x^i \in D$ $(i = 1, 2, 3)$ are given in Table 3.5A. Since in the considered case $x^0 = (5, 0, 1)$, the exact solution $u(x^0, x)$ (see (3.1)) is symmetric with respect to the planes Ox_1x_3 and $x_3 = 1$. For control in the role of the points x^i we took points symmetric with respect to the plane $x_3 = 1$ and $x^i \in Ox_3$. By use of (3.4) and (3.5) we calculated $E_3(x)$ and $E_3^N(x)$ on the axis Ox_3 at the same points x^i for $N = 10^6$, $nq = 200$, $h = 0.03$ and we obtained the following results: $E_3(0, 0, 0.2) = -0.00616876$; $E_3(0, 0, 1) = 0$; $E(0, 0, 1.8) = 0.006161876$; $E_3^N(0, 0, 0.2) = -0.006191$; $E_3^N(0, 0, 1) = 0.62E - 5$; $E_3^N(0, 0, 1.8) = 0.005962$.

The approximate results obtained for $E_3^N(x^i)$ are in excellent agreement with the values of $E_3(x^i)$ at the corresponding points.

The values of approximate solution $u_N(x)$ of Problem B at the points $x^i \in (i = \overline{1, 5})$ are given in Table 3.5B, and $b = 0.25$. From the denoted points the points x^i $(i = 1, 2, 3)$ are same as those in Table 3.5A. Since the boundary function (3.15) is symmetric with respect to the axis Ox_3, for control in the role of the points x^i $(i = 4, 5)$ the points which are symmetric with respect to the axis Ox_3 are taken. The obtained results have sufficient accuracy for many practical problems.

By use of (3.5) we calculated $E_3^N(x)$ at the points x^i $(i = 1, 2, 3)$ for $N = 10^6$, $nq = 200$, $h = 0.03$ and we obtained the following results: $E_3^N(0, 0, 0.2) = 0.4507$; $E_3^N(0, 0, 1) = 0.0343$; $E_3^N(0, 0, 1.8) = 0.4225$. The obtained results are in good agreement with the real physical picture.

In this work we specially solved the problems of type B when boundary functions $g_i(y)$ $(i = \overline{1, m})$ are constants. This was caused by our interest if how much the obtained results were in agreement with real physical picture. It is evident that solving of Problem B under condition (1.6) is not difficult. Indeed, after finding the intersection point y^i of the Wiener process and the surface S, it is easy to establish the part of S in which the point y^i is situated. In general, we can solve Problem B for all such locations of discontinuity curves, which give the possibility to establish the part of surface S where the intersection point is located.

The analysis of the results of numerical experiments shows that the results obtained by suggested algorithm are reliable and it is effective for numerical solution of problems of types A and B. In particular, the algorithm is sufficiently simple for numerical implementation.

Besides, it should be noted that the accuracy of probabilistic solution of problems of types A and B is not significantly increasing (except of some cases, see tables) when $N \to \infty$. The reason of this is fixed nq (the number of the quantification). If we need more accuracy, then calculations for sufficiently large values of nq and N (see [10])

must be realized. In this case, numerical realization on a PC takes very much time. We can avoid this difficulty if we apply the method of parallel calculation. For this suitable computing technique is needed. Respectively, significantly less time will be needed for numerical realization and besides the accuracy of the obtained results will improve.

4. Concluding remarks

1. In this work , we have demonstrated that the method of probabilistic solution (MPS) is ideally suited for numerical solving of both ordinary and generalized (2D and 3D) Dirichlet problems for rather a wide class of domains, in the case of Laplace equation.

2. The MPS does not require an approximation of a boundary function, which is one of its important properties.

3. The MPS is a fast solver for the above noted problems. Besides, it is easy to program, its computational cost is low, it is characterized by an accuracy which is sufficient for many problems.

References

[1] M.A. Lavrent'jev, B.V. Shabat, Methods of the theory of functions of a complex variable, Nauka, Moscow, 1973 (in Russian).
[2] A.N. Tikhonov, A.A. Samarskiĭ, The Equations of Mathematical Physics, fourth ed, Izdat, Nauka, Moscow, 1972 corrected.
[3] V.S. Vladimirov, Equations of Mathematical Physics, second ed., Izdat, Nauka, Moscow, 1971 revised and augmented.
[4] N.S. Koshlyakov, E.B. Gliner, M.M. Smirnov, Equations in Partial Derivatives of Mathematical Physics, Moscow, 1970 (in Russian).
[5] G.A. Grinberg, The selected questions of mathematical theory of electric and magnetic phenomena, Izd. Akad. Nauk SSSR (1948) in Russian.
[6] WilliamR. Smythe, Static and Dynamic Electricity, second ed., New York, Toronto, London, 1950.
[7] H.S. Carslaw, J.C. Jaeger, Conduction of Heat in Solids, Oxford University Press, London, 1959.
[8] B.M. Budak, A.A. Samarskiĭ, A.N. Tikhonov, A collection of problems in mathematical physics, third ed., Nauka, Moscow, 1980, in Russian.
[9] E.B. Duenkin, A.A. Yushkevich, Theorems and problems on Markov's processes, Nauka, Moscow, 1967, (in Russian).
[10] M. Zakradze, M. Kublashvili, Z. Sanikidze, N. Koblishvili, Investigation and numerical solution of some 3D internal Dirichlet generalized harmonic problems in finite domains, Trans. A. Razmadze Math. Inst. 171 (1) (2017) 103–110.
[11] L.V. Kantorovich, V.I. Krylov, Approximate methods of higher analysis, fifth corrected edition, Gosudarstv. Izdat. Fiz.-Mat. Lit. Moscow-Leningrad 1962 (in Russian).
[12] A. Karageorghis, Modified methods of fundamental solutions for harmonic and biharmonic problems with boundary singularities, Numer. Methods Partial Differential Equations 8 (1) (1992) 1–19.
[13] N. Koblishvili, Z. Tabagari, M. Zakradze, On reduction of the Dirichlet generalized boundary value problem to an ordinary problem for harmonic function, Proc. A. Razmadze Math. Inst. 132 (2003) 93–106.
[14] M. Zakradze, N. Koblishvili, A. Karageorghis, Y. Smyrlis, On solving the Dirichlet generalized problem for harmonic function by the method of fundamental solutions, Semin. I. Vekua Inst. Appl. Math. Rep. 34 (2008) 24–32 124.
[15] M. Kublashvili, Z. Sanikidze, M. Zakradze, A method of conformal mapping for solving the generalized Dirichlet problem of Laplace's equation, Proc. A. Razmadze Math. Inst. 160 (2012) 71–89.
[16] N. Koblishvili, M. Zakradze, On solving the Dirichlet generalized problem for a harmonic function in the case of infinite plane with holes, Proc. A. Razmadze Math. Inst. 164 (2014) 71–82.
[17] N. Koblishvili, M. Kublashvili, Z. Sanikidze, M. Zakradze, On solving the Dirichlet generalized problem for a harmonic function in the case of an infinite plane with a crack-type cut, Proc. A. Razmadze Math. Inst. 168 (2015) 53–62.
[18] A.D. Venttsel', A Course in the Theory of Random Processes, Izdat, Nauka, Moscow, 1975 in Russian.
[19] A.Sh. Chaduneli, M.V. Zakradze, Z.A. Tabagari, A method of probabilistic solution to the ordinary and generalized plane Dirichlet problem for the Laplace equation, in: Science and Computing, Proc. Sixth ISTC Scientific Advisory Committee Seminar, vol. 2, Moscow, 2003, pp. 361–366.
[20] A.Sh. Chaduneli, Z.A. Tabagari, M. Zakradze, On solving the internal three-dimensional Dirichlet problem for a harmonic function by the method of probabilistic solution, Bull. Georgian Natl. Acad. Sci. (N.S.) 2 (1) (2008) 25–28.
[21] M. Zakradze, Z. Sanikidze, Z. Tabagari, On solving the external three-dimensional Dirichlet problem for a harmonic function by the probabilistic method, Bull. Georgian Natl. Acad. Sci. (N.S.) 4 (3) (2010) 19–23.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 466–474
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Several series identities involving the Catalan numbers

Li Yin[a], Feng Qi[b,c,*]

[a] *Department of Mathematics, Binzhou University, Binzhou, Shandong Province, 256603, China*
[b] *Institute of Mathematics, Henan Polytechnic University, Jiaozuo, Henan, 454010, China*
[c] *Department of Mathematics, College of Science, Tianjin Polytechnic University, Tianjin, 300387, China*

Received 8 May 2018; received in revised form 10 July 2018; accepted 17 July 2018
Available online 29 July 2018

Abstract

In the paper, the authors discover several series identities involving the Catalan numbers, the Catalan function, the Riemannian zeta function, and the alternative Hurwitz zeta function.

Keywords: Series identity; Catalan number; Catalan function; Riemannian zeta function; Alternative Hurwitz zeta function; Digamma function

1. Introduction and main results

It is well known that the Catalan numbers C_n for $n \ge 0$ form a sequence of natural numbers that occur in tree enumeration problems such as "In how many ways can a regular n-gon be divided into $n-2$ triangles if different orientations are counted separately?" whose solution is the Catalan number C_{n-2}. The Catalan numbers C_n can be generated by

$$\frac{2}{1+\sqrt{1-4x}} = \frac{1-\sqrt{1-4x}}{2x} = \sum_{n=0}^{\infty} C_n x^n = 1 + x + 2x^2 + 5x^3 + \cdots .$$

Three of explicit formulas of C_n for $n \ge 0$ read that

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{4^n\,\Gamma\left(n+\frac{1}{2}\right)}{\sqrt{\pi}\,\Gamma(n+2)} = {}_2F_1(1-n, -n; 2; 1), \tag{1.1}$$

* Corresponding author at: Department of Mathematics, College of Science, Tianjin Polytechnic University, Tianjin, 300387, China.
E-mail addresses: yinli7979@163.com (L. Yin), qifeng618@gmail.com, qifeng618@hotmail.com, qifeng618@qq.com (F. Qi).
URL: http://qifeng618.wordpress.com (F. Qi).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.07.001

where $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,\mathrm{d}\,t$ for $\Re(z) > 0$ is the classical Euler gamma function,

$$ {}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; z) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_p)_n} \frac{z^n}{n!} $$

is the generalized hypergeometric series defined for $a_i \in \mathbb{C}$, $b_i \in \mathbb{C} \setminus \{0, -1, -2, \dots\}$, and $p, q \in \mathbb{N}$, and

$$ (x)_n = \prod_{\ell=0}^{n-1} (x + \ell) = \begin{cases} x(x+1)\cdots(x+n-1), & n \ge 1 \\ 1, & n \ge 0 \end{cases} $$

and $(-x)_n = (-1)^n (x - n + 1)_n$.

In 2014, Beckwith and Harbor [1] proposed a problem: show that

$$ \sum_{n=0}^{\infty} \frac{2^n}{C_n} = 5 + \frac{3}{2}\pi \quad \text{and} \quad \sum_{n=0}^{\infty} \frac{3^n}{C_n} = 22 + 8\sqrt{3}\,\pi. $$

This problem was answered in [1,2] by

$$ \sum_{n=0}^{\infty} \frac{x^n}{C_n} = 1 - \frac{x(x-10)}{(4-x)^2} + \frac{24\sqrt{x}}{(4-x)^{5/2}} \arctan\sqrt{\frac{x}{4-x}}, \quad 0 \le x < 4. \tag{1.2} $$

The editorial comment in [2] listed the formulas

$$ \sum_{n=0}^{\infty} \frac{1}{C_n} = 2 + \frac{4\pi}{9\sqrt{3}}, \qquad \sum_{n=0}^{\infty} \frac{(-1)^n}{C_n} = \frac{14}{25} - \frac{24\sqrt{5}}{125} \ln\frac{1+\sqrt{5}}{2}, $$

$$ \sum_{n=0}^{\infty} \frac{(-2)^n}{C_n} = \frac{1}{3} - \frac{1}{3\sqrt{3}} \ln\bigl(2+\sqrt{3}\bigr), \qquad \sum_{n=0}^{\infty} \frac{(-3)^n}{C_n} = \frac{10}{49} - \frac{36}{49\sqrt{21}} \ln\frac{5+\sqrt{21}}{2}. $$

The editorial comment in [2] pointed out that the result

$$ \sum_{n=0}^{\infty} \frac{x^n}{C_n} = 2\frac{\sqrt{4-x}\,(8+x) + 12\sqrt{x}\,\arctan\frac{\sqrt{x}}{\sqrt{4-x}}}{\sqrt{(4-x)^5}} \tag{1.3} $$

can be found on the website http://planetmath.org/ and that the problem can be solved easily from

$$ \sum_{n=1}^{\infty} \frac{2^n}{\binom{2n}{n}} = \frac{\pi}{2} + 1, \qquad \sum_{n=1}^{\infty} \frac{n2^n}{\binom{2n}{n}} = \pi + 3, \qquad \sum_{n=1}^{\infty} \frac{3^n}{\binom{2n}{n}} = \frac{4\pi\sqrt{3}}{3} + 3, \qquad \sum_{n=1}^{\infty} \frac{n3^n}{\binom{2n}{n}} = \frac{20\pi\sqrt{3}}{3} + 18 $$

which are special cases of the general formula

$$ \sum_{m=1}^{\infty} \frac{(2x)^{2m}}{m\binom{2m}{m}} = \sum_{m=1}^{\infty} \frac{(2x)^{2m}}{m(m+1)C_m} = \frac{2x \arcsin x}{\sqrt{1-x^2}}, \quad |x| < 1 $$

in [3, p. 452, Theorem]. Koshy and Gao [4] obtained

$$ \sum_{n=0}^{\infty} \frac{x^n}{C_n} = \begin{cases} 1 + \dfrac{x(4-x)^{3/2} + 6x(4-x)^{1/2} + 24\sqrt{x}\,\arcsin\frac{\sqrt{x}}{2}}{(4-x)^{5/2}}, & 0 \le x < 4; \\ 1 - \dfrac{|x|(4-x)^{3/2} + 6\sqrt{|x|(4-x)} + 24\sqrt{|x|}\,\ln\frac{\sqrt{-x}+\sqrt{4-x}}{2}}{(4-x)^{5/2}}, & -4 < x \le 0. \end{cases} \tag{1.4} $$

In 2016, motivated by Problem 11765 in [1] mentioned above, Amdeberhan and his four coauthors [5] proposed a general problem: find a closed-form formula for the series $\sum_{n=0}^{\infty} \frac{z^n}{C_n}$. They obtained in [5] that

$$ \sum_{n=0}^{\infty} \frac{z^n}{C_n} = {}_2F_1\left(1, 2; \frac{1}{2}; \frac{z}{4}\right) = \frac{2(z+8)}{(4-z)^2} + \frac{24\sqrt{z}}{(4-z)^{5/2}} \arcsin\frac{\sqrt{z}}{2} \tag{1.5} $$

for $|z| < 4$ by several methods.

It is clear that the formulas (1.2) and (1.3) are the same one and that the formulas (1.4) and (1.5) are also the same one. Since $\arctan\sqrt{\frac{x}{4-x}} = \arcsin\frac{\sqrt{x}}{2}$ for $0 \le x < 4$, the four formulas (1.2) to (1.5) are essentially the same one. It seems that there are close and similar ideas in [1,5] and that the paper [5] is almost an expanded version of [1]. Great minds think alike! For more detailed review, please refer to the survey article [6] and closely related references therein.

The Catalan numbers C_n have a long history [7–11] and have been generalized and developed [6,12–27] in recent years.

In this paper, we will discover several series identities involving the Catalan numbers C_n, the Catalan function $C_x = \frac{4^x\,\Gamma\left(x+\frac{1}{2}\right)}{\sqrt{\pi}\,\Gamma(x+2)}$ for $x \ge 0$, the Riemannian zeta function $\zeta(s) = \sum_{k=1}^{\infty}\frac{1}{k^s}$ for $\Re(s) > 1$, and the alternative Hurwitz zeta function $\zeta_a(s,q) = \sum_{n=0}^{\infty}\frac{(-1)^n}{(q+n)^s}$ for $\Re(s) > 1$ and $\Re(q) > 0$.

Our main results can be stated as the following theorems.

Theorem 1.1. *For $\lambda > 0$, we have*

$$\int_0^1 x^\lambda \ln C_x \,\mathrm{d}\,x = \frac{1}{\lambda+1}\left[\zeta_a(1,\lambda+2) - \sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{k=0}^{\infty}\frac{(-1)^k}{n^k}\frac{2^{k+2}-2}{\lambda+k+3}\right]. \tag{1.6}$$

Theorem 1.2. *For $\lambda < \frac{1}{2}$, we have*

$$\sum_{n=0}^{\infty}\frac{C_n}{4^n} = \sum_{n=0}^{\infty}\frac{1}{n!(1-\lambda)^{n+1/2}}\left(\frac{1}{2}\right)_n\sum_{j=0}^{n}\frac{(-\lambda)^{n-j}}{j+1}\binom{n}{j} = 2. \tag{1.7}$$

Theorem 1.3. *For $x \in (-4, 0]$, we have*

$$\sum_{n=0}^{\infty}\frac{x^n}{C_n} = -\frac{24\sqrt{-x}}{(4-x)^{5/2}}\ln\left(\frac{\sqrt{-x}+\sqrt{4-x}}{2}\right) + \frac{2(8+x)}{(4-x)^2}. \tag{1.8}$$

Theorem 1.4. *For $\lambda < \frac{1}{2}$, we have*

$$C_n = \frac{2^{2n+5}}{\pi}\sum_{m=0}^{\infty}\frac{(-2n)_m}{m!(1-\lambda)^{m-2n}}\sum_{j=0}^{m}(-\lambda)^{m-j}\binom{m}{j}\sum_{k=0}^{\infty}\binom{-2n-3}{k}\frac{1}{j+2k+3}. \tag{1.9}$$

Theorem 1.5. *The zeta function $\zeta(z)$ satisfies*

$$\zeta(3) = \frac{8}{7}\sum_{n=0}^{\infty}\frac{(n+1)(2n)!!C_n}{4^n(2n+1)^2(2n+1)!!}. \tag{1.10}$$

2. Lemmas

In order to prove our main results, we need the following lemmas.

Lemma 2.1 (*[28, Lemma 2.1] and [29, p. 138, 5.5.8]*)**.** *The function $\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is strictly concave on $(0,\infty)$ and satisfies the duplication formula*

$$\psi(2x) = \frac{1}{2}\psi(x) + \frac{1}{2}\psi\left(x+\frac{1}{2}\right) + \ln 2.$$

Lemma 2.2. *For $x, k \in \mathbb{R}$ and $\lambda > 0$, we have*

$$\frac{x^{\lambda+1}}{x+k} = \sum_{j=0}^{\lambda}(-1)^j k^j x^{\lambda-j} + (-1)^{\lambda+1}\frac{k^{\lambda+1}}{x+k}. \tag{2.1}$$

Proof. A simple computation yields

$$\sum_{j=0}^{\lambda}(-1)^j k^{j-1}x^{\lambda-j}+(-1)^{\lambda+1}\frac{k^\lambda}{x+k}=\frac{x^\lambda}{k}\sum_{j=0}^{\lambda}(-1)^j\left(\frac{k}{x}\right)^j+(-1)^{\lambda+1}\frac{k^\lambda}{x+k}$$
$$=\frac{x^\lambda}{k}\frac{1-(-1)^{\lambda+1}\left(\frac{k}{x}\right)^{\lambda+1}}{1+\frac{k}{x}}+(-1)^{\lambda+1}\frac{k^\lambda}{x+k}=\frac{x^{\lambda+1}}{k(x+k)}.$$

The required proof is complete. □

Remark 2.1. Taking $\lambda=1$ in the identity (2.1) leads to $\frac{k^2}{x+k}=k-x+\frac{x^2}{x+k}$ which can be found in [30, p. 96].

Lemma 2.3 *([31, Theorem (The λ-method)]). For $0<\alpha\le 1$, $\eta>0$, $\lambda<\frac{1}{2}$, $0\le j\in\mathbb{N}$, and $\xi\in\mathbb{R}$, suppose that the function G given by $G(x)=\frac{g(x)}{(1-\alpha x^\eta)^\xi}$ satisfies $g, G\in L^1[0,1]$ and that*

$$b_j=b_j(\alpha,\eta)=\alpha^j\int_0^1 t^{j\eta}g(t)\,\mathrm{d}\,t,$$

then

$$\int_0^1\frac{g(x)}{(1-\alpha x^\eta)^\xi}\,\mathrm{d}\,x=\sum_{n=0}^{\infty}\frac{(\xi)_n}{n!(1-\lambda)^{n+\xi}}\sum_{j=0}^{n}\binom{n}{j}(-\lambda)^{n-j}b_j(\alpha,\eta).$$

3. Proofs of main results

We are now in a position to prove our main results.

Proof of Theorem 1.1. Integrating by parts gives

$$\int_0^1 x^\lambda\ln C_x\,\mathrm{d}\,x=\frac{1}{\lambda+1}\int_0^1\ln C_x\,\mathrm{d}\,x^{\lambda+1}=-\frac{1}{\lambda+1}\int_0^1 x^{\lambda+1}\frac{C_x'}{C_x}\,\mathrm{d}\,x.$$

On the other hand, applying Lemma 2.1 leads to

$$\frac{C_x'}{C_x}=2\ln 2+\psi\left(x+\frac{1}{2}\right)-\psi(x+2)=2\psi(2x)-2\psi(x)-\frac{1}{x}-\frac{1}{x+1}. \tag{3.1}$$

Multiplying by $x^{\lambda+1}$ and integrating from 0 to 1 on both sides of (3.1) result in

$$\int_0^1 x^{\lambda+1}\frac{C_x'}{C_x}\,\mathrm{d}\,x=2\int_0^1 x^{\lambda+1}\psi(2x)\,\mathrm{d}\,x-2\int_0^1 x^{\lambda+1}\psi(x)\,\mathrm{d}\,x-\int_0^1 x^\lambda\,\mathrm{d}\,x-\int_0^1\frac{x^{\lambda+1}}{x+1}\,\mathrm{d}\,x.$$

Using the representation

$$\psi(x)=-\gamma-\frac{1}{x}+\sum_{n=1}^{\infty}\frac{x}{n(n+x)}$$

in [29, p. 139, 5.7.6], where γ is the Euler–Mascheroni constant, arrives at

$$\begin{aligned}\int_0^1 x^{\lambda+1}\psi(x)\,\mathrm{d}\,x&=-\gamma\int_0^1 x^{\lambda+1}\,\mathrm{d}\,x-\int_0^1 x^\lambda\,\mathrm{d}\,x+\sum_{n=1}^{\infty}\int_0^1\frac{x^{\lambda+2}}{n(n+x)}\,\mathrm{d}\,x\\&=-\frac{\gamma}{\lambda+2}-\frac{1}{\lambda+1}+\sum_{n=1}^{\infty}\frac{1}{n^2}\int_0^1\frac{x^{\lambda+2}}{1+x/n}\,\mathrm{d}\,x\\&=-\frac{\gamma}{\lambda+2}-\frac{1}{\lambda+1}+\sum_{n=1}^{\infty}\frac{1}{n^2}\int_0^1 x^{\lambda+2}\sum_{k=0}^{\infty}(-1)^k\left(\frac{x}{n}\right)^k\,\mathrm{d}\,x\\&=-\frac{\gamma}{\lambda+2}-\frac{1}{\lambda+1}+\sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{k=0}^{\infty}\frac{(-1)^k}{n^k}\frac{1}{\lambda+k+3}.\end{aligned} \tag{3.2}$$

Similar method results in

$$\begin{aligned}\int_0^1 x^{\lambda+1}\psi(2x)\,\mathrm{d}\,x &= \frac{1}{2^{\lambda+2}}\int_0^2 x^{\lambda+1}\psi(x)\,\mathrm{d}\,x\\ &= -\frac{\gamma}{\lambda+2}-\frac{1}{2(\lambda+1)}+\frac{1}{2^{\lambda+1}}\sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{k=0}^{\infty}\frac{(-1)^k}{n^k}\frac{2^{\lambda+k+3}}{\lambda+k+3}\end{aligned} \tag{3.3}$$

and

$$\int_0^1 \frac{x^{\lambda+1}}{x+1}\,\mathrm{d}\,x = \int_0^1 x^{\lambda+1}\sum_{k=0}^{\infty}(-1)^k x^k\,\mathrm{d}\,x = \sum_{k=0}^{\infty}\frac{(-1)^k}{\lambda+k+2}. \tag{3.4}$$

Combining (3.2) and (3.3) with (3.4) reveals

$$\int_0^1 x^{\lambda+1}\frac{C_x'}{C_x}\,\mathrm{d}\,x = \sum_{n=1}^{\infty}\frac{1}{n^2}\sum_{k=0}^{\infty}\frac{(-1)^k}{n^k}\frac{2^{k+2}-2}{\lambda+k+3}-\zeta_a(1,\lambda+2).$$

The proof of the formula (1.6) is complete. □

Remark 3.1. Taking $\lambda=0$ and using the identity

$$\int_0^1 \ln\Gamma(x+a)\,\mathrm{d}\,x = \frac{1}{2}\ln(2\pi)+a\ln a-a$$

in [32, p. 124, (43a)], we easily obtain

$$\int_0^1 \ln C_x\,\mathrm{d}\,x = -\frac{3}{2}\ln 2-\ln\sqrt{\pi}+\frac{3}{2}.$$

Proof of Theorem 1.2. Using the Maclaurin series

$$\frac{1}{\sqrt{1-4x}} = \sum_{n=0}^{\infty}\binom{2n}{n}x^n$$

and substituting $4x=t$ give

$$\frac{1}{\sqrt{1-t}} = \sum_{n=0}^{\infty}\binom{2n}{n}\frac{t^n}{4^n}. \tag{3.5}$$

Taking $\alpha=1$, $\xi=\frac{1}{2}$, $\eta=1$, and $g(x)=1$ in Lemma 2.3 and integrating on both sides of (3.5) from 0 to 1 arrive at the formula (1.7). The proof of Theorem 1.2 is complete. □

Proof of Theorem 1.3. It is easy to verify that the series converges for $|x|<4$.

For $x\in(-4,0]$, a simple computation yields

$$\sum_{n=0}^{\infty}\frac{x^n}{C_n} = \sum_{n=0}^{\infty}\frac{n!(n+1)!}{(2n)!}x^n = 1+\sum_{n=1}^{\infty}\frac{(n+1)!}{2^n(2n-1)!!}x^n = 1+\frac{1}{2}\sum_{n=1}^{\infty}\frac{(2n+2)!!}{(2n-1)!!}\left(\frac{x}{4}\right)^n.$$

Denoting

$$f(x) = 1+\frac{1}{2}\sum_{n=1}^{\infty}\frac{(2n+2)!!}{(2n-1)!!}\left(\frac{x}{4}\right)^n$$

and substitution $x=-4t$ for $t>0$ produce

$$2[f(-4t)-1] = \sum_{n=1}^{\infty}(-1)^n\frac{(2n+2)!!}{(2n-1)!!}t^n \triangleq g(t).$$

Integrating twice on both sides of

$$\sum_{n=1}^{\infty}\frac{(2n+2)!!}{(2n-1)!!}(-1)^n t^{2n} = g\left(t^2\right)$$

from 0 to t shows

$$h(t) \triangleq \int_0^t g\left(t^2\right)\mathrm{d}\,t = \sum_{n=1}^{\infty}\frac{(2n+2)!!}{(2n+1)!!}(-1)^n t^{2n+1}$$

and

$$\alpha(t) \triangleq \int_0^t h(t)\,\mathrm{d}\,t = \sum_{n=1}^{\infty}\frac{(2n)!!}{(2n+1)!!}(-1)^n t^{2n+2}.$$

Using the well-known Maclaurin series

$$\frac{\ln(x+\sqrt{1+x^2}\,)}{\sqrt{1+x^2}} = \sum_{n=0}^{\infty}(-1)^n\frac{(2n)!!}{(2n+1)!!}x^{2n+1}, \quad x\in[-1,1]$$

in [33, p. 292] gives

$$\alpha(t) = \frac{t\ln\left(t+\sqrt{1+t^2}\right)}{\sqrt{1+t^2}} - t^2,$$

$$h(t) = \frac{\mathrm{d}\,\alpha(t)}{\mathrm{d}\,t} = \frac{t\sqrt{1+t^2}+\ln\left(t+\sqrt{1+t^2}\right)}{\sqrt{(1+t^2)^3}} - 2t,$$

$$g\left(t^2\right) = \frac{\mathrm{d}\,h(t)}{\mathrm{d}\,t} = -\frac{3t\ln\left(t+\sqrt{1+t^2}\right)}{\sqrt{(1+t^2)^5}} - \frac{t^2-2}{(1+t^2)^2} - 2,$$

and

$$f(x) = \frac{1}{2}\left[g\left(-\frac{x}{4}\right)\right] + 1.$$

The formula (1.8) is thus proved. The proof of Theorem 1.3 is complete. □

Proof of Theorem 1.4. Using the integral representation

$$C_n = \frac{2^{2n+5}}{\pi}\int_0^1 \frac{x^2\left(1-x^2\right)^{2n}}{(1+x^2)^{2n+3}}\,\mathrm{d}\,x$$

in [34, p. 10] and letting $\alpha = 1$, $\xi = -2n$, $\eta = 2$, $g(x) = \frac{x^2}{(1+x^2)^{2n+3}}$ in Lemma 2.3 produce

$$\begin{aligned} b_j &= \int_0^1 \frac{t^{j+2}}{(1+t^2)^{2n+3}}\,\mathrm{d}\,x = \frac{1}{2}\int_0^1 u^{(j+1)/2}(1+u)^{-2n-3}\,\mathrm{d}\,u \\ &= \sum_{k=0}^{\infty}\binom{-2n-3}{k}\frac{1}{2}\int_0^1 u^{(j+1+2k)/2}\,\mathrm{d}\,u = \sum_{k=0}^{\infty}\binom{-2n-3}{k}\frac{1}{j+2k+3}. \end{aligned}$$

By virtue of the substitution $t^2 = u$, the formula (1.9) follows immediately. The proof of Theorem 1.4 is complete. □

Proof of Theorem 1.5. Using the Fourier series

$$x = \sum_{n=0}^{\infty}\frac{1}{4^n}\binom{2n}{n}\frac{\sin^{2n+1}x}{2n+1}, \tag{3.6}$$

multiplying (3.7) by $x\cot x$, and integrating over the interval $\left[0,\frac{\pi}{2}\right]$ reveals

$$\int_0^{\pi/2} x^2\cot x\,\mathrm{d}\,x = \sum_{n=0}^{\infty}\frac{(n+1)C_n}{4^n(2n+1)}\int_0^{\pi/2} x\sin^{2n+1}x\cot x\,\mathrm{d}\,x$$

Applying the identity

$$\int_0^{\pi/2} t\cos^{p-1}t\sin at\,\mathrm{d}\,t=\frac{\pi}{2^{p+1}}\Gamma(p)\frac{\psi((p+a+1)/2)-\psi((p-a+1)/2)}{\Gamma((p+a+1)/2)\Gamma((p-a+1)/2)}$$

with the constraints $p>0$ and $|a|<p+1$ in [35, p. 26, (3,8a)] gives

$$\begin{aligned}\int_0^{\pi/2} x\sin^{2n+1}x\cot x\,\mathrm{d}\,x &\overset{x=\frac{\pi}{2}-t}{=\!=} \int_0^{\pi/2}\left(\frac{\pi}{2}-t\right)\cos^{2n}t\sin t\,\mathrm{d}\,t\\ &=\frac{\pi}{2(2n+1)}-\frac{\pi}{2^{2n+2}}\Gamma(2n+1)\frac{\psi((2n+3)/2)-\psi((2n+1)/2)}{\Gamma((2n+3)/2)\Gamma((2n+1)/2)}\\ &=\frac{\pi}{2(2n+1)}-\frac{1}{2n+1}\frac{(2n)!!}{(2n+1)!!}.\end{aligned}$$

On the other hand, applying the identity

$$\int_0^{\pi/2} x\ln\sin x\,\mathrm{d}\,x=\frac{7}{16}\zeta(3)-\frac{\pi^2}{8}\ln 2$$

in [32, p. 144, (7.16)] gives

$$\int_0^{\pi/2} x^2\cot x\,\mathrm{d}\,x=-2\int_0^{\pi/2} x\ln\sin x\,\mathrm{d}\,x=-\frac{7}{8}\zeta(3)+\frac{\pi^2}{4}\ln 2.$$

To prove the formula (3.7), we just need to prove

$$\sum_{n=0}^{\infty}\frac{(n+1)C_n}{4^n(2n+1)^2}=\frac{\pi}{2}\ln 2.$$

In fact, multiplying (3.6) by $\cot x$ results in

$$\sum_{n=0}^{\infty}\frac{(n+1)C_n}{4^n(2n+1)^2}=\int_0^{\pi/2} x\cot x\,\mathrm{d}\,x=-\int_0^{\pi/2}\ln\sin x\,\mathrm{d}\,x=\frac{\pi}{2}\ln 2.$$

The proof of Theorem 1.5 is complete. □

Remark 3.2. There are some results and applications of properties for the Riemannian zeta function $\zeta(s)$ in the papers [36–38] and the closely related references therein.

Remark 3.3. On 9 July 2018, Boyadzhiev recommend his paper [39] on the ResearchGate to the authors. In [39, Theorem 1], the generating function for the numbers $\frac{H_n}{\binom{2n}{n}}$ was computed explicitly in terms of elementary functions and dilogarithms, where H_n denotes harmonic numbers [40]. In the proof of [39, Theorem 1], the sums

$$\sum_{n=0}^{\infty}(-1)^n\frac{(4z)^n}{\binom{2n}{n}}=\frac{1}{1+z}\left(1+\frac{1}{2}\sqrt{\frac{z}{1+z}}\,\ln\frac{1-\sqrt{\frac{z}{1+z}}}{1+\sqrt{\frac{z}{1+z}}}\right)$$

and

$$\sum_{n=0}^{\infty}\frac{(4z)^n}{\binom{2n}{n}}=\frac{1}{1-z}\left(1+\sqrt{\frac{z}{1-z}}\,\arctan\sqrt{\frac{z}{1-z}}\right)$$

in [41, Theorem 2.1] play important roles. From the first relation in (1.1) between the Catalan numbers C_n and central binomial coefficients $\binom{2n}{n}$, we can obtain

$$\sum_{n=0}^{\infty}(-1)^n\frac{(4z)^n}{\binom{2n}{n}}=\sum_{n=0}^{\infty}(-1)^n\frac{(4z)^n}{(n+1)C_n}=\frac{1}{4z}\sum_{n=0}^{\infty}(-1)^n\frac{(4z)^{n+1}}{(n+1)C_n}$$

and

$$\sum_{n=0}^{\infty}\frac{(4z)^n}{\binom{2n}{n}}=\sum_{n=0}^{\infty}\frac{(4z)^n}{(n+1)C_n}=\frac{1}{4z}\sum_{n=0}^{\infty}\frac{(4z)^{n+1}}{(n+1)C_n}.$$

Since

$$\frac{\mathrm{d}}{\mathrm{d}\,z}\left[\sum_{n=0}^{\infty}(-1)^n\frac{(4z)^{n+1}}{(n+1)C_n}\right]=4\sum_{n=0}^{\infty}(-1)^n\frac{(4z)^n}{C_n}\quad\text{and}\quad\frac{\mathrm{d}}{\mathrm{d}\,z}\left[\sum_{n=0}^{\infty}\frac{(4z)^{n+1}}{(n+1)C_n}\right]=4\sum_{n=0}^{\infty}\frac{(4z)^n}{C_n},$$

we can connect some results in this paper with those in the papers [6,13,39,42,43] and closely related references therein.

Remark 3.4. In [39], Boyadzhiev mentioned that Apéry used the representation

$$\zeta(3)=\frac{5}{2}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^3}\frac{1}{\binom{2n}{n}}=\frac{5}{2}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^3}\frac{1}{(n+1)C_n}$$

to show that $\zeta(3)$ is an irrational number. Does this imply that the formula (3.7), which can be rewritten as

$$\zeta(3)=\frac{8}{7}\sum_{n=0}^{\infty}\frac{n!}{(2n+1)^2(2n+1)!!}\binom{2n}{n},\tag{3.7}$$

in our Theorem 1.5 can also be used to show the irrationality of $\zeta(3)$?

Remark 3.5. This paper is a corrected and extended version of the preprint [44].

Acknowledgments

The first author was partially supported by the National Nature Science Foundation of China under Grant Number 11401041, Science Foundation of Binzhou University under Grant Numbers BZXYL1704; BZXYL1104, and the Science and Technology Foundation of Shandong Province under Grant Number J16LI52, China.

The authors appreciate anonymous referees for their careful corrections to and valuable comments on the original version of this paper.

References

[1] D. Beckwith, S. Harbor, Problem 11765, Amer. Math. Monthly 121 (3) (2014) 267–267. Available online at http://dx.doi.org/10.4169/amer.math.monthly.121.03.266.
[2] U. Abel, Reciprocal Catalan sums: Solution to Problem 11765, Amer. Math. Monthly 123 (4) (2016) 405–406. Available online at http://dx.doi.org/10.4169/amer.math.monthly.123.4.399.
[3] D.H. Lehmer, Interesting series involving the central binomial coefficient, Amer. Math. Monthly 92 (7) (1985) 449–457. Available online at http://dx.doi.org/10.2307/2322496.
[4] T. Koshy, Z.-G. Gao, Convergence of a Catalan series, College Math. J. 43 (2) (2012) 141–146. Available online at http://dx.doi.org/10.4169/college.math.j.43.2.141.
[5] T. Amdeberhan, X. Guan, L. Jiu, V.H. Moll, C. Vignat, A series involving Catalan numbers: Proofs and demonstrations, Elem. Math. 71 (3) (2016) 109–121. Available online at http://dx.doi.org/10.4171/EM/306.
[6] F. Qi, B.-N. Guo, Integral representations of the Catalan numbers and their applications, Mathematics 5 (3) (2017) 31. Article 40. Available online at http://dx.doi.org/10.3390/math5030040.
[7] T. Koshy, Catalan Numbers with Applications, Oxford University Press, Oxford, 2009.
[8] P. Larcombe, On the history of the Catalan numbers: a first record in China, Math. Today (Southend-on-Sea) 35 (3) (1999) 89–89.
[9] P.J. Larcombe, The 18th century Chinese discovery of the Catalan numbers, Math. Spectrum 32 (1) (1999/2000) 5–7.
[10] J.J. Luo, Ming Antu, the first discoverer of the Catalan numbers, Neimenggu Daxue Xuebao 19 (2) (1988) 239–245. (in Chinese).
[11] R.P. Stanley, Catalan Numbers, Cambridge University Press, New York, 2015. Available online at http://dx.doi.org/10.1017/CBO9781139871495.
[12] M. Mahmoud, F. Qi, Three identities of the Catalan–Qi numbers, Mathematics 4 (2) (2016) 7. Article 35. Available online at http://dx.doi.org/10.3390/math4020035.
[13] F. Qi, Parametric integrals, the Catalan numbers, and the beta function, Elem. Math. 72 (3) (2017) 103–110. Available online at http://dx.doi.org/10.4171/EM/332.
[14] F. Qi, A. Akkurt, H. Yildirim, Catalan numbers, k-gamma and k-beta functions, and parametric integrals, J. Comput. Anal. Appl. 25 (6) (2018) 1036–1042.
[15] F. Qi, P. Cerone, Some properties of the Fuss–Catalan numbers, Preprints 2017, 2017080056, 14 pages. Available online at http://dx.doi.org/10.20944/preprints201708.0056.v1.
[16] F. Qi, B.-N. Guo, Logarithmically complete monotonicity of a function related to the Catalan–Qi function, Acta Univ. Sapientiae Math. 8 (1) (2016) 93–102. Available online at http://dx.doi.org/10.1515/ausm-2016-0006.

[17] F. Qi, B.-N. Guo, Logarithmically complete monotonicity of Catalan–Qi function related to Catalan numbers, Cogent Math. 3 (2016) 1179379. 6 pages. Available online at http://dx.doi.org/10.1080/23311835.2016.1179379.
[18] F. Qi, B.-N. Guo, Some properties and generalizations of the Catalan, Fuss, and Fuss–Catalan numbers, in: Michael Ruzhansky, Hemen Dutta, Ravi P. Agarwal (Eds.), Mathematical Analysis and Applications: Selected Topics, first ed., John Wiley & Sons, Inc, 2018, pp. 101–133 Chapter 5.
[19] F. Qi, M. Mahmoud, X.-T. Shi, F.-F. Liu, Some properties of the Catalan–Qi function related to the Catalan numbers, SpringerPlus 5 (2016) 1126. 20 pages. Available online at http://dx.doi.org/10.1186/s40064-016-2793-1.
[20] F. Qi, X.-T. Shi, F.-F. Liu, An integral representation, complete monotonicity, and inequalities of the Catalan numbers, Filoma 32 (2) (2018) 575–587. Available online at https://doi.org/10.2298/FIL1802575Q.
[21] F. Qi, X.-T. Shi, F.-F. Liu, D.V. Kruchinin, Several formulas for special values of the Bell polynomials of the second kind and applications, J. Appl. Anal. Comput. 7 (3) (2017) 857–871. Available online at http://dx.doi.org/10.11948/2017054.
[22] F. Qi, X.-T. Shi, M. Mahmoud, F.-F. Liu, Schur-convexity of the Catalan–Qi function related to the Catalan numbers, Tbilisi Math. J. 9 (2) (2016) 141–150. Available online at http://dx.doi.org/10.1515/tmj-2016-0026.
[23] F. Qi, X.-T. Shi, M. Mahmoud, F.-F. Liu, The Catalan numbers: a generalization, an exponential representation, and some properties, J. Comput. Anal. Appl. 23 (5) (2017) 937–944.
[24] F. Qi, Q. Zou, B.-N. Guo, Some identities and a matrix inverse related to the Chebyshev polynomials of the second kind and the Catalan numbers, Preprints 2017, 2017030209, 25 pages. Available online at http://dx.doi.org/10.20944/preprints201703.0209.v2.
[25] X.-T. Shi, F.-F. Liu, F. Qi, An integral representation of the Catalan numbers, Glob. J. Math. Anal. 3 (3) (2015) 130–133. Available online at http://dx.doi.org/10.14419/gjma.v3i3.5055.
[26] Q. Zou, Analogues of several identities and supercongruences for the Catalan–Qi numbers, J. Inequal. Spec. Funct. 7 (4) (2016) 235–241.
[27] Q. Zou, The q-binomial inverse formula and a recurrence relation for the q-Catalan–Qi numbers, J. Math. Anal. 8 (1) (2017) 176–182.
[28] H. Alzer, K.C. Richard, Inequalities for the ratio of complete elliptic integrals, Proc. Amer. Math. Soc. 145 (4) (2017) 1661–1670. Available online at http://dx.doi.org/10.1090/proc/13337.
[29] F.W.J. Olver, D.W. Lozier, R.F. Boisvert, C.W. Clark (Eds.), NIST Handbook of Mathematical Functions, Cambridge University Press, New York, 2010. Available online at http://dlmf.nist.gov/.
[30] D.F. Connon, Some series and integrals involving the Riemann zeta function, binomial coefficients and the harmonic numbers. Volume II(a), arXiv preprint, 2007. Available online at http://arxiv.org/abs/0710.4023.
[31] H. Alzer, K.C. Richard, Series representation for special functions and mathematical constants, Ramanujan J. 40 (2) (2016) 291–310. Available online at http://dx.doi.org/10.1007/s11139-015-9679-7.
[32] D.F. Connon, Some series and integrals involving the Riemann zeta function, binomial coefficients and the harmonic numbers. Volume V, arXiv preprint, 2007. Available online at http://arxiv.org/abs/0710.4047.
[33] C.-H. Xue, S.-L. Xu, Full Solution for Selected Exercises of Mathematical Analysis, Tsinghua University Press, Beijing, China, 2009. (in Chinese).
[34] A. Nkwanta, A. Tefera, Curious relations and identities involving the Catalan generating function and numbers, J. Integer Seq. 16 (9) (2013) 15. Article 13.9.5.
[35] D.F. Connon, Some series and integrals involving the Riemann zeta function, binomial coefficients and the harmonic numbers. Volume I, arXiv preprint, 2007. Available online at http://arxiv.org/abs/0710.4022.
[36] Q.-M. Luo, B.-N. Guo, F. Qi, On evaluation of Riemann zeta function $\zeta(s)$, Adv. Stud. Contemp. Math. (Kyungshang) 7 (2) (2003) 135–144.
[37] Q.-M. Luo, Z.-L. Wei, F. Qi, Lower and upper bounds of $\zeta(3)$, Adv. Stud. Contemp. Math. (Kyungshang) 6 (1) (2003) 47–51.
[38] F. Qi, A double inequality for the ratio of two consecutive Bernoulli numbers, Preprints 2017, 2017080099, 7 pages. Available online at https://doi.org/10.20944/preprints201708.0099.v1.
[39] K.N. Boyadzhiev, Power series with inverse binomial coefficients and harmonic numbers, Tatra Mt. Math. Publ. 70 (2017) 199–206.
[40] B.-N. Guo, F. Qi, Sharp bounds for harmonic numbers, Appl. Math. Comput. 218 (3) (2011) 991–995. Available online at http://dx.doi.org/10.1016/j.amc.2011.01.089.
[41] R. Sprugnoli, Sums of reciprocals of the central binomial coefficients, Integers 6 (A27) (2006) 18.
[42] K.N. Boyadzhiev, Series with central binomial coefficients, Catalan numbers, and harmonic numbers, J. Integer Seq. 15 (1) (2012) 11. Article 12.1.7.
[43] H. Chen, Interesting series associated with central binomial coefficients, Catalan numbers and harmonic numbers, J. Integer Seq. 19 (1) (2016) 11. Article 16.1.5.
[44] L. Yin, F. Qi, Several series identities involving the Catalan numbers, Preprints 2017, 2017030029, 11 pages. Available online at http://dx.doi.org/10.20944/preprints201703.0029.v1.

Available online at www.sciencedirect.com
ScienceDirect
Transactions of A. Razmadze Mathematical Institute 172 (2018) 475–479
Transactions of A. Razmadze Mathematical Institute
www.elsevier.com/locate/trmi

Original article

Linear criteria for hypotheses testing

Z. Zerakidze*, M. Mumladze

Gory State University, Georgia

Received 29 May 2018; received in revised form 25 July 2018; accepted 28 July 2018
Available online 24 September 2018

Abstract

In the present paper we prove necessary and sufficient conditions for the existence of linear consistent criteria.

Keywords: Statistical structure; Linear consistent criteria; Weakly and strongly linear criteria

1. Introduction

When considering statistical problems, the main objects, on which support done statistical conclusions are statistical criteria. In problems of statistics of random processes the sample space is infinite, therefore it plays an important role in the classes of measurable mappings (criteria) given constructively.

Among problems of statistic of random processes is separated the class of problems in which number of observations knowingly limited (then it can be considered a single).

Despite the uniqueness of observation, in many cases we can reliably choose one of the sets of competing hypotheses of exact type of the distribution.

In case, when a hypothesis is the hypothesis determined by one observation they say that for hypothesis exist consistent criteria of testing. To existence of linear criteria for hypotheses testing is devoted to this article.

The theory of linear criteria for hypotheses testing has wide practical application, because it uses minimal information about random process: an average value and correlation function.

2. Linear criteria

Let (X, B) be separable Hilbert space with σ-algebra of Borel sets in X and $\{\mu_h,\ h \in H \subset X\}$ the family of probability measures on separable Hilbert space.

* Corresponding author.
E-mail addresses: zura.zerakidze@mail.ru (Z. Zerakidze), mmumladze@mail.ru (M. Mumladze).
Peer review under responsibility of Journal Transactions of A. Razmadze Mathematical Institute.

https://doi.org/10.1016/j.trmi.2018.07.004

Let a_h be average value:

$$(a_h, z) = \int_X (z, x)\mu_h(dx),$$

for all $z \in X$, B_h is correlation function:

$$(B_h z, u) = \int_X (z, x)(u, x)\mu_h(dx),$$

for all $z, u \in X$, where (z, u) means scalar product in X.

Suppose $B_h = B$ is independent of h. We recall the definitions (see [1]). We assume that the hypotheses are average values:

$$(a_h, z) = \int_X (z, x)\mu_h(dx).$$

In linear theory it is assumed that H is linear manifold. As a criterion we will consider only linear maps from X to H.

Definition 1. An object $\{X, \mathrm{B}, \mu_h, h \in H\}$ is called a statistical structure.

Definition 2. The map $g : X \to X$ for which there exists such sequence of bounded linear maps $g_n : X \to X$, that

$$\lim_{n\to\infty} \int_X (z, g(x) - g_n(x))^2 \mu_h(dx) = 0$$

for all $z \in X$, $h \in H$ is called weakly measurable linear with respect to statistical structure $\{X, \mathrm{B}, \mu_h, h \in H\}$.

Definition 3. The map $g : X \to X$ for which there exists such sequence of bounded linear maps $g_n : X \to X$ that

$$\lim_{n\to\infty} \int_X (g_n(x) - g(x))^2 \mu_h(dx) = 0$$

for all $h \in H$ is called strongly measurable linear with respect to statistical structure $\{X, \mathrm{B}, \mu_h, h \in H\}$.

Definition 4. We will say that the statistical structure $\{X, \mathrm{B}, \mu_h, h \in H\}$ admits weakly sequential consistent linear criteria for hypotheses testing if there exists sequence of continuous linear maps $g_n : X \to X$ that

$$\lim_{n\to\infty} \int_X (z, g_n(x) - h)^2 \mu_h(dx) = 0$$

for all $z \in X, h \in H$.

Definition 5. We will say that the statistical structure $\{X, \mathrm{B}, \mu_h, h \in H\}$ admits strongly sequential consistent linear criteria for hypotheses testing if there exists sequence of continuous linear maps $g_n : X \to X$ that

$$\lim_{n\to\infty} \int_X |g_n(x) - h|^2 \mu_h(dx) = 0$$

for all $h \in H$.

Definition 6. Weakly measurable linear map $g : X \to X$ admits weakly consistent linear criteria for statistical structure $\{E, S, \mu_h, h \in H\}$, if $\mu_h(x : g(x) = h) = 1$ for all $h \in H$.

Definition 7. Strongly measurable linear map $g : X \to X$ admits strongly consistent linear criteria for statistical structure $\{E, S, \mu_h, h \in H\}$, if $\mu_h(x : g(x) = h) = 1$ for all $h \in H$.

Remark 1. It is clear that from consistency follows corresponding sequential consistency. The converse can be shown under additional assumptions on H.

Theorem 2.1. *Let on the set H introduce the norm $\|.\|_H$ so that H in this norm will be separable Banach space, at what functional (h, z) will be continuous in this norm for all $z \in H$ and if statistical structure $\{E, S, \mu_h, h \in H\}$ admits weakly (strongly) sequential consistent criteria, then this statistical structure admits weakly (strongly) consistent criteria.*

Proof. As statistical structure $\{E, S, \mu_h, h \in H\}$ admits weakly sequential consistent criteria then

$$\lim_{n\to\infty} \int_X (z, g_n(x) - h)^2 \mu_h(dx) = 0$$

for all $z \in E$. The value from under the integral sign has form

$$(z, g_n(h) - h)^2 + (g_n B g_n^* z, z),$$

where g_n^* is the linear map conjugate to g in X, $g_n B g_n^*$ the product of linear operators. If it tends to zero, then operator norm $\|g_n B g_n^*\|$ is bounded and

$$\lim_{n\to\infty} (g_n B g_n^* z, z) = 0, \quad z \in X.$$

Except this $\|g_n\|_H$ is bounded and for all $z \in X,\ h \in H$,

$$\lim_{n\to\infty} (z, g_n(h) - h)^2 = 0.$$

Choose sequence n_k so that for some countable, densely in X and H sets series

$$\sum_{k=1}^{\infty} [(z, g_k(h) - h)^2 + (g_{n_k} B g_{n_k}^* z, z)] < \infty$$

will be converged.

Let Z be countable and complete set in X for which holds equality

$$\lim_{n\to\infty} (z, g_n(h) - h)^2 = 0, \quad z \in Z,$$

when $h \in H_1$, where H_1 is the complete set in H. Then sequence $(g_{n_k}(x), z)$ converges for almost all x with respect to the measure μ_h when $z \in Z$. Denote this limit by $g(x, z)$. Because for all $z \in X,\ h \in H$ performed

$$\lim_{n\to\infty} (z, g_n(h) - h)^2 = 0,$$

then is determined value $g(h, z) = (h, z)$. So is determined also value

$$g(x - h, z) = \lim_{n_k\to\infty} (g_{n_k}(x - h), z) = g(x, z) - (h, z).$$

Therefore $g(x - h, z)$ with respect to measure μ_h has the same distribution as $g(x, z)$ with respect to measure μ_0. Because $(g_n B g_n^* z, z) \to 0$, so $\mu_h(\{z : g(z, z) = 0\}) = 1$. It means $\mu_0(\{z : g(x, z) = (h, z)\}) = 1$. From here it is easy to get that $g(x, z) = (g(x), z)$ and $\mu_h(\{z : g(x) = h\}) = 1$ for all $h \in H$. □

Remark 2. The proof for case strongly sequential consistent criteria is similar.

Theorem 1 shows that in many cases for building consistent criteria for hypotheses testing can be used linear criteria. Let us consider the question of existence of such criteria.

Theorem 2.2. *Let* $X_n \subset X_{n+1}$ be *some increasing sequence of finite dimensional subspaces from Hilbert space* X*;* Q_n *projector on* X_n*; operator* Q_n' *satisfies relations*

$$Q_n'(u) = u, \ if\ u \in X_n \Theta Q_n H;$$

$$Q' v_n \in X_n \Theta Q_n H \ for\ all\ v \in X_n;\ Q_n' B = B Q_n' \ on\ X_n.$$

If $z = \lim_{n\to\infty} Q_n z$ *(*z *lies in closure* $\cup X_n$*) then the condition*

$$\lim_{n\to\infty} (B Q_n' Q_n z, Q_n' Q_n z) = (Bz, z)$$

is necessary and sufficient for existence of unbiased consistent criteria coordinated with sequence subspaces $\{X_n\}$.

Proof. Let us suppose that from X allocated sequence subsets X_n such that $X_n \subset X_{n+1}$ and $\bigcup_{n=1}^{\infty} X_n$ densely in X. We denote by Q_n projector on X_n. The sequence of linear criteria $\lambda_n(x)$ is called unbiased criteria for (h, z) coordinated

with the system of subspaces X_n if:

(1) $\lambda_n(x) = \lambda_n(Q_n(x))$,

(2) $\int_X \lambda_n(x)\mu_h(dx) = (Q_n h, z)$ *for all* $h \in H$.

$\lambda_n(x)$ which satisfies condition (1) has form $\lambda_n(x) = (a_n, x)$ where $a_n \in X_n$. If the condition (2) is fulfilled then $(a_n, h) = (Q_n h, z)$ and it means $(h, a_n - Q_n z) = 0$ for all $h \in H$. In particular, it is fulfilled if $a_n = Q_n z$. Thus unbiased criteria for h always exist. If unbiased criteria $\lambda_n(x) = (a_n, x)$ exist then

$$\int_X [\lambda_n(x) - (Q_n h, z)]^2 \mu_h(dx) = \int_X [(x, a_n) - (h, Q_n z)]^2 \mu_h(dx)$$
$$= \int_X [(x, a_n) - (h, a_n)]^2 \mu_h(dx) = \int_X (x - h, a_n)^2 \mu_h(dx) = (Ba_n, a_n).$$

It means that dispersion of unbiased criteria is independent of h.

We will naturally choose the criteria so that depression will be minimal. Such criteria are called the best unbiased criteria.

For building such criteria we need to find $a_n \in X_n$ for which $(a_n, h) = (Q_n z, h)$ for all $h \in H$ and (Ba_n, a_n) is minimal.

The relation $(a_n, h) = (Q_n z, h)$ for all $h \in H$ is equivalent to relation $(a_n, u) = (Q_n z, u), u \in Q_n H$, where $Q_n H$ projection H on X_n. It is clear that $Q_n H$ is the linear subspace in X_n. We denote it by U_n. The set of such a_n for which $(a_n, u) = (Q_n z, u)$, for all $u \in U_n$, has form: $a_n = Q_n z + \nu_n$, where $\nu_n \in X_n \Theta U_n$ (this is orthogonal complement of U_n in X_n).

Let us suppose, that the operator B_n is non degenerate, then in X_n we can introduce scalar product: $(x, y)_n = (B_n x, y)$ and corresponding metric. Minimization of (a_n, a_n) is reduced to finding of vector of form $a_n = Q_n z + \nu_n$, where $\nu_n \in X_n \Theta U_n$, with minimal length in new metric. We denote by Q'_n projector on $X_n \Theta U_n$ in new scalar product $(.,.)_n$.

Let $\nu_n = Q'_n Q_n z$ then $a_n = Q_n z - Q'_n Q_n z$ is vector for which sequence $\lambda_n = (a_n, x)$ is the best unbiased criteria. We note two extreme cases: (1) $U_n = \{0\}$, $X_n \Theta U_n = X_n, a_n = 0$ and it means $\mu_h(\{\lambda_n(x) = (Q_n z, h)\}) = 1$. As $\cup X_n$ is densely in X then this can fulfilled for only finite values of n. (2) $U_n = X_n$, then

$$X_n \Theta U_n = \{0\}, a_n = Q_n z, \lambda_n(x) = (Q_n z, x).$$

In this case the unbiased criterion is only one and it is the best unbiased criterion and $\lambda_n(x) \to (x, z)$ if $n \to \infty$.

We are interested in consistent criteria i.e. such criteria for which

$$\lim_{n\to\infty} \int_X [\lambda_n(x) - (Q_n z, h)]^2 \mu_h(dx) = 0$$

for all $h \in H$.

For existence unbiased criteria for hypotheses testing is necessary and sufficient that

$$\lim_{n\to\infty} [(BQ_n z, Q_n z) - 2(BQ_n z, Q'_n Q_n z) + (BQ'_n Q_n z, Q'_n Q_n z)] = 0. \tag{2.1}$$

If this condition is fulfilled, then the best unbiased criteria will be consistent, because its dispersion is (Ba_n, a_n) and aspiration to zero of dispersion, as follows from equality $a_n = Q_n z - Q'_n Q_n z$, is equivalent to following mathematical expression (2.1)

$$\lim_{n\to 0} [(BQ_n z, Q_n z) - 2(BQ_n z, Q'_n Q_n z) + (BQ'_n Q_n z, Q'_n Q_n z)] = 0.$$

This expression can be simplified. At first, $Q_n z \to z$, secondly by orthogonality a_n and $Q'_n Q_n z$ at scalar product $(.,.)_n$ we have

$$\|Q_n z\|_n^2 = \|Q'_n Q_n z\|_n^2 + \|a_n\|_n^2.$$

So condition $\|a_n\|^2 = (Ba_n, a_n) \to 0$ is equivalent to condition

$$\lim_{n\to\infty} (BQ'_n Q_n z, Q'_n Q_n z) = (Bz, z). \quad \square$$

Remark 3. It should be noted that in the preceding arguments the density of $\cup X_n$ was used only for sowing that has place $Q_n z \to z$. So in the formulation of Theorem 2 the last condition we replace by density condition. If we want to build consistent criteria for (h, z) when z is changing in some dense set, then it is sufficient to consider z in closure H and demands that

$$\lim_{n\to} Q_n h = h \quad for \quad all \quad h \in H.$$

In general, existence or lack of unbiased, consistent criteria depends on choice sequence of subspaces X_n. Can be put the question, as choice vector a that for given z integral

$$\int_X [(a, x) - (h, z)]^2 \mu_h(dx)$$

was as small as possible.

Because this integral is equal

$$\int_X (a, x - h)^2 + (a - z, h)^2 \mu_h(dx) = (Ba, a) + (a - z, h)^2,$$

then the problem is reduced to finding minimum of this expression.

Theorem 2.3. *Let*

$$\inf_a [(Ba, a) + \sup_{\|h\|\le 1} (a - z, h)^2] = 0,$$

then (h, z) has consistent, unbiased criteria $\lambda_n(x)$ coordinated to some increased sequence of finite dimensional subspaces in X_n.

Proof. Let a_n be such sequence that

$$(Ba_n, a_n) + \sup_{\|h\|\le 1} (a_n - z, h)^2 \to 0 \tag{2.2}$$

for all $h \in H$.

Let X_n be subspace generated by $z, a_1, a_2, \ldots, a_n$. It is evident $Q_n z = z$. So general form of unbiased, consistent criteria is following $(z + u_n, x)$, where $(u_n, Q_n h) = 0$ for all $h \in H$. Let us represent $a_n - z = u'_n + v'_n$, where $v'_n \in Q_n H, u'_n \in X_n \Theta Q_n H$. From condition (2.2) follows that $\|v'_n\| \to 0$. It means that $(a_n, x) = (z+u'_n, x)+(v'_n, x)$. Because the second summand tends to zero. Then $(z + u'_n, x)$ is unbiased criteria, for which dispersion tends to zero. □

References

[1] A. Borovkov, Mathematical Statistics, Nauka, Moscow, 1984.

Guide for authors

Types of papers

Proceedings (Transactions) of A. Razmadze Mathematical Institute focus on significant research articles on both pure and applied mathematics. They should contain original new results with complete proofs. The review papers and short communications, which can be published after Editorial Board's decision, are allowed.

All efforts will be made to process papers efficiently within a minimal amount of time.

Ethics in publishing

For information on Ethics on publishing and Ethical guidelines for journal publication see http://www.elsevier.com/editors/publishing-ethics and http://www.elsevier.com/authors/journal-authors/policies-and-ethics .

Author rights

As an author you (or your employer or institution) have certain rights to reuse your work. For more information see www.elsevier.com/copyright

Contact details for Submission

Authors should submit their manuscript via the Elsevier Editorial System (EES), the online submission, peer-review and editorial system for Proceedings (Transactions) of A. Razmadze Mathematical Institute.

Submission declaration

Submission of an article that described has not be published previously (except in the form of abstract or as part of a published lecture or thesis or as an electronic preprint), that it is not under consideration for publication elsewhere, that its publication is approved by all authors.

Copyright

Upon acceptance of an article, authors will be asked to complete "Journal Publishing Agreement" (for more information see www.elsevier.com/copyright). An e-mail will be sent to the corresponding author confirming receipt of the manuscript together with a "Journal Publishing Agreement" form or a link to the online version of this agreement.

Language

Please write your text in good English (American or British is accepted, but not a mixture of these).

Submission

Our online submission system guides you stepwise through the process of entering your article details and uploading your files. The system converts your article files to a single PDF file used in the peer-review process.

All correspondents, including notification of the Editor's decision and requests for revision, is sent by e-mail.

References

Citation in text

Please ensure that every reference cited in the text is also present in the reference list (and wise versa). Any references cited in the abstract must be given full. Unpublished results and personal communications are not recommended in the reference list, but may be mentioned in the text. Citation of a reference "in press" implies that the item has been accepted for the publication.

Web References

The full URL should be given and the date when the reference was last accepted. Any further information, if known (DOI, author names, dates, reference to a source publication, etc.) should be also given.

LaTeX

It is recommended that each submitted article be prepared in camera-ready form using TeX(plain, LaTeX, $\mathcal{A}\mathcal{M}\mathcal{S}$LaTeX) macro page. The type-font for the text is ten point roman with the baselinkship of twelve point. The text area is 190×115 mm excluding page number. The final pagination will be done by the publisher.

Abstracts

The abstract should state briefly the purpose of the research, the principal results and major conclusions. References in the abstract should be avoided, but if essential they must be cited in full, without reference list.

Keywords

Immediately after the abstract, provide a maximum of ten keywords, using American spelling and avoiding general and plural terms and multiple concepts (avoid, for example “and”, or “of”).

Classification codes

Please provide classification codes as per the AMS codes. A full list and information for these can be found at www.ams.orm/msc/ .

Acknowledgements

Collate acknowledgements in a separate section at the end of the article before references and do not, therefore, including them in the title page, as a footnote.

Results

Results should be clear and concise.

Essential title page information

- Title.
  Concise and informative. Avoid abbreviations and formulae where possible.
- Author names and affiliation.
  Please clearly indicate the given name(s) and family name(s) of each author and check that all names are accurately spellt. Present the author’s affiliation addresses below the names. Provide the full postal addresses of each affiliation, including country name and, if available, the e-mail address of each author.
- Corresponding Author.
  Clearly indicate who will handle correspondence at all stages of refereeing and publication.
- Present/permanent address.
  If an author has moved since the work described in the article was done, was visiting at a time, ‘Present address’ (or ‘Permanent address’) may be indicated as a footnote to that author’s name.