Angela Slavova *Editor*

New Trends in the Applications of Differential Equations in Sciences

NTADES 2022, Sozopol, Bulgaria, June 14–17

Springer

Springer Proceedings in Mathematics & Statistics

Volume 412

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including data science, operations research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

Angela Slavova

Editor

New Trends
in the Applications
of Differential Equations
in Sciences

NTADES 2022, Sozopol, Bulgaria,
June 14–17

Springer

Editor
Angela Slavova
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Sofia, Bulgaria

Preface

In this volume, proceedings of the Ninth International Conference on New Trends of the Applications of Differential Equations in Sciences (NTADES 2022), Sozopol, Bulgaria, 14–17 June 2022 are presented.

This conference was devoted to many applications of differential equations in different fields of science. A number of phenomena in nature (physics, chemistry, and biology) and in society (economics) result in problems leading to the study of linear and nonlinear differential equations. During the conference invited and contributed papers were presented. The main topics are given as follows: Applications in Mathematical Physics; Applications in Mathematical Biology; Applications in Mathematical Finance; Applications in Neuroscience; and Applications in Fractional Analysis. The conference provided a wide range of problems concerning recent achievements in both theoretical and applied mathematics. Basic research in mathematics leading to new methods and techniques useful for applications of differential equations was presented. Young researchers, postdocs, and Ph.D. students participated during NTADES 2022 as well.

The Conference was in cooperation with the Society of Industrial and Applied Mathematics. SIAM is the major international organization for Industrial and Applied Mathematics and its role is very important in the Republic of Bulgaria for the promotion of interdisciplinary collaboration between applied mathematics and science, engineering, finance, and neuroscience.

The main goal of the proceedings is to exchange new ideas and research between scientists, who develop and study differential equations, and researchers, who apply them for solving real-life problems. During the conference, more than 70 talks were presented by scientists from universities and institutes from different countries. They represented most of the strongest research groups in the fields. In this volume, 42 chapters in 5 parts are presented on the above topics which were selected after a peer-review process.

In Part I which is the major part of the volume—19 chapters, some new applications of differential equations in mathematical physics are included. These papers concern the study of partial differential equations arising in mathematical physics,

and mechanics, and provided the most recent new methods for their investigation. I would like to stress the readers' attention to Chaps. "Porous-Media Flow and Yamabe Flow on Complete Manifolds" and "On the Square of Laplacian with Inverse Square Potential" which actually were invited plenary talks during the conference.

Part II which is organized into six chapters, deals with some applications in biology, more special studying some models of the Hepatitis B virus with control on the immune system, modeling and simulation of virotherapy in oncology, modeling epilepsy phenomena, etc.

In Part III, consisting of six chapters investigations of forecasting in finance are presented via different models, diversification, and optimization of the financial portfolio. In some of the chapters, stochastic models are investigated for stock price prediction and portfolio formation via extensive simulations.

Part IV which includes five chapters, is devoted to neuroscience applications by studying artificial neural networks, as well as some stochastic approaches based on modified Sobol sequences for Fredholm integral equations. Computer simulations of air pollution models are presented as well.

Last Part V consists of six chapters and provided recent investigations in fractional analysis as well as in approximation theory.

All chapters of this edition are written in an accessible manner and are very well illustrated. The volume can be considered as an informative milestone for young scientists and a comprehensive reference to the experienced reader in order to stimulate further research in applications of differential equations.

Sofia, Bulgaria   Angela Slavova
July 2022

Contents

Applications in Financial Mathematics

Applications in Neuroscience

Applications in Mathematical Physics

Porous-Media Flow and Yamabe Flow on Complete Manifolds

Li Ma

Abstract In this paper, we discuss the recent progress about the (sigma-) porous-media flow and Yamabe flow on the whole space or on a complete Riemannian manifold. The global existence of singular Yamabe flow will be showed. Some results about Yamabe flow related to previous results about scalar curvature problems such as the works of W. M. Ni, Aviles–McOwen, Delanoe, etc. will also be discussed.

Keywords Yamabe flow · Scalar curvature problem · Monotone method · Singular manifolds

1 Introduction to Porous-Media (PM) Equation and Yamabe Flow

The aim of this talk is to present some recent results about Yamabe flow on complete non-compact Riemannian manifolds and porous-media flow on the whole space. We also give outline proofs of the results and pose some new questions.

1.1 Porous-Media Equation

The porous-media equation can be introduced below: Let $\Omega \subset R^n$ be a bounded smooth domain and let $T > 0$. Let $Q = Q_T = \Omega \times [0, T)$ be the space-time domain with parabolic boundary $\partial_p Q = \partial Q \setminus \Omega \times \{T\}$. The typical model equation for porous-media equation is

$$u_t = \Delta u^m, \ u > 0, \ Q_T = \Omega \times [0, T), \ \Omega \subset R^n, \ m > 0. \tag{1}$$

L. Ma (✉)
School of Mathematics and Physics, University of Science and Technology Beijing, 30 Xueyuan Road, Haidian District, Beijing 100083, People's Republic of China
e-mail: lma17@ustb.edu.cn

The study of PM equation has a very long history [1, 37] and the classical result says that it is locally well-posed in weak sense [9].

For $\Omega = \mathbf{R}^n$, we define, for $b, c > 0$, $u_{bc}(x, t) = cu(bx, c^{m-1}b^2 t)$. Then $u_{bc}(x, t)$ is still a solution to (1). In particular, for $u_c(x, t) = cu(x, c^{m-1}t)$ with $c > 1$, by $u_c(x, 0) \geq u(x, 0)$ and by the comparison principle, we know that $u_c \geq u$ for all $t > 0$, which implies that $\frac{d}{dc}|_{c=1}u_c = (m - 1)tu_t + u \geq 0$, i.e., $u_t \leq \frac{u}{(1-m)t}$, $\forall t > 0$, which is the Aronson–Benilan inequality.

Fix $T > 0$ and let $C_n = \frac{2m(n-2-mn)}{1-m}$, $\beta = \frac{n}{n-2-nm}$, and $\gamma = -\beta/n$. Define $u_T(x, s) = (T - t)^{-\beta}u((T - t)^{\gamma}x, t)$, $s = -\log(T - t)$, which are Rescaled solutions. Barenblatt's solutions are of the following form: $B_\lambda(x, t) = \left(\frac{C_n(T-t)}{\lambda(T-t)^{2\gamma}+|x|^2}\right)^{1/(1-m)}$.

The main results from Osher–Ralston, Herrero–Pierre [17], Daskalopoulos–Sesum [11], etc. say that these special solutions are stable in some weighted L^1 space.

Sigma porous-media flow under consideration is given below. Fix $\sigma > 0$. We also consider the σ-(porous-media) flow

$$u^\sigma u_t = Lu, \quad on \ Q_T := \Omega \times [0, T], \tag{2}$$

with the initial-boundary conditions, where $Lu = \Delta u + Vu$ with V being the potential function on $\overline{\Omega}$.

We have the following interesting questions:

Q1: $T > 0$. Classify ancient or eternal solutions to the σ-(porous-media) flow $u^\sigma u_t = Lu$, $on \ bf \ R^n \times (-\infty, T]$ $or \ \mathbf{R}^n \times (-\infty, \infty)$.

Q2: Study the asymptotic behavior of global solutions to the σ (porous media))-flow $u^\sigma u_t = Lu$, $on \ \mathbf{R}^n \times [0, \infty)$.

Q3: Fix $T > 0$. Does the σ (porous media))-flow keeps some weighted Sobolev spaces?

The last question is always true for the standard *heat flow* on $\mathbf{R}^n$. Recall that $u_e(x, t) = e^{x+t}$ and $u_s(x, t) = \frac{1}{2}x^2 + ax + t$, and linear functions $ax + C$ are eternal solutions on $\mathbf{R}$ and they are both L^p stable by $L^p - L^q$ estimate, which says that for any $1 \leq q \leq p \leq \infty$, the solution satisfies $|u(t)|_p \leq (4\pi t)^{-\frac{n}{2}(\frac{1}{q}-\frac{1}{p})}|u_0|_q$, $t > 0$.

One may have weighted version of this inequality. As we mentioned in the last conference, there is a lot of results about porous-media flow in R^n. Still, we may contribute some new results about this flow in latter sections. One may consult the books listed [9] for more examples. For the study of the evolution equation $u_t = u^2(\Delta u + u)$ on Q_T, one may also consult the paper [16]. For porous-media equations, one may also refer to the survey papers [1], etc.

One of *our goals* here is to find some existence results about solutions to PM equations. Based on the understanding of singular Yamabe equation, we shall prove that for some $m \in (0, 1)$, the evolution problem 1 has global solutions with singular initial datum. We will see the geometric character of the related problems is important.

1.2 Scalar Curvature Problems on Compact Manifolds

It is well known that the scalar curvature $S(g)$ of a Riemannian manifold (M, g) is a weak invariant, but it plays an important role in general relativity theory of A. Einstein [19]. The famous Hilbert action is the total scalar curvature defined by $A(g) = \int_M S(g)dV_g$, and its Euler–Lagrange equation is more or less the Einstein equation about the metric g. We now have good understanding about closed 3-manifold with positive scalar curvature. However, we still have a long way to go to classify closed 4-manifold with positive scalar curvature. By the results of Schoen–Yau, Gromov–Lawson, etc. we know that the $n-$ torus T^n cannot have nontrivial nonnegative scalar curvature. One basic problem in scalar curvature geometry is to understand if a metric can be conformally deformed into one with constant scalar curvature, and this problem leads us to consider the geometric evolution of the metric in a conformal class.

We now mention the famous *scalar curvature problems and the Yamabe problem:* The conformal change of the metric g in a Riemannian manifold $(M^n, g), n \geq 3$, says that the scalar curvature of the conformal metric $\tilde{g} = u^{4/(n-2)}g$ is given by $R(\tilde{g}) = u^{-(n+2)/(n-2)}L_g u$, in M, where $L_g u := -a\frac{4(n-1)}{n-2}\Delta_g u + R(g)u$, $a = \frac{4(n-1)}{n-2}$, and $R(g)$ is the metric of the metric g. Sometimes, we let $S(g) = R(g)$. The famous *scalar curvature problem* states that for any closed Riemannian manifold with any given function K, if one can find a conformal metric with the scalar curvature K. One may ask similar question on a complete non-compact or a compact manifold with boundary.

A special case of constant scalar curvature problem is *the famous Yamabe problem* [19], which may be stated as follows: Assume M is closed. Define the Yamabe invariant $Q(M) = \inf\{J(u); u \neq 0, u \in H^1(M)\}$ for the Yamabe functional for $u \in H^1(M), u \neq 0, J(u) = \frac{\int_M (a|\nabla u|^2 + R(g)u^2)}{(\int_M |u|^{2n/(n-2)})^{(n-2)/n}}$.

The famous *Yamabe problem* [38] about the *existence of a minimizer* of the functional $J(u)$ when M is compact. After some important works from N. S. Trudinger [19], Th. Aubin [2], *resolution of the Yamabe problem* was done by R. Schoen [34] in 1984. R. Schoen solved the problem via a use of positive mass theorem.

One may also pose *Yamabe problem with boundary* on compact manifold with boundary [7, 13, 14, 22]. For references about the Yamabe problem on a complete Riemannian manifold, see also [21].

1.3 Scalar Curvature Problem on Complete Manifolds

In case when $M = R^n$, the scalar curvature equation can be rewritten as

$$-\Delta u = K u^{\frac{n+2}{n-2}}, \ u > 0, \ in \ R^n.$$

For this problem, W.M.Ni [31] obtained a remarkable result, which will be stated in latter section. Ni's result was improved by Delaneo (1992) and Ma-McOwen [30]. This problem is largely open on complete non-compact manifolds. In the case when M is an ALE manifold, Brill–Cantor (1981) obtained a beautiful result, where some arguments had some drawback and was corrected by Maxwell, Choquet-Brohat, etc. New results in this direction were obtained by Dilts–Maxwell (2018) [10]. A well-known result is that any compact manifold M^n, $n \geq 3$, admits a Riemannian metric with negative scalar curvature.

The *singular Yamabe problem* is also interesting to know. Given any compact Riemannian manifold (M^n, g), $n \geq 3$, with $Q(M) \leq 0$. Let $\Sigma := \Sigma^d \subset M$ be a closed submanifold of M, $d \leq \frac{n-2}{2}$. Let $N = M \setminus \Sigma$. We ask if there is a complete metric $\hat{g}$ in the conformal class $[g]$ such that $R(\hat{g})$ is a constant on N and such a problem is also called the Yamabe problem on complete non-compact manifolds. There is a counter-example showed to us by Z. R. Jin (1988). On one hand, Aviles–McOwen (1988) proved that N admits no complete metric in the conformal class $[g]$ with negative constant scalar curvature. On the other hand, using the maximum principle, Delanoe (1992) proved that N admits no complete metric in the conformal class $[g]$ with nonnegative scalar curvature. Results about *singular Yamabe problem on S^n* are relatively rich. For $M = S^n$ and $d > \frac{n-2}{2}$, Loewner–Nirenberg (1974) proved that there exists a complete, conformal metric on $S^n \setminus \Sigma$ with constant negative scalar curvature. For $d \leq \frac{n-2}{2}$, Schoen (1988), Mazzeo–Pacard (1999), etc. proved that there exists a complete, conformal metric on $S^n \setminus \Sigma$ with constant positive scalar curvature. There are also some gluing arguments to find singular metrics of constant positive scalar curvature. The moduli space of such metric had also been studied, see works of K. Uhlenbeck, Mazzeo, Pollack, etc.

1.4 Yamabe Flow

We now come to a discussion of *Yamabe flow* [24]. The Yamabe flow was introduced by R.Hamilton in 1989 to understand the Yamabe problem on closed manifolds. Let (M^n, g_b) be the Riemannian manifold of dimension $n \geq 3$ and let $u = u(x, t)$'s are smooth positive functions on M. The *Yamabe flow* is the family of metrics $g(t) = u^{\frac{4}{n-2}} g_b$ satisfying

$$\partial_t g = -R(g)g, \quad g(0) = u_0^{\frac{4}{n-2}} g_b, \tag{3}$$

where $R(g) = u^{-\frac{n+2}{n-2}}(-\frac{4(n-1)}{n-2}\Delta_0 u + R_0 u)$ is the scalar curvature of the metric g and $R_0 = R(g_b)$, where $\Delta_0 = \Delta_{g_b}$ is the Laplacian operator of the metric g_b. We may rewrite (3) by $\partial_t u = u^{\frac{-4}{n-2}}((n-1)\Delta_0 u - \frac{n-2}{4}R_0 u)$.

If we change the background metric $g_b = U^{\frac{4}{n-2}} g_c$, then $g(t) = w^{\frac{4}{n-2}} g_c$ by letting $w = Uu$ and $\partial_t w = w^{\frac{-4}{n-2}}((n-1)\Delta_{g_c} w - \frac{n-2}{4}R(g_c)w)$ with the initial metric Uu_0.

For M being closed, see papers of R. Hamilton, B. Chow, R. Ye, H. Schwetlick–Struwe, S. Brendle, etc. For related Yamabe flow with boundary, one may obtain the global Yamabe flow with the first initial-boundary data. There are some results about *Yamabe flow with boundary data*. For more results, one may refer to the papers of S. Brendle (2002) [4], Ma-Zheng (2021)[23]. The uniqueness of backward Yamabe flow with boundary was obtained by Park Tung Ho [18].

2 Zero and Negative Constant Scalar Curvature Problem and Yamabe Flow

We recall the famous *Brill–Cantor theorem* on ALE [5] in the following way.

Theorem 1 *On an ALE* $(p, 2, \delta)$ *manifold* (M, g) *with* $p > n/2$, $\delta > -\frac{n}{p}$, *there exists an ALE conformal metric* $\tilde{g}$ *such that* $R(\tilde{g}) = 0$ *on* M *if and only if for any nontrivial* $f \in W^p_{2,\tilde{\delta}}$, $\tilde{\delta} > -\frac{n}{p} + \frac{n-2}{2}$, *(* $\tilde{\delta} \geq -1$ *of* $p = 2$ *),*

$$\int_M (a|\nabla f|^2 + R(g) f^2) dV > 0. \tag{4}$$

For a recent proof, one may refer to Maxwell's paper (CMP 2005) or [8].

We prefer to recall Aviles–McOwen's results about the scalar curvature problem.

Theorem 2 (Aviles–McOwen, JDG, 1988) *Assume on the complete Riemannian manifold* (M, g) *that there is some nontrivial nonnegative* $f \in C^1_0(M)$ *satisfying (4), i.e.,*

$$\int_M (a|\nabla f|^2 + R(g) f^2) dV < 0.$$

Then there is a conformal metric $\check{g}$ *with* $R(\check{g}) = -1$. *If we further assume that there are a compact set* $M_0 \subset M$ *and a positive constant* $\delta > 0$ *such that*

$$R(g)(x) \leq -\delta, \quad \forall x \in M \backslash M_0, \tag{5}$$

or for some uniform $\beta \in (0, 2)$ *and* $\alpha \in [0, 1)$ *and* $\beta \in [2\alpha, 1 + \alpha)$, *there hold* $R(g)(x) \leq -C_1 r(x)^{-\beta}$ *for* $x \in M \backslash M_0$ *and* $Rc(g)(\partial_r, \partial_r) \geq -C_2 r(x)^{-2\alpha}$, *the metric* $\check{g}$ *is complete. Here,* $r(x)$ *is the distance function to some point* p *in* (M, g).

The important step in their argument is to find the lower solution of the form $u_-(r) = (r^2 + b^2)^{-\frac{n-2}{4}}$ to the equation $a\Delta_g u - u^{\frac{n+2}{n-2}} - R(g)u = 0$, *on* M.

We may use Aviles–McOwen's conditions to study the Yamabe flow. Assume the conditions (4) and (5), we can prove that for any initial regular data $\varphi(x) \geq u_-(r)$, there is a global Yamabe flow with the initial data φ.

3   Monotone Methods for Yamabe Flows and Sigma PM Flows

In this section, we consider the Yamabe flows and sigma PM flows for M being complete non-compact [28, 29].

First of all, some *recent results about Yamabe flow* [25, 26] can be stated as follows.

1. The existence of global Yamabe flow on asymptotically flat (in short, AF or ALE) manifolds. Note that the ADM mass is preserved in dimensions 3, 4, and 5. *Key step* in the proof: The Yamabe invariant plays an important role.

2. If the initial scalar curvature is nonnegative, the Yamabe flow on ALE manifolds converges to scalar-flat one.

Key step in the proof: there is a bounded positive sub-solution to the corresponding Poisson equation.

We do believe for the Yamabe positive case, we have the convergence of the global Yamabe flow. In the case when the Yamabe constant is nonpositive, the global flow on AF manifolds cannot have any smoothness convergence; otherwise, we get a limit metric with zero scalar curvature, which is impossible by Theorem 5.1 in [12]. It is an open question to normalize the flow to some convergence results. To understand Yamabe flow and porous-media equation well, we need the maximum principle for uniform parabolic equations. Let $\Omega \subset R^m$ be an open, bounded domain. Let $Q_T = \Omega \times [0, T]$ and $\partial Q_T = \partial \Omega \times [0, T] \cup \Omega \times \{0\}$.

Let, for $u \in C^{2,1}(\overline{Q_T})$,

$$Lu = a_{ij}(x, t)u_{ij} + b_j(x, t)u_j + c(x, t)u$$

with $c(x, t) \leq C$ for some $C \in \mathbf{R}$, and bounded continuous coefficients (a_{ij}) and (b_j) such that there is a uniform constant $\lambda_0 > 0$, for any $(x, t) \in Q_T$ and any $\xi = (\xi_j) \in R^n$, it holds $a_{jk}(x, t)\xi_j\xi_k \geq \lambda_0|\xi|^2$.

Define $Bu = u$ or $Bu = \partial_\nu u$ on $\partial \Omega \times [0, T]$). Assume that $a = a(x, t) > 0$ be a bounded continuous on $\overline{Q_T}$. The maximum principle says that if u is the regular solution to $Lu - a_0(x, t)u_t \geq 0$, in Q_T with $Bu \leq 0$ and $u(x, 0) = g(x) \leq 0$ for $x \in \Omega$, then we have $u < 0$ in the interior of Q_T.

We now review the classical *Maximum principle* [16, 20].

Theorem 3 (Linear parabolic maximum principle:) *Assume that $a \geq 0$ on Q_T and ellipticity $a_{ij}\xi_i\xi_j \geq 0$ for all $\xi \in R^m$ holds uniformly. Suppose, u satisfies*

$$\begin{cases} au_t - a_{ij}u_{x_ix_j} - b_ku_{x_k} - cu \leq 0 & in \Omega \times [0, T], \\ u \leq 0 & on \Omega \times \{0\} \text{ and } on \partial\Omega \times [0, T], \end{cases} \tag{6}$$

where the function $c - \lambda a < 0$ (for some $\lambda \in \mathbf{R}$) is bounded from above. Then, $u \leq 0$ in $\Omega \times [0, T]$.

We recall the definitions of *lower and upper solutions*. Let $f : \Omega \times \mathbf{R} \to \mathbf{R}$ be a smooth function on Q_T. Let $\mathbf{L}v = a_{ij}v_{x_i x_j} + b_k v_{x_k}$ be a uniformly elliptic operator on Ω with the complementary boundary operator Bv.

We say that $u_0 = u_0(x, t)$ is a *upper solution* to the evolution equation

$$v_t - \mathbf{L}v = f(x, v), \quad \text{on } Q_T, \tag{7}$$

if $v_t - \mathbf{L}v \geq f(x, v)$, on Q_T, with the initial-boundary condition $Bv \geq h$, where h is a given smooth function. Similarly we may define the *lower solution* to (7) with the initial-boundary data $Bv \leq h$.

To get global Yamabe flow, we need the *Monotone method* [32]. The well-known monotone method may be stated as follows.

Theorem 4 *Let $v_0 \leq u_0$ be a pair of lower upper (or sub-super) solutions to (7). Assume that there exists some constant A such that*

$$\partial_u f(x, u) + A \geq 0 \quad \text{on } \Omega \times [\min v_0, \sup u_0].$$

Then there exists a smooth solution to (7) with $v_0 \leq u \leq u_0$ on Q_T.

This result may be extended to Yamabe-type flow and the lower or upper solutions are constructed by solving linear problems. This is a basic tool for us to solve the existence problems of Yamabe-type flows or porous-media flows.

We remark that *more general formulation* about monotone method is possible.

One may formulate it in a little bit general way. Given a smooth nonlinear function $f(x, v)$. Assume that there is a uniform constant $K > 0$ such that

$$f_v + K > 0, \quad \overline{Q_T} = \overline{\Omega} \times [0, T]. \tag{8}$$

We now consider the nonlinear evolution equation

$$av_t = a_{ij}v_{x_i x_j} + b_k v_{x_k} + f(x, v), \quad a = a(x, t) > 0 \tag{9}$$

with the initial-boundary condition $v = g$ or $v_n + bv = c(x, t)$ $\partial\Omega \times \{t\}$ and $v = g$ at $t = 0$. Given a sub- and super-solution pair (u_1, u_2), $u_1 < u_2$. Then there is a solution $u \in [u_1, u_2]$ to (9).

4 Global Yamabe Flows on R^n and on Singular Manifolds

In 1982, W.M.Ni proved very interesting result about scalar curvature problem on R^n. Write $x = (x_1, x_2) \in R^{n-k} \times R^k$, $k \leq n$ and $n \geq 3$. Let $K(x)$ be a bounded locally Holder continuous function on R^n. Assume that there are some constants $C > 0$ and $l > 2$ such that

$$|K(x_1, x_2)| \leq \frac{C}{|x_2|^l}, \tag{10}$$

for $|x_2|$ large, uniformly in $x_1 \in R^{n-k}$. Then the problem

$$\Delta u + K(x)u^p = 0, \quad p > 1, \ in \ R^n \tag{11}$$

possess infinitely many bounded positive solutions with the property that each of these solutions is also uniformly bounded below by a positive constant. Moreover, if $K(x) \geq 0$ on R^n or $K(x) \leq 0$ on R^n, each of these solutions tends to a positive constant at infinity in the x_2-directions.

We may give some *Extensions*. Ni's result may be thought as a global result. For a local result, one may note the following argument. Let $F(u, K) = \Delta u + K(x)u^p$, $F(1, 0) = 0$ between certain weighted spaces. Notice that $D_1 F(1, 0)v = \Delta v$ and $D_1 F(1, 0) = \Delta$ is an isomorphism by the result of Nirenberg–Walker (1973 JMAA). By the implicit function theorem, we know that for K near to zero in some weighted space, we may solve (11) in certain weighted space. This argument works well for ALE manifolds with zero scalar curvature, see the works from Bartnik, McOwen, Delaneo, Brill–Cantor, Maxwell, Choquet-Brohat, etc.

4.1   Global Yamabe Flow on R^n

Solving elliptic problem has important *impact on Yamabe flow*. In fact, using Ni's result and the monotone method, one may get a global result about Yamabe flow.

Theorem 5 *Take $K_1 \leq 0 \leq K_2$ satisfying (10) on R^n. We may sort of a positive solution u_i to (11) with $K = K_i$ such that $u_1 \leq u_2$ on R^n. Then for any smooth $u_0 \in [u_1, u_2]$ on R^n, we have global Yamabe flow $(u(t))$ for the initial data u_0 and $u(t) \in [u_1, u_2]$.*

We also have similar result for sigma-porous-media flow on R^n and also on ALE manifolds. A similar result on singular metrics will be given in next section.

4.2   Global Yamabe Flow on Singular Manifolds

We first consider the Schoen–Yau's construction in a paper published in 1979 [35]. Let (M^n, g) be an n-dimensional compact Riemannian manifold without boundary. Assume that $R = S(g) > 0$ on M. By a classical result, we have a Green function of the conformal Laplacian operator L. Let $\Sigma := \Sigma^d \subset M$ be a closed submanifold of $M, n - d > 2$. Let $\rho(x) = d(x, \Sigma)$ near to Σ and it may be extended positively and smoothly to $M \backslash \Sigma$. According to Schoen–Yau, the equation $Lu = 0$ has a positive

solution u of the form $u = \rho^{-(n-d-2)} + 0(\rho^{1-(n-d-2)})$, for $d < n - 3$ and $u = \rho^{-(n-d-2)} + 0(\log \rho)$, for $d = n - 3$.

It is interesting to consider *zero scalar curvature problem*.

We remark that for $\frac{n-2}{2} < d < n - 2$, the metric $u^{4/(n-2)}g$ is incomplete. Delanoe (1992) [12] and Ma-McOwen (1992) [30] proved that if $Q(M) > 0$ and $d < n - 2$, for some $\epsilon = c(n, d) > 0$, $c > 0$,, there is a positive solution u of the form $u(x) = c\rho^{d+2-n} + 0(\rho^{d+2-n+\epsilon})$, x near to Σ to $Lu = 0$ on $N := M \setminus \Sigma$; moreover, if $d \leq \frac{n-2}{2}$, the metric $u^{4/(n-2)}g$ is complete.

We now consider the Yamabe flow on the singular manifold $N := M \setminus \Sigma$. We first mention *Schulz's result of Yamabe flow* [33].

Let (M^n, g_0) be a closed manifold. In the interesting paper, M. Schulz proved that if $d > \frac{n-2}{2}$, then an instantaneously complete Yamabe flow $(g(t))$ on $N = M \setminus \Sigma$ with $g(0) = g_0$ exists. He also proved that for $d \leq \frac{n-2}{2}$, any Yamabe flow with initial data (N, g_0) is incomplete and uniquely given by the restriction of the Yamabe flow with initial data (M, g_0). He leaves an open question if such a Yamabe flow exists. We can answer this question and give the existence of such flow below.

We use the barrier functions to get *global Yamabe flow*. Let $\beta \in (0, n - d - 2)$, $\alpha \in (0, 1)$. Let $f \in C_{loc}^{\alpha}(N)$ be a positive smooth function which equals to $\rho^{\beta-(n-d)}$ near to Σ. Since $f \in L^1(M)$, via the use of Green function G, Gf is well defined in $C_{loc}^{2,\alpha}(N)$ and satisfies $Lu = f$ on N with $Gf = 0(\rho^{\beta-(n-d-2)})$ near to Σ. Let, for $C > 0$ and $t \in \mathbf{R}$,

$$u_{C,t} = Cu + tGf > 0, \ in \ M.$$

Then we have

$$Lu_{s,t} = tf, \ in \ M.$$

For $t > 0$, $u_{C,t}$ is a super-solution of $Lu = 0$ and for $t > 0$, $u_{C,t}$ is a lower solution of $Lu = 0$. We may choose $C_1 < C_2$, $t_1 < 0$, $t_2 > 0$ such that $u_1 < u_2$, where $u_i = u_{C_i,t_i}$. Then (u_1, u_2) is a lower and super-solution pair and we may use the monotone method to find global solutions to Yamabe-type flow (and $\sigma-$porous-media flow) with the initial data $u_0 \in [u_1, u_2]$.

One may consult [3] for more results about Yamabe flow on singular manifolds.

References

1. D. G. Aronson, The porous medium equation. *Nonlinear diffusion problems (Montecatini Terme, 1985)*, volume 1224 of *Lecture Notes in Math.*, pages 1–46. Springer, Berlin, 1986.
2. T. Aubin, Équations différentielles non linéaires et problème de Yamabe concernant la courbure scalaire. *J. Math. Pures Appl. (9)*, 55(3):269–296, 1976.
3. E. Bahuaud, B. Vertman, Long-time existence of the edge Yamabe flow. *J. Math. Soc. Japan*, 71(2):651–688, 2019.

4. S. Brendle, A generalization of the Yamabe flow for manifolds with boundary. *Asian J. Math.*, 6(4):625–644, 2002.
5. M. Cantor, D. Brill, The Laplacian on asymptotically flat manifolds and the specification of scalar curvature. *Compositio Math.*, 43(3):317–330, 1981.
6. L. Cheng, A. Zhu, Yamabe flow and ADM mass on asymptotically flat manifolds. *J. Math. Phys.*, 56(10):101507, 21, 2015.
7. P. Cherrier, Problèmes de Neumann non linéaires sur les variétés riemanniennes. *J. Funct. Anal.*, 57(2):154–206, 1984.
8. Y. Choquet-Bruhat, J. Isenberg, D. Pollack, The Einstein-scalar field constraints on asymptotically Euclidean manifolds. *Chinese Ann. Math. Ser. B*, 27(1):31–52, 2006.
9. P. Daskalopoulos, C.E. Kenig, *Degenerate diffusions*, volume 1 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2007. Initial value problems and local regularity theory.
10. P. Daskalopoulos, N. Sesum, The classification of locally conformally flat Yamabe solitons. *Adv. Math.*, 240:346–369, 2013.
11. Ph. Delanoë, About the splitting $\mathbf{R}^N = \mathbf{R}^n \times \mathbf{R}^{N-n}$, $2 < n \leq N$, in a theorem of Wei-Ming Ni. *Comm. Partial Differential Equations*, 14(8-9):1127–1146, 1989.
12. J. Dilts, D. Maxwell, Yamabe classification and prescribed scalar curvature in the asymptotically Euclidean setting. *Comm. Anal. Geom.*, 26(5):1127–1168, 2018.
13. J.F. Escobar, Conformal deformation of a Riemannian metric to a scalar flat metric with constant mean curvature on the boundary. *Ann. of Math. (2)*, 136(1):1–50, 1992.
14. J.F. Escobar, The Yamabe problem on manifolds with boundary. *J. Differential Geom.*, 35(1):21–84, 1992.
15. A. Friedman, *Partial differential equations of parabolic type*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.
16. A. Friedman, B. McLeod, Blow-up of solutions of nonlinear degenerate parabolic equations. *Arch. Rational Mech. Anal.*, 96(1):55–80, 1986.
17. M. Herrero, M. Pierre, The Cauchy problem for $u_t = \Delta u^m$ when $0 < m < 1$. *Trans. Amer. Math. Soc.*, 291(1):145–158, 1985.
18. P.T. Ho, Backwards uniqueness of the Yamabe flow. *Differential Geom. Appl.*, 62:184–189, 2019.
19. O. A. Ladyvzenskaja, V. A. Solonnikov, N. N. Uralcprimeceva, *Linear and quasilinear equations of parabolic type*. Translations of Mathematical Monographs, Vol. 23. American Mathematical Society, Providence, R.I., 1968. Translated from the Russian by S. Smith.
20. J. Lee, T. Parker, The Yamabe problem. *Bull. Amer. Math. Soc. (N.S.)*, 17(1):37–91, 1987.
21. L. Ma, Conformal deformations on a noncompact Riemannian manifold. *Math. Ann.*, 295(1):75–80, 1993.
22. L. Ma, The Yamabe problem with Dirichlet data. *C. R. Acad. Sci. Paris Sér. I Math.*, 320(6):709–712, 1995.
23. L. Ma, Gap theorems for locally conformally flat manifolds. *J. Differential Equations*, 260(2):1414–1429, 2016.
24. L. Ma, Yamabe flow and metrics of constant scalar curvature on a complete manifold. *Calc. Var. Partial Differential Equations*, 58(1):Paper No. 30, 16, 2019.
25. L. Ma, New results about the lambda constant and ground states of the W-functional. *Adv. Nonlinear Stud.*, 20(3):651–661, 2020.
26. L. Ma, Global Yamabe flow on asymptotically flat manifolds. *J. Funct. Anal.*, 281(10):Paper No. 109229, 14, 2021.
27. L. Ma, Y. An, The maximum principle and the Yamabe flow. *In Partial differential equations and their applications (Wuhan, 1999)*, pages 211–224. World Sci. Publ., River Edge, NJ, 1999.
28. L. Ma, L. Cheng, Yamabe flow and Myers type theorem on complete manifolds. *J. Geom. Anal.*, 24(1):246–270, 2014.
29. L. Ma, W. Zheng, Yamabe type flow on bounded domains, new Trends in the Applications of Differential Equations in Sciences (NTADES 2021), Ed. A.Slovova, 2022.

30. X. Ma, R. McOwen, The Laplacian on complete manifolds with warped cylindrical ends. *Comm. Partial Differential Equations*, 16(10):1583–1614, 1991.
31. W.M. Ni, On the elliptic equation $\Delta u + K(x)u^{(n+2)/(n-2)} = 0$, its generalizations, and applications in geometry. *Indiana Univ. Math. J.*, 31(4):493–529, 1982.
32. D. Sattinger, *Topics in stability and bifurcation theory*. Lecture Notes in Mathematics, Vol. 309. Springer-Verlag, Berlin-New York, 1973.
33. R. Schoen, Conformal deformation of a Riemannian metric to constant scalar curvature. *J. Differential Geom.*, 20(2):479–495, 1984.
34. R. Schoen, S.T. Yau, On the structure of manifolds with positive scalar curvature. *Manuscripta Math.*, 28(1-3):159–183, 1979.
35. M. Schulz, Incomplete Yamabe flows and removable singularities. *J. Funct. Anal.*, 278(11):108475, 18, 2020.
36. N. Trudinger, Remarks concerning the conformal deformation of Riemannian structures on compact manifolds. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (3)*, 22:265–274, 1968.
37. J.L. Vázquez, An introduction to the mathematical theory of the porous medium equation. In *Shape optimization and free boundaries (Montreal, PQ, 1990)*, volume 380 of *NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci.*, pages 347–389. Kluwer Acad. Publ., Dordrecht, 1992.
38. H. Yamabe, On a deformation of Riemannian structures on compact manifolds. *Osaka Math. J.*, 12:21–37, 1960.

Simple Equations Method (SEsM): Areas of Possible Applications

Nikolay K. Vitanov

Abstract We discuss the Simple Equations Method (SEsM) for obtaining exact solutions of nonlinear differential equations. The focus of the discussion is on the areas of application of the methodology with respect to various properties of the solved equations. The point of interest here includes the kinds of nonlinearities. We present the result of the application of SEsM to differential equation with nonpolynomial nonlinearity.

Keywords SEsM · Nonlinear differential equations · Exact solutions · Nonpolynomial nonlinearities

1 Introduction

In the last decades, the mathematical instruments for the study of Nature and society become complicated enough. Thus, the researchers started a large amount of studies of complex systems. Just several examples are from the theory of networks, social sciences, economics, dynamics of research groups, etc. [1–15]. In many of these studies, one has to deal with the nonlinearity of the complex systems [16–26]. This can be done by the methods of the time series analysis and by the methods connected to model nonlinear differential or difference equations [27–40]. The methodology for obtaining exact and approximate solutions of these equations attracts much attention. Different ways to deal with the nonlinearity in the model equations have been used. The Hopf–Cole transformation [41, 42] represents such a way. The further study of appropriate transformations led to the Method of Inverse Scattering Transform [43,

N. K. Vitanov (✉)

Institute of Mechanics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 4, 1113 Sofia, Bulgaria

e-mail: vitanov@imbm.bas.bg

44] and to the method of Hirota for obtaining of exact solutions of NPDEs [45]. The truncated Painleve expansions may lead to many appropriate transformations of the nonlinearities of the studied equations [46, 47]. Kudryashov [48, 49] formulated the Method of Simplest Equation (MSE) which is based on the determination of singularity order n of the solved NPDE and searching of a particular solution of this equation as a series containing powers of solutions of a simpler equation called the simplest equation. Several examples of the application of this methodology are provided in [50–52].

30 years ago [53–59], we started work on methodology for obtaining exact and approximate solutions of nonlinear partial differential equations called nowadays Simple Equations Method (SEsM) [60–65]. We have used the ordinary differential equation of Bernoulli as simplest equation [66–68] and we have applied the methodology to ecology and population dynamics [69]. This version of the methodology used the concept of the balance equation for obtaining the kind of the simplest equation and the kind of the solution of the solved equation [70, 71]. Till more recent times (2018) [72–80], we have used the methodology on the basis of one simplest equation and one balance equation. A strategic article from this period was [79]. There we have extended the methodology of the MMSE to the simplest equations of the class $\left(\frac{d^k g}{d\xi^k}\right)^l = \sum_{j=0}^m d_j g^j$, where $k = 1, \ldots, l = 1, \ldots,$ and m and d_j are parameters. In 2018, we extended the methodology for use of more than one simple equation. This modification is called SEsM—Simple Equations Method [81], and the first description of the methodology was made in [60] and then in [62–65]. For more applications of particular cases of the methodology, see [82–86].

Below we will discuss various aspects of the application of SEsM. Some of these aspects occurred in the last year. In addition, we will discuss perspectives of the application of the SEsM.

2 Simple Equations Method (SEsM)

SEsM presents an algorithm for obtaining exact and approximate solutions of systems of n nonlinear differential equations. The solutions are constructed by the solutions of m more simple differential equations. SEsM consists of four steps. Let us consider a system of nonlinear partial differential equations

$$\mathcal{A}_i[u_1(x, \ldots, t), \ldots, u_n(x, \ldots, t)] = 0, \quad i = 1, \ldots, n.$$

$\mathcal{A}_i[u_1(x, \ldots, t), \ldots, u_n(x, \ldots, t), \ldots]$ depend on the functions $u_1(x, \ldots, t), \ldots, u_n(x, \ldots, t)$ and some of their derivatives (u_i can be a function of several spatial coordinates). The steps of SEsM are as follows. At Step 1, we transform the nonlinearity by $u_i(x, ..., t) = T_i[F_i(x, \ldots, t), G_i(x, \ldots, t), \ldots]$. $T_i(F_i, G_i, \ldots)$ is a function of other functions $F_i, G_i, \ldots$. $F_i(x, \ldots, t), G_i(x, \ldots, t), \ldots$ are functions of several spatial variables as well as of the time. The transformations T_i have the goal: to trans-

form the nonlinearity of the solved differential equations to a more treatable kind of nonlinearity (e.g., to polynomial nonlinearity). The application of the transformation to the solved equations leads to nonlinear differential equations for the functions $F_i, G_i, \ldots$. At Step 2 of SEsM, the functions $F_i(x, \ldots, t), G_i(x, \ldots, t), \ldots$ are represented as a composite function of other functions $f_{i1}, \ldots, f_{iN}, g_{i1}, \ldots, g_{iM}, \ldots$. The functions f_i and g_i are connected to solutions of some ordinary or partial differential equations. At Step 3 of SEsM, we choose the simple equations for which $f_{i1}, \ldots, f_{iN}, g_{i1}, \ldots, g_{iM}$ are solutions. At Step 4 of SEsM, we apply steps (1)–(3) to solved equations. This usually leads to a system of nonlinear algebraic equations for the coefficients of the solved nonlinear differential equation and for the coefficients of the solution. Any nontrivial solution of this algebraic system leads to a solution of the studied nonlinear partial differential equation.

3 Kinds of Nonlinearities Treated by SEsM

Step 1 of SesM allows for the treatment of various kinds of nonlinearities [87]. The idea is to reduce the nonpolynomial nonlinearity to polynomial nonlinearity and then to deal with the polynomial nonlinearity by means of Steps 2, 3, and 4 of SEsM. We consider below the problem of searching for exact solutions of nonlinear differential equations containing the function $u(x, \ldots, t)$ and its derivatives. u can depend on several spatial variables $x, \ldots$ and the time t.

We consider a differential equation for the function $u(x, \ldots, t)$ which contains terms of two kinds: (i) terms containing only derivatives of u; (ii) terms containing one or several nonpolynomial nonlinearities of the function u and these nonpolynomial nonlinearities are of the same kind. Let $u = T(F)$ be a transformation with the following properties: (i) The transformation T transforms any of the nonpolynomial nonlinearity to a function which contains only polynomials of F; (ii) The transformation T transforms the derivatives of u to terms containing only polynomials of derivatives of F or polynomials of derivatives of F multiplied or divided by polynomials of F. Then, the transformation T transforms the studied differential equation to a differential equation containing only polynomial nonlinearity of F.

We will skip the proof, and instead of this, we will list a transformation that transforms a nonpolynomial nonlinearity into a polynomial one. Let $N(u) = \sin(u)$. In this case, a possible transformation is $u = 4 \tan^{-1}(F)$. The transformation has Property (i) as $N(u) = \sin[4 \tan^{-1}(F)] = 4 \frac{F(1-F^2)}{(1+F^2)^2}$. $N(u)$ is transformed into a function that contains only polynomials of F. The transformation has also Property (ii) as for an example $u_x = \frac{4}{1+F^2} F_x$.

The list of the appropriate transformations can be continued. Let us now consider one illustrative example. We consider the differential equation with a nonpolynomial nonlinearity

$$bu_{xx}^2 + du_{tt}^2 = l \sin^2(u) \tag{1}$$

where b, d, l are parameters. We use the transformation $u = 4\tan^{-1}(F)$ at the first step of application of SEsM. This transformation leads to an equation for $F(x, t)$ containing only polynomial nonlinearity

$$4F^2(bF_x^4 + dF_t^4) - 4(F + F^3)(bF_x^2 F_{xx} + dF_t^2 F_{tt}) + (1 + 2F^2)(bF_{xx}^2 + dF_{tt}^2) +$$
$$F^4(bF_{xx} + dF_{tt}) - l(F^6 - 2F^4 + F^2) = 0. \tag{2}$$

The step 2 of SEsM requires F to be a composite function of more simple functions: $F(x, t) = F[T_1(x, t), T_2(x, t)]$. In order to consider the general case, we have to use the information from Appendices 1 and 2. In order to keep the example relatively simple, we will consider a particular case of the above composite function: $F(x, t) = AT_1(\mu)T_2(\xi)$, where $\mu = \alpha x$ and $\xi = \gamma t$. The result is a differential equation that contains polynomials constructed of T_1, T_2, and their derivatives.

At Step 3 of SEsM, we have to determine the form of the functions T_1 and T_2. We assume that T_1 and T_2 are solutions of more simple (and ordinary) differential equations which contain polynomial nonlinearity. $T_{1\mu}^2 = \sum_{i=0}^{N_1} \delta_i T_1^i$; $T_{2\xi}^2 = \sum_{j=0}^{N_2} \epsilon_i T_2^j$, where δ_i and ϵ_i are parameters. We substitute these relationships in (2). We obtain a polynomial of T_1 and T_2 which contains monomials of T_1, T_2, and monomials which are combinations of powers of T_1 and T_2. These monomials are multiplied by coefficients which are nonlinear algebraic relationships containing the parameters of the solved equation and the parameters of the more simple equations. We have to ensure that any of these nonlinear algebraic relationships contains at least two terms. This is done by a balance procedure that fixes the values of the parameters $N_1 = 4$ and $N_2 = 4$. For simplicity, we consider a specific case of the obtained simple equations. In such a way, the simple equations for the function T_1 and T_2 become $T_{1\mu}^2 = pT_1^4 + qT_1^2 + r$; $T_{2\xi}^2 = sT_2^4 + vT_2^2 + w$. This form of the simple equations leads to the following system of nonlinear algebraic equations (these are the nonlinear algebraic relationships for the coefficients of the polynomial containing T_1, T_2, and their derivatives).

$$b\alpha^4 q^2 + d\gamma^4 v^2 = l; \quad d\gamma^4 v A^2 w - b\alpha^4 pq = 0$$
$$-4d\gamma^4 sw - b\alpha^4 q^2 - 4b\alpha^4 pr + l - d\gamma^4 v^2 = 0;$$
$$b\alpha^4 A^4 r^2 + d\gamma^4 s^2 = 0; \quad -b\alpha^4 q A^2 + d\gamma^4 sv = 0;$$
$$b\alpha^4 p^2 + d\gamma^4 A^4 w^2 = 0; \quad -d\gamma^4 sv + b\alpha^4 q A^2 r = 0;$$
$$-d\gamma^4 v A^2 + b\alpha^4 pq = 0. \tag{3}$$

The system (3) has the solution

$$p = w = r = s = 0, \quad q = \delta \frac{[-b(d\gamma^2 v^2 - l)]^{1/2}}{\alpha^2 b}, \quad \delta = \pm 1, \quad -b(d\gamma^2 v^2 - l) \geq 0. \tag{4}$$

$v, A, l, b, d, \alpha, \gamma$ are free parameters (they have to satisfy the condition $-b(d\gamma^2 v^2 - l) \geq 0$). Equation (4) corresponds to the following solution of (1)

$$u(x,t) = 4\tan^{-1}\left\{A\exp\left[\delta_1\left(\alpha vx + \gamma\delta\frac{[-b(d\gamma^2 v^2 - l)]^{1/2}}{\alpha^2 b}t\right)\right]\right\}, \quad \delta_1 = \pm 1.$$
(5)

This solution describes traveling waves of kind kink and anti-kink.

4 Concluding Remarks

In this text, we studied the application of SEsM for the case of nonpolynomial nonlinearity of the solved differential equation. We show that by means of appropriate transformations, the nonpolynomial nonlinearities can be reduced to polynomial nonlinearities and then the SEsM can be applied for obtaining the solution of the resulting equations containing polynomial nonlinearities. Thus, the achievements of SEsM for solving differential equations with polynomial nonlinearity can be used to the full extent.

References

1. R. Levin R.: Complexity. Life at the Edge of Chaos. The University of Chicago Press, Chicago (1999)
2. Lambiotte, R., Ausloos, M.: Coexistence of opposite opinions in a network with communities. Journal of Statistical Mechanics: Theory and Experiment, P08026 (2007). https://doi.org/10.1088/1742-5468/2007/08/P08026
3. Dimitrova, Z.I.: On the nonlinear dynamics of interacting populations. Effects of delay on populations substitution. Compt. rend. Acad. bulg. Sci 61, 1541–1548 (2008)
4. Dimitrova, Z. I.: Fluctuations and dynamics of the chaotic attractor connected to an instability in a heated from below rotating fluid layer. Compt. rend. Acad. bulg. Sci 60, 1065–1070 (2007)
5. Vitanov N.K.: Science Dynamics and Research Production. Indicators, Indexes, Statistical Laws and Mathematical Models. Springer, Cham (2016)
6. May, R. M., Levin, S. A., Sugihara, G.: Ecology for bankers. Nature 451, 893–894, (2008). https://doi.org/10.1038/451893a
7. Nikolova, E. V., Serbezov, D. Z., Jordanov, I.: Nonlinear spread waves in population dynamics including a human-induced Allee effect. AIP Conference Proceedings vol. 2075, 150004 (2019). https://doi.org/10.1063/1.5091327
8. Vitanov, N. K., Vitanov, K. N.: Discrete-time model for a motion of substance in a channel of a network with application to channels of human migration. Physica A: Statistical Mechanics and its Applications 509, 635–650 (2018). https://doi.org/10.1016/j.physa.2018.06.076

9. Kutner, R., Ausloos, M., Grech, D., Di Matteo, T., Schinckus, C., Stanley, H. E.: Econophysics and sociophysics: Their milestones & challenges. Physica A 516, 240–253 (2019). https://doi.org/10.1016/j.physa.2018.10.019

10. Z. I. Dimitrova, Z. I., Hoffmann N. P.: On the probability for extreme water levels of the river Elba in Germany. Compt. rend. Acad. bulg. Sci 65, 153–160 (2012)

11. Vitanov, N. K., Vitanov, K. N.: On the motion of substance in a channel of a network and human migration. Physica A: 490, 1277–1294 (2018). https://doi.org/10.1016/j.physa.2017.08.038

12. Vitanov, N. K., Ausloos, M., Rotundo, G.: Discrete model of ideological struggle accounting for migration. Advances in Complex Systems 15, 1250049 (2012). https://doi.org/10.1142/S021952591250049X

13. Amaral, L. A. N., Scala, A., Barthelemy, M., Stanley, H. E.: Classes of small-world networks. PNAS USA 97, 11149–11152 (2000). https://doi.org/10.1073/pnas.20032719

14. Jordanov, I. P., Nikolova, E. V.: On the evolution of nonlinear density population waves in the socio-economic systems. AIP Conference Proceedings vol. 2075, 150002 (2019). https://doi.org/10.1063/1.5091325

15. Simon J. H.: The Economic Consequences of Immigration. The University of Michigan Press, Ann Arbor, MI, USA (1999)

16. Neil Rasband S.: Chaotic Dynamics of Nonlinear Systems. Dover, New York (1990)

17. Jordanov, I., Nikolova, E.: On nonlinear waves in the spatio-temporal dynamics of interacting populations. Journal of Theoretical and Applied Mechanics 43, 69–76 (2013). https://doi.org/10.2478/jtam-2013-0015. arXiv:1208.5465

18. Dimitrova, Z. I.: Numerical investigation of nonlinear waves connected to blood flow in an elastic tube with variable radius. Journal of Theoretical and Applied Mechanics 45, 79–92 (2015). https://doi.org/10.1515/jtam-2015-0025

19. Nikolova E., Goranova, E., Dimitrova Z.: Assessment of rupture risk factors of abdominal aortic aneurysms in Bulgarian patients using a finite element based system. Compt. rend. Acad. bulg. Sci 9, 1213–1222 (2016)

20. Jordanov, I. P.: On the nonlinear waves in (2+ 1)-dimensional population systems. Compt. rend. Acad. bulg. Sci 61, 307–314 (2008)

21. Ganji, D. D., Sabzehmeidani, Y., Sedighiamiri A.: Nonlinear Systems in Heat Transfer Elsevier, Amsterdam (2018)

22. Dimitrova, Z.I.: On travelling waves in lattices: the case of Riccati lattices. Journal of Theoretical and Applied Mechanics 42, 3–22 (2012), https://doi.org/10.2478/v10254-012-0011-2. arXiv:1208.2414

23. Nikolova, E.V.: Evolution Equation for Propagation of Blood Pressure Waves in an Artery with an Aneurysm. In: Georgiev, K., Todorov, M., Georgiev, I. (eds.) Advanced Computing in Industrial Mathematics. BGSIAM 2017. Studies in Computational Intelligence vol. 793, 327–339 (2019). https://doi.org/10.1007/978-3-319-97277-0_27

24. Jordanov, I. P., Dimitrova, Z. I.: On Nonlinear Waves of Migration. Journal of Theoretical and Applied Mechanics 40, 89–96 (2010)

25. Nikolova, E. V.: On nonlinear waves in a blood-filled artery with an aneurysm. In AIP Conference Proceedings vol. 1978, 470050 (2018). https://doi.org/10.1063/1.5044120

26. Dimitrova, Z. I., Ausloos, M.: Primacy analysis in the system of Bulgarian cities. Open Physics 13, 218–225 (2015). https://doi.org/10.1515/phys-2015-0029

27. Kantz, H., Schreiber T.: Nonlinear Time Series Analysis. Cambridge University Press, Cambridge, UK (2004)

28. Struble R.: Nonlinear Differential Equations. Dover, New York (2018)

29. Vitanov, N. K.: Upper bounds on the heat transport in a porous layer. Physica D 136, 322–339 (2000). https://doi.org/10.1016/S0167-2789(99)00165-7

30. Vitanov, N. K., Ausloos, M. R.: Knowledge epidemics and population dynamics models for describing idea diffusion. In: Scharnhorst A., Boerner K., van den Besselaar P. (eds.) Models of science dynamics. Understanding Complex Systems. pp. 69–125. Springer, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-23068-4_3

31. Dimitrova, Z. I., Vitanov, N. K.: Adaptation and its impact on the dynamics of a system of three competing populations. Physica A: Statistical Mechanics and its Applications 300, 91–115 (2001). https://doi.org/10.1016/S0378-4371(01)00330-2
32. Dimitrova, Z., Gogova, D.: Investigation of Differences in Optical Phonons Modes by Principal Component Analysis. Compt. rend. Acad. bulg. Sci 63, 1415–1420 (2010)
33. Mills T.: Applied Time Series Analysis. Academic Press, London (2019)
34. Dimitrova, Z. I., Vitanov, N. K.: Chaotic pairwise competition. Theoretical Population Biology 66, 1–12 (2004). https://doi.org/10.1016/j.tpb.2003.10.008
35. Dimitrova, Z. I.: On the Low-Dimensional Dynamics of Blood Flow in Small Peripheral Human Arteries. Compt. rend. Acad. bulg. Sci 63, 55–60 (2010)
36. Vitanov, N. K., Vitanov, K. N.: Statistical distributions connected to motion of substance in a channel of a network. Physica A 527, 121174 (2019). https://doi.org/10.1016/j.physa.2019.121174
37. Borisov, R., Dimitrova, Z. I., Vitanov, N. K.: Statistical characteristics of stationary flow of substance in a network channel containing arbitrary number of arms. Entropy 22, 553 (2020). https://doi.org/10.3390/e22050553
38. Dimitrova, Z. I., Vitanov, K. N.: Homogeneous balance method and auxiliary equation method as particular cases of simple equations method (SEsM). In AIP Conference Proceedings vol. 2321, 030004 (2021). https://doi.org/10.1063/5.0043070
39. Taulbee, D. B.: An improved algebraic Reynolds stress model and corresponding nonlinear stress model. Physics of Fluids A: Fluid Dynamics 4, 2555–2561 (1992). https://doi.org/10.1063/1.858442
40. Vitanov, N. K., Vitanov, K. N., Kantz, H.: On the motion of substance in a channel of a network: Extended model and new classes of probability distributions. Entropy 22, 1240 (2020). https://doi.org/10.3390/e22111240
41. Hopf E.: The partial differential equation $u_t + uu_x = u_{xx}$. Communications on Pure and Applied Mathematics 3, 201–230 (1950). https://doi.org/10.1002/cpa.3160030302
42. Cole, J. D.: On a quasi-linear parabolic equation occurring in aerodynamics. Quarterly of Applied Mathematics 9, 225–236 (1951). https://doi.org/10.1090/QAM/42889
43. Ablowitz, M. J., Clarkson, P. A.: Solitons, Nonlinear Evolution Equations and Inverse Scattering. Cambridge University Press, Cambridge, UK (1991)
44. Gardner, C. S., Greene, J. M., Kruskal, M. D., Miura, R. M.: Method for solving the Korteweg-deVries equation. Physical Review Letters 19, 1095–1097 (1967). https://doi.org/10.1103/PhysRevLett.19.1095
45. Hirota, R.: The Direct Method in Soliton Theory. Cambridge University Press, Cambridge, UK, (2004)
46. Tabor. M.: Chaos and Integrability in Dynamical Systems. Wiley, New York (1989)
47. Weiss, J., Tabor, M., Carnevale, G.: The Painlevé property for partial differential equations. Journal of Mathematical Physics 24, 522–526 (1983). https://doi.org/10.1063/1.525721
48. Kudryashov, N. A.: Simplest equation method to look for exact solutions of nonlinear differential equations. Chaos, Solitons & Fractals 24, 1217–1231 (2005). https://doi.org/10.1016/j.chaos.2004.09.109
49. Kudryashov, N. A., Loguinova, N. B.: Extended simplest equation method for nonlinear differential equations. Applied Mathematics and Computation 205, 396–402 (2008). https://doi.org/10.1016/j.amc.2008.08.019
50. Kudryashov, N. A.: Exact solitary waves of the Fisher equation. Physics Letters A 342, 99–106 (2005). https://doi.org/10.1016/j.physleta.2005.05.025
51. Kudryashov, N. A.: One method for finding exact solutions of nonlinear differential equations. Communications in Nonlinear Science and Numerical Simulation, 17, 2248–2253 (2012). https://doi.org/10.1016/j.cnsns.2011.10.016
52. Kudryashov, N. A.: Exact solutions of the equation for surface waves in a convecting fluid. Applied Mathematics and Computation, 344, 97–106 (2019). https://doi.org/10.1016/j.amc.2018.10.005

53. Martinov, N., Vitanov, N.: On the correspondence between the self-consistent 2D Poisson-Boltzmann structures and the sine-Gordon waves. Journal of Physics A: Mathematical and General 25, L51–L56 (1992). https://doi.org/10.1088/0305-4470/25/2/004

54. Martinov, N., Vitanov, N.: On some solutions of the two-dimensional sine-Gordon equation. Journal of Physics A: Mathematical and General 25, L419–L426 (1992). https://doi.org/10.1088/0305-4470/25/8/007

55. Martinov, N. K., Vitanov, N. K.: New class of running-wave solutions of the (2+ 1)-dimensional sine-Gordon equation. Journal of Physics A: Mathematical and General 27, 4611–4618 (1994). https://doi.org/10.1088/0305-4470/27/13/034

56. Martinov, N. K., Vitanov, N. K.: On self-consistent thermal equilibrium structures in two-dimensional negative-temperature systems. Canadian Journal of Physics 72, 618–624 (1994). https://doi.org/10.1139/p94-079

57. Vitanov, N. K.: Breather and soliton wave families for the sine-Gordon equation. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454, 2409–2423 (1998). https://doi.org/10.1098/rspa.1998.0264

58. Vitanov, N. K., Martinov, N. K.: On the solitary waves in the sine-Gordon model of the two-dimensional Josephson junction. Zeitschrift fur Physik B Condensed Matter 100, 129–135 (1996). https://doi.org/10.1007/s002570050102

59. Vitanov, N. K.: On travelling waves and double-periodic structures in two-dimensional sine-Gordon systems. Journal of Physics A: Mathematical and General 29, 5195–5207 (1996). https://doi.org/10.1088/0305-4470/29/16/036

60. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Simple Equations Method (SEsM): Algorithm, connection with Hirota method, Inverse Scattering Transform Method, and several other methods. Entropy 23, 10 (2021). https://doi.org/10.3390/e23010010

61. Vitanov, N. K.: Recent developments of the methodology of the modified method of simplest equation with application. Pliska Studia Mathematica Bulgarica 30, 29–42 (2019)

62. Vitanov, N.K.: Modified method of simplest equation for obtaining exact solutions of nonlinear partia differential equations: history, recent developments of the methodology and studied of classes of equations. Journal of Theoretical and Applied Mechanics 49, 107–122 (2019)

63. Vitanov, N. K.: The simple equations method (SEsM) for obtaining exact solutions of nonlinear PDEs: Opportunities connected to the exponential functions. AIP Conference Proceedings vol. 2159, 030038 (2019). https://doi.org/10.1063/1.5127503

64. Vitanov, N. K., Dimitrova, Z. I.: Simple equations method (SEsM) and other direct methods for obtaining exact solutions of nonlinear PDEs. AIP Conference Proceedings vol. 2159, 030039 (2019). https://doi.org/10.1063/1.5127504

65. Dimitrova, Z. I., Vitanov, N. K.: Travelling waves connected to blood flow and motion of arterial walls. Gadomski, A. (ed.) In: Water in Biomechanical and Related Systems pp. 243–263. Springer, Cham (2021)

66. Vitanov, N. K., Jordanov, I. P., Dimitrova, Z. I.: On nonlinear dynamics of interacting populations: Coupled kink waves in a system of two populations. Communications in Nonlinear Science and Numerical Simulation 14, 2379–2388 (2009). https://doi.org/10.1016/j.cnsns.2008.07.015

67. Vitanov, N. K., Jordanov, I. P., Dimitrova, Z. I.: On nonlinear population waves. Applied Mathematics and Computation 215, 2950–2964 (2009). https://doi.org/10.1016/j.amc.2009.09.041

68. Vitanov, N. K.: Application of simplest equations of Bernoulli and Riccati kind for obtaining exact traveling-wave solutions for a class of PDEs with polynomial nonlinearity. Communications in Nonlinear Science and Numerical Simulation 15, 2050–2060 (2010). https://doi.org/10.1016/j.cnsns.2009.08.011

69. Vitanov, N. K., Dimitrova, Z. I.: Application of the method of simplest equation for obtaining exact traveling-wave solutions for two classes of model PDEs from ecology and population dynamics. Communications in Nonlinear Science and Numerical Simulation 15, 2836–2845 (2010). https://doi.org/10.1016/j.cnsns.2009.11.029

70. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: Modified method of simplest equation and its application to nonlinear PDEs. Applied Mathematics and Computation 216, 2587–2595 (2010). https://doi.org/10.1016/j.amc.2010.03.102

71. Vitanov, N. K.: Modified method of simplest equation: powerful tool for obtaining exact and approximate traveling-wave solutions of nonlinear PDEs. Communications in Nonlinear Science and Numerical Simulation 16, 1176–1185 (2011). https://doi.org/10.1016/j.cnsns.2010.06.011

72. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: On the class of nonlinear PDEs that can be treated by the modified method of simplest equation. Application to generalized Degasperis-Processi equation and b-equation. Communications in Nonlinear Science and Numerical Simulation 16, 3033–3044 (2011). https://doi.org/10.1016/j.cnsns.2010.11.013

73. Vitanov, N. K.: On modified method of simplest equation for obtaining exact and approximate solutions of nonlinear PDEs: the role of the simplest equation. Communications in Nonlinear Science and Numerical Simulation 16, 4215–4231 (2011). https://doi.org/10.1016/j.cnsns.2011.03.035

74. Vitanov, N. K.: On modified method of simplest equation for obtaining exact solutions of nonlinear PDEs: case of elliptic simplest equation. Pliska Studia Mathematica Bulgarica 21, 257–266 (2012)

75. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: Application of the method of simplest equation for obtaining exact traveling-wave solutions for the extended Korteweg-de Vries equation and generalized Camassa-Holm equation. Applied Mathematics and Computation 219, 7480–7492 (2013). https://doi.org/10.1016/j.amc.2013.01.035

76. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Traveling waves and statistical distributions connected to systems of interacting populations. Computers & Mathematics with Applications 66, 1666–1684 (2013). https://doi.org/10.1016/j.camwa.2013.04.002

77. Vitanov, N. K., Vitanov, K. N.: Population dynamics in presence of state dependent fluctuations. Computers & Mathematics with Applications 68, 962–971 (2014). https://doi.org/10.1016/j.camwa.2014.03.006

78. Dimitrova, Z.I.: Relation between G'/G-expansion method and the modified method of simplest equation. Compt. rend. Acad. bulg. Scie 65, 1513–1520 (2012)

79. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Modified method of simplest equation for obtaining exact analytical solutions of nonlinear partial differential equations: further development of the methodology with applications. Applied Mathematics and Computation 269, 363–378 (2015). https://doi.org/10.1016/j.amc.2015.07.060

80. Vitanov, N. K., Dimitrova, Z. I., Ivanova, T. I.: On solitary wave solutions of a class of nonlinear partial differential equations based on the function $1/cosh^n$ (αx+ βt). Applied Mathematics and Computation 315, 372–380 (2017). https://doi.org/10.1016/j.amc.2017.07.064

81. Vitanov, N. K., Dimitrova, Z. I.: On the modified method of simplest equation and the nonlinear Schrödinger equation. Journal of Theoretical and Applied Mechanics 48, 59–68 (2018)

82. Nikolova, E. V., Jordanov, I. P., Dimitrova, Z. I., Vitanov, N. K.: Evolution of nonlinear waves in a blood-filled artery with an aneurysm. AIP Conference Proceedings vol. 1895, 070002 (2017). https://doi.org/10.1063/1.5007391

83. Jordanov, I.P., Vitanov, N.K.: On the Exact Traveling Wave Solutions of a Hyperbolic Reaction-Diffusion Equation. In: Georgiev, K., Todorov, M., Georgiev, I. (eds.) Advanced Computing in Industrial Mathematics. BGSIAM 2017. Studies in Computational Intelligence, vol. 793, pp. 199–210. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-97277-0_16

84. Nikolova, E. V., Chilikova-Lubomirova, M., Vitanov, N. K.: Exact solutions of a fifth-order Korteweg-de Vries-type equation modeling nonlinear long waves in several natural phenomena. AIP Conference Proceedings vol. 2321, 030026 (2021). https://doi.org/10.1063/5.0040089

85. Dimitrova, Z. I.: Several examples of application of the simple equations method (SEsM) for obtaining exact solutions of nonlinear PDEs. AIP Conference Proceedings vol. 2459, 030005 (2022). https://doi.org/10.1063/5.0083572

86. Dimitrova, Z. I.: On several specific cases of the simple equations method (SEsM): Jacobi elliptic function expansion method, F-expansion method, modified simple equation method,

trial function method, general projective Riccati equations method, and first intergal method. AIP Conference Proceedings, vol. 2459, 030006 (2022). https://doi.org/10.1063/5.0083573

87. Vitanov, N. K., Dimitrova, Z. I.: Simple Equations Method and non-linear differential equations with non-polynomial non-linearity Entropy 23, 1624 (2021). https://doi.org/10.3390/e23121624

An Example for Application of the Simple Equations Method for the Case of Use of a Single Simple Equation

Zlatinka I. Dimitrova

Abstract We discuss the application of the methodology of the Simple Equations Method. Discussion is focused on an example where the methodology is applied on the basis of a single simple equation. We explain in detail each step of the methodology and discuss the obtained exact solutions of the solved nonlinear partial differential equation.

Keywords SEsM · Methodology based on one simple equation

1 Introduction

An essential feature of many complex systems is their nonlinearity [1–18]. The corresponding effects are studied, for an example, by the methods of the time series analysis and/or by models based on differential or difference equations [19–36]. There are various methods for obtaining exact solutions of nonlinear differential equations. We mention the Hopf-Cole transformation [37, 38]. It transforms the nonlinear Burgers equation to the linear heat equation. Well-known methods for obtaining soliton solutions of classes of nonlinear differential equations are the Method of Inverse Scattering Transform [39, 40] and the method of Hirota [41].

Below we apply the SEsM (Simple Equations Method) for obtaining exact solutions of nonlinear differential equations. The line of research resulting in SEsM begins with the work of Kudryashov and then Kudryashov and Loguinova [42, 43]. The development of SEsM [44–47] started many years ago [48–56]. This research was continued in 2009 [57, 58] and in 2010, when the ordinary differential equation of Bernoulli was used as simplest equation [59] with application to ecology and population dynamics [60]. SEsM uses balance equations [61, 62] to determine the kind of the simplest equations and truncation of the series of solutions of the simplest

Z. I. Dimitrova (✉)

Institute of Mechanics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 4, 1113 Sofia, Bulgaria

e-mail: zdim@imbm.bas.bg

25

equation [63–69]. An important milestone in the development of the methodology was [70] where the methodology was extended to simplest equations of the class $\left(\frac{d^k g}{d\xi^k}\right)^l = \sum_{j=0}^{m} d_j g^j$, where $k = 1, \ldots, l = 1, \ldots,$ and m and d_j are parameters. In addition, in [70] it was mentioned that the composite functions and their derivatives play important role in the methodology. In the last years, the methodology was extended to SEsM—Simple Equations Method. A version of SEsM based on two simple equations was applied in [71] and the first description of the methodology was made in [44] and then in [45–47]. For more applications of specific cases of the methodology see [72–85].

The goal of this article is to show one example of application of SEsM. These examples will be connected to the use of composite function which is a function of a single simple equation. We discuss the SEsM in Sect. 2 and supply the needed information for the use of derivatives of composite functions in SEsM. The example is shown in Sect. 3. Section 4 summarizes some concluding remarks.

2   Methodology of the Method of Simple Equations (SEsM)

Below we solve a single differential equation. Because of this we present the version of the method for solving a single nonlinear differential equation (Fig. 1).

We consider the nonlinear differential equation $\mathcal{P}[u(x, \ldots, t), \ldots,] = 0$. $\mathcal{P}[u(x, \ldots, t), \ldots]$ depends on the function $u(x, \ldots, t)$ and some of their derivatives (u_i can be a function of more than 1 spatial coordinates). Step 1 of SEsM is to perform transformations $u(x, \ldots, t) = T[F(x, \ldots, t), G(x, \ldots, t), \ldots]$.. Here $T(F, G, \ldots)$ is some function of another functions $F, G, \ldots$. In general $F(x, \ldots, t),$ $G(x, \ldots, t), \ldots$ are functions of several spatial variables as well as of the time. The
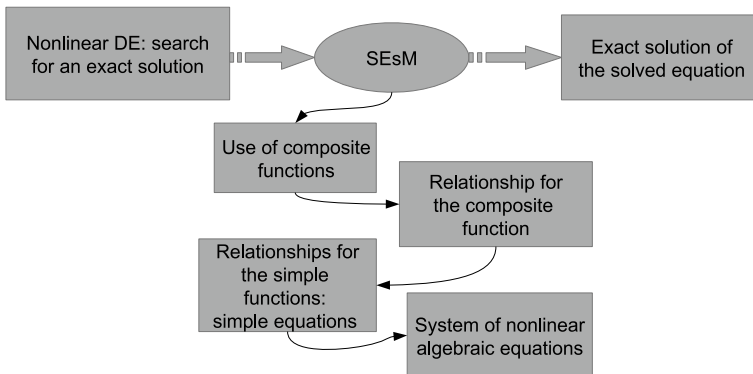
Fig. 1 The example of SEsM with use of composite functions

goal of the transformations is to transform the nonlinearity of the solved differential equations to more treatable kind of nonlinearity.

The nonlinearities in the solved equations can be of different kinds. Because of this there is no universal transformation of the nonlinearity which is valid for all possible cases. The application of the transformation T to the solved equation leads to nonlinear PDEs for the functions $F, G, \ldots$.

In Step 2 of SEsM the functions $F(x, \ldots, t), G(x, \ldots, t), \ldots$ are represented as a function of other functions $f_1, \ldots, f_N, g_1, \ldots, g_M, \ldots$, which are connected to solutions of some differential equations that are more simple than the solved equation. F_i and G_j become composite functions and we have to use the Faa-di-Bruno formula in order to write the derivatives of these functions [86].

Step 3 of SEsM is connected to the representation of the functions used in $F, G, \ldots$—the functions $f_1, \ldots, f_N, g_1, \ldots, g_M$ are solutions of some differential equations. These equations are more simple than the solved nonlinear partial differential equation. If the simple equations contain only polynomial nonlinearities then their form can be determined by balance equations. Balance equations may be needed in order to ensure that the system of algebraic equations from Step 4 contains a nontrivial solution.

At Step 4 of SEsM we apply the steps 1–3 to the solved equation. In the most cases the result of this are functions which are sum of terms where each term contains some function multiplied by a coefficient. We can obtain a nontrivial solution of the solved equation if all coefficients mentioned in Step 4 are set to 0. This usually leads to a system of nonlinear algebraic equations for the coefficients of the solved nonlinear PDE and for the coefficients of the solution. Any nontrivial solution of this algebraic system leads to a solution to the studied nonlinear partial differential equation.

3   An Example of Use of Composite Functions in the Methodology of SEsM

We consider an example where a composite function will be used to find an appropriate transformation that transforms the solved nonlinear equation to an equation for which we can easily find simple equations. The solved nonlinear differential equation is

$$uu_t + u^3 u_x + p u_x^2 + q u u_{xx} = 0 \tag{1}$$

We are going to transform the nonlinearity in (1) by means of a transformation which is a simple composite function of the kind $u(x, t) = u\,[\phi(x, t), \phi_x(x, t)]$ .. The substitution of this relationship in (1) leads to the equation

$$uu_\phi(\phi_t + q\phi_{xx}) + uu_{\phi_x}(\phi_t + q\phi_{xx})_x + u^3(u_\phi \phi_x + u_{\phi_x}\phi_{xx}) +$$
$$p(u_\phi \phi_x + u_{\phi_x}\phi_{xx})^2 + qu(u_{\phi\phi}\phi_x^2 + u_{\phi_x\phi_x}\phi_{xx}^2) = 0 \tag{2}$$

Equation (2) can be decomposed again into two equations

$$u^3(u_\phi \phi_x + u_{\phi_x} \phi_{xx}) + p(u_\phi \phi_x + u_{\phi_x} \phi_{xx})^2 + qu(u_{\phi\phi} \phi_x^2 + u_{\phi_x \phi_x} \phi_{xx}^2) = 0 \quad (3)$$

$$\phi_t + q\phi_{xx} = 0 \tag{4}$$

Equation (4) can be treated as simplest equation again and (3) is an equation for the transformation.

Equation (3) has the following solution

$$u(\phi, \phi_x) = \left(r \frac{\phi_x}{\phi} \right)^\pi, \quad \pi = 1/2, \quad r = -2\nu \tag{5}$$

and in addition $p = q = -\nu$. We assume $\tau = qt$ and the solution of (5) is

$$\phi(x, \tau) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} d\lambda \int\limits_{-\infty}^{\infty} d\xi \ f(\xi) \ \cos[\lambda(x - \xi)] \ \exp(-\lambda^2 \tau), \tag{6}$$

where the initial condition is $\phi \mid_{\tau=0} = f(x)$ (Cauchy problem).

By means of the solution (6) of the simple equation (4) we obtain the solution

$$u(x, t) = \left\{ -2\nu \frac{\frac{\partial}{\partial x} \left[\int\limits_{-\infty}^{\infty} d\lambda \int\limits_{-\infty}^{\infty} d\xi \ f(\xi) \ \cos[\lambda(x - \xi)] \ \exp(-\lambda^2 \tau) \right]}{\int\limits_{-\infty}^{\infty} d\lambda \int\limits_{-\infty}^{\infty} d\xi \ f(\xi) \ \cos[\lambda(x - \xi)] \ \exp(-\lambda^2 \tau)} \right\}^{1/2} \tag{7}$$

4   Concluding Remarks

In this article we discuss the application of the Simple Equations Method (SEsM) for obtaining exact solutions of nonlinear differential equations. We present a formulation of the methodology based on composite function and on the possibility for use of more than one simple equation. The composite functions of the methodology can be used in principle when one: (a) searches for the appropriate transformation in order to transform the nonlinearity of the equation; (b) constructs the solution of the equation by means of the solution of more simple equations; (c) searches for solution of the more simple equation (we note that the more simple equation can be quite complicated). We presented an example where a separate equation is available for the transformation which has to transform the nonlinearity of the equation. The solution of this equation for the transformation leads to the possibility that the solution of the searched nonlinear equation can be constructed on the basis of solution of a simple equation which is a linear partial differential equation. This allows us to

present a solution of the Cauchy problem for the solved nonlinear partial differential equation.

Our goal in this text was just to present the methodology of SEsM and to illustrate its use. The methodology has many applications of the kind which has been demonstrated in the example above. We will discuss these applications in our future work.

References

1. Drazin. P.G.: Nonlinear Systems. Cambridge University Press, Cambridge, UK (1992)
2. Lambiotte, R., Ausloos, M.: Coexistence of opposite opinions in a network with communities. Journal of Statistical Mechanics: Theory and Experiment, P08026 (2007). https://doi.org/10.1088/1742-5468/2007/08/P08026
3. Vitanov, N., Busse, F.: Bounds on the heat transport in a horizontal fluid layer with stress-free boundaries. Z. angew. Math. Phys. 48, 310–324 (1997). https://doi.org/10.1007/PL00001478
4. Kutner, R., Ausloos, M., Grech, D., Di Matteo, T., Schinckus, C., Stanley, H. E.: Econophysics and sociophysics: Their milestones & challenges. Physica A 516, 240–253 (2019). https://doi.org/10.1016/j.physa.2018.10.019
5. Vitanov, N. K., Sakai, K., Jordanov, I. P., Managi, S., Demura, K.: Analysis of a Japan government intervention on the domestic agriculture market. Physica A: 382, 330–335 (2007). https://doi.org/10.1016/j.physa.2007.02.025
6. Sheard, S. A., Mostashari, A.: Principles of complex systems for systems engineering. Systems Engineering 12, 295–311 (2009). https://doi.org/10.1002/sys.20124
7. Vitanov N.K.: Science Dynamics and Research Production. Indicators, Indexes, Statistical Laws and Mathematical Models. Springer, Cham (2016)
8. Bahrami, M., Chinichian, N., Hosseiny, A., Jafari, G., Ausloos, M.: Optimization of the post-crisis recovery plans in scale-free networks. Physica A 540, 123203 (2020). https://doi.org/10.1016/j.physa.2019.123203
9. Vitanov, N. K.: Upper bound on the heat transport in a horizontal fluid layer of infinite Prandtl number. Physics Letters A 248, 338–346 (1998). https://doi.org/10.1016/S0375-9601(98)00674-4
10. Lempert, R. J.: A new decision sciences for complex systems. Proceedings of the National Academy of Sciences 99 (suppl 3), 7309–7313 (2002). https://doi.org/10.1073/pnas.082081699
11. Vitanov, N. K., Chabchoub, A., Hoffmann N.: Deep-Water Waves: on the Nonlinear Schrödinger Equation and its Solutions. Journal of Theoretical and Applied Mechanics 43, 43–54 (2013). https://doi.org/10.2478/jtam-2013-0013
12. Nikolova, E. V., Vitanov, N. K.: On the Possibility of chaos in a ceneralized model of three interacting sectors. Entropy 22, 1388 (2020). https://doi.org/10.3390/e22121388
13. Vitanov, N. K., Ausloos, M., Rotundo, G.: Discrete model of ideological struggle accounting for migration. Advances in Complex Systems 15, 1250049 (2012). https://doi.org/10.1142/S021952591250049X
14. Vitanov, N. K., Vitanov, K. N.: Box model of migration channels. Mathematical Social Sciences 80, 108–114 (2016). https://doi.org/10.1016/j.mathsocsci.2016.02.001
15. Vitanov, N. K., Vitanov, K. N.: On the motion of substance in a channel of a network and human migration. Physica A: 490, 1277–1294 (2018). https://doi.org/10.1016/j.physa.2017.08.038

16. Amaral, L. A. N., Scala, A., Barthelemy, M., Stanley, H. E.: Classes of small-world networks. PNAS USA 97, 11149–11152 (2000). https://doi.org/10.1073/pnas.20032719

17. Vitanov, N. K., Vitanov, K. N., Kantz, H.: On the motion of substance in a channel of a network: Extended model and new classes of probability distributions. Entropy 22, 1240 (2020). https://doi.org/10.3390/e22111240

18. Larsen-Freeman, D., Cameron, L.: Complex systems and applied linguistics. Oxford: Oxford University Press (2008)

19. Kantz, H., T. Schreiber, T.: Nonlinear Time Series Analysis. Cambridge University Press, Cambridge, UK (2004)

20. Struble R.: Nonlinear Differential Equations. Dover, New York (2018)

21. Jordanov, I., Nikolova, E.: On nonlinear waves in the spatio-temporal dynamics of interacting populations. Journal of Theoretical and Applied Mechanics 43, 69–76 (2013). arXiv preprint arXiv:1208.5465

22. Boeck, T., Vitanov, N. K.: Low-dimensional chaos in zero-Prandtl-number Benard-Marangoni convection. Physical Review E 65, 037203 (2002). https://doi.org/10.1103/PhysRevE.65.037203

23. Jordanov, I. P.: On the nonlinear waves in (2+ 1)-dimensional population systems. Comptes rendus de l'Académie bulgare des Sciences 61, 307–314 (2008)

24. Vitanov, N.K.: Results Connected to Time Series Analysis and Machine Learning. In: Atanassov, K.T. (eds) Research in Computer Science in the Bulgarian Academy of Sciences. Studies in Computational Intelligence, vol 934, pp. 363–384. Springer, Cham. (2021). https://doi.org/10.1007/978-3-030-72284-5_17

25. Brockwell P.J., Davis R.A, Calder M.V.: Introduction to Time Series and Forecasting. Springer, New York (2002)

26. Vitanov, N. K., Ausloos, M. R.: Knowledge epidemics and population dynamics models for describing idea diffusion. In: Scharnhorst A., Boerner K., van den Besselaar P. (eds.) Models of science dynamics. Understanding Complex Systems. pp. 69–125. Springer, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-23068-4_3

27. Kantz, H., Holstein, D., Ragwitz, M., Vitanov, N. K.: Markov chain model for turbulent wind speed data. Physica A 342, 315–321 (2004). https://doi.org/10.1016/j.physa.2004.01.070

28. Ashenfelter, K. T., Boker, S. M., Waddell, J. R., Vitanov, N.: Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation. Journal of Experimental Psychology: Human Perception and Performance 35, 1072–1091 (2009). https://doi.org/10.1037/a0015017

29. Vitanov, N. K.: Upper bounds on the heat transport in a porous layer. Physica D 136, 322–339 (2000). https://doi.org/10.1016/S0167-2789(99)00165-7

30. Fuchs A.: Nonlinear Dynamics in Complex Systems. Springer, Berlin (2014)

31. Vitanov, N. K., Hoffmann, N. P., Wernitz, B.: Nonlinear time series analysis of vibration data from a friction brake: SSA, PCA, and MFDFA. Chaos, Solitons & Fractals 69, 90–99 (2014). https://doi.org/10.1016/j.chaos.2014.09.010

32. Goldstein, J.: Social psychology and nonlinear dynamical systems theory. Psychological Inquiry 8, 125–128 (1997). https://doi.org/10.1207/s15327965pli0802_6

33. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: On the trap of extinction and its elimination. Physics Letters A 349, 350–355 (2006). https://doi.org/10.1016/j.physleta.2005.09.043

34. Borisov, R., Dimitrova, Z. I., Vitanov, N. K.: Statistical characteristics of stationary flow of substance in a network channel containing arbitrary number of arms. Entropy 22, 553 (2020). https://doi.org/10.3390/e22050553

35. Vitanov, N. K., Vitanov, K. N.: Statistical distributions connected to motion of substance in a channel of a network. Physica A 527, 121174 (2019). https://doi.org/10.1016/j.physa.2019.121174

36. Vitanov, N. K., Borisov, R., Vitanov, K. N.: On the motion of substance in a channel and growth of random networks. Physica A 581, 126207 (2021). https://doi.org/10.1016/j.physa.2021.126207

37. Hopf E.: The partial differential equation $u_t + uu_x = u_{xx}$. Communications on Pure and Applied Mathematics 3, 201–230 (1950). https://doi.org/10.1002/cpa.3160030302
38. Cole, J. D.: On a quasi-linear parabolic equation occurring in aerodynamics. Quarterly of applied mathematics 9, 225–236 (1951). https://doi.org/10.1090/QAM/42889
39. Ablowitz, M. J., Clarkson, P. A.: Solitons, Nonlinear Evolution Equations and Inverse Scattering. Cambridge University Press, Cambridge, UK (1991)
40. Gardner, C. S., Greene, J. M., Kruskal, M. D., Miura, R. M.: Method for solving the Korteweg-deVries equation. Physical review letters, 19(19), 1095–1897 (1967). https://doi.org/10.1103/PhysRevLett.19.1095
41. Hirota, R.: The Direct Method in Soliton Theory. Cambridge University Press, Cambridge, UK, (2004)
42. Kudryashov, N. A.: Simplest equation method to look for exact solutions of nonlinear differential equations. Chaos, Solitons & Fractals 24, 1217–1231 (2005). https://doi.org/10.1016/j.chaos.2004.09.109
43. Kudryashov, N. A., Loguinova, N. B.: Extended simplest equation method for nonlinear differential equations. Applied Mathematics and Computation 205, 396–402 (2008). https://doi.org/10.1016/j.amc.2008.08.019
44. Vitanov, N. K.: Recent developments of the methodology of the modified method of simplest equation with application. Pliska Studia Mathematica Bulgarica 30, 29–42 (2019)
45. Vitanov, N.K.: Modified method of simplest equation for obtaining exact solutions of nonlinear partia differential equations: history, recent developments of the methodology and studied of classes of equations. Journal of Theoretical and Applied Mechanics 49, 107–122 (2019)
46. Vitanov, N. K.: The simple equations method (SEsM) for obtaining exact solutions of nonlinear PDEs: Opportunities connected to the exponential functions. AIP Conference Proceedings vol. 2159, 030038 (2019). https://doi.org/10.1063/1.5127503
47. Vitanov, N. K., Dimitrova, Z. I.: Simple equations method (SEsM) and other direct methods for obtaining exact solutions of nonlinear PDEs. AIP Conference Proceedings vol. 2159, 030039 (2019). https://doi.org/10.1063/1.5127504
48. Martinov, N., Vitanov, N.: On the correspondence between the self-consistent 2D Poisson-Boltzmann structures and the sine-Gordon waves. Journal of Physics A: Mathematical and General 25, L51–L56 (1992). https://doi.org/10.1088/0305-4470/25/2/004
49. Martinov, N., Vitanov, N.: On some solutions of the two-dimensional sine-Gordon equation. Journal of Physics A: Mathematical and General 25, L419–L426 (1992). https://doi.org/10.1088/0305-4470/25/8/007
50. Martinov, N., Vitanov, N.: Running wave solutions of the two-dimensional sine-Gordon equation. Journal of Physics A: Mathematical and General 25, 3609–3613 (1992). https://doi.org/10.1088/0305-4470/25/12/021
51. Martinov, N. K., Vitanov, N. K.: New class of running-wave solutions of the (2+1)-dimensional sine-Gordon equation. Journal of Physics A: Mathematical and General 27, 4611–4618 (1994). https://doi.org/10.1088/0305-4470/27/13/034
52. Martinov, N. K., Vitanov, N. K.: On self-consistent thermal equilibrium structures in two-dimensional negative-temperature systems. Canadian Journal of Physics 72, 618–624 (1994). https://doi.org/10.1139/p94-079
53. Vitanov, N. K., Martinov, N. K.: On the solitary waves in the sine-Gordon model of the two-dimensional Josephson junction. Zeitschrift fur Physik B Condensed Matter 100, 129–135 (1996). https://doi.org/10.1007/s002570050102
54. Vitanov, N. K.: On travelling waves and double-periodic structures in two-dimensional sine-Gordon systems. Journal of Physics A: Mathematical and General 29, 5195–5207 (1996). https://doi.org/10.1088/0305-4470/29/16/036
55. Vitanov, N. K.: Complicated exact solutions to the 2+ 1-dimensional sine-Gordon equation. Zeitschrift für angewandte Mathematik und Mechanik, 78, S787–S788 (1998)
56. Vitanov, N. K.: Breather and soliton wave families for the sine-Gordon equation. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454, 2409–2423 (1998). https://doi.org/10.1098/rspa.1998.0264

57. Vitanov, N. K., Jordanov, I. P., Dimitrova, Z. I.: On nonlinear dynamics of interacting populations: Coupled kink waves in a system of two populations. Communications in Nonlinear Science and Numerical Simulation 14, 2379–2388 (2009). https://doi.org/10.1016/j.cnsns.2008.07.015

58. Vitanov, N. K., Jordanov, I. P., Dimitrova, Z. I.: On nonlinear population waves. Applied Mathematics and Computation 215, 2950–2964 (2009). https://doi.org/10.1016/j.amc.2009.09.041

59. Vitanov, N. K.: Application of simplest equations of Bernoulli and Riccati kind for obtaining exact traveling-wave solutions for a class of PDEs with polynomial nonlinearity. Communications in Nonlinear Science and Numerical Simulation 15, 2050–2060 (2010). https://doi.org/10.1016/j.cnsns.2009.08.011

60. Vitanov, N. K., Dimitrova, Z. I.: Application of the method of simplest equation for obtaining exact traveling-wave solutions for two classes of model PDEs from ecology and population dynamics. Communications in Nonlinear Science and Numerical Simulation 15, 2836–2845 (2010). https://doi.org/10.1016/j.cnsns.2009.11.029

61. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: Modified method of simplest equation and its application to nonlinear PDEs. Applied Mathematics and Computation 216, 2587–2595 (2010). https://doi.org/10.1016/j.amc.2010.03.102

62. Vitanov, N. K.: Modified method of simplest equation: powerful tool for obtaining exact and approximate traveling-wave solutions of nonlinear PDEs. Communications in Nonlinear Science and Numerical Simulation 16, 1176–1185 (2011). https://doi.org/10.1016/j.cnsns.2010.06.011

63. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: On the class of nonlinear PDEs that can be treated by the modified method of simplest equation. Application to generalized Degasperis-Processi equation and b-equation. Communications in Nonlinear Science and Numerical Simulation 16, 3033–3044 (2011). https://doi.org/10.1016/j.cnsns.2010.11.013

64. Vitanov, N. K.: On modified method of simplest equation for obtaining exact and approximate solutions of nonlinear PDEs: the role of the simplest equation. Communications in Nonlinear Science and Numerical Simulation 16, 4215–4231 (2011). https://doi.org/10.1016/j.cnsns.2011.03.035

65. Vitanov, N. K.: On modified method of simplest equation for obtaining exact solutions of nonlinear PDEs: case of elliptic simplest equation. Pliska Studia Mathematica Bulgarica 21, 257–266 (2012)

66. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: Application of the method of simplest equation for obtaining exact traveling-wave solutions for the extended Korteweg-de Vries equation and generalized Camassa-Holm equation. Applied Mathematics and Computation 219, 7480–7492 (2013). https://doi.org/10.1016/j.amc.2013.01.035

67. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Traveling waves and statistical distributions connected to systems of interacting populations. Computers & Mathematics with Applications 66, 1666–1684 (2013). https://doi.org/10.1016/j.camwa.2013.04.002

68. Vitanov, N. K., Dimitrova, Z. I.: Solitary wave solutions for nonlinear partial differential equations that contain monomials of odd and even grades with respect to participating derivatives. Applied Mathematics and Computation 247, 213–217 (2014). https://doi.org/10.1016/j.amc.2014.08.101

69. Vitanov, N. K., Dimitrova, Z. I., Ivanova, T. I.: On solitary wave solutions of a class of nonlinear partial differential equations based on the function $1/cosh^n$ (αx+ βt). Applied Mathematics and Computation 315, 372–380 (2017). https://doi.org/10.1016/j.amc.2017.07.064

70. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Modified method of simplest equation for obtaining exact analytical solutions of nonlinear partial differential equations: further development of the methodology with applications. Applied Mathematics and Computation 269, 363–378 (2015). https://doi.org/10.1016/j.amc.2015.07.060

71. Vitanov, N. K., Dimitrova, Z. I.: On the modified method of simplest equation and the nonlinear Schrödinger equation. Journal of Theoretical and Applied Mechanics 48, 59–68 (2018)

72. Nikolova, E. V., Jordanov, I. P., Dimitrova, Z. I., Vitanov, N. K.: Evolution of nonlinear waves in a blood-filled artery with an aneurysm. AIP Conference Proceedings vol. 1895, 070002 (2017). https://doi.org/10.1063/1.5007391

73. Jordanov, I.P., Vitanov, N.K.: On the Exact Traveling Wave Solutions of a Hyperbolic Reaction-Diffusion Equation. In: Georgiev, K., Todorov, M., Georgiev, I. (eds.) Advanced Computing in Industrial Mathematics. BGSIAM 2017. Studies in Computational Intelligence, vol. 793, pp. 199–210. Springer, Cham. (2019). https://doi.org/10.1007/978-3-319-97277-0_16

74. Nikolova, E. V., Chilikova-Lubomirova, M., Vitanov, N. K.: Exact solutions of a fifth-order Korteweg-de Vries-type equation modeling nonlinear long waves in several natural phenomena. AIP Conference Proceedings vol. 2321, 030026 (2021). https://doi.org/10.1063/5.0040089

75. Vitanov, N. K.: Simple equations method (SEsM) and its connection with the inverse scattering transform method. AIP Conference Proceedings vol. 2321, 030035 (2021). https://doi.org/10.1063/5.0040409

76. Nikolova, E.V., Serbezov, D.Z., Jordanov, I.P., Vitanov, N.K.: Non-linear Waves of Interacting Populations with Density-Dependent Diffusion. In: Georgiev, I., Kostadinov, H., Lilkova, E. (eds.) Advanced Computing in Industrial Mathematics. BGSIAM 2018. Studies in Computational Intelligence, vol. 961, pp. 324–332. Springer, Cham. (2021). https://doi.org/10.1007/978-3-030-71616-5_29

77. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Simple Equations Method (SEsM): Algorithm, connection with Hirota method, Inverse Scattering Transform Method, and several other methods. Entropy 23 10 (2021). https://doi.org/10.3390/e23010010

78. Vitanov. N.K.: Schrödinger Equation and Nonlinear Waves. Simpao V., Little H. (eds.). In: Understanding the Schrödinger Equation. pp. 37–92. Nova Science Publishers, New York (2020)

79. Vitanov, N. K., Dimitrova, Z. I.: Simple Equations Method and non-linear differential equations with non-polynomial non-linearity Entropy 23, 1624 (2021). https://doi.org/10.3390/e23121624

80. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: On the use of composite functions in the Simple Equations Method to obtain exact solutions of nonlinear differential equations. Computation 9, 104 (2021). https://doi.org/10.3390/computation9100104

81. Vitanov, N. K., Dimitrova, Z. I.: Simple equations method (SEsM) and its particular cases: Hirota method. AIP Conference Proceedings vol. 2321, 030036 (2021). https://doi.org/10.1063/5.0040410

82. Vitanov, N. K.: Simple equations method (SEsM) and nonlinear PDEs with fractional derivatives. AIP Conference Proceedings vol. 2459, 030040 (2022). https://doi.org/10.1063/5.0083566

83. Dimitrova, Z. I.: Several examples of application of the simple equations method (SEsM) for obtaining exact solutions of nonlinear PDEs. AIP Conference Proceedings vol. 2459, 030005 (2022). https://doi.org/10.1063/5.0083572

84. Vitanov, N. K.: Simple equations method (SEsM): Review and new results. AIP Conference Proceedings, vol. 2459, 020003 (2022). https://doi.org/10.1063/5.0083565

85. Dimitrova, Z. I.: On several specific cases of the simple equations method (SEsM): Jacobi elliptic function expansion method, F-expansion method, modified simple equation method, trial function method, general projective Riccati equations method, and first intergal method. AIP Conference Proceedings, vol. 2459, 030006 (2022). https://doi.org/10.1063/5.0083573

86. Constantine, G.M., Savits, T.H.: A multivariate Faa di Bruno formula with applications. Transactions of the American Mathematical Society 348, 503–520 (1996). https://doi.org/10.1090/S0002-9947-96-01501-2

Boundary Value Problems for the Polyharmonic Operators

Petar Popivanov and Angela Slavova

Abstract This paper deals with Dirichlet-type elliptic and degenerate oblique derivative boundary value problem for the polyharmonic operator Δ^m in the ball B_1 and in the half space. Pointwise estimations from above and from below for the Green function to the homogeneous Dirichlet problem for $(-\Delta)^m u = f$ in B_1 are found and applied to the corresponding solutions u in the case $f \geq 0$. Similar results are obtained for the Navier problem with homogeneous data on $S_1 = \partial B_1$. Solutions in closed (explicit) form for the equation of $\Delta^m u = 0$ in B_1 with Lopatinskii-type boundary conditions are constructed too. By using the subelliptic estimates and the hypoellipticity of several elliptic degenerate PDO in Examples 1–4, boundary value problems violating Lopatinskii conditions are studied in $\mathbf{R}_+^{n+1}$. They turn out to be hypoelliptic and Fredholm ones.

Keywords Polyharmonic operator · Dirichlet problem · Degenerate oblique derivative problem · Lopatinskii condition · Navier problem

1 Introduction

This paper deals with Dirichlet-type and degenerate oblique derivative-type boundary value problems for the polyharmonic operator Δ^m in the ball and in the half space. Certainly, the boundary value problems for elliptic operators satisfying the Lopatinskii condition are well studied both in Hölder and Sobolev spaces [6, 7, 9]. There are many papers and monographs devoted to the Dirichlet, Navier, and mixed Dirichlet–Navier boundary value problems for the polyharmonic equation and its generalizations, including the construction of Green functions [1, 5]. Here

P. Popivanov · A. Slavova (✉)

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia 1113, Bulgaria
e-mail: slavova@math.bas.bg

we concentrate on the explicit formulas for the solutions of several boundary value problems and more specially on the pointwise estimations from above and below of the solutions satisfying Lopatinskii-type boundary conditions in a ball (Propositions 1 and 2). By using some results of Zhao [14] we can obtain very precise result for the Green function in the ball of the Poisson equation with homogeneous Dirichlet data on the sphere. It is worth mentioning that the solution of the Dirichlet problem for the equation $\Delta^m u = f$ in B_1 with zero data on the unit sphere S_1 is located between two convex rotational cones with $\overrightarrow{0u}$ axes and having in the base B_1 zero of sharp order m on S_1. Similar result is valid if we replace "cone" by "smooth rotational convex–concave surface" tangential of sharp order m to S_1. Further on, we propose short comment on the subelliptic estimates for first-order differential (pseudo-differential) operators (see [4, 7, 10]) and study in the half space $x_{n+1} \geq 0$, $n \geq 2$ different boundary value problems violating Lopatinskii condition. The corresponding examples 1, 2, 3, 4 deal with hypoelliptic boundary value problems for the biharmonic and three harmonic operators. Concerning the notion of hypoellipticity see [7]. In the case of $\Delta^3 u = 0$, $x \in \mathbf{R}_+^{n+1}$, we find out subelliptic estimates for the solutions of the boundary value problem. Assuming that $u \in H_{loc}^{s_0}(\mathbf{R}_+^{n+1})$, $s_0 > 2m - 1/2$ ($s_0 > m + 1/2$) and the boundary data are $C_0^\infty(\mathbf{R}^n)$ smooth we get that $u \in C^\infty(\overline{\mathbf{R}_+^{n+1}})$ in the hypoelliptic case.

We write down the explicit formulas for the solutions of the Dirichlet and Neumann problems for the Laplace equation in half space. Applying Faa di Bruno formula [13], we obtain an explicit integral formula for the polyharmonic functions in $\mathbf{R}_+^{n+1}$ (see Proposition 4). It enables us to find the traces $u|_{x_{n+1}=0}$, ..., $D_n^j u|_{x_{n+1}=0}$ for $j = 0, \ldots, m - 1$.

In Proposition 3, we study a boundary value problem of elliptic (Fredholm) type for $\Delta^m u = 0$ in a ball and reduce it to the solvability of m-Dirichlet problems for the Laplace equation in the unit ball B_1. This way formulas can be written into explicit (closed) form for the polyharmonic operator equipped with Lopatinskii-type boundary conditions given on the unit sphere S_1.

The paper is organized as follows. Section 1 contains the introduction. In Sect. 2, we formulate our main results, illustrative examples and comments. The proofs are given in Sect. 3, while Sect. 4 contains the References. This paper can be considered as a further development of our investigations in [11, 12] on the biharmonic operator in B_1 in the case of non-Lopatinskii (not complementing in other terminology) boundary conditions. We put here the stress on the polyharmonic operators and on the explicit formulas for the solutions in the half space and in the ball. Integral representations of the solutions of the Dirichlet problem in the unit ball for $\Delta^m u = 0$ can be found in [3].

2 Formulation of the Main Results, Illustrative Examples, and Comments

At first we consider the Dirichlet boundary value problem

$$
\left|
\begin{array}{l}
(-\Delta)^m u = f \text{ in } B_1 \text{ in } \mathbf{R}^n, |x| < 1, f \in C^0(\bar{B}_1) \\
D^\alpha u|_{\partial B_1} = 0, |\alpha| \le m - 1, S_1 = \partial B_1.
\end{array}
\right.
\tag{1}
$$

One can easily see that the boundary condition (1) is equivalent to $u|_{S_1} = \ldots = \frac{\partial^{m-1} u}{\partial n^{m-1}}|_{S_1} = 0$, where $\frac{\partial}{\partial n}|_{S_1} = \frac{\partial}{\partial r}|_{S_1}$ is the unit outward normal to S_1, $r = |x|$.

Boggio proved in 1905 in [1] that

$$
u(x) = \int_{B_1} G_{(-\Delta)^m, B}(x, y) f(y) dy,
\tag{2}
$$

where the Green function $G_{m,n} = G_{(-\Delta)^m, B}$ under the integral sign is given by

$$
G_{m,n}(x, y) = k_{m,n} |x - y|^{2m-n} \int_1^{[XY]/|x-y|} (v^2 - 1)^{m-1} v^{1-n} dv > 0
\tag{3}
$$

and the constant $k_{m,n} = \frac{\pi^{-n/2} \Gamma(\frac{n}{2})}{2 \cdot 4^{m-1} [(m-1)!]^2}$, $\Gamma(x)$ being the Euler gamma function.

As it is shown in [5]

$$
G_{m,n}(x, y) \simeq
\begin{cases}
|x - y|^{2m-n} min(1, \frac{d^m(x) d^m(y)}{|x-y|^{2m}}), n > 2m \\
log(1 + \frac{d^m(x) d^m(y)}{|x-y|^{2m}}), n = 2m \\
d^{m-\frac{n}{2}}(x) d^{m-\frac{n}{2}}(y) min(1, \frac{d^{\frac{n}{2}}(x) d^{\frac{n}{2}}(y)}{|x-y|^n}), n < 2m.
\end{cases}
\tag{4}
$$

As usual, $d(x) = dist(x, 0) = 1 - |x|, x \in B_1, [XY] = ||x|y - \frac{x}{|x|}|; [XY]^2 = |x - y|^2 + (1 - |x|^2)(1 - |y|^2) > 0, \forall x, y \in B_1$, $d(x)$ being Lipschitz function in $\bar{B}_1$, $d(x) \in C^\infty(B_1 \setminus 0)$. Evidently, $[XY]^2 \simeq |x - y|^2 + d(x) d(y)$.

Geometrically, if $x \in B_1 \setminus 0$ and x^* is its inverse point to x with respect to S_1, then $x^* = \frac{x}{|x|^2} \notin \bar{B}_1$. Then $[XY] = |x||y - x^*|, y \in B_1$.

In the special case $m = 1, n \ge 3$ we have that $k_{1,n} = \frac{\pi^{-n/2} \Gamma(\frac{n}{2})}{2}$. Then from (3) we have

$$
0 < G_{1,n}(x, y) = \frac{k_{1,n}}{n - 2} |x - y|^{2-n} (|x - y|^{2-n} - [XY]^{2-n}), x, y \in B_1.
\tag{5}
$$

Proposition 1 *Consider the bvp (1). Then*
  (a) for $m = 1$, $n \geq 3$ the corresponding Green function

$$G_{1,n}(x, y) \geq \frac{k_{1,n}}{2(n-2)} \frac{|x-y|^{2-n}d(x)d(y)}{|x-y|^2 + d(x)d(y)} \geq \frac{k_{1,n}}{n-2} \frac{2^{1-n}}{5} d(x)d(y), x, y \in B_1.$$
(6)

Suppose that $f \geq 0$, $f \not\equiv 0$ in (1). Then

$$\frac{k_{1,n}d(x)2^{1-n}}{5(n-2)} \int_{B_1} d(y)f(y)dy \leq \frac{k_{1,n}d(x)}{2(n-2)} \int_{B_1} \frac{|x-y|^{2-n}d(y)f(y)dy}{|x-y|^2 + d(x)d(y)} \leq u(x)$$
(7)

$$\leq ||f||_{C^0(\bar{B}_1)} \frac{d(x)}{n}, \forall x \in B_1.$$

  (b) There exists a constant $C_{m,n}$ depending on n, m and B_1 only and such that if $f \geq 0$, $f \not\equiv 0$

$$C_{m,n}d^m(x) \int_{B_1} f(y)d^m(y) \leq u(x) \leq \frac{||f||_{C^0(\bar{B}_1)}}{C_m}(1-r^2)^m \leq \frac{||f||_{C^0(\bar{B}_1)}}{C_m} d^m(x)2^m.$$
(8)

The constant $C_m = 2^m m!(n+2m-2)(n+2m-4)\ldots(n+2)n$.

We point out that with appropriate constant $D_{m,n} > 0$ $G_{m,n}(x, y) \geq D_{m,n}d^m(x)$ $d^m(y), \forall x, y \in \bar{B}_1$.

Starting from formula (4) one can easily see that for $n > 2m$, $0 < G_{m,n}(x, y) \leq E_{m,n}\frac{d^m(x)d^m(y)}{|x-y|^n}, x \neq y; x, y \in \overline{B_1}$. Therefore, for $n > 2m$ the double-sided estimate $D_{m,n}d^m(x)d^m(y) \leq G_{m,n}(x, y) \leq E_{m,n}\frac{d^m(x)d^m(y)}{|x-y|^n}, \quad x \neq y; \quad x, y \in \overline{B_1}$ holds and $D_{m,n}, E_{m,n} > 0$ are some absolute constants.

Consider now the Navier problem [5] in B_1:

$$(-\Delta)^m = f \text{ in } B_1, f \in C^0(\bar{B}_1)$$
$$u|_{S_1} = \Delta^m u = \ldots = \Delta^{m-1}u|_{S_1} = 0.$$
(9)

(9) is reduced to m Dirichlet-type bvp for Laplace operator, namely,

$$\begin{vmatrix} -\Delta u = u_1 \\ u|_{S_1} = 0 \end{vmatrix}, \begin{vmatrix} -\Delta u_1 = u_2 \\ u_1|_{S_1} = 0 \end{vmatrix} \cdots \begin{vmatrix} \Delta u_{m-1} = -f \\ u_{m-1}|_{S_1} = 0. \end{vmatrix}$$
(10)

Proposition 2 *Suppose that the right-hand side of (9) is continuous and $f \geq 0$, $f \not\equiv 0$. Then the unique solution of (9) can be estimated from below and above as follows:*

$$C_{m,n}d(x)(\int_{B_1} d^2(y)dy)^{m-1} \int_{B_1} f(z)d(z)dz \leq u(x) \leq \frac{1}{n^m}||f||_{C^0(\bar{B}_1)}d(x), \forall x \in B_1.$$

The constant $C_{m,n}$ is absolute, $\int_{B_1} d^2(y)dy = meas\,S_1 \frac{2}{n(n+1)(n+2)}$, $meas\,S_1 = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})}$.

If we consider the solutions of (1), (9) as smooth surfaces they are located in the case $m = 1$ between two rotational cones $K_{1,2}$ with vertices $k_{1,2}$ at $\overrightarrow{0u}$ axis, while in the case (1), $m \geq 2$ they are located between two convex cones $L_{1,2}$ with vertices $l_{1,2}$ tangential to S_1, respectively, smooth rotational convex–concave surfaces $M_{1,2}$ tangential to S_1 of sharp order m.

Several remarks about the subelliptic estimates in $\mathbf{R}_+^{n+1}$, $n \geq 2$ are given below. Consider the boundary first-order differential operator M in $\{x_{n+1} = 0\}$ with real coefficients

$$M = L(x, \partial_x) + a(x)\frac{\partial}{\partial x_{n+1}}, \mathbf{R}_+^{n+1} = \{(x, x_{n+1}), x = (x_1, \ldots, x_n) \in \mathbf{R}^n, x_{n+1} > 0\},$$
(11)

L-nondegenerate, real vector field, $a \in C^\infty(\mathbf{R}^n)$, $a(x) \geq 0$, ($a(x) \leq 0$) everywhere and $a(x) = 0 \Rightarrow \sum_{j=1}^k |L^j a(x)| > 0$. Evidently, k is even. Assume that

$$\begin{vmatrix} \Delta u = 0, \ \text{in} \ \mathbf{R}_+^{n+1}, u \in H^{s_0}(\mathbf{R}_+^{n+1}), s_0 > \frac{3}{2} \\ Mu|_{x_{n+1}=0} = \varphi(x). \end{vmatrix}$$
(12)

Certainly, (12) is not a Lopatinskii-type boundary value problem. Then $u|_{x_{n+1}=0} = c(x)$ satisfies for $s > \frac{1}{2}$ a first-order pseudo-differential equation and the a priori estimate called subelliptic estimate:

$$||c||_{H^{s+\frac{1}{k+1}}} \leq C_s(||\varphi||_s + ||c||_{s-1}).$$
(13)

Moreover, (13) is taken in $H^{s+\frac{1}{k+1}}_{loc}(\mathbf{R}^n)$ and the solution of the Fredholm-type bvp (12) satisfies the a priori estimate in $\mathbf{R}_+^{n+1}$:

$$||u||_{H^s(\mathbf{R}_+^{n+1})} \leq C(||\varphi||_{s-3/2+\frac{k}{k+1}} + ||u||_{H^{s-1/2}(\mathbf{R}_+^{n+1})}).$$
(14)

The norm of φ is taken in the Sobolev space $H^{s-3/2+\frac{k}{k+1}}(\mathbf{R}^n)$.

We propose a geometrical interpretation of the condition imposed on the vector field M and on the coefficient $a \geq 0$ ($a \leq 0$) everywhere. In fact, $-\frac{\partial}{\partial x_{n+1}}$ is the unit outward normal to the boundary $\mathbf{R}_x^n$ of the half space $\mathbf{R}_+^{n+1}$, the real-valued vector field M is nondegenerate and becomes tangential to $\mathbf{R}_x^n$ at the points x, where $a(x) = 0$. Moreover, M is pointing either outward ($a(x) > 0$) or inward ($a(x) < 0$) to $\mathbf{R}_+^{n+1}$. We point out that $\{a(x) = 0\}$ is not obliged to be smooth submanifold.

Here are several examples of bvp for biharmonic and three harmonic operators which are non-elliptic but of Fredholm type.

Example 1.

$$\left|\begin{array}{l} \Delta^2 u = 0 \text{ in } \mathbf{R}_+^{n+1}, n \geq 2 \\ B_1(u)|_{x_{n+1}=0} = Mu|_{x_{n+1}=0} = \varphi_1 \\ B_2(u)|_{x_{n+1}=0} = \frac{\partial^2 u}{\partial x_{n+1}^2}|_{x_{n+1}=0} = \varphi_2. \end{array}\right. \tag{15}$$

(15) is reduced to a subelliptic pseudo-differential equation for $c_0 = u|_{x_{n+1}=0}$, while $c_1 = \frac{|D|c_0 - |D|^{-1}\varphi_2}{2}$ and $|D|$ has symbol $|\xi| \geq 1$ belonging to $S_{1,0}^1$, $u(x, x_{n+1}) = (2\pi)^{-n} \int e^{ix\xi - x_{n+1}|\xi|} (\hat{c}_0(\xi) + x_{n+1}\hat{c}_1(\xi)) d\xi$. The bvp (15) is hypoelliptic.

Example 2.

$$\left|\begin{array}{l} \Delta^2 u = 0 \text{ in } \mathbf{R}_+^{n+1}, n \geq 2 \\ B_1(u)|_{x_{n+1}=0} = \frac{\partial u}{\partial x_{n+1}}|_{x_{n+1}=0} = \varphi_1 \\ B_2(u)|_{x_{n+1}=0} = g(x)\frac{\partial^2 u}{\partial x_{n+1}^2} + \frac{\partial^2 u}{\partial x_1^2}|_{x_{n+1}=0} = \varphi_2. \end{array}\right. \tag{16}$$

Then c_0 satisfies the hypoelliptic equation on the boundary of the type $g(x)\Delta_x c_0 + \partial_{x_1}^2 c_0 = \psi$ with $g(x) = x_1^2$. The bvp is hypoelliptic one [7].

Example 3.

$$\left|\begin{array}{l} \Delta^2 u = 0 \text{ in } \mathbf{R}_+^{n+1}, n \geq 2 \\ B_1(u)|_{x_{n+1}=0} = \frac{\partial u}{\partial x_{n+1}}|_{x_{n+1}=0} = \varphi_1 \\ B_2(u)|_{x_{n+1}=0} = g(x)\frac{\partial^2 u}{\partial x_{n+1}^2} + A(x)\frac{\partial u}{\partial x_1}|_{x_{n+1}=0} = \varphi_2, \end{array}\right. \tag{17}$$

with $A(x) \in C^\infty$, $A(x) > 0$, $g(x) > 0$ for $|x| > 0$, $g \in C^\infty$.

Then c_0 satisfies on the boundary the inverse parabolic equation $A(x)\partial_1 c_0 + g(x)\Delta_x c_0 = \psi$. Taking $A(x) \equiv 1$ and $g(x)$ flat at the origin we conclude that (17) is hypoelliptic boundary value problem.

Example 4.

$$\left|\begin{array}{l} \Delta^3 u = 0 \text{ in } \mathbf{R}_+^{n+1}, n \geq 2 \\ B_1(u)|_{x_{n+1}=0} = Mu|_{x_{n+1}=0} = \varphi_1 \in H^{s-3/2+\frac{k}{k+1}}(\mathbf{R}^n) \\ B_2(u)|_{x_{n+1}=0} = \frac{\partial^2 u}{\partial x_{n+1}^2}|_{x_{n+1}=0} = \varphi_2 \in H^{s-\frac{5}{2}+\frac{k}{k+1}}(\mathbf{R}^n) \\ B_3(u)|_{x_{n+1}=0} = \frac{\partial^3 u}{\partial x_{n+1}^3}|_{x_{n+1}=0} = \varphi_3 \in H^{s-\frac{7}{2}+\frac{k}{k+1}}(\mathbf{R}^n), s \geq 6. \end{array}\right. \tag{18}$$

Then (18) is reduced to 3×3 system of pseudo-differential equations on the boundary $x_{n+1} = 0$, the equation for c_0 being first-order subelliptic one and $a(x) = 0 \Rightarrow \sum_{j=1}^k |L^j a(x)| > 0$.

Then $u(x, x_{n+1}) = (2\pi)^{-n} \int e^{ix\xi - x_{n+1}|\xi|} (\hat{c}_0 + x_{n+1}\hat{c}_1 + x_{n+1}^2 \hat{c}_2(\xi)) d\xi$, where $\hat{c}_j(\xi)$ stands for the Fourier transform of the tempered distribution $c_j(x)$.

One can guess that (18) is hypoelliptic bvp of Fredholm type and $u \in H_{loc}^s(\mathbf{R}_+^{n+1})$.

For the sake of completeness, we give the exact formulas for the solutions of Dirichlet (Neumann) problems for Laplace equation in $\mathbf{R}_+^{n+1}$:

$$\left|\begin{array}{l} \Delta u = 0 \text{ in } \mathbf{R}_+^{n+1}, n \geq 2 \\ u|_{x_{n+1}=0} = \varphi(x) \in C_b^0(\mathbf{R}^n). \end{array}\right. \tag{19}$$

Then there exists unique solution $u \in C^\infty(\mathbf{R}_+^{n+1}) \cap C_b^0(\bar{\mathbf{R}}^{n+1})$ given by the Poisson formula

$$u(x, x_{n+1}) = \frac{\Gamma(\frac{n+1}{2})}{\pi^{\frac{n+1}{2}}} x_{n+1} \int_{\mathbf{R}^n} \frac{\varphi(x-y)dy}{(x_{n+1}^2 + |y|^2)^{\frac{n+1}{2}}} = \frac{\Gamma(\frac{n+1}{2})}{\pi^{\frac{n+1}{2}}} x_{n+1} \int_{\mathbf{R}^n} \frac{\varphi(x+yx_{n+1}))dy}{(1+|y|^2)^{\frac{n+1}{2}}} \tag{20}$$

and $||u||_{C^{2,\alpha}(\bar{\mathbf{R}}_+^{n+1})} \leq C_\alpha ||\varphi||_{C^{2,\alpha}(\mathbf{R}^n)}, 0 < \alpha < 1$.

It is interesting to point out that $u(x, x_{n+1}) \rightarrow_{x_{n+1}\to 0} \varphi(x), \forall x$, while $\lim_{x_{n+1}\to 0} x_{n+1} \frac{\partial u}{\partial x_{n+1}} = 0$.

We will give examples of Hölder $C^{0,\alpha}$ and Lipschitz $C^{0,1}$ Dirichlet data such that $\frac{\partial u}{\partial x_{n+1}}(0, x_{n+1}) \sim c \ln \frac{1}{x_{n+1}}, c > 0$ for $x_{n+1} \to 0$ in the Lipschitz case. As it concerns the Neumann bvp

$$\left|\begin{array}{l} \Delta u = 0, \text{ in } \mathbf{R}_+^{n+1}, n \geq 2, u \in C^\infty(\mathbf{R}_+^{n+1}) \\ -\frac{\partial u}{\partial x_{n+1}}|_{x_{n+1}=0} = h(x) \in C_b^0(\mathbf{R}^n) \cap L^1(\mathbf{R}^n), \end{array}\right. \tag{21}$$

we obtain that

$$u(x, x_{n+1}) = \frac{\Gamma(\frac{n+1}{2})}{\pi^{\frac{n+1}{2}}(n-1)} \int_{\mathbf{R}_y^n} \frac{h(y)dy}{(x_{n+1}^2 + |x-y|^2)^{\frac{n-1}{2}}} + C, \tag{22}$$

$x_{n+1} > 0$ and (22) holds for $|h(y)| \leq C(1+|y|^2)^{-1}$. We remind that the convolution of $f, g \in L^1(\mathbf{R}^n)$ is defined by $f * g(x) = \int_{\mathbf{R}^n} f(x-y)g(y)dy$.

Example 5. Consider the polyharmonic operator

$$\Delta^m u = 0, \text{ in } \mathbf{R}_+^{n+1}, m \geq 2, n \geq 2. \tag{23}$$

After a partial Fourier transform with respect to $x \in \mathbf{R}^n$ we get that $\hat{u}(\xi, x_{n+1}) = F_{x\to\xi}(u(x, x_{n+1}))$, $\hat{u}(\xi, x_{n+1}) = \sum_{j=1}^m x_{n+1}^{j-1} \hat{c}_j(\xi) e^{-x_{n+1}|\xi|}$, $x_{n+1} > 0$, $\hat{c}_j(\xi)$ being tempered distributions in $\mathbf{R}_x^n$.

By using (20) we come to the formula

$$u(x, x_{n+1}) = B_n \sum_{j=1}^m x_{n+1}^j \frac{1}{(x_{n+1}^2 + |x|^2)^{\frac{n+1}{2}}} * c_j(x) = \sum_{j=1}^m x_{n+1}^{j-1} u_j(x, x_{n+1}), \tag{24}$$

$B_n = \frac{\Gamma(\frac{n+1}{2})}{\pi^{\frac{n+1}{2}}}, u_j = B_n \frac{x_{n+1}}{(x_{n+1}^2+|x|^2)^{\frac{n+1}{2}}} * c_j(x), \Delta u_j = 0 \text{ for } x_{n+1} > 0, \Delta^m(x_{n+1}^{j-1} u_j) = 0$
for $x_{n+1} > 0$ and $j = 1, \ldots, m$.

The convolution is taken in $\mathbf{R}_x^n$ with respect to x, x_{n+1} being a parameter.

One can see that (23) is satisfied by the function

$$u = x_{n+1}^m \sum_{j=0}^{m-1} \partial_{x_{n+1}}^j \left(\frac{1}{(x_{n+1}^2 + |x|^2)^{\frac{n+1}{2}}} * c_j(x) \right). \tag{25}$$

The simplest case is when $c_j \in C_0^\infty(\mathbf{R}_x^n)$. If $I_j = \partial_{x_{n+1}}^j \left(\frac{1}{(x_{n+1}^2+|x|^2)^{\frac{n+1}{2}}} \right) * c_j(x)$ it is interesting to find out an explicit formula for the corresponding derivative $\partial_{x_{n+1}}^j$ of the Poisson kernel. Then we must consider two cases: j even and j odd. We shall apply Faa di Bruno formula [13] to the Poisson kernel for $j \geq 1$, j even. After several computations we come to the expression:

$$\partial_{x_{n+1}}^j \left(\frac{1}{(x_{n+1}^2 + |x|^2)^{\frac{n+1}{2}}} \right) = \sum_{p=0}^{j/2} (-1)^{j-p} (j-p) \ldots (j-2p+1) \tag{26}$$

$$\frac{\frac{n+1}{2} \frac{n+3}{2} \ldots (\frac{n+1}{2} + j - p - 1)}{(x_{n+1}^2 + |x|^2)^{j-p+\frac{n+1}{2}}} (2x_{n+1})^{j-2p} \binom{j}{p}.$$

We point out that $\partial_{x_{n+1}}^j \left(\frac{x_{n+1}}{(x_{n+1}^2+|x|^2)^{\frac{n+1}{2}}} \right)$ is reduced to the computation of $\partial_{x_{n+1}}^j \left(\frac{1}{(x_{n+1}^2+|x|^2)^{\frac{n+1}{2}}} \right)$ and $|x|^2 \partial_{x_{n+1}}^j \left(\frac{1}{(x_{n+1}^2+|x|^2)^{\frac{n+3}{2}}} \right)$.

And now we shall study the following elliptic boundary value problem in the unit ball B_1 finding out the solution into explicit form via the Poisson formula. Thus,

$$\begin{vmatrix} \Delta^m u = 0 \text{ in } B_1, m \geq 2, n \geq 3 \\ B_0(u) = \varphi_1 \text{ on } S_1 \\ \ldots \\ B_{m-1}(u) = \varphi_m \text{ on } S_1, \end{vmatrix} \tag{27}$$

where $B_0(u) = u, B_1(u) = \sum_{l=1}^n a_l(x) D_{x_l} + b(x) D_r u, \ldots B_j(u) = \sum_{k+|\alpha|=j} a_{kl}(x)$ $D_x^\alpha D_r^k u$; $j = 1, 2, \ldots, m-1$. Evidently, $B_j(u)$ for $j \geq 1$ are differential operators on the boundary with smooth coefficients. As usual, $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbf{N}_0^n$, $D_{x_j} = \frac{1}{i} \frac{\partial}{\partial x_j}$, $D_r = \frac{1}{i} \frac{\partial}{\partial r}$, r being the unit outward normal to S_1 for $r = 1$. We can take the radial vector field $L = \sum_{j=1}^n x_j \frac{\partial}{\partial x_j} = r \frac{d}{dr}, L \neq 0$ for $x \neq 0$ and $L|_{S_1} = \frac{\partial}{\partial r}|_{S_1}$ instead of $\frac{\partial}{\partial r}$, respectively D_r. The vector fields $\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}, L$ are linearly dependent in $\mathbf{R}^n$. According to [1] each sufficiently smooth m-polyharmonic function in B_1 can be written in the form

$$u = \sum_{j=1}^{m} (r^2 - 1)^{j-1} u_j, \ \Delta u_j = 0 \ \text{in} \ B_1. \tag{28}$$

In fact, one can easily prove by induction the identities

$$\Delta L = L\Delta + 2\Delta, \ \Delta L^2 = L^2 \Delta + 4L\Delta + 4\Delta, \tag{29}$$

$$\Delta u = 0 \Rightarrow \Delta L^k u = 0 \tag{30}$$

and

$$\Delta u = 0 \Rightarrow \Delta(r^k \frac{d^k u}{dr^k}) = 0, \ \Delta^{k+1}((r^2 - 1)^k u) = 0. \tag{31}$$

(29), (30), (31) hold in B_1 for $k \in \mathbf{N}$.

Proposition 3 *Consider the boundary value problem (27) with sufficiently smooth right-hand sides φ_j, $j = 0, 1, \ldots, m-1$ and assume that for each $j \in [1, \ldots, m-1]$ the expression*

$$A_j(x) = \sum_{k+|\alpha|=j} a_{k\alpha}(x) D_x^\alpha D_r^k((r^2 - 1)^j) \neq 0 \tag{32}$$

at S_1.

Then (27) possesses a unique solution of the form (28), where the functions u_j satisfy the Dirichlet problem for Laplace equation in B_1, namely,

$$\begin{vmatrix} \Delta u_1 = 0, B_1 \\ u_1|_{S_1} = \varphi_1 \end{vmatrix} \begin{vmatrix} \Delta u_2 = 0, B_1 \\ u_2|_{S_1} = \psi_2 \end{vmatrix} \cdots \begin{vmatrix} \Delta u_m = 0, B_1 \\ u_m|_{S_1} = \psi_m. \end{vmatrix} \tag{33}$$

$\psi_1 = \varphi_1$ and ψ_j is expressed for $j \geq 2$ by φ_j as well as by $\varphi_1, \ldots, \varphi_{j-1}$ and their derivatives.

This is the formula for the unique solution of the Dirichlet problem to the Poisson equation [6]:

$$\begin{vmatrix} -\Delta u = f \ \text{in} \ B_1 \subset \mathbf{R}^n, n \geq 3, f \in C^\alpha(\bar{B}_1) \\ u|_{S_1} = \varphi \in C^0(S_1) \end{vmatrix} \tag{34}$$

$$u(x) = \int_{B_1} G_{1,n}(x, y) f(y) dy + \frac{1 - |x|^2}{n\omega_n} \int_{S_1} \frac{\varphi(y)}{|x - y|^n},$$

where $G_{1,n}$ is given by formula (5), $n\omega_n = \frac{2\pi^{n/2}}{\Gamma(\frac{n}{2})}$.

Example 6. $\begin{vmatrix} \Delta^2 u = 0 \ \text{in} \ B_1, n \geq 3 \\ B_0(u) \equiv u = \varphi_1 \ \text{at} \ S_1 \\ B_1(u) = \varphi_2 = \sum_{l=1}^{n} a_l(x) D_{x_l} u + b(x) D_r u|_{S_1}. \end{vmatrix}$

Then condition $A_1(x) \neq 0$ (32) takes the form $A_1(x) = \sum_{j=1}^{n} x_j a_j(x) + b(x) \neq$
0 at S_1, respectively, $u = (r^2 - 1)u_1 + u_2$, $\begin{vmatrix} \Delta u_1 = 0, B_1 \\ u_1|_{S_1} = \varphi_1, \end{vmatrix} \begin{vmatrix} \Delta u_2 = 0, B_1 \\ u_2|_{S_1} = i\frac{\varphi_2 - B_1(u_1)}{2A_1}|_{S_1} \end{vmatrix}$.
Our last Proposition 4 concerns the Dirichlet problem for the polyharmonic operator
in $\mathbf{R}_+^{n+1}$.

Proposition 4 *Consider in $\mathbf{R}_+^{n+1}$, $n \geq 2$ the boundary value problem*

$$\begin{vmatrix} \Delta^m u = 0, & \text{in } \mathbf{R}_+^{n+1} \\ u|_{x_{n+1}=0} = \varphi_1 \in H^{m-1/2}(\mathbf{R}^n) \\ \dots \\ \frac{\partial^{m-1} u}{\partial x_{n+1}^{m-1}}|_{x_{n+1}=0} = \varphi_m \in H^{1/2}(\mathbf{R}^n). \end{vmatrix} \tag{35}$$

Then there exists a solution of (35) $u \in H^m_{loc}(\mathbf{R}_+^{n+1})$ that can be written as $\hat{u}(\xi, x_{n+1}) = \sum_{j=1}^{m} x_{n+1}^{j-1} e^{-x_{n+1}|\xi|} \hat{c}_j(\xi), c_j$ being unknown tempered distributions in $\mathbf{R}^n$. The distributions c_j, $j = 1, \dots, m$ can be expressed as

$$\begin{vmatrix} c_1 = \varphi_1 \\ c_2 = \varphi_2 + |D|\varphi_1 \\ c_3 = \frac{1}{2!}\varphi_3 + \frac{1}{2}|D|^2\varphi_1 + |D|\varphi_2 \\ \dots \\ c_j = \frac{1}{(j-1)!}\varphi_j + A_{1,j}|D|^{j-1}\varphi_1 + \dots + A_{j-1,j}|D|\varphi_{j-1} \\ \dots, \end{vmatrix} \tag{36}$$

the constants A_{kj} being uniquely determined.

In the standard variables $(x, x_{n+1}) \in \mathbf{R}_+^{n+1}$, we have

$$u(x, x_{n+1}) = \sum_{j=1}^{m} x_{n+1}^{j-1} [D_{1j} \frac{x_{n+1}}{(x_{n+1}^2 + |x|^2)^{\frac{n+1}{2}}} * \varphi_j + \tag{37}$$

$$B_{1j} \frac{\partial^{j-1}}{\partial x_{n+1}^{j-1}} (\frac{x_{n+1}}{(x_{n+1}^2 + |x|^2)^{\frac{n+1}{2}}} * \varphi_1) +$$

$$\dots + B_{j-1,j} \frac{\partial}{\partial x_{n+1}} (\frac{x_{n+1}}{(x_{n+1}^2 + |x|^2)^{\frac{n+1}{2}}} * \varphi_{j-1})]$$

and $B_{k,j} = A_{k,j}(-1)^{j-k}\frac{\Gamma(\frac{n+1}{2})}{\pi^{\frac{n+1}{2}}}$, $k = 1, \dots, j-1$, while $D_{1j} = \frac{1}{(j-1)!}\frac{\Gamma(\frac{n+1}{2})}{\pi^{\frac{n+1}{2}}}$.

3 Proofs of Main Results, the Examples, and Comments

We shall begin with Proposition 1 (a). Thus,

$$0 < G_{1,n}(x, y) = \frac{k_{1,n}}{n-2}(|x - y|^{2-n} - [XY]^{2-n}) =$$

$$\frac{k_{1,n}}{n-2} \frac{(]XY] - |x - y|)([XY]^{n-3} + \ldots + |x - y|^{n-3})}{|x - y|^{n-2}[XY]^{n-3}[XY]} \geq$$

$$\geq \frac{k_{1,n}}{n-2} \frac{[XY]^2 - |x - y|^2}{|x - y|^{n-2}[XY]([XY] + |x - y|)}.$$

Having in mind that $|x - y| \leq [XY]$ and therefore $[XY] + |x - y| \leq 2[XY]$ we get

$$G_{1,n}(x, y) \geq \frac{k_{1,n}}{2(n-2)}(|x - y|^{n-2} + \frac{|x - y|^n}{[XY]^2 - |x - y|^2})^{-1} \geq$$

$$\frac{k_{1,n}}{2(n-2)}|x - y|^{2-n}d(x)d(y)(|x - y|^2 + d(x)d(y))^{-1},$$

$$d(x)d(y) \leq [XY]^2 - |x - y|^2 \leq 4d(x)d(y).$$

We are interested in the estimation from below:

$$\frac{|x - y|^{2-n}d(x)d(y)}{|x - y|^2 + d(x)d(y)} \geq cd(x)d(y) \tag{38}$$

for some constant $c > 0$ in the ball: $x, y \in B_1$. Consequently,

$$1 \geq c|x - y|^{n-2}(|x - y|^2 + d(x)d(y)). \tag{39}$$

On the other hand, $|x - y|^{n-2}(|x - y|^2 + d(x)d(y) \leq 2^{n-2}.5$ for each $x, y \in B_1$. Taking $c = \frac{2^{2-n}}{5}$ we come to (38). Thus, $G_{1,n}(x, y) \geq \frac{k_{1,n}}{5(n-2)}2^{1-n}d(x)d(y), \forall x, y \in B_1$, i.e., we obtain (6).

The estimations from below of the Green functions $G_{m,n}$ (4) are similar, the results are the same and we shall investigate only the case $n > 2m$. We are looking for such a positive constant $C > 0$ that

$$|x - y|^{2m-n}min\{1, \frac{d^m(x)d^m(y)}{|x - y|^{2m}}\} \geq Cd^m(x)d^m(y), \forall x, y \in \bar{B}_1. \tag{40}$$

Thus, in an equivalent form

$$\frac{1}{C}min\{1, \frac{d^m(x), d^m(y)}{|x - y|^{2m}}\} \geq |x - y|^{n-2m}d^m(x)d^m(y), \forall x, y \in B_1. \tag{41}$$

Evidently, $0 \leq |x - y|^{n-2m}d^m(x)d^m(y) \leq 2^{n-2m}, \forall x, y \in \bar{B}_1$.

There are two possibilities in (41):

(a) $0 \leq \frac{d(x)d(y)}{|x-y|^2} \leq 1 \iff 0 \leq \frac{d^m(x)d^m(y)}{|x-y|^{2m}} \leq 1$

(b) $\frac{d(x)d(y)}{|x-y|^2} \geq 1 \iff \frac{d^m(x)d^m(y)}{|x-y|^{2m}} \geq 1$.

In the case (a) (41) takes the form

$$\frac{1}{C}\frac{d^m(x)d^m(y)}{|x-y|^{2m}} \geq |x-y|^{n-2m}d^m(x)d^m(y), \tag{42}$$

i.e., (41) holds if $C = 2^{-n}$.

If (b) (40) holds we have

$$\frac{1}{C} \geq |x-y|^{n-2m}d^m(y)d^m(y). \tag{43}$$

Therefore, (43) is valid with $C = 2^{-n+2m}$. Taking the constant $C = 2^{-n}$ we get (40). As it concerns the estimation from above of the solution of (1), Proposition 1 points out that $r^{2p+1} = |x|^{2p+1} \notin C^\infty(|x| \leq 1)$ but only to $C^\infty(B_1 \setminus 0)$. By using formula (2) we reduce our problem to the construction of C^∞ smooth solution of $(-\Delta u)^m = 1$ with zero Dirichlet data in B_1. Because of this we are trying to find the corresponding solution as radial symmetric one, i.e., $v = \frac{1}{C_m}(1-r^2)^m$, $C_m > 0$. The boundary conditions are obviously fulfilled. The inductive formula $\Delta^m r^\alpha = \alpha(\alpha - 2)\ldots(\alpha - 2(m-1))(\alpha - 2 + n)(\alpha - 4 + n)\ldots(\alpha - 2m + n)r^{\alpha - 2m}$ enables us to conclude that

$$\left|\begin{array}{l} (-\Delta)^m v = 1 \text{ in } B_1 \\ v|_{S_1} = \ldots = \frac{\partial^{m-1}v}{\partial r^{m-1}}|_{S_1} = 0 \end{array}\right. \tag{44}$$

and $v = \frac{1}{C_m}(1-r^2)^m$ with $C_m = 2^m m!(n + 2m - 2)(n + 2m - 4)\ldots(n + 2)n$.

Proposition 1 is verified.

Remark 1 We can construct Green function in $\mathbf{R}^n_+$, $n \geq 3$. As it is mentioned in [5] for $n > 2m$ $G_{(-\Delta)^m, \mathbf{R}^n_+} = k_{m,n}|x-y|^{2m-n}\int_1^{\frac{|x^*-y|}{|x-y|}}(v^2-1)^{m-1}v^{1-n}dv$ and $\mathbf{R}^n_+ = \{x \in \mathbf{R}^n, x_1 > 0\}$, while if $x \in \mathbf{R}^n_+$ then $x^* = (-x_1, x_2, \ldots, x_n) \in \mathbf{R}^n_-$. It is easy to see that $|x^* - y|^2 = |x-y|^2 + 4d(x)d(y)$, $d(x) = x_1$, $d(y) = y_1$, $x_1 = dist(x, \partial\mathbf{R}^n_+)$. Similar considerations as in the case of Proposition 1 can be done with the Green function $G_{(-\Delta)^m, \mathbf{R}^n_+}(x, y)$ but we omit the details.

Several words about the Navier problem (9), based on (7). With some constant $c > 0$ for the Poisson equations (10), we guess

$$\begin{array}{l} cd(x)\int_{B_1}d(y)u_1(y)dy \leq u(x) \leq \frac{||u_1||_{C^0}}{n}d(x) \\ cd(x)\int_{B_1}d(y)u_2(y)dy \leq u_1(x) \leq \frac{||u_2||_{C^0}}{n}d(x) \\ \ldots \\ cd(x)\int_{B_1}d(y)u_{j+1}(y)dy \leq u_j(x) \leq \frac{||u_{j+1}||_{C^0}}{n}d(x) \\ \ldots \\ cd(x)\int_{B_1}d(y)f(y)dy \leq u_{m-1}(x) \leq \frac{||f||_{C^0}}{n}d(x). \end{array} \tag{45}$$

We observe that $f \geq 0 \Rightarrow u_{m-1} \geq 0 \Rightarrow u_{m-2} \geq 0$ and we conclude inductively that each $u_j \geq 0$ as well $u \geq 0$. As $0 \leq u(x) \leq \frac{1}{n}||u_1||_{C^0}d(x), \ldots, ||u_{m-2}||_{C^0} \leq \frac{1}{n}||u_{m-1}||_{C^0}$ and $||u_{m-1}||_{C^0} \leq \frac{1}{n}||f||_{C^0} \Rightarrow 0 \leq u(x) \leq \frac{1}{n^m}||f||_{C^0}d(x)$, i.e., $u(x) \leq \frac{1}{n^m}||f||_{C^0(\bar{B}_1)}d(x)$ from Proposition 2 is shown. The left-hand side of the inequality is proved inductively, as the first two lines from (45) give $u(x) \geq cd(x)\int_{B_1} u_1(y)d(y) dy \geq c^2 d(x) \int_{B_1} d^2(y)dy \int_{B_1} u_2(y)d(y)dy$, etc.

Therefore, $u(x) \geq c^m d(x)(\int_{B_1} d^2(y)dy)^{m-1} \int_{B_1} f(y)d(y)dy$ and one can take $c = \frac{k_{1,n}2^{1-n}}{5(n-2)}$. This way Proposition 2 is proved.

To find formula (20) we remind that $u(x, x_{n+1}) = F_{\xi \to x}^{-1}(e^{-x_{n+1}|\xi|}\hat{\varphi})$. On the other hand, if we put $\hat{v} = e^{-x_{n+1}|\xi|}$ we obtain that $u(x, x_{n+1}) = F_{\xi \to x}^{-1}(\hat{\varphi}\hat{v}) = v * \varphi$. It is well known (Eskin, Shilov) that $v = F_{\xi \to x}^{-1}(\hat{v}) = \frac{\Gamma(\frac{n+1}{2})}{\pi^{\frac{n+1}{2}}} \frac{x_{n+1}}{(x_{n+1}^2+|x|^2)^{\frac{n+1}{2}}}$, $n \geq 3$. Thus, $u = \frac{\Gamma(\frac{n+1}{2})}{\pi^{\frac{n+1}{2}}} \frac{x_{n+1}}{(x_{n+1}^2+|x|^2)^{\frac{n+1}{2}}} * \varphi(x)$. To obtain Neumann formula (22) for (21) we put $v(x, x_{n+1}) = -\frac{\partial u}{\partial x_{n+1}} \Rightarrow \begin{vmatrix} \Delta v = 0, x_{n+1} > 0 \\ v|_{x_{n+1}=0} = h(x) \end{vmatrix}$. Poisson formula (20) gives us that $v = \frac{\Gamma(\frac{n+1}{2})x_{n+1}}{\pi^{\frac{n+1}{2}}} \int_{\mathbf{R}^n} \frac{h(y)dy}{(x_{n+1}^2+|x|^2)^{\frac{n+1}{2}}}$. According to Lebesgue's dominated convergence theorem $u(x, x_{n+1})$, given by (20), tends to 0 for $x_{n+1} \to \infty$. Therefore, $u(x, x_{n+1}) - u(x, +\infty) = -\frac{\Gamma(\frac{n+1}{2})}{\pi^{\frac{n+1}{2}}} \int_{\mathbf{R}^n} h(y)[\int_{+\infty}^{x_{n+1}} \frac{\lambda d\lambda}{(\lambda^2+|x-y|^2)^{\frac{n+1}{2}}}]dy$. After a simple integration we come to (22).

The proof of (14) for the operator $L(x, D) + ia(x)|D|$ relies on (13) and Parseval's identity. In fact, in the case $s = 2$ $\int_{\mathbf{R}_x^n} \int_0^\infty |D_x^\alpha D_{x_{n+1}}^\beta u(x, x_{n+1})|^2 dx dx_{n+1} = (2\pi)^{-n} \int_{\mathbf{R}_x^n} |\xi^\alpha|^2 |D_{x_{n+1}}^\beta \hat{u}(\xi, x_{n+1})|^2 d\xi dx_{n+1}$. Certainly, $\Delta u = 0$ in $\mathbf{R}_+^{n+1}$ implies that $\hat{u}(\xi, x_{n+1}) = e^{-x_{n+1}|\xi|}\hat{c}(\xi)$, i.e., $D_{x_{n+1}}^\beta \hat{u}(\xi, x_{n+1}) = const|\xi|^\beta \hat{c}(\xi)e^{-x_{n+1}|\xi|}$, i.e., $||D_x^\alpha D_{x_{n+1}}^\beta u||_{L^2(\mathbf{R}_+^{n+1})}^2 \leq const \int_0^\infty \int_{\mathbf{R}_x^n} |\xi|^{2|\alpha+\beta|}|\hat{c}(\xi)|^2 e^{-2x_{n+1}|\xi|} d\xi dx_{n+1}$. Thus, for $s = 2$, i.e., $|\alpha + \beta| \leq 2$, we get $||u||_{H^s(\mathbf{R}_+^{n+1})}^2 \leq const||c||_{H^{s-1/2}(\mathbf{R}^n)}$ as $\int_0^\infty e^{-2x_{n+1}|\xi|} dx_{n+1} = \frac{1}{2|\xi|}$. Applying (13) to $||c||_{H^{s-1/2}}$ we come to (14). We repeat that in those estimates the H_{loc}^s Sobolev norms are used; $c(x) = u(x, 0)$. Because of the lack of space and the same pseudo-differential approach used in Examples 1–4 we shall study in details only Example 4 that seems to be more complicated. Then its solution exists of the form

$$u = (2\pi)^{-n} \int e^{ix\xi - x_{n+1}|\xi|}(\hat{c}_0(\xi) + x_{n+1}\hat{c}_1(\xi) + x_{n+1}^2\hat{c}_2(\xi))d\xi.$$

We find consecutively $\frac{\partial u}{\partial x_{n+1}}|_{x_{n+1}=0} = c_1 - |D|c_0, \varphi_2 = \frac{\partial^2 u}{\partial x_{n+1}^2}|_{x_{n+1}=0} = |D|^2 c_0 - 2|D|c_1 + 2c_2$, $\varphi_3 = \frac{\partial^3 u}{\partial x_{n+1}^3}|_{x_{n+1}=0} = -|D|^3 c_0 + 3|D|^2 c_1 - 6|D|c_2$. From the pseudo-differential system containing the last two equations, we eliminate c_2 and get

$$c_1 = \frac{1}{3}(-|D|^{-2}\varphi_3 - 3|D|^{-1}\varphi_2 + 2|D|c_0). \tag{46}$$

Therefore,

$$c_2 = \frac{1}{2}(-\varphi_2 + \frac{1}{3}|D|^2 c_0 - \frac{2}{3}|D|^{-1}\varphi_3). \tag{47}$$

On the other hand, $B_1(u)|_{x_{n+1}=0} = \varphi_1$ leads to the pseudo-differential equation on the boundary $(L(x, D_x) + ia(x)|D|)c_0 - ia(x)c_1 = -i\varphi_1$, L being defined in (11). Eliminating c_1 we conclude that

$$(L(x, D) + \frac{i}{3}a(x)|D|)c_0 = -i\varphi_1 - i\frac{a(x)}{3}|D|^{-2}\varphi_3 - ia(x)|D|^{-1}\varphi_2 \equiv \psi \quad (48)$$

is a subelliptic equation.

One can see from (18) that $\psi \in H^{s-\frac{3}{2}+\frac{k}{k+1}}$ and according to the subelliptic estimate (13) for the boundary operator $\left|\begin{array}{l} \Delta u = 0, \mathbf{R}_+^{n+1} \\ M_1 u = L(x, \partial_x)u + \frac{a(x)}{3}\frac{\partial u}{\partial x_{n+1}}|_{x_{n+1}=0} = \psi \end{array}\right.$ we have that $c_0 = u|_{x_{n+1}=0} \in H^{s-1/2}(\mathbf{R}^n)$. Then according to (46) $c_1 \in H^{s-3/2}$ and (47) implies $c_2 \in H^{s-5/2}$.

Concluding remark:

$$\begin{aligned} &\Delta^3 u = 0 \text{ in } \mathbf{R}_+^{n+1} \\ &u|_{x_{n+1}=0} \in H^{s-1/2}(\mathbf{R}^n) \\ &\frac{\partial u}{\partial x_{n+1}}|_{x_{n+1}=0} \in H^{s-3/2} \\ &\frac{\partial^2 u}{\partial x_{n+1}^2}|_{x_{n+1}=0} \in H^{s-5/2+\frac{k}{k+1}} \subset H^{s-\frac{5}{2}}. \end{aligned}$$

As u satisfies the Dirichlet (i.e., elliptic) problem for $\Delta^3 u = 0$ we conclude that $u(x, x_{n+1}) \in H_{loc}^s(\mathbf{R}_+^{n+1})$, $s \geq 6$.

Concerning Example 2 we shall mention that in this case $c_1 = \varphi_1 + |D|c_0$, while c_0 satisfies the second-order PDE on the boundary

$$g(x)\Delta_x c_0 + \partial_{x_1}^2 c_0 = \varphi_2 + 2g(x)|D|\varphi_1, g(x) = x_1^2. \tag{49}$$

The operator (49) is Hörmander's sum of squares of smooth linear real vector fields, namely, it has the form $Nc_0 \equiv x_1^2(\partial_1^2 + \partial_2^2 + \ldots + \partial_n^2)c_0 + \partial_1^2 c_0 = \psi(x)$. N can be written as $N = \sum_{j=1}^{n+1} X_j^2 + X_0$, where $X_{n+1} = \partial_1$, $X_j = x_1\partial_j$ for $n \geq j \geq 1$ and $X_0 = -x_1\partial_1$. Evidently, $[\partial_1, x_1\partial_1] = \partial_1$, $[\partial_1, x_1\partial_j] = \partial_j$ for $j \geq 2$, $[x_1\partial_j, x_1\partial_k] = 0$, $j \geq 2, k \geq 2$, $[x_1\partial_1, x_1\partial_j] = x_1\partial_j$, $j \geq 2$.

As the system of the vector fields has rank n the hypoellipticity of N holds.

We shall prove now Proposition 3. We are looking for the solution of (27) of the form

$$u = \sum_{l=1}^m (r^2 - 1)^{l-1} u_l, \Delta u_l = 0, r = |x|. \tag{50}$$

Evidently, $\begin{vmatrix} \Delta u_1 = 0 & \text{in } B_1, \\ u_1|_{S_1} = \varphi_1. \end{vmatrix}$

One can easily see that for $j \geq 1$

$$\varphi_{j+1} = B_j(u)|_{S_1} = A_j u_{j+1} + \tag{51}$$

a linear combination of $u_1, \ldots, u_j$ and their derivatives of some order.

The boundary value problem (27) decomposes to the solvability of m-Dirichlet boundary value problems of the type (33). As u_1 can be found directly via Poisson formula the other solutions can be constructed inductively via (51), respectively, the solution u of (27) is written in the form (50).

The proof of Proposition 4 is standard and we omit it.

We shall begin now with several concluding remarks.

At first we prove that $\lim_{x_{n+1} \to 0} x_{n+1} \frac{\partial u}{\partial x_{n+1}} = 0$ for u from (19), (20), where $\frac{\partial u}{\partial x_{n+1}} = C_n \int_{\mathbf{R}^n} \frac{\varphi(x-y)(|y|^2 - n x_{n+1}^2)}{(x_{n+1}^2 + |y|^2)^{\frac{n+3}{2}}} dy$.

Example 7. Consider (20) with $\varphi(x) = \begin{cases} |x|^\alpha, |x| \leq 1 \\ 1, |x| \geq 1 \end{cases}$ and $0 < \alpha \leq 1$. Then for $x_{n+1} > 0$, $u(0, x_{n+1}) = C_n \int_{\mathbf{R}^n} \frac{\varphi(x_{n+1}z)dz}{(1+|z|^2)^{\frac{n+1}{2}}} = C_n x_{n+1}^\alpha \int_{|z| < \frac{1}{x_{n+1}}} \frac{|z|^\alpha dz}{(1+|z|^2)^{\frac{n+1}{2}}} + C_n \int_{|z| > \frac{1}{x_{n+1}}} \frac{dz}{(1+|z|^2)^{\frac{n+1}{2}}}$. So with some $Q > 0$

$$u(0, x_{n+1}) = C_n x_{n+1}^\alpha \left(\frac{meas\, S_1}{\alpha - 1} x_{n+1}^{1-\alpha} + Q + o(x_{n+1}^{1-\alpha}) \right) + C_n meas\, S_1 x_{n+1} + o(x_{n+1}), x_{n+1} \to 0,$$

i.e., $u(0, x_{n+1})$ is Hölder function with exponent $\alpha < 1$.

In the Lipschitz case $\alpha = 1$, we have that

$$u(0, x_{n+1}) = C_n meas\, S_1 x_{n+1} \left(ln \frac{1}{x_{n+1}} + 1 \right) + o(x_{n+1}), x_{n+1} \to 0,$$

i.e., the solution u is losing Lipschitz property near the boundary. Certainly, $|\frac{\partial u}{\partial x_{n+1}}(0, x_{n+1})|_{x_{n+1} \to 0} \to \infty$ for each exponent $0 < \alpha \leq 1$.

References

1. T. Boggio. Sulle funzioni di Green d'ordine m. Circ. Mat. Palermo, 20; 97–135, 1905.
2. H. Dwight. Tables of integrals and other mathematical data, IV Edition. The Mcmillan company, NY, 1961.
3. J. Edenhofer. Eine Integraldarstellung der Lösung der Dirichletschen Aufgabe bei der Polypotentialgleichung im Falle eine Hyperkugel, Math. Nachr. 69:149–162, 1975.
4. Y.V. Egorov. Linear differential equations of principal type, Consultants Bureau, NY, 1986.
5. F. Gazzola, H.-C. Grunau, G. Sweers, Positivity preserving and nonlinear higher order elliptic equations in bounded domains, Lecture Notes in Mathematics v.1991, Springer, 2010.

6. D. Gilbarg, N. Trudinger. Elliptic partial differential equations of second order, Grundlehren der Math. Wissenschaften, 224, Springer, Berlin, 1983.
7. L. Hörmander. The Analysis of Linear Partial Differential Operators, vol. I–IV, Springer, Berlin, 1983–1985.
8. T. Komorowski, A. Bobrowski. A quantitative Hopf-type maximum principle for subsolutions of elliptic PDEs, Discrete and continuous dynamical systems series S, 13:12, 2020.
9. J.-L. Lions, E. Magenes, Problémes aux limites non-homogénes et applications, vol. I, Travaux et Recherches Mathematics, No. 17, Dunod, Paris, 1968.
10. P. Popivanov, D. Palagachev. The degenerate oblique derivative problem for elliptic and parabolic equations, Akademie Verlag, Berlin, 1997.
11. P. Popivanov. Boundary value problems for the biharmonic operator violating Lopatinskii conditions, C. R. Acad. Bulg. Sci., v.66:9, 1231–1238, 2013.
12. P. Popivanov. Boundary value problems for the biharmonic operator in the unit ball, Sixth International Conference NTADES, AIP Conference Proceedings, 2159, 030028, 2019.
13. I. Rijik, M. Gradstein. Table of integrals, series and products, Academic Press, 1980.
14. Z. Zhao. Green function for Schrödinger operator and conditioned Feynman-Kac gauge, J. Math. Anal. Appl. 116, 309–334, 1986.

Search of Complex Transcendental Roots of a Combination of a Nonlinear Equation and a Polynomial

Yoshihiro Mochimaru

Abstract Target nonlinear equation is limited within ordinary differential equations with one-point boundary value or two-point boundary value problems, and the order of a polynomial is assumed to be finite and it is assumed that at least one complex solution exists (which may be real). The purpose is to find numerically at least one root (not necessarily all of them). Examples are similar equations to Chandrasekhar's white dwarf equation, nonlinear differential equation exactly expressible through Weierstrass function, and some internal gas explosion equations.

Keywords Transcendental root · Combination · Nonlinear · Monic equation

1 Introduction

Root(s) of a single real nonlinear equation would be found numerically in principle such as by means of Newton's method if it is analytic. Complex roots of some Bessel functions, Hankel functions, their cross-products, and Mathieu functions are reported in [1, 2].

Here treated are complex transcendental roots of a combination of a nonlinear equation and a polynomial. Target equations: Specific nonlinear functions such as those given by derivatives through their polynomials, and nonlinear functions defined by nonlinear ordinary differential equation(s). Among possibilities of multiple zero points, one zero point on a prescribed branch is sought; two-point boundary value problem function, nonlinear integral function explicitly defined uniquely through such a principal value, nonlinear integro-differential equation, boundary integration types are not included.

Finally, zeros of single specified forms of a Gauss Hypergeometric Function are given, which are the cases corresponding to no additional polynomial.

Y. Mochimaru (✉)
Tokyo Institute of Technology, Tokyo, Japan
e-mail: ymochima-1947@cx.117.cx

2   Analysis

To find the zero(s) of the target equation regarding initial value parameter(s) or
pure parameter(s), it becomes necessary to solve a truncated polynomial equation,
any function including special functions, infinite series, or their mixed, inclusive of
an additive finite polynomial. For such purposes, use is made of the Taylor series
expansion, the coefficients of which can be obtained analytically or through finite
difference approximation, with a suitable shift to truncate some higher terms com-
pletely, to change to a monic polynomial. Two ways are possible, the first is to divide
all the coefficients of the polynomial by that of the highest degree (if not zero), the
second is to divide all the coefficients by the constant (the lowest, if not zero), and
change the variable of the polynomial by its inverse. The zeros of the monic polyno-
mial are computed by the Jenkins–Traub algorithm; some Fortran routine based on
that is found in NAPACK [3], where slight modification is required. However, many
(not necessary all) of the zeros computed there do not satisfy the original equation
(equality) even approximately, since the said truncation is based on the assumption
(as necessary condition) that the higher terms are small.

2.1   From the 7th Power Equation

$$\left(\frac{dw}{dx}\right)^7 - w^2 \left(\frac{dw}{dx}\right) - w = 0, \tag{1}$$

$$\frac{1}{W}\left(1 + z^2\right) + F(z) = 0, \tag{2}$$

$$F(z) \equiv \left(\frac{dw}{dx}\right)_{x=1}, \quad w(0) = z, \tag{3}$$

where W is a constant given a priori (hereafter). If $\dfrac{dw}{dx} \equiv y\,(w)^{1/3}$, then

$$y^7 - y - \frac{w}{w^{7/3}} = 0. \tag{4}$$

Assuming $\phi(\equiv w/w^{7/3})$, $|\phi| \ll 1$, one root is given by

$$y = 1 + \frac{1}{6}\phi - \frac{7}{72}\phi^2 + \frac{7}{81}\phi^3 + \frac{4515}{31104}\phi^4 + \cdots . \tag{5}$$

Thus, Eq.(1) can be integrated numerically using the Runge–Kutta method. For
getting Taylor coefficients of expansion, select a suitable point(s) (which may depend

on functions and parameters) and choose small difference interval values such as ± 0.002 (which may be complex, in double precision) to get function values to apply the difference formula. If $W = \dfrac{-0.1 + 0.01i}{(0.05 - 0.05\,i)^2}$, then $z = -1.968 - 0.765\,i$.

2.2 Form of Quintic Equation

$$\left(\frac{dw}{dx}\right)^5 - \left(\frac{dw}{dx}\right) - 2 + \frac{w}{2} = 0, \tag{6}$$

$$\frac{1}{W}\left(1 + z^2\right) + F(z) = 0 \;,\; x \in [0, 1], \tag{7}$$

$$w(0) = z\,,\;\; F(z) \equiv w(1). \tag{8}$$

Equation (6) regarding dw/dx can be solved by Modular Function as long as $-2 + w/2$ is given a priori [4], so that if the branch of the solution is fixed, a continuous solution is found as long as $\left(\dfrac{dw}{dx}\right) \neq 0$, i.e., $w \neq 4$ (necessary condition). If $W = \dfrac{0.02 - 0.01\,i}{(0.1 - 0.1\,i)^2}$, then $z = -0.604 + 0.908\,i$.

2.3 Form of Cubic Equation

$$\left(\frac{dw}{dx}\right)^3 - z\,x\left(\frac{dw}{dx}\right) + x^3 = 0, \tag{9}$$

$$\frac{1}{W}\left(1 + z^2\right) + F(z) = 0, \tag{10}$$

$$F(z) \equiv w(1)\,;\; x \in [0, 1], \tag{11}$$

provided that

$$\frac{dw}{dx} \sim z^{0.5}\,x^{0.5} - \frac{x^2}{2\,z} \;(z \neq 0, |x| \ll 1), \tag{12}$$

$$\frac{dw}{dx} = 2\sqrt{\frac{z}{3}}\,x\,\sin\theta, \tag{13}$$

$$\theta = \frac{1}{3i} \ln\left(\sqrt{1 - \omega^2} + i\,\omega\right) \text{ (principal values)}, \tag{14}$$

$$\omega \equiv \frac{3^{1.5}}{2\,z \times z^{0.5}}\, x^{1.5} \ , \ (x \geq 0). \tag{15}$$

If $W = \dfrac{0.5 + 0.1\,i}{(0.5 - 0.2\,i)^2}$, then $z = 0.071 + 1.186\,i$.

2.4 Briot–Bouquet Differential Equation

$$x\,y' = \lambda\,y + a\,x + a_{02}\,y^2 + a_{03}\,y^3 + x\,(a_{11}\,y + a_{12}\,y^2), \tag{16}$$

$$a = z \ , \ x \in [0, 1] \ , \ a_{03} \neq 0, \tag{17}$$

$$\frac{1}{W}\left(1 + z^2\right) + F(z) = 0 \ , \ F(z) \equiv y(1). \tag{18}$$

To be regular at $x = 0$,

$$y(0) = \frac{1}{2\,a_{03}}\left(-a_{02} + \sqrt{a_{02}{}^2 - 4\,\lambda\,a_{03}}\right), \tag{19}$$

$$y'(0) = \left\{\lambda + 2\,a_{02}\,y(0) + 3\,a_{03}\,y^2(0)\right\} y'(0) + \left\{a + a_{11}\,y(0) + a_{12}\,y^2(0)\right\}. \tag{20}$$

If $W = \dfrac{-0.12 + 0.15\,i}{(0.15 - 0.3\,i)^2}$, $\lambda = 2$, $a_{02} = 0.5$, $a_{03} = 1.5$, $a_{11} = 1.2$, $a_{12} = 1.7$, then $z = 0.256 - 1.657\,i$.

2.5 Briot–Bouquet Differential Equation of Order 2 Type

$$x^2\,y' = x f(y) + x^2\,g(y) + h(y), \tag{21}$$

$$f(y) \equiv a_0 + a_1\,y + a_2\,y^2, \tag{22}$$

$$g(y) \equiv z + c_1\,y, \tag{23}$$

$$h(y) \equiv b_1\,y + b_2\,y^2 + b_3\,y^3. \tag{24}$$

To be regular at $x = 0$,

$$f(y_0) = 0, \ y_0 \equiv y(0), \tag{25}$$

$$y_0 = -b_2/(2 b_3) \ , 4 b_1 b_3 - b_2^2 = 0, \tag{26}$$

$$\{f'(y_0) - 1\} \ y_0' + g(y_0) + \frac{1}{2} h''(y_0) \ y_0'^2, \tag{27}$$

$$\frac{1}{W} \left(1 + z^2\right) + F(z) = 0 \ , F(z) \equiv y(1). \tag{28}$$

If $W = \dfrac{-0.12 + 0.15\,i}{(0.15 - 0.3\,i)^2}$, $a_0 = 1, a_1 = 1, a_2 = 1.5, b_1 = 1, b_2 = 1.7, c_1 = 1.3$, then $z = 0.400 - 0.265\,i$.

2.6 Definite Integral of a Special Function

$$F(z) \equiv \int_0^{+\infty} \frac{1}{x^{0.5}} \, {}_2F_1(a, b; c; -z\,x) \, {}_2F_1(a, b; c; -z/x) \, dx, \tag{29}$$

$$\Re(a) > 1/2 \, , \Re(b) > 1/2, \, |\arg(z)| < \pi. \tag{30}$$

(Singular integral at $x = 0$) ${}_2F_1$: Gauss Hypergeometric Function.

$$\frac{1}{W} \left(1 + z^2\right) + F(z) = 0, \tag{31}$$

$$F(z) = \sqrt{\frac{\pi}{z}} \, \Gamma \begin{bmatrix} a - \frac{1}{2}, \ b - \frac{1}{2}, & c \\ a, & b, & c - \frac{1}{2} \end{bmatrix} \times {}_2F_1 \left(2a - 1, 2b - 1; 2c - 1; -z\right), \tag{32}$$

$$\Gamma \begin{bmatrix} \alpha, \ \beta, \ \gamma \\ \alpha^*, \ \beta^*, \ \gamma^* \end{bmatrix} \equiv \frac{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\alpha^*)\Gamma(\beta^*)\Gamma(\gamma^*)}. \tag{33}$$

If $W = \dfrac{0.005 - 0.003\,i}{(0.2 + 0.7\,i)^2}$, $a = 1.5, b = 0.7, c = 1.2$, then $z = 0.024 - 0.092\,i$.

2.7 Definite Integral Including a Special Function

$$F(z) \equiv \int_0^{+\infty} e^{-z\,x} \, {}_1F_1 (a; b; x) \, dx. \tag{34}$$

$_1 F_1$: Confluent Hypergeometric Function

$\Re(z) > 1$ (sufficient condition of integration)

$$\frac{1}{W}\left(1 + z^2\right) + F(z) = 0. \tag{35}$$

If $W = \dfrac{-0.2 + 0.3\,i}{(0.2 - 0.1\,i)^2}$, $a = 2.5$, $b = 0.8$, then $z = 3.259 - 0.050\,i$.

2.8 Inverse Function

Inverse function at particular values of Gauss Hypergeometric Function

$$F(z) \equiv {}_2F_1^{-1}\left(\frac{1}{8}, 1; \frac{9}{8}; z\right), \tag{36}$$

$$\frac{1}{W}\left(1 + z^2\right) + F(z) = 0, \tag{37}$$

$$
\begin{aligned}
{}_2F_1\left(\frac{1}{8}, 1; \frac{9}{8}; z\right) &= F^{-1} \\
&= \frac{1}{8\,z^{1/8}}\left[\ln\frac{1 + z^{1/8}}{1 - z^{1/8}} + \frac{1}{i}\ln\frac{1 + i\,z^{1/8}}{1 - i\,z^{1/8}}\right. \\
&\quad + \frac{1}{2^{0.5}\,i}\ln\frac{1 + i2^{0.5}\,z^{1/8}/(1 - z^{1/4})}{1 - i2^{0.5}\,z^{1/8}/(1 - z^{1/4})} \\
&\quad \left. - \frac{1}{2^{0.5}}\ln\frac{1 - 2^{0.5}\,z^{1/8} + z^{1/4}}{1 + 2^{0.5}\,z^{1/8} + z^{1/4}}\right].
\end{aligned}
\tag{38}
$$

Numerical inverse function values $F(z)$ at complex points are also obtained by the Newton method from F^{-1}. If $W = \dfrac{0.004 - 0.006\,i}{(0.1 - 0.1\,i)^2}$, then $z = 0.343 - 0.147\,i$.

2.9 Nonlinear Second-Order Ordinary Differential Equation, Resulting in Two-Point Boundary Value Problem

$$w'' = w^2 + w + x \ , \ x \in [0, 1], \tag{39}$$

$$w(0) = z\,, \ w(1) = 1\,, \ F(z) = w'(1), \tag{40}$$

$$\frac{1}{W}\left(1+z^2\right) + F(z) = 0. \tag{41}$$

If $W = \dfrac{-0.04 - 0.01\,i}{(0.12 + 0.12\,i)^2}$, then $z = -0.264 + 1.113\,i$.

2.10 Nonlinear Differential Equation Exactly Expressible Through Weierstrass Function

$$y_{zz} + y\,y_z - y^3 - b_0\,y_z - 3\,b_0\,y^2 - 3\,b_0^2\,y - b_0^3 = 0. \tag{42}$$

The suffix z stands for the derivative with respect to z [5].

$$y = -b_0 + \frac{R_z}{R}\ , \quad F(z) \equiv y(z). \tag{43}$$

$R \equiv R(z)$: Weierstrass Function.

$$\frac{1}{W}\left(1+z^2\right) + F(z) = 0. \tag{44}$$

If $\quad \omega_1 = 2,\ \omega_3 = 1 + i$ (half periods), $W = \dfrac{0.2 + 0.1\,i}{(0.2 - 0.25\,i)^2}$, $b_0 = 1.5$, $\quad$ then $z = 0.804 + 1.231\,i$.

2.11 One Internal Gas Explosion Equation

$$r\,\frac{d^2 r}{dt^2} + \frac{3}{2}\left(\frac{dr}{dt}\right)^2 - k^2 \left(\frac{r_0}{r}\right)^{3\gamma} = 0, \tag{45}$$

where r: explosion radius; t: time. Mathematical model: k, r_0, γ: real. Initial condition (at $t = 0$)

$$r = r_0, \tag{46}$$

$$\frac{dr}{dt} = z\ , \quad (\Re(z) > 0), \tag{47}$$

$$\frac{1}{W}\left(1+z^2\right) + F(z) = 0,\ F(z) \equiv \frac{d}{dt}\,r(2). \tag{48}$$

If $W = \dfrac{0.2 + 0.1\,i}{(0.2 - 0.25\,i)^2}$, $r_0 = 10$, $k = 300$, $\gamma = \dfrac{4}{3}$, then $z = 5.470 - 2.63\,i$.

2.12 Mathematical Differential Model From Modified Chandrasekhar's White Dwarf Equation

[Two-point boundary value problem:]

$$y'' + \frac{2}{x}\,y' + \beta\left(y^2 - z\right)^{(3/2)} = 0, \tag{49}$$

$$0 < x_0 \leq x \leq 1, \quad y(x_0) = 1, \ y(1) = z, \ x_0 = 0.03, \ \beta = 0.2, \tag{50}$$

$$\frac{1}{W}\left(1 + z^2\right) + F(z) = 0, \quad F(z) \equiv \int_{x_0}^{1} \left(y - z^{0.5}\right) dx. \tag{51}$$

Originally, β is set as $\beta = 1$ [4], however, different value is introduced to exclude spatial limitations. If $W = \dfrac{-2 + 0.15\,i}{(0.9 - 0.4\,i)^2}$, then $z = 0.399 + 0.200\,i$.

2.13 Mathematical Coupled Differential Model From Laminar Natural Vertical Convection along a Constant Temperature Wall

$$f''' + 3\,f\,f'' - 2\left(f'\right)^2 + T = 0, \tag{52}$$

$$T'' + 3\,Pr\,f\,T' = 0, \quad ' \equiv \frac{d}{dx}, x \geq 0. \tag{53}$$

T: originally from dimensionless temperature, Pr stands for a Prandtl number, where $Pr = 0.7$ is used. With boundary conditions (two-point boundary value problem)

$$f(0) = f'(0) = 0, T(0) = 1, T'(4) = 0, \tag{54}$$

(4 stands for substantially large location value, far away condition) (Fig. 1)

$$\frac{1}{W}\left(1 + z^2\right) + F(z) = 0, \quad F(z) \equiv f(4), \ f''(0) = z. \tag{55}$$

If $W = \dfrac{0.06 - 0.08\,i}{(0.2 + 0.3\,i)^2}$, then $z = 0.730 - 0.287\,i$.

Fig. 1 Solution

$Pr = 0.7$
$z = 0.730 - 0.287\ i$

$\Re\ (T)$
$\Im\ (T)$
$\Re\ (f)$
$\Im\ (f)$
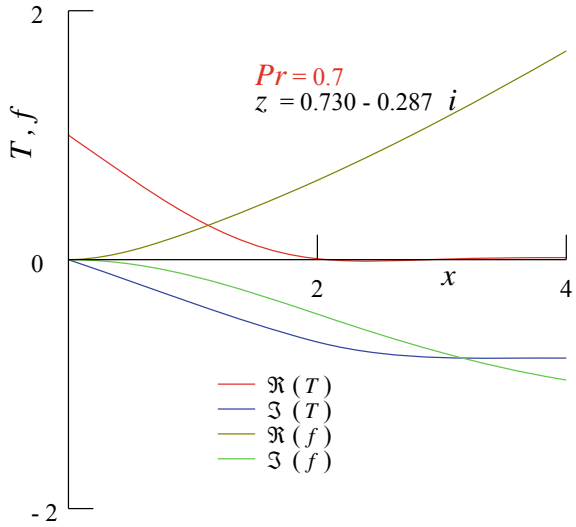
2.14 Integro-Differential Problem of Volterra

$$\frac{dy}{dt} = a\,y - b\,y^2 + y \int_0^t c\,(t-s)\,y(s)\,ds, \quad a,b,c : \text{constant}, \qquad (56)$$

$$\frac{1}{W}\left(1+z^2\right) + F(z) = 0,\ y(0) = z(\neq 0),\ F(z) \equiv y(x_{\max}),\ x_{\max} = 5. \qquad (57)$$

Finally,

$$y''' = 3\,\frac{y'\,y''}{y} - 2\,\frac{(y')^3}{y^2} - b\,y\,y'' + c\,y^2, \qquad (58)$$

$$y'(0) = a\,y(0) - b\,\{y(0)\}^2, \qquad (59)$$

$$y''(0) = a\,y'(0) - 2\,b\,y(0)\,y'(0). \qquad (60)$$

If $W = \dfrac{0.1 + 0.1\,i}{(0.5 - 0.2\,i)^2}$, $a = 1, b = 1, c = -0.1$ (Fig. 2), then $z = -0.073 + 1.072\,i$.
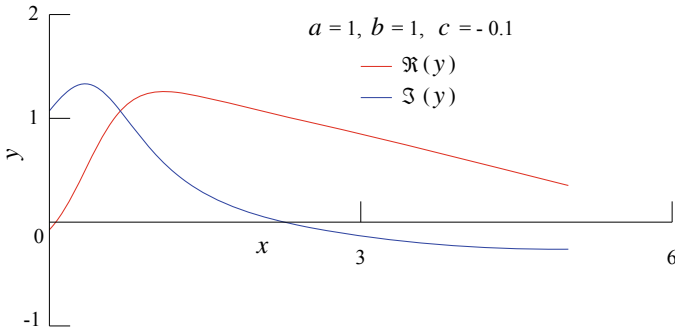
Fig. 2 Solution

Fig. 3 Zeros
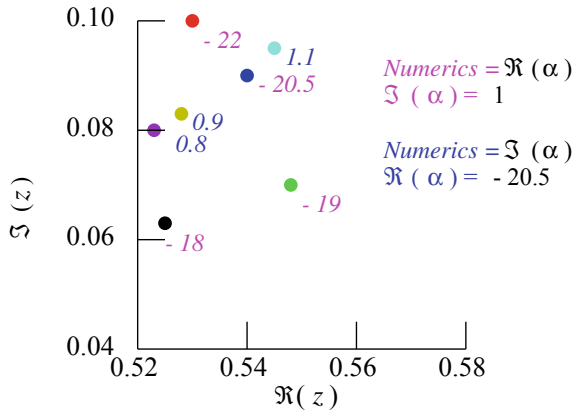
Zeros of $_2F_1(\alpha, \alpha; \alpha + 0.5; z) = 0$

3 Zero-Points of Some Gauss Hypergeometric Functions

They correspond to no additional or missing polynomial case. The first case: $_2F_1(\alpha, \alpha; \alpha + 1/2; z) = 0$. Examples of typical zeros are shown in Fig. 3.

The second case: $_2F_1(\alpha, 1.2\alpha, 1.5\alpha; z) = 0$.

Examples of typical zeros are shown in Fig. 4.

4 Conclusions

Complex transcendental roots of a combination of nonlinear ordinary equation and a polynomial are searched numerically for various kinds of equations. As long as the

Fig. 4 Zeros

Zeros of $_2F_1(\alpha, 1.2\,\alpha; 1.5\,\alpha; z) = 0$
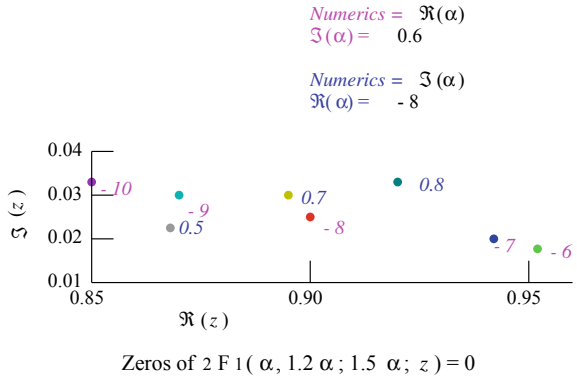
coefficients (appearing explicitly) are moderate (after normalized in some sense), it is possible to find approximately one of the zeros (not necessarily all) if any.

References

1. F.W.J. Olver et.al. (ed.), NIST Handbook of Mathematical Functions, NIST and Cambridge Univ., (2010).
2. N.W. McLachlan, Theory and Application of Mathieu Functions, Dover, (1964).
3. NAPACK, https://www.netlib.org/napack (13/03/2022)
4. H.T. Davis, Introduction to Nonlinear Differential and Integral Equations, Dover, (1962).
5. N.A. Kadryashov, Nonlinear Differential Equations with Exact Solutions Expressed via the Weierstrass Functions, Z. Naturforsh, 59a, 443–454, (2004).

Null Non-controllability for Singular and Degenerate Heat Equation with Double Singular Potential

Nikolai Kutev and Tsviatko Rangelov

Abstract Initial boundary value problem for singular and degenerate heat equation with double singular potential is considered. Such problems appear in many physical and mechanical models. Global existence of weak solutions, as well as null non-controllability in dependance on the parameter of the potential, are obtained. Hardy inequalities with double singular kernels at an interior point and on the boundary of the domain and with optimal constant are used in the proof of the main results.

Keywords Singular and degenerate heat equations · Null non-controllability · Global existence · Hardy inequality

1 Introduction

We consider the parabolic problem

$$
\begin{cases}
u_t - \operatorname{div}(|x|^l \nabla u) - \mu \Psi(x) u = f(t, x), \ \text{in } Q = (0, T) \times \Omega, \\
u(t, x) = 0, \ \text{for } x \in \partial\Omega, \ \ u(0, x) = u_0(x),
\end{cases}
\tag{1}
$$

where $f(t, x) \in L^2\left((0, T); W_{-l}^{-1,2}(\Omega)\right)$, $u_0(x) \in L^2(\Omega)$, $\Omega \subset \mathbb{R}^n$, $0 \in \Omega$, $n > 2$, and Ω is a bounded star-shaped domain with respect to a small ball centered at the origin, i.e.

$$
\Omega = \{x \in \mathbb{R}^n, |x| \leq \varphi(x)\},
\tag{2}
$$

where $\varphi(x) \in C^{0,1}(\mathbb{R}^n)$ is a positive homogeneous function of 0-th order.

N. Kutev · T. Rangelov (✉)
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl.8, 1113 Sofia, Bulgaria
e-mail: rangelov@math.bas.bg

N. Kutev
e-mail: kutev@math.bas.bg

The potential $\Psi(x)$

$$\Psi(x) = |x|^{l-2} \left[1 - \left(\frac{|x|}{\varphi(x)} \right)^{\frac{l+n-2}{2}} \right]^{-2}, l+n-2 > 0, l < 2,$$

is singular on the boundary $\partial\Omega$ and at the origin.

For $l > 0$, Eq. (1) is a degenerate parabolic equation, for $l < 0$, it is a singular parabolic equation, and for $l = 0$, it is a uniformly parabolic one.

Problem (1) occurs in many physical problems, in the non-relativistic quantum mechanics [6, 19], molecular physics [16], in quantum cosmology [6], in the study of near-horizon structure of black holes and other.

After the remarkable paper [4], for existence or blow-up of the solutions to the heat equation with singular at a point potential, their results are extended in different directions. One of them is the null controllability for linear heat equations [15], and sub-linear ones [12], see also the references therein for inverse-square potentials. More general than quadratic singularities at an interior point, motivated in the book [20], are studied in [3]. The case of singular on the whole boundary potential in the n-dimensional case is treated in [5].

The threshold between the global existence of weak solutions and the null non-controllability in the papers mentioned above is the optimal constant in the Hardy inequality with the corresponding singular potential. When the coefficient before the singular term is less than the Hardy constant, by means of Carleman's estimates, global existence holds, while for greater constant, blow-up or null non-controllability occurs.

Finally, let us mention that problem (1) with singular potential only at an interior point is studied in [1, 7] for a degenerate or singular heat equation, where global existence or blow-up of the solutions is proved. Nonexistence of positive solution to more general reaction-diffusion equations with Hardy type potential are studied in [2, 13], see also the references therein.

In the present paper, we extend our result in [14] for the heat equation with double singular potential to degenerate or singular heat equation with the same potential. When the coefficient before the singular term is less than the corresponding optimal Hardy constant, we prove global existence of weak solutions and null non-controllability when the coefficient is greater than the Hardy constant.

2   Preliminaries

Let us denote by $W_{l,0}^{1,2}(\Omega)$, $l \in \mathbb{R}$, $l + n > 2$ the completion of $C_0^\infty(\Omega)$ functions with respect to the norm

$$\|z\|_{W_{l,0}^{1,2}(\Omega)} = \left(\int_\Omega |x|^l |\nabla z|^p dx \right)^{\frac{1}{2}} < \infty. \tag{3}$$

We also consider the condition

$$\lim_{\varepsilon \to 0} \varepsilon^{l+1} \int_{S_\varepsilon} |\text{Tr}z|^2 dS = 0, \quad S_\varepsilon = \{x \in \mathbb{R}^n, |x| = \varepsilon\}, \tag{4}$$

where $\text{Tr}z$ is the trace operator, see [9, Sect. 5.5] and [18, Sect. 1.1.6]. The space $W_{l,0}^{1,2}(\Omega)$ is reflexive and we can define the dual space $W_{-l}^{-1,2}(\Omega)$ of $W_{l,0}^{1,2}(\Omega)$ as $W_{-l}^{-1,2}(\Omega) = \{v \in D'(\Omega); z = \text{div}v, z \in L_{-l}^2(\Omega; \mathbb{R}^n)\}$ where $L_{-l}^2(\Omega; \mathbb{R}^n) = \{z : \Omega \to R \text{ measurable, such that } x^{-\frac{l}{2}}z(x) \in L^2(\Omega)\}$ with norm $\|z\|_{L_{-l}^2(\Omega)} = \left(\int_\Omega |x|^{-l}|z|^2 dx\right)^{\frac{1}{2}}$.

Note that $W_{l,0}^{1,2}(\Omega) \subset W_0^{1,2}(\Omega)$ for $l < 0$ and $W_0^{1,2}(\Omega) \subset W_{l,0}^{1,2}(\Omega)$ for $l \geq 0$. Moreover, if $\bar{\omega} \subset \Omega\setminus\{0\}$, then $W_{l,0}^{1,2}(\omega)$ coincides with $W_0^{1,2}(\omega)$.

In what follows we will use the notation $C_{l,n} = \left(\frac{l+n-2}{2}\right)^2$.

For the proof of null non-controllability in Sect. 3, we need the inequality with potential, singular only at the origin, see (1.5) and Corollary 1.4 in [11].

Theorem 1 *Suppose $\Omega \subset \mathbb{R}^n$, $0 \in \Omega$, $n > 2$, is a bounded domain. For every constant $l + n > 2$, $l < 2$ and every $w(x) \in W_{l,0}^{1,2}(\Omega)$ the Hardy inequality*

$$\int_\Omega |x|^l |\nabla z(x)|^2 dx \geq C_{l,n} \int_\Omega |x|^{l-2} |z(x)|^2 dx, \tag{5}$$

holds. The constant $C_{l,n}$ is optimal.

In the proof of global existence in Sect. 4, the following result for potential $\psi(x)$ is used.

Theorem 2 *Suppose $l + n > 2$ and $\Omega \subset \mathbb{R}^n$, $n > 2$, $l < 2$, $0 \in \Omega$, is bounded starshaped domain with respect to a small ball centered at the origin. For every $z(x) \in W_{l,0}^{1,2}(\Omega)$ satisfying (4), the Hardy inequality*

$$\int_\Omega |x|^l |\nabla z(x)|^2 dx \geq C_{l,n} \int_\Omega \psi(x) z^2(x) dx,$$

holds. The constant $C_{l,n}$ is optimal.

The proof of Theorem 2 follows from Theorem 1.1 of [10].

3 Null Non-controllability for $\mu > C_{l,n}$

In this section, we prove that there is no control for problem (1). First, let us recall some definitions, see [8].

Definition 1 The null controllability means that for every $u_0(x) \in L^2(\Omega)$, there exists a control $f(t, x) \in L^2\left((0, T) \times W_{-l}^{-1,2}(\Omega)\right)$ such that the solution $u(t, x)$ of (1) satisfies $u(T, x) = 0$ for $x \in \Omega$.

If

$$u(t, x) \in L^\infty\left((0, T); L^2(\Omega)\right) \cap L^2\left((0, T); W_{l,0}^{1,2}(\Omega)\right), \tag{6}$$

then following the idea of the optimal control, we consider for every $u_0(x) \in L^2(\Omega)$ the functional

$$J_{u_0}(u, f) = \frac{1}{2} \int_Q u^2(t, x) dx dt + \frac{1}{2} \int_0^T \|f\|_{W_{-l}^{-1,2}(\Omega)}^2 dt,$$

defined in the set

$$D(u_0) = \left\{ (u, f) \in L^2\left((0, T); W_{l,0}^{1,2}(\Omega)\right) \times L^2\left((0, T); W_{-l}^{-1,2}(\Omega)\right) \right\}.$$

Definition 2 A control $f(t, x) \in L^2\left((0, T); W_{-l}^{-1,2}(\Omega)\right)$ is localized in ω if $f(t, x) = 0$ in $\Omega \backslash \bar\omega$ in the sense that for every $\theta(x) \in C^\infty(\Omega \backslash \bar\omega)$, we have $\theta(x) f(t, x) = 0$ in $L^2\left((0, T); W_{-l}^{-1,2}(\Omega)\right)$.

Definition 3 Problem (1) can be stabilized if there exists a constant $C > 0$, such that

$$\inf_{(u, f) \in D(u_0)} J_{u_0}(u, f) \le C \|u_0\|_{L^2(\Omega)}^2, \quad \text{for every } u_0 \in L^2(\Omega).$$

Let us define

$$\Psi_\varepsilon(x) = \left(|x|^2 + \varepsilon\right)^{\frac{l-2}{2}} \left(1 + \varepsilon - \left(\frac{|x|}{\varphi(x)}\right)^{\frac{l+n-2}{2}}\right)^{-2}, \tag{7}$$

and consider the regularized problem

$$\begin{cases} u_{\varepsilon t} - \text{div}\left((|x|^2 + \varepsilon)^{\frac{l}{2}} \nabla u_\varepsilon\right) - \mu \Psi_\varepsilon(x) u_\varepsilon = f(t, x), & \text{in } Q, \\ u_\varepsilon(t, x) = 0, & \text{for } t \in [0, T], x \in \partial\Omega, \\ u(0, x) = u_0(x), & \text{for } x \in \Omega. \end{cases} \tag{8}$$

For $\varepsilon > 0$, problem (8) is well-posed.

We consider the functional

$$J_{u_0}(f) = \frac{1}{2} \int_Q u_\varepsilon^2(t, x) dx dt + \frac{1}{2} \int_0^T \|f\|_{W_{-l}^{-1,2}(\Omega)}^2 dt,$$

for every $f(t, x)$ localized in $[0, T] \times \omega, \bar\omega \subset \Omega$.

Theorem 3 *Suppose $\mu > C_{l,n}$, $n > 2$, $l > 2 - n$, $l < 2$, $f(t, x)$ is localized in ω, $\bar{\omega} \subset \Omega \backslash \{0\}$. Then there is no constant C such that for every $\varepsilon > 0$ and every $u_0(x) \in L^2(\Omega)$ the estimates*

$$\inf_{f \in D_1} J_{u_0}(f) \leq C \|u_0\|_{L^2(\Omega)}^2,$$

holds, where $D_1 = \left\{ f \in L^2\left((0, T); W_{-l}^{-1,2}(\Omega) \right) \right\}$.

For the proof of Theorem 3, we need spectral estimates for the following eigenvalue problem

$$\begin{cases} -\text{div}\left((|x|^2 + \varepsilon)^{\frac{l}{2}} \nabla \phi_1^\varepsilon \right) - \mu \Psi_\varepsilon \phi_1^\varepsilon = \lambda_1^\varepsilon \phi_1^\varepsilon \ \text{in}\ \Omega, \\ \phi_1^\varepsilon = 0, \ \text{on}\ \partial \Omega, \ \|\phi_1^\varepsilon\|_{L^2(\Omega)} = 1, \end{cases} \tag{9}$$

where λ_1^ε is the first eigenvalue and ϕ_1^ε is the corresponding first eigenfunction.

Lemma 1 *If $\mu > C_{l,n}$, $n > 2$, $l > 2 - n$, $l < 2$, then*

$$\lim_{\varepsilon \to 0} \lambda_1^\varepsilon = -\infty, \tag{10}$$

and for every $\bar{\omega} \subset \Omega \backslash \{0\}$

$$\lim_{\varepsilon \to 0} \|\phi_1^\varepsilon\|_{W^{1,2}(\omega)} = 0, \tag{11}$$

Proof If we assume that λ_1^ε is bounded from below with a constant C_1, then from the Reyleigh identity, it follows that

$$\int_\Omega (|x|^2 + \varepsilon)^{\frac{l}{2}} |\nabla z|^2 dx \geq \mu \int \Psi_\varepsilon(x) z^2(x) dx + C_1 \int_\Omega z^2 dx, \tag{12}$$

for every $z \in C_0^\infty(\Omega)$.

For the function $z = \varepsilon^{-\frac{n}{4}} z\left(\varepsilon^{-\frac{1}{4}} x \right)$, $\varepsilon \leq 1$, we get from (12)

$$\mu \varepsilon^{-\frac{n}{2}} \int_\Omega (|x|^2 + \varepsilon)^{\frac{l-2}{2}} \left(1 + \varepsilon - \left(\frac{|x|}{\varphi(x)} \right)^{\frac{l+n-2}{2}} \right)^{-2} z^2\left(\varepsilon^{-\frac{1}{4}} x \right) dx \tag{13}$$
$$\leq \varepsilon^{-\frac{n+1}{2}} \int_\Omega (|x|^2 + \varepsilon)^{\frac{l}{2}} |\nabla z(\varepsilon^{-\frac{1}{4}} x)|^2 dx - C_1 \varepsilon^{-\frac{n}{2}} \int_\Omega z^2(\varepsilon^{-\frac{1}{4}} x) dx$$

Since $\varphi(\varepsilon^{-\frac{1}{4}} x) = \varphi(x)$, after the change of the variables $\varepsilon^{-\frac{1}{4}} x = y$, (13) becomes

$$\mu \varepsilon^{\frac{l-2}{4}} \int_\Omega \left(|y|^2 + \varepsilon^{\frac{1}{2}} \right)^{\frac{l-2}{2}} \left(1 + \varepsilon - \left(\frac{\varepsilon^{\frac{1}{4}} |y|}{\varphi(y)} \right)^{\frac{l+n-2}{2}} \right)^{-2} z^2(y) dy \tag{14}$$
$$\leq \varepsilon^{\frac{l-2}{4}} \int_\Omega \left(|y|^2 + \varepsilon^{\frac{1}{2}} \right)^{\frac{l}{2}} |\nabla z(y)|^2 dy - C_1 \int_\Omega z^2(y) dy.$$

After the limit $\varepsilon \to 0$, we have from (14) and $l + n - 2 > 0, l < 2$ the final inequality

$$\int_\Omega |y|^l |\nabla z(y)|^2 dy \geq \mu \int_\Omega |y|^{l-2} z^2(y) dy, \tag{15}$$

for every $z \in C_0^\infty(\Omega)$, and hence for every $z \in W_{l,0}^{1,2}(\Omega)$. Since $\mu > C_{l,n}$ inequality (15) contradicts the Hardy inequality (5) and (10) is proved.

In order to prove (11), for $\bar{\omega} \subset \Omega \backslash \{0\}$, we fix the sets $\bar{\omega} \subset \bar{\omega}_1 \subset \bar{\omega}_2 \subset \Omega \backslash \{0\}$ and choose $C^2(\Omega)$ smooth function $\eta_1(x)$, $\eta_1 \geq 0$, $\|\eta_1\|_{L^\infty(\Omega)} \leq 1$ satisfying

$$\begin{cases} \eta_1(x) = 1 \ \text{ for } x \in \omega_1, \\ \eta_1(x) = 0 \ \text{ for } x \in \Omega \backslash \omega_2. \end{cases} \tag{16}$$

After multiplication of (9) with $\eta_1 \phi_1^\varepsilon$ and integration, we get

$$\begin{aligned} &- \lambda_1^\varepsilon \int_\Omega \eta_1(\phi_1^\varepsilon)^2 dx + \int_\Omega \eta_1(|x|^2 + \varepsilon)^{\frac{l}{2}} |\nabla \phi_1^\varepsilon|^2 dx \\ &= \mu \int_\Omega \eta_1 \Psi_\varepsilon(x)(\phi_1^\varepsilon)^2 dx - \int_\Omega \eta_1(|x|^2 + \varepsilon)^{\frac{l}{2}} \phi_1^\varepsilon \langle \nabla \phi_1^\varepsilon, \nabla \eta_1 \rangle dx \\ &\leq \mu \sup_{x \in \omega_2} \Psi(x) \int_\Omega \eta_1(\phi_1^\varepsilon)^2 dx + \int_\Omega \eta_1(|x|^2 + \varepsilon)^{\frac{l}{2}} \phi_1^\varepsilon |\nabla \phi_1^\varepsilon||\nabla \eta_1| dx \\ &\leq \mu \sup_{x \in \omega_2} \Psi(x) \int_\Omega \eta_1(\phi_1^\varepsilon)^2 dx + \frac{1}{2} \int_\Omega \eta_1(|x|^2 + \varepsilon)^{\frac{l}{2}} |\nabla \phi_1^\varepsilon|^2 dx \\ &+ \sup_{x \in \omega_2} \left[\sigma(\eta_1)(|x|^2 + \varepsilon)^{\frac{l}{2}} \right] \int_{\omega_2} (\phi_1^\varepsilon)^2 dx, \end{aligned}$$

because $|\nabla \eta_1|^2 \leq 2\eta_1 \sup |\sigma(\eta_1)|$, where $\sigma(z) = \Sigma_{i=1}^n |z_{x_i x_i}|$.

Hence

$$\begin{aligned} &- \lambda_1^\varepsilon \int_\Omega \eta_1(\phi_1^\varepsilon)^2 dx \leq -\lambda_1^\varepsilon \int_\Omega \eta_1(\phi_1^\varepsilon)^2 dx + \frac{1}{2} \int_\Omega \eta_1(|x|^2 + \varepsilon)^{\frac{l}{2}} |\nabla \phi_1^\varepsilon|^2 dx \\ &\leq \left[\mu \sup_{x \in \omega_2} \Psi(x) + \sup_{x \in \omega_2} \left(|\sigma(\eta_1)|(|x|^2 + \varepsilon)^{\frac{l}{2}} \right) \right] \int_{\omega_2} (\phi_1^\varepsilon)^2 dx, \\ &\leq \mu \sup_{x \in \omega_2} \Psi(x) + \sup_{x \in \omega_2} \left[\sigma(\eta_1)(|x|^2 + \varepsilon)^{\frac{l}{2}} \right] < \infty, \end{aligned} \tag{17}$$

and from (10) it follows that $\lim_{\varepsilon \to 0} \int_\Omega \eta_1(\phi_1^\varepsilon)^2 dx = 0$. From the choice of η_1 in (16) we get

$$\lim_{\varepsilon \to 0} \int_{\omega_1} (\phi_1^\varepsilon)^2 dx = 0. \tag{18}$$

Now repeating the above considerations for ω and ω_1, instead of ω_1, ω_2 and function $\eta(x)$ satisfying

$$\begin{cases} \eta(x) = 1 \ \text{ for } x \in \omega, \\ \eta(x) = 0 \ \text{ for } x \in \Omega \backslash \omega_1, \\ \eta(x) \geq 0 \ \|\eta\|_{L^\infty(\Omega)} \leq 1, \end{cases}$$

inequality (17) becomes

$$\frac{1}{2} \int_\omega (|x|^2 + \varepsilon)^{\frac{l}{2}} |\nabla \phi_1^\varepsilon|^2 dx$$
$$\leq \frac{1}{2} \int_\Omega \eta (|x|^2 + \varepsilon)^{\frac{l}{2}} |\nabla \phi_1^\varepsilon|^2 dx - \lambda_1^\varepsilon \int_\Omega \eta (\phi_1^\varepsilon)^2 dx \tag{19}$$
$$\leq \left[\mu \sup_{x \in \omega_1} \Psi(x) + \sup_{x \in \omega_1} \left(\sigma(\eta)(|x|^2 + \varepsilon)^{\frac{l}{2}} \right) \right] \int_{\omega_1} (\phi_1^\varepsilon)^2 dx.$$

From (18), (19), and $0 \notin \omega$, we obtain

$$\lim_{\varepsilon \to 0} \int_\omega (|x|^2 + \varepsilon)^{\frac{l}{2}} |\nabla \phi_1^\varepsilon|^2 dx = \lim_{\varepsilon \to 0} \int_\omega |x|^l |\nabla \phi_1^\varepsilon|^2 dx = 0,$$

which proves (11).

Proof (*Proof of Theorem* 3). We choose $u_0(x) = \phi_1^\varepsilon(x)$ and define the functions

$$a_\varepsilon(t) = \int_\Omega u_\varepsilon(t, x) \phi_1^\varepsilon(x) dx, \quad b_\varepsilon(t) = \int_\Omega f(t, x) \phi_1^\varepsilon(x) dx,$$

where ε is small enough such that $\lambda_1^\varepsilon < 0$ according to (10) and $u_\varepsilon(t, x)$ is solution of (8) satisfying (6). Simple computations give us after integration by parts the equalities

$$\begin{aligned}
a_\varepsilon(t) &= \int_\Omega \phi_1^\varepsilon(x) \left[\text{div} \left((|x|^2 + \varepsilon)^{\frac{l}{2}} \nabla u_\varepsilon(t, x) \right) + \mu \Psi_\varepsilon(x) u_\varepsilon(t, x) + f(t, x) \right] \\
&= \int_\Omega u_\varepsilon(t, x) \left[\text{div} \left((|x|^2 + \varepsilon)^{\frac{l}{2}} \nabla \phi_1^\varepsilon(x) \right) + \mu \Psi_\varepsilon(x) \phi_1^\varepsilon(x) \right] dx \\
&\quad + \int_\Omega \phi_1^\varepsilon(x) f(t, x) dx = -\lambda_1^\varepsilon \int_\Omega \phi_1^\varepsilon(x) u_\varepsilon(t, x) dx + \int_\Omega \phi_1^\varepsilon(x) f(t, x) dx \\
&= -\lambda_1^\varepsilon a_\varepsilon(t) + b_\varepsilon(t).
\end{aligned} \tag{20}$$

From (20) we get

$$a_\varepsilon(t) = e^{-\lambda_1^\varepsilon t} + \int_0^t e^{-\lambda_1^\varepsilon(t-s)} b_\varepsilon(s) ds.$$

Since $f(t, x)$ is localized in $[0, T] \times \omega$, $\bar{\omega} \subset \Omega \backslash \{0\}$, we obtain the estimates

$$b_\varepsilon^2(t) \leq \|f\|^2_{W^{-1,2}_{-l}(\omega)} \|\phi_1^\varepsilon\|^2_{W^{1,2}(\omega)}, \quad \int_0^T a^2(t) dt \leq \int_Q u_\varepsilon^2(t, x) dx dt \tag{21}$$

and

$$\begin{aligned}
&\int_0^T \left[e^{-\lambda_1^\varepsilon t} + \int_0^t e^{-\lambda_1^\varepsilon(t-s)} b_\varepsilon(s) ds \right]^2 dt \\
&\geq \frac{1}{2} \int_0^T e^{-2\lambda_1^\varepsilon t} dt - \int_0^T \left[\int_0^t e^{-\lambda_1^\varepsilon(t-s)} b_\varepsilon(s) ds \right]^2 dt \\
&\geq -\frac{1}{4\lambda_1^\varepsilon} \left(e^{-2\lambda_1^\varepsilon T} - 1 \right) - \int_0^T \left[\int_0^t e^{-2\lambda_1^\varepsilon(t-s)} ds \int_0^t b_\varepsilon^2(s) ds \right] dt \\
&= -\frac{1}{4\lambda_1^\varepsilon} \left(e^{-2\lambda_1^\varepsilon T} - 1 \right) - \frac{1}{\lambda_1^\varepsilon} \int_0^T \left(1 - e^{-2\lambda_1^\varepsilon t} \right) \int_0^t b_\varepsilon^2(s) ds dt \\
&\geq -\frac{1}{4\lambda_1^\varepsilon} \left(e^{-2\lambda_1^\varepsilon T} - 1 \right) \left[1 - \frac{1}{4\lambda_1^\varepsilon} \|\phi_1^\varepsilon\|^2_{W^{1,2}(\omega)} \int_0^T \|f\|^2_{W^{-1,2}_{-l}(\omega)} dt \right].
\end{aligned} \tag{22}$$

From (21) and (22), we get the inequality

$$-\frac{1}{4\lambda_1^\varepsilon}\left(e^{-2\lambda_1^\varepsilon T}-1\right)\le\int_Q u_\varepsilon^2 dxdt$$
$$+\frac{1}{4(\lambda_1^\varepsilon)^2}\left(e^{-2\lambda_1^\varepsilon T}-1\right)\|\phi_1^\varepsilon\|_{W^{1,2}(\omega)}^2\int_0^T\|f(t,.)\|_{W_{-l}^{-1,2}(\omega)}^2 dt, \tag{23}$$

and (23) means that either

$$-\frac{1}{8\lambda_1^\varepsilon}\left(e^{-2\lambda_1^\varepsilon T}-1\right)\le\int_Q u_\varepsilon^2 dxdt \ \ \text{or}$$
$$-\frac{1}{8\lambda_1^\varepsilon}\left(e^{-2\lambda_1^\varepsilon T}-1\right)\le\frac{1}{4(\lambda_1^\varepsilon)^2}\left(e^{-2\lambda_1^\varepsilon T}-1\right)\|\phi_1^\varepsilon\|_{W^{1,2}(\omega)}^2 \tag{24}$$
$$\times\int_0^T\|f(t,.)\|_{W_{-l}^{-1,2}(\omega)}^2 dt.$$

From (23) and (24), we have the final estimate.

$$J_{u_0}^\varepsilon(f)\ge\min\left\{-\frac{1}{16\lambda_1^\varepsilon}\left(e^{-2\lambda_1^\varepsilon T}-1\right),-\frac{\lambda_1^\varepsilon}{4\|\phi_1^\varepsilon\|_{W^{1,2}(\omega)}}\right\},$$

and from (10) and (11), it follows that $\lim_{\varepsilon\to 0}J_{u_0}^\varepsilon(f)=\infty$ which proves Theorem 3.

4 Global Existence for $\mu < C_{l,n}$

In this section, we prove existence of a weak global solution $u(t,x)$

$$u(t,x)\in C^0\left((0,\tau);L^2(\Omega)\right)\cap L^2\left((0,\tau);W_{l,0}^{1,2}(\Omega)\right),\ \ \text{for all }\tau>0, \tag{25}$$

of the problem (1) when $l+n>2$, $l<2$, $\mu<C_{l,n}$ and Ω satisfies (2).

Theorem 4 *Let $\Omega\subset\mathbb{R}^n$, $n>2$, $0\in\Omega$, be a star-shaped domain with respect to a small ball centered at the origin, (2) holds and $\mu<C_{l,n}$, $l+n>2$, $l<2$. If $f(t,x)\in L^2\left((0,T);W_{-l}^{-1,2}(\Omega)\right)$, $u_0(x)\in L^2(\Omega)$, then problem (1) has a weak distributional solution $u(t,x)$ satisfying (25).*

Proof We define with $N\ge 1$ the functions

$$h_N(x)=\begin{cases}|x|^l, & \text{for }l\le 0,\\ |x|^l+\frac{1}{N}, & \text{for }l>0,\end{cases}$$
$$\Psi_N(x)=\left(|x|^{2-l}+\frac{1}{N}\right)^{-1}\left(1+\frac{1}{N}-\left(\frac{|x|}{\varphi(x)}\right)^{\frac{l+n-2}{2}}\right)^{-2}.$$

The problem

$$\begin{cases}u_{N,t}-\text{div}\,(h_N(x)\nabla u_N)-\mu\Psi_N(x)u_N=f(t,x), & \text{in }Q,\\ u_N(t,x)=0, & \text{for }(t,x)\in(0,T)\times\partial\Omega, \\ u_N(0,x)=u_0(x), & \text{for }x\in\Omega,\end{cases} \tag{26}$$

has unique solution in the distributional sense, see for example Proposition 1 in [7] and Theorem 1.2 in [17].

Multiplying (26) with $u_N(t, x)$ and integrating we get from the Cauchy–Bunyakovski inequality the estimate

$$
\begin{aligned}
&\tfrac{1}{2} \int_\Omega \left(u_N^2(T, x) - u_0^2(x) \right) dx + \int_Q h_N(x) |\nabla u_N(t, x)|^2 dx dt \\
&- \mu \int_Q \Psi_N(x) u_N^2(t, x) dx dt = \int_Q f(t, x) u_N(t, x) dx dt \\
&\leq \int_0^T \| f(t, .) \|_{W_{-l}^{-1,2}(\Omega)} \| u_N(t, .) \|_{W_{l,0}^{1,2}(\Omega)} dt \\
&\leq \tfrac{\delta}{2m} C_{l,n}^{-1} \int_0^T \| u_N(t, .) \|_{W_{l,0}^{1,2}(\Omega)}^2 dt \\
&+ \tfrac{1}{2} \left(\tfrac{m}{\delta} \right) C_{l,n} \int_0^T \| f(t, .) \|_{W_{-l}^{-1,2}(\Omega)}^2 dt,
\end{aligned}
$$

with

$$
m = \min_{x \in \Omega} \Psi_1(x) \text{ and } \delta = \frac{m}{2} \left[C_{l,n} - \mu \right] > 0.
$$

From Theorem 2 and the definition (3) of $W_{l,0}^{1,2}(\Omega)$, we obtain the following chain of inequalities:

$$
\begin{aligned}
0 &\leq \tfrac{1}{2} \int_\Omega u_0^2(x) dx - \int_Q h_N(x) |\nabla u_N(t, x)|^2 dx dt + \left(\mu + \tfrac{\delta}{2m} \right) \int_Q \Psi(x) |u_N(t, x)|^2 dx dt \\
&- \tfrac{\delta}{2m} \int_Q \Psi_N(x) |u_N(t, x)|^2 dx dt + \tfrac{\delta}{2m} C_{l,n}^{-1} \int_Q |x|^l |\nabla u_N(t, x)|^2 dx dt \\
&+ \tfrac{1}{2} \left(\tfrac{m}{\delta} \right) C_{l,n} \int_0^T \| f(t, .) \|_{W_{-l,0}^{-1,2}(\Omega)}^2 dt \\
&\leq C_2 - \tfrac{\delta}{2} \int_Q |u_N(t, x)|^2 dx dt + \left(\mu + \tfrac{2\delta}{2m} \right) C_{l,n}^{-1} \int_Q |x|^l |\nabla u_N(t, x)|^2 dx dt \\
&\leq C_2 - \int_Q \left[1 - \left(\mu + \tfrac{\delta}{2m} \right) C_{l,n}^{-1} \right] |x|^l |\nabla u_N(t, x)|^2 dx dt - \tfrac{\delta}{2} \int_Q |u_N(t, x)|^2 dx dt,
\end{aligned}
$$

where the constant $C_2 = \tfrac{1}{2} \int_\Omega u_0^2(x) dx + \tfrac{1}{2} \left(\tfrac{m}{\delta} \right) C_{l,n} \int_0^T \| f(t, .) \|_{W_{-l}^{-1,2}(\Omega)}^2 dt$ is independent of N and $u_N(t, x)$.

From the choice of δ and $h_N(x)$, we get the final estimate

$$
\frac{3}{4} C_{l,n} \left[C_{l,n} - \mu \right] \int_Q |x|^l |\nabla u_N(t, x)|^2 dx dt + \frac{m}{4} \left[C_{l,n} - \mu \right] \int_Q |u_N(t, x)|^2 dx dt \leq C_2,
$$

i.e., the norms

$$
\| u_N \|_{L^2\left(0,T; W_{l,0}^{1,2}(\omega)\right)} \leq C_3, \quad \| u_N \|_{L^\infty\left(0,T; L^2(\omega)\right)} \leq C_3, \tag{27}
$$

are uniformly bounded with a constant C_3 independent of N. By means of (27), the limit of a subsequence of $\{u_N(t, x)\}$, $N \to \infty$ gives a weak solution $u(t, x)$ in the distributional sense of (1). The convergence of $\{u_{N_k}(t, x)\}$ under the estimates (27) without changes follows from the proof of Theorem 3.1 in [7] and we omit it. Theorem 4 is proved.

References

1. Abdellaoui, B., Colorado, E., Peral, I.: Existence and nonexistence results for a class of linear and semi-linear parabolic equations related to some Caffarelli-Khon- Nirenberg inequalities. J. Eur. Math. Soc. **6**, 119—148 (2004)
2. Abdellaoui, B., Miri, S.E., Peral, I., Touaoula, T.M.: Some remarks on quasilinear parabolic problems with singular potential and a reaction term. Nonlinear Differ. Equ. Appl. **21**, 453–490 (2014)
3. Anita, S., Tataru, D.: Exact controllability of the superlinear heat equation. Appl.Math. Optim. **42**, 73–89 (2000)
4. Baras, P., Goldstein, J.A.: The heat equation with singular potential. Trans. Am. Math. Soc. **284**, 121–139 (1984)
5. Boccardo, L., Murat, F.: Almost everywere convergence of the gradients of solutions to elliptic and parabolic equations. Nonlinear Anal. **19**(6), 437–477 (1992)
6. de Castro, A.: Bound states of the Dirac equation for a class of effective quadratic plus inversely quadratic potentialss. Ann. Physics **311**, 170–181 (2004)
7. Dall'Aglio, A., Giachetti, D., Peral, I.: Results on parabolic equations related to Caffarelli-Khon-Nirenberg inequalities. SIAM J. Math. Anal. **36**(3), 691—716 (2004)
8. Ervedoza, S.: Control and stabilization properties for a singular heat equation with an inverse-square potential. Comm. Part. Diff. Eq. **33**(11), 1996—2019 (2008)
9. Evans, L.: Partial Differential Equations, Graduate Studies of Mathematics, v.19. AMS, Providence, Rhode Island (1998)
10. Fabricant, A., Kutev, N., Rangelov, T.: Hardy-type inequality with double singular kernels. Centr. Eur. J. Math. **11**(9), 1689–1697 (2013)
11. Fabricant, A., Kutev, N., Rangelov, T.: Note on sharp Hardy-type inequality. Mediterr. J. Math. **11**, 31–44 (2014)
12. Fursikov, A.V., Imanuvilov, O.Y.: Controllability of Evolution Equations, Lecture Notes Ser., vol. 34. Seoul National University, Seoul (1996)
13. Iagar, R.G., Sńchez, A.: Instantaneous and finite time blow-up of solutions to a reaction-diffusion equation with Hardy-type singular potential. J. Math. Anal. Appl. **491**(1), 124,244:1—10 (2020)
14. Kutev, N., Rangelov, T.: Control properties for heat equation with double singular potential. Compt. Rend. Acad. Sci. Bulgar. **75**(2), 187–196 (2022)
15. Lebeau, G., Robbiano, L.: Controle exact de l'équation de la chaleur. Comm Part. Diff. Eq. **30**, 335–357 (2012)
16. Levy-Leblond, J.M.: Electron capture by polar molecules. Phys. Rev. **153**(1), 1–4 (1967)
17. Lions, J.L.: Quelques méthodes de résolution des problèmes aux limites non linéaires. Dunod, Gauthier-Villar, Paris (1969)
18. Maz'ja, V.G.: Sobolev Spaces. Springer Verlag, Berlin (1985)
19. Peral, I., Vazquez, J.: On the stability or instability of the singular solution of the semilinear heat equation with exponential reaction term. Arch. Ration. Mech. Anal. **129**, 201–224 (1995)
20. Reed, M., Simon, B.: Methods of Morden Mathematical Physics: II Fourier Analysis, Self-Adjointness. Academic Press, INC, London (1975)

Special Functions and Polynomials Connected to the Simple Equations Method (SEsM)

Nikolay K. Vitanov

Abstract We discuss the Simple Equations Method (SEsM) for obtaining exact solutions of nonlinear differential equations. The discussion is focused on two topics: (i) Specific special function (called V-function) which occurs when the solved nonlinear differential equation possesses polynomial nonlinearities; (ii) Two sets of polynomials which occur for the case when the differential equation which is used as a simple equation in the methodology has specific simple form. We present several results on the properties of the V-function and on the properties of the sets of the studied polynomials.

Keywords SEsM · Nonlinear differential equations · Exact solutions · V-function · Special sets of polynomials

1 Introduction

The mathematical instruments for study of Nature and society became powerful enough in order to allow studies of the numerous characteristics of the complex systems from the Nature and society [1–16] such as their nonlinearity [17–23]. The methods of the time series analysis and the methods based on model nonlinear differential or difference equations [24–36] are very suitable for study of nonlinear systems. Various approaches to deal with the nonlinearity of the model equations are used. Examples are the Method of Inverse Scattering Transform [37, 38] and the method of Hirota [39]. In addition Kudryashov formulated the Method of Simplest Equation (MSE) [40, 41]. MSE is based on determination of singularity order n of the solved NPDE and on searching of specific solution of this equation as series containing powers of solutions of an equation called simplest equation (two examples are [42, 43]).

N. K. Vitanov (✉)

Institute of Mechanics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 4, 1113 Sofia, Bulgaria

e-mail: vitanv@imbm.bas.bg

12 years ago we have proposed a methodology for obtaining exact and approximate solutions of nonlinear partial differential equations known today as Simple Equations Method (SEsM) [44–49]. Our work on elements of this methodology began many years ago [50–56]. In [57, 58] and in 2010 we have used the ordinary differential equation of Bernoulli as simplest equation [59] and we have applied the simplest version of SEsM: the Modified Method of Simplest Equation (MMSE) to ecology and population dynamics [60]. In MMSE we have used the concept of the balance equation in order to determine the kind of the simplest equation and the kind of the solution of the solved equation [61, 62]. Till 2018 our work on the methodology and its application have been based on the MMSE [63–72]. An important article from this period was [71]. There, the methodology of the MMSE was extended to to simplest equations of the class $\left(\frac{d^l h}{d\xi^l}\right)^k = \sum_{i=0}^{m} d_i h^i$ where $k = 1, \ldots, l = 1, \ldots,$ and m and d_i are parameters.

The last version of the methodology (called SEsM—Simple Equations Method) allows us to use of more than one simplest equation. Application of SEsM based on two simple equations can be seen in [73]. The first description of SEsM was in [44] and then in [46–49]. For more applications of SEsM see [74–78].

Below we will discuss two aspects of the application of SEsM. One aspect is connected to a special function which arises in the application of SEsM for obtaining solutions of nonlinear equations having polynomial nonlinearity. For the same class of equations specific series of polynomials arise. This is the second aspect of the methodology which will be discussed below.

2   Simple Equations Method (SEsM)

SEsM is an algorithm for obtaining exact and approximate solutions of systems of nonlinear differential equations. The solutions are constructed by known solutions of m more simple differential equations. Here, we describe SEsM for the case of obtaining an exact solution of a single nonlinear differential equation.

SEsM has 4 steps. We want to solve a nonlinear partial differential equation $\mathcal{G}[v(x, \ldots, t), \ldots,] = 0$. Here $\mathcal{G}[v(x, \ldots, t), \ldots]$ depends on the functions $v(x, \ldots, t)$ and some of their derivatives (v can be a function of several spatial coordinates). The steps of SEsM are as follows. Step 1 is to transform the nonlinearity. We apply transformation $v(x, ..., t) = R[F_1(x, \ldots, t), \ldots]$. $R(F_1, \ldots)$ is a composite function of other functions $F_1, F_2, \ldots$. $F_i(x, \ldots, t)$ are functions of several spatial variables as well as of the time. The goal of R is to transform the nonlinearity of the solved equation to more simple kind of nonlinearity The transformation leads to nonlinear differential equations for the functions F_i. At Step 2 the functions $F_i(x, ..., t)$ are chosen as a composite functions of other functions $g_{i1}, ..., g_{iN}$. The functions g_i are constructed by known solutions of differential equations which are more simple than the solved equation. At Step 3 we determine the form of the simple equations and their solutions. For the case of polynomial nonlinearities in the solved equations,

the methodology determines the form of the simple equations by means of use of additional relationships called balance equations. At Step 4, we apply the steps 1–3 to the solved equation and obtain a system of nonlinear algebraic equations for the coefficients of the solved nonlinear differential equation and for the coefficients of the solution. Any nontrivial solution of this system corresponds to a solution the solved nonlinear differential equation.

3   The V-Function

Let us consider a special class of equations which have polynomial nonlinearities with respect of the unknown function and its derivatives. Here, we consider the specific case of one spatial variable and time and we search for a solution of the kind $g(x, t) = g(\xi)$; $\xi = \mu x + \nu t$, where μ and ν are parameters. The basis of our search will be a solution $h(\xi)$ of a certain simple equation. Then, $g = f[h(\xi)]$. We assume that f is a polynomial of h. Then $f = \sum_{r=0}^{q} b_r h^r$. We use the simple equation

$$h_{(l)}^k = \left(\frac{d^l h}{d\xi^l} \right)^k = \sum_{i=0}^{m} a_i h^i. \tag{1}$$

In (1) k, l, m are integers. We denote the solution of (1) as function

$$V_{a_0, a_1, \ldots, a_m}(\xi; l, k, m),$$

where l is the order of derivative of h; k is the degree of derivative, and m is the highest degree of the polynomial of g. The trigonometric, hyperbolic, elliptic functions of Jacobi, etc. are specific case of the function V.

Let us consider several features of the function V. We set first $k = 1$ and $l = 1$. (1) becomes $h_{(1)} = \left(\frac{dh}{d\xi} \right) = \sum_{i=0}^{m} a_i h^i$. For the case $m = 0$ we obtain $V_{a_0}(\xi; 1; 1; 0) = a_0 \xi + C$, where C is a constant of integration. For the case $m = 1$ we have $h_{(1)} = \left(\frac{dh}{d\xi} \right) = a_0 + a_1 h$. For the case $a_0 = 0$, $a_1 \neq 0$ the solution is $V_{0,a_1}(\xi; 1; 1; 1) = C \exp[a_1 \xi]$, where C is a constant of integration. The solution for the case $a_0 \neq 0, a_1 \neq 0$ is $V_{a_0,a_1}(\xi; 1; 1; 1) = \exp[a_1 \xi] \left[C - \frac{a_0}{a_1} \exp[-a_1 \xi] \right]$. For the case $m = 2$ we obtain $h_{(1)} = \left(\frac{dh}{d\xi} \right) = a_0 + a_1 h + a_2 h^2$. This is an equation of Riccati kind. One solution of this equation is $h^*(\xi) = -\frac{a_1}{2a_2} - \frac{\theta}{2a_2} \tanh \left[\frac{\theta(\xi + \xi_0)}{2} \right]$. Above ξ_0 is a constant of integration. We can use this solution in order to obtain the general solution $V_{a_0,a_1,a_2}(\xi; 1; 1; 2)$ of the above equation of Riccati kind. The substitution $u = \frac{1}{h - h^*}$, transforms the equation to the linear equation $\frac{du}{d\xi} + (2a_2 h^* + a - 1)u + a_2 = 0$. The solution of this equation is $u = \exp \left[- \int d\xi (2a_2 h^* + a_1) \right] \left[C - a_2 \exp \left[\int d\xi (2a_2 h^* + a_1) \right] \right]$. Then

$$V_{a_0,a_1,a_2}(\xi; 1; 1; 2) = -\frac{a_1}{2a_2} - \frac{\theta}{2a_2} \tanh\left[\frac{\theta(\xi + \xi_0)}{2}\right] +$$

$$\tanh\left[\frac{\theta(\xi + \xi_0)}{2}\right]^2 \left\{C - a_2 \tanh\left[\frac{\theta(\xi + \xi_0)}{2}\right]^2\right\}^{-1}, \tag{2}$$

where C is a constant of integration.

Let us now consider the case $k = 2$, $l = 1$. In this case (1) becomes $h_{(1)}^2 = \left(\frac{dh}{d\xi}\right)^2 = \sum_{i=0}^{m} a_i h^j$. Let $m = 0$. Then $\left(\frac{dh}{d\xi}\right)^2 = a_0$. The solution is $V_{a_0}(\xi; 1; 2; 0) = a_0^{1/2}\xi + C$, where C is a constant of integration. Thus, we have the following relationship between different V-functions.

$$V_{a_0^2}(\xi; 1; 2; 0) = V_{a_0}(\xi; 1; 1; 0) \tag{3}$$

We can continue the calculation of the V functions and their specific cases. This will be done elsewhere.

4 The Polynomials

Here we mention a theorem (for details about other related propositions and their proofs see [71, 79]). The theorem is for the case of application of SEsM where the composite function depends on a function of a single variable and the last function is a solution of a differential equation containing polynomial non-linearity. We consider a nonlinear partial differential equation with nonlinearities which are polynomials of the unknown function $h(x, t)$ and its derivatives. We search for a solution of the kind $i(x, t) = i(\xi)$; $\xi = \epsilon x + \delta t$, where ϵ and δ are parameters. The basis of our search will be a solution $h(\xi)$ of a certain simple equation. Then, $i = f[h(\xi)]$. We assume that f is a polynomial of h. Then $f = \sum_{s=0}^{r} b_s h^s$. We use the simple equation (1). In (1) k, l, m are integers. Below we note a theorem in which the function $V_{a_0,a_1,\ldots,a_m}(\xi; 1, 2, m)$ participates. This function is solution of the simple equation $h_{(1)}^2 = \left(\frac{dh}{d\xi}\right)^2 = \sum_{j=0}^{m} a_j h^j$. The theorem is [71]

Theorem 1 *If $h_{(1)}^2$ is given by equation $h_{(1)}^2 = \left(\frac{dh}{d\xi}\right)^2 = \sum_{j=0}^{m} a_j h^j$ and f is a polynomial of h given by equation $f = \sum_{s=0}^{r} b_s h^s$, then for $i[f(h)]$ the following relationship holds*

$$i_{(n)} = X_n(q, m)(h) + h_{(1)} Y_n(q, m)(h)$$

where $X_n(q, m)(g)$ and $Y_n(q, m)(g)$ are polynomials of the function $h(\xi)$.

The theorem allows us to calculate the derivatives of composite functions. The polynomials are calculated as follows. $X_0 = \sum_{s=0}^{r} b_s h^s$; $Y_0 = 0$. Starting from here, we obtain $X_{n+1} = \frac{Y_n}{2} \sum_{j=0}^{m} j a_j h^{j-1} + \frac{dY_n}{dh} \sum_{j=0}^{m} a_j H^j$, $Y_{n+1} = \frac{dX_n}{dh}$.

The equations of Bernoulli and Riccati are specific cases of the simple equation

$$h_{(1)} = \sum_{j=0}^{n} d_j h^j. \tag{4}$$

In (4) n and d_j are constant parameters. (4) is specific case of $f = \sum_{s=0}^{r} b_s h^s$.

For the case when the simple equation has the specific form (4) we have to calculate a single kind of polynomial M_n. In other words, when the simple equation is of the kind (4) $i_{(n)}$ is a polynomial of h: $i_{(n)} = M_n(h)$. M_n can be calculated as follows. We start from $M_0 = \sum_{s=0}^{r} b_s h^s$. Then we use the recurrence relationship $M_{j+1} = \frac{dM_j}{dh} \sum_{k=0}^{m} c_k h^k$.

Below we calculate several of the polynomials X_n and Y_n. We start from $X_0 = \sum_{s=0}^{r} b_s h^s$; $Y_0 = 0$. By the recurrence relationships we obtain $X_1 = 0$; $Y_1 = \sum_{s=0}^{r} s b_s h^{s-1}$. We can continue without any problems. We can calculate also several of the polynomials M_i for the case (4) We start from $M_0 = \sum_{s=0}^{r} b_s h^s$. The application of the recurrence relationship for these polynomials leads to the following relationships for M_1, ... $M_1 = \sum_{s=0}^{r} \sum_{j=0}^{m} b_s s c_j h^{s+j-1}$, $M_2 = \sum_{s=0}^{r} \sum_{j=0}^{m} \sum_{k=0}^{m} b_s s(s+j-1) c_j c_k h^{s+j+k-2}$, etc.

5 Concluding Remarks

In this text we discussed two aspects of the application of the Simple equations Method: (i) the occurrence and selected features of a special function (the V function) and (ii) the occurrence of two classes of polynomials connected to the class of possible simple equations which posses polynomial nonlinearities. We derived relationships for the simplest cases of the V-function. We obtained relationships for the specific cases of the discussed polynomials. Obviously, this research open a large field for new results. These results will be reported elsewhere.

References

1. Nicolis G., Nicolis C.: Foundations of Complex Systems. World Scientific, New Jersey (2012)
2. Levin. R.: Complexity. Life at the Edge of Chaos. The University of Chicago Press, Chicago (1999)
3. Ivanova, K., Ausloos, M.: Application of the detrended fluctuation analysis (DFA) method for describing cloud breaking. Physica A: Statistical Mechanics and its Applications 274, 349–354 (274). https://doi.org/10.1016/S0378-4371(99)00312-X

4. Dimitrova, Z. I.: Fluctuations and dynamics of the chaotic attractor connected to an instability in a heated from below rotating fluid layer. Compt. rend. Acad. bulg. Sci 60, 1065–1070 (2007)

5. Vitanov N.K.: Science Dynamics and Research Production. Indicators, Indexes, Statistical Laws and Mathematical Models. Springer, Cham (2016)

6. Nikolova, E. V., Serbezov, D. Z., Jordanov, I.: Nonlinear spread waves in population dynamics including a human-induced Allee effect. AIP Conference Proceedings vol. 2075, 150004 (2019). https://doi.org/10.1063/1.5091327

7. Vitanov, N. K., Vitanov, K. N.: Discrete-time model for a motion of substance in a channel of a network with application to channels of human migration. Physica A: Statistical Mechanics and its Applications 509, 635–650 (2018). https://doi.org/10.1016/j.physa.2018.06.076

8. Dimitrova, Z.I.: On the nonlinear dynamics of interacting populations. Effects of delay on populations substitution. Compt. rend. Acad. bulg. Sci 61, 1541–1548 (2008)

9. Vitanov, N. K., Ausloos, M., Rotundo, G.: Discrete model of ideological struggle accounting for migration. Advances in Complex Systems 15, 1250049 (2012). https://doi.org/10.1142/S021952591250049X

10. Vitanov, N. K., Vitanov, K. N.: Box model of migration channels. Mathematical Social Sciences 80, 108–114 (2016). https://doi.org/10.1016/j.mathsocsci.2016.02.001

11. Kutner, R., Ausloos, M., Grech, D., Di Matteo, T., Schinckus, C., Stanley, H. E.: Econophysics and sociophysics: Their milestones & challenges. Physica A 516, 240–253 (2019). https://doi.org/10.1016/j.physa.2018.10.019

12. Dimitrova, Z. I., Hoffmann N. P.: On the probability for extreme water levels of the river Elba in Germany. Compt. rend. Acad. bulg. Sci 65, 153–160 (2012)

13. Vitanov, N. K., Vitanov, K. N.: On the motion of substance in a channel of a network and human migration. Physica A: 490, 1277–1294 (2018). https://doi.org/10.1016/j.physa.2017.08.038

14. Chen, W.-K.: Theory of Nets. Flows in Networks. Imperial College Press, London, UK (2003)

15. Jordanov, I. P., Nikolova, E. V.: On the evolution of nonlinear density population waves in the socio-economic systems. AIP Conference Proceedings vol. 2075, 150002 (2019). https://doi.org/10.1063/1.5091325

16. Simon J. H.: The Economic Consequences of Immigration. The University of Michigan Press, Ann Arbor, MI, USA (1999)

17. Torokhti A., Howlett P.: Computational Methods for Modelling of Nonlinear Systems. Elsevier, Amsterdam (2007)

18. Jordanov, I., Nikolova, E.: On nonlinear waves in the spatio-temporal dynamics of interacting populations. Journal of Theoretical and Applied Mechanics 43, 69–76 (2013). https://doi.org/10.2478/jtam-2013-0015. arXiv:1208.5465

19. Jordanov, I. P.: On the nonlinear waves in (2+ 1)-dimensional population systems. Compt. rend. Acad. bulg. Sci 61, 307–314 (2008)

20. Dimitrova, Z.I.: On travelling waves in lattices: the case of Riccati lattices. Journal of Theoretical and Applied Mechanics 42, 3–22 (2012), https://doi.org/10.2478/v10254-012-0011-2). arXiv:1208.2414

21. Jordanov, I. P., Dimitrova, Z. I.: On Nonlinear Waves of Migration. Journal of Theoretical and Applied Mechanics 40, 89–96 (2010)

22. Nikolova, E. V.: On nonlinear waves in a blood-filled artery with an aneurysm. In AIP Conference Proceedings vol. 1978, 470050 (2018). https://doi.org/10.1063/1.5044120

23. Dimitrova, Z. I., Ausloos, M.: Primacy analysis in the system of Bulgarian cities. Open Physics 13, 218–225 (2015). https://doi.org/10.1515/phys-2015-0029

24. Kantz, H., Schreiber T.: Nonlinear Time Series Analysis. Cambridge University Press, Cambridge, UK (2004)

25. Struble R.: Nonlinear Differential Equations. Dover, New York (2018)

26. Vitanov, N. K.: Upper bounds on the heat transport in a porous layer. Physica D 136, 322–339 (2000). https://doi.org/10.1016/S0167-2789(99)00165-7

27. Vitanov, N. K., Ausloos, M. R.: Knowledge epidemics and population dynamics models for describing idea diffusion. In: Scharnhorst A., Boerner K., van den Besselaar P. (eds.) Models of science dynamics. Understanding Complex Systems. pp. 69–125. Springer, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-23068-4_3

28. Dimitrova, Z. I., Vitanov, N. K.: Adaptation and its impact on the dynamics of a system of three competing populations. Physica A: Statistical Mechanics and its Applications 300, 91–115 (2001). https://doi.org/10.1016/S0378-4371(01)00330-2

29. Dimitrova, Z., Gogova, D.,: Investigation of Differences in Optical Phonons Modes by Principal Component Analysis. Compt. rend. Acad. bulg. Sci 63, 1415–1420 (2010)

30. Mills T.: Applied Time Series Analysis. Academic Press, London (2019)

31. Dimitrova, Z. I., Vitanov, N. K.: Chaotic pairwise competition. Theoretical Population Biology 66, 1–12 (2004). https://doi.org/10.1016/j.tpb.2003.10.008

32. Dimitrova, Z. I.: On the Low-Dimensional Dynamics of Blood Flow in Small Peripheral Human Arteries. Compt. rend. Acad. bulg. Sci 63, 55–60 (2010)

33. Borisov, R., Dimitrova, Z. I., Vitanov, N. K.: Statistical characteristics of stationary flow of substance in a network channel containing arbitrary number of arms. Entropy 22, 553 (2020). https://doi.org/10.3390/e22050553

34. Vitanov, N. K., Vitanov, K. N.: Statistical distributions connected to motion of substance in a channel of a network. Physica A 527, 121174 (2019). https://doi.org/10.1016/j.physa.2019.121174

35. Dimitrova, Z. I., Vitanov, K. N.: Homogeneous balance method and auxiliary equation method as particular cases of simple equations method (SEsM). In AIP Conference Proceedings vol. 2321, 030004 (2021). https://doi.org/10.1063/5.0043070

36. Vitanov, N. K., Vitanov, K. N., Kantz, H.: On the motion of substance in a channel of a network: Extended model and new classes of probability distributions. Entropy 22, 1240 (2020). https://doi.org/10.3390/e22111240

37. Ablowitz, M. J., Kaup, D. J., Newell, A. C., Segur, H.: The inverse scattering transform-Fourier analysis for nonlinear problems. Studies in Applied Mathematics 53, 249–315 (1974). https://doi.org/10.1002/sapm1974534249

38. Gardner, C. S., Greene, J. M., Kruskal, M. D., Miura, R. M.: Method for solving the Korteweg-deVries equation. Physical review letters 19, 1095–1097 (1967). https://doi.org/10.1103/PhysRevLett.19.1095

39. Hirota, R.: The Direct Method in Soliton Theory. Cambridge University Press, Cambridge, UK, (2004)

40. Kudryashov, N. A.: Simplest equation method to look for exact solutions of nonlinear differential equations. Chaos, Solitons & Fractals 24, 1217–1231 (2005). https://doi.org/10.1016/j.chaos.2004.09.109

41. Kudryashov, N. A., Loguinova, N. B.: Extended simplest equation method for nonlinear differential equations. Applied Mathematics and Computation 205, 396–402 (2008). https://doi.org/10.1016/j.amc.2008.08.019

42. Kudryashov, N. A.: Exact solitary waves of the Fisher equation. Physics Letters A 342, 99–106 (2005). https://doi.org/10.1016/j.physleta.2005.05.025

43. Kudryashov, N. A.: Exact solutions and integrability of the Duffing-Van der Pol equation. Regular and Chaotic dynamics 23, 471–479 (2018). https://doi.org/10.1134/S156035471804007X

44. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Simple Equations Method (SEsM): Algorithm, connection with Hirota method, Inverse Scattering Transform Method, and several other methods. Entropy 23, 10 (2021). https://doi.org/10.3390/e23010010

45. Vitanov, N. K.: Recent developments of the methodology of the modified method of simplest equation with application. Pliska Studia Mathematica Bulgarica 30, 29–42 (2019)

46. Vitanov, N.K.: Modified method of simplest equation for obtaining exact solutions of nonlinear partia differential equations: history, recent developments of the methodology and studied of classes of equations. Journal of Theoretical and Applied Mechanics 49, 107–122 (2019)

47. Vitanov, N. K.: The simple equations method (SEsM) for obtaining exact solutions of nonlinear PDEs: Opportunities connected to the exponential functions. AIP Conference Proceedings vol. 2159, 030038 (2019). https://doi.org/10.1063/1.5127503

48. Vitanov, N. K., Dimitrova, Z. I.: Simple equations method (SEsM) and other direct methods for obtaining exact solutions of nonlinear PDEs. AIP Conference Proceedings vol. 2159, 030039 (2019). https://doi.org/10.1063/1.5127504

49. Dimitrova, Z. I., Vitanov, N. K.: Travelling waves connected to blood flow and motion of arterial walls. Gadomski, A. (ed.) In: Water in Biomechanical and Related Systems pp. 243–263. Springer, Cham. (2021)

50. Martinov, N., Vitanov, N.: On the correspondence between the self-consistent 2D Poisson-Boltzmann structures and the sine-Gordon waves. Journal of Physics A: Mathematical and General 25, L51–L56 (1992). https://doi.org/10.1088/0305-4470/25/2/004

51. Martinov, N., Vitanov, N.: On some solutions of the two-dimensional sine-Gordon equation. Journal of Physics A: Mathematical and General 25, L419–L426 (1992). https://doi.org/10.1088/0305-4470/25/8/007

52. Martinov, N. K., Vitanov, N. K.: New class of running-wave solutions of the (2+ 1)-dimensional sine-Gordon equation. Journal of Physics A: Mathematical and General 27, 4611–4618 (1994). https://doi.org/10.1088/0305-4470/27/13/034

53. Martinov, N. K., Vitanov, N. K.: On self-consistent thermal equilibrium structures in two-dimensional negative-temperature systems. Canadian Journal of Physics 72, 618–624 (1994). https://doi.org/10.1139/p94-079

54. Vitanov, N. K.: Breather and soliton wave families for the sine-Gordon equation. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454, 2409–2423 (1998). https://doi.org/10.1098/rspa.1998.0264

55. Vitanov, N. K., Martinov, N. K.: On the solitary waves in the sine-Gordon model of the two-dimensional Josephson junction. Zeitschrift fur Physik B Condensed Matter 100, 129–135 (1996). https://doi.org/10.1007/s002570050102

56. Vitanov, N. K.: On travelling waves and double-periodic structures in two-dimensional sine-Gordon systems. Journal of Physics A: Mathematical and General 29, 5195–5207 (1996). https://doi.org/10.1088/0305-4470/29/16/036

57. Vitanov, N. K., Jordanov, I. P., Dimitrova, Z. I.: On nonlinear dynamics of interacting populations: Coupled kink waves in a system of two populations. Communications in Nonlinear Science and Numerical Simulation 14, 2379–2388 (2009). https://doi.org/10.1016/j.cnsns.2008.07.015

58. Vitanov, N. K., Jordanov, I. P., Dimitrova, Z. I.: On nonlinear population waves. Applied Mathematics and Computation 215, 2950–2964 (2009). https://doi.org/10.1016/j.amc.2009.09.041

59. Vitanov, N. K.: Application of simplest equations of Bernoulli and Riccati kind for obtaining exact traveling-wave solutions for a class of PDEs with polynomial nonlinearity. Communications in Nonlinear Science and Numerical Simulation 15, 2050–2060 (2010). https://doi.org/10.1016/j.cnsns.2009.08.011

60. Vitanov, N. K., Dimitrova, Z. I.: Application of the method of simplest equation for obtaining exact traveling-wave solutions for two classes of model PDEs from ecology and population dynamics. Communications in Nonlinear Science and Numerical Simulation 15, 2836–2845 (2010). https://doi.org/10.1016/j.cnsns.2009.11.029

61. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: Modified method of simplest equation and its application to nonlinear PDEs. Applied Mathematics and Computation 216, 2587–2595 (2010). https://doi.org/10.1016/j.amc.2010.03.102

62. Vitanov, N. K.: Modified method of simplest equation: powerful tool for obtaining exact and approximate traveling-wave solutions of nonlinear PDEs. Communications in Nonlinear Science and Numerical Simulation 16, 1176–1185 (2011). https://doi.org/10.1016/j.cnsns.2010.06.011

63. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: On the class of nonlinear PDEs that can be treated by the modified method of simplest equation. Application to generalized Degasperis-Processi equation and b-equation. Communications in Nonlinear Science and Numerical Simulation 16, 3033–3044 (2011). https://doi.org/10.1016/j.cnsns.2010.11.013

64. Vitanov, N. K.: On modified method of simplest equation for obtaining exact and approximate solutions of nonlinear PDEs: the role of the simplest equation. Communications in Nonlinear Science and Numerical Simulation 16, 4215–4231 (2011). https://doi.org/10.1016/j.cnsns.2011.03.035

65. Vitanov, N. K.: On modified method of simplest equation for obtaining exact solutions of nonlinear PDEs: case of elliptic simplest equation. Pliska Studia Mathematica Bulgarica 21, 257–266 (2012)

66. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: Application of the method of simplest equation for obtaining exact traveling-wave solutions for the extended Korteweg-de Vries equation and generalized Camassa-Holm equation. Applied Mathematics and Computation 219, 7480–7492 (2013). https://doi.org/10.1016/j.amc.2013.01.035

67. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Traveling waves and statistical distributions connected to systems of interacting populations. Computers & Mathematics with Applications 66, 1666–1684 (2013). https://doi.org/10.1016/j.camwa.2013.04.002

68. Vitanov, N. K., Vitanov, K. N.: Population dynamics in presence of state dependent fluctuations. Computers & Mathematics with Applications 68, 962–971 (2014). https://doi.org/10.1016/j.camwa.2014.03.006

69. Vitanov, N. K., Dimitrova, Z. I.: Solitary wave solutions for nonlinear partial differential equations that contain monomials of odd and even grades with respect to participating derivatives. Applied Mathematics and Computation 247, 213–217 (2014). https://doi.org/10.1016/j.amc.2014.08.101

70. Dimitrova, Z.I.: Relation between G'/G-expansion method and the modified method of simplest equation. Compt. rend. Acad. bulg. Scie 65, 1513–1520 (2012).

71. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Modified method of simplest equation for obtaining exact analytical solutions of nonlinear partial differential equations: further development of the methodology with applications. Applied Mathematics and Computation 269, 363–378 (2015). https://doi.org/10.1016/j.amc.2015.07.060

72. Vitanov, N. K., Dimitrova, Z. I., Ivanova, T. I.: On solitary wave solutions of a class of nonlinear partial differential equations based on the function $1/cosh^n$ (αx+ βt). Applied Mathematics and Computation 315, 372–380 (2017). https://doi.org/10.1016/j.amc.2017.07.064

73. Vitanov, N. K., Dimitrova, Z. I.: On the modified method of simplest equation and the nonlinear Schrödinger equation. Journal of Theoretical and Applied Mechanics 48, 59–68 (2018)

74. Nikolova, E. V., Jordanov, I. P., Dimitrova, Z. I., Vitanov, N. K.: Evolution of nonlinear waves in a blood-filled artery with an aneurysm. AIP Conference Proceedings vol. 1895, 070002 (2017). https://doi.org/10.1063/1.5007391

75. Vitanov, N.K., Dimitrova, Z.I.: Simple Equations Method and Non-Linear Differential Equations with Non-Polynomial Non-Linearity. Entropy 23, 1624 (2021). https://doi.org/10.3390/e23121624

76. Nikolova, E. V., Chilikova-Lubomirova, M., Vitanov, N. K.: Exact solutions of a fifth-order Korteweg-de Vries-type equation modeling nonlinear long waves in several natural phenomena. AIP Conference Proceedings vol. 2321, 030026 (2021). https://doi.org/10.1063/5.0040089

77. Dimitrova, Z. I.: Several examples of application of the simple equations method (SEsM) for obtaining exact solutions of nonlinear PDEs. AIP Conference Proceedings vol. 2459, 030005 (2022). https://doi.org/10.1063/5.0083572

78. Dimitrova, Z. I.: On several specific cases of the simple equations method (SEsM): Jacobi elliptic function expansion method, F-expansion method, modified simple equation method, trial function method, general projective Riccati equations method, and first integral method. AIP Conference Proceedings, vol. 2459, 030006 (2022). https://doi.org/10.1063/5.0083573

79. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: On the use of composite functions in the Simple Equations Method to obtain exact solutions of nonlinear differential equations. Computation 9, 104 (2021). https://doi.org/10.3390/computation9100104

Finite Time Blow Up of the Solutions to Nonlinear Wave Equations with Sign-Changing Nonlinearities

Nikolai Kutev, Milena Dimova, and Natalia Kolkovska

Abstract We study the initial boundary value problem for the nonlinear wave equation in $\Omega \subset \mathbb{R}^n$ with power-type nonlinearity and sign-changing coefficients. We investigate the non-existence of global solutions for different energy levels. When the energy is supercritical, we give a new sufficient condition on the initial data, which guarantees finite time blow up of the corresponding solution.

Keywords Nonlinear wave equation · Sign-changing nonlinearity · Blow up

1 Introduction

We consider the initial boundary value problem for the nonlinear wave equation with sign-changing nonlinearity

$$u_{tt} - \Delta u = f(x, u), \qquad t > 0, \ \ x \in \Omega, \tag{1}$$

$$u(0, x) = u_0(x), \ \ u_t(0, x) = u_1(x), \qquad x \in \Omega, \tag{2}$$

$$u(t, x) = 0, \qquad t \geq 0, \ \ x \in \partial\Omega. \tag{3}$$

Here Ω is a bounded open subset of $\mathbb{R}^n$ ($n \geq 1$) with smooth boundary $\partial\Omega$ and

$$u_0(x) \in \mathrm{H}_0^1(\Omega), \qquad u_1(x) \in \mathrm{L}^2(\Omega). \tag{4}$$

N. Kutev · M. Dimova (✉) · N. Kolkovska
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Street, Bl.8, 1113 Sofia, Bulgaria
e-mail: mdimova@unwe.bg

M. Dimova
University of National and World Economy, Students' Town 1700, Sofia, Bulgaria

The nonlinear term $f(x, u)$ has one of the following forms:

$$f(x, u) = a(x)|u|^{p-1}u + b(x)|u|^{q-1}u, \tag{5}$$

$$f(x, u) = a(x)|u|^p + b(x)|u|^{q-1}u, \tag{6}$$

$$1 < q < p \quad \text{and} \quad p < \infty \text{ for } n = 1, 2; \quad p < \frac{n+2}{n-2} \text{ for } n \geq 3. \tag{7}$$

We assume that functions $a(x)$ and $b(x)$ in (5) and (6) satisfy conditions

$$a(x) \in C(\overline{\Omega}), \quad b(x) \in C(\overline{\Omega}), \quad |a(x)| \leq A, \quad |b(x)| \leq A \ \forall \ x \in \overline{\Omega}, \tag{8}$$

$$\{x \in \Omega : a(x) > 0\} \neq \emptyset, \quad b(x) \leq 0 \ \forall \ x \in \overline{\Omega}. \tag{9}$$

Let us note that $a(x)$ may change its sign in Ω.

The global existence and finite time blow up of the solutions to the wave equation

$$u_{tt} - \Delta u = f(u), \quad t > 0, \ x \in \Omega \subset \mathbb{R}^n, \ n \geq 1 \tag{10}$$

with superlinear nonlinearity $f(u)$ have been intensively investigated in the last decades. The first results concern the finite time blow up of the solutions to (10), (2)–(4) with negative or zero initial energy, see [1] and the earlier papers [2, 3]. A powerful method of proving blow up of the solutions to abstract nonlinear equations is developed by Levine in [4]. The main idea of the concavity method of Levine is the reduction of the wave equation to an ordinary differential inequality with respect to the L^2 norm of the solution.

In the pioneering paper [5], the authors introduce the so-called potential well method. By means of this method, the global behavior of the solutions is completely studied for subcritical initial energy, $0 < E(0) < d$. Here $E(0)$ is the energy of the initial data, see (12), and d is the depth of the potential well, see (15). The authors prove that when $0 < E(0) < d$ and the Nehari functional is negative $I(u_0) < 0$ (see (14)), then the solution to problem (10), (2)–(4) with

$$f(u) = a|u|^{p-1}u, \quad p > 1 \tag{11}$$

blows up for a finite time. In the critical energy case, $E(0) = d$, it is proved that if $I(u_0) < 0$ and $(u_0, u_1) \geq 0$, then the solution blows up for a finite time, see, e.g., [6]. Later on, the results for problem (10), (2)–(4) with the single nonlinearity (11) are extended for more general combined power-type nonlinearities with constant coefficients, see [7, 8] and the references therein.

For supercritical initial energy, $E(0) > d$, the concavity method of Levine is improved in [9], where the finite time blow up of the solutions to (10) is proved for arbitrary positive energy under the conditions

$$E(0) < \frac{1}{2} \frac{\left(\int_{\Omega} u_0(x)u_1(x)\, dx\right)^2}{\int_{\Omega} u_0^2(x)\, dx}, \quad \int_{\Omega} u_0(x)u_1(x)\, dx > 0.$$

Further generalizations of the concavity method are suggested in [10, 11].

Let us also mention the results in [7, 12] for blow up of the solutions to problem (10), (2)–(4) with arbitrary positive energy. For nonlinearity (11) (see [12]) and combined power-type nonlinearities with constant coefficients (see [7]), the blow up of the solutions is proved under the conditions

$$E(0) < \frac{\mathcal{C}(p-1)}{2(p+1)} \int_{\Omega} u_0^2(x)\, dx, \quad \int_{\Omega} u_0(x)u_1(x)\, dx \geq 0,$$

where $\mathcal{C}$ is the constant of the Poincare inequality, see (26).

In the present paper, we focus our investigation on equation (1) with more general nonlinear term $f(x, u)$ given in (5)–(9) with a sign-changing coefficient $a(x)$. The finite time blow up of the solutions to (1)–(9) with non-positive and subcritical initial energy is completely solved. For supercritical initial energy, a new sufficient condition for finite time blow up is found. This sufficient condition holds without any requirements for the sign of the scalar product of the initial data. Thus, for the first time in the literature, we prove finite time blow up for initial data with supercritical energy and negative sign of their scalar product.

The paper is organized in the following way. In Sect. 2, some definitions and preliminary lemmas are given. In Sect. 3, finite time blow up of the solutions to (1)–(9) with subcritical energy is proved by the potential well method. A new sufficient condition for finite time blow up of the solutions with supercritical energy is derived in Sect. 4.

2 Preliminary

Further on, we will use the following short notations for the functions depending on t and x

$$\|u\|_p = \|u(t, \cdot)\|_{\mathrm{L}^p(\Omega)}, \ p > 0, \quad (u, v) = (u(t, \cdot), v(t, \cdot)) = \int_{\Omega} u(t, x)v(t, x)\, dx.$$

For convenience, we will write $\|u\|$ instead of $\|u\|_2$.

Let us recall the definition of blow up of the weak solution to problem (1)–(9).

Definition 1 The solution $u(t, x)$ of (1)–(9) defined in the maximal existence time interval $[0, T_m)$, $0 < T_m \leq \infty$ blows up at T_m if

$$\limsup_{t \to T_m, t < T_m} \|u\| = \infty.$$

The solution $u(t, x)$ to (1)–(9) satisfies the conservation law $E(0) = E(t)$ for every $t \in [0, T_m)$, where

$$E(t) := E(u(t, \cdot), u_t(t, \cdot)) = \frac{1}{2} \left(\|u_t\|^2 + \|\nabla u\|^2 \right) - \int_\Omega \int_0^{u(t,x)} f(x, z) \, dz \, dx.$$
$$(12)$$

Below, we recall some preliminary results useful for proving blow up via the modifications of the concavity method proposed in [8, 11].

Lemma 1 (Theorem 3.2 in [8]) *Suppose $\psi(t) \in C^2([0, T_m))$ is a nonnegative solution to the problem*

$$\psi''(t)\psi(t) - \gamma\psi'^2(t) = Q(t), \quad t \in [0, T_m), \quad 0 < T_m \leq \infty,$$
$$\gamma > 1, \quad Q(t) \in C([0, \infty)), \quad Q(t) \geq 0 \text{ for } t \in [0, \infty).$$
$$(13)$$

If $\psi(t)$ blows up at T_m, then $T_m < \infty$.

Lemma 2 (Theorem 2.3 in [11]) *Suppose $\psi(t) \in C^2([0, T_m))$ is a nonnegative solution of the problem*

$$\psi''(t)\psi(t) - \gamma\psi'^2(t) = \alpha\psi^2(t) - \beta\psi(t) + H(t), \quad t \in [0, T_m) \ 0 < T_m \leq \infty,$$
$$\gamma > 1, \ \alpha > 0, \ \beta > 0, \ H(t) \in C([0, \infty)), \ H(t) \geq 0 \text{ for } t \in [0, \infty).$$

If $\psi(t)$ blows up at T_m, then $T_m < \infty$.

To apply the potential well method, proposed in [5], we introduce the functionals

$$J(u) = J(u(t, \cdot)) = J(t) := \frac{1}{2}\|\nabla u\|^2 - \int_\Omega \int_0^{u(t,x)} f(x, z) \, dz \, dx,$$

the Nehari functional $I(u)$

$$I(u) = I(u(t, \cdot)) = I(t) := \|\nabla u\|^2 - \int_\Omega u f(x, u(t, x)) \, dx, \quad (14)$$

and the critical energy constant d (the depth of the potential well)

$$d = \inf_{u \in \mathcal{N}} J(u), \quad (15)$$

where $\mathcal{N}$ is the Nehari manifold, defined by

$$\mathcal{N} = \{u \in H_0^1(\Omega) : I(u) = 0, \ \|\nabla u\| \neq 0\}.$$

3 Subcritical Initial Energy, $E(0) < d$

In case of subcritical initial energy, $E(0) < d$, we employ the potential well method in proving blow up of the weak solutions to (1)–(9). We start with statements about the structure of the Nehary manifold, the estimate of the depth of the potential well d as well as the relations between functionals I, J and constant d.

Lemma 3 *If $z \in H_0^1(\Omega)$ and $\|\nabla z\| \neq 0$, then there exists a constant $\lambda^* \neq 0$ such that $\lambda^* z$ lies in $\mathcal{N}$ iff*

$$\int_\Omega a(x) \, |z(x)|^{p+1} \, dx > 0 \qquad \text{when (5) holds} \qquad (16)$$

and

$$\int_\Omega a(x) \, |z(x)|^p z(x) \, dx \neq 0 \qquad \text{when (6) holds.} \qquad (17)$$

Proof **Case (5) Proof of Sufficiency.** From (14), we get

$$I(\lambda z) = \lambda^2 \left(\|\nabla z\|^2 - |\lambda|^{p-1} \int_\Omega a(x)|z(x)|^{p+1} \, dx - |\lambda|^{q-1} \int_\Omega b(x)|z(x)|^{q+1} dx \right)$$
$$= \lambda^2 h(\lambda),$$

$$(18)$$

where $h(\lambda) = \|\nabla z\|^2 - |\lambda|^{p-1} \int_\Omega a(x)|z(x)|^{p+1} \, dx - |\lambda|^{q-1} \int_\Omega b(x)|z(x)|^{q+1} \, dx$.

Since $h(\lambda)$ is an even function we investigate the behavior of $h(\lambda)$ for $\lambda > 0$. Obviously

$$\lim_{\lambda \to 0^+} h(\lambda) > 0, \qquad \lim_{\lambda \to +\infty} h(\lambda) = -\infty.$$

If $\int_\Omega b(x)|z(x)|^{q+1} \, dx \neq 0$, then function $h(\lambda)$ has a positive maximum at the point λ_0, where

$$\lambda_0 = \left(-\frac{(q-1) \int_\Omega b(x)|z(x)|^{q+1} \, dx}{(p-1) \int_\Omega a(x)|z(x)|^{p+1} \, dx} \right)^{\frac{1}{p-q}} > 0.$$

Thus, $h(\lambda)$ is strictly increasing for $\lambda \in (0, \lambda_0)$ and strictly decreasing for $\lambda \in (\lambda_0, +\infty)$. If $\int_\Omega b(x)|z(x)|^{q+1} \, dx = 0$, then $h(\lambda)$ is strictly decreasing for $\lambda \in (0, +\infty)$. In both cases, there exists a constant $\lambda^* > 0$ such that $h(\lambda^*) = 0$. Since $h(\lambda)$ is even then $h(-\lambda^*) = 0$ and hence $I(-\lambda^* z) = I(\lambda^* z) = 0$, i.e. $-\lambda^* z$ and $\lambda^* z$ belong to $\mathcal{N}$.

  Case (5) Proof of Necessity. Let $\lambda^* z \in \mathcal{N}$ for some constant $\lambda^* \neq 0$. We suppose by contradiction that

$$\int_\Omega a(x) \, |z(x)|^{p+1} \, dx \leq 0. \qquad (19)$$

From (18) for $\lambda = \lambda^*$ and (19), we get the following impossible chain of inequalities $0 = I(\lambda^* z) \geq (\lambda^*)^2 \|\nabla z\|^2 > 0$. Hence, condition (16) is fulfilled.

Case (6) Proof of Sufficiency. Suppose (17) holds. Then

$$I(\lambda z) = \lambda^2 \left(\|\nabla z\|^2 - |\lambda|^{p-2} \lambda \int_\Omega a(x)|z(x)|^p z(x) dx - |\lambda|^{q-1} \int_\Omega b(x)|z(x)|^{q+1} dx \right)$$
$$= \lambda^2 g(\lambda),$$

where $g(\lambda) = \|\nabla z\|^2 - |\lambda|^{p-2} \lambda \int_\Omega a(x)|z(x)|^p z(x)\, dx - |\lambda|^{q-1} \int_\Omega b(x)|z(x)|^{q+1}\, dx$.

If $\int_\Omega a(x)|z(x)|^p z(x)\, dx > 0$, then $g(\lambda) > 0$ for every $\lambda < 0$. For $\lambda \in (0, +\infty)$, similarly to Case (5), we conclude that there exist a unique constant $\lambda^* > 0$, such that $g(\lambda^*) = 0$ and hence $\lambda^* z \in \mathcal{N}$.

Let $\int_\Omega a(x)|z(x)|^p z(x)\, dx < 0$. Since $g(\lambda) > 0$ for every $\lambda > 0$, we study the case $\lambda \in (-\infty, 0)$. For $\int_\Omega b(x)|z(x)|^{q+1}\, dx \neq 0$ we obtain that $g(\lambda)$ has a unique local positive maximum at the point

$$\lambda_0 = - \left(\frac{(q-1) \int_\Omega b(x)|z(x)|^{q+1}\, dx}{(p-1) \int_\Omega a(x)|z(x)|^p z(x)\, dx} \right)^{\frac{1}{p-q}} < 0$$

and $\lim_{\lambda \to 0^-} g(\lambda) > 0$, $\lim_{\lambda \to -\infty} g(\lambda) = -\infty$. Hence, there exists a unique constant $\lambda^* < \lambda_0$, such that $g(\lambda^*) = 0$. For $\int_\Omega b(x)|z(x)|^{q+1}\, dx = 0$, function $g(\lambda)$ is strictly decreasing, so there exists a uinique λ^* such that $g(\lambda^*) = 0$. Thus, for every $b(x)$ satisfying (8) and (9), there exists a unique constant λ^* for which $\lambda^* z \in \mathcal{N}$.

The proof of **Case (6) Necessity** is analogous to the proof of **Case (5) Necessity**, so we omit it. Lemma 3 is proved.

Corollary 1 *Suppose $f(x, z)$ satisfies (5)–(9). Then the Nehari manifold $\mathcal{N}$ is not empty.*

Proof **Case (5)** From (8) and (9), it follows that there exists $\Omega_1 \subset \Omega$ such that $a(x) > 0$ for $x \in \Omega_1$. If we choose $w \in H_0^1(\Omega_1)$, $w > 0$ in Ω_1 and $w \equiv 0$ in $\Omega \setminus \Omega_1$, then (16) is satisfied. Lemma 3 gives us that $\lambda^* w \in \mathcal{N}$ for some constant $\lambda^* \neq 0$.

Case (6) Similarly, from (8) and continuity of $a(x)$, it follows that there exists $\Omega_2 \subset \Omega$ such that $a(x) \neq 0$ for $x \in \Omega_2$. Hence, for $w \in H_0^1(\Omega_2)$, $w > 0$ in Ω_2 and $w \equiv 0$ in $\Omega \setminus \Omega_2$, condition (17) in Lemma 3 holds and $\lambda^* w \in \mathcal{N}$ for some constant $\lambda^* \neq 0$.

Lemma 4 *Suppose $f(x, z)$ satisfies (5)–(9), $z \in H_0^1(\Omega)$, $\|\nabla z\| \neq 0$ and let*

$$r = \left(A C_{p+1}^{p+1} \right)^{-\frac{1}{p-1}} > 0, \quad \text{where } C_{p+1} = \sup_{u \in H_0^1(\Omega) \setminus \{0\}} \frac{\|u\|_{p+1}}{\|\nabla u\|}. \tag{20}$$

(i) If $I(z) < 0$, then $\|\nabla z\| > r$;

(ii) If $I(z) = 0$, then $\|\nabla z\| \geq r$.

Proof (i) Since $I(z) < 0$, we obtain the following chain of inequalities:

$$\|\nabla z\|^2 < \int_\Omega zf(x,z)\,dx \leq \int_\Omega |a(x)||z(x)|^{p+1}\,dx + \int_\Omega b(x)|z(x)|^{q+1}\,dx$$

$$\leq \int_\Omega |a(x)||z(x)|^{p+1}\,dx \leq A\|z\|_{p+1}^{p+1} \leq AC_{p+1}^{p+1}\|\nabla z\|^{p+1}.$$

Hence, $AC_{p+1}^{p+1}\|\nabla z\|^{p-1} > 1$, i.e. $\|\nabla z\| > r$. The proof of (ii) is analogous to (i) and we omit it. Lemma 4 is completed.

Lemma 5 *Suppose $f(x,z)$ satisfies (5)–(9). Then the following estimate holds*

$$d \geq \frac{p-1}{2(p+1)}r^2 > 0, \quad \text{where } r \text{ is defined in (20)}.$$

Proof Corollary 1 gives us that $\mathcal{N} \neq \emptyset$. Hence, from (15), Lemma 4 (ii), and the relation

$$J(z) = \frac{1}{p+1}I(z) + \frac{p-1}{2(p+1)}\|\nabla z\|^2 - \frac{p-q}{(p+1)(q+1)}\int_\Omega b(x)|z(x)|^{q+1}\,dx,$$
$$(21)$$

we get

$$d = \inf_{z \in \mathcal{N}} J(z) \geq \frac{p-1}{2(p+1)} \inf_{z \in \mathcal{N}} \|\nabla z\|^2 \geq \frac{p-1}{2(p+1)}r^2 > 0,$$

which proves Lemma 5.

Lemma 6 *Suppose $f(x,z)$ satisfies (5)–(9), $z \in H_0^1(\Omega)$, $\|\nabla z\| \neq 0$, and $I(z) < 0$. Then there exists a constant $|\lambda^*| < 1$, $\lambda^* \neq 0$ such that $I(\lambda^*z) = 0$.*

Proof **Case (5)** From $I(z) < 0$ and Lemma 4 (i), we get

$$0 < r^2 < \|\nabla z\|^2 < \int_\Omega a(x)|z(x)|^{p+1}\,dx + \int_\Omega b(x)|z(x)|^{q+1}\,dx$$

$$\leq \int_\Omega a(x)|z(x)|^{p+1}\,dx.$$

Hence, condition (16) is satisfied, and from Lemma 4, there exists λ^*, $|\lambda^*| \in (0, 1)$ such that $I(\lambda^*u) = 0$.

Case (6) The case $\int_\Omega a(x)\,|z(x)|^p z(x)\,dx > 0$ is treated in the same way as Case (5). Now we consider the case $\int_\Omega a(x)\,|z(x)|^p z(x)\,dx < 0$. Then

$$I(-z) = \|\nabla(-z)\|^2 - \int_{\Omega} a(x)|z(x)|^p(-z(x))\,dx - \int_{\Omega} b(x)|z(x)|^{q+1}\,dx$$

$$= I(z) + 2\int_{\Omega} a(x)|z(x)|^p z(x)\,dx < 0.$$

Therefore, the result in Case (5) is applicable to $-z$, i.e., there exists λ^*, $|\lambda^*| \in (0, 1)$ such that $I(\lambda^*(-z)) = I(-\lambda^* z) = 0$. The proof of Lemma 6 is completed.

Lemma 7 *Suppose $f(x, z)$ satisfies* (5)–(9), $z \in H_0^1(\Omega)$, $\|\nabla z\| \neq 0$ *and* $I(z) < 0$. *Then the following inequality holds:*

$$I(z) < (p+1)(J(z) - d). \tag{22}$$

Proof From Lemma 6, it follows that there exists $|\lambda^*| < 1$, $\lambda^* \neq 0$ such that $I(\lambda^* u) = 0$. From (15) and (21), we get the following chain of inequalities:

$$d \leq J(\lambda^* z) = \frac{1}{p+1} I(\lambda^* z) + \frac{p-1}{2(p+1)} (\lambda^*)^2 \|\nabla z\|^2$$

$$- \frac{p-q}{(p+1)(q+1)} |\lambda^*|^{q+1} \int_{\Omega} b(x)|z(x)|^{q+1}\,dx$$

$$< \frac{p-1}{2(p+1)} \|\nabla z\|^2 - \frac{p-q}{(p+1)(q+1)} \int_{\Omega} b(x)|z(x)|^{q+1}\,dx = J(z) - \frac{1}{p+1} I(z).$$

Thus (22) is proved.

In the framework of the potential well method, there are two important subsets of $H_0^1(\Omega)$:

$$W = \left\{ z \in H_0^1(\Omega) : I(z) > 0 \right\} \cup \{0\} \quad \text{and} \quad V = \left\{ z \in H_0^1(\Omega) : I(z) < 0 \right\}.$$

In the following theorem, we formulate the sign preserving properties of the Nehari functional $I(z)$, i.e., the invariance of W and V under the flow of (1)–(9) when $E(0) < d$.

Theorem 1 *Suppose $u(t, x)$ is the weak solution of* (1)–(9) *defined in the maximal existence time interval* $[0, T_m)$, $0 < T_m \leq \infty$ *and* $E(0) < d$.

(i)   *If $u_0 \in W$, then $u(t, x) \in W$ for every $t \in [0, T_m)$;*
(ii) *If $u_0 \in V$, then $u(t, x) \in V$ for every $t \in [0, T_m)$.*

The proof of Theorem 1 is identical to the proof of the invariance of W and V for nonlinearities (5) and (6) with $a(x) \equiv a > 0$ and $b(x) \equiv b < 0$, see Theorem 3.2 in [13]. That is why we omit it.

The following corollary will allow us to treat the case of non-positive energy within the potential well method.

Corollary 2 *Suppose $E(0) < 0$ or $E(0) = 0$ and $\|\nabla u\| \neq 0$. Then every weak solution of* (1)–(9) *belongs to V during its existence time.*

Proof From the conservation law, we get $E(0) = E(t) \leq 0$. Since

$$
\begin{aligned}
E(t) = & \frac{1}{p+1} I(u) + \frac{1}{2} \|u_t\|^2 + \frac{p-1}{2(p+1)} \|\nabla u\|^2 \\
& - \frac{p-q}{(p+1)(q+1)} \int_\Omega b(x)|u(t,x)|^{q+1}\, dx,
\end{aligned}
\tag{23}
$$

we get that $I(u) < 0$ and $u(t,x) \in V$ for every $t \in [0, T_m)$.

Theorem 2 *Suppose $u(t,x)$ is the weak solution to* (1)–(9) *defined in the maximal existence time interval* $[0, T_m)$, $0 < T_m \leq \infty$. *If $E(0) < d$ and $u_0 \in V$, then $u(t,x)$ blows up at $T_m < \infty$.*

Proof We suppose by contradiction that there exists a global weak solution $u(t,x)$ of (1)–(9). Applying Theorem 1 (ii) and Lemma 7, for function $\psi(t) = \|u(t,\cdot)\|^2$, we obtain

$$
\begin{aligned}
\psi''(t) &= (p+3)\|u_t\|^2 - 2(p+1)(E(0) - J(t)) - 2I(t) \\
&= (p+3)\|u_t\|^2 + 2[(p+1)(J(t) - E(0)) - I(t)] \geq 2(p+1)(d - E(0)) > 0.
\end{aligned}
\tag{24}
$$

Integrating (24) twice, we get $\psi(t) \geq (p+1)(d - E(0))t^2 + \psi'(0)t + \psi(0)$ and $\lim_{t \to \infty} \psi(t) = \infty$.

On the other hand, one can show that $\psi(t)$ satisfies conditions in Lemma 1. Indeed,

$$
\begin{aligned}
\psi''(t)\psi(t) - \frac{p+3}{4}\psi'^2(t) = & (p+3)\left(\|u\|^2\|u_t\|^2 - (u, u_t)^2\right) \\
& + 2[(p+1)(J(t) - E(0)) - I(t)]\|u\|^2.
\end{aligned}
$$

Hence, $\psi(t)$ is a solution of (13) for $\gamma = \frac{p+3}{4} > 1$ and for $t \in [0, \infty)$

$$
Q(t) = (p+3)\left(\|u\|^2\|u_t\|^2 - (u, u_t)^2\right) + 2[(p+1)(J(t) - E(0)) - I(t)]\|u\|^2 \geq 0.
$$

According to Lemma 1 function $\psi(t)$, or equivalently, $u(t,x)$, blows up for a finite time, which contradicts our assumption that $u(t,x)$ is globally defined. The proof of Theorem 2 is completed.

Theorem 3 *Suppose $E(0) < d$ and $u_0 \in W$. Then problem* (1)–(9) *has no blowing up weak solutions.*

Proof Suppose by contradiction that $u(t,x)$, defined in $[0, T_m)$, $0 < T_m \leq \infty$, blows up at T_m. Since $I(0) > 0$, from Theorem 1 (i) it follows that $I(t) > 0$ for $t \in (0, T_m)$. From (23) we have for every $t \in (0, T_m)$ the estimate $E(0) \geq \frac{p-1}{2(p+1)}\|\nabla u\|^2$ holds, which contradicts Definition 1. Thus Theorem 3 is proved. $\square$

4 Supercritical Initial Energy, $E(0) \geq d$

Theorem 4 *Suppose $u(t, x)$ is the weak solution to (1)–(9) defined in the maximal existence time interval $[0, T_m)$, $0 < T_m \leq \infty$. If*

$$E(0) < \frac{\mathcal{C}(p-1)}{2(p+1)}\|u_0\|^2 + \frac{\sqrt{\mathcal{C}(p-1)}}{p+1}(u_0, u_1), \tag{25}$$

then $u(t, x)$ blows up at $T_m < \infty$. Here $\mathcal{C}$ is the constant of the Poincaré inequality

$$\|\nabla u\|^2 \geq \mathcal{C}\|u\|^2. \tag{26}$$

Proof Analogously to the proof of Theorem 2, we assume that $u(t, x)$ is globally defined. Simple computations give us that $\psi(t) = \|u(t, \cdot)\|^2$ is a solution of the equation

$$\psi''(t) = (p-1)\mathcal{C}\psi(t) - 2(p+1)E(0) + G(t), \tag{27}$$

where $\mathcal{C}$ is defined in (26), and

$$G(t) = (p-1)(\|\nabla u\|^2 - \mathcal{C}\|u\|^2) + (p+3)\|u_t\|^2 - \frac{2(p-q)}{q+1}\int_\Omega b(x)|u(t, x)|^{q+1}\,dx \geq 0. \tag{28}$$

If we set $\alpha = \mathcal{C}(p-1) > 0$ and $\beta = 2(p+1)E(0) > 0$, then the classical solution of (27) is

$$\begin{aligned}
\psi(t) = &\frac{1}{2}\left(\psi(0) + \frac{1}{\sqrt{\alpha}}\psi'(0) - \frac{\beta}{\alpha}\right)e^{\sqrt{\alpha}t} \\
&+ \frac{1}{2}\left(\psi(0) - \frac{1}{\sqrt{\alpha}}\psi'(0) - \frac{\beta}{\alpha}\right)e^{-\sqrt{\alpha}t} \\
&+ \frac{\beta}{\alpha} + \frac{1}{\sqrt{\alpha}}\int_0^t G(s)\sinh(\sqrt{\alpha}(t-s))\,ds.
\end{aligned} \tag{29}$$

Since (25) is equivalent to $\psi(0) + \frac{1}{\sqrt{\alpha}}\psi'(0) - \frac{\beta}{\alpha} > 0$, from (28) and (29) we conclude that $\lim_{t\to\infty}\psi(t) = \infty$.

Obviously, $\psi(t)$ satisfies also the equation

$$\psi''(t)\psi(t) - \frac{p+3}{4}\psi'^2(t) = \alpha\psi^2(t) - \beta\psi(t) + H(t), \quad \text{where}$$

$$\begin{aligned}
H(t) = &(p+3)(\|u_t\|^2\|u\|^2 - (u, u_t)^2) + (p-1)(\|\nabla u\|^2 - \mathcal{C}\|u\|^2)\|u\|^2 \\
&- \frac{2(p-q)}{q+1}\|u\|^2\int_\Omega b(x)|u(t, x)|^{q+1}\,dx \geq 0.
\end{aligned}$$

From Lemma 2, it follows that $u(t, x)$ blows up for a finite time, which contradicts our assumption $T_m = \infty$. Theorem 4 is proved.

References

1. Ball, J.M.: Remarks on blow-up and nonexistence theorems for nonlinear evolution equations. Quart. J. Math. 28, 473–486 (1977). https://doi.org/10.1093/qmath/28.4.473
2. Tsutsumi M.: On solutions of semilinear differential equations in a Hubert space. Math. Japon. 17, 173–193 (1972).
3. Glassey R.T.: Blow-up theorems for nonlinear wave equations. Math. Z. 132, 183–203 (1973). https://doi.org/10.1007/BF01213863
4. Levine H.A.:, Instability and nonexistence of global solutions to nonlinear wave equations of the form $Pu_{tt} = -Au + F(u)$. Trans. Amer. Math. Soc. 192, 1–21 (1974). https://doi.org/10.2307/1996814
5. Payne L.E., Sattinger D.H.: Saddle points and instability of nonlinear hyperbolic equations. Israel. J. Math. 22, 273–303 (1975). https://doi.org/10.1007/BF02761595
6. Esquivel-Avila J.: Blow up and asymptotic behavior in a nondissipative nonlinear wave equation. Appl. Anal. 93, 1963–1978 (2014). https://doi.org/10.1080/00036811.2013.859250
7. Xu R., Chen Y., Yang Y., Chen S., Shen J., Yu T., Xu Z.: Global well-posedness of semilinear hyperbolic equations, parabolic equations and Schrodinger equations. Electron. J. Differ. Equ. 2018, 1–52 (2018). http://ejde.math.unt.edu
8. Dimova M., Kolkovska N., Kutev N.: Global behavior of the solutions to nonlinear Klein–Gordon equation with critical initial energy. Electron. Res. Arch. 28(2), 671–689 (2020). https://doi.org/10.3934/era.2020035
9. Straughan B.: Further global nonexistence theorems for abstract nonlinear wave equations. Proc. Amer. Math. Soc. 48, 381–390 (1975). https://doi.org/10.2307/2040270
10. Kalantarov V.K., Ladyzhenskaya O.A.: The occurrence of collapse for quasilinear equations of parabolic and hyperbolic types. J. Soviet Math. 10, 53–70 (1978). https://doi.org/10.1007/BF01109723
11. Dimova M., Kolkovska N., Kutev N.: Blow up of solutions to ordinary differential equations arising in nonlinear dispersive problems. Electron. J. Differ. Equ. 2018(68), 1–16 (2018). http://ejde.math.unt.edu
12. F. Gazzola F., Squassina M.: Global solutions and finite time blow up for damped semilinear wave equations. Ann. I. H. Poincaré—AN 23, 185–207 (2006). https://doi.org/10.1016/j.anihpc.2005.02.007
13. Kutev N., Kolkovska N., Dimova M.: Global existence to generalized Boussinesq equation with combined power–type nonlinearities. J. Math. Anal. Appl. 410, 427–444 (2014). https://doi.org/10.1016/j.jmaa.2013.08.036

An Example for Application of the Simple Equations Method for the Case of Use of Two Simple Equations

Zlatinka I. Dimitrova

Abstract We discuss the methodology of the Simple Equations Method. Discussion is focused not only on the methodology but also on an example where the methodology is applied on the basis of a two simple equations for solution of a differential equation of a nonlinear Schrödinger kind. We explain the steps of the methodology and discuss the obtained exact solutions of the solved nonlinear partial differential equation.

1   Introduction

Complexity is closely related to the nonlinearity [1–16]. Methods of the time series analysis and modeling based on differential or difference equations are appropriate to study these effects [17–36]. Methods for obtaining exact solutions of nonlinear differential equations such as the Hopf–Cole transformation [37, 38] are of interest for us. Other interesting methods for obtaining exact solutions of classes of nonlinear differential equations are the Method of Inverse Scattering Transform [39], and the method of Hirota [40].

Below we discuss a methodology called SEsM (simple equations method). It is designed for obtaining exact solutions of nonlinear differential equations. SEsM is the successor of the methodology called MMSE (modified method of simplest equation). MMSE was designed on the basis of research of Kudryashov and Loguinova who

Z. I. Dimitrova (✉)
Institute of Mechanics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 4, 1113 Sofia, Bulgaria
e-mail: zdim@imbm.bas.bg

presented the Method of Simplest Equation (MSE) [41, 42]. The development of
SEsM [43–46] started many years ago [47–55]. This research continued in 2009
[56, 57] and in 2010 by the use of the ordinary differential equation of Bernoulli as
simplest equation [58] and application of MMSE to ecology and population dynamics
[59]. Characteristic feature of MMSE is the use of a balance equation [60, 61]
and till 2018 MMSE [62–65, 67, 68] was based on one simplest equation and one
balance equation. An important addition to the methodology was made in [69]. There,
MMSE was extended to simplest equations of the class $\left(\frac{d^k f}{d\xi^j}\right)^k = \sum_{i=0}^{m} c_i f^i$, where
$j = 1, \ldots, k = 1, \ldots,$ and m and c_i are parameters. In the last years, the methodology
was extended by the possibility of use of more than one simplest equation. The
modification is called SEsM—Simple Equations Method [43–46, 70]. For more
applications of specific cases of the methodology see [71, 84].

The goal of this article is to discuss SEsM and to show one example of its appli-
cation. SEsM is discussed in Sect 2. In Sect. 3 we present an application of SEsM for
the case of use of composite functions of two solutions of simple equations. Several
concluding remarks are summarized in Sect. 4.

2 Methodology of the Simple Equations Method (SEsM)

SEsM is designed for search of solution of systems of nonlinear differential equations
but below we shall use it in order to obtain a solution of a single differential equation
(Fig. 1).

We search for an exact solution of the nonlinear differential equation
$\mathcal{D}[u(x, \ldots, t), \ldots,] = 0$. $\mathcal{D}[u(x, \ldots, t), \ldots]$ depends on the function $u(x, \ldots, t)$
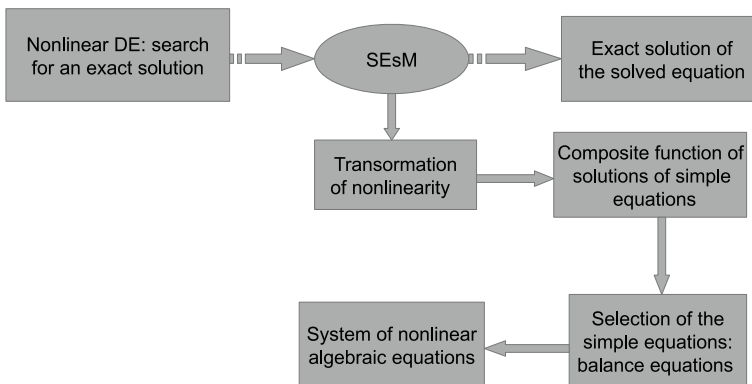
Fig. 1 The example of SEsM with use of composite functions. Composite functions can be used in:
the transformation of the nonlinearity; the construction of the solution as function of the solutions
of the simple equations; in the process of obtaining solutions of the simple equations

and some of their derivatives (u_i can be a function of more than 1 spatial coordinates). Step 1 of SEsM is connected to the transformation $u(x, \ldots, t) = W[G(x, \ldots, t),$ $H(x, \ldots, t), \ldots]$.. $W(F, G, \ldots)$ is a composite function of another functions F, G, …. $G(x, \ldots, t)$, $H(x, \ldots, t)$, … are functions of several spatial variables as well as of the time. The goal of the transformation is to transform the nonlinearity of the solved differential equation to more treatable kind of nonlinearity. The application of the transformation leads to nonlinear PDEs for the functions G, H, ….

In Step 2 of SEsM the functions $G(x, \ldots, t)$, $H(x, \ldots, t)$, … are represented as a composite function of other functions $g_1, \ldots, g_N, h_1, \ldots, h_M, \ldots$. The last functions are connected to solutions of some differential equation which are more simple than Eq. (2).

The solved differential equation contains derivatives. Then, we use the Faa di Bruno's formula in order to write the general form of the derivatives of the composite functions F and G [85].

Step 3 of SEsM is connected to determination of the functions $g_1, \ldots, g_N$, $h_1, \ldots, h_M, \ldots$. The equations for these functions are more simple than the solved nonlinear partial differential equation.

At Step 4 of SEsM, the steps 1–3 are applied to Eq. (2). The result is a system of nonlinear algebraic equations. Any nontrivial solution of the algebraic system corresponds to a solution of the studied nonlinear partial differential equation.

3 An Example of Use of Composite Functions in the Methodology of SEsM for the Case of Two Simple Equations

The example will be connected to a class of equations of the kind of the nonlinear Schrödinger equation, namely,

$$i\frac{\partial \psi}{\partial t} + a\frac{\partial^2 \psi}{\partial x^2} + \psi \sum_{l=-B; l \neq -1}^{C} b_l \mid \psi \mid^{2l} = 0. \tag{1}$$

Above, B and C are nonnegative integers. We skip Step 1 of SEsM where the transformation is used and proceed to the Step 2 where we have to construct the solution ψ as composite function of solutions of two simple equations $\psi(x, t) = \psi(p(x, t), q(x, t))$. (1) can have many solutions. Below we will search for specific solutions which are represented by a product of the solutions of the two simple equations $\psi(h, g) = p(x, t)q(x, t)$. Then (1) becomes

$$i(pq_t + qp_t) + a(pq_{xx} + qp_{xx} + 2p_xq_x) + pq \sum_{l=-B; l \neq -1}^{C} b_l \mid pq \mid^{2l} = 0. \tag{2}$$

We focus our interest on the class of solutions which possess a traveling-wave component. $\psi(x, t) = q(x, t)p(x, t)$. $p(\xi)$ is a real function ($\xi = \alpha x + \beta t$) and $q(x, t)$ is a complex function. The substitution of this relationship for ψ in the solved equation leads to following relationship for the first simple equation: $\frac{dq}{d\zeta} = iq$, $\zeta = \kappa x + \omega t + \pi$. The solution of this equation is $q(\zeta) = \exp(i\zeta) = \exp[i(\kappa x + \omega t + \pi)]$.

For the function $p(\xi)$, we obtain the equation

$$\alpha^2 a p_{\xi\xi} + (2\alpha\kappa a + \beta)i p_\xi - (\omega + \kappa^2 a)p + p \sum_{l=-B;l\neq-1}^{C} b_l p^{2l} = 0. \qquad (3)$$

(3) is solved by means of the second simple equation. $g(\xi)$ has to be a real function and, because of this, we have $\beta = -2\alpha\kappa a$ We substitute this in (3) and multiply the result by p'. The obtained relationship is integrated with respect to ξ (here a constant of integration c occurs). Then, it is assumed that $v = p^\sigma$, where σ is a parameter. The form of the equation for v becomes

$$v'^2 = \frac{\sigma^2(\omega + \kappa^2 a)}{\alpha^2 a} v^2 + \frac{\sigma^2 c}{\alpha^2 a} v^{2(\sigma-1)/\sigma} - \sum_{l=-B;l\neq-1}^{C} \frac{\sigma^2 b_l}{\alpha^2 a(l+1)} v^{2(l+\sigma)/\sigma}. \qquad (4)$$

Many specific cases of (4) are possible. Here we discuss the case $c = 0$, σ: arbitrary positive integer that is different from 1, $C = D\sigma$, $l = n\sigma$, $n = -1, \ldots, D$. In this case, (4) is reduced to the equation

$$v'^2 = \sum_{n=-1}^{D} c_n v^{2(n+1)}, \qquad (5)$$

where $c_{-1} = -\frac{\sigma^2 b_{-\sigma}}{\alpha^2 a(1-\sigma)}$, $c_0 = \frac{\sigma^2(\omega+\kappa^2 a-b_0)}{\alpha^2 a}$, $c_n = -\frac{\sigma^2 b_{n\sigma}}{\alpha^2 a(\sigma n+1)}, n = 1, \ldots, D$.

(5) defines function which is particular case of the special function $V_{\mathbf{a}}(\xi; q, l, m)$ discussed in [69]. This function is defined as the solution of the nonlinear ODE $\left(\frac{d^q v}{d\xi^q}\right)^l = \sum_{j=0}^{m} a_j u^j$,, where q, l, and m are positive integers and $\mathbf{a} = (a_0, \ldots, a_m)$. For the considered specific case (5) the solution of (1) is

$$q(x, t) = [V_{\mathbf{a}_3}(\xi; 1, 2, 2D + 2)]^{1/\sigma} \exp[i(kx + \omega t + \theta)] \qquad (6)$$

where $\mathbf{a}_3 = (a_0, \ldots, a_{2D+2})$ and $a_{2i} = c_{i-1}, a_{2i+1} = 0, i = 0, D + 1$.

4 Concluding Remarks

In this article, we show an application of the Simple Equations Method (SEsM) for obtaining exact solutions of nonlinear differential equations. The methodology is based on use of composite functions. The discussed example is connected to a class of equations of Schrödinger kind. The example illustrates the use of two simple equations within the methodology of SEsM. The first simple equations describe the envelope of the solution and the oscillations within the envelope are described by the second simple equation. The solution of the last equation can be expressed by a special function which is frequently encountered in the process of application of the SEsM.

References

1. Axelrod R., Cohen M.: Harnessing Complexity. Basic Books, New York (2001)
2. Vitanov, N., Busse, F.: Bounds on the heat transport in a horizontal fluid layer with stress-free boundaries. Z. angew. Math. Phys. 48, 310–324 (1997). https://doi.org/10.1007/PL00001478
3. Kutner, R., Ausloos, M., Grech, D., Di Matteo, T., Schinckus, C., Stanley, H. E.: Econophysics and sociophysics: Their milestones & challenges. Physica A 516, 240–253 (2019). https://doi.org/10.1016/j.physa.2018.10.019
4. Vitanov, N. K., Sakai, K., Jordanov, I. P., Managi, S., Demura, K.: Analysis of a Japan government intervention on the domestic agriculture market. Physica A: 382, 330–335 (2007). https://doi.org/10.1016/j.physa.2007.02.025
5. May, R. M., Levin, S. A., Sugihara, G.: Ecology for bankers. Nature 451, 893–894, (2008). https://doi.org/10.1038/451893a
6. Nikolova, E. V., Vitanov, N. K.: On the Possibility of chaos in a ceneralized model of three interacting sectors. Entropy 22, 1388 (2020). https://doi.org/10.3390/e22121388
7. Vitanov N.K.: Science Dynamics and Research Production. Indicators, Indexes, Statistical Laws and Mathematical Models. Springer, Cham (2016)
8. Bahrami, M., Chinichian, N., Hosseiny, A., Jafari, G., Ausloos, M.: Optimization of the post-crisis recovery plans in scale-free networks. Physica A 540, 123203 (2020). https://doi.org/10.1016/j.physa.2019.123203
9. Vitanov, N. K.: Upper bound on the heat transport in a horizontal fluid layer of infinite Prandtl number. Physics Letters A 248, 338–346 (1998). https://doi.org/10.1016/S0375-9601(98)00674-4
10. Lempert, R. J.: A new decision sciences for complex systems. Proceedings of the National Academy of Sciences 99 (suppl 3), 7309–7313 (2002). https://doi.org/10.1073/pnas.082081699
11. Vitanov, N. K., Ausloos, M., Rotundo, G.: Discrete model of ideological struggle accounting for migration. Advances in Complex Systems 15, 1250049 (2012). https://doi.org/10.1142/S021952591250049X
12. Vitanov, N. K., Vitanov, K. N.: Box model of migration channels. Mathematical Social Sciences 80, 108–114 (2016). https://doi.org/10.1016/j.mathsocsci.2016.02.001
13. Larsen-Freeman, D., Cameron, L.: Complex systems and applied linguistics. Oxford: Oxford University Press (2008)

14. Vitanov, N. K., Vitanov, K. N.: On the motion of substance in a channel of a network and human migration. Physica A: 490, 1277–1294 (2018). https://doi.org/10.1016/j.physa.2017.08.038
15. Amaral, L. A. N., Scala, A., Barthelemy, M., Stanley, H. E.: Classes of small-world networks. PNAS USA 97, 11149–11152 (2000). https://doi.org/10.1073/pnas.20032719
16. Vitanov, N. K., Vitanov, K. N., Kantz, H.: On the motion of substance in a channel of a network: Extended model and new classes of probability distributions. Entropy 22, 1240 (2020). https://doi.org/10.3390/e22111240
17. Kantz, H., T. Schreiber, T.: Nonlinear Time Series Analysis. Cambridge University Press, Cambridge, UK (2004)
18. Jordanov, I., Nikolova, E.: On nonlinear waves in the spatio-temporal dynamics of interacting populations. Journal of Theoretical and Applied Mechanics 43, 69–76 (2013).
19. Boeck, T., Vitanov, N. K.: Low-dimensional chaos in zero-Prandtl-number Benard-Marangoni convection. Physical Review E 65, 037203 (2002). https://doi.org/10.1103/PhysRevE.65.037203
20. Jordanov, I. P.: On the nonlinear waves in (2+ 1)-dimensional population systems. Comptes rendus de l'Académie bulgare des Sciences 61, 307-314 (2008)
21. Vitanov, N.K.: Results Connected to Time Series Analysis and Machine Learning. In: Atanassov, K.T. (eds) Research in Computer Science in the Bulgarian Academy of Sciences. Studies in Computational Intelligence, vol 934, pp. 363–384. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72284-5_17
22. Brockwell P.J., Davis R.A, Calder M.V.: Introduction to Time Series and Forecasting. Springer, New York (2002)
23. Vitanov, N. K., Ausloos, M. R.: Knowledge epidemics and population dynamics models for describing idea diffusion. In: Scharnhorst A., Boerner K., van den Besselaar P. (eds.) Models of science dynamics. Understanding Complex Systems. pp. 69–125. Springer, Berlin, Heidelberg (2012). https://doi.org/10.1007/978-3-642-23068-4_3
24. Kantz, H., Holstein, D., Ragwitz, M., Vitanov, N. K.: Markov chain model for turbulent wind speed data. Physica A 342, 315–321 (2004). https://doi.org/10.1016/j.physa.2004.01.070
25. Ashenfelter, K. T., Boker, S. M., Waddell, J. R., Vitanov, N.: Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation. Journal of Experimental Psychology: Human Perception and Performance 35, 1072–1091 (2009). https://doi.org/10.1037/a0015017
26. Vitanov, N. K., Chabchoub, A., Hoffmann N.: Deep-Water Waves: on the Nonlinear Schrödinger Equation and its Solutions. Journal of Theoretical and Applied Mechanics 43, 43–54 (2013). https://doi.org/10.2478/jtam-2013-0013
27. De Gooijer N.G.: Elements of Nonlinear Time Series Analysis and Forecasting. Springer, New York (2017)
28. Vitanov, N. K.: Upper bounds on the heat transport in a porous layer. Physica D 136, 322–339 (2000). https://doi.org/10.1016/S0167-2789(99)00165-7
29. Fuchs A.: Nonlinear Dynamics in Complex Systems. Springer, Berlin (2014)
30. Vitanov, N. K., Hoffmann, N. P., Wernitz, B.: Nonlinear time series analysis of vibration data from a friction brake: SSA, PCA, and MFDFA. Chaos, Solitons & Fractals 69, 90–99 (2014). https://doi.org/10.1016/j.chaos.2014.09.010
31. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: On the trap of extinction and its elimination. Physics Letters A 349, 350–355 (2006). https://doi.org/10.1016/j.physleta.2005.09.043
32. Hale J. K.: Oscillations in Nonlinear Systems. Dover, New York (1991)
33. Borisov, R., Dimitrova, Z. I., Vitanov, N. K.: Statistical characteristics of stationary flow of substance in a network channel containing arbitrary number of arms. Entropy 22, 553 (2020). https://doi.org/10.3390/e22050553
34. Hegger, R., Kantz, H., Schreiber, T.: Practical implementation of nonlinear time series methods: The TISEAN package. Chaos: An Interdisciplinary Journal of Nonlinear Science 9, 413–435 (1999). https://doi.org/10.1063/1.166424
35. Vitanov, N. K., Vitanov, K. N.: Statistical distributions connected to motion of substance in a channel of a network. Physica A 527, 121174 (2019). https://doi.org/10.1016/j.physa.2019.121174

36. Vitanov, N. K., Borisov, R., Vitanov, K. N.: On the motion of substance in a channel and growth of random networks. Physica A 581 126207 (2021). https://doi.org/10.1016/j.physa.2021.126207
37. Hopf E.: The partial differential equation $u_t + uu_x = u_{xx}$. Communications on Pure and Applied Mathematics 3, 201–230 (1950). https://doi.org/10.1002/cpa.3160030302
38. Cole, J. D.: On a quasi-linear parabolic equation occurring in aerodynamics. Quarterly of applied mathematics 9, 225–236 (1951). https://doi.org/10.1090/QAM/42889
39. Ablowitz, M. J., Kaup, D. J., Newell, A. C., Segur, H.: The inverse scattering transform-Fourier analysis for nonlinear problems. Studies in Applied Mathematics 53, 249–315 (1974). https://doi.org/10.1002/sapm1974534249
40. Hirota, R.: The Direct Method in Soliton Theory. Cambridge University Press, Cambridge, UK (2004)
41. Kudryashov, N. A.: Simplest equation method to look for exact solutions of nonlinear differential equations. Chaos, Solitons & Fractals 24, 1217–1231 (2005). https://doi.org/10.1016/j.chaos.2004.09.109
42. Kudryashov, N. A., Loguinova, N. B.: Extended simplest equation method for nonlinear differential equations. Applied Mathematics and Computation 205, 396–402 (2008). https://doi.org/10.1016/j.amc.2008.08.019
43. Vitanov, N. K.: Recent developments of the methodology of the modified method of simplest equation with application. Pliska Studia Mathematica Bulgarica 30, 29–42 (2019).
44. Vitanov, N.K.: Modified method of simplest equation for obtaining exact solutions of nonlinear partial differential equations: history, recent developments of the methodology and studied of classes of equations. Journal of Theoretical and Applied Mechanics 49, 107–122 (2019).
45. Vitanov, N. K.: The simple equations method (SEsM) for obtaining exact solutions of nonlinear PDEs: Opportunities connected to the exponential functions. AIP Conference Proceedings vol. 2159, 030038 (2019). https://doi.org/10.1063/1.5127503
46. Vitanov, N. K., Dimitrova, Z. I.: Simple equations method (SEsM) and other direct methods for obtaining exact solutions of nonlinear PDEs. AIP Conference Proceedings vol. 2159, 030039 (2019). https://doi.org/10.1063/1.5127504
47. Martinov, N., Vitanov, N.: On the correspondence between the self-consistent 2D Poisson-Boltzmann structures and the sine-Gordon waves. Journal of Physics A: Mathematical and General 25, L51–L56 (1992). https://doi.org/10.1088/0305-4470/25/2/004
48. Martinov, N., Vitanov, N.: On some solutions of the two-dimensional sine-Gordon equation. Journal of Physics A: Mathematical and General 25, L419–L426 (1992). https://doi.org/10.1088/0305-4470/25/8/007
49. Martinov, N., Vitanov, N.: Running wave solutions of the two-dimensional sine-Gordon equation. Journal of Physics A: Mathematical and General 25, 3609–3613 (1992). https://doi.org/10.1088/0305-4470/25/12/021
50. Martinov, N. K., Vitanov, N. K.: New class of running-wave solutions of the (2+ 1)-dimensional sine-Gordon equation. Journal of Physics A: Mathematical and General 27, 4611–4618 (1994). https://doi.org/10.1088/0305-4470/27/13/034
51. Martinov, N. K., Vitanov, N. K.: On self-consistent thermal equilibrium structures in two-dimensional negative-temperature systems. Canadian Journal of Physics 72, 618–624 (1994). https://doi.org/10.1139/p94-079
52. Vitanov, N. K., Martinov, N. K.: On the solitary waves in the sine-Gordon model of the two-dimensional Josephson junction. Zeitschrift für Physik B Condensed Matter 100, 129–135 (1996). https://doi.org/10.1007/s002570050102
53. Vitanov, N. K.: On travelling waves and double-periodic structures in two-dimensional sine-Gordon systems. Journal of Physics A: Mathematical and General 29, 5195–5207 (1996). https://doi.org/10.1088/0305-4470/29/16/036
54. Vitanov, N. K.: Complicated exact solutions to the 2+ 1-dimensional sine-Gordon equation. Zeitschrift für angewandte Mathematik und Mechanik, 78, S787–S788 (1998).
55. Vitanov, N. K.: Breather and soliton wave families for the sine-Gordon equation. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 454, 2409–2423 (1998). https://doi.org/10.1098/rspa.1998.0264

56. Vitanov, N. K., Jordanov, I. P., Dimitrova, Z. I.: On nonlinear dynamics of interacting populations: Coupled kink waves in a system of two populations. Communications in Nonlinear Science and Numerical Simulation 14, 2379–2388 (2009). https://doi.org/10.1016/j.cnsns.2008.07.015

57. Vitanov, N. K., Jordanov, I. P., Dimitrova, Z. I.: On nonlinear population waves. Applied Mathematics and Computation 215, 2950–2964 (2009). https://doi.org/10.1016/j.amc.2009.09.041

58. Vitanov, N. K.: Application of simplest equations of Bernoulli and Riccati kind for obtaining exact traveling-wave solutions for a class of PDEs with polynomial nonlinearity. Communications in Nonlinear Science and Numerical Simulation 15, 2050–2060 (2010). https://doi.org/10.1016/j.cnsns.2009.08.011

59. Vitanov, N. K., Dimitrova, Z. I.: Application of the method of simplest equation for obtaining exact traveling-wave solutions for two classes of model PDEs from ecology and population dynamics. Communications in Nonlinear Science and Numerical Simulation 15, 2836–2845 (2010). https://doi.org/10.1016/j.cnsns.2009.11.029

60. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: Modified method of simplest equation and its application to nonlinear PDEs. Applied Mathematics and Computation 216, 2587–2595 (2010). https://doi.org/10.1016/j.amc.2010.03.102

61. Vitanov, N. K.: Modified method of simplest equation: powerful tool for obtaining exact and approximate traveling-wave solutions of nonlinear PDEs. Communications in Nonlinear Science and Numerical Simulation 16, 1176–1185 (2011). https://doi.org/10.1016/j.cnsns.2010.06.011

62. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: On the class of nonlinear PDEs that can be treated by the modified method of simplest equation. Application to generalized Degasperis-Processi equation and b-equation. Communications in Nonlinear Science and Numerical Simulation 16, 3033–3044 (2011). https://doi.org/10.1016/j.cnsns.2010.11.013

63. Vitanov, N. K.: On modified method of simplest equation for obtaining exact and approximate solutions of nonlinear PDEs: the role of the simplest equation. Communications in Nonlinear Science and Numerical Simulation 16, 4215–4231 (2011). https://doi.org/10.1016/j.cnsns.2011.03.035

64. Vitanov, N. K.: On modified method of simplest equation for obtaining exact solutions of nonlinear PDEs: case of elliptic simplest equation. Pliska Studia Mathematica Bulgarica 21, 257–266 (2012).

65. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: Application of the method of simplest equation for obtaining exact traveling-wave solutions for the extended Korteweg-de Vries equation and generalized Camassa-Holm equation. Applied Mathematics and Computation 219, 7480–7492 (2013). https://doi.org/10.1016/j.amc.2013.01.035

66. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Traveling waves and statistical distributions connected to systems of interacting populations. Computers & Mathematics with Applications 66, 1666–1684 (2013). https://doi.org/10.1016/j.camwa.2013.04.002

67. Vitanov, N. K., Dimitrova, Z. I.: Solitary wave solutions for nonlinear partial differential equations that contain monomials of odd and even grades with respect to participating derivatives. Applied Mathematics and Computation 247, 213–217 (2014). https://doi.org/10.1016/j.amc.2014.08.101

68. Vitanov, N. K., Dimitrova, Z. I., Ivanova, T. I.: On solitary wave solutions of a class of nonlinear partial differential equations based on the function $1/cosh^n$ (αx+ βt). Applied Mathematics and Computation 315, 372–380 (2017). https://doi.org/10.1016/j.amc.2017.07.064

69. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Modified method of simplest equation for obtaining exact analytical solutions of nonlinear partial differential equations: further development of the methodology with applications. Applied Mathematics and Computation 269, 363–378 (2015). https://doi.org/10.1016/j.amc.2015.07.060

70. Vitanov, N. K., Dimitrova, Z. I.: On the modified method of simplest equation and the nonlinear Schrödinger equation. Journal of Theoretical and Applied Mechanics 48, 59–68 (2018).

71. Nikolova, E. V., Jordanov, I. P., Dimitrova, Z. I., Vitanov, N. K.: Evolution of nonlinear waves in a blood-filled artery with an aneurysm. AIP Conference Proceedings vol. 1895, 070002 (2017). https://doi.org/10.1063/1.5007391

72. Jordanov, I.P., Vitanov, N.K.: On the Exact Traveling Wave Solutions of a Hyperbolic Reaction-Diffusion Equation. In: Georgiev, K., Todorov, M., Georgiev, I. (eds.) Advanced Computing in Industrial Mathematics. BGSIAM 2017. Studies in Computational Intelligence, vol. 793, pp. 199–210. Springer, Cham. (2019). https://doi.org/10.1007/978-3-319-97277-0_16

73. Nikolova, E. V., Chilikova-Lubomirova, M., Vitanov, N. K.: Exact solutions of a fifth-order Korteweg-de Vries-type equation modeling nonlinear long waves in several natural phenomena. AIP Conference Proceedings vol. 2321, 030026 (2021). https://doi.org/10.1063/5.0040089

74. Vitanov, N. K.: Simple equations method (SEsM) and its connection with the inverse scattering transform method. AIP Conference Proceedings vol. 2321, 030035 (2021). https://doi.org/10.1063/5.0040409

75. Nikolova, E.V., Serbezov, D.Z., Jordanov, I.P., Vitanov, N.K.: Non-linear Waves of Interacting Populations with Density-Dependent Diffusion. In: Georgiev, I., Kostadinov, H., Lilkova, E. (eds.) Advanced Computing in Industrial Mathematics. BGSIAM 2018. Studies in Computational Intelligence, vol. 961, pp. 324–332. Springer, Cham. (2021). https://doi.org/10.1007/978-3-030-71616-5_29

76. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Simple Equations Method (SEsM): Algorithm, connection with Hirota method, Inverse Scattering Transform Method, and several other methods. Entropy, 23, 10 (2021). https://doi.org/10.3390/e23010010

77. Vitanov. N.K.: Schrödinger Equation and Nonlinear Waves, Simpao, V., H. Little H.(eds.) In: Understanding the Schrödinger Equation. pp. 37–92. Nova Science Publishers, New York (2020)

78. Vitanov, N. K., Dimitrova, Z. I.: Simple Equations Method and non-linear differential equations with non-polynomial non-linearity. Entropy 23, 1624 (2021). https://doi.org/10.3390/e23121624

79. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: On the use of composite functions in the Simple Equations Method to obtain exact solutions of nonlinear differential equations. Computation 9, 104 (2021). https://doi.org/10.3390/computation9100104

80. Vitanov, N. K., Dimitrova, Z. I.: Simple equations method (SEsM) and its particular cases: Hirota method. AIP Conference Proceedings vol. 2321, 030036 (2021). https://doi.org/10.1063/5.0040410

81. Vitanov, N. K.: Simple equations method (SEsM) and nonlinear PDEs with fractional derivatives. AIP Conference Proceedings vol. 2459, 030040 (2022). https://doi.org/10.1063/5.0083566

82. Dimitrova, Z. I.: Several examples of application of the simple equations method (SEsM) for obtaining exact solutions of nonlinear PDEs. AIP Conference Proceedings vol. 2459, 030005 (2022). https://doi.org/10.1063/5.0083572

83. Vitanov, N. K.: Simple equations method (SEsM): Review and new results. AIP Conference Proceedings vol. 2459, 020003 (2022). https://doi.org/10.1063/5.0083565

84. Dimitrova, Z. I.: On several specific cases of the simple equations method (SEsM): Jacobi elliptic function expansion method, F-expansion method, modified simple equation method, trial function method, general projective Riccati equations method, and first intergal method. AIP Conference Proceedings vol. 2459, 030006 (2022). https://doi.org/10.1063/5.0083573

85. Constantine, G.M., Savits, T.H.: A multivariate Faa di Bruno formula with applications. Transactions of the American Mathematical Society 348, 503–520 (1996). https://doi.org/10.1090/S0002-9947-96-01501-2

A Note on the Steady Poiseuille Flow of Carreau–Yasuda Fluid

Nikolay Kutev and Sonia Tabakova

Abstract The steady Poiseuille flow of Carreau–Yasuda fluid in a pipe, caused by a constant pressure gradient, is studied theoretically. It is proved that at some values of the viscosity model parameters, the problem has a generalized solution, while at others—classical solution. For the latter, a necessary and sufficient condition is found, which depends on the pressure gradient and Carreau–Yasuda model parameters.

Keywords Carreau–Yasuda fluid · Steady Poiseuille flow · Classical solution · Necessary and sufficient condition

1 Introduction

The Poiseuille flow problem is one of the fundamental problems of fluid mechanics, corresponding to a parabolic velocity profile in a pipe or channel when the flow is laminar and fully developed in the axial direction. For a Newtonian fluid, the problem has a well-known analytical solution [1]. However, for non-Newtonian fluids, it has no analytical solution, except for some special cases for generalized Newtonian fluid models, such as the power law model of viscosity dependence on shear rate [2]. The behavior of the so-called shear-thinning fluids (viscosity is decreasing function of shear rate), for example, polymer solutions, polymer melts, suspensions, emulsions, and some biological fluids, is described by different viscosity models, such as the power law model, Carreau model, Carreau–Yasuda model, and others [2–4].

N. Kutev
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria
e-mail: kutev@math.bas.bg

S. Tabakova (✉)
Institute of Mechanics, Bulgarian Academy of Sciences, Sofia, Bulgaria
e-mail: stabakova@gmail.com

The general unsteady and oscillatory cases of the Poiseuille flow in an infinite channel or pipe using the Carreau or Carreau–Yasuda models are studied in our previous papers [5–8]. The most interesting case of negative power index in these models, $n < 0$, corresponds to severe pseudoplastic behavior [9–11]. It has been studied theoretically in [7] and [8], but the sufficient and necessary condition for solution existence is still a challenge. The aim of the present work is to prove the solution existence or non-existence for the different parameters of the steady Poiseuille flow problem of the Carreau–Yasuda fluid and to find the sufficient and necessary condition for classical solution existence at $n < 0$.

The dimensionless equations of motion and continuity for the steady pipe flow are as follows:

$$\mathbf{v} \cdot \nabla \mathbf{v} = -\nabla p + \nabla \cdot \mathbf{T}, \tag{1}$$

$$\nabla \cdot \mathbf{v} = 0, \tag{2}$$

where $\mathbf{v}$ is velocity vector, p—pressure, $\mathbf{T} = \mu_{app} \dot{S}$—viscous stress tensor with $\dot{S} = \dfrac{1}{2} \left(\nabla \mathbf{v} + \nabla \mathbf{v}^T \right)$ as strain rate tensor, and μ_{app}—apparent viscosity (constant for Newtonian fluid and a non-linear function of shear rate $\dot{\gamma} = \sqrt{2\dot{S} : \dot{S}}$ for generalized Newtonian fluid).

The flow is assumed fully developed laminar driven by a constant pressure gradient b in the axial direction. Then, in the adopted cylindrical coordinate system (x, Y, φ) with x as the axial coordinate, the flow velocity is simplified as $\mathbf{v} = (U(Y), 0, 0)$ and $\mathbf{T}$ reduces to the scalar $\mu_{app}(\dot{\gamma})\dot{\gamma}$, where $\dot{\gamma} = | U_Y |$. The Carreau–Yasuda viscosity model is assumed for the apparent viscosity μ_{app}:

$$\mu_{app} = 1 - c + c[1 + Cu^\alpha \dot{\gamma}^\alpha]^{(n-1)/\alpha}, \tag{3}$$

where $c = 1 - \dfrac{\mu_\infty}{\mu_0}$, with μ_0, μ_∞ as the upper and lower limits of the viscosity corresponding to the low and high shear rates, Cu is the Carreau number (Weissenberg number), and α and n are empirically determined for the considered fluid.

Applying the upper assumptions, the system (1)–(3) is reduced to one elliptic equation for the axial velocity:

$$L(U(Y)) = \frac{1}{Y} \frac{d}{dY} \left\{ \left[1 - c + c \left(1 + Cu^\alpha \mid U_Y \mid^\alpha \right)^{\frac{n-1}{\alpha}} \right] Y U_Y \right\} = b \tag{4}$$

with the axisymmetrical and no-slip boundary conditions:

$$U_Y(0) = U(R) = 0, \tag{5}$$

where R is the dimensionless pipe radius (the pipe radius or diameter is usually taken as characteristic length). Here $Cu > 0$, $b \in \mathbb{R}$, $\alpha > 0$, $n \in \mathbb{R}$, $c \in [0, 1]$ are empirically determined parameters.

If $b = 0$, then (4), (5) has only the trivial solution $U(Y) \equiv 0$. It is clear that $U(Y, -b) = -U(Y, b)$, where $U(Y, b), U(Y, -b)$ are solutions of (4), (5) with right-hand side b and $-b$, correspondingly. That is why further on we consider only the case $b > 0$.

The outline of the paper is as follows. In Sect. 2, the main results are formulated in four theorems to be proved in Sect. 3. An application, connected with polymer pipe flows with a lower limit of viscosity tending to zero, i.e., at $c = 1$, is presented by one theorem and found in Sect. 4.

2 Main Results

In order to formulate the main results, we introduce the function

$$F(\zeta) = \left[1 - c + c \left(1 + Cu^\alpha \zeta^\alpha\right)^{\frac{n-1}{\alpha}}\right] \zeta \quad \text{for} \quad \zeta \geq 0, \tag{6}$$

where $\zeta = | U_Y |$.

Theorem 1 *Suppose $b > 0$, $\alpha > 0$, and either $c = 0$, $Cu = 0$, or $n = 1$. Then problem (4), (5) is Newtonian one (classical Poiseuille problem in a pipe) and has a unique classical solution:*

$$U(Y) = \frac{b}{4}(Y^2 - R^2) \quad for \quad every \quad Y \in [0, R]. \tag{7}$$

Theorem 2 *Suppose $b > 0$, $\alpha > 0$, $c \in (0, 1)$, and either one of the following conditions holds:*

$$n \geq 0; \tag{8}$$

$$n < 0 \quad and \quad \left(1 - \frac{\alpha + 1}{n}\right)^{\frac{n-1-\alpha}{\alpha}} < \frac{1 - c}{\alpha c}. \tag{9}$$

Then problem (4), (5) has a unique classical solution $U(Y) \in C^2([0, R])$ satisfying

$$U(Y) = -\int_Y^R F^{-1}\left(\frac{bs}{2}\right) ds \quad and \tag{10}$$

$$0 \leq U_Y(Y) \leq F^{-1}\left(\frac{bR}{2}\right) \quad for \quad Y \in [0, R]. \tag{11}$$

Theorem 3 *Suppose $b > 0$, $\alpha > 0$, $n < 0$, $c \in (0, 1)$, and*

$$\left(1 - \frac{\alpha + 1}{n}\right)^{\frac{n-1-\alpha}{\alpha}} = \frac{1 - c}{\alpha c}. \tag{12}$$

If

$$\left[1 - c + c \left(1 - \frac{\alpha + 1}{n} \right)^{\frac{n-1}{\alpha}} \right] Cu^{-1} \left(-\frac{\alpha + 1}{n} \right)^{\frac{1}{\alpha}} \geq \frac{bR}{2}, \tag{13}$$

then problem (4), (5) has a unique classical solution $U(Y)$ satisfying (10), (11);
If

$$\left[1 - c + c \left(1 - \frac{\alpha + 1}{n} \right)^{\frac{n-1}{\alpha}} \right] Cu^{-1} \left(-\frac{\alpha + 1}{n} \right)^{\frac{1}{\alpha}} < \frac{bR}{2}, \tag{14}$$

then problem (4), (5) has a unique generalized solution $U(Y) \in C^2([0, R] \setminus Y_0) \bigcap$
$C^1([0, R])$ satisfying (10), (11), $U_{YY}(Y_0) = \infty$, where

$$Y_0 = \frac{2}{b} \left[1 - c + c \left(1 - \frac{\alpha + 1}{n} \right)^{\frac{n-1}{\alpha}} \right] Cu^{-1} \left(-\frac{\alpha + 1}{n} \right)^{\frac{1}{\alpha}} \quad \in (0, R). \tag{15}$$

Remark 1 If (13) is strict inequality, then $U(Y) \in C^2([0, R])$, while when (13) is
an equality, then $U(Y) \in C^2([0, R)) \bigcap C^1([0, R])$, $U_{YY}(R) = \infty$.

Theorem 4 *Suppose $b > 0$, $\alpha > 0$, $n < 0$, $c \in (0, 1)$, and*

$$\left(1 - \frac{\alpha + 1}{n} \right)^{\frac{n-1-\alpha}{\alpha}} > \frac{1 - c}{\alpha c}. \tag{16}$$

Then problem (4), (5) has a unique classical solution iff

$$\frac{bR}{2} \leq \left[1 - c + c \left(1 + Cu^\alpha \zeta_1^\alpha \right)^{\frac{n-1}{\alpha}} \right] \zeta_1 = F(\zeta_1), \tag{17}$$

where ζ_1 is the first positive zero of $F'(\zeta) = 0$. Moreover, $U(Y)$ satisfies (10), (11)
and

$$0 \leq U_Y(Y) \leq F^{-1} \left(\frac{bR}{2} \right) \leq \zeta_1 \quad for \quad Y \in [0, R]. \tag{18}$$

Remark 2 If (17) is a strict inequality, then $U(Y) \in C^2([0, R])$. When (17) is an
equality, then $U(Y) \in C^2([0, R)) \bigcap C^1([0, R])$, $U_{YY}(R) = \infty$.

3   Proofs of the Main Results

Integrating $Y.L(U(Y))$ from 0 to $Y \in (0, R]$, we get the identity

$$\left[1 - c + c \left(1 + Cu^\alpha \mid U_Y \mid^\alpha \right)^{\frac{n-1}{\alpha}} \right] U_Y = \frac{bY}{2} \quad for \quad Y \in [0, R]. \tag{19}$$

Hence $U_Y(Y) > 0$ for $Y \in (0, R]$ and (19) is equivalent to

$$F(U_Y(Y)) = \left[1 - c + c\left(1 + Cu^\alpha U_Y^\alpha\right)^{\frac{n-1}{\alpha}}\right] U_Y = \frac{bY}{2} \quad \text{for} \quad Y \in [0, R]. \quad (20)$$

Proof of Theorem 1. For $c = 0$ or $n = 1$, we have $F(\zeta) = \zeta$ and (20) becomes

$$U_Y(Y) = \frac{bY}{2} \quad \text{for} \quad Y \in [0, R]. \quad (21)$$

Integrating (21) from 0 to $Y \in (0, R]$, we obtain from (5) that $U(Y) = \frac{b}{4}(Y^2 - R^2)$, which proves Theorem 1.

Proof of Theorem 2. Simple computations give us for $\zeta \geq 0$

$$F'(\zeta) = 1 - c + c\left(1 + Cu^\alpha \zeta^\alpha\right)^{\frac{n-1-\alpha}{\alpha}}\left(1 + nCu^\alpha \zeta^\alpha\right), \quad (22)$$

$$F''(\zeta) = c(n-1)Cu^\alpha \zeta^{\alpha-1}\left(1 + Cu^\alpha \zeta^\alpha\right)^{\frac{n-1-2\alpha}{\alpha}}\left(\alpha + 1 + nCu^\alpha \zeta^\alpha\right). \quad (23)$$

Thus, from (22), we get

$$F'(\zeta) \geq 1 - c + c\left(1 + Cu^\alpha \zeta^\alpha\right)^{\frac{n-1-\alpha}{\alpha}} > 0 \quad \text{for} \quad \zeta \geq 0, \quad n \geq 0, \quad \text{and} \quad c \in (0, 1]. \quad (24)$$

If (9) holds, then from (23) we obtain

$$F''(\zeta_0) = 0 \quad \text{for} \quad Cu^\alpha \zeta_0^\alpha = -\frac{\alpha + 1}{n} > 0, \quad (25)$$

$$F''(\zeta) < 0 \quad \text{for} \quad \zeta < \zeta_0 \quad \text{, i.e., for} \quad Cu^\alpha \zeta^\alpha < -\frac{\alpha + 1}{n}, \quad \text{and} \quad (26)$$

$$F''(\zeta) > 0 \quad \text{for} \quad \zeta > \zeta_0 \quad \text{, i.e., for} \quad Cu^\alpha \zeta^\alpha > -\frac{\alpha + 1}{n}. \quad (27)$$

From (25)–(27), it follows that the function $F'(\zeta)$ has a global minimum in the interval $[0, \infty)$ at the point $\zeta = \zeta_0$. Since from (9), we have

$$F'(\zeta_0) = 1 - c - \alpha c\left(1 - \frac{\alpha + 1}{n}\right)^{\frac{n-1-\alpha}{\alpha}}, \quad (28)$$

$$F'(\zeta) \geq F'(\zeta_0) > 0 \quad \text{for} \quad \zeta \geq 0. \quad (29)$$

Under the conditions (8), (9) and from (24), (19), it is seen that $F'(\zeta) > 0$ and $F(\zeta)$ is strictly monotone increasing. Moreover,

$$F(0) = 0, \quad \lim_{\zeta \to \infty} F(\zeta) = \infty, \tag{30}$$

so that there exits the inverse function $F^{-1}(\zeta)$. From

$$(F^{-1}(\zeta))' = \frac{1}{F'(F^{-1}(\zeta))} > 0 \tag{31}$$

it follows that $F^{-1}(\zeta) : [0, \infty) \to [0, \infty)$ is strictly monotone increasing function and (20) is equivalent to

$$U_Y(Y) = F^{-1}\left(\frac{bY}{2}\right) \quad \text{for} \quad Y \in [0, R], \tag{32}$$

where

$$F^{-1}\left(\frac{bY}{2}\right) : [0, R] \to \left[0, F^{-1}\left(\frac{bR}{2}\right)\right]. \tag{33}$$

Integrating (32) from $Y \in [0, R)$ to R, we get from (5) that the function

$$U(Y) = -\int_Y^R F^{-1}\left(\frac{bs}{2}\right) ds \quad \text{for} \quad Y \in [0, R] \tag{34}$$

is the unique classical solution of (4), (5), when one of the conditions (8), (9) is satisfied. Moreover, from the monotonicity of $F^{-1}(\zeta)$ and (32), the estimate (11) holds and Theorem 2 is completed.

Proof of Theorem 3. From (12), the inequalities (28), (29) become

$$F'(\zeta_0) = 0, \quad F'(\zeta) > 0 \quad \text{for} \quad \zeta \geq 0, \quad \zeta \neq \zeta_0. \tag{35}$$

Thus, $F(\zeta)$ is a monotone increasing function, and the inverse function $F^{-1}(\zeta)$ exists and is monotone increasing, which is evident from (31). Moreover, (30) holds and $F(\zeta) : [0, \infty) \to [0, \infty)$, $F^{-1}(\zeta) : [0, \infty) \to [0, \infty)$. Hence, equation (20) is equivalent to (32), and after integration, as in the proof of Theorem 2, the unique solution of (4), (5) is given by (34).

Let us analyze the regularity of $U(Y)$. It is clear that $U(Y) \in C^1([0, R])$ and from the monotonicity of F^{-1} the estimate (11) holds. Since (13) is equivalent to

$$F(\zeta_0) \geq \frac{bR}{2}, \quad \text{i.e.,} \quad \zeta_0 \geq F^{-1}\left(\frac{bR}{2}\right) \tag{36}$$

and after differentiating (32), we get from (31), (26) the estimates

$$U_{YY}(Y) = \frac{\partial}{\partial Y}\left(F^{-1}\left(\frac{bY}{2}\right)\right) = \frac{b}{2F'(F^{-1}\left(\frac{bY}{2}\right))} \leq \frac{b}{2F'(F^{-1}\left(\frac{bR}{2}\right))} \tag{37}$$

$$\leq \frac{b}{2F'(F^{-1}(\zeta_0))} = \infty \quad \text{for} \quad Y \in [0, R].$$

Thus

$$0 < U_{YY}(Y) < \infty \quad \text{for} \quad Y \in [0, R) \quad \text{and} \quad U_{YY}(R) = \infty, \tag{38}$$

when (13), equivalently (36), is a strict inequality, because

$$\frac{b}{2F'(F^{-1}\left(\frac{bR}{2}\right))} < \frac{b}{2F'(F^{-1}(\zeta_0))} = \infty.$$

However, when (13), equivalently (36), is equality, then

$$U_{YY}(R) = \frac{b}{2F'(F^{-1}\left(\frac{bR}{2}\right))} = \frac{b}{2F'(F^{-1}(\zeta_0))} = \infty.$$

Remark 1 and the proof of (13) are completed.

Suppose that (14) holds. As in the proof of (13) $U(Y) \in C^1([0, R])$ is the unique solution of (4), (5) satisfying (10), (11). Conditions (14), (15) are equivalent to

$$F(\zeta_0) < \frac{bR}{2}, \quad F(\zeta_0) = \frac{bY_0}{2}, \quad \text{i.e.,} \quad \zeta_0 < F^{-1}\left(\frac{bR}{2}\right), \quad \zeta_0 = F^{-1}\left(\frac{bY_0}{2}\right) \tag{39}$$

and therefore $Y_0 \in (0, R)$.

From the monotonicity of F^{-1}, we get from (39)

$$F^{-1}\left(\frac{bY}{2}\right) < F^{-1}\left(\frac{bY_0}{2}\right) = \zeta_0 \quad \text{for} \quad Y \in [0, Y_0), \tag{40}$$

$$F^{-1}\left(\frac{bY}{2}\right) > F^{-1}\left(\frac{bY_0}{2}\right) = \zeta_0 \quad \text{for} \quad Y \in (Y_0, R]. \tag{41}$$

Repeating the estimate (37), from (41) and the monotonicity of $F'(\zeta)$, see (26), (27), we have

$$U_{YY}(Y_0) = \infty, \quad 0 < U_{YY}(Y) < U_{YY}(Y_0) = \infty \quad \text{for} \quad Y \in [0, R] \setminus Y_0, \tag{42}$$

i.e., $U(Y) \in C^2([0, R] \setminus Y_0) \bigcap C^1([0, R])$, which proves (14) and Theorem 3.

Proof of Theorem 4. Under conditions (16), (28), it follows that $F'(\zeta)$ has a strictly negative minimum in the interval $[0, \infty)$ at the point ζ_0. Since

$$F'(0) = 0, \quad \lim_{\zeta \to \infty} F'(\zeta) = 1 - c > 0 \tag{43}$$

the function $F'(\zeta)$ has two positive roots $0 < \zeta_1 < \zeta_2 < \infty$, $F'(\zeta_i) = 0$, $i = 1, 2$, $\zeta_0 \in (\zeta_1, \zeta_2)$. Moreover, from (43) and (26), it follows that

$$F'(\zeta) > 0, \quad F''(\zeta) < 0 \quad \text{for} \quad \zeta \in [0, \zeta_1), \quad F'(\zeta_1) = 0, \tag{44}$$

i.e., $F'(\zeta)$ is strictly monotone decreasing, while $F(\zeta)$ is strictly monotone increasing for $\zeta \in [0, \zeta_1)$. Hence, the inverse function

$$F^{-1}(\zeta) : [0, \zeta_1) \to [0, F^{-1}(\zeta_1)] \tag{45}$$

is well defined and from (31), (44) is strictly monotone increasing.

Sufficiency: From (17), we get

$$\zeta_1 \geq F^{-1}\left(\frac{bR}{2}\right) \quad \text{and} \quad F^{-1}\left(\frac{bY}{2}\right) : [0, R] \to \left[0, F^{-1}\left(\frac{bR}{2}\right)\right]. \tag{46}$$

Since $\left[0, F^{-1}\left(\frac{bR}{2}\right)\right] \subset [0, \zeta_1)$, from (45) equation (20) is equivalent to (32) and after integration $U(Y) \in C^1([0, R])$ satisfies (34). Moreover, from the monotonicity of $F^{-1}(\zeta)$ and (46), the estimate (18) is satisfied because

$$0 \leq U_Y(Y) \leq U_Y(R) = F^{-1}\left(\frac{bR}{2}\right) \leq \zeta_1.$$

Repeating the estimate (37), we get

$$0 < U_{YY}(Y) < \infty \quad \text{for} \quad Y \in [0, R], \tag{47}$$

when (17), equivalently (46), is a strict inequality, i.e., $U(Y) \in C^2([0, R])$.

If (17), equivalently (46), is an equality, then

$$U_{YY}(R) = \frac{b}{2F'(F^{-1}\left(\frac{bR}{2}\right))} = \frac{b}{2F'(\zeta_1)} = \infty, \quad U(Y) \in C^2([0, R)) \cap C^1([0, R]), \tag{48}$$

which proves the sufficiency in Theorem 4 and Remark 2.

Necessity: Suppose $U(Y) \in C^2((0, R)) \cap C^1([0, R])$ is a classical solution of (4), (5). Then $U(Y)$ satisfies (20). We assume by contradiction that (17) fails, i.e.,

$$\frac{bR}{2} > F(\zeta_1). \tag{49}$$

If $Y_0 = \dfrac{2F(\zeta_1)}{b}$, then from (49), we have $Y_0 \in (0, R)$ and from (44), (45)

$$\zeta_1 = F^{-1}\left(\frac{bY_0}{2}\right). \tag{50}$$

Thus, (50) and (37) for $Y = Y_0$ give us

$$U_{YY}(Y_0) = \frac{b}{2F'(F^{-1}\left(\frac{bY_0}{2}\right))} = \frac{b}{2F'(\zeta_1)} = \infty. \tag{51}$$

Since Y_0 is an interior point of the interval $(0, R)$, it follows from (51) that $U(Y)$ is not a classical $C^2(0, R)$ smooth solution of (4), (5) and Theorem 4 is completed.

4 Applications. Case c = 1

In the applications of polymer pipe flows, the special case of $c = 1$ is important $(\mu_\infty \to 0 \,[2])$.

Theorem 5 *Suppose $b > 0$, $\alpha > 0$, $n \in \mathbb{R}$, and $c = 1$. Then*

(a) for

$$n > 0 \tag{52}$$

problem (4), (5) has a unique classical solution $U(Y) \in C^2([0, R])$ satisfying (10), (11);
(b) for

$$n = 0 \tag{53}$$

problem (4), (5) has a unique classical solution $U(Y) \in C^2((0, R)) \cap C^1([0, R])$ iff

$$\frac{bR}{2} \leq Cu^{-1} \tag{54}$$

and $U(Y)$ satisfies (10), (11);
(c) for

$$n < 0 \tag{55}$$

problem (4), (5) has a unique classical solution $U(Y) \in C^2((0, R)) \cap C^1([0, R])$ iff

$$\frac{bR}{2} \leq Cu^{-1}\left(\frac{n-1}{n}\right)^{\frac{n-1}{\alpha}}\left(-\frac{1}{n}\right)^{\frac{1}{\alpha}} = F\left(Cu^{-1}\left(-\frac{1}{n}\right)^{\frac{1}{\alpha}}\right) \tag{56}$$

and $U(Y)$ satisfies (10), (11).

Remark 3 If (54) or (56) is strict inequality, then $U(Y) \in C^2([0, R])$. When (54) or (56) is equality then $U(Y) \in C^2((0, R)) \cap C^1([0, R])$ and $U_{YY}(R) = \infty$.

Proof (a) Suppose (52) holds. Then from (24)

$$F'(\zeta) \geq (1 + Cu^\alpha \zeta^\alpha)^{\frac{n-1-\alpha}{\alpha}} > 0 \quad \text{for} \quad \zeta \geq 0 \tag{57}$$

and $F(\zeta)$ is strictly monotone increasing function.

Since $F(0) = 0$, $\lim_{\zeta \to \infty} F(\zeta) = \infty$, there exists the inverse function $F^{-1}(\zeta)$: $[0, \infty) \to [0, \infty)$, which is strictly monotone increasing function from (31) and (57). Thus problem (4), (5) is equivalent to (20), and after integration, $U(Y)$ satisfies (10), (11).

(b) Suppose (53) holds. Here, again (57) is valid, from which $F(\zeta)$ is strictly monotone increasing function for $\zeta \geq 0$. Since $F(0) = 0$, $\lim_{\zeta \to \infty} F(\zeta) = Cu^{-1} = \sup_{\zeta \geq 0} F(\zeta)$, there exists the inverse function

$$F^{-1}(\zeta) : [0, Cu^{-1}) \to [0, \infty), \tag{58}$$

which is strictly monotone increasing function from (31) and (57).

Sufficiency: From (54) $F^{-1}\left(\dfrac{bY}{2}\right) : \left[0, \dfrac{bR}{2}\right] \to \left[0, F^{-1}\left(\dfrac{bR}{2}\right)\right]$ is well defined for $Y \in [0, R]$ because $\left[0, \dfrac{bR}{2}\right] \subset [0, Cu^{-1}]$. The rest of the proof is identical to the proof of *(a)* of Theorem 5.

Necessity: If $U(Y) \in C^2((0, R)) \cap C^1([0, R])$ is a classical solution of (4), (5), we suppose by contradiction that $\dfrac{bR}{2} > Cu^{-1}$.

Then $Y_0 = \dfrac{2}{bCu} \in (0, R)$ and from (20) and the monotonicity of $F(\zeta)$ we have for $Y \in (Y_0, R)$ the following impossible chain of inequalities:

$$Cu^{-1} = \frac{bY_0}{2} < \frac{bY}{2} = F(U_Y(Y)) \leq \sup_{\zeta \geq 0} F(\zeta) = Cu^{-1}. \tag{59}$$

(c) Suppose (55) holds. From (22), (23), we get

$$F'(\zeta) = (1 + Cu^\alpha \zeta^\alpha)^{\frac{n-1-\alpha}{\alpha}} (1 + nCu^\alpha \zeta^\alpha), \tag{60}$$

$$F''(\zeta) = (n-1)Cu^\alpha \zeta^{\alpha-1} (1 + Cu^\alpha \zeta^\alpha)^{\frac{n-1-\alpha}{\alpha}} (\alpha + 1 + nCu^\alpha \zeta^\alpha), \tag{61}$$

and $F'(\zeta) > 0$ for $Cu^\alpha \zeta^\alpha < -\dfrac{1}{n}$, $F'(\zeta) = 0$ for $Cu^\alpha \zeta^\alpha = -\dfrac{1}{n}$, and $F'(\zeta) < 0$ for $Cu^\alpha \zeta^\alpha > -\dfrac{1}{n}$. Thus, the function $F(\zeta)$ has a global maximum in the interval $[0, \infty)$ at the point $\zeta_1 = Cu^{-1}\left(-\dfrac{1}{n}\right)^{\frac{1}{\alpha}}$ and

$$F(\zeta_1) = Cu^{-1} \left(\frac{n-1}{n}\right)^{\frac{n-1}{\alpha}} \left(-\frac{1}{n}\right)^{\frac{1}{\alpha}} > 0.$$

The rest of the proof is identical to the proof of Theorem 4 and we omit it. The proof of Remark 1 is identical to the proof of Remarks 1 and 2.

In practice, the necessary and sufficient condition (56) corresponds to three different scenarios for solution existence of (4), (5), which depend on the values of the parameters n, α, Cu, b, and R: (i) if (56) is strict inequality, then ζ_1 is not reached for any $Y \in [0, R]$ and there exist two solutions of (4), but only one of them is the solution of (5); (ii) if (56) is equality, then there is only one solution of (4), (5); (iii) if (56) is not fulfilled, then ζ_1 is reached at some inner point $Y_1 \in [0, R)$, thus there is no solution of (4), (5).

References

1. Schlichting, H., Gersten, K.: Boundary Layer Theory. Springer, Berlin (2000).
2. Bird, R.B., Stewart, W.E., Lightfoot, E.N.: Transport phenomena. John Wiley & Sons, Inc., New York (2002).
3. Chhabra, R.P.: Bubbles, drops, and particles in non-Newtonian fluids. CRC Press, Taylor & Francis Group, Boca Raton (2007).
4. Boyd, J., Buick, J. M., Green, S.: Analysis of the Casson and Carreau-Yasuda non-Newtonian blood models in steady and oscillatory flows using the lattice Boltzmann method, Physics of Fluids **19**, 093103 (2007). https://doi.org/10.1063/1.2772250
5. Tabakova, S., Kutev, N., Radev, St.: Oscillatory Carreau flows in straight channels, Royal Soc. Open Sci. **7**, 191305 (2020). http://dx.doi.org/10.1098/rsos.191305
6. Kutev, N., Tabakova, S., Radev, St.: Unsteady flow of Carreau fluid in a pipe, Z. Angew. Math. Phys. **72**, 196. (2021). https://doi.org/10.1007/s00033-021-01624-5
7. Kutev, N., Tabakova, S.: Oscillatory flow of Carreau-Yasuda fluid in a pipe, *AIP Conference Proceedings* **2522**, 030005 (2022). https://doi.org/10.1063/5.0100980
8. Kutev, N., Tabakova, S.: Analysis of general unsteady flow of Carreau-Yasuda fluid in a pipe, *AIP Conference Proceedings* **2459**, 030019 (2022). https://doi.org/10.1063/5.0083551
9. R. Kotsilkova, S. Tabakova, R. Ivanova, Effect of graphene nanoplatelets and multiwalled carbon nanotubes on the viscous and viscoelastic properties and printability of polylactide nanocomposites, Mech Time-Depend Mater (2021). https://doi.org/10.1007/s11043-021-09503-2
10. A. Arzumand, S. Srinivas, Y. Yuan, C. Zou, D. Sarkar, Mechano - Morphological Characterization of Polyethylene - Glycol Based Polyurethane Microgel, Macromol. Mater. Eng. **301**, 1158–1171 (2016). https://doi.org/10.1002/mame.201600186
11. D Brabazon, D.J Browne, A.J Carr, Experimental investigation of the transient and steady state rheological behaviour of Al-Si alloys in the mushy state, Materials Science and Engineering: A, **356**, 69–80 (2003). https://doi.org/10.1016/S0921-5093(03)00158-8

Green's Function and Wave Scattering in Inhomogeneous Anti-plane PEM Half-Plane

Tsviatko Rangelov and Petia Dineva

Abstract Half-plane Green's functions and wave scattering solutions in quadratically and exponentially graded in respect to depth piezoelectric half-plane with and without taking into consideration the surface elasticity properties are derived analytically. The surface elasticity model is considered in the frame of Gurtin and Murdoch (Arch. Ration Mach Anal 57:291–323, 1975). Anti-plane stress–strain state holds. The mechanical model considers materials with the same functional behavior of all material properties. The obtained solutions recover a wide class of Green's functions for homogeneous/graded, elastic/piezoelectric materials with/without surface elasticity properties.

Keywords Graded piezoelectric half-plane · Surface elasticity · Green's function · Free-field SH wave

1 Introduction

Two kinds of analytical solutions are essential in elastodynamic BEM models. For the whole plane problems they are fundamental solutions and plane-wave solutions. For the half-plane models they are Green's function and free-field wave motion solution. The key role played by them is to reduce a Boundary Value Problem (BVP)

T. Rangelov (✉)
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 8, 1113 Sofia, Bulgaria
e-mail: rangelov@math.bas.bg

P. Dineva
Institute of Mechanics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 4, 1113 Sofia, Bulgaria
e-mail: petia@imbm.bas.bg

formulated by the governing partial differential equation together with boundary and initial conditions into a system of boundary integral equations through the use of reciprocal theorems.

The aim of this work is to derive analytically 2D fundamental solutions, half-plane Green's functions and free-field wave motion solutions for a family of functionally graded Piezoelectric Materials (PEM) without and with surface elasticity effects. Recall that Green's function in the half-plane is a fundamental solution that satisfies additionally boundary conditions on the horizontal surface, which are: (a) classical traction-free boundary conditions; (b) non-classical boundary conditions accounting for the surface effect. There are restricted set of results concerning Green's function for homogeneous half-plane, see [7] and Green's function for graded materials, see [4], Chaps. 2 and 5 and references therein.

To the authors' knowledge there are no results for dynamic Green's function in the homogeneous or graded half-plane and free-field solution with non-classical boundary conditions for the aim of nanomechanics, see [2] and [3]. The derived here solutions recover additionally the fundamental solutions and Green's functions for (a) homogeneous PEM; (b) homogeneous and graded pure elastic isotropic materials.

2 Preliminaries

In a Cartesian coordinate system $Ox_1x_2x_3$ we consider a piezoelectric half-plane $\mathbb{R}^2_- = \{x = (x_1, x_2), x_2 < 0\}$ poled in x_3—direction and subjected to time-harmonic incident SH-wave with frequency ω. The only non-vanishing displacements are the anti-plane mechanical displacement $u_3(x, \omega)$ and the in-plane electrical displacements $D_1(x, \omega)$, $D_2(x, \omega)$, $x = (x_1, x_2)$. In all formulas below capital indexes $K, J, \ldots$ vary 3, 4, while small indexes $i, j, \ldots$ vary 1, 2, also comma denotes partial differentiation and the summation convention under repeating indexes is applied. Assuming quasi–static approximation of piezoelectricity, the field equations in the frequency domain in absence of body forces and free volume charges are given by the balance equation, see [1]

$$\sigma_{iJ,i} + \rho_{JK}\omega^2 u_K = 0, \tag{1}$$

where $u_K = (u_3, \phi)$ is the generalized displacement and σ_{iJ} is the generalized stress, $\sigma_{iJ} = C_{iJKl}u_{K,l} = C_{iJKl}s_{Kl}$, where

$$C_{i33l} = \begin{cases} c_{44}, i = l \\ 0, \quad i \neq l \end{cases}; \quad C_{i34l} = \begin{cases} e_{15}, i = l \\ 0 \quad i \neq l \end{cases}; \quad C_{i44l} = \begin{cases} -\varepsilon_{11}, i = l \\ 0, \quad i \neq l \end{cases}.$$

Here c_{44} is shear stiffness, e_{15} is dielectric permittivity and ε_{11} is piezoelectric permittivity. The strain–displacement and electric field–potential relations are $s_{i3} = u_{3,i}$, $E_i = -\phi_{,i}$ where s_{ij}, ϕ are the strain tensor and electric potential, the

generalized strain is denoted with $s_{Ji} = (u_{3,i}, E_i)$. In (1) it is used the notation $\rho_{JK} = \begin{cases} \rho, J = K = 3 \\ 0, J = 4 \text{ or } K = 4 \end{cases}$, where ρ is the mass density.

We further assume that the mass density and material parameters vary in the same manner with respect to x, by a function $h(x) \in C^2(\mathbb{R}^2_-)$ and $h(x) \geq 1$ such that $c_{44}(x) = c_{44}^0 h(x)$, $e_{15}(x) = e_{15}^0 h(x)$, $\varepsilon_{11}(x) = \varepsilon_{11}^0 h(x)$, $\rho(x) = \rho^0 h(x)$. What follows is an analytical derivation of 2D fundamental solutions and half-plane Green's functions for a family of functionally graded piezoelectric materials.

Fundamental solution $u^*(x, \xi, \omega)$ of Eq. (1) is solution of the equation

$$\sigma_{iJM,i}^*(x, \xi, \omega) + \rho_{JK}\omega^2 u_{KM}^*(x, \xi, \omega) = \delta_{JM}\delta(x, \xi), \quad x, \xi \in \mathbb{R}^2_-, \qquad (2)$$

where $\sigma_{iJM}^* = C_{iJKl} u_{KM,l}^*$, $x = (x_1, x_2)$ and $\xi = (\xi_1, \xi_2)$. Here δ_{JM} is Kroneker's symbol, $\delta(x, \xi)$ is Dirak's delta function. For simplicity if there is no misunderstanding we will omit the arguments of the functions.

In order to derive the fundamental solution $u^*(x, \xi, \omega)$ and then Green's function we first transform equation in (2) by a suitable change of functions to an equation with constant coefficients. This can be done if certain restrictions on function $h(x)$ are supposed. In a second step we apply Fourier transform, which allows the construction of a set of fundamental solutions depending on the roots of the characteristic equation of the obtained ordinary differential equations. Next, using both the inverse Fourier transform and the inverse change of functions, the fundamental solutions of Eq. (1) are obtained for specific types of graded piezoelectric material. Finally, applying partial Fourier transform with respect to x_1 and using the representation of the fundamental solution with special functions, see [6], we construct a correction term, i.e., function $w(x, \xi, \omega)$, such that $g(x, \xi, \omega) = u^*(x, \xi, \omega) + w(x, \xi, \omega)$ is a Green's function and $g(x, \xi, \omega)$ satisfies the boundary condition along the half-plane boundary.

In the first step we use a method that successfully was applied by [5] for elastic continua. The smooth transformation

$$u_K(x, \omega) = h^{-1/2}(x)U_K(x, \omega) \qquad (3)$$

leads to a wave equation with constant coefficients, having in mind that

$$\begin{aligned} u_{K,l} &= -\tfrac{1}{2}h^{-3/2}h_{,l}U_K + h^{-1/2}U_{K,l}, \\ \sigma_{iJ} &= C_{iJKl}^0 h u_{K,l} = C_{iJKl}^0[-(h^{1/2})_{,l}U_K + h^{1/2}U_{K,l}], \text{ and} \\ \sigma_{iJ,q} &= C_{iJKl}^0[-(h^{1/2})_{,lq}U_K - (h^{1/2})_{,l}U_{K,q} + (h^{1/2})_{,q}U_{K,l} + h^{1/2}U_{K,lq}], \end{aligned}$$

where $C_{iJKl}(x) = h(x)C_{iJKl}^0$. Since for $q = i = l$, the terms with the first derivatives of U_K vanish, Eq. (1) after dividing it by $h^{1/2}(x)$ takes the form

$$C_{iJKi}^0 U_{K,ii} + [\rho_{JK}^0 \omega^2 - C_{iJKi}^0 h^{-1/2}(h^{1/2})_{,ii}]U_K = 0, \qquad (4)$$

Correspondingly, using the property $h^{-1/2}(x)\delta(x,\xi) = h^{-1/2}(\xi)\delta(x,\xi)$ of the Dirac function, Eq. (2) with the transformation

$$u_{KM}^*(x,\xi,\omega) = h^{-1/2}(x)U_{KM}^*(x,\xi,\omega) \tag{5}$$

yields

$$C_{iJKi}^0 U_{KM,ii}^* + [\rho_{JK}^0\omega^2 - C_{iJKi}^0 h^{-1/2}(h^{1/2})_{,ii}]U_{KM}^* = h^{-1/2}(\xi)\delta_{JM}\delta(x,\xi). \tag{6}$$

Equations (4) and (6) are equations with constant coefficients, if the condition

$$C_{iJKi}^0 h^{-1/2}(h^{1/2})_{,ii} = p_{JK} = const, \quad x \in \mathbb{R}_-^2, \tag{7}$$

is fulfilled, what constitutes certain restrictions for the inhomogeneity function $h(x)$.

Let us consider two types of inhomogeneous piezoelectric materials for which condition (7) is fulfilled.

Case (q) For $p_{JK} = 0$—quadratic inhomogeneous function $h_q(x) = (bx_2 + 1)^2$, $b < 0, x \in \mathbb{R}_-^2$;

Case (e) For $p_{JK} = C_{2JK2}^0 a^2$—exponential inhomogeneous function $h_e(x) = e^{2ax_2}, a < 0, x \in \mathbb{R}_-^2$;

3 Fundamental Solution and Plane-Wave Solution

3.1 Fundamental Solution

The fundamental solution matrix is not unique, and we need at least one to use it for derivation of Green's function. In [1], Chap. 3, a fundamental solution matrix is obtained with Radon transform, but now for the aim of derivation of Green's function we need to obtain a fundamental solution with Fourier transform. Below we will consider both cases (q) and (e) simultaneously having in mind that: in case (q) $a = 0$; in case (e) $a \neq 0$.

After some simplifications from matrix equation (4) we obtain the following equations and relation for U_{KM}^*

$$\begin{aligned}
\Delta U_{33}^* + k^2 U_{33}^* &= \frac{1}{\gamma}h^{-1/2}(\xi)\delta(x,\xi), \quad U_{43}^* = U_{34}^* = \frac{e_{15}^0}{\varepsilon_{11}^0}U_{33}^*, \\
U_{44}^* &= \left(\frac{e_{15}^0}{\varepsilon_{11}^0}\right)^2 U_{33}^* + V_{44}^*(x,\xi), \quad \text{with } \Delta V_{44}^* - a^2 V_{44}^* = -\frac{h^{-1/2}(\xi)}{\varepsilon_{11}^0}\delta(x,\xi),
\end{aligned} \tag{8}$$

where $\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}$ is the Laplace operator, $\gamma = c_{44}^0 + \frac{(e_{15}^0)^2}{\varepsilon_{11}^0}$ and $k^2 = \frac{\rho^0\omega^2}{\gamma} - a^2$.

Equations in (8) are of Helmholtz type and let us consider the model equation

$$\Delta v(x, \xi) + k^2 v(x, \xi) = c(\xi)\delta(x, \xi)$$

where k^2 is a constant and $c(\xi) \neq 0$. Let us introduce the notation $r(x, \xi) = |x - \xi|$. There are the following 3 types of solutions with respect to k see [9]:

(−) if $k^2 < 0$, $v(x, \xi) = \frac{c(\xi)}{2\pi} K_0(|k|r(x, \xi)) = \frac{ic(\xi)}{4} H_0^{(1)}(i|k|r(x, \xi))$,

(0) if $k^2 = 0$, $v(x, \xi) = \frac{c(\xi)}{2\pi} \ln r(x, \xi)$,

(+) if $k^2 > 0$, $v(x, \xi) = -\frac{c(\xi)}{2\pi} K_0(-i|k|r(x, \xi)) = -\frac{ic(\xi)}{4} H_0^{(1)}(|k|r(x, \xi))$,

where K_0 is Kelvin function of 0-th order and $H_0^{(1)}$ is Hankel function of the first kind and 0-th order.

Now solving Helmholtz type equations in (8) and equation for V_{44}^* for different values of k^2 we obtain fundamental solution u^* via $\{U_{KM}^*\}$ in (5) as follows, where we use the notations

$$M^0 = \begin{pmatrix} 1 & \frac{e_{15}^0}{\varepsilon_{11}^0} \\ \frac{e_{15}^0}{\varepsilon_{11}^0} & \left(\frac{e_{15}^0}{\varepsilon_{11}^0}\right)^2 \end{pmatrix}, \quad N^0 = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{\varepsilon_{11}^0} \end{pmatrix}.$$

3.1.1 Case (q)

Here $k^2 = \frac{\rho^0 \omega^2}{\gamma} > 0$, i.e., $a = 0$, $h_q^{-1/2}(x)h_q^{-1/2}(\xi) = \frac{1}{bx_2+1}\frac{1}{b\xi_2+1} = P_q(x, \xi)$ and

$$u_q^*(x, \xi, \omega) = -P_q(x, \xi)\left[\frac{i}{4\gamma}M^0 H_0^{(1)}(k_q r(x, \xi)) + \frac{1}{2\pi}N^0 \ln r(x, \xi)\right]. \quad (9)$$

3.1.2 Case (e)

In this case $h_e^{-1/2}(x)h_e^{-1/2}(\xi) = e^{-a(x_2+\xi_2)} = P_e(x, \xi)$ and let us define $\omega_0 = \frac{\gamma}{\rho^0}|a|$, $a \neq 0$ and then with respect to ω we have the following cases for k^2 and corresponding fundamental solutions: For $k^2 < 0$, i.e., $\omega < \omega_0$, this is case (-) that corresponds to simple vibration behavior:

$$u_e^*(x, \xi, \omega) = -P_e(x, \xi)\left[\frac{i}{4\gamma}M^0 H_0^{(1)}(i|k_e|r(x, \xi)) - \frac{1}{4}N^0 H_0^{(1)}(i|a|r(x, \xi))\right]. \quad (10)$$

For $k^2 = 0$, i.e., $\omega = \omega_0$, this is case (0) that corresponds to quazi-static behavior;

$$u_e^*(x, \xi, \omega) = P_e(x, \xi)\left[\frac{1}{2\pi\gamma}M^0 \ln r(x, \xi) - \frac{1}{4}N^0 H_0^{(1)}(i|a|r(x, \xi))\right]. \quad (11)$$

For $k^2 > 0$, i.e., $\omega > \omega_0$, this is case (+) that corresponds to wave propagation behavior:

$$u_e^*(x, \xi, \omega) = -P_e(x, \xi) \left[\frac{i}{4\gamma} M^0 H_0^{(1)}(k_e r(x, \xi)) + \frac{1}{4} N^0 H_0^{(1)}(i|a|r(x, \xi)) \right]. \quad (12)$$

The fundamental solution of the equation of motion (1) for different types of the material gradient depends on the type and properties of the material gradient, on the reference material piezoelectric properties and on the frequency of the dynamic load.

3.2 Plane-Wave Solution

In order to find SH plane-wave solution of (1) in a quadratically and exponentially graded whole plane we will use the transformation (3), so that U_K^{in} satisfies Eq. (4) with constant coefficients and the incident generalized displacement solution is $u_K^{in} = h^{-1/2}(x)U_K^{in}$. Vector function $U^{in} = \begin{Bmatrix} U_3^{in} \\ U_4^{in} \end{Bmatrix}$ is obtained by the wave decomposition method, following Chap. 11 in [1], which is expressed as follows

$$U^{in} = \begin{pmatrix} 1 \\ \frac{e_{15}^0}{\varepsilon_{11}^0} \end{pmatrix} e^{ik(x_1\zeta_1 + x_2\zeta_2)} \quad (13)$$

where $\zeta_1 = \cos\theta$, $\zeta_2 = \sin\theta$ and $\theta \in (0, \pi/2]$ is the direction of the SH wave.

4 Green's Function and Free-Field Solution

4.1 Green's Function

Green's function $g(x, \xi, \omega)$ of Eq. (1) in the half-plane $\mathbb{R}_-^2$ is a fundamental solution that satisfies in addition prescribed traction boundary condition on $x_2 = 0$, i.e., $g(x, \xi, \omega)$ is defined as the solution of the problem

$$\begin{aligned} Lg &\equiv \Sigma_{iJM,i} + \rho_{JK}\omega^2 g_{KM} = \delta_{JM}\delta(x, \xi), \quad x, \xi \in \mathbb{R}_-^2 \\ T_{JM}^g &= -\delta_{J3}\mu^S g_{3M,11} \text{ on } x_2 = 0, \end{aligned} \quad (14)$$

where the corresponding to Green's function $g(x, \xi, \omega)$ stress is defined as $\Sigma_{iJM} = C_{iJKl}g_{KM,l}$, the corresponding to Green's function $g(x, \xi, \omega)$ traction is $T_{JM}^g = \Sigma_{iJM}n_i$ and $n = (0, 1)$ is outward normal vector to the boundary $x_2 = 0$. Note that boundary condition (14) concerns only the pure mechanical stress and traction, although $g(x, \xi, \omega)$ is time-harmonic Green's function for piezoelectric half-plane. Constant $\mu^S \neq 0$ if there is a surface effect on $x_2 = 0$, see [2, 3] and $\mu^S = 0$ if the classical traction-free boundary condition on $x_2 = 0$ is satisfied in the case of absence of surface elasticity effect. In the case of piezoelectricity with taking into account

the surface elasticity effects it is assumed: (a) existence of the infinitely thin layer along the half-plane free surface which is elastic isotropic with own surface elastic characteristics; (b) the thin free surface layer is electrically impermeable and inside it the electric field is ignored. It may be thought as a low-capacitance medium with a potential jump $\Delta\phi = \phi^+ - \phi^-$. The boundary condition for the normal components of the electrical displacement in this case is $D_n^S = 0$ along the free surface thin layer S. In order to derive Green's functions we will proceed as in [8]. Let us denote $\beta = \sqrt{\eta^2 - k^2}$ and use the following representation of the Hankel function and $\ln r$, see [6]:

$$
\begin{aligned}
H_0^{(1)}(kr) &= \tfrac{1}{i\pi} \int_{i\alpha-\infty}^{i\alpha+\infty} \tfrac{1}{\beta} e^{-\beta|x_2-\xi_2|} e^{i\eta(x_1+\xi_1)} d\eta, \quad x_2 < 0, \xi_2 < 0, \\
\ln r &= \tfrac{1}{2} \int_{-\infty}^{+\infty} \tfrac{1}{|\eta|} e^{-|\eta||x_2-\xi_2|} e^{i\eta(x_1+\xi_1)} d\eta, \quad x_2 < 0, \xi_2 < 0,
\end{aligned}
\tag{15}
$$

where $r = |x - \xi|$, $k = k_1 + ik_2$, $k_1 = \frac{\rho\omega^2}{\gamma} - a^2$, $k_2 > 0$, $-k_2 < \alpha < k_2$. Note that in the cases we consider, we can pass $k_2 \to 0$, i.e., $\alpha = 0$ in (15) and integrals are over $\mathbb{R}$.

The expression (15) is replaced in (9) for case (q) and in (10)—(12) for case (e).

Recall that for the fundamental solution it holds $Lu^*(x, \xi, \omega) = \delta(x, \xi)I_2$, I_2 is unit matrix in $\mathbb{R}^2$. As far as Green's function is presented by $g(x, \xi, \omega) = u^*(x, \xi, \omega) + w(x, \xi, \omega)$, we have to find 2×2 matrix function $w(x, \xi, \omega) = \{w_{IJ}\}$, $I, J = 3, 4$ such that

$$
\begin{aligned}
&Lw(x, \xi, \omega) = 0, \\
&T_{JM}^w - \delta_{J3}\mu^S w_{3M,11} = -\left(T_{JM}^{u^*} - \delta_{J3}\mu^S u_{3M,11}^*\right), \quad \text{for } x_2 = 0.
\end{aligned}
\tag{16}
$$

Note that for matrix function $z_{JM} = h^{-\frac{1}{2}}(x) Z_{JM}(x, \xi, \omega)$, the traction matrix on $x_2 = 0$ is defined as

$$
T_{JM}^z|_{x_2=0} = \sigma_{2JM}|_{x_2=0} = C_{2JK2}^0 \left(-\tfrac{1}{2}h^{-\frac{1}{2}}h'_{,2}Z_{KM} + h^{\frac{1}{2}}Z_{KM,2}\right)|_{x_2=0}.
$$

So, in the considered inhomogeneity cases we have

$$
\begin{aligned}
&\text{Case (q): } T_{JM}^z|_{x_2=0} = C_{2JK2}^0 \left[-bZ_{KM} + Z_{KM,2}\right]|_{x_2=0}, \\
&\text{Case (e): } T_{JM}^z|_{x_2=0} = C_{2JK2}^0 \left[-aZ_{KM} + Z_{KM,2}\right]|_{x_2=0}.
\end{aligned}
\tag{17}
$$

The matrix function $w(x, \xi, \omega)$ is derived separately for cases (q) and (e).

4.2 Quadratic Inhomogeneity—Case (q)

We are asking for matrix function $w(x, \xi, \omega)$ as an inverse Fourier transform in respect to η with components $w_{IJ} = \frac{1}{bx_2+1} W_{IJ}$ where

$$W_{IJ} = \frac{1}{b\xi_2 + 1} \int_{\mathbb{R}} \left(S_{IJ} \frac{1}{\beta} e^{\beta(x_2+\xi_2)} + D_{IJ} \frac{1}{|\eta|} e^{|\eta|(x_2+\xi_2)} \right) e^{i\eta(x_1-\xi_1)} d\eta, \qquad (18)$$

here $\beta^2 = \eta^2 - k^2$ and S_{IJ}, D_{IJ} depend on η, β, μ^S, γ, b, but does not depend on x_1, x_2.

The columns of the matrix function W_{IJ} should satisfies Eq. (4) with constant coefficients, then w_{IJ} satisfies Eq. (16). This holds under the conditions $S_{4J} = \frac{e_{15}^0}{\varepsilon_{11}^0} S_{3J}$, $D_{3J} = 0$, $J = 3, 4$. Independent functions S_{3J} and D_{4J}, $J = 3, 4$ are determined from the boundary conditions (16) on $x_2 = 0$. Applying (17) for w and u^* it is obtained finally the following expressions

$$S_{33} = \frac{1}{4\pi\gamma} \frac{\beta+b+\mu^S\eta^2}{\beta-b-\mu^S\eta^2}, S_{43} = \frac{e_{15}^0}{\varepsilon_{11}^0} S_{33}, S_{34} = \frac{e_{15}^0}{\varepsilon_{11}^0} \frac{1}{4\pi\gamma} \frac{\beta+b}{\beta-b}, S_{44} = \frac{e_{15}^0}{\varepsilon_{11}^0} S_{34},$$

$$D_{33} = D_{34} = 0, D_{34} = \frac{1}{2\pi\gamma} \frac{|\eta|+b}{|\eta|-b}, D_{44} = \frac{e_{15}^0}{\varepsilon_{11}^0} D_{34}.$$

4.3 Exponential Inhomogeneity—Case (e)

By the use of the fundamental solutions (10), (11), (12) and representations (15), we will find the corresponding matrix function $w(x, \xi, \omega)$ for different cases of k. Let us define together with $\beta = \sqrt{\eta^2 - k^2}$ also $\beta_k = \sqrt{\eta^2 + k^2}$ and $\beta_a = \sqrt{\eta^2 + a^2}$.

4.3.1   Simple Vibration Case $k^2 < 0$

We are asking for matrix function $w^-(x, \xi, \omega)$ as an inverse Fourier transform in respect to η with components $w_{IJ}^- = e^{-ax_2} W_{IJ}^-$, where

$$W_{IJ}^- = e^{-a\xi_2} \int_{\mathbb{R}} \left(S_{IJ}^- \frac{1}{\beta_k} e^{\beta_k(x_2+\xi_2)} + D_{IJ}^- \frac{1}{\beta_a} e^{\beta_a(x_2+\xi_2)} \right) e^{i\eta(x_1-\xi_1)} d\eta.$$

We find that Eq. (16) is satisfies for w_{IJ}^-, if

$$S_{33}^- = \frac{1}{4\pi\gamma} \frac{\beta_k+a-\mu^S\eta^2}{\beta_k-a+\mu^S\eta^2}, S_{43}^- = \frac{e_{15}^0}{\varepsilon_{11}^0} S_{33}^-, S_{34}^- = \frac{e_{15}^0}{\varepsilon_{11}^0} \frac{1}{4\pi\gamma} \frac{\beta_k+a}{\beta_k-a}, S_{44}^- = \frac{e_{15}^0}{\varepsilon_{11}^0} S_{34}^-,$$

$$D_{33}^- = D_{34}^- = 0, D_{34}^- = \frac{1}{2\pi\gamma} \frac{\beta_a+a}{\beta_a-a}, D_{44}^- = \frac{e_{15}^0}{\varepsilon_{11}^0} D_{34}^-.$$

4.3.2   Quasi-Static Case $k^2 = 0$

We are asking for matrix function $w^0(x, \xi, \omega)$ as an inverse Fourier transform in respect to η with components $w_{IJ}^0 = e^{-ax_2} W_{IJ}^0$, where

$$W_{IJ}^0 = e^{-a\xi_2} \int_{\mathbb{R}} \left(S_{IJ}^0 \frac{1}{|\eta|} e^{|\eta|(x_2+\xi_2)} + D_{IJ}^0 \frac{1}{\beta_a} e^{\beta_a(x_2+\xi_2)} \right) e^{i\eta(x_1+\xi_1)} d\eta.$$

We find that Eq. (16) is satisfied for w_{IJ}^-, if

$$S_{33}^0 = \frac{1}{4\pi\gamma} \frac{|\eta|+a-\mu^S\eta^2)}{|\eta|-a+\mu^S\eta^2)}, \ S_{43}^0 = \frac{e_{15}^0}{\varepsilon_{11}^0} S_{33}^0, \ S_{34}^0 = \frac{e_{15}^0}{\varepsilon_{11}^0} \frac{1}{4\pi\gamma} \frac{|\eta|+a}{|\eta|-a}, \ S_{44}^0 = \frac{e_{15}^0}{\varepsilon_{11}^0} S_{34}^0,$$
$$D_{33}^0 = D_{34}^0 = 0, \ D_{34}^0 = \frac{1}{2\pi\gamma} \frac{\beta_a+a}{\beta_a-a}, \ D_{44}^0 = \frac{e_{15}^0}{\varepsilon_{11}^0} D_{34}^0.$$

4.3.3 Wave Propagation Case $k^2 > 0$

We are asking for matrix function $w^+(x, \xi, \omega)$ as an inverse Fourier transform in respect to η with components $w_{IJ}^+ = e^{-ax_2} W_{IJ}^+$, where

$$W_{IJ}^+ = e^{-a\xi_2} \int_{\mathbb{R}} \left(S_{IJ}^+ \frac{1}{\beta} e^{\beta(x_2+\xi_2)} + D_{IJ}^+ \frac{1}{\beta_a} e^{\beta_a(x_2+\xi_2)} \right) e^{i\eta(x_1+\xi_1)} d\eta.$$

We find that Eq. (16) is satisfied for w_{IJ}^+, if

$$S_{33}^+ = \frac{1}{4\pi\gamma} \frac{\beta+a-\mu^S\eta^2}{\beta-a+\mu^S\eta^2}, \ S_{43}^+ = \frac{e_{15}^0}{\varepsilon_{11}^0} S_{33}^+, \ S_{34}^+ = \frac{e_{15}^0}{\varepsilon_{11}^0} \frac{1}{4\pi\gamma} \frac{\beta+a}{\beta-a}, \ S_{44}^+ = \frac{e_{15}^0}{\varepsilon_{11}^0} S_{34}^+,$$
$$D_{33}^+ = D_{34}^+ = 0, \ D_{34}^+ = \frac{1}{2\pi\gamma} \frac{\beta_a+a}{\beta_a-a}, \ D_{44}^+ = \frac{e_{15}^0}{\varepsilon_{11}^0} D_{34}^+.$$

4.4 Free–Field Wave Solution

The traction-free boundary conditions along the boundary of the inhomogeneous half-plane are expressed by the following formulas defined by the vector-valued function $z = h^{-1/2}(x)Z(x, \omega)$

$$\text{Case (q): } T_J^z|_{x_2=0} = C_{2JK2}^0 \left[-bZ_K + Z_{K,2}\right]|_{x_2=0}$$
$$\text{Case (e): } T_J^z|_{x_2=0} = C_{2JK2}^0 \left[-aZ_K + Z_{K,2}\right]|_{x_2=0}$$

Let us denote $z_K^{ff} = z_K^{in} + z_K^{sc}$, where z_K^{in} is incident SH wave generalized displacement and z_K^{sc} is the generalized displacement of SH wave scattering by the boundary $x_2 = 0$. For the boundary condition with the surface effect it holds $T_J^{zff} - \delta_{J3}\mu^S z_{3,11}^{ff}|_{x_2=0} = 0$. Note that the surface effect can present only for the mechanical displacement component z_3. As is mentioned in [3] there exists only scattering SH wave displacement in the case of incident SH-wave displacement, propagating in half-plane with surface effect along its boundary. We suppose that this is true also for the considered here quadratically and exponentially graded in depth piezoelectric half-plane.

We are asking the free–field wave solution in the form
$u_K^{ff} = h^{-1/2}(x)\left(A_1 U_K^{in} + A_2 U_K^{sc}\right)$, where U^{in} is defined in (13) and

$$U^{sc} = \begin{pmatrix} 1 \\ \frac{e_{15}^0}{\varepsilon_{11}^0} \end{pmatrix} e^{ik(x_1\zeta_1 - x_2\zeta_2)}.$$

Our aim is to determine constants A_1, A_2 such that the boundary condition (13) holds. Note that the amplitude of the incident wave is usually assumed to be unit. Both inhomogeneity cases will be considered separately.

4.4.1   Quadratic Inhomogeneity—Case (q)

Recall that in this case $h_q^{-1/2}(x) = \frac{1}{bx_2+1}$, $b \le 0$ and

$$\begin{pmatrix} u_3^{ff} \\ u_4^{ff} \end{pmatrix} = \frac{1}{bx_2+1}\begin{pmatrix} 1 \\ \frac{e_{15}^0}{\varepsilon_{11}^0} \end{pmatrix}\left[A_1^q e^{-ik(x_1\zeta_1 + x_2\zeta_2)} + A_2^q e^{-ik(x_1\zeta_1 - x_2\zeta_2)}\right].$$

After some simplifications we obtain that the boundary conditions $T_3^{u^{ff}} - \mu^S u_{3,11}^{ff}\big|_{x_2=0} = 0$, and $T_4^{u^{ff}}\big|_{x_2=0} = 0$ are satisfy if $A_2^q = \frac{\gamma(b+ik\zeta_2)-\mu^S\zeta_1^2}{\gamma(-b+ik\zeta_2)+\mu^S\zeta_1^2}A_1^q$.

4.4.2   Exponential Inhomogeneity—Case (e)

Recall that in this case $h_e^{-1/2}(x) = e^{-ax_2}$, and we will consider only the case of wave propagation, i.e., $k^2 = \frac{\rho^0\omega^2}{\gamma} - a^2 > 0$

$$\begin{pmatrix} u_3^{ff} \\ u_4^{ff} \end{pmatrix} = e^{-ax_2}\begin{pmatrix} 1 \\ \frac{e_{15}^0}{\varepsilon_{11}^0} \end{pmatrix}\left[A_1^e e^{-ik(x_1\zeta_1 + x_2\zeta_2)} + A_2^e e^{-ik(x_1\zeta_1 - x_2\zeta_2)}\right].$$

After some simplifications we obtain that boundary conditions $T_3^{u^{ff}} - \mu^S u_{3,11}^{ff}\big|_{x_2=0} = 0$, and $T_4^{u^{ff}}\big|_{x_2=0} = 0$ are satisfy if $A_2^e = \frac{\gamma(a+ik\zeta_2)-\mu^S k^2\zeta_1^2}{\gamma(-a+ik\zeta_2)+\mu^S k^2\zeta_1^2}A_1^e$.

5   Concluding Remarks

2D half-plane Green's functions and free-field SH wave solutions for a family of homogeneous and inhomogeneous in depth piezoelectric and elastic isotropic materials at macro and nano-level are derived analytically. They are the key components of the mesh-reducing BEM-based models with well-known advantages discussed in details in [1, 4] for solution of dynamic problems for inhomogeneous/homogeneous

and heterogeneous elastic and multifunctional materials with coupled properties. Different types of half-plane domains under dynamic loads with or without heterogeneities (layers, cracks, inclusions, relief) can be studied via BEM based on the derived here Green's functions and free-field SH wave solutions.

The above-described package of analytically derived dynamic Green's functions and free-field SH wave solutions presents an excellent basis for development of innovative dynamic BEM models in solid mechanics, fracture mechanics, mechanics of heterogeneous, and functional graded materials at macro and nano level.

References

1. Dineva, P., Gross, D., Müller, R., Rangelov, T.: Dynamic Fracture of Piezoelectric Materials. Solutions of Time-harmonic problems via BIEM. Solid Mechanics and its Applications, v. 212, Springer Int. Publ., Switzerland (2014)
2. Gurtin, M.E., Murdoch, A.I.: A continuum theory of elastic material surfaces. Arch. Ration. Mach. Anal. **57**, 291–323 (1975)
3. Gurtin, M.E., Murdoch, A.I.: Effect of surface stress on wave propagation in solids. J. Appl. Phys. **47**(10), 4416–4421 (1976)
4. Manolis, G.D., Dineva, P.S., Rangelov, T.V., Wuttke, F.: Seismic Wave Propagation in Non–Homogeneous Elastic Media by Boundary Elements. Solid Mechanics and its Applications, v. 240, Springer Int. Publ., Cham, Switzerland (2017)
5. Manolis, G.D., Shaw, R.: Green's function for a vector wave equation in mildly heterogeneous continuum. Wave Motion **24**, 59–83 (1996)
6. Noble, B.: Methods Based on the Wiener–Hopf Technique. Pergamon Press, New York (1958)
7. Rajapakse, R.K.N.D., Wang, Y.: Elastodynamic Green's functions of orthotropic half plane. ASCE, J. Eng. Mech. **117**(3), 588–604 (1991)
8. Rangelov, T.V., Manolis, G.D.: Point force and dipole solutions in the inhomogeneous half-plane under time-harmonic conditions. Mech. Res. Commun. **56**, 90–97 (2014)
9. Vladimirov, V.: Equations of Mathematical Physics. Marcel Dekker, Inc., New York (1971)

On the Well-Posedness for The Complex Ginzburg–Landau Equation on the Product Manifold $\mathbb{R}^d \times \mathbb{T}$

Elena Nikolova, Mirko Tarulli, and George Venkov

Abstract We prove both local and global well-posedness results for the solution to the complex Ginzburg–Landau equation with pure power mass-energy intercritical nonlinearity, posed on the product space $\mathbb{R}^d \times \mathbb{T}$.

1 Introduction

The aim of this paper is to display local and global well-posedness in energy space for the solution to the complex nonlinear Ginzburg–Landau equation (GLE) on $\mathbb{R}^d \times \mathbb{T}$, with $d \geq 1$ and $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ is the one-dimensional standard torus endowed with the flat metric. Namely, we consider the following Cauchy problem:

$$
\begin{cases}
u_t - (b_1 + ib_2)\Delta_{x,y}u = (c_1 + ic_2)|u|^\alpha u, & (t, x, y) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{T}, \\
u(0, x, y) = f(x, y) \in H^1(\mathbb{R}^d \times \mathbb{T}).
\end{cases}
\tag{1.1}
$$

E. Nikolova · M. Tarulli (✉) · G. Venkov
Department of Mathematical Analysis and Differential Equations, Technical University of Sofia, Kliment Ohridski Blvd. 8, 1000 Sofia, Bulgaria
e-mail: mta@tu-sofia.bg

E. Nikolova
e-mail: elenikolova@tu-sofia.bg

G. Venkov
e-mail: gvenkov@tu-sofia.bg

M. Tarulli
Institute of Mathematics and Informatics, Bulgarian Academy of Science, Acad. Georgi Bonchev Str., Block 8, 1113 Sofia, Bulgaria

Dipartimento di Matematica, Università di Pisa, Largo Bruno Pontecorvo 5 I, 56127 Pisa, Italy

Here $b_1 \geq 0$, b_2, c_1, c_2 are real parameters, $\Delta_{x,y} = \Delta_x + \partial_y^2$, and the nonlinearity parameter α satisfies the assumption

$$0 < \alpha < \alpha^*(d), \quad \alpha^*(d) = \begin{cases} \frac{4}{d-1} & \text{if } d \geq 2, \\ +\infty & \text{if } d = 1, \end{cases} \tag{1.2}$$

such that $\alpha^*(d)$ corresponds to the H^1-critical nonlinearity in $\mathbb{R}^{d+1}$. Notice that the Cauchy problem (1.1) inherits the properties of the gauge invariant nonlinear Schrödinger equation (NLS), but lacks its Hamiltonian structure. In view of that, we shall explore local and global existence, assuming the initial data are chosen in the space $H^1(\mathbb{R}^d \times \mathbb{T})$. Our first main result in this direction is given in the following:

Theorem 1 *Let $d \geq 1$ and α satisfying the assumption* (1.2) *be fixed. Then we have, for any initial data $f \in H^1(\mathbb{R}^d \times \mathbb{T})$, that the problem* (1.1) *has a unique local solution $u(t, x, y) \in C([0, T); H^1(\mathbb{R}^d \times \mathbb{T}))$, where $T = T(\|f\|_{H^1(\mathbb{R}^d \times \mathbb{T})}) > 0$.*

Our second result sheds light on the conservation of mass and energy for the GLE when it behaves like a Schrödinger-type equation (that is, when $b_1 = 0$). We have

Theorem 2 *The solution $u(t, x, y) \in C((-T, T); H^1(\mathbb{R}^d \times \mathbb{T}))$ of*

$$\begin{cases} iu_t + b_2 \Delta_{x,y} u = (ic_1 - c_2)|u|^\alpha u, & (t, x, y) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{T}, \\ u(0, x, y) = f(x, y) \in H^1(\mathbb{R}^d \times \mathbb{T}), \end{cases} \tag{1.3}$$

with $b_2 > 0$, $c_1, c_2 < 0$ satisfies the following inequalities

$$\|u(t)\|_{L^2(\mathbb{R}^d \times \mathbb{T})} \leq \|f\|_{L^2(\mathbb{R}^d \times \mathbb{T})} \tag{1.4}$$

and

$$E(u(t)) \leq E(u(0)), \tag{1.5}$$

for any $t > 0$, where

$$E(u(t)) = \frac{b_2}{2} \int_{\mathbb{R}^d \times \mathbb{T}} \left(|\nabla_x u(t, x, y)|^2 + |\partial_y u(t, x, y)|^2\right) dx dy \tag{1.6}$$

$$+ \frac{|c_2|}{\alpha + 2} \int_{\mathbb{R}^d \times \mathbb{T}} |u(t, x, y)|^{\alpha+2} dx dy.$$

Moreover $u(t, x, y)$ can be extended globally in time.

The Ginzburg–Landau type equations have an important role in several different models of mathematical physics, quantum mechanics as well as biomedicine. We refer to [7] for more detailed information on their applications. Moreover, we mainly address [4, 14], were all the mathematical analysis associated to the L^r-theory for the GLE of type (1.1) is developed in details. On the other hand, we recall [3] and [1] (and the included references), in which well-posedness and again blow up for the half

Ginzburg–Landau–Kuramoto equation with rough coefficients are examined. In the previous two papers, the nonlinearity considered is of the type $ic|u|^\alpha u$, with $c \in \mathbb{R}$, which transposes in our frame to selecting $c_1 \neq 0$ and $c_2 = 0$ in (1.3). A key tool to study the properties for solutions to the nonlinear problem of type of (1.1) are the Strichartz estimates (see for example [5, 6, 13] and references therein). Our paper has two different aims. First we show, the unconditional local well-posedness of the solution to (1.1) in the energy space by generalizing the classical L^2-based Strichartz estimates and the nonlinear extended Strichartz ones (see [2, 8, 10] and [11]) to the free propagator $e^{(b_1 + ib_2)t\Delta_{x,y}}$. We highlight that our method used in the various proofs guarantees to treat the lack of endpoint estimates, especially for $d = 1, 2$. In addition it permits to avoid, in the fixed point argument, the use of the fractional Leibniz rule to the term $|u|^\alpha u$ with respect to the x-variable. Then we exhibit that, unlike the classical NLS with pure power nonlinearity, the mass and energy are not conserved quantities for the GLE, but decay, adding a parabolic effect to the system. This method permits to bypass the technical difficulties associated with the classical regularization argument (see [4]).

2 Notations and Preliminaries

We indicate by $f \in L_x^{r_1} L_y^{r_2}$, for $1 \leq r_1, r_2 < \infty$, if

$$\|f\|_{L_x^{r_1} L_y^{r_2}}^{r_1} = \int_{\mathbb{R}^d} \|f(x, \cdot)\|_{L_y^{r_2}}^{r_1} dx < +\infty, \quad \text{where} \quad \|f(x, \cdot)\|_{L_y^{r_2}}^{r_2} = \int_{\mathbb{T}} |f(x, y)|^{r_2} dy,$$

with obvious modification for $r_1, r_2 = \infty$. Moreover, we introduce, for any $s, \gamma \in \mathbb{R}$, the Sobolev spaces

$$H_x^s = (1 - \Delta_x)^{-\frac{s}{2}} L^2(\mathbb{R}^d), \qquad H_y^\gamma = (1 - \partial_y^2)^{-\frac{\gamma}{2}} L^2(\mathbb{T})$$

and the anisotropic space

$$H_x^s H_y^\gamma = (1 - \Delta_x)^{-\frac{s}{2}} (1 - \partial_y^2)^{-\frac{\gamma}{2}} L_{x,y}^2.$$

Given any Banach space X we define, for any $1 \leq q < \infty$,

$$\|f\|_{L_t^q X} = \left(\int_{\mathbb{R}^+} \|f(x)\|_X^q dt \right)^{1/q},$$

with obvious modification for $q = \infty$. For its version local in time we embrace the notation $L_T^q X$ when $t \in (0, T)$, for $T > 0$. In addition, we shall need:

Proposition 1 *For every $\frac{1}{2} < \gamma < 1$, there exists $\widetilde{C} = C(\alpha, \gamma) > 0$ such that:*

$$\||u|u|^{\alpha}\|_{H_y^{\gamma}} \leq \widetilde{C}\|u\|_{H_y^{\gamma}}^{\alpha+1}. \tag{2.1}$$

Proof The proof can be found in [11], so we will omit it.   □

3   Strichartz Estimates for the Complex GLE

The Schrödinger group $e^{it\Delta_x}$ and the heat semigroup $e^{t\Delta_x}$ are both parts of the analytic semigroup defined via the Fourier transform as

$$e^{z\Delta_x}\varphi = \mathcal{F}^{-1}\left(e^{-4z\pi^2|\cdot|^2}\mathcal{F}\varphi\right),$$

for all complex z with Re $z \geq 0$. It follows that for $t > 0$, $e^{(b_1+ib_2)t\Delta_x}$ is a continuous map on the space of tempered distributions $\mathcal{S}'\left(\mathbb{R}^d\right)$ and that for each $\varphi \in \mathcal{S}'\left(\mathbb{R}^d\right)$, the map $t \mapsto e^{(b_1+ib_2)t\Delta_x}\varphi$ is continuous from $\mathbb{R}_+$ into $\mathcal{S}'\left(\mathbb{R}^d\right)$ and the operator $e^{(b_1+ib_2)t\Delta_x} = \mathscr{F}^{-1}e^{-(b_1+ib_2)t|\xi|^2}\mathscr{F}$ enjoys several decay estimates that are summarized in the following:

Proposition 2 *Let $2 \leq r \leq p \leq \infty$. Then, for any $t > 0$, $b_1 \geq 0$ and $b_2 \in \mathbb{R}$ we have*

$$\left\|e^{(b_1+ib_2)t\Delta_x}f\right\|_{L_x^p} \lesssim t^{-\frac{d}{2}\left(\frac{1}{r}-\frac{1}{p}\right)}\|f\|_{L_x^r}. \tag{3.1}$$

Proof The proof is based on an application of the Young and Hölder's inequalities and can be found in [4] (see also [12]).   □

As a consequence of the above result, we can obtain generalized Strichartz estimates (see [2] and [6]) on the geometric setting $\mathbb{R}^d \times \mathbb{T}$, adapted to the nonlinear problem (1.1) which are fundamental for the proof of well-posedness results. Namely, inspired by [11], one gets

Lemma 1 *For any $d \geq 3$, $\gamma \in (\frac{1}{2}, 1)$ and α satisfying the assumption (1.2) there exist real $s^* = \frac{d\alpha-4}{2\alpha} \in \left(0, \frac{1}{2}\right)$ and $s^* \leq s < \frac{1}{2}$, at least a pair $(q, r) \in \mathbb{R}_+^2$ and a corresponding pair $(\tilde{q}, \tilde{r}) \in \mathbb{R}_+^2$ that fulfill*

$$0 < \frac{1}{r}, \frac{1}{\tilde{r}} < \frac{1}{2}, \quad 0 < \frac{1}{q} < \frac{1}{2}, \quad 0 < \frac{1}{\tilde{q}} < 1, \tag{3.2}$$

$$\frac{1}{q} + \frac{1}{\tilde{q}} < 1, \quad \frac{d-2}{d} < \frac{r}{\tilde{r}} < \frac{d}{d-2}, \tag{3.3}$$

$$\frac{1}{q} + \frac{d}{r} < \frac{d}{2}, \quad \frac{1}{\tilde{q}} + \frac{d}{\tilde{r}} < \frac{d}{2}, \tag{3.4}$$

$$\frac{2}{q} + \frac{d}{r} = \frac{d}{2} - s, \quad \frac{2}{q} + \frac{d}{r} + \frac{2}{\tilde{q}} + \frac{d}{\tilde{r}} = d, \tag{3.5}$$

$$\frac{1}{\tilde{r}'} = \frac{\alpha+1}{r}, \quad \frac{1}{\tilde{q}'} \geq \frac{\alpha+1}{q}, \tag{3.6}$$

such that the solution to the GLE (1.1) *satisfies the estimate*

$$\|e^{(b_1+ib_2)t\Delta_{x,y}}f\|_{L_t^q L_x^r H_y^\gamma} + \left\| \int_0^t e^{(b_1+ib_2)(t-\tau)\Delta_{x,y}} F(\tau)d\tau \right\|_{L_t^q L_x^r H_y^\gamma}$$

$$\leq C\left(\|f\|_{H_x^s H_y^\gamma} + \|F\|_{L_t^{\tilde{q}'} L_x^{\tilde{r}'} H_y^\gamma} \right). \tag{3.7}$$

For $d = 1, 2$ *the above estimate remains valid, provided that we drop the conditions* (3.3).

Proof We introduce the operator

$$u(t, x, y) = e^{(b_1+ib_2)t\Delta_{x,y}} f(x, y) + \int_0^t e^{(b_1+ib_2)(t-\tau)\Delta_{x,y}} F(\tau, x, y)d\tau$$

and observe that

$$\partial_t u - (b_1 + ib_2)\Delta_x u + (b_1 + ib_2)\partial_y^2 u = F, \quad (t, x, y) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{T}, \tag{3.8}$$

with $u(0, x, y) = f(x, y)$. We have the following decomposition with respect to the orthonormal basis $\{e^{iky}\}$ of $L^2(\mathbb{T})$,

$$u(t, x, y) = \sum_k u_k(t, x)e^{iky}, \tag{3.9}$$

$$F(t, x, y) = \sum_k F_k(t, x)e^{iky}, \tag{3.10}$$

$$f(x, y) = \sum_k f_k(x)e^{iky}. \tag{3.11}$$

Notice that the functions $u_k(t, x)$, $F_k(t, x)$ and $f_k(x)$ satisfy the following Cauchy problem

$$\partial_t u_k - (b_1 + ib_2)\Delta_x u_k + (b_1 + ib_2)k^2 u_k = F_k, \quad (t, x) \in \mathbb{R}_+ \times \mathbb{R}^d, \quad (3.12)$$

with $u_k(0, x) = f_k(x)$. Let us prove the homogeneous estimates. Applying the classical homogeneous Strichartz estimates (see [6]) to (3.12) with $F_k(x, y) = 0$, we achieve

$$\|e^{(b_1+ib_2)t(\Delta_x - k^2)} f_k\|_{L_t^q L_x^r} \leq C \|f_k\|_{H_x^s},$$

with (q, r) and s as in (3.2) and (3.5). We see that the presence of the factor $e^{-(b_1+ib_2)tk^2}$ does not affect the above estimates, moreover the constants $C > 0$ does not depend on k. Summing the squares over k we achieve

$$\|e^{(b_1+ib_2)t(\Delta_x - k^2)} f_k\|_{\ell_k^2 L_t^q L_x^r} \leq C \|f_k\|_{\ell_k^2 H_x^s}.$$

Furthermore, because of $q, r \geq 2$, by the Minkowski inequality and Plancherel identity we infer

$$\|e^{(b_1+ib_2)t\Delta_{x,y}} f\|_{L_t^q L_x^r L_y^2} \leq C \|f\|_{H_x^s L_y^2},$$

Now it is sufficient to commute the equation (3.8) with the operator $(-\partial_y^2)^{\frac{\gamma}{2}}$, then the estimate (3.7) follows. The proof of inhomogeneous case in (3.7) is similar: by picking $f_k(x, y) = 0$ in (3.12) and using now the extended inhomogeneous Strichartz estimates obtained in [8] and [11] one arrives at

$$\left\| \int_0^t e^{(b_1+ib_2)(t-\tau)(\Delta_x - k^2)} F_k(s, \cdot) d\tau \right\|_{L_t^q L_x^r \ell_k^2} \leq C \|F_k\|_{L_t^{\tilde{q}'} L_x^{\tilde{r}'} \ell_k^2}, \quad (3.13)$$

with (q, r) and $(\tilde{q}, \tilde{r})$ satisfying the relations (3.2)–(3.5). To complete the proof of the lemma we want not only that the above estimates work together, but also in a way well fitted to handle the nonlinear problem (1.1). This is guaranteed by Proposition 1 in [10] (see also [8] and [11]), which states that the set of real numbers (q, r), $(\tilde{q}, \tilde{r})$ and s satisfying all the conditions (3.2)–(3.5) and (3.6) is not empty. $\square$

We look also at the full classical Strichartz estimates.

Proposition 1 *Let be $d \geq 1$. Indicate by D^k both ∇_x, and ∂_y. Then we have for $k = 0, 1$ and $\gamma \in \mathbb{R}$ the following estimates:*

$$\|D^k e^{(b_1+ib_2)t\Delta_{x,y}} f\|_{L_t^\ell L_x^p H_y^\gamma} + \left\| D^k \int_0^t e^{(b_1+ib_2)(t-\tau)\Delta_{x,y}} F(\tau) d\tau \right\|_{L_t^\ell L_x^p H_y^\gamma} \quad (3.14)$$
$$\leq C \big(\|D^k f\|_{L_x^2 H_y^\gamma} + \|D^k F\|_{L_t^{\tilde{\ell}'} L_x^{\tilde{p}'} H_y^\gamma} \big),$$

provided that

$$\frac{2}{\ell} + \frac{d}{p} = \frac{2}{\tilde{\ell}} + \frac{d}{\tilde{p}} = \frac{d}{2}, \quad \ell \geq 2, \quad (\ell, 2) \neq (2, 2). \quad (3.15)$$

Proof We can proceed as in the proof of Lemma 1, making use now of the classical Strichartz estimates available in [4, 12, 14], (see also [13] and references therein). □

4 Proof of Main Theorems

4.1 Unconditional Well-Posedness in $C([0, T); H^1(\mathbb{R}^d \times \mathbb{T}))$

First, we will need the next useful result.

Proposition 3 *Let $d \geq 1$ and α satisfying the assumption (1.2) be fixed. Then there exist pairs (q, r) and (ℓ, p) satisfying (3.15) and such that:*

$$\frac{1}{p'} = \frac{1}{p} + \frac{\alpha}{r}, \qquad \frac{1}{\ell'} > \frac{1}{\ell} + \frac{\alpha}{q}. \tag{4.1}$$

Proof The proof of the case $\alpha \geq 4/d$, with (q, r) any couple given by Lemma 1, can be found in [8]. For $0 < \alpha < 4/d$, instead, it is enough to select $(\frac{1}{q}, \frac{1}{r}) = (\frac{1}{\ell}, \frac{1}{p}) = (\frac{d\alpha}{4(\alpha+2)}, \frac{1}{\alpha+2})$ satisfying (3.15). □

We can proceed with:

4.1.1 Proof of Theorem 1.

We will perform a contraction argument, following [8] and [9]. Let be defined the integral operator associated to (1.1) for all $f \in H^1_{x,y}$,

$$\mathcal{T}_f(u) = e^{(b_1+ib_2)\Delta_{x,y}} f + (c_1 + ic_2) \int_0^t e^{(b_1+ib_2)(t-\tau)\Delta_{x,y}} (|u|^\alpha u)(\tau, x, y) \, d\tau. \tag{4.2}$$

One wants to show that there exist a $T = T(\|f\|_{H^1_{x,y}}) > 0$ and an unique $u(t, x, y) \in C((0, T); H^1_{x,y})$, satisfying the property $\mathcal{T}_f(u(t)) = u(t)$, for any $t \in (0, T)$. Let us set the following auxiliary norm

$$\|w\|_{\mathcal{Z}_T(q,r)} = \|w\|_{L^q_T L^r_x H^\gamma_y} + \|w\|_{L^\ell_T L^p_x L^2_y} + \|\nabla_x w\|_{L^\ell_T L^p_x L^2_y} + \|\partial_y w\|_{L^\ell_T L^p_x L^2_y}, \tag{4.3}$$

where (q, r) are as in Lemma 1 and $(\ell, p) = (\ell(q, r), p(q, r))$ as in Proposition 3. We split the proof in several different steps as follows.

Step One: For any $f \in H^1_{x,y}$, there exist $T = T(\|f\|_{H^1_{x,y}}) > 0$ and $R = R(\|f\|_{H^1_{x,y}}) > 0$ such that for any $T' < T$:

$$\mathcal{T}_f B_{\mathcal{Z}_{T'}(q,r)}(0, R)) \subset B_{\mathcal{Z}_{T'}(q,r)}(0, R).$$

We start with estimating the nonlinear term by using (3.6) in combination with (2.1), Sobolev embedding and the Hölder inequality. Namely

$$\|(c_1 + ic_2)u|u|^{\alpha}\|_{L_T^{\tilde{q}'} L_x^{\tilde{r}'} H_y^{\gamma}} \lesssim \|\|u\|_{H_y^{\gamma}}^{\alpha+1}\|_{L_T^{\tilde{q}'} L_x^{\tilde{r}'}} \lesssim T^{\frac{1}{\tilde{q}'} - \frac{\alpha+1}{q}} \|u\|_{L_T^q L_x^r H_y^{\gamma}}^{\alpha+1}, \qquad (4.4)$$

for some $C > 0$. Hence, the estimate (3.7) with $s + \gamma \leq 1$ allows us to conclude

$$\|\mathcal{T}_f u\|_{L_T^q L_x^r H_y^{\gamma}} \lesssim \|f\|_{H_x^s H_y^{\gamma}} + T^{\frac{1}{\tilde{q}'} - \frac{\alpha+1}{q}} \|u\|_{L_T^q L_x^r H_y^{\gamma}}^{\alpha+1}. \qquad (4.5)$$

An use of Propositions 1, 3, in conjunction with the bound (2.1), Hölder inequality, and Sobolev embedding yields the following:

$$\|\mathcal{T}_f u\|_{L_T^{\ell} L_x^p L_y^2} + \|\nabla_x \mathcal{T}_f u\|_{L_T^{\ell} L_x^p L_y^2} + \|\partial_y \mathcal{T}_f u\|_{L_T^{\ell} L_x^p L_y^2}$$

$$\lesssim \sum_{k=0,1} \left(\|D^k f\|_{L_x^2 L_y^2} + \|D^k(u|u|^{\alpha})\|_{L_T^{\ell'} L_x^{p'} L_y^2} \right)$$

$$\lesssim \sum_{k=0,1} \left(\|D^k f\|_{L_x^2 L_y^2} + \left\| \|D^k u(t,x)\|_{L_y^2} \|u(t,x)\|_{L_y^{\infty}}^{\alpha} \right\|_{L_T^{\ell'} L_x^{p'}} \right)$$

$$\lesssim \sum_{k=0,1} \left(\|D^k f\|_{L_x^2 L_y^2} + \left\| \|D^k u(t,x)\|_{L_x^p L_y^2} \|u(t,x)\|_{L_x^r H_y^{\gamma}}^{\alpha} \right\|_{L_T^{\ell'}} \right)$$

$$\lesssim \sum_{k=0,1} \left(\|D^k f\|_{L_x^2 L_y^2} + T^{\frac{1}{\ell'} - \frac{1}{\ell} - \frac{\alpha}{q}} \|D^k u(t,x)\|_{L_T^{\ell} L_x^p L_y^2} \|u(t,x)\|_{L_T^q L_x^r H_y^{\gamma}}^{\alpha} \right). \qquad (4.6)$$

Combining the above result with (4.5) and the natural embedding $H_x^1 \subseteq H_x^s H_y^{\gamma}$ we get

$$\|\mathcal{T}_f u\|_{\mathcal{Z}_T(q,r)} \lesssim \|f\|_{H_{x,y}^1} + T^{\frac{1}{\ell'} - \frac{1}{\ell} - \frac{\alpha}{q}} \|u\|_{\mathcal{Z}_T(q,r)} \|u\|_{L_T^q L_x^r H_y^{\gamma}}^{\alpha}, \qquad (4.7)$$

that concludes the proof of the step.

Step Two: Let T, $R > 0$ be as in the previous step, then there exists $\overline{T} = \overline{T}(\|f\|_{H_{x,y}^1}) < T$ such that $\mathcal{T}_f$ is a contraction on $B_{\mathcal{Z}_{\overline{T}}(q,r)}(0, R)$, equipped with the norm $\|.\|_{L_T^q L_x^r L_y^2}$. For any $v_1, v_2 \in B_{L_T^q L_x^r H_y^{\gamma}}(0, R)$, by estimate (3.7), Minkowski, Hölder and Sobolev inequalities, we arrive to

$$\|\mathcal{T}_f v_1 - \mathcal{T}_f v_2\|_{L_T^q L_x^r L_y^2} \leq C \|v_1|v_1|^{\alpha} - v_2|v_2|^{\alpha}\|_{L_T^{\tilde{q}'} L_x^{\tilde{r}'} L_y^2}$$

$$\leq C \left\| \|v_1 - v_2\|_{L_y^2} (\|v_1\|_{L_y^{\infty}}^{\alpha} + \|v_2\|_{L_y^{\infty}}^{\alpha}) \right\|_{L_T^{\tilde{q}'} L_x^{\tilde{r}'}}$$

$$\leq CT^{\frac{1}{\tilde{q}'} - \frac{\alpha+1}{q}} \left(\|v_1\|_{L_T^q L_x^r H_y^{\gamma}}^{\alpha} + \|v_2\|_{L_T^q L_x^r H_y^{\gamma}}^{\alpha} \right) \|v_1 - v_2\|_{L_T^q L_x^r L_y^2}.$$

This is enough to complete this step.

Step Three: The solution exists and is unique in $\mathcal{Z}_{\overline{T}}(q, r)$, where $\overline{T}$ is as in the former step.

We are in position to show existence and uniqueness of the solution by applying the contraction principle to the map $\mathcal{T}_f$ defined on the complete metric space $B_{\mathcal{Z}_{\overline{T}}(q,r)}(0, R)$, equipped with the topology induced by $\|.\|_{L_{\overline{T}}^q L_x^r L_y^2}$.

Step Four: Regularity of the solution: $u(t, x) \in C((0, T); H_{x,y}^1)$.

Arguing as in the proof of (4.6) with $(\ell, p) = (\infty, 2)$ provides $u(t) \in C((0, T); H_{x,y}^1)$.

Step Five: Assuming that $u_1, u_2 \in C((0, T); H_{x,y}^1)$ are fixed points of $\mathcal{T}_f$, then $u_1 = u_2$.

The technique utilized here is a generalization of the one presented in [9–11] (and references therein). If u_1, u_2 are fixed points of $\mathcal{T}_f$ then $\mathcal{T}_f(u_1 - u_2) = u_1 - u_2$. By (3.14), we get by Sobolev embedding and the property $2 < \frac{\alpha p}{p-2} < \frac{2d}{d-1}$, true for any $d \geq 1$,

$$\|\mathcal{T}_f(u_1 - u_2)\|_{L_T^\ell L_x^p L_y^2} \lesssim \||u_1|^\alpha u_1 - |u_2|^\alpha u_2\|_{L_T^{\ell'} L_x^{p'} L_y^2}$$

$$\lesssim \|u_1 - u_2\|_{L_T^\ell L_x^p L_y^2} \left(\|u_1\|^\alpha_{L_T^{\frac{\alpha \ell}{\ell-2}} L_x^{\frac{\alpha p}{p-2}} L_y^\infty} + \|u_2\|^\alpha_{L_T^{\frac{\alpha \ell}{\ell-2}} L_x^{\frac{\alpha p}{p-2}} L_y^\infty} \right)$$

$$\|u_1 - u_2\|_{L_T^\ell L_x^p L_y^2} T^{\frac{\ell-2}{\ell}} \left(\|u_1\|^\alpha_{L_T^\infty L_x^{\frac{\alpha p}{p-2}} H_y^\gamma} + \|u_2\|^\alpha_{L_T^\infty L_x^{\frac{\alpha p}{p-2}} H_y^\gamma} \right)$$

$$\lesssim \|u_1 - u_2\|_{L_T^\ell L_x^p L_y^2} T^{\frac{\ell-2}{\ell}} \left(\|u_1\|^\alpha_{L_T^\infty H_{x,y}^1} + \|u_2\|^\alpha_{L_T^\infty H_{x,y}^1} \right), \tag{4.8}$$

for T suitably small. A continuity argument completes the proof of the final step. $\square$

4.2 Decay of Mass and Energy

In this setting, for each $\varphi \in \mathcal{S}'(\mathbb{R}^d)$, the map $t \mapsto e^{ib_2 t \Delta_x} \varphi$ is continuous from $\mathbb{R}$ into $\mathcal{S}'(\mathbb{R}^d)$ and by the above arguments we know now that $u(t, x, y) \in C((-T, T); H^1(\mathbb{R}^d \times \mathbb{T}))$ because $b_1 = 0$ (see [11]). We are in position to give the proof of the second main result.

4.2.1 Proof of Theorem 2

Throughout this section we shall denote the nonlinearity as $h(u) = (|c_2| - i|c_1|)|u|^\alpha u$. We follow the technique presented in [10] that relies entirely on Duhamel's representation of the solution to (1.3)

$$u(t) = U(t)f - i \int_0^t U(t-s)h(u)\,ds, \tag{4.9}$$

where $U(t) = e^{itb_2\Delta}$ is a Schrödinger-type operator. We have

$$\text{Im}\,\langle h(u),\, u\rangle = -\,\text{Im}\,i\langle |c_1||u|^\alpha u,\, u\rangle = -|c_1|\langle |u|^\alpha u,\, u\rangle. \tag{4.10}$$

Thanks to the Eq. (1.3) we obtain

$$\text{Im}\,\langle h(u),\, b_2\Delta_{x,y}u\rangle = \text{Re}\,\langle (|c_2| - i|c_1|)|u|^\alpha u,\, \partial_t u\rangle$$
$$= \frac{1}{2}\frac{|c_2|}{\alpha+2}\partial_t \int_{\mathbb{R}^d\times\mathbb{T}} |u|^{\alpha+2}dxdy - |c_1|\,\text{Re}\,i\langle |u|^\alpha u,\, \partial_t u\rangle. \tag{4.11}$$

We apply the operator $U(-t)$ to u, $\nabla_x u$ and $\partial_y u$. Because of its unitary property, we get for $k = 0, 1,$[1]

$$\left\| D^k u(t) \right\|_{L^2_{x,y}}^2 = \left\| U(-t)D^k u(t) \right\|_{L^2_{x,y}}^2 = \left\| D^k f - i\int_0^t U(-s)D^k h(u(s))\,ds \right\|_{L^2_{x,y}}^2$$
$$= \left\| D^k f \right\|_{L^2_{x,y}}^2 + 2\,\text{Im}\int_0^t \langle D^k h(u(s)),\, U(s)D^k f\rangle\,ds + \left\| \int_0^t U(-s)D^k h(u(s))\,ds \right\|_{L^2_{x,y}}^2. \tag{4.12}$$

A differentiation and then a further integration of the last term w.r.t. the time variable yield

$$\left\| \int_0^t U(-s)D^k h(u(s))\,ds \right\|_{L^2_{x,y}}^2 = 2\,\text{Im}\int_0^t \langle D^k h(u(s)),\, -i\int_0^s U(s-\sigma)D^k h(u(\sigma))d\sigma\rangle\,ds. \tag{4.13}$$

Plugging (4.13) into (4.12) we arrive to

$$\left\| D^k u(t) \right\|_{L^2_{x,y}}^2 = \left\| D^k f \right\|_{L^2_{x,y}}^2 + 2\,\text{Im}\int_0^t \langle D^k h(u(s)),\, D^k u\rangle\,ds. \tag{4.14}$$

Finally, using property (4.10) we obtain, for $k = 0$, the mass

$$\| u(t) \|_{L^2_{x,y}}^2 = \| f \|_{L^2_{x,y}}^2 - 2|c_1|\int_0^t \langle |u|^\alpha u,\, u\rangle\,ds, \tag{4.15}$$

---

[1] We underline here that the inner products are understood, for $k = 0, 1,$ in

$$(L^\infty_t H^1_{x,y} \cap L^\ell_t H^{k,\alpha}_x L^2_y \cap L^\ell_t L^\alpha_x H^k_y) \times (L^1_t H^1_{x,y} + L^{\ell'}_t H^{k,\alpha/(\alpha-1)}_x L^2_y + L^{\ell'} L^{\alpha/(\alpha-1)} H^k_y),$$

with $\ell = 4\alpha/(d(\alpha-2))$.

which immediately leads to (1.4). From (4.11), for $k = 1$, we get instead

$$E(u(t)) = E(u(0)) + |c_1| \operatorname{Re} i \int_0^t \langle (|u|^\alpha u)(s), \, \partial_s u(s) \rangle \, ds, \tag{4.16}$$

with $E(u(t))$ as in (1.6). By inserting (1.3) into (4.16) we achieve

$$E(u) - E(u(0)) = |c_1| \int_0^t \operatorname{Re} \langle |u|^\alpha u, \, -i\partial_s u \rangle \, ds$$

$$= |b_2 c_1| \int_0^t \operatorname{Re} \langle |u|^\alpha u, \, \Delta_{x,y} u \rangle \, ds - |c_1 c_2| \int_0^t \operatorname{Re} \langle |u|^\alpha u, \, |u|^\alpha u \rangle \, ds. \tag{4.17}$$

Then it follows that

$$E(u(t)) - E(u(0)) \le |b_2 c_1| \int_0^t \operatorname{Re} \langle |u|^\alpha u, \, \Delta_{x,y} u \rangle \, ds$$

$$= -|b_2 c_1| \int_0^t \operatorname{Re} \langle |u|^\alpha (\nabla_x, \partial_y) u + u(\nabla_x, \partial_y)|u|^\alpha), \, (\nabla_x, \partial_y) u \rangle \, ds$$

$$= -|b_2 c_1| \int_0^t \int_{\mathbb{R}^d \times \mathbb{T}} |u|^\alpha \left| (\nabla_x, \partial_y) u \right|^2 + \alpha |u|^{\alpha-2} (\operatorname{Re} (u(\nabla_x, \partial_y))^2 \overline{u}) \, dx \, dy \, ds \le 0. \tag{4.18}$$

This shows that the energy is a decreasing function w.r.t the time variable and the proof of the theorem is complete. $\qquad\square$

References

1. L. Forcella, K. Fujiwara, V. Georgiev, T. Ozawa, Local Well-posedness and Blow-up for the Half Ginzburg-Landau-Kuramoto equation with rough coefficients and potential, Discrete Continuous Dynamical Systems, 39, pp. 2661–2678 (2018).
2. D. Foschi, Inhomogeneous Strichartz estimates, J. Hyper. Differential Equations, 02(1), pp. 1–24 (2005).
3. K. Fujiwara, V. Georgiev, T. Ozawa, Blow-up for self-interacting fractional Ginzburg-Landau equation, Dyn. Partial Differ. Equ. 15(3), pp. 175–182 (2018).
4. J. Ginibre and G. Velo, The Cauchy problem in local spaces for the complex Ginzburg-Landau equation (II) Contraction Methods, Commun. Math. Phys., 187, pp. 45–79, (1997).
5. S. Gustafson and D. Roxanas, Global, decaying solutions of a focusing energy-critical heat equation in $\mathbb{R}^4$, J. Differential Equations, 264(9), pp. 5894–5927 (2018).
6. M. Keel and T. Tao, Endpoint Strichartz estimates, Amer. J. Math., 120, pp. 955–980 (1998).
7. Y. Kuramoto, Chemical Oscillations, Waves, and Turbulence, Dover books on chemistry, xiv+323 (2003).
8. E. Nikolova, M. Tarulli, and G, Venkov, On the extended Strichartz estimates for the nonlinear heat equation, AIP Conference Proceedings, 2172, 030015 (2019).
9. E. Nikolova, M. Tarulli, and G, Venkov, On the Cauchy Problem For The Nonlinear Heat Equation, AIP Conference Proceedings, 2159, 030025 (2019).

10. E. Nikolova, M. Tarulli, and G, Venkov, Unconditional well-posedness in the energy space for the Ginzburg-Landau equation, AIP Conference Proceedings 2159, 030024 (2019).
11. N. Tzvetkov and N. Visciglia, Well-posedness and scattering for NLS on $\mathbb{R}^d \times \mathbb{T}$ in the energy space, Rev. Mat. Iberoamericana, 32(4), pp. 1163–1188 (2016).
12. B. Wang, Z. Huo, C. Hao, and Z. Guo, Harmonic analysis method for nonlinear evolution equations I, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, (2011).
13. Z. Zhai, Strichartz type estimates for fractional heat equations, J. Math. Anal. Appl., 356, pp. 642–658 (2009).
14. B. X. Wang, Exponential Besov spaces and their applications to certain evolution equations with dissipation, Commun. Pure and Appl. Anal., 3, pp. 883–919 (2004).

Exact Traveling Wave Solutions of the Generalized Rosenau–Kawahara-RLW Equation via Simple Equations Method

Elena V. Nikolova and Mila Chilikova-Lubomirova

Abstract We apply the Simple Equations Method (SEsM) for obtaining exact solutions of a non-linear partial differential equation (PDE) from the kind of the generalized Rosenau–Kawahara-RLW equation. The elliptic equation of Jacobi and the elliptic equation of Weierstrass are used as simple equations. Various exact solutions of the studied equation are obtained. These solutions are expressed by a special function V, which takes various forms depending on the used simple equation.

Keywords Exact traveling wave solutions · Rosenau–Kawahara-RLW equation · Simple equations method

1 Introduction

The complex dynamics of many processes in nature and human society is often presented by non-linear PDEs. Various analytical and numerical methods can be applied to the analysis of such models. Finding exact analytical solutions to these models is a powerful and effective tool for a better understanding of the space-temporal dynamics of these systems. Many years ago, the methods for obtaining exact solutions of non-linear PDEs were appropriate only for the case of integrable non-linear equations, as the method of inverse scattering transform or the method of Hirota [1–3]. Recently, other methods for obtaining exact special solutions of non-integrable non-linear PDEs have been developed (for examples, see [4–6]). Another famous method in this branch is the Method of Simplest Equation (MSE) developed

E. V. Nikolova (✉) · M. Chilikova-Lubomirova
Institute of Mechanics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Block 4, 1113, Sofia, Bulgaria
e-mail: elena@imbm.bas.bg

M. Chilikova-Lubomirova
e-mail: milachl@imbm.bas.bg

Institute for Climate, Atmosphere and Water Research, Bulgarian Academy of Sciences, 66 Tsarigradsko Shose Blvd., Sofia, Bulgaria

by Kudryashov [7]. Modification of this method is made in [8, 9], as it is known as Modified Method of Simplest Equation (MMSE). The MMSE follows the concept of the balance equation. It is based on the determination of the kind of the simplest equation and truncation of series constructed from solutions of the simplest equation by means of application of a balance equation. This methodology was successfully applied for obtaining exact traveling wave solutions of various non-linear evolution equations [10, 11]. In the last few years, MMSE has been extended, as the last version is called Simple Equations Methodology (SesM) (see [12, 13]). SEsM is based on the possibility of the use of more than one simple equation. For this case, the used simple equations are more simple than the solved non-linear PDE, but these simple equations in fact can be quite complicated. The SEsM and its particular cases were successfully applied in [14–19].

Below, we shall apply the SesM for obtaining exact analytical solutions of the generalized Rosenau–Kawahara-RLW equation. A short description of the studied equation and the basic algorithm of SesM is presented in Sect. 2. In Sect. 3, various analytical solutions of one variant of the generalized Rosenau–Kawahara-RLW are derived. Conclusions based on the obtained results are made in Sect. 4.

2 Problem Formulation and Methodology

In this study, we consider the generalized Rosenau–Kawahara-RLW equation, presented in the form [20]:

$$u_t + \alpha u_x + \beta u^p u_x - \gamma u_{xxt} + \epsilon u_{xxx} + \lambda u_{xxxxt} + \mu u_{xxxxx} = 0. \tag{1}$$

Here, $u(x, t)$ is the non-linear wave profile, where x and t are spatial and temporal variables, respectively. Also, $\alpha, \beta, \gamma, \epsilon, \lambda, \mu$ are arbitrary parameters, and p is a positive integer presenting power law non-linearity. Equation (1) is the generalization of three basic models available in the literature depending on the numerical values of the model parameters. For the case when $\epsilon = \lambda = \mu = 0$, Eq. (1) is reduced to the general regularized long-wave (RLW) equation [21]. When $\gamma = \epsilon = \mu = 0$, Eq. (1) is reduced to the general Rosenau equation [22]. When $\gamma = \lambda = 0$, it gives rise to the general Kawahara equation [23]. All of these equations have extensive applications in many scientific areas such as fluid mechanics, electro magnetics, plasma, and non-linear optics.

Here, we shall search for an analytical solution of Eq. (1) applying the SEsM. The SEsM can be used for obtaining analytical solutions of non-linear PDEs of the kind

$$\Phi(u(x, t), \ldots\ldots\ldots) = 0 \tag{2}$$

where the left-hand side of Eq. (2) is a relationship containing the function $u(x, t)$ and some of its derivatives. The algorithm of SEsM includes the following four steps [19]:

(1) The transformation

$$u(x, ..., t) = Tr(F_1(x, ..., t), F_2(x, ..., t), ...F_N(x, ..., t)) \tag{3}$$

is made, where $Tr(F_1, F_2, ..., F_N(x, ..., t))$ is a composite function of other functions $F_i\ i = 1...N$. $F_1(x, ..., t), F_2(x, ..., t), ..., F_N(x, ..., t)$ are functions of several spatial variables, as well as of time. The transformations Tr have two goals: (1) They can remove some non-linearities if possible (an example is the Hopf–Cole transformation, which leads to the linearization of the Burgers equation); (2) They can transform the non-linearity of the solved differential equations to a more treatable kind of non-linearity (e.g., to polynomial non-linearity). In many particular cases, one may skip this step (then we have just $u(x, ..., t) = F(x, ..., t)$), but in numerous cases this step is necessary for obtaining a solution of the studied non-linear PDE. The substitution of Eq. (3) in Eq. (2) leads to a non-linear PDE for the function $F(x, ..., t)$. In many cases, the general form of the transformation $Tr(F)$ is not known.

(2) This step is based on the use of composite functions. In this step, the functions $F_1(x, ..., t), F_2(x, ..., t), ...$ are chosen as composite functions of the functions $f_{i1}, ..., f_{iN}, ...$, which are solutions of simpler differential equations. There are two possibilities: (1) The construction relationship for the composite function is not fixed. Then, the Faa di Bruno relationship for the derivatives of a composite function is used; (2) The construction relationship for the composite function is fixed. For example, for the case of one solved equation and one function F, the construction relationship can be chosen to be

$$F = \hat{\alpha} + \sum_{i_1=1}^{N} \hat{\beta}_i f_{i_1} + \sum_{i_1=1}^{N} \sum_{i_2=1}^{N} \hat{\gamma}_{i_1} f_{i_1} f_{i_2} + ... + \sum_{i_1=1}^{N} ..... \sum_{i_N=1}^{N} \hat{\sigma}_{i_1...i_n} f_{i_1} .... f_{i_N}. \tag{4}$$

Then one can directly calculate the corresponding derivatives from the solved differential equation.

(3) In this step, the simple equations for the functions $f_{i_1}, ..., f_{i_N}$ must be selected. In addition, in accordance with the hypothesis of Point (1) of Step 2, the relationship between the composite functions $F_1(x, ..., t), F_2(x, ..., t), ...$, and the functions $f_{i_1}, ..., f_{i_N}$ must be fixed. The fixation of the simple equations and the fixation of the relationships for the composite functions are connected. The fixations transform the left-hand sides of Eq. (2). The result of this transformation can be functions that are the sum of terms. Each of these terms contains some function multiplied by a coefficient. This coefficient is a relationship containing some of the parameters of the solved equations and some of the parameters of the solutions of the simple equations used. The fixation mentioned above is performed by a

balance procedure that ensures that the relationships for the coefficients contain more than one term. This balance procedure leads to one or more additional relationships among the parameters of the solved equation and parameters of the solutions of the simple equations used. These additional relationships are known as balance equations.

(4) A non-trivial solution of Eq. (2) is obtained if all coefficients mentioned in Step 3 are set to zero. This condition usually leads to a system of non-linear algebraic equations. The unknown variables in these equations are the coefficients of the solved non-linear differential equation and the coefficients of the solutions of the simple equations. Any non-trivial solution of this algebraic system leads to a solution of the studied non-linear PDE.

Below, we shall apply the above-given methodology to obtain exact solutions of Eq. (1). We shall consider u as a composite function of one function of one variable $\xi = \kappa x + \omega t$ (κ and ω are parameters), i.e.,

$$u(\xi) = f[g(\xi)]. \tag{5}$$

Let us assume that f is a polynomial of g. Then

$$f = \sum_{i}^{n} a_i g(\xi)^i. \tag{6}$$

In this way, Eq. (5) becomes

$$u = \sum_{i}^{n} a_i g(\xi)^i, \tag{7}$$

where a_i, $i = 0, ..., n$ are parameters, and n shall be determined by means of a balance equation. We use the following simple equation:

$$\left(\frac{d^k g}{d\xi^k}\right)^l = \sum_{j=0}^{m} b_j g^j, \tag{8}$$

where k is the order of derivative of g, l is the degree of derivative in the defining ODE, and m is the highest degree of the polynomial of g in the defining ODE. We shall present the solutions of Eq. (5) by the special function $V(\xi, k, l, m)$, which has interesting properties. This function can be trigonometric, hyperbolic, elliptic function of Jacobi, etc. For our study, we choose one specific case of the function V, the function $V_{b_0, b_1, ..., b_m}(\xi; 1, 2, m)$, which is a solution of the simple equation:

$$\left(\frac{dg}{d\xi}\right)^2 = \sum_{j=0}^{m} b_j g^j, \tag{9}$$

where b_j, $j = 0, ..., m$ are parameters, and m shall be determined by means of a balance equation.

Below, we shall present several examples of the application of V depending on the numerical value of m.

3 Exact Solutions of the Generalized Rosenau–Kawahara-RLW Equation

Following the above-given algorithm, we skip Step 1 of SEsM (no additional transformation of non-linearity). In Step 2, we consider u as a composite function of one variable $\xi = \kappa x + \omega t$ (see Eq. 5). Then Eq. (1) is reduced to

$$(\omega + \alpha\kappa)\frac{du}{d\xi} + \beta\kappa u^p \frac{du}{d\xi} - \gamma\kappa^2\omega\frac{d^3u}{d\xi^3} + \epsilon\kappa^3\frac{d^3u}{d\xi^3} + (\lambda\omega\kappa^4 - \mu\kappa^5)\frac{d^5u}{d\xi^5} = 0. \quad (10)$$

In Step 3 of SEsM, we have to select the equation for $u(f)$ (the relationship for the composite function) and the equation for $g(\xi)$ (the simple equation). We assume that the expression for u is of kind (7). In addition, the simple equation g is assumed to be of kind (9). The substitution of Eqs. (7) and (9) in Eq. (10) leads to a polynomial of the function g. In order to obtain the system of non-linear algebraic equations, we have to write balance equations for these powers, i.e., in this case we have to balance the largest powers. This procedure leads us to the balance equation

$$n = \frac{2m - 4}{p}. \quad (11)$$

We fix $p = 2$. Then Eq. (1) may have solutions of the kind

$$u(x, t) = \sum_{i=0}^{m-2} a_i g(\xi)^i, \quad \xi = \kappa x + \omega t \quad (12)$$

and the function g is a solution of the simple equation

$$\left(\frac{dg}{d\xi}\right)^2 = \sum_{j=0}^{m} b_j g^j. \quad (13)$$

We note that the solution of Eq. (1) is presented by the V-function $V_{b_0,...,b_m}(\xi; 1, 2, m)$. Below, we give two examples of the solution of this equation depending on the numerical value of m.

3.1   Case M = 4

According to the balance equation (11), $n = 2$. In this case

$$u(\xi) = a_0 + a_1 g + a_2 g^2, \quad \left(\frac{dg}{d\xi}\right)^2 = b_0 + b_1 g + b_2 g^2 + b_3 g^3 + b_4 g^4. \quad (14)$$

The substitution of Eq. (14) in Eq. (10) leads to a system of six non-linear algebraic equations (Step 4. of the SesM). One non-trivial solution of this system is

$$a_0 = \frac{1}{40} \frac{1}{b_4 \sqrt{\beta \kappa(-\lambda \omega + \mu \kappa)}} (80 \mu \kappa^3 b_4 b_2 - 4(480 \mu \kappa^5 b_4^2 b_2^2 \lambda \omega$$

$$-1440 \kappa^5 b_4^3 b_0 \lambda \omega \mu - 10 \kappa^2 \alpha b_4^2 \mu - 240 \mu^2 \kappa^6 b_4^2 b_2^2 - 10 \omega b_4^2 \mu \kappa$$

$$+720 \mu^2 \kappa^6 b_4^3 b_0 + 720 \kappa^4 b_4^3 b_0 \lambda^2 \omega^2 + 10 \omega^2 b_4^2 \lambda + 10 \kappa \alpha b_4^2 \lambda \omega$$

$$-240 \kappa^4 b_4^2 b_2^2 \lambda^2 \omega^2)^{1/2} - 80 \kappa^2 b4 b_2 \lambda \omega) \sqrt{10}; \quad b_1 = \frac{1}{8} \frac{(4b_4 b_2 - b_3^2)b_3}{b_4^2}$$

$$a_1 = 1008 \frac{\kappa^4 \sqrt{10}\sqrt{\beta \kappa (-\lambda \omega + \mu \kappa)} b_3}{\beta (24 \kappa^3 \lambda \omega b_4^2 - 24 \kappa^4 \mu b_4^2 + 360 \kappa^3)}$$

$$a_2 = 6 \frac{\sqrt{10}\sqrt{\beta \kappa (-\lambda \omega + \mu k)} b_4}{\beta}^{\smile}, \quad (15)$$

$$\epsilon = \frac{\kappa^2}{4\alpha^3 b_4} (4 b_4 \omega \gamma + 4(-240 \kappa^6 \mu^2 b_2^2 b_4^2 + 720 \kappa^6 \mu^2 b_4^3 b_0 - 1440 \kappa^5 b_4^3 b_0 \lambda \omega \mu$$

$$+480 \kappa^5 b_2^2 b_4^2 \lambda \omega \mu - 240 \kappa^4 b_2^2 b_4^2 \lambda^2 \omega^2 + 720 \kappa^4 b_4^3 b_0 \lambda^2 \omega^2$$

$$-10 \kappa^2 \alpha b_4^2 \mu - 10 \omega b_4^2 k \mu + 10 \kappa \alpha b_4^2 \lambda \omega + 10 \omega^2 b_4^2 \lambda)^{1/2}).$$

Then the used simple equation becomes

$$\left(\frac{dg}{d\xi}\right)^2 = b_0 + \frac{1}{8} \frac{(4b_4 b_2 - b_3^2)b_3}{b_4^2} g + b_2 g^2 + b_3 g^3 + b_4 g^4, \quad (16)$$

where b_0, b_2, b_3, and b_4 are free parameters, and b_1 is dependent on b_2, b_3 and b_4. The solution of Eq. (16) can be written as

$$g = V_{b_0, \frac{1}{8} \frac{(4b_4 b_2 - b_3^2)b_3}{b_4^2}, b_2, b_3, b_4} (\xi; 1, 2, 4). \quad (17)$$

Then, the solution of Eq. (10) (and Eq. 1 respectively) is

$$u(\xi) = \frac{1}{40} \frac{1}{b_4 \sqrt{\beta \kappa (-\lambda \omega + \mu \kappa)}} (80 \, \mu \, \kappa^3 b_4 \, b_2 - 4 \, (480 \, \mu \, \kappa^5 b_4^2 b_2^2 \lambda \omega$$

$$1440 \, \kappa^5 b_4^3 b_0 \, \lambda \omega \, \mu - 10 \, \kappa^2 \alpha \, b_4^2 \mu - 240 \, \mu^2 \kappa^6 b_4^2 b_2^2 - 10 \, \omega \, b_4^2 \mu \, \kappa$$

$$+720 \, \mu^2 \kappa^6 b_4^3 b_0 + 720 \, \kappa^4 b_4^3 b_0 \, \lambda^2 \omega^2 + 10 \, \omega^2 b_4^2 \lambda + 10 \, \kappa \, \alpha \, b_4^2 \lambda \omega$$

$$-240 \, \kappa^4 b_4^2 b_2^2 \lambda^2 \omega^2)^{1/2} - 80 \, \kappa^2 b4 \, b_2 \, \lambda \omega) \sqrt{10} \qquad (18)$$

$$+1008 \frac{\kappa^4 \sqrt{10} \sqrt{\beta \kappa (-\lambda \omega + \mu \kappa)} b_3}{\beta \, (24 \, \kappa^3 \lambda \omega \, b_4^2 - 24 \, \kappa^4 \mu \, b_4^2 + 360 \, \kappa^3)} V_{b_0, \frac{1}{8} \frac{(4b_4 b_2 - b_3^2) b_3}{b_4^2}, b_2, b_3, b_4} (\xi; 1, 2, 4)$$

$$+6 \frac{\sqrt{10} \sqrt{\beta \kappa (-\lambda \omega + \mu \, k)} b_4}{\beta} \check{V}^2_{b_0, \frac{1}{8} \frac{(4b_4 b_2 - b_3^2) b_3}{b_4^2}, b_2, b_3, b_4} (\xi; 1, 2, 4), \quad \xi = \kappa x + \omega t.$$

Let us consider the specific case $b_1 = 0$ and $b_3 = 0$. Then the non-trivial solution of the system of non-linear algebraic equations reduces to

$$a_0 = \frac{1}{b_4 \sqrt{\beta \kappa (-\lambda \omega + \mu \kappa)}} (80 \, \mu \, \kappa^3 b_4 \, b_2 - (160 \omega^2 b_4^2 \lambda - 3840 \kappa^4 b_4^2 b_2^2 \lambda^2 \omega^2$$

$$-160 \omega b_4^2 \mu \kappa - 160 \kappa^2 \alpha b_4^2 \mu - 3840 \mu^2 \kappa^6 b_4^2 b_2^2 + 11520 \mu^2 \kappa^6 b_4^3 b_0$$

$$+11520 \kappa^4 b_4^3 b_0 \lambda^2 \omega^2 + 160 \kappa \alpha b_4^2 \lambda \omega - 23040 \kappa^5 b_4^3 b_0 \lambda \omega \mu$$

$$+7680 \mu \kappa^5 b_4^2 b_2^2 \lambda \omega)^{1/2} - 80 \, \kappa^2 b_4 \, b_2 \, \lambda \omega) \sqrt{10} \qquad (19)$$

$$a_1 = 0; \quad a_2 = 6 \frac{\sqrt{10} \sqrt{\beta \kappa (-\lambda \omega + \mu \kappa)} b_4 \kappa}{\beta};$$

$$\epsilon = \frac{\kappa^2}{\alpha^3} (\omega \gamma + (-240 \, \kappa^6 \mu^2 b_2^2 + 720 \, k^6 \mu^2 b_4 b_0 - 1440 \, \kappa^5 \mu \, b_4 b_0 \lambda \omega$$

$$+480 \, \kappa^5 \mu \, b_2^2 \lambda \omega - 240 \, \kappa^4 \lambda^2 \omega^2 b_2^2 + 720 \, \kappa^4 \lambda^2 \omega^2 b_4 b_0 - 10 \, \alpha \, \mu \, \kappa^2 + 10 \, \omega^2 \lambda)^{1/2}).$$

Then the used simple equation becomes

$$\left(\frac{dg}{d\xi}\right)^2 = b_0 + b_2 g^2 + b_4 g^4, \qquad (20)$$

where b_0, b_2, and b_4 are free parameters. Equation (20) is of the Jacobi elliptic function kind, and its solution can be presented as

$$g = V_{b_0, 0, b_2, 0, b_4}(\xi; 1, 2, 4). \qquad (21)$$

Then, the solution of Eq. (10) (and Eq. 1 respectively) reduces to

$$u(\xi) = \frac{1}{b_4 \sqrt{\beta \kappa \, (-\lambda \omega + \mu \kappa)}} (80 \, \mu \, \kappa^3 b_4 \, b_2 - (160\omega^2 b_4^2 \lambda - 3840\kappa^4 b_4^2 b_2^2 \lambda^2 \omega^2$$

$$-160\omega b_4^2 \mu\kappa - 160\kappa^2 \alpha b_4^2 \mu - 3840\mu^2 \kappa^6 b_4^2 b_2^2 + 11520\mu^2 \kappa^6 b_4^3 b_0$$

$$+11520\kappa^4 b_4^3 b_0 \lambda^2 \omega^2 + 160\kappa\alpha b_4^2 \lambda\omega - 23040\kappa^5 b_4^3 b_0 \lambda\omega\mu \qquad (22)$$

$$+7680\mu\kappa^5 b_4^2 b_2^2 \lambda\omega)^{1/2} - 80 \, \kappa^2 b_4 \, b_2 \, \lambda \, \omega)\sqrt{10}$$

$$+6 \frac{\sqrt{10}\sqrt{\beta \kappa \, (-\lambda \omega + \mu \kappa)}b_4\kappa}{\beta} V_{b_0,0,b_2,0,b_4}^2(\xi; 1, 2, 4), \quad \xi = \kappa x + \omega t.$$

In the context of Jacobi elliptic functions, the solution (22) presented by the special function $V_{b_0,0,b_2,0,b_4}(\xi; 1, 2, 4)$ may have various forms depending on the numerical values of the free coefficients b_0, b_2, and b_4. The most popular variants of solutions are

$$V_{1,0,-(1+k^{*2}),0,k^{*2}}(\xi; 1, 2, 4) = sn(\xi, k^*);$$
$$V_{(1-k^{*2}),0,(2k^{*2}-1),0,k^{*2}}(\xi; 1, 2, 4) = cn(\xi, k^*); \qquad (23)$$
$$V_{-(1-k^{*2}),0,(2-k^{*2}),0,-1}(\xi; 1, 2, 4) = dn(\xi, k^*),$$

where k^* $(0 < k^* < 1)$ is the modulus of the elliptic function. In addition, when $k^* \to 0$, the Jacobi elliptic functions degenerate to the trigonometric functions. For example

$$sn(\xi) \to \sin(\xi), \quad cn(\xi) \to \cos(\xi), \quad dn(\xi) \to 1. \qquad (24)$$

Moreover, when $k^* \to 1$, the Jacobi elliptic functions degenerate to the hyperbolic functions. For example

$$sn(\xi) \to \tanh(\xi), \quad cn(\xi) \to sech(\xi), \quad dn(\xi) \to sech(\xi). \qquad (25)$$

All possible solutions of the Jacobi elliptic equation (and $V_{b_0,0,b_2,0,b_4}(\xi; 1, 2, 4)$, respectively) depending on the values of b_0, b_2, and b_4 can be found in [24].

3.2   Case M = 3

According to the balance equation (11), $n = 1$. In this case

$$u(\xi) = a_0 + a_1 g, \quad \left(\frac{dg}{d\xi}\right)^2 = b_0 + b_1 g + b_2 g^2 + b_3 g^3. \qquad (26)$$

The substitution of Eq. (26) in Eq. (10) leads to a system of three non-linear algebraic equations (Step 4 of the SesM). One non-trivial solution of this system is

$$a_0 = \frac{1}{10} \frac{\left(\gamma \omega - 5\kappa^2 \lambda \omega b_2 + 5\kappa^3 \mu b_2 - \epsilon \kappa\right) \sqrt{10}}{\sqrt{\beta \kappa \left(-\lambda \omega + \mu \kappa\right)}};$$

$$a_1 = \frac{3}{2} \frac{\sqrt{10}\sqrt{\beta \kappa \left(-\lambda \omega + \mu \kappa\right)}\kappa b_3}{\beta} \tag{27}$$

$$b_1 = (-10\,\alpha\,\kappa\,\lambda\,\omega - 2\,\gamma\,\omega\,\epsilon\,\kappa + 15\,\kappa^4\lambda^2\omega^2 b_2^2 - 30\,\kappa^5\mu\,b_2^2\lambda\,\omega - 10\,\lambda\,\omega^2$$
$$+\gamma^2\omega^2 + \epsilon^2\kappa^2 + 15\,\kappa^6\mu^2 b_2^2 + 10\,\omega\,\mu\,\kappa + 10\,\alpha\,\kappa^2\mu)/(45\kappa^4 b_3(\lambda^2\omega^2$$
$$-2\,\lambda\,\omega\,\mu\,\kappa + \mu^2\kappa^2)).$$

Then the used simple equation becomes

$$\left(\frac{dg}{d\xi}\right)^2 = b_0 + b_1 g + b_2 g^2 + b_3 g^3, \tag{28}$$

where b_0, b_2, b_3 are free parameters, and b_1 is dependent on b_2 and b_3 (see Eq. (27)). The solution of Eq. (28) can be written as

$$g = V_{b_0,b_1,b_2,b_3}(\xi; 1, 2, 3) \tag{29}$$

where b_1 is presented in Eq. (27). Then, the solution of Eq. (10) (and Eq. 1, respectively) is

$$u(\xi) = \frac{1}{10} \frac{\left(\gamma \omega - 5\kappa^2 \lambda \omega b_2 + 5\kappa^3 \mu b_2 - \epsilon \kappa\right) \sqrt{10}}{\sqrt{\beta \kappa \left(-\lambda \omega + \mu \kappa\right)}} \tag{30}$$
$$+\frac{3}{2} \frac{\sqrt{10}\sqrt{\beta \kappa \left(-\lambda \omega + \mu \kappa\right)}\kappa b_3}{\beta} V_{b_0,b_1,b_2,b_3}(\xi; 1, 2, 3)$$

as $\xi = \kappa x + \omega t$.

Let us consider the specific case $b_2 = 0$ and $b_3 = 4$. We rewrite also $b_0 = -g_3$ and $b_1 = -g_2$. Then the used simple equation is reduced to an equation of the Weierstrass elliptic function kind $\wp(\xi, g_2, g_3)$

$$\left(\frac{d\wp}{d\xi}\right)^2 = 4\wp^3 - g_2\wp - g_3, \tag{31}$$

where g_2 and g_3 are parameters. Then the non-trivial solution of the system of non-linear algebraic equations reduces to

$$a_0 = \frac{1}{10} \frac{(\gamma \omega - \epsilon \kappa) \sqrt{10}}{\sqrt{\beta \kappa \left(-\lambda \omega + \mu \kappa\right)}}; \quad a_1 = 6 \frac{\sqrt{10}\sqrt{\beta \kappa \left(-\lambda \omega + \mu \kappa\right)}\kappa}{\beta} \tag{32}$$

$$\epsilon = \frac{1}{\kappa}(\gamma \omega + (10\,\alpha\,\kappa\,\lambda\,\omega - 10\,\omega\,\mu\,\kappa + 10\,\lambda\,\omega^2 + 360\,\kappa^5\lambda\,\omega\,\mu$$
$$-180g_2\kappa^6\mu^2 - 10\,\alpha\,\kappa^2\mu - 180g_2\kappa^4\lambda^2\omega^2)^{1/2})$$

where g_2 is a free parameter (The parameter g_3 is not presented in the formulas given above.). For this case, the V function can be presented as

$$V_{b_0,b_1,0,4}(\xi; 1, 2, 3) = \wp(\xi, g_2, g_3). \tag{33}$$

Then, the solution of Eq. (10) (and Eq. 1, respectively) is reduced to

$$u(\xi) = \frac{1}{10} \frac{(\gamma\omega - \epsilon\kappa)\sqrt{10}}{\sqrt{\beta\kappa\,(-\lambda\omega + \mu\kappa)}} + 6 \frac{\sqrt{10}\sqrt{\beta\kappa\,(-\lambda\omega + \mu\kappa)}\kappa}{\beta}\wp(\xi, g_2, g_3), \ \ \xi = \kappa x + \omega t. \tag{34}$$

In order to show the forms of the Weierstrass elliptic functions in a more familiar (popular) format, we rewrite them in terms of Jacobi's elliptic functions. The corresponding relations are

$$\begin{aligned}
\wp(\xi, g_2, g_3) &= e_3 + \frac{e_1 - e_3}{sn^2(\xi\sqrt{e_1 - e_3}, k^*)}; \\
\wp(\xi, g_2, g_3) &= e_2 + (e_1 - e_3)\frac{dn^2(\xi\sqrt{e_1 - e_3}, k^*)}{sn^2(\xi\sqrt{e_1 - e_3}, k^*)}; \\
\wp(\xi, g_2, g_3) &= e_1 + (e_1 - e_3)\frac{cn^2(\xi\sqrt{e_1 - e_3}, k^*)}{sn^2(\xi\sqrt{e_1 - e_3}, k^*)},
\end{aligned} \tag{35}$$

where e_1, e_2, and e_3 are the roots of the polynomial $4\wp^3 - g_2\wp - g_3$, and the modulus of Jacobi's elliptic functions is $k^* = \sqrt{\frac{e_2 - e_3}{e_1 - e_3}}$ ($e_1 > e_2 > e_3$). The Weierstrass elliptic functions can take other forms considered in the context of Jacobi elliptic functions. For example, when $k^* \to 0$, i.e., $e_2 \to e_3$, the Jacobi elliptic functions (and the Weierstrass elliptic functions, respectively) can transform to the trigonometric functions (for instance, see Eq. 24). Moreover, when $k^* \to 1$, i.e., $e_2 \to e_1$, the Jacobi elliptic functions (and the Weierstrass elliptic functions, respectively) can transform to the hyperbolic functions (for instance, see Eq. 25).

4   Conclusion

In this paper, we obtain various exact traveling wave solutions of a variant of the generalized Rosenau–Kawahara-RLW equation by applying a particular case of SEsM, including the use of one simple equation. The elliptic equation of Jacobi and the elliptic equation of Weierstrass are used as simple equations. The general solutions of the studied equation are presented by the special function V. It is shown that for specific values of the coefficients of the used simple equations, exact solutions of the elliptic kind, of triangular kind, or of hyperbolic kind are possible.

References

1. Ablowitz, M. J., Clarkson, P. A.: Solitons, Nonlinear Evolution Equations and Inverse Scattering. Cambridge: Cambridge University Press (1991).
2. Gardner, C. S., Greene, J. M., Kruskal, M. D., Miura, R. R.: Method for solving the Korteweg-deVries equation. Phys. Rev. Lett. 19, 1095 – 1097 (1967). https://doi.org/10.1103/PhysRevLett.19.1095
3. Hirota, R.: Exact Solution of the Korteweg-de Vries Equation for Multiple Collisions of Solitons. Phys. Rev. Lett. 27 1192 – 1194 (1971). https://doi.org/10.1103/PhysRevLett.27.1192
4. Fan, E., Hon, Y. C.: A series of travelling wave solutions for two variant Boussinesq equations in shallow water waves.Chaos, Solitons & Fractals 15, 559 – 566 (2003). https://doi.org/10.1016/S0960-0779(02)00144-3
5. Wazwaz, A. M.: The tanh method for traveling wave solutions of nonlinear equations. Applied Mathematics and Computation 154, 713 – 723 (2004). https://doi.org/10.1016/S0096-3003(03)00745-8
6. Wazwaz, A. M.: Partial differential equations and solitary waves theory, Springer, Dordrecht (2009).
7. Kudryashov, N. A.: Simplest equation method to look for exact solutions of nonlinear differential equations.Chaos Solitons & Fractals 24, 1217 – 1231 (2005). https://doi.org/10.1016/j.chaos.2004.09.109
8. Vitanov, N. K.: Application of simplest equations of Bernoulli and Riccati kind for obtaining exact traveling-wave solutions for a class of PDEs with polynomial nonlinearity.Communications in Nonlinear Science and Numerical Simulations, 15, 2050–2060 (2010). https://doi.org/10.1016/j.cnsns.2009.08.011
9. Vitanov, N. K.: On modified method of simplest equation for obtaining exact and approximate solutions of nonlinear PDEs: The role of the simplest equation. Communications in Nonlinear Science and Numerical Simulations 16, 4215–4231 (2011). https://doi.org/10.1016/j.cnsns.2011.03.035
10. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: On the class of nonlinear PDEs that can be treated by the modified method of simplest equation. Application to generalized Degasperis-Processi equation and b-equation.Communications in Nonlinear Science and Numerical Simulation 16, 3033–3044 (2011). https://doi.org/10.1016/j.cnsns.2010.11.013
11. Vitanov , N. K., Dimitrova, Z. I., Kantz, H.: Application of the method of simplest equation for obtaining exact traveling-wave solutions for the extended Korteweg-de Vries equation and generalized Camassa-Holm equation. Applied Mathematics and Computation 219 7480-7492 (2013). https://doi.org/10.1016/j.amc.2013.01.035
12. Vitanov, N. K.: The simple equations method (SEsM) for obtaining exact solutions of nonlinear PDEs: Opportunities connected to the exponential functions. AIP Conference Proceedings 2159, 030038 (2019). https://doi.org/10.1063/1.5127503
13. Vitanov, N. K., Dimitrova, Z. I.: Simple equations method (SEsM) and other direct methods for obtaining exact solutions of nonlinear PDEs. AIP Conference Proceedings 2159, 030039 (2019). https://doi.org/10.1063/1.5127504
14. Vitanov, N. K., Dimitrova, Z. I.: Simple Equations Method and Non-Linear Differential Equations with Non-Polynomial Non-Linearity. Entropy 23, 1624 (2021). https://doi.org/10.3390/e23121624

15. Jordanov, I. P.: JSimple equations method applied to equations of nonlinear Schrödinger kind. AIP Conference Proceedings 2459, 030016 (2022). https://doi.org/10.1063/5.0084340
16. Vitanov, N. K.: Simple equations method (SEsM): Review and new results. AIP Conference Proceedings 2459, 020003 (2022). https://doi.org/10.1063/5.0083565
17. Dimitrova, Z. I.: On several specific cases of the simple equations method (SEsM): Jacobi elliptic function expansion method, F-expansion method, modified simple equation method, trial function method, general projective Riccati equations method, and first intergal method. AIP Conference Proceedings 2459, 030006 (2022). https://doi.org/10.1063/5.0083573
18. Dimitrova, Z. I.: Several examples of application of the simple equations method (SEsM) for obtaining exact solutions of nonlinear PDEs. AIP Conference Proceedings 2459, 030005 (2022), https://doi.org/10.1063/5.0083572
19. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: On the Use of Composite Functions in the Simple Equations Method to Obtain Exact Solutions of Nonlinear Differential Equations. Computation 9, 104 (2021). https://doi.org/10.3390/computation9100104
20. He, D.D., Pan, K.J.: Exact solitary solution and a three-level linearly implicit conservative finite difference method for the generalized Rosenau-Kawahara-RLW equation with generalized Novikov type perturbation. Appl. Math. Comput. 271, 323–336 (2015). https://doi.org/10.1007/s11071-016-2700-x
21. Peregrine, D. H.: Calculations of the development of an undular bore.J. Fluid Mech. 25, 321–330 (1966). https://doi.org/10.1017/S0022112066001678
22. Rosenau, P.: A Quasi-Continuous Description of a Nonlinear Transmission Line.Phys. Scr. 34, 827–829 (1986). https://doi.org/10.1088/0031-8949/34/6B/020/meta
23. Kawahara, T.: Oscillatory Solitary Waves in Dispersive Media. Journal of thePhysical Society of Japan 33, 260–264 (1972). https://doi.org/10.1143/JPSJ.33.260
24. Filiz, A., Ekici, M., Sonmezoglu, A.: F-expansion method and new exact solutions of the Schrödinger-KdV equation. Hindawi Publishing Corporation, The Scientific World Journal 2014, Article ID 534063 (2014).

Several Properties of the Solutions of Linear and Semilinear Harmonic and Polyharmonic Equations

Petar Popivanov and Angela Slavova

Abstract This paper deals with several necessary conditions for the existence of nontrivial classical non-negative solutions to the Dirichlet problem for the semilinear polyharmonic equation and for the existence of weak solutions to the Dirichlet problem for Liouville equation with power type singularity $|x|^{-2}$. By using variants of Harnack inequality for some classes of second-order elliptic PDEs (say of Cordes type), quantitative Hopf's principle is verified and the corresponding constant in the case of harmonic functions is written explicitly.

Keywords Semilinear polyharmonic equation · Liouville equation · Harnack inequality · Quantitative Hopf's principle · Explicit solutions

1 Introduction and Formulation of the Main Results

We have two aims in this paper. At first, we find necessary conditions for the solvability of the homogeneous Dirichlet problem for the semilinear polyharmonic equation $(-\Delta)^m u = f(u) + g(x)$ in the unit ball B_1 in $\mathbf{R}^n$ with polynomial nonlinearity $f(u)$ and right-hand side $g(x) < 0$; $f(u)$ contains the term $c|u|^{u-1}.s$, $s = \frac{n+2m}{n-2m}$, and $n > 2m$, being the critical embedding Sobolev exponent. For a special choice of $f(u)$ and $g(x)$, the Dirichlet problem under consideration does not possess classical positive (non-negative) in B_1 nontrivial solution u.

Our first result is illustrated by the well-known example having $f(u) = \lambda u + |u|^s s$, $\lambda \leq 0$. We shall formulate in a more general form Proposition 1, while more refined results will be proposed and commented on the process of proving.

Proposition 1 *Consider the boundary value problem (bvp)*

$$(-\Delta)^m u = f(u) + g(x), f \in C^1(\mathbf{R}^1), g(x) \in C^1 \text{in } \bar{B}_1, \tag{1}$$

P. Popivanov · A. Slavova (✉)
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia 1113, Bulgaria
e-mail: slavova@math.bas.bg

$B_1 = \{x \in \mathbf{R}^n : |x| < 1\}$, $S_1 = \partial B_1$, $m > 1$, $n > 2m$ and $u \in C^{2m}(B_1) \cap C^{2m-1}$ $(\bar{B}_1)$, $u|_{S_1} = \frac{\partial u}{\partial n}|_{S_1} = \ldots = \frac{\partial^{m-1} u}{\partial n^{m-1}}|_{S_1} = 0$, $\frac{\partial}{\partial n}|_{S_1} = \frac{\partial}{\partial r}|_{r=1}$. Put $I_1 = \int_{B_1} [F(u(x)) - \frac{n-2m}{2n} u(x) f(u(x))] dx$, where $F = \int_0^z f(\mu) d\mu$ and $I_2 = \frac{1}{n} \int_{B_1} u(x)(Lu + \frac{n+2m}{2})$ $g(x) dx$, where $L = \sum_{j=1}^n x_j \frac{\partial}{\partial x_j}$ is the radial vector field.

Then (1) does not possess the classical solution u if $I_1 \leq 0$, $I_2 \leq 0$ and $I_1 + I_2 < 0$. Under the requirements (15), (18) proposed below (1) does not have classical non-negative non trivial solutions.

Later on, in Example 2, we construct a 1 parametric family of non-classical solutions $0 > u_c \in C^\infty(B_1 \setminus 0)$ of the Liouville-type equation, which develop logarithmic singularities at the origin, while $\frac{e^u}{r^2} \in C^\infty(\bar{B}_1)$ for the values of the parameter $c = 4k^2$, $k \in \mathbf{N}$. More precisely, u can be prolonged at 0 in such a way that $\frac{e^u}{r^2}$ becomes C^∞ smooth everywhere.

Then we prove that

$$\left| \begin{array}{l} -\Delta u = \frac{e^u}{2|x|^2} + f(x), 0 \in \Omega \subset \mathbf{R}^n, n \geq 2, \Omega - \text{ bdd }, \partial\Omega \in C^2 \\ u|_{\partial\Omega} = 0 \end{array} \right. \tag{2}$$

with $f \geq 0$, $f \not\equiv 0$, $d(x)f(x) \in L_1(\Omega)$, $d(x) = dist(x, \partial\Omega)$, $x \in \Omega$ does not possess weak solutions.

In the second part of the paper, we prove the quantitative Hopf lemma for Cordes-type second-order elliptic equation

$$Lu = \sum_{i,j=1}^n a_{ij}(x)u_{x_i x_j} = 0, a_{ij}(x) = a_{ji}(x) \tag{3}$$

in the bounded domain Ω with $\partial\Omega \in C^2$. The coefficients $|a_{ij}(x)| \leq M$, $M = const > 0$ and if we denote by $A(x) = (a_{ij}(x))_{i,j=1}^n$, according to Cordes condition, the eigenvalues $0 < \lambda_1(x) \leq \ldots \leq \lambda_n(x)$ of $A(x)$ do not scatter too much, i.e.,

$$tr A(x) = \sum_{i=1}^n \lambda_i(x) \leq \lambda \lambda_1(x) \tag{4}$$

with $n \leq \lambda < n + 2$.

We shall begin with the Laplace equation.

Proposition 2 Suppose that $u \in C^1(\bar{B}_1)$ is a non-constant harmonic function in $B_1(0)$ and it attains its maximum at $x_0 \in \partial B_1(0)$. Then

$$\frac{\partial u}{\partial n}(x_0) \geq (2n+1)e^{-(2n+1)} \frac{1}{3} \left(\frac{2}{3} \right)^{n-2} (u(x_0) - u(0)). \tag{5}$$

This is the quantitative Hopf principle.

Proposition 3 *Consider the non-constant classical solution $u(x)$ of (3), $u \in C^0(\bar{\Omega})$ and assume that $\frac{\partial u}{\partial n}$ exists at the point $x_0 \in \partial\Omega$ such that $\max_{\bar{\Omega}} u = u(x_0)$. Then there exists a positive constant $C(\lambda, n, M, \partial\Omega)$ for which $\frac{\partial u}{\partial n}(x_0) \geq C(u(x_0) - u(z_0))$, $z_0 \in \Omega$.*

As usual, n stands for the unit outward normal to $\partial\Omega$ at x_0 and the location of z_0 can be specified—z_0 belongs to some ball $B_R(y_0) \subset \Omega$ which is tangential to $\partial\Omega$ at x_0.

Detailed proofs of Proposition 1–3 are given in Sect. 2. The proof of Proposition 1 is based on Pohozaev identity [4], the study of (2) relies on [1], the proof of Proposition 2 is standard but uses Harnack inequality for non-negative harmonic functions in the ball. In the proof of Proposition 3, we apply [11] and Harnack inequality for positive solutions of (2) under Cordes condition (4) [2, 3, 8, 9, 12, 13].

2 Proofs and Comments of Propositions 1–3

1. To find several necessary conditions for the solvability of the Dirichlet problem for the polyharmonic equation in the ball, we shall use Pohozaev-type identity [4]. Thus, let $B_1 = \{x \in \mathbf{R}^n : |x| < 1\}$, $S_1 = \partial B_1$, $n > 2m$, $m > 1$ and $u \in C^{2m}(B_1) \cap C^{2m-1}(\bar{B}_1)$ satisfies the Dirichlet problem (1) into the equivalent form:

$$\left| \begin{array}{l} (-\Delta)^m u = f(u) + g(x), \ f \in C^1(\mathbf{R}^1), g \in C^1(\bar{B}_1) \\ D^\alpha u|_{S_1} = 0, |\alpha| \leq m - 1. \end{array} \right. \tag{6}$$

Evidently, $u \equiv 0$ in $(1) \Rightarrow g(x) = -f(0)$, $\forall x \in B_1$.

There are two different cases for $m : m$ even, i.e., $m = 2k$, $k \geq 1$, m-odd, i.e. $m = 2k + 1$. Then it is proved in [4] that if $L = \sum_{j=1}^n x_j \frac{\partial}{\partial x_j} = r \frac{d}{dr}$ is the radial vector field

$$\int_{B_1} (-\Delta)^{2k} u L u \, dx = -\frac{n-2m}{2} \int_{B_1} (\Delta^k u)^2 dx - \frac{1}{2} \int_{S_1} (\Delta^k u)^2 dw \tag{7}$$

for m-even and

$$\int_{B_1} (-\Delta)^{2k+1} u L u \, dx = -\frac{n-2m}{2} \int_{B_1} |\nabla \Delta^k u|^2 dx - \frac{1}{2} \int_{S_1} \left(\frac{\partial}{\partial n} \Delta^k u \right)^2 dw \tag{8}$$

for m-odd.

Remark 1 (1) If $x \in S_1$ and n_x is the unit outward normal to S_1 at x, then $< x, n_x >= 1$.

(2) Define $s = \frac{n+2m}{n-2m} > 1$ as critical Sobolev exponent.

In fact, the embedding $H_0^m(\Omega)$ into $L^{p+1}(\Omega)$ for Ω-smooth bounded domain in $\mathbf{R}^n$ is compact for $p < s$. It remains continuous but not compact for $p = s$, namely, $H_0^m(\Omega) \subset L^{s+1}$, $s + 1 = \frac{2n}{n-2m} > 2$, $n > 2$ ([4]).

There are no difficulties to guess that

$$\int_{B_1} g(x)Lu\,dx = -\int_{B_1} uLg - n\int_{B_1} u(x)g(x)dx = -\int_{B_1} u(x)(L+n)g(x)dx, \tag{9}$$

$$\int_{B_1} f(u)Lu\,dx = -n\int_{B_1} F(u(x))dx, \tag{10}$$

where $F(z) = \int_0^z f(\mu)d\mu$, i.e. $F'(z) = f(z)$.

From (6) with testing function $\frac{n-2m}{2}u$, we conclude that for $m = 2k$

$$\frac{n-2m}{2}\int_{B_1}(\Delta^k u)^2 dx = \frac{n-2m}{2}\left[\int_{B_1} uf(u)dx + \int_{B_1} ug(x)dx\right], \tag{11}$$

while for $m = 2k+1$

$$\frac{n-2m}{2}\int_{B_1}|\nabla\Delta^k u|^2 dx = \frac{n-2m}{2}\left[\int_{B_1} uf(u)dx + \int_{B_1} ug(x)dx\right]. \tag{12}$$

In fact, $\int_{B_1}\Delta v.v + \int_{B_1}|\nabla v|^2 = \int_{S_1} v\frac{\partial v}{\partial n}dw = 0$ if $v = \Delta^k u \Rightarrow v|_{S_1} = 0$.

Combining (7), (9), (10), and (11), respectively, with (8), (9), (10), and (12), we deduce that for $m = 2k$

$$0 \le \frac{1}{2n}\int_{S_1}(\Delta^k u)^2 dw = \int_{B_1}\left[F(u(x)) - \frac{n-2m}{2n}u(x)f(u(x))\right]dx + \tag{13}$$
$$\frac{1}{n}\int_{B_1} u(x)\left(L + \frac{n+2m}{2}\right)g(x)dx \equiv I_1 + I_2$$

and, for $m = 2k+1$,

$$0 \le \frac{1}{2n}\int_{S_1}\left(\frac{\partial}{\partial n}\Delta^k u\right)^2 dw = \int_{B_1}\left[F(u(x)) - \frac{n-2m}{2n}u(x)f(u(x))\right]dx + \tag{14}$$
$$\frac{1}{n}\int_{B_1} u(x)\left(L + \frac{n+2m}{2}\right)g(x)dx = I_1 + I_2$$

Certainly, $0 < \frac{n-2m}{2n} = \frac{1}{s+1} < 1/2$. (6) does not possess classical solution if $I_1 \le 0$, $I_2 \le 0$ and $I_1 + I_2 < 0$. (13), (14) are identities for $u = 0$.

Our first step is to solve the linear first-order PDE

$$(L + A)g = h(x) \in C^0(\bar{B}_1) \tag{15}$$

with (a) $h \leq 0$, $h \not\equiv 0$, possibly (b) $h(x) < 0$ for $x \neq 0$, $A = \frac{n+2m}{2} > 2$.

The method of characteristics gives us that (15) possesses the solution

$$g(x) = \int_0^1 t^{A-1} h(tx) dt \leq 0, \; g \not\equiv 0, \tag{16}$$

and $g(x) < 0$ for $x \neq 0$ if $h(x) < 0$ for $x \neq 0$; $g(0) = \frac{h(0)}{A}$, $h \in C^p \Rightarrow g \in C^p$, $p \in \mathbf{N}_0$.

One can obtain the same result by writing (15) in polar coordinates, i.e., $r\frac{dg}{dr} + Ag = h(r, \omega)$. This is linear ODE of first order, etc.

Our next step is to investigate I_1. So, we consider the function

$$F(z) - \frac{1}{s+1} z f(z) = F(z) - \frac{z}{s+1} F^{'}(z). \tag{17}$$

(i.e., we have put $z = u(x)$). We suppose that (17) satisfies ODE with negative right-hand side, i.e.,

$$F(z) - \frac{1}{(s+1)} z F^{'}(z) = q(z) < 0 \; \text{ for } \; 0 < |z| \leq 1, q(0) = 0, 0 < \frac{1}{s+1} < \frac{1}{2}. \tag{18}$$

Thus,

$$F(z) = C|z|^{s+1} - |z|^{s+1}(s+1) \int \frac{q(z) dz}{z|z|^{s+1}}. \tag{19}$$

If $q(z) \sim -z^2$ near 0 then $F(z) \sim C|z|^{s+1} + \frac{s+1}{1-s}|z|^2$ and if $q \equiv 0 \Rightarrow F(z) = C|z|^{s+1}$. Certainly, in the case $q(z) \sim -z^2$ near 0, $q(z) < 0$ for $|z| > 0$, we have that $f(z) = F^{'}(z) \sim C(s+1)|z|^{s-1}z + 2z\frac{s+1}{1-s}$, $z \sim 0$, $C > 0$, $1 - s < 0$. Formally, if $q(z) \sim -z^{2p}$, $z \to 0$ then $F(z) \sim C|z|^{s+1} + (s+1)\frac{|z|^{2p}}{2p-(s+1)}$, i.e., $f(z) \sim C(s+1)|z|^{s-1}z + \frac{2p(s+1)|z|^{2p-2}}{2p-(s+1)}z$, $z \to 0$.

Combining the previous results, we come to Proposition 1 from Sect. 1. Thus, the Dirichlet problem (6) under conditions (15), (18) imposed on the right-hand side $g(x)$ and, on the nonlinearity $f(z)$, $|z| \geq 0$ does possess classical positive solution in the case (15) (a) and non-negative not nontrivial solution in the case (15) (b).

Example 1 As in [4] take $f(u) = \lambda u + |u|^{s-1}u$, $\lambda \in \mathbf{R}^1$-parameter. Then $F(u) = \lambda\frac{u^2}{2} + \frac{|u|^{s+1}}{s+1}$, according to (16) $g(x) = \int_0^1 t^{A-1}h(tx)dt \leq 0$, $g(x) \not\equiv 0$, $I_1 = \lambda(\frac{1}{2} - \frac{1}{s+1})\int_{B_1} u^2 dx$, $u \not\equiv 0$, $\lambda < 0 \Rightarrow I_1 < 0$. Then (6) does not possess classical ≥ 0 solution.

If $g \equiv 0$ (6) does not possess classical nontrivial solution u for $\lambda < 0$ (the sign of u is not important).

Suppose now that $\lambda = 0$ but (15) (b) holds. Then (6) does not possess non-negative nontrivial classical solution in B_1 for $g(x) < 0$, $x \neq 0$ given by (16). Here, the trivial solutions are: $u \equiv 0$; $u \equiv 0$ satisfies (6) with $f(u) = \lambda u + |u|^{s-1}u$ if and only if $g(x) \equiv 0$.

Existence and regularity results for weak and classical solutions of (6) can be found in [4], Sects. 7, 7.2, 7.4, and 7.5.

2. H.Brezis and X.Cabré considered in [1] the Dirichlet problem

$$-\Delta u = \frac{u^2}{|x|^2} + f(x), \ \text{in} \ \Omega, u = 0 \ \text{on} \ \partial\Omega, \tag{20}$$

where Ω is a smooth bounded domain in $\mathbf{R}^n$, $0 \in \Omega$ with $d(x)f(x) \in L_1(\Omega)$, $f \geq 0$, i.e., $f \not\equiv 0$. They proved that (20) does not possess a weak solution in the following sense. Assume that $h(x, u)$ is a Carathéodory function in $\Omega \times \mathbf{R}$. Then $u(x)$ is a weak solution of $-\Delta u = h(x, u)$ in $\Omega, u|_{\partial\Omega} = 0$ if $u \in L_1(\Omega), d(x)h(x, u) \in L_1(\Omega)$ and $-\int_\Omega u\Delta\xi = \int_\Omega h(x, u)\xi$ for each $\xi \in C_0^2(\bar{\Omega})$. As usual, $d(x) = dist(x, \partial\Omega)$.

The problem for solvability of (20) is rather delicate, as $\Delta u \leq 0$ in Ω, i.e., u is superharmonic function that attains its minimum at $\partial\Omega$. If $min_\Omega u = u(\hat{x})$, $\hat{x} \in \Omega \Rightarrow u \equiv const$.

Certainly, it is interesting to find out other solution, that does not enter into the frames of the weak solutions.

Example 2 Consider the Dirichlet problem for the Liouville equation with power singularity in the unit circle $B_1 \subset \mathbf{R}^2$:

$$\Delta u + \frac{e^u}{2|x|^2} = 0, x \in B_1, u|_{S_1} = 0. \tag{21}$$

At first, we construct formally a radial solution u of (21) (see [10]). In polar coordinates, we have $r^2 u''(r) + ru'(r) + \frac{1}{2}e^{u(r)} = 0, 0 < r \leq 1, u(1) = 0$. The standard change $r = e^t, t \in (-\infty, 0]$ and the substitution $t = lnr, u(r) = \tilde{u}(t)$ lead to $\tilde{u}'' + \frac{1}{2}e^{\tilde{u}} = 0, \tilde{u}(0) = 0$, i.e., $(\tilde{u}')^2 = C_1 - e^{\tilde{u}}, C_1 > 0, -\infty < \tilde{u} \leq lnC_1$. We shall investigate only the case $\tilde{u}'(t) = \sqrt{C_1 - e^{\tilde{u}}}$. As $\tilde{u}(0) = 0$, we must have $lnC_1 \geq 0 \Rightarrow C_1 \geq 1$. Suppose that $C_1 > 1$, Thus, $F(\tilde{u}) = \int_0^{\tilde{u}(t)} \frac{d\lambda}{\sqrt{C_1-e^\lambda}} = t = lnr, \tilde{u}(0) = 0$. If

$$F(\tilde{u}) = \int_0^{\tilde{u}} \frac{d\lambda}{\sqrt{C_1 - e^\lambda}} \Rightarrow F'(\tilde{u}) > 0, \tilde{u} \leq lnC_1, F'(lnC_1) = +\infty, \tag{22}$$

$0 < F(lnC_1) < \infty, F(\tilde{u}) \sim \frac{\tilde{u}}{\sqrt{C_1}}, \tilde{u} \to -\infty, F(\tilde{u}) > 0$ for $0 < \tilde{u} \leq lnC_1, F'(0) = \frac{1}{\sqrt{C_1-1}} > 0$. The inverse function $\tilde{u}(t) = F^{-1}(t), -\infty < t \leq F(lnC_1), \tilde{u}(t) \sim \sqrt{C_1}t$, $t \to -\infty, \tilde{u}'(t) > 0, \tilde{u}(0) = 0, \tilde{u}'|_{F(lnC_1)} = 0$. Therefore, $u(r) = F^{-1}(lnr), 0 < r \leq e^{F(lnC_1)}$.

We can compute (22) via the change $z = (C_1 - e^\lambda)^{1/2}$ (see [10]). The indefinite integral $F(\tilde{u}) = \int \frac{d\tilde{u}}{\sqrt{C_1-e^{\tilde{u}}}} = -\frac{1}{\sqrt{C_1}}ln|\frac{\sqrt{C_1}+\sqrt{C_1-e^{\tilde{u}}}}{\sqrt{C_1}-\sqrt{C_1-e^{\tilde{u}}}}|, 0 < e^{\tilde{u}} \leq C_1$ gives that $lnr = F(u) - F(0) = -\frac{1}{\sqrt{C_1}}ln\frac{\frac{\sqrt{C_1}+\sqrt{C_1-e^u}}{\sqrt{C_1}-\sqrt{C_1-e^u}}}{\frac{\sqrt{C_1}+\sqrt{C_1-1}}{\sqrt{C_1}-\sqrt{C_1-1}}}$. Thus,

$$u = ln4C_1 + ln\frac{r^{\sqrt{C_1}}}{(r^{\sqrt{C_1}}(\sqrt{C_1} - \sqrt{C_1 - 1}) + \sqrt{C_1} + \sqrt{C_1 - 1})^2}. \qquad (23)$$

Evidently, $r = 1 \Rightarrow u(1) = 0$, $u \sim \sqrt{C_1}lnr$, $r \to 0$, $u \leq lnC_1$, $u = lnC_1 \Longleftrightarrow$ $r_0 = (\sqrt{C_1} + \sqrt{C_1 - 1})^{\frac{2}{\sqrt{C_1}}} > 1, 0 < r \leq r_0$.

Consider $\frac{e^u}{2r^2} = 2C_1\frac{A}{(A+1)^2 r^2}$. If $\sqrt{C_1} = 2k, k \in \mathbf{N}$, then $\frac{e^u}{2r^2} \in C^\infty(\bar{B}_1)$ as $r^{\sqrt{C_1}} = r^{2k} \in C^\infty[0, \infty)$. Otherwise, the regularity of $\frac{e^u}{2r^2}$ is increasing with $C_1 > 1$. Evidently, $u \in L_p(B_1)$ for each $p \geq 1$. Certainly, $u \notin L^\infty(B_1)$.

Conclusion. We have constructed 1-parametric family of solutions $u = u_{C_1}(r)$, $C_1 > 1$ of (21). u_{C_1} are superharmonic functions, for $C_1 = 4k^2$, we have that $\frac{e^u}{2|x|^2} \in C^\infty(B_{r_0})$, $u \in C^\infty(B_{r_0} \setminus \{0\})$, $B_{r_0} = \{|x| < r_0, x \in \mathbf{R}^2\}$ but u_{C_1} develop logarithmic singularity at the origin and $u_{C_1} < 0$ in B_1. u can be prolonged smoothly in $[r_0, 2r_0]$ by the formula $u(2r_0 - r) = u(r)$ for $0 \leq r \leq r_0$.

We shall say several words about the weak solutions of the Dirichlet problem (2), where $d(x)f(x) \subset L_1(\Omega)$, $f \geq 0$ almost everywhere, $f \not\equiv 0$. Repeating the proof of Corollary 1.3. of [1], we conclude that (2) does not possess weak solutions. In fact, at first, one proves that each weak solution $u = 0$ is almost everywhere. This is a contradiction with (2) as $f(x) \geq 0$ almost everywhere, $f \not\equiv 0$, while $f(x) = -\frac{1}{2|x|^2}$ almost everywhere in Ω. In the two-dimensional case, the proof is rather simple, while if $n \geq 3$, we apply Lemma 1.7 from [1] taking the smooth concave function $\Phi(u) = \begin{cases} e^{-\varepsilon} - e^{-u}, u \geq \varepsilon > 0 \\ \frac{u-\varepsilon}{e^\varepsilon}, u \leq \varepsilon, \end{cases}$, the solution u of $-\Delta u \geq \frac{e^u}{2|x|^2}, u \geq 0$, being $\geq \varepsilon >$ 0 in some ball $B_\eta(0) \subset \Omega$. Then we define $v = \Phi(u), u = u(x)$, and $x \in B_\eta(0)$ and find that $0 \leq v \leq e^{-\varepsilon}$ in $B_\eta(0)$. According to Lemma 1.7, $-\Delta v \geq \frac{1}{2|x|^2}$ in $B_\eta(0)$ in distribution sense $D^{'}(B_\eta(0))$. The function $w = v - \frac{1}{2(n-2)}log\frac{1}{|x|} \in L_1(B_\eta(0))$ and one checks that $-\Delta w \geq 0$ in $D^{'}(B_\eta(0))$. Therefore, there exists a constant $C > 0$ and such that $w \geq -C$ in $B_\eta(0)$. Letting $x \to 0$ (i.e., $r \to 0$), we obtain contradiction $v(x) \to +\infty$, i.e., $u = 0$ almost everywhere.

3. We shall prove now the quantitative Hopf lemma for harmonic functions [6] with an explicit form of the constants participating in the corresponding formula (5). So assume that $u \in C^2(B_1) \cap C^0(\bar{B}_1)$ is harmonic function in the unit ball. Then if $u \geq 0$ in $B_1 \subset \mathbf{R}^n, n \geq 3$, the Harnack inequality [5] claims that

$$\frac{1 - |x|}{(1 + |x|)^{n-1}}u(0) \leq u(x) \leq \frac{1 + |x|}{(1 - |x|)^{n-1}}u(0), x \in B_1. \qquad (24)$$

Having in mind that $f_1(r) = \frac{1-r}{(1+r)^{n-1}}, 0 \leq r \leq 1$ is monotonically decreasing, while $f_2(r) = \frac{1+r}{(1-r)^{n-1}}$ is monotonically increasing function of r we conclude that $f_1(r) \geq \frac{1}{3}(\frac{2}{3})^{n-2}$ in $\bar{B}_{1/2}$ and $0 < f_2(r) \leq 3.2^{n-2}$ in $\bar{B}_{1/2}$. Thus, (24) implies

$$\frac{1}{3}\left(\frac{2}{3}\right)^{n-2} u(0) \leq u(x) \leq 3.2^{n-2}u(0) \text{ in } \bar{B}_{1/2}. \qquad (25)$$

From (25), we deduce that $inf_{B_{1/2}} u \leq sup_{B_{1/2}} u \leq \inf_{B_{1/2}} u 2.3^n$. This is another form of Harnack inequality.

We suppose that $u(x) < u(x_0), \forall x \in B_1, x_0 \in S_1, u(x) \leq u(x_0)$ at S_1, as otherwise $u(x) = u(x_0)$ for some $x \in B_1 \Rightarrow u \equiv u(x_0)$. Another additional condition is $u \in C^1(\bar{B}_1)$ guaranteeing the existence of $\frac{\partial u}{\partial n}(x_0) = \frac{\partial u}{\partial r}|_{r=1}(x_0)$.

As usual consider the auxiliary function $v(x) = e^{-\alpha|x|^2} - e^{-\alpha}$ with $\alpha > 0$. Certainly, $0 < r < 1 \Rightarrow 1 - e^{-\alpha} > v(r) > 0, v(1) = 0$, i.e. $0 < v(r) < 1$. Writing $\Delta = \frac{\partial^2}{\partial r^2} + \frac{n-1}{r}\frac{\partial}{\partial r}$ in radial coordinates we obtain $\Delta v = e^{-\alpha r^2}(4\alpha^2 r^2 - 2\alpha n)$ and therefore $\Delta v > 0 \iff r > \sqrt{\frac{n}{2\alpha}}$. Suppose that $\alpha \geq 2n+1 \Rightarrow \sqrt{\frac{n}{2\alpha}} < \frac{1}{2}$. Taking $r \geq \frac{1}{2}$, we see that v is subharmonic function, i.e., $\Delta v > 0$ in $B_1 \setminus B_{1/2} = A$. So $max_{\bar{A}} u = max_{\partial A} u = max_{S_1 \cup S_{1/2}} u$.

Define in A the function $h_\varepsilon(x) = u(x) - u(x_0) + \varepsilon v(x)$, $\varepsilon > 0$. Then $\Delta h_\varepsilon = \Delta u(x) + \varepsilon \Delta v = \varepsilon \Delta v > 0$ in A and h_ε is subharmonic in A, $max_{\bar{A}} h_\varepsilon = max_{S_1 \cup S_{1/2}} h_\varepsilon$. Thus, $h_\varepsilon|_{S_1} = u(x) - u(x_0)|_{S_1} \leq 0$, $h_\varepsilon(x_0) = 0$ and

$$h_\varepsilon|_{S_{\frac{1}{2}}} = u|_{S_{1/2}} - u(x_0) + \varepsilon v(x)|_{S_{1/2}} < max_{S_{1/2}} u - u(x_0) + \varepsilon, \tag{26}$$

as $0 < v|_{S_{1/2}} < 1$.

Consider now the non-negative harmonic function $0 < w(x) = u(x_0) - u(x)$ in B_1. According to (25), $inf_{S_{1/2}} w \geq c(n)w(0)$, $c(n) = \frac{1}{3}(\frac{2}{3})^{n-2}$. Therefore, $u(x_0) - u(x) \geq c(n)w(0), \forall x \in S_{1/2} \Rightarrow$

$$max_{S_{1/2}} u \leq u(x_0) - c(n)w(0). \tag{27}$$

From (26) and (27), we get $h_\varepsilon|_{S_{1/2}} < \varepsilon - c(n)w(0)$. Taking $\varepsilon \leq c(n)w(0)$, we conclude that $h_\varepsilon|_{S_{1/2}} < 0$. It is easy to see that $max_{\bar{A}} h_\varepsilon = max_{S_1 \cup S_{1/2}} h_\varepsilon = 0$ and maximum is attained in $x_0 \in S_1 = \partial B_1$. Consequently, $\frac{\partial h_\varepsilon}{\partial n}(x_0) \geq 0$, i.e., $\frac{\partial u}{\partial n}(x_0) \geq -\varepsilon \frac{\partial v}{\partial r}(x_0) = 2\varepsilon \alpha e^{-\alpha}$. Conclusion: $\frac{\partial u}{\partial n}(x_0) \geq 2\alpha e^{-\alpha} c(n)(u(x_0) - u(0))$ and we may take here $\alpha = 2n+1, c(n) = \frac{1}{3}(\frac{2}{3})^{n-2}$. This way Proposition 2 is proved. In fact, (5) holds with the above-given values of α and $c(n)$.

Bellow, we shall generalize Proposition 2 for the Cordes-type second-order elliptic equation (3) (see [2, 3]): $Lu = 0$ in Ω, where Ω is bounded domain in $\mathbf{R}^n, \partial \Omega \in C^2$.

These are the conditions, imposed on L.

(a) $a = inf_{x \in \Omega, |\xi|=1} \sum_{j,k=1}^n a_{jk}(x)\xi_j\xi_k, 0 < a < \infty$. Thus, $\sum_{j,k=1}^n a_{jk}(x)\xi_j\xi_k \geq a|\xi|^2$ and therefore $\sum_{i=1}^n a_{ii}(x) \geq na$ as $a_{ii}(x) \geq a$.
(b) $\sum_{j,k} |a_{jk}(x)| \leq M$, i.e., $\sum_{i=1}^n a_{ii}(x) \leq M$.
(c) $0 < \lambda = sup_{x \in \Omega, |\xi|=1} \frac{\sum_{i=1}^n a_{ii}(x)}{\sum_{j,k=1}^n a_{jk}(x)\xi_j\xi_k}$.

Thus, $\sum_{j,k=1}^n a_{jk}(x)\xi_j\xi_k \geq \frac{1}{\lambda}\sum_{i=1}^n a_{ii}(x)|\xi|^2$, i.e.,

$$\sum_{j,k=1}^n a_{jk}(x)\xi_j\xi_k \geq \frac{na}{\lambda}|\xi|^2, x \in \Omega, \xi \in \mathbf{R}^n. \tag{28}$$

λ is called constant of ellipticity of L.

The Cordes condition (c) has the following interpretation. Consider the symmetric positive matrix $A(x) = (a_{jk}(x))_{j,k=1}^n$ and denote by $\lambda_j(x)$ its eigenvalues: $0 < \lambda_1(x) \le \lambda_2(x) \le \ldots \le \lambda_n(x)$. As we know, $\lambda_1(x) = inf_{|\xi|=1}(A\xi, \xi)$, $\lambda_n(x) = sup_{|\xi|=1}(A\xi, \xi)$, $\sum_{i=1}^n \lambda_i(x) = tr A(x) = \sum_{i=1}^n a_{ii}(x)$. Therefore, $sup_{|\xi|=1} \frac{1}{(A\xi,\xi)} = \frac{1}{\lambda_1(x)} \Rightarrow \lambda = sup_{x\in\Omega, |\xi|=1} \frac{\sum_{i=1}^n \lambda_i(x)}{\lambda_1(x)} \Rightarrow \sum_{i=1}^n \lambda_i(x) \le \lambda\lambda_1(x)$. So $\lambda \ge n$.

Further, we assume that supplementary

(d) $n \le \lambda < n + 2$.

Cordes condition means that the eigenvalues λ_i of $A(x)$ do not scatter too much. This is the result to be used further on.

Theorem 1 (10.2 from Chap. 1 of [8]). *Suppose that $Lu = 0$ in $B_R(y_0) \subset \Omega$, $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$, $\lambda < n + 2$ and $u > 0$ there. Then for each r, $0 < r < R$*

$$sup_{B_r(y_0)} u \le C_1 inf_{B_r(y_0)} u \tag{29}$$

and the constant $C_1 > 0$ depends on λ, n, a, M, $\frac{r}{R}$, i.e., $C_1(\lambda, n, a, M, \frac{r}{R})$. Evidently, $C_1 \ge 1$.

The proof of Proposition 3 imitates the proof of Proposition 2 up to application of Theorem 1 in estimating w. We shall compare Proposition 3 with Theorem 2.2 from [7]. Consider the bounded domain Ω, $\partial\Omega \in C^2$ and the elliptic operator

$$L(u) = \frac{1}{2} \sum_{j,k=1}^n a_{jk}(x) \frac{\partial^2 u}{\partial x_j \partial x_k} + \sum_{i=1}^n b_i(x) \frac{\partial u}{\partial x_i} + c(x)u, \tag{30}$$

$u \in C^2(\Omega) \cap C^0(\bar{\Omega})$, $a_{jk} = a_{kj}$, b_i, $c \in C^0(\bar{\Omega})$, where $\sum_{j,k=1}^n a_{jk}(x)\xi_j\xi_k \ge \lambda|\xi|^2$, $\lambda = const > 0, \forall x \in \Omega, \forall \xi \in \mathbf{R}^n, \sum_{j,k=1}^n |a_{jk}(x)| + \sum_{i=1}^n |b_i(x)| + |c| \le M, M = const > 0, x \in \Omega$.

Theorem 2.2 from [7] assumes that $c(x) \le -c_*$ for some $c_* > 0, \forall x \in \Omega$ and u is non-constant subsolution $L(u) \ge 0$, attaining its non-negative maximum at $\hat{x} \in \partial\Omega$, where $\frac{\partial u}{\partial n}(\hat{x})$ exists. Then there exists a positive constant $\gamma(c_*, \lambda, M, \Omega)$ and such that $\frac{\partial u}{\partial n}(\hat{x}) > \gamma u(\hat{x})$. In the proof of Proposition 3, we rely heavily on Harnack inequality (29) for positive solutions of the Cordes-type elliptic PDE (4), while the key ingredient in the proof of Theorem 2.2 is Feynman–Kac representation formula of a subsolution to the elliptic equation (30).

References

1. Brezis, H., Cabré, X.:Some simple nonlinear PDE's without solutions, Bolletino U.M.I., (8) 1-B, 1998, 223-262.
2. Cordes, H.O.: Die erste Randwerafgabe bei Differentialgleichungen Zweiter ordnung in mehr als zwei Variabeln, Math.Ann. Bd. 131, Hit 3(156), 278-318.
3. Cordes, H.O.: Zero order a-priori estimates for solutions of elliptic differential equations, Proc. Sympos. Pure math. 4, 1961, 157-166.
4. Gazzola, F., Grunau, H.-Ch., Sweers, G.: Positivity preserving and nonlinear higher order elliptic equations in bounded domains, Lecture Notes in Math., v.1991, Springer, 2010.
5. Gillbarg D., Trudinger N.: Elliptic partial differential equations of second order, Grundlehren der Math. Wissenschafter, 224, Springer, Berlin, 1983.
6. Han, Q., Lin, F.: Elliptic partial differential equations, Second edition, Courant Lecture Notes in Math. 1, AMS, 2011.
7. Komorowski, T., Bobrowski, A.: A quantitative Hopf maximum principle for subsolutions of elliptic PDE's, Discrete and continuous dynamical systems series S, 13:12, 2020.
8. Landis, E.:Elliptic and Parabolic Equations of Second Order, Translations of Mathematical Monographs, AMS, v.171, 1997.
9. Moser, J.: On Harnack's theorem for elliptic differential equations, Comm. Pure Appl. Math., 14:3, 1961, 577-591.
10. Popivanov, P.: Explicit formulas to the solutions of Dirichlet problem for equations arising in geometry and physics, C.R.Acad.Bulg.Sci., 68:1, 2015, 19-24.
11. Protter, M., Weinberger H.: Maximum Principles in Differential Equations, Prentice Hall Inc., Englewood Cliffs N.J., 1967.
12. Serrin, J.: On the Harnack inequality for linear elliptic equations, Journal d'Analyse Math., 4:2, 1954, 292-308.
13. Stampacchia, L.G.: Le probléme de Dirichlet pour les équations elliptiques du 2nd ordre á coefficients discontinues, Ann. Inst. Fourier, Grenoble, 15:1, 1965, 189-258.

Global Existence Result of Solutions for a Riser Equation with Logarithmic Source Term and Damping Term

Nazlı Irkıl and Erhan Pişkin

Abstract This work deals with the influence of the logarithmic source term on solutions to quasilinear riser equation with nonlinear damping term. We established global results of solutions with negative initial energy.

Keywords Global existence · Second-order equation · Logarithmic nonlinearity

1 Introduction

In this work, we consider the following riser equation with logarithmic nonlinearity:

$$
\begin{cases}
|u_t|^\mu u_{tt} + u_t\,|u_t|^{k-1} + 2\beta u_{xxxx} - 2\left[(ax+b)\,u_x\right]_x + \dfrac{\beta}{3}\left(u_x^3\right)_{xxx} & \\
\quad - \left[(ax+b)\,u_x^3\right]_x - \left(\beta u_{xx}^2 u_x\right)_x = |u|^{p-2}\,u\ln|u|, & (x,t) \in [0,1] \times (0,T), \\
u\,(0,t) = u\,(1,t) = u_{xx}\,(0,t) = u_{xx}\,(1,t) = 0, t \in (0,T), & \\
u\,(x,0) = u_0\,(x), \quad u_t\,(x,0) = u_1\,(x), & x \in [0,1],
\end{cases}
\tag{1}
$$

where a, b, α, β are nonnegative constants. The exponents were taken $\mu \geq 0$, $0 \leq k \leq \min\{\mu+1, 1\}$ and $p \geq 2$.

When there is no logarithmic source term in (1), the problem becomes such that

$$
u_{tt} + \alpha u_t + 2\beta u_{xxxx} - 2\left[(ax+b)\,u_x\right]_x + \frac{\beta}{3}\left(u_x^3\right)_{xxx} - \left[(ax+b)\,u_x^3\right]_x
$$
$$
- \left(\beta u_{xx}^2 u_x\right)_x = f\,(u).
\tag{2}
$$

The problem (2) which was discussed by many authors (see [1, 3, 5, 7, 12]) is related to the dynamics of a riser vibrating due to effects of current and waves [10, 13].

N. Irkıl (✉) · E. Pişkin
Dicle University, Diyarbakır 21280, Turkey
e-mail: nazliirkil@gmail.com

In recent years, hyperbolic wave equations with logarithmic source term were discussed by many mathematicians. Moreover, there have been some works on the logarithmic source term (see [2, 4, 9, 11]). To the best of our knowledge, in the absence of the damping term and taking $\mu = 0$ in the problem (1) was studied by [8]. The aim of this work is to have a global existence of solutions to the problem (1). More precisely, we consider the effect of the damping term on the quasilinear riser problem with the logarithmic source term.

2 Preliminaries

This section includes some definitions and lemmas which will be used in the proof of our results. We denote $(., .)_2$ the inner product in $L^2 = L^2 [0, 1]$ and $\|.\|_p$ is the norm in $L^p = L^p [0, 1]$.

The total energy functional was defined as

$$E(t) = \frac{1}{\mu + 2} \|u_t\|_{\mu+2}^{\mu+2} + \beta \|u_{xx}\|^2 + \int_0^1 (ax + b) u_x^2 dx + \frac{\beta}{2} \|u_x u_{xx}\|^2$$

$$+ \frac{1}{4} \int_0^1 (ax + b) u_x^4 dx + \frac{1}{p^2} \|u\|_p^p - \frac{1}{p} \int_0^1 u^p \ln |u| \, dx. \tag{3}$$

By using initial condition and (3), the initial energy functional can be considered, such that

$$E(0) = \frac{1}{\mu + 2} \|u_1\|_{\mu+2}^{\mu+2} + \beta \|u_{0xx}\|^2 + \int_0^1 (ax + b) u_{0x}^2 dx + \frac{\beta}{2} \|u_{0x} u_{0xx}\|^2$$

$$+ \frac{1}{4} \int_0^1 (ax + b) u_{0x}^4 dx + \frac{1}{p^2} \|u_0\|_p^p - \frac{1}{p} \int_0^1 |u_0|^p \ln |u_0| \, dx. \tag{4}$$

Definition 1 A function $u \in C \left([0, T], H_0^2 [0, 1]\right)$ is said to be a weak solution of (1) on $[0, T]$ if

$$
\int_0^1 |u_t|^\mu u_{tt} w dx + \int_0^1 u_t |u_t|^{k-1} w dx + \int_0^1 2\beta u_{xx} w_{xx}
$$

$$
+ \int_0^1 2 \left[(ax + b) u_x \right] w_x - \int_0^1 \frac{\beta}{3} \left(u_x^3 \right)_{xx} w_x
$$

$$
+ \int_0^1 \left[(ax + b) u_x^3 \right] w_x + \int_0^1 \left(\beta u_{xx}^2 u_x \right) w_x \tag{5}
$$

$$
= \int_0^1 |u|^{p-2} u \ln |u| \, w, \%
$$

Lemma 1 *Assume that u is the solution for the problem (1). Then the $E(t)$ is decreasing fo $t > 0$ and*

$$
E'(t) = - \|u_t\|_k^k \leq 0. \tag{6}
$$

Proof We multiply both sides of the equation (1) by u_t and then integrating from 0 to 1, we obtain (6).

3 Global Existence

In this part, we state the global existence of the problem (1).

Theorem 1 *Let $u_0 \in H_0^2 [0, 1]$ and $u_1 \in L^2 [0, 1]$ hold. Assume further that $\int_0^1 u^p \ln |u| \, dx \leq 0$, $\|u_t\|_{\mu+2} > 1$ and $E(0) < 0$. Then, the problem (1) has a global weak solution.*

Proof We define a functional, such that

$$
B(t) = E(t) + \frac{1}{p} \int_0^1 u^p \ln |u| \, dx - \frac{1}{p^2} \int_0^1 u^p dx. \tag{7}
$$

Thanks to definition of the $E(t)$, (7) can be rewritten as

$$
B(t) = \frac{1}{\mu + 2} \|u_t\|_{\mu+2}^{\mu+2} + \beta \|u_{xx}\|^2 + \int_0^1 (ax + b) u_x^2 dx
$$

$$
+ \frac{\beta}{2} \|u_x u_{xx}\|^2 + \frac{1}{4} \int_0^1 (ax + b) u_x^4 dx
$$

$$
\geq 0. \tag{8}
$$

By taking derivative of equality (7), it yields that

$$B'(t) = -\|u_t\|_{k+1}^{k+1} + \int_0^1 u^{p-1} u_t \ln |u|\, dx + \frac{1}{p} \int_0^1 u^{p-1} u_t dx - \frac{1}{p} \int_0^1 u^{p-1} u_t dx$$

$$= -\|u_t\|_{k+1}^{k+1} + \int_0^1 u^{p-1} u_t \ln |u|\, dx$$

$$\leq -\|u_t\|_{k+1}^{k+1} + \int_0^1 u^p u_t dx. \tag{9}$$

By using basic inequality $x > \ln |x|$, Hölder and Young's inequality, (9) was obtained such that

$$\left| B'(t) \right| < \int_0^1 u^p u_t dx$$

$$\leq \|u\|_p^{\frac{kp}{(k+1)}} \|u_t\|_{k+1}$$

$$\leq \frac{1}{4\theta^2} \|u\|_p^p + \theta^2 \|u_t\|_{k+1}^{k+1} \tag{10}$$

where θ was defined as a positive constant.

By using (7), the inequality (10) is obtained as

$$\left| B'(t) \right| \leq \frac{p^2}{4\theta^2} \left[E(t) - B(t) + \frac{1}{p} \int_0^1 u^p \ln |u|\, dx \right] + \theta^2 \|u_t\|_{k+1}^{k+1}. \tag{11}$$

Thanks to $E(0) < 0$ and Lemma 1, we know

$$\frac{dE(t)}{dt} = -\|u_t\|_{k+1}^{k+1} \leq 0, \tag{12}$$

$$E(t) \leq E(0) < 0. \tag{13}$$

By combination of (8) and (13) and using the condition $\int_0^1 u^p \ln |u|\, dx \leq 0$, we have

$$E(t) - B(t) + \frac{1}{p} \int_0^1 u^p \ln |u|\, dx \leq 0. \tag{14}$$

Therefore, (11) becomes

$$\left| B'(t) \right| \leq \theta^2 \|u_t\|_{k+1}^{k+1}. \tag{15}$$

On the other hand, from (8), we conclude that

$$(\mu + 2) B(t) = \|u_t\|_{\mu+2}^{\mu+2} + (\mu + 2) \beta \|u_{xx}\|^2 + (\mu + 2) \int_0^1 (ax + b) u_x^2 dx$$

$$+ \frac{(\mu + 2)}{2} \|u_x u_{xx}\|^2 + \frac{(\mu + 2)}{4} \int_0^1 (ax + b) u_x^4 dx.$$

Consequently, we have

$$\|u_t\|_{\mu+2}^{\mu+2} \leq (\mu + 2) B(t). \tag{16}$$

We can suppose that $\|u_t\|_{\mu+2} \geq \|u_t\|_{k+1} > 1$, then, by $k \leq \mu + 1$ and embedding theorem, we can conclude that

$$\|u_t\|_{k+1}^{k+1} \leq C \left(\|u_t\|_{\mu+2}^{\mu+2} \right)^{\frac{k+1}{\mu+2}} \leq C \|u_t\|_{\mu+2}^{\mu+2} \tag{17}$$

where C is positive constant.

Making using of the (15)–(17), we get

$$\left| B'(t) \right| \leq \theta^2 \|u_t\|_{k+1}^{k+1} \leq C\theta^2 \|u_t\|_{\mu+2}^{\mu+2} \leq (\mu + 2) \theta^2 B(t), \tag{18}$$

where $k \leq \mu + 1$.

Therefore, we have

$$\left| B'(t) \right| \leq (\mu + 2) \theta^2 B(t). \tag{19}$$

It is clear that

$$B(t) \exp\left(-(\mu + 2) \theta^2 t \right) < B(t) < B(t) \exp\left((\mu + 2) \theta^2 t \right). \tag{20}$$

With the (20), the definition of $B(t)$ and continuation principle, the proof was completed.

References

1. Bayrak, V., Can, M.: Global nonexistence of solutions of the vibrations of a riser. Math. Comput. Appl. 2(1), 45-52 (1997). https://doi.org/10.3390/mca2010045
2. Bialynicki-Birula, I., Mycielski, J.: Wave equations with logarithmic nonlinearities. Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys. 23(4), 461-466 (1975).

3. Çelebi, A.O., Gür, Ş., Kalantaro v, V.K.:Structural stability and decay estimate for marine riser equations. Math. Comput Model. 54(11-12), 3182-3188 (2011). https://doi.org/10.1016/j.mcm.2011.08.014

4. Chen, Y., Xu, R.: Global well-posedness of solutions for fourth order dispersive wave equation with nonlinear weak damping, linear strong damping and logarithmic nonlinearity. Nonlinear Anal. 192, (2020). https://doi.org/10.1016/j.na.2019.111664

5. Esquivel-Avila, J.A.: On nonexistence of global solutions of a quasilinear riser equation. Adv. Pure Appl. Math. 9(3), 195-203 (2018). https://doi.org/10.1515/apam-2017-0048

6. Gorka, P.: Logarithmic Klein-Gordon equation. Acta Phys. Pol. B. 40(1),(2009).

7. Hao, J., Li, S., Zhang, Y. Blow up and global solutions for a quasilinear riser problem. Nonlinear Anal. 67(3), 974-980 (2007). https://doi.org/10.1016/j.na.2006.06.042

8. N. Irkıl, N., Pişkin, E.: Blow up and global existence of solution for a riser problem with logarithmic nonlinearity. Sigma J. Eng. and Nat. Sci. 39(5), 56-63 (2021).

9. N. Irkıl, N., Pişkin, E., Agarwal, P.: Global existence and decay of solutions for a system of viscoelastic wave equations of Kirchhoff type with logarithmic nonlinearity. Math. Methods Appl. Sci. 45(5), 2921-2948 (2022). https://doi.org/10.1002/mma.7964

10. Köhl, M.: An extended Liapunov approach to the stability assessment of marine risers. Z. Angew. Math. Mech. 73(2), 85-92 (1993). https://doi.org/10.1002/zamm.19930730208

11. Liu, G.: The existence, general decay and blow-up for a plate equation with nonlinear damping and a logarithmic source term. Electron. Res. Arch. 28(1), 263-289 (2020). 10.3934/era.2020016

12. Meyvacı, M.: Energy decay rate of the solutions of a marine riser equation with a variable coefficient. Commun. Fac. Sci. Univ. Ank. Ser. A1. Math. Stat. 67(1), 286-292 (2018).

13. Sun, M.G.: The stress boundary layers of a slender riser in a steady flow. Adv. Hydrodyn. 4, 2-43 (1986).

Parabolic Equations with Causal Operators

Tzanko Donchev, Dimitar Kolev, Mihail Kolev, and Alina I. Lazu

Abstract We study an initial boundary value problem (IBVP) with a parabolic PDE with a nonlinear reaction function $F(t, x, p, \alpha)$, where $\alpha = (Qu)(t, x)$, Q being a causal operator (CO), and F is Hölder in t and x, continuous in p, Lipschitz in α, and monotone increasing in (p, α). We state that under the above-mentioned conditions and assuming in addition that F contains a locally Lipschitz summand, then the solution can be qualitatively estimated. Moreover, this solution blows up in some t-interval. To this end, we use the monotone iterative method.

Keywords Parabolic equation · Causal operators · Monotone-iterative technique

1 Introduction

Consider the parabolic partial differential equation in the general form

$$\frac{\partial u}{\partial t} = F(t, x, u, (Qu)(t, x)) \quad \text{in} \quad I \times \Omega \subset \mathbf{R} \times \mathbf{R}^n, \tag{1}$$

with the initial condition $u(t, x) = \varphi(x)$, $x \in \mathbf{R}^n$, $t \in [-h, 0]$ $(h > 0)$, where $Q : \mathcal{B} \to \mathcal{B}_1$ is a causal operator, $\mathcal{B}$ and $\mathcal{B}_1$ being Banach spaces, the function G is

T. Donchev (✉)
Department of Mathematics UACG, 1 Hr. Smirnenski Bvd, 1046 Sofia, Bulgaria
e-mail: tzankodd@gmail.com

D. Kolev
Abdus Salam School of Mathematical Sciences, 68-B, New Muslim Town, Lahore, Pakistan

M. Kolev
Institute of Metal Science, Equipment and Technologies with Hydroaerodynamics Centre, Bulgarian Academy of Sciences, 67 Shipchenski prohod Str., 1574 Sofia, Bulgaria

A. I. Lazu
Department of Mathematics, "Gh. Asachi" Technical University, Iasi 700506, Romania

169

continuous w.r.t. all arguments in the domain $\tilde{\Omega} = I \times \Omega$, $I \subset \mathbf{R}$ is an interval such that $[-h, 0] \subset I$, and Ω is a bounded domain with a smooth boundary $\partial\Omega$.

The theory of CO started with the work of C. Corduneanu [5]. Some evolution processes can be modeled by PDEs in the form $F(t, x, u(t, x), Qu(t, x))$. Here, the unknown function u takes part as an argument of F depending on t and x and secondly as transformed by a causal operator Q. An example of the casual operator is the delay (deviation), that is, $Qu = u(t - \sigma)$, where $\sigma > 0$. Here, u depends smoothly on the time moment $t - \sigma$, where the delay of time is expressed by σ. The same function u is taken in a position at σ units back, that is to say, as though the equation under consideration has past memory. Such an equation is called functional differential equation (see, e.g., [2, 6]). It turns out that the general class of CO includes the Niemytzki operator, Volterra integral operator, Fredholm operator (see, e.g., [5]), and also the problems with "maxima" (see [3]). Another type of CO is time lag, comprehensively studied in [4, 8, 9].

In this paper, we first show a technique for the existence of solutions to (1). Given the Robin boundary condition $Bu = h(t, x)$ imposed on the boundary ∂D, where B is defined by $B \equiv \alpha_0(t, x)\dfrac{\partial}{\partial \nu} + \beta_0(t, x)$, $\alpha_0(t, x)$ and $\beta_0(t, x)$ are nonnegative Hölder continuous functions for $t \in [t_0, T)$. The initial condition is given by $u_0(x) \equiv u(t_0, x)$ in Ω, where t_0 is the initial time, for instance $t_0 = 0$. The idea of the existence and uniqueness of solutions for a similar class of PDEs of parabolic type with delay can be seen in [11] (see, e.g., [2, 6]). Some methods such as quasi-linearization, comparison, and some modifications by which one is able to investigate the existence and uniqueness, stability, and decay for nonlinear parabolic problems are given in [1, 6, 10, 11]. Here, we are interested in the behavior of the solutions of the IBVP for (1) including the blow-up property, under Hölder continuous boundary and initial data.

2 Preliminaries

Let us introduce the function space $E = E([0, T), \mathbf{R}^n)$ with a fixed number $T > 0$. It contains all maps acting from $[0, T)$ into $\mathbf{R}^n$ with the norm $\|\cdot\|_c$ by which we obtain a certain topology. Let $Q : E \to E$ be a map which we call an operator on E.

Definition 1 Q is a causal operator if for each pair $\{x(\cdot), y(\cdot)\} \in E$ such that $x(s) = y(s)$ for $0 \le s \le t$, the following equality holds:

$$(Qx)(s) = (Qy)(s) \text{ for } 0 \le s \le t, \tag{2}$$

with $t < T$ arbitrary.

Example 1 Consider the spaces

$$\Omega_1 = \{u : u = \phi(t)(x^2 y + \lambda_1(t)xy) + a_1(t)x + b_1(t)y + c_1(t), \ (x, y) \in M \subset \mathbf{R}^2\},$$
$$\Omega_2 = \{v : v = \phi(t)(\tfrac{y^3}{3} + \lambda_2(t)xy) + a_2(t)x + b_2(t)y + c_2(t), \ (x, y) \in M \subset \mathbf{R}^2\},$$

where $M = X \cup Y$, $X = \{(x, y) : \sqrt{3}x + y = 0\}$, $Y = \{(x, y) : \sqrt{3}x - y = 0\}$, and $\phi(t)$, $\lambda_i(t)$, $a_i(t)$, $b_i(t)$, and $c_i(t)$ $(i = 1, 2)$ are arbitrary functions, continuous on some open interval $J \subset \mathbf{R}$. Let us construct the space $\Omega = \Omega_1 \cup \Omega_2$. Consider the differential operator $L = D(t, x, y)(\partial_{xx}^2 + \partial_{yy}^2)$, where $D(t, x, y)$ is an arbitrary function continuous in $J \times \mathbf{R}^2$.

Now we may take an element $z \in \tilde{\Omega} \subset \Omega$, where $\lambda_1(t) = \lambda_2(t)$, $a_1(t) = a_2(t)$, $b_1(t) = b_2(t)$, $c_1(t) = c_2(t)$ and then see that $Lz(t, x, y)$ satisfies the stated property in Definition 1. Hence, we may conclude that L is a causal operator.

Definition 2 Let $0 < \theta < 1$ be a fixed number. Then the map $z : \Theta \subset \mathbf{R}^{n+1} \to \mathbf{R}$ is said to be θ Hölder (with a constant κ) if $|z(x) - z(y)| \le \kappa \|x - y\|^\theta$ for any $x, y \in \Theta$. We write $z \in C^{1+\theta}$ if $z(\cdot)$ admits partial derivatives which are θ Hölder.

Introduce the notations $D_T \equiv [0, T) \times \Omega$, $S_T \equiv (0, T] \times \partial\Omega$, $D_{-\sigma} \equiv [-\sigma, 0] \times \Omega$, $E_T \equiv [-\sigma, T) \times \bar{\Omega}$, $\mathcal{E} = C(D_{-\sigma}, \mathbf{R})$, $\mathcal{E}^+ = C(D_{-\sigma}, \mathbf{R}^+)$.

Consider the parabolic IBVP with the causal operator Q in the form

$$
\begin{array}{lrl}
(a) & \frac{\partial u}{\partial t} - Lu = F(t, x, u, Qu(t, x)) & \text{in} \ \ D_T, \\
(b) & Bu = h(t, x) & \text{on} \ \ S_T, \\
(c) & u(t, x) = \eta_0(t, x) & \text{in} \ \ D_{-\sigma}, \\
(d) & \eta_0(0, x) = u_0(x) & \text{in} \ \ \Omega,
\end{array}
\tag{3}
$$

where $\sigma > 0$ is a given number, $\eta_0(t, x) \in \mathcal{E}^+$ is a given Hölder continuous function in $D_{-\sigma}$, $u_0(\cdot) \in C^\theta(\bar{\Omega})$. Moreover, $h(t, x) \ge 0$ is assumed to be in the class $C^{1+\theta}(S_T)$. We note that the operator

$$L \equiv \sum_{i,j=1}^n a_{ij}(t, x) \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} + \sum_{j=1}^n b_j(t, x) \frac{\partial}{\partial x_j} \tag{4}$$

in (3)(a) is uniformly elliptic in the sense that the matrix $\{a_{ij}(t, \cdot)\}$ is positive definite on $\bar{\Omega}$ and uniformly on t. We assume that the coefficients of L are in the class $C^{1+\theta}(\bar{\Omega})$ $(0 < \theta < 1)$. The boundary operator B is defined by $B \equiv \alpha_0(t, x)\frac{\partial}{\partial v} + \beta_0(t, x)$, where $\alpha_0(t, x)$ and $\beta_0(t, x)$ are nonnegative functions in $C^{1+\theta}(\partial\Omega)$ for $t \in [0, \infty)$ and not identically zero on $[0, \infty) \times \partial\Omega$; $\partial/\partial v$ is the outward normal derivative on $\partial\Omega$. Next we assume that F has the form $F = f(t, x, u) + R(t, x, Qu)$, then both functions f and R are Hölder continuous in $D_T \times \mathbf{R}$. In addition, $f(t, x, \cdot)$ and $R(t, x, \cdot)$ are C^1-functions. The solution of (3) is a function $u \in C^{1,2}(D_T)$.

Everywhere in this paper, we assume that the problem (3) satisfies the compatibility condition $\left[\alpha_0(0, x)\dfrac{\partial u_0}{\partial \nu} + \beta_0(0, x)u_0 \right]_{x \in \partial\Omega} = h(0, x)\Big|_{x \in \partial\Omega}$.

Definition 3 A function $\tilde{u} \in C^\theta \cap C^{1,2}(D_T)$ is called an upper solution of IBVP (3) if

> (a) $\frac{\partial \tilde{u}}{\partial t}(t, x) \geq L\tilde{u}(t, x) + F(t, x, u, Qu(t, x))$ in D_T,
> (b) $B\tilde{u}(t, x) \geq h(t, x)$ on S_T,
> (c) $\tilde{u}(t, x) \geq \eta_0(t, x)$ in $D_{-\sigma}$.

Similarly, $\hat{u} \in C^\theta \cap C^{1,2}(D_T)$ is called a lower solution of (3) if it satisfies the reversed inequalities.

Definition 4 A pair $\{\tilde{u}(t, x), \hat{u}(t, x)\}$ is called ordered if $\tilde{u}(t, x) \geq \hat{u}(t, x)$ in E_T. Then the set of all functions $u = u(t, x)$ such that $\hat{u}(t, x) \leq u(t, x) \leq \tilde{u}(t, x)$ in E_T is denoted by $\langle \hat{u}, \tilde{u} \rangle$ and is called sector.

Next we will make use of a standard elliptic boundary value problem. Let $\lambda_0 > 0$ and $\Phi(x)$, always normalized by $\max \{\Phi(x) : x \in \overline{\Omega}\} = 1$, in $\overline{\Omega}$ be the principal eigenvalue and correspondent normalized eigenfunction, respectively, to

$$-Lu = \lambda u \text{ in } \Omega, Bu = 0 \text{ on } \partial\Omega. \tag{5}$$

Define the following linear differential operator of parabolic type:

$$\mathcal{L}_c \equiv \left(\frac{\partial}{\partial t} - L + c \right) \text{ in } (0, T] \times \mathbf{R}^n, \tag{6}$$

where T and L are the same as those in (3) and (4), respectively, and $c = c(t, x)$ is a bounded function in $(0, T] \times \mathbf{R}^n$. Now we consider the following IBVP:

> (a) $\mathcal{L}_c v = q(t, x)$ in D_T,
> (b) $Bv = h(t, x)$ on S_T, $\tag{7}$
> (c) $v(0, x) = u_0$ in Ω,

where $\mathcal{L}_c$ is defined by (6), $Bu \equiv \alpha_0(t, x)\frac{\partial u}{\partial \nu} + \beta_0(t, x)u, \alpha_0(t, x) \geq 0, \beta_0(t, x) \geq 0$, $\alpha_0(t, x) + \beta_0(t, x) > 0$, and $h \geq 0$ on S_T. Assume that the coefficients of L, q, h, u_0, and c are Hölder continuous in their respective domains. Then due to the above-stated assumptions, there exists a positive constant $C = C(T) > 0$ such that

$$|F(t, x, u_1, \alpha_1(t, x)) - F(t, x, u_2, \alpha_2(t, x))| \leq C(|u_1 - u_2| + |\alpha_1 - \alpha_2|). \tag{8}$$

Let the number σ be $0 < \sigma < 1$. We formulate the following hypothesis.

(H) Suppose that

(i) $F(t, x, p, \alpha)$ is Hölder in $(t, x) \in D_T$, uniformly $\forall (p, \alpha) \in \mathbf{R}^+ \times \mathbf{R}^+$ and globally nonnegative;

(ii) $F(t, x, p, \alpha)$ is C^1-function in $p \in \mathbf{R}^+$, uniformly $\forall\ (t, x, \alpha) \in D_T \times \mathbf{R}^+$;

(iii) $F(t, x, p, \alpha)$ is Lipschitz in $\alpha \in \mathcal{E}^+$, uniformly $\forall\ (t, x, p) \in D_T \times \mathbf{R}^+$;

(iv) $F(t, x, p, \alpha)$ is monotone increasing in $(p, \alpha) \in \mathbf{R}^+ \times \mathcal{E}^+$.

The last implies that if $p \geq q$ and $\alpha(t, x) \geq \beta(t, x)$, then $F(t, x, p, \alpha) \geq F(t, x, q, \beta)$. Next, we need the following auxiliary results.

Lemma 1 *Let (H) be satisfied. Then there exists a pair of ordered upper and lower solutions of (3).*

Proof From [7, Theorem 7, Ch. 2], we know that the problem (7) admits a unique solution $z(t, x)$. Clearly, z is a lower solution of (3). Let $M = \|\eta_0\|_{\mathcal{E}}$ and fix $N > 0$. As before due to [7, Theorem 7], with $F = F(t, x, M + N, M + N)$ the considered problem admits a unique solution $w(t, x)$ on E_T. Since $w(\cdot, \cdot)$ is continuous, one has that there exists $T_1 > 0$ such that $|w(t, x)| \leq M + N$ on E_{T_1}. Clearly, $w(\cdot, \cdot)$ is an upper solution of (3). Let $t > T_1$. Replacing t by t_θ, we consider again (3) on E_T. Using the same method, one can prove the existence of the upper solution of (3) with changed initial condition on $[0, T_2]$ for some $T_2 > 0$. By virtue of Zorn's lemma, there exists a maximal interval $[0, T)$ with $0 < T \leq +\infty$, where the upper solution w is defined. The proof is therefore complete. $\square$

Lemma 2 *Let $u^{(k)}$ be a lower solution, and $v^{(k)}$ be an upper for (3) such that the following iteration process for lower solution holds:*

$$
\begin{aligned}
&(a)\ \frac{\partial u^{(k)}}{\partial t} = Lu^{(k-1)} + F(t, x, u^{(k-1)}, (Qu^{(k-1)})(t, x))\ \ in\ \ D_T,\\
&(b)\ Bu^{(k)} = h(t, x)\ \ on\ \ S_T,\\
&(c)\ u^{(k)}(t, x) = \eta_0(t, x)\ \ in\ \ D_{-\sigma},\\
&(d)\ \eta_0(0, x) = u_0(x)\ \ in\ \ \Omega,
\end{aligned}
\tag{9}
$$

and similar to upper solutions v^k. Then the following inequalities hold:

$$
u^{(0)} \leq \cdots \leq u^{(k)} \leq u^{(k+1)} \leq \cdots \leq v^{(k+1)} \leq v^{(k)} \leq \cdots \leq v^{(0)}.
\tag{10}
$$

Proof First, assume that $u^{(k)} \geq u^{(k-1)}$, and then set $u^{(k)} - u^{(k-1)} \equiv y^{(k-1)} \geq 0$. We should prove that $y^{(k)} \geq 0$. Therefore, from the monotonicity of F follow the inequalities

$$
\begin{aligned}
&\partial y^{(k)}/\partial t = Ly^{(k-1)} + F(t, x, u^{(k)}, (Qu^{(k)})(t, x)) - -F(t, x, u^{(k-1)}, u_t^{(k-1)}(x)) =\\
&Ly^{(k-1)} + F(t, x, u^{(k)}, (Qu^{(k)})(t, x)) - F(t, x, u^{(k-1)}, (Qu^{(k)})(t, x)) +\\
&+F(t, x, u^{(k-1)}, (Qu^{(k)})(t, x)) - -F(t, x, u^{(k-1)}, (Qu^{(k-1)})(t, x)) \geq\\
&\geq Ly^{(k-1)} + F(t, x, u^{(k-1)}, (Qu^{(k)})(t, x)) - -F(t, x, u^{(k-1)}, (Qu^{(k-1)})(t, x))\\
&\geq Ly^{(k-1)}.
\end{aligned}
\tag{11}
$$

Obviously, $By^{(k)} = 0$, $y^{(k)}(0, x) = 0$ and having in mind Corollary 2.2.1 in [10], we conclude that $y^{(k)} \geq 0$ and $u^{(k+1)}(t, x) \geq u^{(k)}(t, x)$. Next, we consider $v^{(k)}$ and assume that $v^{(k)} \leq v^{(k-1)}$. Then setting $v^{(k)} - v^{(k-1)} \equiv z^{(k-1)} \leq 0$ get the inequalities

$$
\begin{aligned}
\partial z^{(k)}/\partial t &= Lz^{(k-1)} + F(t, x, v^{(k)}, (Qv^{(k)})(t, x)) \\
&\quad -F(t, x, v^{(k-1)}, (Qv^{(k-1)})(t, x)) = Lz^{(k-1)} + F(t, x, v^{(k)}, (Qv^{(k)})(t, x)) \\
&\quad -F(t, x, v^{(k-1)}, (Qv^{(k)})(t, x)) + F(t, x, v^{(k-1)}, (Qv^{(k)})(t, x)) \\
&\quad -F(t, x, v^{(k-1)}, (Qu^{(k)})(t, x)) \le Lz^{(k-1)}.
\end{aligned}
\tag{12}
$$

Obviously, $Bz^{(k)} = 0$, $z^{(k)}(0, x) = 0$, and by the same argument as it was before we conclude that $z^{(k)} \le 0$. Hence $v^{(k+1)}(t, x) \le v^{(k)}(t, x)$. Now assume that $u^{(k)} \le v^{(k)}$. Therefore, we have to prove that $w^{(k+1)} \equiv v^{(k+1)} - u^{(k+1)} \ge 0$. To this end, we use the same argument as before

$$
\begin{aligned}
\partial w^{(k+1)}/\partial t &= Lw^{(k)} + F(t, x, v^{(k)}, (Qv^{(k)})(t, x)) \\
&\quad -F(t, x, u^{(k)}, (Qu^{(k)})(t, x)) \ge Lw^{(k)}.
\end{aligned}
\tag{13}
$$

It is clear that $Bw^{(k+1)} = 0$, $w^{(k)}(0, x) = 0$, hence $w^{(k+1)} \ge 0$ and $v^{(k+1)}(t, x) \ge u^{(k+1)}(t, x)$ by which the proof is complete.   □

Having in mind the above stated lemmas, it follows.

Corollary 1 *There exist the sequences:*

$$
\begin{aligned}
u^{(0)} &\le u^{(1)} \le \cdots \le u^{(m-1)} \le u^{(m)} \le \cdots \\
&\le v^{(m)} \le v^{(m-1)} \le \cdots \le v^{(1)} \le v^{(0)},
\end{aligned}
\tag{14}
$$

such that $u^{(m)}(t, x) \to u(t, x)$ *and* $v^{(m)}(t, x) \to v(t, x)$ *as* $m \to \infty$. *Moreover,* $u(t, x) \equiv v(t, x)$ *is the unique solution of* (3).

The next statement concerns the existence and uniqueness of the solution of (3).

Theorem 1 *Suppose that* **(H)** *holds and the reaction function* F *has the general form* $F = F(t, x, u, u_t)$. *Then there exists a unique solution of* (3).

Proof From the above-stated results in Lemmas 1, 2, and Corollary 1, we conclude that there exist the sequences (14) $u^{(0)} \le u^{(1)} \le \cdots \le u^{(m-1)} \le u^{(m)} \le \cdots v^{(m)} \le v^{(m-1)} \le \cdots \le v^{(1)} \le v^{(0)}$ having limits $u \equiv v$. Further, $u \equiv v$ is the unique solution of the problem under consideration.   □

3 Main Result

We assume that the condition **(H)** holds and suppose in addition the following:

   (H1) The reaction function has the form $F = f(t, x, u) + R(t, x, Qu)$, where $f(t, x, \cdot)$ is locally Lipschitz in t and x, i.e., it belongs to $L_{loc}(\mathbf{R}^+)$ (the set of all locally Lipschitz functions on $\mathbf{R}^+$), while $R(t, x, \cdot)$ is bounded on the bounded sets, where $(t, x) \in D_T$.
   Since $Qu \in \mathcal{B}_1$, we use the norm $\|Qu\| = \max_{(t,x) \in D_T} |Qu(t, x)|$.

Theorem 2 *Suppose that **(H1)** holds and $u(t, x)$ is the nonnegative solution of (3).*
If there exist constants $\alpha > 0$, $\beta > 0$, and $A > 0$ such that

$$
\begin{aligned}
(a) \quad & f(t, x, \eta) \geq \left(\frac{\lambda_0}{2} + \alpha + \frac{A}{2}\right)\eta, \quad \forall \eta \geq 0, \\
(b) \quad & R(t, x, Qu(t, x)) \geq \left(\frac{\lambda_0}{2} + \beta + \frac{A}{2}\right)\|Qu\|,
\end{aligned}
\tag{15}
$$

where $Qu(t, x) \in C([-\sigma, 0] \times \Omega, R)$, then for any $\delta > 0$ and $\eta_0(t, x) \geq \delta$ in $D_{-\sigma}$,
the solution of (3) satisfies the inequality:

$$
u(t, x) \geq \delta e^{(\alpha+\beta+A)t}\Phi(x).
\tag{16}
$$

Proof First, we define the function $\hat{u} \equiv \delta e^{(\alpha+\beta+A)t}\Phi(x)$ and verify that it is a lower
solution of (3). According to the definition of lower solution, we have to prove that
the quantity on the left-hand sides of (3) is less or equal to those on the right-hand
sides for $\hat{u}$ instead of u. From the elliptic boundary value problem (5) with $h \geq 0$,
we infer that $\hat{u}$ can be a lower solution of (3) if the following estimates hold true:

$$
\begin{cases}
\left(\dfrac{\lambda_0}{2} + \alpha + \dfrac{A}{2}\right)\delta e^{(\alpha+\beta+A)t}\Phi(x) \leq f(t, x, \delta e^{(\alpha+\beta+A)t}\Phi(x)), \\
\left(\dfrac{\lambda_0}{2} + \beta + \dfrac{A}{2}\right)\delta\|\eta(x)\| \leq R(t, x, \delta\eta(x)),
\end{cases}
\tag{17}
$$

where $\eta(x)(s) = e^{(\alpha+\beta+A)s}\Phi(x)$ for every $s \in [-\sigma, 0]$. Taking into account (3)(a)
and in view of (15), we get

$$
\begin{aligned}
\frac{\partial\hat{u}}{\partial t} - L\hat{u} &= \delta(\alpha + \beta + A)e^{(\alpha+\beta+A)t}\Phi(x) + \delta\lambda_0\Phi(x)e^{(\alpha+\beta+A)t} \\
&= \delta\Phi(x)e^{(\alpha+\beta+A)t}(\alpha + \beta + A + \lambda_0) \\
&\leq \left(\frac{\lambda_0}{2} + \alpha + \frac{A}{2}\right)\delta\Phi(x)e^{(\alpha+\beta+A)t} + \left(\frac{\lambda_0}{2} + \beta + \frac{A}{2}\right)\delta\|\eta(x)\| \\
&\leq f(t, x, \delta e^{(\alpha+\beta+A)t}\Phi(x)) + R(t, x, \delta\Phi(x)\eta(x)e^{(\alpha+\beta+A)(t-\sigma)}).
\end{aligned}
$$

The boundary condition (3)(b) leads to $\hat{u}|_{S_T} = \delta\Phi(x)e^{(\alpha+\beta+A)t}|_{S_T} = 0 \leq h(t, x)|_{S_T}$.
For the initial data (3)(c) and (3)(d), we have $\hat{u}(t, x) \equiv \delta\Phi(x)e^{(\alpha+\beta+A)t} \leq \delta \leq$
$\eta_0(t, x)$ in $D_{-\sigma}$, and $\hat{u}(0, x) = \delta\Phi(x) \leq \delta \leq u_0(x) = \eta_0(0, x)$. Thus, we veri-
fied that $\hat{u}(t, x)$ is a lower solution of (3). Our goal is to find an upper solution of
(3). To this end, we fix $T > 0$ and choose both numbers $M, N \geq \delta e^{(\alpha+\beta+A)T}$. Taking
into account the assumptions for both functions f and R (and therefore F) which
are bounded on bounded sets, we define the auxiliary function F_{NM} by

$$
F_{NM}(t, x, u, \alpha(x)) = \begin{cases}
F(t, x, u, v(x)) & u < M, \ \|v(x)\| < N \\
F(t, x, u, \tilde{v}(x)) & u < M, \ \|v(x)\| \geq N \\
F(t, x, M, v(x)) & u \geq M, \ \|v(x)\| < N \\
F(t, x, M, \tilde{v}(x)) & u \geq M, \ \|v(x)\| \geq N,
\end{cases}
\tag{18}
$$

where $\tilde{\nu}(x) = N \dfrac{\nu(x)}{\|\nu(x)\|}$. The definition of F_{NM} was suggested by our former coauthor K. Nakagawa. Consider (3)(a) with right-hand side F_{NM} instead of F, that is,

$$
\begin{aligned}
&(a)\ \tfrac{\partial u}{\partial t} - Lu = F_{NM}(t, x, u, Qu(t, x))\ \text{ in } D_T,\\
&(b)\ Bu = h(t, x)\ \text{ on } S_T,\\
&(c)\ u(t, x) = \eta_0(t, x)\ \text{ in } D_{-\sigma},\\
&(d)\ \eta_0(0, x) = u_0(x)\ \text{ in } \Omega.
\end{aligned}
\tag{19}
$$

Next, we consider again the linear problem (7). Then having in mind the above-stated requirements, we conclude that the IBVP (7) has a unique solution $r = r(t, x)$, $r \in C(\overline{D}_T) \cap C^{1,2}(D_T)$. Let's assume that $c = 0$ and $q \geq F_{NM}$ in (7). Then we denote the solution of (7) in this case by $r_0(t, x)$ that is an upper solution of the modified problem (19). Since $F_{NM}(t, x, \cdot, \cdot) \in L_{loc}(\mathbf{R}^+ \times \mathbf{R}^+)$, and at the same time $F_{NM}(t, x, 0, 0) \geq 0$, then Theorem 1 ensures the existence of a unique solution z of (19), and $0 \leq z \leq r_0(t, x)$. By the choice of N, M, the function $\hat{u} = \delta e^{(\alpha+\beta+A)t} \Phi(x)$ is a lower solution of the modified problem (19) in the domain $\overline{D}_T$, and $\hat{u} \leq r_0$. Therefore, $\hat{u} \leq z \leq r_0$. Hence, by the definition of F_{NM}, z is the solution of the original problem for as long as $z \leq \min\{N, M\}$. Also, it follows that by taking N, M sufficiently large, the solution of problem (3) satisfies the inequality (16) $\forall t > 0$ unless it blows up in finite time. Thus in the case of blow-up, the inequality (16) is evidently satisfied. This completes the proof. $\qquad\square$

Theorem 3 *Suppose that **(H1)** holds and $u(t, x)$ is the nonnegative solution of (3). If there exist constants $\alpha > 0$, $\beta > 0$, and $A > 0$ such that*

$$
\begin{aligned}
&(a)\ \ f(t, x, \eta) \geq \left[\left(\alpha + \frac{A}{2}\right)(1+t)^{-1} + \frac{\lambda_0}{2}\right]\eta\ \ \forall\, \eta \geq 0,\\
&(b)\ \ R(t, x, Qu(t, x)) \geq \left[\left(\beta + \frac{A}{2}\right)(1+t)^{-1} + \frac{\lambda_0}{2}\right]\|Qu\|,
\end{aligned}
\tag{20}
$$

then, for any $\delta > 0$ and for any $\eta_0(t, x) \geq \delta$,

$$
u(t, x) \geq \delta(1+t)^{\alpha+\beta+A}\Phi(x)\ \text{for}\ t > 0,\ x \in \overline{\Omega}.
\tag{21}
$$

Proof Let $\hat{u} = \delta(1+t)^{\alpha+\beta+A}\Phi(x)$ be a function of t and x, where $A > 0$. We will prove that $\hat{u}$ is a lower solution of (3). To this end, we have to verify that

$$
\tfrac{\partial \hat{u}}{\partial t} - L\hat{u} \leq f(t, x, \hat{u}) + R(t, x, Q\hat{u})\ \text{ in }\ D_T.
$$

We note that

$$\frac{\partial \hat{u}}{\partial t} - L\hat{u} = \delta(\alpha + \beta + A)(1+t)^{\alpha+\beta+A-1}\Phi(x) + \delta\lambda_0\Phi(x)(1+t)^{\alpha+\beta+A}$$
$$= \delta\Phi(x)(1+t)^{\alpha+\beta+A}[(\alpha+\beta+A)(1+t)^{-1}+\lambda_0]$$
$$= \delta\Phi(x)(1+t)^{\alpha+\beta+A}[(\alpha+A/2)(1+t)^{-1}+\lambda_0/2+(\beta+A/2)(1+t)^{-1}+\lambda_0/2]$$
$$= \delta\Phi(x)(1+t)^{\alpha+\beta+A}[(\alpha+A/2)(1+t)^{-1}+\lambda_0/2]$$
$$+\delta\Phi(x)(1+t)^{\alpha+\beta+A}[(\beta+A/2)(1+t)^{-1}+\lambda_0/2]$$
$$\leq \delta\Phi(x)[(\alpha+A/2)(1+t)^{-1}+\lambda_0/2](1+t)^{\alpha+\beta+A}$$
$$+\|\hat{u}_t\|[(\beta+A/2)(1+t)^{-1}+\lambda_0/2]$$
$$\leq f(t,x,\delta\Phi(x)(1+t)^{\alpha+\beta+A})+R(t,x,Q\hat{u}).$$

From (3)(b) and (5), we have $B\hat{u} = \delta(1+t)^{\alpha+\beta+A}B\Phi(x) = 0 \leq h(t,x)$ on S_T, and using (3)(c) obtain $\hat{u}(t,x) = \delta(1+t)^{\alpha+\beta+A}\Phi(x) \leq \eta_0(t,x)$ on $D_{-\sigma}$.

Our next goal is to find an upper solution for (3). Further for some $T > 0$, we choose numbers $N, M \geq \delta(1+T)^{\alpha+\beta+A}$ and define the function $F_{NM} = F_{NM}(t,x,u,v)$ as in (18). Having in mind (19) and by the same reasoning as in the proof of Theorem 2, we obtain that the solution u of (3) satisfies the inequality (21).

The second result concerns the existence of blow-up.

Theorem 4 *Let (**H1**) hold and let z be a nonnegative function defined on $[0, T_0) \times \overline{\Omega}$ and unbounded at some point in $\overline{\Omega}$ as $t \to T_0$. If z is a lower solution of (3) in $\overline{D}_T$ for every $T < T_0$, then there exists another positive number $T^* \leq T_0$ such that a unique solution $u = u(t,x)$ of (3) exists in $(0, T_0] \times \overline{\Omega}$ and $\lim_{t \to T^*}[\max_{x \in \overline{\Omega}} u] = \infty$.*

Proof Consider the function $F_{NM}(t,x,u,v)$ in (18) defined in D_T and for each $N > 0, M > 0$. Here, we will use the modified IBVP (19). Now observe that the condition (**H1**) holds for the modified function F_{NM} for any $u_1 \geq u_2 \geq 0$. Choose arbitrary $N, M > \eta_0(t,x)$ in $D_{-\sigma}$. Then there is a number $T_{NM} < T_0$ such that $v \leq N$, $u \leq M$ on $D_{NM} \equiv [0, T_{NM}) \times \overline{\Omega}$, and $v(t_0, x_0) = N$, $u(t_1, x_1) = M$ for some points $(t_0, x_0), (t_1, x_1) \in D_{NM}$, respectively. Accept $\hat{u} = \max\{u, v\}$ as a lower solution of the modified problem (19) in D_{NM}. Assume that $\underline{u}^{(0)} = \hat{u}$ is the initial iteration and then the corresponding lower sequence $\underline{u}^{(k)}$ is monotone nondecreasing in D_{NM}. The uniform boundedness of F_{NM} ensures that $\underline{u}^{(k)}$ is bounded in D_{NM}. One may prove (see, e.g., [11]) that $\underline{u}^{(k)}$ converges monotonically from below to a unique solution z^* of the modified IBVP (19) and $z^* \geq \hat{u}$. From $\eta_0 < S \equiv \min\{N, M\}$ in $\overline{\Omega}$, we conclude that there exists $t_2 \leq T_{NM}$ such that $0 \leq z^* \leq S$ on $[0, t_2] \times \overline{\Omega}$ and $z^*(t_2, x_2) = S$ at $x_2 \in \overline{\Omega}$. We note that F_{NM} coincides with F when $0 \leq v \leq N$ and $0 \leq u \leq M$ and then z^* is solution of the IBVP (3) in $[0, t_2] \times \overline{\Omega}$. Hence, we infer that there exists a solution of (3) for as long as it remains bounded by S. If the solution is bounded in $[0, T_0] \times \overline{\Omega}$ by some constant C, then it is bounded in the set $[0, T_{NM}] \times \overline{\Omega}$ by C for any $T_{NM} \leq T_0$. Therefore, for some $N, M > C$ it follows that for some $t_2 \leq T_{NM}$, IBVP (3) has a unique solution z in $[0, t_2] \times \overline{\Omega}$ such that $z(t_2, x_2) = S$ for $x_2 \in \overline{\Omega}$. This contradicts to $z \leq C < S$ in $[0, T_0] \times \overline{\Omega}$ hence z should be unbounded in $[0, T_0] \times \overline{\Omega}$. $\square$

References

1. Agarwal R., Arshad S., Lupulescu V., O'Regan D., Evolution equations with CO, *Differ. Equ. Appl.* **7** (2015) 15–26.
2. Ahmad B., Khan R., Sivasundaram S., Generalized quasilinearization method for nonlinear functional differential equations, *J. Appl. Math. Stoch. Anal.* **16** (2003) 33–43.
3. Bainov, D., Hristova, S., Differential Equations with Maxima, Taylor & Francis Group, Chapman & Hall/CRC, 2011.
4. Burlică M., Necula M., Roşu D., Vrabie I. I., *Delay Differential Evolutions Subjected to Nonlocal Initial Conditions*. Monographs and Research Notes in Mathematics. CRC Press, New York, 2016.
5. Corduneanu, C., Functional Equations with Causal Operators, Taylor & Francis 2002.
6. Drici Z., McRae F., Vasundhara Devi J., Quasilinearization for functional differential equations with retardation and anticipation, *Nonlinear Anal.* **70** (2009) 1763–1775.
7. Friedman, A., Partial differential equations of parabolic type, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1964.
8. Jabeen, T., Agarwal, R. , Lupulescu, V., O'Regan, D., Impulsive evolution equations with causal operators, Symmetry, 2020, 12, no. 1: 48.
9. Jabeen T., Lupulescu V., Existence of mild solutions for a class of non-autonomous evolution equations with nonlocal initial conditions, *J.Nonlinear Sci. Appl.* **10** (2017) 141–153.
10. Lakshmikantham, V., Köksal, S., Monotone Flows and Rapid Convergence for Nonlinear Partial Differential Equations, Taylor & Francis, 2003.
11. Pao, C. V., Nonlinear Parabolic and Elliptic Equations, Plenum Press, 1992.

Well Posedness of Solutions for a Degenerate Viscoelastic Equation for Kirchhoff-Type with Logarithmic Nonlinearity

Nazli Irkıl and Erhan Pişkin

Abstract This paper deals with Kirchhoff-type equation with degenerate viscoelastic term and logarithmic nonlinearity. We established the global existence of solutions by using the potential well method.

Keywords Degenerate viscoelastic wave equation · Kirchhoff-type equation · Logarithmic nonlinearity

1 Introduction

In this paper, we study the following viscoelastic Kirchhoff-type problem:

$$
\begin{cases}
v_{tt} - M\left(\|\nabla v\|^2\right)\Delta v - div\left[b\left(x\right)\nabla v\right] + \int_0^t g\left(t-s\right) div\left[b\left(x\right)\nabla v\right] ds + v_t = |v|^{p-2}\, v \ln|v|, \\
v\left(x, 0\right) = v_0\left(x\right), v_t\left(x, 0\right) = v_1\left(x\right), x \in \Omega, \\
v\left(x, t\right) = 0, x \in \partial\Omega \times \left(0, \infty\right),
\end{cases}
\tag{1}
$$

where Ω is a bounded domain in R^n with smooth boundary $\partial\Omega$. $M\left(s\right)$ is a continuous function and $g\left(t\right)$ is a positive kernel function which was defined on R^+.

The problems which include logarithmic source term are related to different branches of physics [2, 6]. Because of that, these type problems are new and important area for authors. And they showed different mathematical behavior of problems. For more results, we refer the readers to [3, 7, 10].

The following equation

$$
u_{tt} - M\left(\|\nabla u\|^2\right)\Delta u - \int_0^t h\left(t-s\right)\triangle u\left(s\right) ds - \triangle u_t = |u|^{p-2}\, u \ln|u|. \tag{2}
$$

N. Irkıl (✉) · E. Pişkin
Dicle University, Diyarbakır 21280, Turkey
e-mail: nazliirkil@gmail.com

has been considered by Ferreira et al. [5]. They established blow up results of solutions for negative initial energy. Later Pişkin et al. [8] studied initial boundary problem (2) with $p = 2$. They proved the local existence of weak solution by applying Banach fixed theorem. And they obtained blow up results.

Motivated by the papers, we are led naturally to generalize and consider a degenerate viscoelastic equation in Kirchhoff-type subject to the global existence.

2 Set Up of Potential Well

In this section, we will outline the general framework of the definitions of the functions in the problem (1) and give some related lemma used to prove the conclusions later. First of all, we define $\|.\|_2$ and $\|.\|_p$ to the usual $L^2(\Omega)$ norm and $L^p(\Omega)$ norm, respectively. We denote by C and C_i $(i = 1, 2, \ldots)$ various positive constants.

Let us begin with defining the $b(x) \in C^1(\Omega)$ is a positive function such that

$$b(x) \geq b_1^2 > 0, \tag{3}$$

and

$$H_b := \left\{ v \in H_0^1(\Omega) : \int_\Omega b(x) |\nabla v|^2 \, dx < \infty \right\}, \tag{4}$$

is a Hilbert space endowed with a norm

$$\|\nabla v\|_b^2 = \int_\Omega b(x) |\nabla v|^2 \, dx. \tag{5}$$

By combination of (3) and (5), is clear that

$$\|\nabla v\|_2^2 \leq \frac{1}{b_1^2} \|\nabla v\|_b^2. \tag{6}$$

Now we state some hypotheses and lemmas which will be used to prove our results.

$(H1)$ $g : R^+ \to R^+$ is a C^1 nonincreasing function satisfying

$$g(0) \geq 0, \, g'(0) \leq 0, \, 1 - \int_0^t g(s) \, ds = l > 0. \tag{7}$$

$(H2)$ There exist positive differentiable function $\vartheta(s)$, such that

$$g'(s) \leq \vartheta(s) g(s), \, for \, s \geq 0. \tag{8}$$

$(H3)$ $M(s)$ is a C^1 function and satisfies

$$M(s) \geq -m_0 \text{ and } M(s)\, s \geq \widehat{M}(s) \tag{9}$$

where $l - 1 \geq \frac{m_0}{b_1}$, where b_1 the first eigenvalue of the problem $\Delta v = -b$ $(div\,[b\,(x)\,\nabla v])$ and $\widehat{M}(s) = \int\limits_0^s M(\tau)\,d\tau$.

$(H4)$ The constant p satisfies

$$\begin{cases} p \geq 1, n = 1, 2, \\ \min\{1, b-1\} < p \leq \frac{n+2}{n-2}, n \geq 3. \end{cases} \tag{10}$$

Corollary 1 *The first eigenvalue b_1 of the problem*

$$\begin{cases} \Delta v = -b\,(div\,[b\,(x)\,\nabla v]) \\ v = 0, \partial\Omega \end{cases} \tag{11}$$

satisfies

$$b_1 = \inf_{v \in H_b(\Omega)} \frac{\|\nabla v\|_b^2}{\|\nabla v\|_2^2} > 0, \|\nabla v\|_2^2 \leq \frac{1}{b_1^2}\,\|\nabla v\|_b^2. \tag{12}$$

Lemma 1 [1] *Let C_1 be the smallest positive constant satisfying for $\forall v \in H_0^1(\Omega)$*

$$\|v\|_q \leq C_1\,\|\nabla v\|, \tag{13}$$

where $2 \leq q < \infty$, if $n = 1, 2$; and $2 \leq q \leq \frac{2n}{n-2}$, if $n \geq 3$.

Now, we introduce the total energy function fot the problem (1) at time $t \geq 0$ such that

$$E(t) = \frac{1}{2}\,\|v_t\|^2 + \frac{1}{2}\left(l - \frac{m_0}{b_1}\right)\|\nabla v\|_b^2 + \frac{1}{2}\,(g \circ \nabla v)\,(t)$$
$$+ \frac{1}{p^2}\,\|v\|_p^p - \frac{1}{p}\int\limits_\Omega v^p \ln|v|\,dx. \tag{14}$$

where

$$(g \circ \nabla v)\,(t) = \int\limits_0^t g\,(t - \tau)\,\|\nabla v\,(t) - \nabla v\,(\tau)\|_b^2\,d\tau. \tag{15}$$

The potential energy functional

$$J(v) = \frac{1}{2}\left(l - \frac{m_0}{b_1}\right)\|\nabla v\|_b^2 + \frac{1}{2}\left(g \circ \nabla v\right)(t)$$
$$+ \frac{1}{p^2}\|v\|_p^p - \frac{1}{p}\int_\Omega v^p \ln|v|\,dx. \tag{16}$$

and the Nehari functional

$$I(v) = \left(l - \frac{m_0}{b_1}\right)\|\nabla v\|_b^2$$
$$+ (g \circ \nabla v)(t) - \int_\Omega v^p \ln|v|\,dx, \tag{17}$$

for $v \in H_0^1(\Omega)$.

Thanks to (14)–(17), it tells us that for $v \in H_0^1(\Omega)$,

$$J(v) = \frac{I(v)}{p} + \frac{(p-2)}{2p}\left(l - \frac{m_0}{b_1}\right)\|\nabla v\|_b^2$$
$$+ \frac{(p-2)}{2p}\left(g \circ \nabla v\right)(t) + \frac{1}{p^2}\|v\|_p^p, \tag{18}$$

and

$$E(t) = \frac{1}{2}\|v_t\|^2 + J(v). \tag{19}$$

And then, by (17), we state potential set as follows:

$$W = \left\{v \in H_0^1(\Omega) \mid I(v) > 0\right\} \cup \{0\}, \tag{20}$$

and the outer space of the potential well

$$U = \left\{v \in H_0^1(\Omega) \mid I(v) < 0\right\}. \tag{21}$$

The depth of potential well is defined as

$$d = \inf_{u \in N} J(v). \tag{22}$$

where the Nehari manifold was denited as

$$N = \left\{v \in H_0^1(\Omega) \mid I(v) = 0\right\} \cup \{0\}. \tag{23}$$

Lemma 2 *Suppose that* (A1) *and* (A2) *hold. Then the energy functional* $E(t)$ *is decresing with respect to t and*

$$E'(t) = \frac{1}{2} \left[(g' \circ \nabla v)(t) - g(t) \|\nabla v(t)\|_b^2 \right] - \|v_t\|^2 \leq 0, \tag{24}$$

where

$$(g' \circ \nabla v)(t) = \int_0^t g'(t-s) \|\nabla v(t) - \nabla v(\tau)\|_b^2 \, d\tau. \tag{25}$$

Lemma 3 [8, 9] *For any* $v \in H_b(\Omega)$, $\|v\| \neq 0$, *then we have*
   (i)

$$\lim_{\lambda \to 0} J(\lambda v) = 0, \tag{26}$$

and

$$\lim_{\lambda \to \infty} J(\lambda v) = -\infty \tag{27}$$

   (ii) *For* $0 < \lambda < \infty$, *there exists a unique* λ_1, *such that*

$$\frac{d}{d\lambda} J(\lambda v) |_{\lambda = \lambda_1} = 0 \tag{28}$$

   where λ_1 *is the unique root of equation* $l \left(l - \frac{m_0}{b_1} \right) \|\nabla v\|_b^2 + (g \circ \nabla v)(t) = -\lambda^{p-2} \int_\Omega v^p \ln |v| \, dx - \lambda^{p-2} \ln \lambda \int_\Omega v^p dx$
   (iii) $J(\lambda v)$ *is strictly decreasing on* $\lambda_1 < \lambda < \infty$, *strictly increasing on* $0 < \lambda < \lambda_1$ *and attains the maximum at* $\lambda = \lambda_1$
   (iv) $I(\lambda v) > 0$ *for* $0 < \lambda < \lambda_1$, $I(\lambda v) > 0$ *for* $\lambda_1 < \lambda < \infty$, *and* $I(\lambda_1 v) = 0$

$$I(\lambda v) = \lambda \frac{d}{d\lambda} J(\lambda v) \begin{cases} > 0, & 0 \leq \lambda \leq \lambda_1, \\ = 0, & \lambda = \lambda_1, \\ < 0, & \lambda_1 \leq \lambda. \end{cases} \tag{29}$$

Lemma 4 [4, 8] *For* $t \geq 0$, *the potential well depth d represented as*

$$d = \frac{(p-2)}{2p} \left[(l - m_0)^{\frac{p+1}{2}} \frac{b_1^{\frac{p2+p+1}{p-1}}}{C_1^{p+1}} \right]^{\frac{2}{p-1}}, \tag{30}$$

where C_1 *is the optimal constant of Lemma 1* $(H_b(\Omega) \hookrightarrow L^{p+1})$ *and*

$$\begin{cases} b - 1 \leq p \leq \frac{n+2}{n-2} n \geq 3, \\ b - 1 \leq p \leq \infty, n = 1, 2. \end{cases} \tag{31}$$

Definition 1 (*Local weak solution*) Let $T > 0$. Assume that $(H1) - (H4)$ hold. We say that for any $(v_0, v_1) \in H_b(\Omega) \times L^2(\Omega)$, there is a unique local weak solution of problem (1) such that

$$v \in C\left([0, T]; H_0^2(\Omega) \cap H_b(\Omega)\right), \tag{32}$$

and

$$(v_t) \in \left(C[0, T]; L^2(\Omega)\right), \tag{33}$$

leads to

$$\begin{cases} (v_{tt}, v_1) + M\left(\|\nabla v\|^2\right)(\nabla v, \nabla v_0) \\ - \int_0^t g_1(t - s)\, \nabla\,[b(x)\, \nabla v]\, v_0 ds + \int_\Omega v_0 v_t dx \\ \qquad = \int_\Omega |v|^{p-2}\, v \ln |v|\, v_0 dx. \end{cases} \tag{34}$$

3 Global Existence

The global existence of solution for the problem (1) was considered in this part. We give some useful lemma and later we proved Theorem 1.

Lemma 5 *Let* $(H1) - (H4)$ *hold. Suppose that for any* $(v_0, v_1) \in H_b(\Omega) \times L^2(\Omega)$, *such that*

$$I(0) > 0, \varphi = \frac{C_1^{p+1} b_1}{b_1^{p+1}(lb_1 - m_0)}\left(\frac{2pb_1}{(p-2)(lb_1 - m_0)} E(0)\right)^{\frac{p-1}{2}} \tag{35}$$

then for all $t > 0$

$$I(v(t)) > 0. \tag{36}$$

Proof Onwards $I(0) > 0$, then by continuity of $v(t)$, we conclude that, there is a time $t^* < T$, such that for $\forall t \in [0, t^*]$

$$I(v(t)) \geq 0. \tag{37}$$

Thanks to (18)and (36), we obtain that

$$\begin{aligned} J(v(t)) \geq{}& \frac{(p-2)}{2p}\left(l - \frac{m_0}{b_1}\right)\|\nabla v\|_b^2 \\ &+ \frac{(p-2)}{2p}(g \circ \nabla v)(t) + \frac{1}{p^2}\|v\|_p^p. \end{aligned} \tag{38}$$

for any $\alpha > 0$. Since taking $l - 1 \geq \frac{m_0}{b_1}$ and using Lemma 2 and (38), we arrive that for $\forall t \in [0, t^*]$

$$\left(l - \frac{m_0}{b_1}\right) \|\nabla v\|_b^2 \leq \frac{2p}{(p-2)} E(0). \tag{39}$$

Exploiting $\log x < x$ inequality, Lemma 1, (35), and (36), we conclude that

$$\int\limits_{\Omega} |v|^p \ln |v| \, dx \leq \|v\|_{p+1}^{p+1}$$

$$= \varphi \left(l - \frac{m_0}{b_1}\right) \|\nabla v\|_b^2 < \left(l - \frac{m_0}{b_1}\right) \|\nabla v\|_b^2. \tag{40}$$

therefore, by definition of $I(u(t))$, we show that for any $t \in [0, t^*]$

$$I(v(t)) > 0. \tag{41}$$

$$\lim_{t \to t^*} \frac{C_1^{p+1} b_1}{b_1^{p+1} (lb_1 - m_0)} \left(\frac{2pb_1}{(p-2)(lb_1 - m_0)} E(v(t), v_t(t))\right)^{\frac{p-1}{2}} \leq \alpha < 1 \tag{42}$$

we can take $t^* = T$.

Lemma 6 [8] *Suppose that $(v_0, v_1) \in H_b(\Omega) \times L^2(\Omega)$ and $(H1) - (H4)$, $0 < E(0) < d$. If $I(v_0) > 0$, then $v(t) \in W$, where v was defined as a weak solution of the problem (1).*

Theorem 1 *Let $(v_0, v_1) \in H_b(\Omega) \times L^2(\Omega)$ and $(H1) - (H4)$ hold. Suppose that $E(0) < d$ and $v(t) \in W$, then problem (1) admits a globally and bounded weak solution.*

Proof It suffices to show that $\|v_t\|^2 + \|\nabla v\|_b^2$ is bounded independently of t. Under the condition of Theorem 1 and thanks to Lemma 6, we conclude that $v \in W$ on $[0, T)$. Thus, we obtain

$$C_2 \left(\|v_t\|^2 + \|\nabla v\|_b^2\right) \leq E(0), \tag{43}$$

where $C_2 = \min\left\{\frac{1}{2}, \left(l - \frac{m_0}{b_1}\right) \frac{(p-2)}{2p}, 1\right\}$ is a positive constant, which depends on p, l, m_0, b_1.

By condition $E(0) < d$, (43) can be rewritten as

$$\|v_t\|^2 + \|\nabla v\|_b^2 \leq \frac{1}{C_2} E(0) \leq \frac{1}{C_2} d < \infty. \tag{44}$$

By using of the (44) and continuation principle, we have the global existence of the solution v for the problem (1).

Notes and Comments This work was partly supported by the Scientific and Technological Research Institution of Turkey (TUBÏTAK).

References

1. Adams, R.A., Fournier, J.J.F.: Sobolev Spaces. Academic Press, Elsevier, USA(2003).
2. Bialynicki-Birula, I., Mycielski, J.:Wave equations with logarithmic nonlinearities. Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys. (23)(4), 461-466 (1975).
3. Boulaaras, S., Draifia, A., Alnegga, M.:Polynomial decay rate for Kirchhoff type in viscoelasticity with logarithmic nonlinearity and not necessarily decreasing kernel.Symmetry.(11)(2), 226 (2019). https://doi.org/10.3390/sym11020226
4. Boumaza, N., Gheraibia, B.: General decay and blowup of solutions for a degenerate viscoelastic equation of Kirchhoff type with source term. J. Math. Anal. Appl. (489)(2), 124185 (2020). https://doi.org/10.1016/j.jmaa.2020.124185
5. Ferreira, J., Irkıl, N., Raposo, C.A.: Blow up results for a viscoelastic Kirchhoff-type equation with logarithmic nonlinearity and strong damping. Math. Morav.(25)(2), 125-141 (2021). 10.5937/MatMor2102125F
6. De Martino,S., Falanga, M., Godano, C., Lauro, G.:Logarithmic Schrödinger-like equation as a model for magma transport. Europhys. Lett.(63)(3), 472 (2003). https://doi.org/10.1209/epl/i2003-00547-6
7. Al-Gharabli, M., Guesmia, A., Messaoudi, S.: Existence and a general decay results for a viscoelastic plate equation with a logarithmic nonlinearity. Commun. Pure Appl. Anal. (18)(1), 159-180 (2019). 10.3934/cpaa.2019009
8. Pişkin, E., Boulaaras, S., Irkıl, N.: Existence and blow up for a nonlinear viscoelastic Kirchhoff-type equation with logarithmic nonlinearity. Kragujevac J. Math. (49)(3), 335-351 (2025).
9. Yang, Y., Li, J., Yu, T. Qualitative analysis of solutions for a class of Kirchhoff equation with linear strong damping term, nonlinear weak damping term and power-type logarithmic source term. Appl. Numer. Math. (141), 263-285 (2019). https://doi.org/10.1016/j.apnum.2019.01.002
10. Ye, Y. Global solution and blow-up of logarithmic Klein-Gordon equation. Bull. Korean Math. Soc. (57)(2), 281-294 (2020). https://doi.org/10.4134/BKMS.B190190

An Application of Simplest Equations Method to Nonlinear Equations of Schrödinger Kind

Ivan P. Jordanov

Abstract In this study, we apply the Simplest Equations Method (SEsM) to obtain the exact solution of equation which is connected to the nonlinear Equation of Schrödinger. We use a specific case of SEsM which is based on the use of two simple equations. By means of the first simple equation, we construct a complex exponential function. The solution of the second simple equation is derived by applying the classical modified method of the simplest equation.

Keywords Simple equations method (SEsM) · Nonlinear equation of Schrödinger

1   Introduction

The dynamics of many social and economic systems is nonlinear [1–18]. During the past few decades, many non-linear phenomena are modeled by systems of nonlinear partial differential equations (PDEs). Such model systems require applications of the methods of non-linear dynamics, chaos theory, and theory of stochastic processes [19–39].

In more detail, reaction–diffusion equations have many applications for the description of different kinds of processes in physics, chemistry, biology, economics, and social sciences. Most systems in our environment contain components that interact through competition or cooperation which in some cases can lead to adaptation of the interactions. It is of special importance to study the behavior of such systems and to develop and apply new appropriate mathematical methods for the description of processes in these systems. In addition, the agent models are an important tool for the analysis of complex systems [40]. Depending on the nature of the system, the agents may have a variety of properties, as well as they can interact in different

I. P. Jordanov (✉)

Institute of Mechanics, Bulgarian Academy of Sciences, Akad. G. Bonchev Str., Bl. 4, 1113 Sofia, Bulgaria

e-mail: jordanovip@gmail.com

Studentski Grad Hristo Botev, University of National and World Economy, 1700 Sofia, Bulgaria

ways. In the recent years, we observe a rapid growth in publications related to agent models because they may explain adequately the complex processes in a number of social systems, such as migrations of various human populations [41–45].

When the number of agents is small and the chaos in their behavior is insignificant, the dynamics of systems of interacting agents can be described by less complicated systems of ordinary or partial differential equations. Therefore, the deterministic systems of PDEs are widely used to represent the dynamics of many economic and social processes. Analysis of these systems by applying various analytical and numerical methods can lead us to a better understanding of the main dynamic features of agent systems. For example, extracting exact particular analytical solutions of such equations can help in understanding and explaining the migration of individuals in space, i.e., so-called population waves [41–49].

The paper is organized as follows. We discuss the SEsM methodology for obtaining exact solutions of nonlinear partial differential equations in Sect. 2. In Sect. 3, we apply a simple specific case of this methodology, based on two simple equations, in order to obtain solutions of nonlinear partial differential equation connected to the nonlinear equation of Schrödinger. Several concluding remarks are summarized in Sect. 4.

2 The Simple Equations Method (SEsM)

Below we apply a specific case of a methodology for obtaining exact and approximate solutions of nonlinear partial differential equations called Simple Equations Method (SEsM) [71–74]. Vitanov used elements of SEsM in his early works almost 30 years ago [50–58]. In 2009 [48, 49] and in 2010, Vitanov and co-authors have used the ordinary differential equation of Bernoulli as simplest equation [59] and applied a version of SEsM called Modified Method of Simplest Equation (MMSE) to ecology and population dynamics [60]. This method uses a balance equation [61, 62] to determine the kind of the simplest equation and truncation of the series of solutions of the simplest equation. Up to 2018, the applications of the methodology have been based on MMSE [63–70]. Recently Vitanov extended the capacity of the methodology by the inclusion of the possibility of use of more than one simplest equation. This modification is called SEsM—Simple Equations Method—as the used simple equations are more simple than the solved nonlinear partial differential equation but these simple equations in fact can be quite complicated. We note that the first description of the methodology was made in [71] and then in [72–74]. For more applications of specific cases of the methodology, see [75–80].

A short description of SEsM is as follows. We consider the specific case of SEsM for obtaining the exact solution of one nonlinear partial differential equation. We consider the nonlinear partial differential equation

$$\mathcal{Y}[u_1(x, \ldots, t), \ldots, u_n(x, \ldots, t)] = 0. \tag{1}$$

$\mathcal{Y}[u_1(x, \ldots, t), \ldots, u_n(x, \ldots, t)]$ depends on the functions $u_i(x, \ldots, t)$ and some of their derivatives (u_i can be a function of more than 1 spatial coordinates). Step 1 of SEsM is to perform transformation

$$u(x, \ldots, t) = T[F_i(x, \ldots, t), G_i(x, \ldots, t), \ldots]. \tag{2}$$

Here $T(F, G, \ldots)$ is some function of another functions $F, G, \ldots$ In general, $F(x, \ldots, t), G(x, \ldots, t), \ldots$ are functions of several spatial variables as well as of the time. The goal of the transformations is to remove the nonlinearity of the solved equation or to transform the nonlinearity of the solved differential equation to more treatable kind of nonlinearity (such as polynomial nonlinearity). The nonlinearities in the solved equations can be of different kinds. Because of this, there is no universal transformation of the nonlinearity which is valid for all possible cases. In many specific cases, one may skip this step. Then $u(x, \ldots, t) = F(x, \ldots, t))$. But, in numerous cases, the step is necessary for obtaining a solution of the studied nonlinear partial differential equation. The application of (2) to (1) leads to a nonlinear PDEs for the functions $F, G, \ldots$.

In Step 2 of SEsM, the functions $F(x, \ldots, t), G(x, \ldots, t), \ldots$ are represented as a function of other functions $f_1, \ldots, f_N, g_1, \ldots, g_M, \ldots$, which are connected to solutions of some differential equations (these equations can be partial or ordinary differential equations) that are more simple than Eq. (2). The possible values of N and M are N = 1, 2, …, M = 1,2,... (there may be an infinite number of functions f too). The forms of the functions $F(f_1, \ldots, f_N), G(x, \ldots, t), \ldots$ can be different. For example, for the case of single solved equation, the function F can have the form

$$F = \alpha + \sum_{i_1=1}^{N} \beta_{i_1} f_{i_1} + \sum_{i_1, i_2=1}^{N} \gamma_{i_1, i_2} f_{i_1} f_{i_2} + \cdots + \sum_{i_1, \ldots, i_N=1}^{N} \sigma_{i_1, \ldots, i_N} f_{i_1} \cdots f_{i_N}, \tag{3}$$

where $\alpha, \beta_{i_1}, \gamma_{i_1, i_2}, \sigma_{i_1, \ldots, i_N} \ldots$ are parameters.

Step 3 of SEsM is connected to the representation of the functions used in $F, G, \ldots$ - the functions $f_1, \ldots, f_N, g_1, \ldots, g_M$ which are solutions of some differential equations. We shall consider the case when these differential equations are ordinary differential equations.

We shall skip Step 4 of SEsM. These steps deal with the case when the above differential equations are complicated and we have to construct their solution again on the basis of more simple differential equations. We assume that the simple differential equation has the needed exact solutions which allow us to construct the solution of the solved more complicated differential equation.

Step 5 of SEsM is connected to the determination of the used simple differential equations. In many cases, the form of the simple equations is determined by balance equations. Balance equations may be needed in order to ensure that the system of algebraic equations from Step 6 contains more than one term in any of the equations. This is needed if one searches for a nontrivial solution of the solved nonlinear partial differential equation.

At Step 6 of SEsM, we apply the steps (1)–(5) to Eqs. (2) and this transforms the left-hand side of these equations into functions which are the sum of terms where each term contains some function multiplied by a coefficient. This coefficient contains some of the parameters of the solved equations and some of the parameters of the solution. In most cases, a balance procedure must be applied in order to ensure that the above-mentioned relationships for the coefficients contain more than one term. The balance procedure leads to additional relationships between the parameters of the solved equation and the parameters of the solution. These relationships are called balance equations.

Finally, at Step 7 of SEsM, We can obtain a nontrivial solution of Eq. (2) if all coefficients mentioned in Step (6) are set to 0. This condition leads to a system of nonlinear algebraic equations for the coefficients of the solved nonlinear PDE and for the coefficients of the solution. Any nontrivial solution of this algebraic system leads to a solution to the studied nonlinear partial differential equation.

3   Application to Nonlinear Differential Equation of Schrödinger Kind

In this section, we shall find an analytical solution of an equation of the nonlinear Schrödinger kind. We shall use a simple version of SEsM based on two simple equations.

Let us consider the equation of nonlinear Schrödinger kind:

$$i\frac{\partial q(x,t)}{\partial t} + a\frac{\partial^2 q(x,t)}{\partial x^2} + b_{-2}\, q(x,t)\, |q(x,t)|^{-4} + b_0\, q(x,t) + \quad (4)$$
$$+b_1\, q(x,t)\, |q(x,t)|^2 + b_2\, q(x,t)\, |q(x,t)|^4 + b_3\, q(x,t)\, |q(x,t)|^6 = 0,$$

where $i = \sqrt{-1}$, a and b_i are parameters, and $q(x,t)$ is a complex function. The solution of Eq. 4 will be searched as

$$q(x,t) = g(\xi)\, h(x,t), \qquad (5)$$

where $g(\xi)$ is a real function ($\xi = \alpha x + \beta t$) and $h(x,t)$ is a complex function. The two simplest equations will be for the functions $g(\xi)$ and $h(x,t)$, respectively. Let us denote as $\bar{h}(x,t)$ the complex conjugate function of $h(x,t)$. The substitution of Eq. 5 in Eq. 4 leads to (we denote $\frac{dg(x,t)}{d\xi}$ as g' and $\frac{d^2 g(x,t)}{d\xi^2}$ as g'')

$$i\beta\, g'\, h(x, t) + i\, g\, \frac{\partial h(x, t)}{\partial t} + \alpha^2\, a\, g''\, h(x, t) + 2\,\alpha\, a\, g'\, \frac{\partial h(x, t)}{\partial x} +$$

$$+a\, g\, \frac{\partial^2 h(x, t)}{\partial x^2} + b_{-2}\, g^{-3}\, h^{-1}(x, t)\, \bar{h}^{-2}(x, t) b_0\, g\, h(x, t) + \quad (6)$$

$$+b_1\, g^3\, h^2(x, t)\, \bar{h}(x, t) + b_2\, g^5\, h^3(x, t)\, \bar{h}^2(x, t) + b_3\, g^7\, h^4(x, t)\, \bar{h}^3(x, t) = 0.$$

The first simplest equation is for the function $h(x, t)$. Taking into an account the presence of h and its derivatives in Eq. 6 as well as the presence of the term $h(x, t)\, \bar{h}(x, t)$ there and aiming to choose such simplest equation that will lead to the reduction of Eq. 6 to an equation for $g(\xi)$, we arrive at the simplest equation

$$\frac{\partial h(\zeta)}{\partial \zeta} = i\, h(\zeta), \quad \zeta = k\, x + \omega\, t + \theta \quad (7)$$

which solution is

$$h(\zeta) = \exp(i\,\zeta) = \exp[i\,(k\,x + \omega\,t + \theta)]. \quad (8)$$

The substitution of Eq. 7 in Eq. 6 reduces Eq. 6 to an equation for the function $g(\xi)$:

$$\alpha^2\, a\, g'' + i\,(\beta + 2\, a\,\alpha\, k)\, g' + b_{-2}\, g^{-3} + \quad (9)$$

$$+(b_0 - \omega - a\, k^2)\, g + b_1\, g^3 + b_2\, g^5 + b_3\, g^7 = 0.$$

$g(\xi)$ has to be a real function and then

$$\beta = -2\, a\,\alpha\, k. \quad (10)$$

The substitution of Eq. 10 in Eq. 9 followed by multiplication of the result by g' and integration with respect to ξ leads to the equation

$$\alpha^2\, a\, g'^2 - b_{-2}\, g^{-2} - C + (b_0 - \omega - a\, k^2)\, g^2 + \frac{b_1}{2}\, g^4 + \frac{b_2}{3}\, g^6 + \frac{b_3}{4}\, g^8 = 0, \quad (11)$$

where C is a constant of integration. Further, we set

$$u(\xi) = g^\sigma(\xi), \quad (12)$$

where σ is a parameter that will be determined below. Then the substitution of Eq. 12 in Eq. 11 leads to the following equation for $u(\xi)$:

$$u'^2 = \frac{\sigma^2}{a\,\alpha^2}\Big[b_{-2}\, u^{\frac{2\,(\sigma-2)}{\sigma}} + C\, u^{\frac{2\,(\sigma-1)}{\sigma}} - (b_0 - \omega - a\, k^2)\, u^2 +$$

$$-\frac{b_1}{2}\, u^{\frac{2\,(\sigma+1)}{\sigma}} - \frac{b_2}{3}\, u^{\frac{2\,(\sigma+2)}{\sigma}} - \frac{b_3}{4}\, u^{\frac{2\,(\sigma+3)}{\sigma}}\Big]. \quad (13)$$

The cases where $b_3 = 0$ and $\sigma = \pm 2, \pm 4$ are discussed in detail in [44, 76], where a solution was obtained using elliptic functions.

For the case $(b_3 \neq 0)$ $\sigma = 2$, Eq. 13 becomes

$$u'^2 = \frac{\sigma^2}{a\,\alpha^2}[b_{-2} + C\,u - (b_0 - \omega - a\,k^2)\,u^2 - \frac{b_1}{2}\,u^3 - \frac{b_2}{3}\,u^4 - \frac{b_3}{4}\,u^5]. \quad (14)$$

Now we search for solution of Eq. (14) of kind $u = u(\xi) = \sum_{i=0}^{n} a_i\,\varphi^i(\xi)$, where $\varphi'(\xi) = \sum_{j=0}^{r} c_j\varphi^j(\xi)$. Here a_i and c_j are parameters, and $\varphi(\xi)$ is a solution of some ordinary differential equation, referred to as the simplest equation. The balance equation is $2r - 2 = 3n$. We assume that $n = 1$ and $r = 5/2$, i.e., the equation $\varphi'(\xi) = [c_1\varphi(\xi) + c_0]^{\frac{5}{2}}$ will play the role of simplest equation and $u(\xi) = a_1\varphi(\xi) + a_0$. We substitute the last expression in Eq. (14). Relationships between the coefficients of the solution and the coefficients of the model are derived by solving a system of six algebraic equations, and they are

$$a_0 = -\frac{\sqrt[3]{4}(\sqrt[3]{16}\,\sigma\,b_2 + 15\,c_0\sqrt[3]{a\,\alpha^2\,c_1^2\,\sigma\,b_3^2})}{15\,\sigma\,b_3}, \quad b_1 = \frac{16\,b_2^2}{45\,b_3},$$

$$a_1 = -\frac{c_1\sqrt[3]{4\,a\,\alpha^2\,c_1^2\,\sigma\,b_3^2}}{\sigma\,b_3}, \quad C = -\frac{64\,b_2^4}{10125\,b_3^3}, \quad (15)$$

$$b_{-2} = -\frac{256\,b_2^5}{759375\,b_3^4}, \quad b_0 = \frac{675\,b_3^2(\omega + 675\,a\,k^2) + 32\,b_2^3}{675\,b_3^2},$$

and $c_0,\ c_1,\ b_2,\ b_3 \neq 0$, $a,\ k,\ \alpha,\ \sigma \neq 0$, and ω are free parameters.

Substituting (15) in the solution of equation $\varphi'(\xi) = [c_1\,\varphi(\xi) + c_0]^{\frac{5}{2}}$, the solution of Eq. (14) obtains the form (C_1 is integration constant):

$$u(x,\,t) = \frac{\sqrt[3]{4\,a\,\alpha^2\,c_1^2\,\sigma\,b_3^2}}{\sigma\,b_3}\left\{c_0 - \frac{\sqrt[3]{4}}{\sqrt[3]{[3\,c_1\,(\alpha\,x - 2\,a\,\alpha\,k\,t + C_1)]^2}}\right\} +$$

$$-\frac{\sqrt[3]{4}(\sqrt[3]{16}\,\sigma\,b_2 + 15\,c_0\sqrt[3]{a\,\alpha^2\,c_1^2\,\sigma\,b_3^2})}{15\,\sigma\,b_3}, \quad (16)$$

and the corresponding solution (5) of Eq. (4) is

$$q(x,\,t) = \left(\frac{\sqrt[3]{4\,a\,\alpha^2\,c_1^2\,\sigma\,b_3^2}}{\sigma\,b_3}\left\{c_0 - \frac{\sqrt[3]{4}}{\sqrt[3]{[3\,c_1\,(\alpha\,x - 2\,a\,\alpha\,k\,t + C_1)]^2}}\right\} +\right.$$

$$\left.-\frac{\sqrt[3]{4}(\sqrt[3]{16}\,\sigma\,b_2 + 15\,c_0\sqrt[3]{a\,\alpha^2\,c_1^2\,\sigma\,b_3^2})}{15\,\sigma\,b_3}\right)^{1/2} \exp[i\,(k\,x + \omega\,t + \theta)]. \quad (17)$$

4   Concluding Remarks

In this study, we show an example of the application of the Simple equations Methods (SEsM) for obtaining the exact solution of a nonlinear partial differential equation of Schrödinger kind. The used version of SEsM is based on two simple equations. We have obtained a solution that contains two parts. The first part is the envelope described by a complex exponential function. The second part is derived by applying the classical modified method of the simplest equation. The obtained results show that SEsM is an effective method for obtaining exact analytical solutions of nonlinear partial differential equations. We are going to apply this methodology for a further study of equations of nonlinear Schrödinger kind in our future research.

References

1. P. G. Drazin: Nonlinear Systems, Cambridge University Press, Cambridge, UK, (1992)
2. R. Kutner, M. Ausloos, D. Grech, T. Di Matteo, C. Schinckus, H. E. Stanley: Econophysics and sociophysics: Their milestones & challenges, Physica A 516, 240 - 253 (2019). https://doi.org/10.1016/j.physa.2018.10.019
3. R. Lambiotte, M. Ausloos: Natural clustering: the modularity approach, Journal of Statistical Mechanics: Theory and Experiment, P08026 (2007). https://doi.org/10.1088/1742-5468/2007/08/L08001
4. N. K. Vitanov, K. Sakai, I. P. Jordanov, S. Managi, K. Demura: Analysis of a Japan government intervention on the domestic agriculture market, Physica A, 382, 330 – 335 (2006). https://doi.org/10.1016/j.physa.2007.02.025
5. R. M. May, S. A. Levin, G. Sugihara. Complex systems: ecology for bankers, Nature, 451, 893 – 895 (2008). https://doi.org/10.1038/451893a
6. S. A. Sheard, A. Mostashari: Principles of Complex Systems for Systems Engineering, Systems Engineering, 12, 295 - 311 (2009). http://www.sebokwiki.org/w/index.php?title=Principles_of_Complex_Systems_for_Systems_Engineering&oldid=65114
7. N. K. Vitanov: Science Dynamics and Research Production. Indicators, Indexes, Statistical Laws and Mathematical Models, Springer, Cham (2016).
8. M. Bahrami, N. Chinichian, A. Hosseiny, G. Jafari, M. Ausloos: Optimization of the post-crisis recovery plans in scale-free networks, Physica A, 540, 123203 (2020). https://doi.org/10.1016/j.physa.2019.123203
9. N.K. Vitanov, F. H. Busse: Upper bounds on heat transport in a horizontal fluid layer with stress-free boundaries, ZAMP, 48, Birkhäuser Verlag, Basel, 310 – 324 (1997). https://doi.org/10.1007/PL00001478
10. R. J. Lempert: A new decision sciences for complex systems, PNAS USA, 99, Suppl. 3, 7309 - 7313 (2002). https://doi.org/10.1073/pnas.082081699
11. E. V. Nikolova, N. K. Vitanov: On the Possibility of Chaos in a Generalized Model of Three Interacting Sectors, Entropy, 22(12), 1388 (2020). https://doi.org/10.3390/e22121388
12. N. K. Vitanov, M. Ausloos, G. Rotundo: Discrete model of ideological struggle accounting fot migration, Advances in Complex Systems, 15, Supp. 01, 1250049 (2012). https://doi.org/10.1142/S021952591250049X
13. J. Foster: From simplistic to complex systems in economics, Cambridge Journal of Economics, 29, 873 - 892 (2005). https://doi.org/10.1093/cje/bei083

14. N.K. Vitanov: Upper bound on the heat transport in a horizontal fluid layer of infinite Prandtl number, Physics Letters A, 248, 338-346, (1998). https://doi.org/10.1016/S0375-9601(98)00674-4

15. N. K.Vitanov, K. N. Vitanov: Box model of migration channels, Mathematical Social Sciences, 80, 108 – 114 (2016). https://doi.org/10.1016/j.mathsocsci.2016.02.001

16. L. Cameron, D. Larsen-Freeman: Complex Systems and Applied Linguistics, Journal of Applied Linguistics, 17, 226 - 239 (2007). https://doi.org/10.1111/j.1473-4192.2007.00148.x

17. L. A. N. Amaral, A. Scala, M. Barthelemy, H. E. Stanley: Classes of Small-World Networks, Proceedings of the National Academy of Sciences, 97(21), 11149 – 11152 (2000). https://doi.org/10.1073/pnas.200327197

18. N. K. Vitanov, K. N. Vitanov, H. Kantz: On the Motion of Substance in a Channel of a Network: Extended Model and New Classes of Probability Distributions, Entropy, 22, 1240 (2020). https://doi.org/10.3390/e22111240

19. H. Kantz, T. Schreiber: Nonlinear Time Series Analysis, Cambridge University Press, Cambridge, UK (2004)

20. A. S. Pikovsky, D. L. Shepelyansky: Destruction of Anderson Localization by a Weak Nonlinearity, Phys. Rev. Lett., 100, American Physical Society, 094101 (2008). https://doi.org/10.1103/PhysRevLett.100.094101

21. T. Boeck, N. K. Vitanov: Low-dimensional chaos in zero-Prandtl-number Bénard-Marangoni convection, Physical Review E, 65, 037203 (2002). https://doi.org/10.1103/PhysRevE.65.037203

22. Z. I. Dimitrova: Numerical Investigation Of Nonlinear Waves Connected To Blood Flow In An Elastic Tube With Variable Radius, Journal of Theoretical and Applied Mechanics **45**, No. 4, 79 – 92 (2015). https://doi.org/10.1515/jtam-2015-0025

23. Y. Niu, S. Gong: Enhancing Kerr nonlinearity via spontaneously generated coherence, Phys. Rev. A, 73, 053811 (2006). https://doi.org/10.1103/PhysRevA.73.053811

24. N. K. Vitanov: Resulat connected to time series analysis and machine learning, Studies in Computational Intelligence, 934, 363 - 383 (2021). https://doi.org/10.1007/978-3-030-72284-5_17

25. R. Struble: Nonlinear Differential Equations, Dover, New York (2018)

26. N. K. Vitanov, M. Ausloos. Knowledge epidemics and population dynamics models for describing idea diffusion, in Models of Science Dynamics, edited by A.Scharnhorst, K Boerner, P. van den Besselaar, Springer, Berin, p.p. 65 – 129, (2012). https://doi.org/10.48550/arXiv.1201.0676

27. P. J. Brockwell, R. A. Davis, M. V. Calder: Introduction to Time Series and Forecasting, Springer, New York, (2002)

28. H. Kantz, D. Holstein, M. Ragwitz, N. K. Vitanov: Markov chain model for turbulent wind speed data, Physica A, 342, 315 – 321 (2004). https://doi.org/10.1016/j.physa.2004.01.070

29. A. B. Chambel: Applied chaos theory: A paradigm for complexity, Academic Press, Boston (1993)

30. K. T. Ashenfelter, S. M. Boker, J. R Waddell, N. Vitanov: Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation, Journal of Experimental Psychology: Human Perception and Performance, 35, 1072 – 1091 (2009). https://doi.org/10.1037/a0015017

31. N. K. Vitanov: Upper bounds on the heat transport in a porous layer, Physica D, 136, 322 – 339 (2000). https://doi.org/10.1016/S0167-2789(99)00165-7

32. S. Grossberg: Nonlinear neural networks: Principles, mechanisms, and architectures, Neural Networks, 1, 17-61 (1988). https://doi.org/10.1016/0893-6080(88)90021-4

33. N. K. Vitanov, N. Hoffmann, B. Wernitz: Nonlinear time series analysis of vibration data from a friction brake: SSA, PCA, and MFDFA, Chaos Solitons and Fractals, 69, 90 – 99 (2014). https://doi.org/10.1016/j.chaos.2014.09.010

34. S. Goldstein: Social Psychology and Nonlinear Dynamical Systems Theory, Phychological Inquiry, 8 , 125 - 128 (1997). https://doi.org/10.1207/s15327965pli0802_6

35. N. K. Vitanov, Z. Dimitrova, H. Kantz: On the trap of extinction and its elimination, Physics Letters A, 346, 350-355 (2006). https://doi.org/10.1016/j.physleta.2005.09.043
36. N. K. Vitanov, K. N. Vitanov: Statistical distributions connected to motion of substance in a channel of a network, Physica A, 527, 121174 (2019). https://doi.org/10.1016/j.physa.2019.121174
37. R. Borisov, Z. I. Dimitrova, N. K. Vitanov: Statistical Characteristics of Stationary Flow of Substance in a Network Channel Containing Arbitrary Number of Arms, Entropy, 22, 553 (2020). https://doi.org/10.3390/e22050553
38. N. K. Vitanov, R. Borisov, K. N. Vitanov: On the motion of substance in a channel and growth of random networks, Physica A, **581**, 126207 (2021). https://doi.org/10.1016/j.physa.2021.126207
39. S. Goldstein: Social Psychology and Nonlinear Dynamical Systems Theory, Phychological Inquiry, 8, 125 - 128 (1997). https://doi.org/10.1207/s15327965pli0802_6
40. L. Hammyl, N. Gilbert, *Agent-based models in economics*. (Wiley, Chichester, UK, 2016).
41. I. P. Jordanov: On the nonlinear waves in (2+ 1)-dimensional population systems, Comptes rendus de l'Academie Bulgare des sciences **61**, 307–314 (2008)
42. I. P. Jordanov: Nonlinear waves caused by diffusion of population members, Comptes rendus de l'Academie Bulgare des sciences **62**, 33–40 (2009)
43. I. P. Jordanov: Coupled Kink Population Waves, Journal of Theoretical and Applied Mechanics, **40**, n.2,93-98 (2010)
44. I. P. Jordanov, AIP Conference Proceedings, 2459, 030016 (2022)
45. I. P. Jordanov, Z. I. Dimitrova: On Nonlinear Waves of Migration, Journal of Theoretical and Applied Mechanics, **40**, n.1 89–96 (2010)
46. I. P. Jordanov, E. V. Nikolova: On nonlinear waves in the spatio-temporal dynamics of interacting populations, Journal of Theoretical and Applied Mechanics, **43**, 69–76 (2013). https://doi.org/10.2478/jtam-2013-0015
47. I. N. Dushkov, I. P. Jordanov, N. K. Vitanov: Numerical modeling of dynamics of a population system with time delay, Mathematical Methods in the Applied Sciences, 41, 8377 – 8384 (2017). https://doi.org/10.1002/mma.4553
48. N. K. Vitanov, I. P. Jordanov, Z. I. Dimitrova: On nonlinear population waves, Applied Mathematics and Computation, 215, 2950–2964 (2009). https://doi.org/10.1016/j.amc.2009.09.041
49. N. K. Vitanov, I .P. Jordanov, Z. I. Dimitrova: On nonlinear dynamics of interacting populations: Coupled kink waves in a system of two populations, Communications in Nonlinear Science and Numerical Simulation, 14, 2379–2388 (2009). https://doi.org/10.1016/j.cnsns.2008.07.015
50. N. Martinov, N. Vitanov: On the correspondence between the self-consistent 2D Poisson-Boltzmann structures and the sine-Gordon waves, Journal of Physics A: Mathematical and General, **25**, L51 – L56 (1992). https://doi.org/10.1088/0305-4470/25/2/004
51. N. Martinov, N. Vitanov: On some solutions of the two-dimensional sine-Gordon equation, Journal of Physics A: Mathematical and General, 25, L419 – L426 (1992). https://doi.org/10.1088/0305-4470/25/8/007
52. N. Martinov, N. Vitanov: Running wave solutions of the two-dimensional sine-Gordon equation, J. Phys A: Math. Gen., 25(12), 3609 – 3613 (1992). https://doi.org/10.1088/0305-4470/25/12/021
53. N. K. Martinov, N. K. Vitanov: New class of running-wave solutions of the (2+1)-dimensional sine-Gordon equation, Journal of Physics A: Mathematical and General, 27, 4611 – 4618 (1994). https://doi.org/10.1088/0305-4470/27/13/034
54. N. K. Martinov, N. K. Vitanov: On self-consistent thermal equilibrium structures in two-dimensional negative-temperature systems, Canadian Journal of Physics, 72, 618 – 624 (1994). https://doi.org/10.1139/p94-079
55. N. Vitanov, N. Martinov, : On the solitary waves in the sine-Gordon model of the two-dimensional Josephson junction, Zeitschrift für Physik B, 100, 129 – 135 (1996). https://doi.org/10.1007/s002570050102
56. N. K. Vitanov: On travelling waves and double-periodic structures in two-dimensional sine-Gordon systems, Journal of Physics A: Mathematical and General, 29, 5195 – 5207 (1996). https://doi.org/10.1088/0305-4470/29/16/036

57. N. K. Vitanov: Complicated exact solutions to the 2+ 1-dimensional sine-Gordon equation, ZAMM, 78, S787 - S788 (1998)
58. N. K. Vitanov: Breather and soliton wave families for the sine-Gordon equation, Proc. Roy. Soc. London A, 454, 2409 – 2423 (1998). https://doi.org/10.1098/rspa.1998.0264
59. N. K. Vitanov: Application of simplest equations of Bernoulli and Riccati kind for obtaining exact traveling-wave solutions for a class of PDEs with polynomial nonlinearity, Communications in Nonlinear Science and Numerical Simulation, 15, 2050 – 2060 (2010). https://doi.org/10.1016/j.cnsns.2009.08.011
60. N. K. Vitanov, Z. I. Dimitrova: Application of the method of simplest equation for obtaining exact traveling-wave solutions for two classes of model PDEs from ecology and population dynamics, Communications in Nonlinear Science and Numerical Simulation 15, 2836 – 2845 (2010). https://doi.org/10.1016/j.cnsns.2009.11.029
61. N. K. Vitanov, Z. I. Dimitrova, H. Kantz: Modified method of simplest equation and its application to nonlinear PDEs, Applied Mathematics and Computation 216, 2587 - 2595, 216, 2587 – 2595 (2010). https://doi.org/10.1016/j.amc.2010.03.102
62. N. K. Vitanov: Modified method of simplest equation: powerful tool for obtaining exact and approximate traveling-wave solutions of nonlinear PDEs, Communications in Nonlinear Science and Numerical Simulation, 16, 1176 – 1185 (2011). https://doi.org/10.1016/j.cnsns.2010.06.011
63. N. K. Vitanov, Z. I. Dimitrova, K. N. Vitanov: On the class of nonlinear PDEs that can be treated by the modified method of simplest equation. Application to generalized Degasperis-Processi equation and b-equation, Communications in Nonlinear Science and Numerical Simulation, 16, 3033 – 3044 (2011). https://doi.org/10.1016/j.cnsns.2010.11.013
64. N. K. Vitanov: On modified method of simplest equation for obtaining exact and approximate solutions of nonlinear PDEs: the role of the simplest equation, Communications in Nonlinear Science and Numerical Simulation, 16, 4215 – 4231 (2011). https://doi.org/10.1016/j.cnsns.2011.03.035
65. N. K. Vitanov: On modified method of simplest equation for obtaining exact solutions of nonlinear PDEs: case of elliptic simplest equation, Pliska Stud. Math. Bulgar., 21, 257 – 266 (2012)
66. N. K. Vitanov, Z. I. Dimitrova, H. Kantz: Application of the method of simplest equation for obtaining exact traveling-wave solutions for the extended Korteweg-de Vries equation and generalized Camassa-Holm equation, Applied Mathematics and Computation, 219, 7480 – 7492 (2013). https://doi.org/10.1016/j.amc.2013.01.035
67. N. K. Vitanov, Z. I. Dimitrova, K. N. Vitanov: Traveling waves and statistical distributions connected to systems of interacting populations, Computers & Mathematics with Applications, 66, 1666 – 1684 (2013). https://doi.org/10.1016/j.camwa.2013.04.002
68. N. K. Vitanov, Z. I. Dimitrova. Applied Mathematics and Computation, **247**, 213 – 217 (2014). https://doi.org/10.1098/rspa.1998.0264
69. N. K. Vitanov, Z. I. Dimitrova, T. I. Ivanova. Applied Mathematics and Computation, **315**, 372 – 380 (2017).
70. N. K. Vitanov, Z. I. Dimitrova, K. N. Vitanov. Applied Mathematics and Computation, **269**, 363 – 378 (2015).
71. N. K. Vitanov. Pliska Studia Mathematica Bulgarica, **30**, 29 – 42 (2019).
72. N. K. Vitanov. Journal of Theoretical and Applied Mechanics, **49**, 107 – 122 (2019).
73. N. K. Vitanov. AIP Conference Proceedings **2159**, 030038 (2019).
74. N. K. Vitanov, Z. I. Dimitrova. AIP Conference Proceedings **2159**, 030039 (2019).
75. E. V. Nikolova, I. P. Jordanov, Z. I. Dimitrova, N. K. Vitanov. AIP Conference Proceedings, **1895**, 07002 (2017).
76. N. K. Vitanov,Z. I. Dimitrova. Journal of Theoretical and Applied Mechanics, Sofia, **48**, No. 1, 59 – 68 (2018).
77. I. P. Jordanov, N. K. Vitanov. Studies in Computational Intelligence **793**, 199-210 (2019).
78. N. K. Vitanov, Z. I. Dimitrova, K. N. Vitanov. Entropy, **23**, 10 (2021).

79. N. K. Vitanov. Schrödinger Equation and Nonlinear Waves, in Understanding the Schrödinger Equation edited by V. Simpao, H. Little, Nova Science Publishers, New York, pp. 37 - 92 (2020)
80. N. K. Vitanov, Z. I. Dimitrova, AIP Conference Proceeding, **2321**, 030036 (2021).

On the Square of Laplacian with Inverse Square Potential

Vladimir Georgiev and Mario Rastrelli

Abstract The work treats the domain of 3D Laplace operator A with inverse square potential and its square. In the case when this operator is essentially self-adjoint, its domain has an explicit representation in terms of classical Sobolev spaces and the value of the function at the origin. If the coefficient of the inverse square potential is sufficiently large, then a similar characterization is obtained for A^2.

Keywords Self-adjoint operator · Inverse square potential · Rellich inequality

1 Introduction

In this work, we study the domain of the operator

$$A = -\Delta + \frac{q(|x|)}{|x|^2} \tag{1}$$

and its square. Here $x \in \mathbb{R}^3$ and $q(|x|) \in C^1(\mathbb{R}^3)$ is a radial positive decreasing function that satisfies

$$\inf q(|x|) > \beta_1 = \frac{3}{4}. \tag{2}$$

For $q(|x|) = \beta > \beta_1$, it is well-known that the operator is essentially self-adjoint (see [1, 2]). Hence, one can start with domain $C_0^\infty(\mathbb{R}^n \backslash \{0\})$, where the operator is

V. Georgiev (✉) · M. Rastrelli
Dipartimento di Matematica, Universitá di Pisa, Largo B. Pontecorvo 5, 56127 Pisa, Italy
e-mail: georgiev@dm.unipi.it

V. Georgiev
Faculty of Science and Engineering, Waseda University, 3-4-1, Tokyo, Okubo, Shinjuku-ku 169-8555, Japan

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. Georgi Bonchev Str. Block 8, 1113 Sofia, Bulgaria

symmetric and its graph closure gives the unique self-adjoint extension that shall be denoted again by A.

So, we know that the operator $-\Delta + \beta/|x|^2$ is essentially self-adjoint if $\beta > \beta_1 = \frac{3}{4}$. In particular, A coincides with the Friedrichs extension of the symmetric operator operator $-\Delta + \beta/|x|^2$ with domain $C_0^\infty(\mathbb{R}^n \backslash \{0\})$.

In [3], the following inclusion is established

Theorem 1 (Characterization of $D(A)$, see *[3]*) *If $q(|x|)$ satisfies* (2), *then $D(A) = H^2(\mathbb{R}^3) \cap \{u(0) = 0\}$.*

The key question that we treat in this work is to give a characterization of the square of A, when $q(x) = \beta$.

Our main result is the following.

Theorem 2 (Characterization of $D(A^2)$) *If*

$$\beta > \beta_1, \ \beta \neq \beta_2, \ \beta_2 = \frac{35}{4} = 8 + \beta_1 \beta \neq \beta_2 - 2, \beta \neq \beta_2 - 6 \qquad (3)$$

then

$$D(A^2) = H^4(\mathbb{R}^3) \cap \{u(0) = u'(0) = u''(0) = 0\}. \qquad (4)$$

2 Domain of $D(A)$

Here for completeness, we give a proof of Theorem 1. From now on, we will call

$$C_{(0)}^\infty = C_0^\infty(\mathbb{R}^3 \backslash \{0\}) \qquad (5)$$

Our first step is the proof of the inclusion $D(A) = H^2(\mathbb{R}^3) \cap \{u(0) = 0\}$.

Proposition 1 ($D(A) \subseteq H^2$) *Let $u \in D(A)$, then $u \in H^2$.*

Proof From the definition of $D(A)$, we know that $u \in D(A)$ means that there exists a sequence $u_k \in C_{(0)}^\infty$ such that

$$u_k \longrightarrow u \text{ in } L^2$$
$$Au_k \longrightarrow Au \text{ in } L^2.$$

To prove that u is in H^2, it is sufficient to prove that Δu is in L^2. So we will prove that u/r^2 is in L^2. Let's write the formula of A for u_k

$$-\Delta u_k + \frac{q(r)}{r^2} u_k = f_k \in L^2$$

Let's scalar multiply for u_k/r^2 in L^2. From now on, we will omit the ω_n term for the radial integrals. We obtain

$$-\int_{\mathbb{R}^3} \Delta u_k \cdot \frac{u_k}{r^2}\, dx + \int_{\mathbb{R}^3} \frac{q(r)}{r^4} u_k^2\, dx = \int_{\mathbb{R}^3} f_k \frac{u_k}{r^2}\, dx. \tag{6}$$

Using the relation $\nabla r^{-k} = -k\frac{x}{r^{k+2}}$, we find

$$-\int_{\mathbb{R}^3} \Delta u_k \cdot u_k\, r^{-2} dx = \int_{\mathbb{R}^3} |\nabla u_k|^2 \frac{dx}{r^2} - \int_{\mathbb{R}^3} |u_k|^2 \frac{dx}{r^4}.$$

Therefore, we arrive at

$$\left(A u_k, \frac{u_k}{r^2}\right)_{L^2} = \left\|\frac{\nabla u_k}{r}\right\|_{L^2}^2 - \left\|\frac{u_k}{r^2}\right\|_{L^2}^2 + \int_{\mathbb{R}^3} \frac{q(r)}{r^4} |u_k|^2 dx \geq$$

$$\geq \left\|\frac{\nabla u_k}{r}\right\|_{L^2}^2 + (\beta - 1)\left\|\frac{u_k}{r^2}\right\|_{L^2}^2.$$

Due to Hardy's inequality (see (33) in the Appendix), with $p = 2$ and $q = 1$, the following inequality holds: $\left\|\frac{\nabla u_k}{r}\right\|_{L^2}^2 \geq \frac{1}{4}\left\|\frac{u_k}{r^2}\right\|_{L^2}^2$ so we arrive at $\left(A u_k, \frac{u_k}{r^2}\right)_{L^2} \geq \left(\beta - \frac{3}{4}\right)\left\|\frac{u_k}{r^2}\right\|_{L^2}^2$ and hence we have the estimate

$$\left\|\frac{u_k}{r^2}\right\|_{L^2} \leq \frac{1}{\beta - 3/4}\|A u_k\|_{L^2}. \tag{7}$$

This estimate implies

$$\|\Delta u_k\|_{L^2} \leq \|A u_k\|_{L^2} + \frac{1}{\beta - 3/4}\|A u_k\|_{L^2}. \tag{8}$$

Taking the limit $k \to \infty$ we complete the proof. $\qquad\square$

In the previous proof, we have obtained the following key estimates:

$$\|\Delta u\|_{L^2} \lesssim \|A u\|_{L^2}, \tag{9}$$

and

$$\left\|\frac{u}{r^2}\right\|_{L^2} \lesssim \|A u\|_{L^2}. \tag{10}$$

Proof (*Proof of Theorem* 1) Once we have the inclusion $D(A) \subset H^2$, we use the Sobolev embedding, and from $u \in D(A)$, we see that u is a continuous function, and from the convergence of the integral

$$\int_{\mathbb{R}^3} \frac{|u(x)|^2}{|x|^2} dx$$

, we get $u(0) = 0$. So we have the inclusion $D(A) \subseteq H^2(\mathbb{R}^3) \cap \{u(0) = 0\}$.

It remains to prove the opposite inclusion $H^2(\mathbb{R}^3) \cap \{u(0) = 0\} \subseteq D(A)$.

But this follows from Rellich's inequality (see (34) in the Appendix).

Let's consider $u \in H^2$, it is sufficient to check that u/r^2 is in L^2. Thanks to Hardy's inequality with $p = 2$ and $q = n - 2$, and after the Rellich's one with $k = 1$, we have

$$\left\| \frac{1}{r^2} u \right\|_{L^2}^2 \leq \left(\frac{4}{3} \right)^2 \|\Delta u\|_{L^2}^2 .$$

This completes the proof.   □

Observation 1 *In the proof, we have obtained the following fundamental inequality:*

$$\left\| \frac{1}{r^2} u \right\|_{L^2} \lesssim \|\Delta u\|_{L^2} . \tag{11}$$

3   Definition and Domain of A^2

Thanks to the definition of $D(A)$ we can define A^2. The square is for the composition and, because we know how A^2 behaves on $C_{(0)}^\infty$, the closure is well defined and is essentially self-adjoint too.

If we consider $u \in C_{(0)}^\infty$, it can be easily checked that

$$A^2 u = \Delta^2 u - \beta \Delta \left(\frac{u}{r^2} \right) - \frac{\beta}{r^2} \Delta u + \frac{\beta^2}{r^4} u. \tag{12}$$

For simplicity, we shall reduce the proof of the main result to the case of radial functions, since adding integration in angular variables we can get the general result.

Proposition 2 $(H_{rad}^4 \subseteq D(A^2))$ *Let* $u \in H_{rad}^4$, *then* $u \in D(A^2)$.

Proof It is sufficient to prove that $\|A^2 u\|_{L^2}$ is finite. So we compute the L^2 norm of every term.

First of all, $\|\Delta^2 u\|_{L^2}$ is finite because u is in H^4. Calling $\Delta u = g \in H^2$ and using (11), we obtain

$$\left\| \frac{1}{r^2} \Delta u \right\|_{L^2} \lesssim \|\Delta^2 u\|_{L^2} \tag{13}$$

and we have the estimate on the third term. We can repeat the computation done in Proposition 1 for the forth term, Hardy with $p = 2$ and $q = -3$ and Rellich with

$k = 3$.

$$\left\|\frac{1}{r^4}u\right\|_{L^2}^2 = \int_0^\infty \left|\frac{1}{r}u(r)\right|^2 r^{-4}dr \leq \left(\frac{2}{5}\right)^2 \left\|\frac{1}{r^3}\partial_r u\right\|_{L^2}^2$$
$$\leq \left(\frac{4}{35}\right)^2 \left\|\frac{1}{r^2}\Delta u\right\|_{L^2}^2, \tag{14}$$

and thanks to (13), we can write

$$\left\|\frac{1}{r^4}u\right\|_{L^2} \lesssim \left\|\frac{1}{r^2}\Delta u\right\|_{L^2} \lesssim \left\|\Delta^2 u\right\|_{L^2} \tag{15}$$

and

$$\left\|\frac{1}{r^3}\partial_r u\right\|_{L^2} \lesssim \left\|\frac{1}{r^2}\Delta u\right\|_{L^2} \lesssim \left\|\Delta^2 u\right\|_{L^2} \tag{16}$$

We conclude with the Leibnitz formula for the Laplacian $\Delta(fg) = \Delta f \cdot g + f \cdot \Delta g + 2\nabla f \nabla g$, we have, indeed

$$\Delta\left(\frac{u}{r^2}\right) = \frac{1}{r^2}\Delta u + u\Delta\left(\frac{1}{r^2}\right) + 2\nabla u \cdot \nabla\left(\frac{1}{r^2}\right).$$

So, because $2\nabla u \cdot \nabla\left(\frac{1}{r^2}\right) = -2\frac{2}{r^3}\nabla u \cdot \hat{r} = -\frac{4}{r^3}\partial_r u$ and $\Delta\left(\frac{1}{r^2}\right) = \frac{1}{r^2}\partial_r\left(r^2\partial_r\frac{1}{r^2}\right) = \frac{2}{r^4}$, the following estimate can be computed

$$\left\|\Delta\left(\frac{u}{r^2}\right)\right\|_{L^2} \leq \left\|\frac{1}{r^2}\Delta u\right\|_{L^2} + 2\left\|\frac{1}{r^4}u\right\|_{L^2} + 4\left\|\frac{1}{r^3}\partial_r u\right\|_{L^2}.$$

Finally, thanks to (13), (15), and (16), we can conclude that

$$\|A^2 u\|_{L^2} \lesssim \left\|\Delta^2 u\right\|_{L^2}. \tag{17}$$

Lemma 1 If $\beta > \beta_2 = 8 + \beta_1$, then $D_{rad}(A^2) = H_{rad}^4 \cap \{u(0) = u'(0) = u''(0) = 0\}$.

Proof Let's consider $u \in D(A^2)$, in particular, $u \in D(A)$ and $Au = f \in D(A)$. So, thanks to (10), we have

$$\left\|\frac{1}{r^2}f\right\|_{L^2} \lesssim \|Af\|_{L^2} = \|A^2 u\|_{L^2} \tag{18}$$

Let's consider a sequence $\{u_k\}_{k=1}^\infty \subseteq C_{(0)}^\infty$ such that $u_k \to u$ in H^2, $f_k = Au_k \to f$ in H^2 and $A^2 u_k \to A^2 u$ in L^2, and let's compute the following scalar product in L^2:

$$\left(-\frac{1}{r^2}\Delta u_k + \frac{\beta}{r^4}u_k, \frac{1}{r^4}u_k\right)_{L^2}. \tag{19}$$

With Cauchy–Schwartz and (18), we obtain

$$(19) \leq \left\|\frac{1}{r^2}f_k\right\|_{L^2}\left\|\frac{1}{r^4}u_k\right\|_{L^2} \lesssim \|A^2 u_k\|_{L^2}\left\|\frac{1}{r^4}u_k\right\|_{L^2} \tag{20}$$

On the other hand

$$(19) = -\int_0^\infty \Delta u_k\, u_k\, r^{n-7}dr + \beta\left\|\frac{1}{r^4}u_k\right\|_{L^2}^2. \tag{21}$$

Now, as done for the $D(A)$ case, let's study the first term

$$-\int_0^\infty \Delta u_k\, u_k\, r^{-4}dr = -\int_0^\infty \partial_r^2 u_k\, u_k\, r^{-4}dr - (n-1)\int_0^\infty \partial_r u_k\, u_k\, r^{-5}dr,$$

We integrate by parts

$$-\int_0^\infty \partial_r^2 u_k\, u_k\, r^{-4}dr = \int_0^\infty (\partial_r u_k)^2\, r^{-4}dr + (n-7)\int_0^\infty \partial_r u_k\, u_k\, r^{-5}dr,$$

and we substitute

$$-\int_0^\infty \Delta u_k\, u_k\, r^{-4}dr = \int_0^\infty (\partial_r u_k)^2\, r^{-4}dr - 6\int_0^\infty \partial_r u_k\, u_k\, r^{-5}dr.$$

Thanks to Hardy, with $p = 2$ and $q = -3$, we can write

$$\left.\begin{aligned}
\left\|\frac{1}{r^3}\partial_r u_k\right\|_{L^2}^2 &= \int_0^\infty (\partial_r u_k)^2\, r^{-4}dr \geq \\
&\geq \left(\frac{5}{2}\right)^2\int_0^\infty \left|\frac{u_k}{r}\right|^2\, r^{-3}dr = \left(\frac{5}{2}\right)^2\left\|\frac{1}{r^4}u_k\right\|_{L^2}^2,
\end{aligned}\right\} \tag{22}$$

and, integrating by parts again

$$-6\int_0^\infty \partial_r u_k\, u_k\, r^{-5}dr = -3\int_0^\infty \partial_r(u_k^2)\, r^{-5}dr =$$
$$= 3(-5)\int_0^\infty u_k^2\, r^{-6}dr = -15\left\|\frac{1}{r^4}u_k\right\|_{L^2}^2.$$

Adding all terms, the following inequality holds:

$$-\int_0^\infty \Delta u_k\, u_k\, r^{-4} dr \geq -\frac{35}{4}\left\|\frac{1}{r^4}u_k\right\|_{L^2}^2$$

so we have two estimates on the scalar product

$$\left(\beta - \frac{35}{4}\right)\left\|\frac{1}{r^4}u_k\right\|_{L^2}^2 \leq (19) \lesssim \|A^2 u_k\|_{L^2}\left\|\frac{1}{r^4}u_k\right\|_{L^2}. \tag{23}$$

Simplifying, we have computed

$$\left\|\frac{1}{r^4}u_k\right\|_{L^2} \lesssim \|A^2 u_k\|_{L^2}. \tag{24}$$

Moreover, due to the convergence, we have obtained

$$\left\|\frac{1}{r^4}u\right\|_{L^2} \lesssim \|A^2 u\|_{L^2}. \tag{25}$$

Let's now write respect to u the inequality (18):

$$\left\|-\frac{1}{r^2}\Delta u + \frac{\beta}{r^4}u\right\|_{L^2} \lesssim \|A^2 u\|_{L^2},$$

adding (25), we have estimated also

$$\left\|\frac{1}{r^2}\Delta u\right\|_{L^2} \lesssim \|A_F^2 u\|_{L^2}. \tag{26}$$

Repeating the same computation as in (14) with Rellich, we have

$$\left\|\frac{1}{r^3}\partial_r u\right\|_{L^2} \lesssim \left\|\frac{1}{r^2}\Delta u\right\|_{L^2} \lesssim \|A^2 u\|_{L^2}. \tag{27}$$

Now the thesis follows because, thanks to triangular inequality, we can write

$$\|\Delta^2 u\|_{L^2} \lesssim \|A^2 u\|_{L^2}. \tag{28}$$

The conditions for u and its derivatives in 0 follow from Sobolev embeddings. Indeed, the convergence for $u, \partial_r u$ and $\partial_r^2 u$ is uniform, in particular pointwise.

Lemma 2 *If $\beta_1 < \beta < \beta_2$, then $D_{rad}(A^2) = H_{rad}^4 \cap \{u(0) = u'(0) = u''(0) = 0\}$;*

Proof It is sufficient to consider the following scalar product in L^2, instead of (19),

$$\left(-\frac{1}{r^2}\Delta u_k + \frac{\beta}{r^4}u_k, -\frac{1}{r^3}\partial_r u_k\right)_{L^2}. \tag{29}$$

Integrating by parts, we compute $\left(-\frac{1}{r^2}\Delta u_k + \frac{\beta}{r^4}u_k, -\frac{1}{r^3}\partial_r u_k\right)_{L^2} = \frac{7}{2}\left\|\frac{1}{r^3}\partial_r u\right\|_{L^2}^2 -$
$\beta\frac{5}{2}\left\|\frac{1}{r^4}u\right\|_{L^2}^2$, and, due to Cauchy-Schwartz, we obtain:

$$\frac{7}{2}\left\|\frac{1}{r^3}\partial_r u\right\|_{L^2}^2 - \frac{5\beta}{2}\left\|\frac{1}{r^4}u\right\|_{L^2}^2 \leq \left\|\frac{1}{r^2}f\right\|_{L^2}\left\|\frac{1}{r^3}\partial_r u\right\|_{L^2}. \tag{30}$$

Because $\beta > \beta_1 > 0$ and (22), we can write

$$-\frac{5\beta}{2}\left\|\frac{1}{r^4}u_k\right\|_{L^2}^2 \geq -\frac{2\beta}{5}\left\|\frac{1}{r^3}\partial_r u_k\right\|_{L^2}^2 \tag{31}$$

and find

$$\left(\frac{7}{2} - \frac{2\beta}{5}\right)\left\|\frac{1}{r^3}\partial_r u\right\|_{L^2}^2 \leq \left\|\frac{1}{r^2}f\right\|_{L^2}\left\|\frac{1}{r^3}\partial_r u\right\|_{L^2}. \tag{32}$$

so, for $\beta < \beta_2$, we have the boundedness of $\left\|\frac{1}{r^3}\partial_r u\right\|_{L^2}^2$, $\left\|\frac{1}{r^4}u\right\|_{L^2}^2$.

Proof *(Proof of Theorem 2)* It is sufficient to join the results of Lemmas 1 and 2.

4   Appendix: Hardy and Rellich Inequalities

Proposition 3 (Weighed Hardy's inequality) *The inequality*

$$\int_0^\infty \left|\frac{1}{r}f(r)\right|^p r^{q-1}\, dr \leq \left(\frac{p}{|p-q|}\right)^p \int_0^\infty |f'(r)|^p\, r^{q-1}dr \tag{33}$$

holds:

– *if $p > q$ and $f(0) = 0$,*
– *if $p < q$ and $f(\infty) = 0$.*

Proof The Proof is in Corollary 1.2.9 in [4]

Thanks to Hardy's inequality we can prove a particular case of the Rellich's inequality.

Proposition 4 (Radial Rellich's inequality) *Let u be a sufficiently smooth radial function, then the following inequality holds for every real numbers $n > 0$, $k \geq 1 - \frac{n}{2}$:*

$$\int_0^\infty \left|\frac{\partial_r u(r)}{r^k}\right|^2 r^{n-1}dr \leq \left(\frac{2}{n+2k-2}\right)^2 \int_0^\infty \left|\frac{\Delta u(r)}{r^{k-1}}\right|^2 r^{n-1}dr. \tag{34}$$

Proof

$$\int_0^\infty \left| \frac{\partial_r u(r)}{r^k} \right|^2 r^{n-1} dr = \int_0^\infty \left| \frac{r^{n-1}\partial_r u(r)}{r} \right|^2 r^{3-n-2k} dr$$

$$\leq \left(\frac{2}{n+2k-2} \right)^2 \int_0^\infty \left| \frac{1}{r^{k-1}} \frac{1}{r^{n-1}} \partial_r \left(r^{n-1}\partial_r u(r) \right) \right|^2 r^{n-1} dr$$

$$= \left(\frac{2}{n+2k-2} \right)^2 \int_0^\infty \left| \frac{1}{r^{k-1}} \Delta u(r) \right|^2 r^{n-1} dr,$$

where, in the second line, we used the Hardy's inequality for $p = 2, q = 4 - n - 2k$ and $f(r) = r^{n-1}\partial_r u(r)$, which in zero is equal to zero.

References

1. Kalf, H. and Schmincke, U.-W. and Walter, J. and Wüst, R., *On the spectral theory of Schrödinger and Dirac operators with strongly singular potentials*, Spectral theory and differential equations, pp. 182–226, Springer, Berlin (1975).
2. Simon, B., *Essential self-adjointness of Schrödinger operators with singular potentials*, Arch. Rational Mech. Anal. Vol. 52, (1973), pp. 44–48, https://doi.org/10.1007/BF00249091.
3. Georgiev, V. and Kubo, H, *Global solvability for nonlinear wave equations with singular potential*, arxiv, 2022, arXiv:2201.07128.
4. Balinsky, Alexander A. and Evans, W. Desmond and Lewis, Roger T. *The Analysis and Geometry of Hardy's Inequality*, Springer (2010), ISBN 978-3-319-22869-3, https://doi.org/10.1007/978-3-319-22870-9.

Applications in Mathematical Biology

Parameter Recovery Study of Honeybee Colony Failure Due to Nutritional Deficiency

Atanas Atanasov and Slavi Georgiev

Abstract The global decline of honeybee colony number, called Colony Collapse Disorder, is a long-lasting and severe problem. It has been researched a lot, but some of the causal factors are still unknown. Nevertheless, it is well agreed that poor nutrition often leads to disruption in the colony routine and it is indeed a primary reason for population decline. In this study, we adopt a compartment model that considers the distinct population of brood, hive, and forager bees and the amount of stored pollen and nectar. The paper proposes a robust approach to calibrate the mathematical model from experimental data in order to study the rates of change in the food consumption. The obtained values of the dynamics parameters are not directly observable but vital to explore the impact of the behavioral changes occurring in the case of nutritional stress. Such information is of extreme importance to choose and apply the best beekeeping policy to prevent or cure colony failure.

Keywords Honeybee population dynamics · Colony Collapse Disorder · Nutrition · Parameter identification

A. Atanasov
Department of Agricultural Machinery, Agrarian and Industrial Faculty, University of Ruse, 8 Studentska Str., 7004 Ruse, Bulgaria
e-mail: aatanasov@uni-ruse.bg

S. Georgiev (✉)
Department of Informational Modeling, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 8 Acad. Georgi Bonchev Str., 1113 Sofia, Bulgaria
e-mail: sggeorgiev@math.bas.bg; sggeorgiev@uni-ruse.bg

Department of Applied Mathematics and Statistics, Faculty of Natural Sciences and Education, University of Ruse, 8 Studentska Str., 7004 Ruse, Bulgaria

1 Introduction

The honeybee is one of the most important species on the Earth. It is the main known pollinator of plants around the world. The life on the planet is impossible in case of bees absence. Honeybees collect food in the form of nectar and pollen from the surrounding environment, then process it, and store it for the whole colony population. Recently, massive colony declines were observed all over the world, while they were caused by multiple but unclear factors. This phenomenon, characterized by the absence of forage bees (those bees which fly and gather food) with the presence of a small number of hive bees (those who care for the hive and the brood) and the unattended queen, is called Colony Collapse Disorder (CCD) [2, 3, 15, 18]. Since it is agreed that not a single factor causes this syndrome, many reasons are associated to the colony failure, such as pesticides, viruses, bacteria, fungi, harsh weather conditions, and nutritional stress. It is concluded that the joint effect of the multiple stressors causes the colony collapse [10–12, 16].

In this paper, we adopt the model proposed in [4]. It distinguishes five classes of bees—eggs and larvae, pupae, hive bees, pollen forager bees, and nectar forager bees. The other two compartments are the amount of stored pollen and nectar. The model explores the consumption and impact of the food on the honey bees during different stages of their lives. It provides a simple theoretical framework to investigate how the presence of pollen and nectar influences the colony population dynamics.

While the mathematical models provide a tool for simulation, system behavior is driven by the model parameters. Unfortunately, most of them are not directly observable or measurable in practice. Their values, however, together with the respective sensitivities, determine the outcome of the experiment and thus—the possible scenario in the real world. So the knowledge of the parameter values is of paramount importance for the proper honeybee management. What is usually measurable, though, is the values of the model functions. In our case, they are the population size of the different classes and the amount of the stored food. There exists appropriate hardware which is capable of measuring the respective quantities with satisfying precision. Thus, we are able to define the *inverse problem* of deriving the parameter values given observations of the model functions, which will be rigorously defined in the subsequent sections.

Such inverse problems, concerning similar models, are solved by different mathematical methods in [1–3, 10].

The next section is dedicated to the detailed explanation of the mathematical model. In Sect. 3, the inverse problem is defined, while in Sect. 4 the direct and inverse problems are solved, respectively. The experiments are presented in Sect. 5 and the paper is concluded in the last section.

2   Mathematical Model

Almost all compartment models regard the rates of change in the population size of the different classes of honeybees. The stages of the bee life are four—an egg, a larva, a pupa, and a mature bee. Usually, there is a single queen in a colony and her main job is to lay eggs. A healthy queen lays up to 1500–2000 eggs per day. There are three types of adult bees—a queen, female workers, and male drones. The latter emerge from non-fertilized eggs, comprise less than 1/20 from the adults, and do not contribute to the colony work, thus being usually neglected in modeling. The queen and workers emerge from fertilized eggs and all of them are fed with royal jelly during the first three days of their lives. Then, those larvae, chosen to become queens, are continued to be fed with royal jelly, while the others are fed with nectar or diluted honey and pollen. After pupation, the young hive bees clean the hive and feed the larvae and the queen, following the responsibilities of building and guarding the nest. When they mature enough (usually after about twenty days), they could be recruited to become forage bees. The foragers fly outside the hive to gather nectar, pollen, and water. The behavioral maturation of the hive bees is regulated by a pheromone and it is known as social inhibition. If there is a plenty of foragers, the hive bees are recruited at a later age, and vice versa. It is also possible for foragers to go back to hive duties.

In the present model, the forage bees are differentiated between pollen and nectar foragers. The first ones are known for their large pollen loads, while the second ones are recognized by their extended abdomens. The forager behavior primarily depends on the colony needs. The collected amount of pollen and nectar heavily depends on the available quantity in the environment. As mentioned, consuming pollen, the hive bees produce royal jelly, consumed on its turn by the queen, adult bees, and the uncapped brood. The latter is composed of the eggs and larvae. After a couple of days, the hive bees seal the cells and the brood becomes capped or pupae.

Let us denote the population size of the uncapped brood with B_u and of the capped brood with B_c. Further, as usual, the population size of the hive bees is denoted with H, and the population size of the pollen forage bees and the nectar forage bees—with F_p and F_n, respectively. Similarly, the amount (in grams) of stored pollen and nectar is denoted with f_p and f_n. The time t is measured in days. The model is constituted by the following system of ordinary differential equations (ODEs):

$$\frac{dB_u}{dt} = L \cdot S(H, f_p, f_n) - \phi_u B_u, \tag{1a}$$

$$\frac{dB_c}{dt} = \phi_u B_u - (\varphi_c + m_c) B_c, \tag{1b}$$

$$\frac{dH}{dt} = \varphi_c B_c - H \cdot \big(R_p(H, F_p, F_n, f_p) + R_n(H, F_p, F_n, f_n)\big), \tag{1c}$$

$$\frac{dF_p}{dt} = H \cdot R_p(H, F_p, F_n, f_p) - m_p F_p, \tag{1d}$$

$$\frac{\mathrm{d}F_n}{\mathrm{d}t} = H \cdot R_n(H, F_p, F_n, f_n) - m_n F_n, \tag{1e}$$

$$\frac{\mathrm{d}f_p}{\mathrm{d}t} = \mu_p(t) c F_p - \gamma_u B_u - \gamma_H H, \tag{1f}$$

$$\frac{\mathrm{d}f_n}{\mathrm{d}t} = \mu_n(t) c F_n - \lambda_u B_u - \lambda_A (H + F_p + F_n). \tag{1g}$$

Equation (1a) of model (1a–1g) describes the rate of change of the uncapped brood, constituted by eggs and larvae. L is the number of eggs laid by the queen per day. $S(\cdot)$ is a saturation function determining the survival rate of the uncapped brood. It is assumed to depend on the number of hive bees that take care after the uncapped brood and the amount of the stored food. It is based on the function suggested in [14], but is extended to

$$S(H, f_p, f_n) = \frac{H}{H + \nu} \cdot \frac{f_p^2}{f_p^2 + K \cdot H} \cdot \frac{f_n}{f_n + b}. \tag{2}$$

The first term in (2) models the impact of the hive bees on the survival of the uncapped brood. Since the hive bees feed and keep warm the brood, if they are low in number, the survival rate declines. The half-saturation constant ν determines the speed of the decline.

The second and the third terms in (2) consider the decline of the survival rate if the food amounts are low. Their sigmoid forms are suggested in [14]. The parameter b is again a half-saturation constant that controls the effect of nectar shortage. Most of the protein is provided from the pollen, so the former would be enough if the amount of the collected pollen is proportional to the population size of the hive bees. The parameter K denotes the maximal amount of pollen that could be consumed by a hive bee to be saturated to produce enough royal jelly.

The second term in (1a) is the daily capping rate.

Equation (1b) models the rate of change of the capped brood. The first term describes the size of the uncapped brood that develops into the capped brood. The first part of the second term shows the number of pupae that develop into young hive bees and the constant φ_c is the emerging rate. The second part is the capped brood mortality rate.

The rate of change in the population size of the hive bees is modeled by Eq. (1c). The first constituent term is the number of the capped brood that develop into hive bees. The second one is the number of bees, recruited for pollen and nectar foragers, respectively.

The transition from hive bees to foragers is high if there is a shortage of foragers and low in case of enough foragers. The pollen recruitment function is assumed to be

$$R_p(H, F_p, F_n, f_p) = \alpha_p^{\min} + \alpha_p^{\max}\left(1 - \frac{f_p^2}{f_p^2 + K \cdot H}\right) - \delta \cdot \frac{F_p}{H + F_p + F_n}, \tag{3}$$

where $\alpha_p^{\min}$ is the minimal pollen recruitment rate when there is enough stored pollen in the hive. The second term in (3) expresses how the shortage of collected pollen increases the pollen recruitment rate, as the constant $\alpha_p^{\max}$ controls that effect. The last term in (3) models how the presence of pollen foragers decreases the recruitment rate, as the constant δ regulates the strength of the social inhibition.

Analogously, the nectar recruitment function (4) is assumed to be

$$R_n(H, F_p, F_n, f_n) = \alpha_n^{\min} + \alpha_n^{\max}\left(1 - \frac{f_n}{f_n + b}\right) - \delta \cdot \frac{F_n}{H + F_p + F_n}. \quad (4)$$

The rate of change of the pollen foragers is modeled by Eq. (1d), where the first term represents the pollen recruitment rate, and the second term—the pollen forager mortality rate. Similarly, the rate of change of the nectar forage bees is modeled by Eq. (1e).

The rate of change in the stored pollen is described with Eq. (1f). The first term accounts for the amount of pollen collected by the pollen foragers. This directly depends on the pollen availability, which is not constant throughout the year. Functions that describe the variation of the availability, for instance the peaks during the plant flowering and the dips during the winter, are proposed in [17]. We consider the following expression for $\mu_p(t)$:

$$\mu_p(t) = \frac{1}{2}\left(\cos\left(\frac{\pi}{180}t\right) + \frac{5}{2}\right).$$

The constant c denotes the maximal amount of food collected from a forager per day. The second and third terms in (1f) denote the daily consumption of pollen from the uncapped brood and the hive bees, respectively.

In a similar manner, the rate of change in the stored nectar is modeled by Eq. (1g). The nectar availability is described with the function

$$\mu_n(t) = \frac{1}{2}\left(\sin\left(\frac{\pi}{180}t\right) + \frac{11}{2}\right).$$

Again, λ_u and λ_A are the nectar consumption rates from the uncapped brood and the adult bees.

3 Definition of the Inverse Problem

This section is dedicated to the formulation of the inverse coefficient problem. This is essentially finding the "fair" values of the unknown parameters, provided with additional information about the solution to the problem (1a–1g). In practice, it is very important to know how fast the bees consume their food stores, which is given

by the levels of the rates γ_u, γ_H, λ_u, and λ_A. They are not directly measurable, but it is possible to reconstruct their values using mathematical methods.

The problem (1a–1g) is subjected to the initial condition

$$B_u(t_0) = B_u^0, \; B_c(t_0) = B_c^0, \; H(t_0) = H^0, \; F_p(t_0) = F_p^0, \; F_n(t_0) = F_n^0, \; f_p(t_0) = f_p^0, \; f_n(t_0) = f_n^0, \quad (5)$$

where $\boldsymbol{p} = (p^1, p^2, p^3, p^4)^\top$, $p^1 := \gamma_u$, $p^2 := \gamma_H$, $p^3 := \lambda_u$, $p^4 := \lambda_A$ and

$$\boldsymbol{p} \in \mathbb{S}_{\text{adm}} \equiv \left\{ \mathbb{R}^4 : \; 0 < p^r < P^r, \; r = \overline{1, 4} \right\}. \tag{6}$$

The solutions $\{ B_u(t; \boldsymbol{p}), B_c(t; \boldsymbol{p}), H(t; \boldsymbol{p}), F_p(t; \boldsymbol{p}), F_n(t; \boldsymbol{p}), f_p(t; \boldsymbol{p}), f_n(t; \boldsymbol{p}) \}, \boldsymbol{p} \in \mathbb{S}_{\text{adm}}$ to (1a–1g) are defined on the interval $t_0 \le t \le T$. The admissible set $\mathbb{S}_{\text{adm}}$ follows the biology of the honey bee [20] and the particular model [4]. If the values of the parameters γ_u, γ_H, λ_u, and λ_A are known, the problem (1a–1g)–(5) is well-posed and it is called a *direct problem*.

In the paper, it is studied the *inverse problem* of recovering the parameters $\boldsymbol{p} \in \mathbb{S}_{\text{adm}}$, which are assumed to be unknown. Further, the provided additional information is in the form of *observations* of the values of the functions

$$\left\{ B_u^{\text{obs}}(t^i), B_c^{\text{obs}}(t^i), H^{\text{obs}}(t^i), F_p^{\text{obs}}(t^i), F_n^{\text{obs}}(t^i), f_p^{\text{obs}}(t^i), f_n^{\text{obs}}(t^i) \right\},$$
$$i = 1, \ldots, I_{\text{obs}}; \; t_0 = t^1 < \ldots < t^{I_{\text{obs}}} = T \quad (7)$$

of the system (1a–1g)–(5).

Via solving the inverse problem, the goal is to reconstruct the unknown parameters $\boldsymbol{p}$ from the observation data. It is done through minimization of the functional

$$\min_{\boldsymbol{p} \in \mathbb{S}_{\text{adm}}} \Phi(\boldsymbol{p}), \qquad \boldsymbol{p} = (p^1, p^2, p^3, p^4).$$

Our target is to find the point $\boldsymbol{p} = (p^1, p^2, p^3, p^4)$ of the local minimum of the functional $\Phi(\boldsymbol{p})$. The functional $\Phi(\boldsymbol{p})$ could be written as

$$\Phi(\boldsymbol{p}) = \frac{1}{2} \sum_{i=1}^{I_{\text{obs}}} \left[\left(B_u(t^i; \boldsymbol{p}) - B_u^{\text{obs}}(t^i) \right)^2 + \left(B_c(t^i; \boldsymbol{p}) - B_c^{\text{obs}}(t^i) \right)^2 + \right.$$
$$\left(H(t^i; \boldsymbol{p}) - H^{\text{obs}}(t^i) \right)^2 + \left(F_p(t^i; \boldsymbol{p}) - F_p^{\text{obs}}(t^i) \right)^2 + \left(F_n(t^i; \boldsymbol{p}) - F_n^{\text{obs}}(t^i) \right)^2 +$$
$$\left. \left(f_p(t^i; \boldsymbol{p}) - f_p^{\text{obs}}(t^i) \right)^2 + \left(f_n(t^i; \boldsymbol{p}) - f_n^{\text{obs}}(t^i) \right)^2 \right]. \quad (8)$$

In (8) $\left\{ B_u^{\text{obs}}(t^i), B_c^{\text{obs}}(t^i), H^{\text{obs}}(t^i), F_p^{\text{obs}}(t^i), F_n^{\text{obs}}(t^i), f_p^{\text{obs}}(t^i), f_n^{\text{obs}}(t^i) \right\}$ are the experimental data (7) and $\left\{ B_u(t^i; \boldsymbol{p}), B_c(t^i; \boldsymbol{p}), H(t^i; \boldsymbol{p}), F_p(t^i; \boldsymbol{p}), F_n(t^i; \boldsymbol{p}), f_p(t^i; \boldsymbol{p}), f_n(t^i; \boldsymbol{p}) \right\}$ is the theoretical solution to problem (1a–1g)–(6).

4   Solution to the Direct and Inverse Problems

This section begins with a brief description of the numerical algorithm to solve the direct problem (1a–1g)–(5). Such kind of sophisticated models do not possess analytic solutions, so a computational approach is needed here. What is more, in the synthetic data framework, the direct problem is solved in order to take observations to explore the solution to the inverse problem.

We use the nonuniform mesh

$$\overline{\omega}_\tau = \left\{ t_0, t_1, t_2, \ldots, t_j, \ldots, t_J = T \right\} \text{ for } j = 0, \ldots, J \tag{9}$$

and the boolean mask

$$m_j = \mathbb{1}_{\{t_j \text{ is an observation point}\}} = \begin{cases} 1, & t_j = t^i \text{ for some } i \le j, \\ 0, & \text{otherwise,} \end{cases}$$

where, as mentioned, t_j, $j = \overline{0, J}$ are the nodes for the direct problem and t^i, $i = \overline{1, I_{\text{obs}}}$ is the subset of nodes where measurements are taken at. Without loss of generality, it is assumed that the observation moments coincide with some of the time nodes. If this is not the case in practice, a simple interpolation would help.

The standard numerical methods are able to solve the initial problem (1a–1g)–(5). Similar problems are solved in [2, 10], including honeybee feeding [1]. In the present paper, we employ the MATLAB$^\circledR$ ode45 subroutine with the mesh (9). This solver is based on an explicit Runge–Kutta–Dormand–Prince formula [9, 19], which could achieve fourth and fifth order of accuracy.

The inverse problem solution regards the coefficient recovery in (1a–1g)–(6), (7). It consists of finding those values of the parameters p, which reproduce the empirical dynamics as far as possible. In other words, we are looking for the point of minimum of $\Phi(p)$, which is denoted with $\check{p}$.

The argmin vector $\check{p}$ is called *nonlinear least squares estimator (LSE)*. There are a couple of methods to find $\check{p}$, and we will employ the Trust Region Reflexive algorithm [7, 8], using the MATLAB$^\circledR$ subroutine lsqnonlin.

More information about the metrics used to estimate the *goodness-of-fit* could be found in [5, 6].

5   Computational Experiments

In this section, we supply numerical simulations to validate the proposed algorithm. First, we tackle the direct problem. Further, we use the results as measurements to solve the inverse problem.

Let us demonstrate the solution to the direct problem. The values of the parameters are taken from [13, 14, 17]. The maximum number of eggs laid by the queen per day is
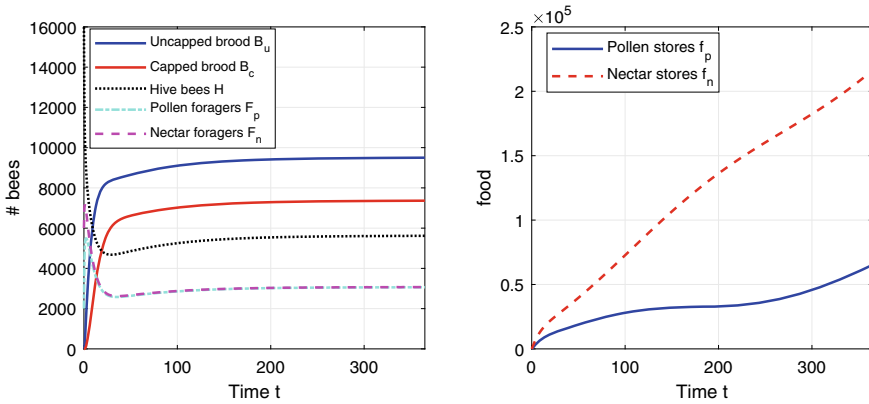
Fig. 1 Honey bee population dynamics

$L = 2000$. The number of hive bees required for surviving half the eggs is $\nu = 5000$ and the required mass of the stored nectar for surviving half the eggs is $b = 500$ g. The maximum amount of pollen that could be consumed by a hive bee to saturate the food is $K = 8$. The pupation rate of the uncapped brood is $\phi_u = 1/9$ and the emerging rate of the capped brood is $\varphi_c = 1/12$. The mortality rate of the capped brood is $m_c = 0.006$. The minimal recruitment rates of pollen and nectar foragers in the case of plenty of foragers are $a_p^{\min} = a_n^{\min} = 0.25$. The maximal recruitment rates in absence of foragers are $a_p^{\max} = a_n^{\max} = 0.25$. The parameter controlling the effect of excess of foragers is $\delta = 0.75$. The mortality rates of the pollen and nectar foragers are $m_p = m_n = 0.1$. The maximum amount of food brought to the colony from a forager per day is $c = 0.1$ g. The consumption of pollen by uncapped brood and hive bees is $\gamma_u = 0.018$ g/day and $\gamma_H = 0.007$ g/day. The consumption of nectar by uncapped brood and adult bees is $\lambda_u = 0.018$ g/day and $\lambda_A = 0.007$ g/day, respectively.

A colony with no brood in the beginning is assumed, $B_u^0 = B_c^0 = 0$, with $H^0 = 16000$ hive bees, $F_p^0 = 2000$ pollen foragers, $F_n^0 = 6000$ nectar foragers, and no food stores, $f_p^0 = f_n^0 = 0$. The considered period is one year, $t_0 = 0$ and $T = 365$. The result from the simulation is given in Fig. 1.

In the case of relative low mortality rate $m_p = m_n = 0.1$, the colony easily achieves an equilibrium state.

Now let us proceed to the solution to the inverse problem. As we discussed earlier, we aim to reconstruct the unknown food consumption rates $\boldsymbol{p} = (\gamma_u, \gamma_H, \lambda_u, \lambda_A)^\top = (0.018, 0.007, 0.018, 0.007)^\top$ (6). The observations (7) are assumed to be taken equidistantly, once in ten days. The initial values are the neutral $\boldsymbol{p}_0 = (0.01, 0.01, 0.01, 0.01)^\top$.

To begin with, a test with exact observations is conducted. This simply means that the functions are measured without any error or noise. The results are given in Table 1.

Table 1 Test with exact observations

Parameter	p_0^r	p^r	$\check{p}^r$	$\left\| p^r - \check{p}^r \right\|$	$\dfrac{\left\| p^r - \check{p}^r \right\|}{p^r}$	$\widehat{NSE}$
γ_u	0.01	0.018	0.0180	5.8981e−17	3.2767e−15	2.0624e−14
γ_H	0.01	0.007	0.0070	9.1940e−17	1.3134e−14	8.6642e−14
λ_u	0.01	0.018	0.0180	2.1858e−16	1.2143e−14	2.0487e−14
λ_A	0.01	0.007	0.0070	1.8562e−16	2.6516e−14	4.0398e−14

Table 2 Goodness-of-fit metrics with exact measurements

$\left\| \nabla \Phi(\check{\boldsymbol{p}}) \right\|_\infty$	$\delta \boldsymbol{p}_k$	$k+1$	$\Phi(\check{\boldsymbol{p}})$	$\tilde{\sigma}^2$	$\hat{\sigma}$	R^2
1.73e−5	2.7318e−17	10	5.9325e−20	1.6034e−21	4.1771e−11	1−4.0542e−31

Table 3 Tests with perturbed observations

Par	p_0^r	p^r	1%		2%		3%	
			$\check{p}^r$	$\widehat{NSE}$	$\check{p}^r$	$\widehat{NSE}$	$\check{p}^r$	$\widehat{NSE}$
γ_u	0.01	0.018	0.0182	0.4048	0.0183	0.8019	0.0185	1.1915
γ_H	0.01	0.007	0.0067	1.7824	0.0065	3.7091	0.0062	5.8016
λ_u	0.01	0.018	0.0185	0.3939	0.0191	0.7650	0.0196	1.1154
λ_A	0.01	0.007	0.0065	0.8573	0.0061	1.8460	0.0056	2.9990

The unknown parameters are accurately reconstructed, which is apparent from the negligible errors. The goodness-of-fit metrics are shown in Table 2, demonstrating an ideal fit.

Let us proceed with experiments with perturbed observations. In practice, every electronic device has its instrumental error thus testing with noisy measurements is meaningful. We add Gaussian noise to the observations (7). By $m\%$ noise we mean the relative error in the observation does not exceed $m\%$ with 95% confidence. We also define the relative parameter error (10) as follows.

$$RE_{\boldsymbol{p}} := \frac{\left\| \check{\boldsymbol{p}} - \check{\boldsymbol{p}}^{\text{pert}} \right\|_\infty}{\left\| \check{\boldsymbol{p}} \right\|_\infty}. \tag{10}$$

For $m = 1, 2, 3$, the results are presented in Table 3. It is obvious that the higher the noise, the higher the error in the recovered parameters.

What is more, the normalized standard errors also increase, but the inaccuracies in the recovered values are not that high.

The gof metrics are given in Table 4. The non-monotonicity of the first-order optimality measure $\left\| \nabla \Phi(\check{\boldsymbol{p}}) \right\|_\infty$ is because of the randomness of the noise. Moreover, the coefficients of determination R^2 are high enough. Nevertheless, the relative parameter errors are proportional to the level of noise.

Table 4 Goodness-of-fit metrics with perturbed measurements

Noise (%)	$\left\|\nabla\Phi(\check{\boldsymbol{p}})\right\|_{\infty}$	$\delta\boldsymbol{p}_k$	$\Phi(\check{\boldsymbol{p}})$	$RE_{\boldsymbol{p}}$	$\tilde{\sigma}^2$	$\hat{\sigma}$	R^2
1	15.9	3.92959e−14	2.32864e7	0.0304	6.2936e5	827.5836	0.9998
2	0.633	2.50024e−14	9.31455e7	0.0608	2.5174e6	1.6552e3	0.9994
3	1.38	4.25102e−14	2.09577e8	0.0913	5.6642e6	2.4827e3	0.9986

6 Conclusion

In the paper, a robust and effective algorithm for parameter reconstruction of honey bee population dynamics is suggested. The adopted model balances between simplicity and sophistication. It consists itself of a number of first-order ordinary differential equations, which describe the rates of change in honey bee compartments and food stores.

We recover the parameters related to the consumption rates of the brood and the bees. They are important in view of adequate beekeeping practice and management, but not directly observable in the nature. This means that knowing their fair values provides an advantage in taking appropriate precocious measures. The parameters are recovered via a square cost functional minimization, consisting of the squared difference between the theoretical and the experimental quantities. The solution to the direct problem is standard and is given in brief, and the solution to the inverse coefficient identification problem is thoroughly described and the computational algorithm is presented in detail. The numerical simulations with exact and perturbed observations support the theoretical findings. The results are acceptable even in the case of noisy measurements. Finally, it is shown that the algorithm is resource-efficient and fast to compute.

Among the ways this study could be continued, considering even more elaborate models, using the memory property of the system (through fractional-order derivatives), or involving time-varying parameters seem promising to explore. Moreover, a bifurcation analysis of the biologically meaningful equilibria points is useful from both theoretical and practical perspectives.

References

1. A.Z. Atanasov, S.G. Georgiev, L.G. Vulkov, Parameter identification analysis of food and population dynamics in honey bee colonies, to appear in Studies in Computational Intelligence, Springer, (2022)
2. A.Z. Atanasov, S.G. Georgiev, L.G. Vulkov, Parameter identification of colony collapse disorder in honeybees as a contagion, Comm. Comp. Inf. Sci., **1341**, Springer, 363–377 (2021)

3. A.Z. Atanasov, S.G. Georgiev, L.G. Vulkov, Reconstruction analysis of honeybee colony collapse disorder modeling, Optim. Eng., **22**, 2481–2503 (2021)
4. S. Bagheri, M. Mirzaie, A mathematical model of honey bee colony dynamics to predict the effect of pollen on colony failure, PLoS ONE, **14**(11), e0225632 (2019)
5. H.T. Banks, M. Davidian, J.R. Samuels, An inverse problem statistical methodology summary, Mathematical and statistical estimation approaches in epidemiology, Springer, 249–302 (2009)
6. H.T. Banks, S. Dediu, S. Ernstberger, Sensitivity functions and their uses in inverse problems, J. Inverse Ill-Posed Problems, **15**(7), 683–708 (2008)
7. T.F. Coleman, Y. Li, An interior, trust region approach for nonlinear minimization subject to bounds, SIAM J. Opt., **6**, 418–445 (1996)
8. T.F. Coleman, Y. Li, On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds, Math. Progr., **67**(2), 189–224 (1994)
9. J.R. Dormand, P.J. Prince, A family of embedded Runge-Kutta formulae, J. Comp. Appl. Math., **6**, 19–26 (1980)
10. S.G. Georgiev, L.G. Vulkov, Parameter identification approach for a fractional dynamics model of honeybee population, Lecture Notes in Computer Science, **13127**, Springer, 40–48 (2022)
11. P. Hristov, R. Shumkova, N. Palova, B. Neov, Factors associated with honeybee colony losses: a mini-review, Vet. Sci., **7**(166) (2020)
12. P. Hristov, R. Shumkova, N. Palova, B. Neov, Honey bee colony losses: why are honey bees disappearing, Sociobiology, **68**(1), e-5851 (2021)
13. D.S. Khoury, M.R. Myerscough, A.B. Barron, A quantitative model of honey bee colony population dynamics, PLoS ONE, **6**(4), e18491 (2011)
14. D.S. Khoury, A.B. Barron, M.R. Meyerscough, Modelling food and population dynamics honey bee colonies, PLoS ONE, **8**(5), e0059084 (2013)
15. C.M. Kribs-Zaleta, C. Mitchell, Modeling Colony Collapse Disorder in honeybees as a contagion, Math. Biol. Eng., **11**(6), 1275–1294 (2014)
16. M.R. Myerscough, D.S. Khoury, S. Ronzani, A.B. Barron, Why do hives die? Using mathematics to solve the problem of honey bee colony collapse, Mathematics for Industry 25, Springer, 35–50 (2017)
17. J.P.L.M. Paiva, H.M. Paiva, E. Esposito, M.M. Morais, On the effects of artificial feeding on bee colony dynamics: a mathematical model, PLoS ONE, **11**(11), e0167054 (2016)
18. S. Russel, A.B. Barron, D. Harris, Dynamics modelling of honeybee (Apis mellifera) colony growth and failure, Ecolog. Model., **265**, 138–169 (2013)
19. L.F. Shampine, M.W. Reichelt, The MATLAB ODE Suite, SIAM J. Sci. Comp., **18**, 1–22 (1997)
20. M.L. Winston, The Biology of the Honey Bee, Harvard University Press, (1991)

Chaos in a Dynamical Model of Competition Between Three Basic Power Stations Types

Elena V. Nikolova

Abstract In this study we propose a mathematical model based on a system of three non-linear ordinary differential equations to describe possible dynamical effects from competition between three basic power stations types depending on the primary energy source for electricity generation (nuclear energy, fossil fuels (non-renewable) energy and renewable sources). We choose the average price of electricity per kWh to be the common variable indicator for the proposed system. We consider a particular case of the reduced number of system parameters. For this case we analyze the dynamical properties of the equilibrium points of the three-dimensional system. On the basis of this analysis we found that for selected values of the parameters of the model of equations, chaos of Shilnikov kind in the considered system is possible. We illustrate numerically the development of a chaotic system attractor with increasing the value of an appropriate control parameter.

Keywords Dynamical model · Competition between three power stations · Shilnikov chaos

1 Introduction

The use of systems of differential equations is a widespread method for modeling dynamics of many natural processes from the past centuries to the present [1–18]. In the recent years, however, the focus of scientific interest has shifted to understanding the chaotic dynamics appeared in systems of three or more ordinary differential equations. A such dynamics includes sequences of bifurcations, periodic doubling

E. V. Nikolova (✉)
Institute of Mechanics, Bulgarian Academy of Sciences, Akad. G. Bonchev Str., Bl. 4, BG-1113 Sofia, Bulgaria
e-mail: elena@imbm.bas.bg

cascades and appearance of complex attractors, creating a 'butterfly effect' chaos (in the well-known Lorenz system [19]) or a spiral chaos (in the well-known Shilnikov system [20]). The most important tool for analyzing the appearance of chaotic motion of a Shilnikov kind in a 3D non-linear dynamical system is the well-known Shilnikov theorem. According to this theorem, chaotic behavior is exhibited in the neighborhoods of parameter space where certain homoclinic orbits appear, surrounding the saddle–focus equilibrium point [20]. The spiral chaos (or so-called also Shilnikov chaos) has been observed in many dynamical systems, starting from the Rössler system [21] and the Arneodo–Coullet–Tresser system [22]. Shilnikov chaos has been found also in various dynamical models in physics [23–27], in ecology and population dynamics [28–33], in economics [34, 35] and in medicine [36, 37].

In this paper we present a system of three non-linear ordinary differential equations to describe the competition between three different power stations depending on the primary energy source for electricity generation. The primary energy sources can be classified into three categories: nuclear energy, fossil fuels (non-renewable) sources, like coal, natural gas, and petroleum, and renewable energy sources, like wind, solar, geothermal, and hydro-power. These primary energy sources are converted to electricity in the corresponding power stations. On the basis of this classification, in our dynamical model, we assume that a nuclear power station (NPS), a fossil fuel power station (FFPS) (It can use coal or natural gas to produce electricity), and a hydro-electric power station (HPS) are competed with respect to the average price of the produced electricity per kWh. Next our goal is to demonstrate that chaos based on Shilnikov's theorem can exist in the considered three-dimensional system.

The paper is structured as follows. A short description of the proposed model is presented in Sect. 2 of the paper. The basic criteria for the appearance of chaos based on Shilnikov's theorem are formulated in the same section. In Sect. 3 conditions for existence of Shilnikov chaos in the three-dimensional model are analytically determined. 3D phase portraits, which illustrate the appearance and the evolution of a system chaotic attractor are presented in the Sect. 4. Conclusions based on the obtained results are made in Sect. 5.

2 Formulation of the Problem

We consider the following system of ordinary differential equations

$$\dot{x} = a_{10}x - a_{11}x^2 - a_{12}xy - a_{13}xz$$
$$\dot{y} = a_{20}y - a_{21}yx - a_{22}y^2 - a_{23}yz$$
$$\dot{z} = a_{30}z - a_{31}zx - a_{32}zy - a_{33}z^2, \tag{1}$$

assuming that x ($x > 0$) is a quantity accounting for the average price of electricity per kWh produced by a NPS; y ($y > 0$) is the quantity accounting for average price of electricity per kWh produced by a FFPS; and z ($z > 0$) is the quantity accounting

for average price of electricity per kWh produced by a HPS. All the coefficients of a system (1) are nonnegative. The time derivative of a quantity corresponds to the growth of the average price of electricity produced by the corresponding power station per unit time. In the model (1), the competition between power stations is taken into account through interactions between every two power stations. For an example, such competition can be expressed as follows: the increase of the average price of electricity produced by a FFPS and the increase of the average price of electricity produced by a HPS lead to a decrease of the growth of the average price of electricity produced by a NPS per unit time. This is accounted for by the last two terms in the first equation of (1). In the same way, the increase of the average price of electricity produced by a NPS and the increase of the average price of electricity produced by a HPS lead to a decrease of the growth of the average price of electricity produced by a FFPS per unit time. This is accounted for by the second term and the last term in the second equation of (1). In addition, the presence of the NPS and FFPS decreases the growth of the average price of electricity produced by a HPS per unit time. This is accounted for by the second term and the third term in the third equation of (1). Finally, in the absence of interactions (competition) between the power stations, the increase of the average price of electricity produced by each of the three power stations is assumed to follow a logistic law, where $a_{10}, a_{11}, a_{20}, a_{22}, a_{30}$ and a_{33} are the corresponding logistic growth coefficients.

Chaotic behavior in the above-formulated system can be realized if it satisfies the requirements of the theorem of Shilnikov [20].

Theorem 1 (Shilnikov) *If for the system*

$$\dot{x} = \rho x - \omega y + P(x, y, z)$$
$$\dot{y} = \omega x + \rho y + Q(x, y, z)$$
$$\dot{z} = \gamma z + R(x, y, z) \tag{2}$$

where $(P, Q, R$ are C^r functions $(1 < r < \infty)$ vanishing together with their first derivative at $O = (0, 0, 0))$ an unstable orbit Γ exists, which is a homoclinic connection, and if

$$\gamma > -\rho > 0 \tag{3}$$

then every neighborhood of the orbit Γ contains a denumerable set of unstable periodic solutions of saddle type.

In order to obtain chaos in the system (1) according to the Shilnikov theorem we have to satisfy the following Shilnikov conditions:

C 1*a*: There a saddle focus equilibrium point of the system (1) exists, where the characteristic eigenvalues connected to its linear stability are $\lambda_3 > 0$, and $\lambda_{1,2} = \rho \pm i\omega$ with negative real part ρ and imaginary ω;

C 1*b*: The condition $\lambda_3 > -\rho$ between the characteristic eigenvalues of the saddle focus equilibrium point must be satisfied.

C 2: There is a homoclinic orbit based at the saddle focus equilibrium point.

3 Appearance of Shilnikov Chaos in the System (1)

In order to obtain chaotic behavior for the system (1) according to the requirements of the Shilnikov theorem we analyze the properties of the equilibrium points of (1). In analogy with [29–33] we find that one of the possibilities for developing the Shilnikov chaos in the system (1) is reduced to the case in which the system parameters are described by the following relationships:

$$a_{11} = a_{13} = a_{21} = \kappa_1, \quad a_{12} = a_{22} = a_{23} = a_{31} = a_{32} = \kappa_2, \quad a_{33} = \kappa_3 \quad (4)$$

and

$$a_{10} = 2\kappa_1 + \kappa_2, \quad a_{20} = \kappa_1 + 2\kappa_2, \quad a_{30} = \kappa_3 + 2\kappa_2 \quad (5)$$

Thus, we reduce the number of system parameters to 3: κ_i ($i = 1, 2, 3$) and the system (1) becomes:

$$\begin{aligned}
\dot{x} &= (2\kappa_1 + \kappa_2)x - \kappa_1 x^2 - \kappa_2 xy - \kappa_1 xz \\
\dot{y} &= (\kappa_1 + 2\kappa_2)y - \kappa_1 yx - \kappa_2 y^2 - \kappa_2 yz \\
\dot{z} &= (\kappa_3 + 2\kappa_2)z - \kappa_2 zx - \kappa_2 zy - \kappa_3 z^2
\end{aligned} \quad (6)$$

Next we shall analyze the stability of the equilibrium points connected to the existence of Shilnikov chaos. We note again that the Shilnikov theorem [20] states that a system of ordinary differential equations exhibits Shilnikov chaos if there exists: a saddle focus equilibrium point of the system with $\lambda_3 > -\rho$, and a homoclinic orbit based at this equilibrium point. We shall focus only on the first Shilnikov condition (defined as *C* 1 in the previous section) for the system (6). The second condition (defined as *C* 2 in the previous section), connected to the existence of a homoclinic orbit, is hard to show analytically, but we shall visualize its appearance numerically in the next section of the paper.

The equilibrium points of (6) are

$$E_1 : \ x = y = z = 0, \ E_2 : \ x = y = 0, \ z = \frac{2\kappa_2 + \kappa_3}{\kappa_3}, \ E_3 : \ x = y = z = 1,$$

$$E_4 : \ x = \frac{2\kappa_1 + \kappa_2}{\kappa_1}, \ y = z = 0, \ E_5 : \ x = 0, \ y = \frac{2\kappa_2 + \kappa_1}{\kappa_2}, \ z = 0, \quad (7)$$

$$E_6 : \ x = 0, y = \frac{\kappa_2 \kappa_3 + \kappa_1 \kappa_3 - 2\kappa_2^2}{\kappa_2(\kappa_3 - \kappa_2)}, z = \frac{\kappa_3 - \kappa_1}{\kappa_3 - \kappa_2},$$

$$E_7: \quad x = \frac{\kappa_1\kappa_3 + \kappa_2\kappa_3 - 2\kappa_1\kappa_2}{\kappa_1(\kappa_3 - \kappa_2)}, \, y = 0, z = \frac{\kappa_2^2 - \kappa_1\kappa_3}{\kappa_1(\kappa_3 - \kappa_2)}$$

The equilibrium states (7) are realistic (nonnegative) when $\kappa_3 > \kappa_2 > \kappa_1$, $\kappa_2\kappa_3 + \kappa_1\kappa_3 > 2\kappa_2^2$, $\kappa_1\kappa_3 + \kappa_2\kappa_3 > 2\kappa_1\kappa_2$ and $\kappa_2^2 > \kappa_1\kappa_3$.

Now in order to obtain realistic equilibria of the system (6) for the considered problem, we shall focus only on the parameter region

$$\kappa_3 > \kappa_2 > \kappa_1, \tag{8}$$

and we shall determine the area of validity of the Shilnikov criteria C_1 in this region.

Proposition 1 *When $\frac{\kappa_2}{\kappa_1} > 1$ and $\frac{\kappa_3}{\kappa_1} > \frac{\kappa_2}{\kappa_1}$, the equilibrium point $E_6 = (0, \frac{\kappa_2\kappa_3 + \kappa_1\kappa_3 - 2\kappa_2^2}{\kappa_2(\kappa_3 - \kappa_2)}, \frac{\kappa_3 - \kappa_1}{\kappa_3 - \kappa_2})$ of the system (6) is a saddle focus with $\lambda_3 > -\rho$, where λ_3 is its real characteristic eigenvalue and ρ is the real part of its conjugated complex characteristic eigenvalues.*

Proof The linear stability of the equilibrium points of (6) is determined by the Jacobian matrix:

$$M_{ij} = \begin{pmatrix} M_{11} - \lambda & -\kappa_2 x & -\kappa_1 x \\ -\kappa_1 y & M_{22} - \lambda & -\kappa_2 y \\ -\kappa_2 z & -\kappa_2 z & M_{33} - \lambda \end{pmatrix} \tag{9}$$

where

$$\begin{aligned} M_{11} &= 2\kappa_1 + \kappa_2 - 2\kappa_1 x - \kappa_2 y - \kappa_1 z \\ M_{22} &= \kappa_1 + 2\kappa_2 - \kappa_1 x - 2\kappa_2 y - \kappa_2 z \\ M_{33} &= 2\kappa_2 + \kappa_3 - \kappa_2 x - \kappa_2 y - 2\kappa_3 z \end{aligned} \tag{10}$$

According to the requirements of the theorem of Shilnikov, chaotic motion for the system (6) will be observed if there an equilibrium point of a saddle focus type exists and characteristic eigenvalues connected to linear stability of this equilibrium point have to satisfy the condition (3) in some region of the space of the system parameters.

We denote $\alpha = \frac{\kappa_2}{\kappa_1}$ and $\beta = \frac{\kappa_3}{\kappa_1}$, assuming $\alpha > 1$ and $\beta > 1$ to satisfy the condition (8) for the system parameter region. Then, the equilibrium point E_6 from (7) can be rewritten as $E_6 = \left(0, \frac{\alpha\beta + \beta - 2\alpha^2}{\alpha(\beta - \alpha)}, \frac{\beta - 1}{\beta - \alpha}\right)$. The Jacobian matrix at this equilibrium point is

$$M'_{ij} = \begin{pmatrix} M'_{11} - \lambda & -\alpha\kappa_1 x & -\kappa_1 x \\ -\kappa_1 y & M'_{22} - \lambda & -\alpha\kappa_1 y \\ -\alpha\kappa_1 z & -\alpha\kappa_1 z & M'_{33} - \lambda \end{pmatrix} \tag{11}$$

where

$$M'_{11} = \kappa_1(2 + \alpha - 2x - \alpha y - z)$$
$$M'_{22} = \kappa_1(2\alpha + 1 - x - 2\alpha y - \alpha z) \tag{12}$$
$$M'_{33} = \kappa_1(2\alpha + \beta - \alpha x - \alpha y - 2\beta z)$$

According to Eq. (11), the characteristic eigenvalues connected to linear stability of this equilibrium point are:

$$\lambda_{1,2} = \kappa_1\left(\frac{-2\alpha^2 + \alpha\beta + \beta^2}{2(\alpha - \beta)} \pm \frac{\sqrt{A}}{2(\alpha - \beta)}\right); \ \lambda_3 = -\frac{\kappa_1(\alpha^2 - 2\alpha + 1)}{\alpha - \beta} \tag{13}$$

where

$$A = 4\alpha^4 - 12\alpha^3\beta + 9\alpha^2\beta^2 - 2\alpha\beta^3 + \beta^4 + 8\alpha^3 - 12\alpha^2\beta + 8\alpha\beta^2 - 4\alpha\beta - 4\beta^3 + 4\beta^2 \tag{14}$$

Thus, when the system of inequalities:

$$\frac{A}{2(\alpha - \beta)} < 0, \ \frac{-2\alpha^2 + \alpha\beta + \beta^2}{2(\alpha - \beta)} < 0, \ -\frac{\alpha^2 - 2\alpha + 1}{\alpha - \beta} > 0 \tag{15}$$

and

$$-\frac{\alpha^2 - 2\alpha + 1}{\alpha - \beta} > -\frac{-2\alpha^2 + \alpha\beta + \beta^2}{2(\alpha - \beta)} \tag{16}$$

hold true, the equilibrium point E_6 of the system (6) is of a saddle focus type, and the condition (3) for the characteristic eigenvalues connected to its stability, is satisfied too. There are several solutions of the system of inequalities (14)–(15), but only one solution satisfies the condition ((8)) for the system parameter region. This solution is

$$\alpha > 1, \ \beta > \alpha, \tag{17}$$

Thereby the Shilnikov condition $C\ 1$ for the system (6) is satisfied in the parameter region (16), i.e., the above-formulated Proposition follows.

In addition, the system (6) shall turn to a chaotic regime if there is another equilibrium point of Eq. (6), which must undergo a supercritical Hopf bifurcation in the same region of parameters (see Eq. (16)). To obtain the Hopf bifurcation in our three-dimensional system we use the center-manifold theorem. In accordance with this theorem we must reduce the considered system to a normal form in which the bifurcation occurs when the system parameters approach 0 and the equilibrium point is located in the origin with pure imaginary characteristic eigenvalues. In our case this equilibrium point is $E_3 = (1, 1, 1)$ and its characteristic eigenvalues presented in terms of α and β are

$$\lambda_{1,2} = \left(-\frac{1}{3}(\alpha + \beta) - \frac{5}{12} + \frac{\beta - 5\alpha + \alpha\beta - 4\alpha^2 - \beta^2 - 1}{3B}\right.$$
$$\left.\pm \frac{\sqrt{3}}{2}i\left(\frac{1}{6} + \frac{2\alpha\beta - 10\alpha + 2\beta - 8\alpha^2 - 2\beta^2 - 2}{3B}\right)\right)\kappa_1 \tag{18}$$

$$\lambda_3 = \left(-\frac{1}{3}(\alpha + \beta) - \frac{1}{6} - \frac{\beta - 5\alpha + \alpha\beta - 4\alpha^2 - \beta^2 - 1}{3B} \right)\kappa_1$$

where

$$B = \left(-168\alpha + 120\alpha^2 - 12\alpha\beta + 12\beta + 12\beta^2 + 12\alpha\beta^2 - 152\alpha^3 - 7 - 24\alpha^2\beta \right.$$

$$-8\beta^3 + 1 + 2(6\beta^3 - 372\alpha^3 + 108\alpha^3\beta - 75\alpha^2\beta^2 + 36\alpha\beta^3 + 300\alpha^4 - 3\beta^4$$

$$-48\alpha^4\beta^2 - 72\alpha^3\beta^2 + 6\alpha^2\beta^3 + 60\alpha^4\beta - 360\alpha^5 + 132\alpha^6 - 6\alpha\beta^4 + 24\alpha^3\beta^3 \qquad (19)$$

$$\left. -3\alpha^2\beta^4 + 72\alpha^5\beta + 12\alpha - 12\alpha\beta + 144\alpha^2 + 108\alpha^2\beta - 54\alpha\beta^2 - 3\beta^2)^{1/2} \right)^{1/3}$$

We note that

$$-\frac{1}{3}(\alpha + \beta) - \frac{5}{12} + \frac{\beta - 5\alpha + \alpha\beta - 4\alpha^2 - \beta^2 - 1}{3B} \rightarrow 0 \qquad (20)$$

in the parameter region (16) as well as the inequality

$$\frac{1}{6} + \frac{2\alpha\beta - 10\alpha + 2\beta - 8\alpha^2 - 2\beta^2 - 2}{3B} > 0 \qquad (21)$$

is satisfied in the same parametric region.

4 Numerical Results

We illustrate the process of appearance of Shilnikov chaos in Fig. 1. We use κ_2 as a control parameter for obtaining the attracting manifold.

As Fig. 1 shows, initially, the system (6) has a stable equilibrium state (E_3) for $\kappa_2 < 0.9$ (Fig.1a). Whit increasing κ_2, the equilibrium point becomes unstable and undergoes a supercritical Hopf bifurcation at $\kappa_2 = 0.9$ as a stable periodic orbit bifurcates from it (Fig. 1b). Next we illustrate the evolution of this limit cycle with increasing κ_2. We observe that when $\kappa_2 > 0.9$ initially the trajectory spirals onto the orbit decreasing the size of the limit cycle (Fig. 1c) whereas at $\kappa_2 \geq 1$ this orbit becomes a boundary of the unstable manifold of the saddle focus (E_6) that spirals onto it (Fig. 1d). Finally, when $\kappa_2 \geq 1.1$ period double cascade is observed, and the mechanism of Shilnikov holds: The unstable 1D manifold touch the stable 2D manifold thereby forming the aforementioned homoclinic loop of the saddle focus (Fig. 1e). After that numerous almost periodic orbits around the homoclinic loop occur, as the attracting whirlpool already contains a set of complex structures, i.e., the so-called spiral strange attractor has appeared (Fig. 1f).
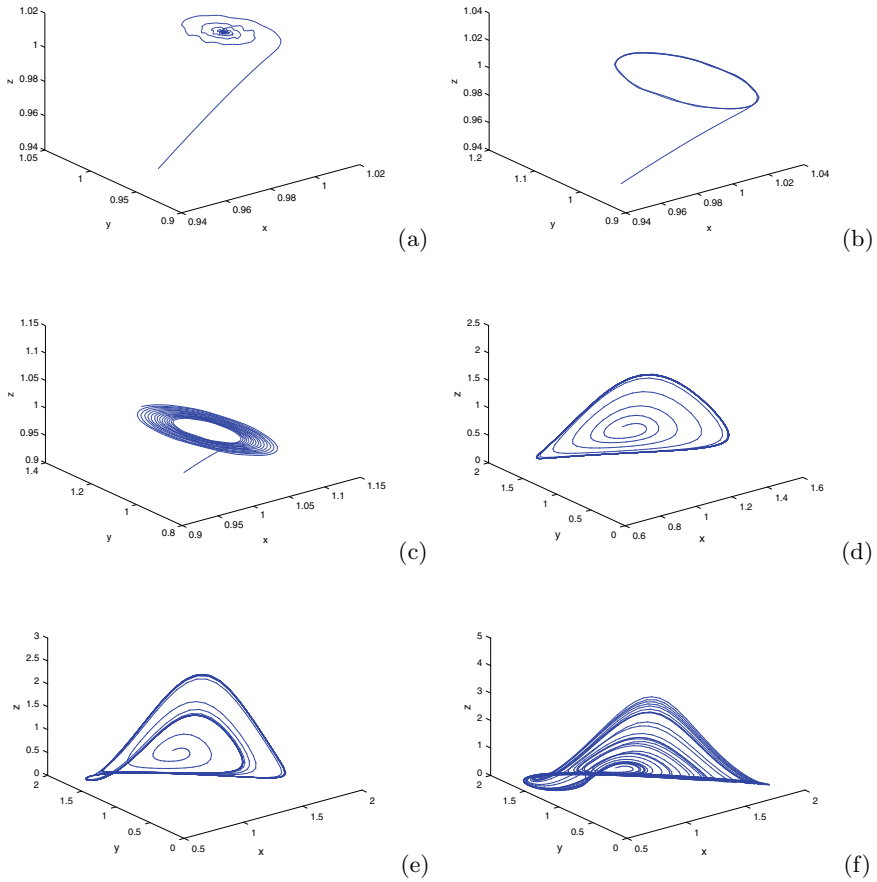
Fig. 1 Transition to Shilnikov chaos in the system (6). The values of system parameters are $\kappa_1 = 0.05$, $\kappa_3 = 1.3$. We vary the value of κ_2. Panel **a**: $\kappa_2 = 0.8$. The equilibrium point E_3 is stable. Panel **b** $\kappa_2 = 0.9$. The cyclic state after the Hopf bifurcation is purely presented. Panel **c** $\kappa_2 = 0.95$. The trajectory spirals onto the limit cycle. Panel **d** $\kappa_2 = 1$. The saddle focus E_6 already has appeared and the unstable manifold of the saddle focus is clearly visible. Panel **e** $\kappa_2 = 1.1$. Period double cascade is observed. A homoclinic loop appears. Panel **f** $\kappa_1 = 1.2$. Chaotic motion already exists. With further increasing κ_1 the attractor again is reduced to a periodic cycle

5 Conclusion

In this paper we show that when competition between three main electricity producers exists with respect to the average price of electricity, the behavior of the corresponding dynamical model becomes complicated. For example, the equilibrium point E_3 describes the case at which the average prices of electricity produced by the three power stations are the same. On the other hand, the saddle focus E_6 corresponds

to situation in which the average price of electricity produced by the NPS is 0 and the average prices of electricity produced by the FFPS and by the HPS are larger than those at the point E_3. In other words, the Shilnikov chaos corresponds to a situation in which the competition between three electricity producers with respect to the average price of produced electricity leads to the attenuation of one of them, whereas the other two power producers oscillate in irregular ways. In this case, the large changes in the average price of electricity produced by NPS are observed, as this price can even become minimal. We note that the FFPS or the HPS (taken alone) can also demonstrate such attenuation in its average price of produced electricity depending on the choice in the arrangement of the variables in the system of equations. In addition, we identify numerically the region of parameter space where the chaotic motion of Shilnikov kind in the studied dynamical system exists.

References

1. Axelrod, R., Cohen, M.: Harnessing Complexity. Basic Books, New York (2001)
2. Vitanov, N. K., Dimitrova, Z. I., Ausloos, M.: Verhulst - Lotka - Volterra (VLV) model of ideological struggle. Physica A 389, 4970–4980 (2010). https://doi.org/10.1016/j.physa.2010.06.032
3. Flaut, C., Hoskova-Mayerova, S., Flaut D. (Eds.): Models and Theories in Social Systems. Springer, Cham, (2019).
4. Vitanov, N. K., Sakai, K., Dimitrova, Z.: SSA, PCA, TDPSC, ACFA: Useful combination of methods for analysis of short and nonstationary time series. Chaos Solitons and Fractals 37, 187–202 (2008). https://doi.org/10.1016/j.chaos.2006.08.043
5. Forester, J. W.: Nonlinearity in high-order models of social systems. European Journal of Operational Research 30, 104–109 (1987). https://doi.org/10.1016/0377-2217(87)90086-5
6. Jordanov, I., Nikolova, E.: On nonlinear waves in the spatio–temporal dynamics of interacting populations. Journal of Theoretical and Applied Mechanics, 43, No. 2, 69–76 (2013).
7. Vitanov, N. K., Sakai, K., Jordanov, I. P., Managi, S., Demura, K.: Analysis of a Japan government intervention on the domestic agriculture market. Physica A 382, 330–335 (2006). https://doi.org/10.1016/j.physa.2007.02.025
8. Vitanov, N. K., Ausloos, M.: Knowledge epidemics and population dynamics models for describing idea diffusion, in *Models of Science Dynamics*, edited by A. Scharnhorst, K Boerner, P. van den Besselaar, Springer, Berin, 65–129 (2012).
9. Dimitrova, Z. I., Ausloos, M.: Primacy analysis in the system of Bulgarian cities. Open Physics 13, 218–225 (2015). https://doi.org/10.1515/phys-2015-0029
10. Vitanov, N. K., Dimitrova, Z. I., Vitanov, K. N.: Modified method of simplest equation for obtaining exact analytical solutions of nonlinear partial differential equations: further development of the methodology with applications. Applied Mathematics and Computation 269, 363–378 (2015). https://doi.org/10.1016/j.amc.2015.07.060
11. Vitanov, N. K., Jordanov, I. P., Dimitrova, Z. I.: On nonlinear dynamics of interacting populations: Coupled kink waves in a system of two populations. Commun. Nonl. Sci. Numerical Simulat. 14, 2379–2388 (2009). https://doi.org/10.1016/j.cnsns.2008.07.015

12. Vitanov, N. K., Jordanov, I. P., Dimitrova, Z. I.: On nonlinear population waves. Applied Mathematics and Computation 215, 2950–2964 (2009). https://doi.org/10.1016/j.amc.2009.09.041

13. Vitanov, N. K., Dimitrova, Z. I., Kantz, H.: Modified method of simplest equation and its application to nonlinear PDEs. Applied Mathematics and Computation 216, 2587–2595 (2010). https://doi.org/10.1016/j.amc.2010.03.102

14. Vitanov, N. K., Ausloos, M., Rotundo, G.: Discrete Model of Ideological Struggle Accounting for Migration. Advances in Complex Systems 15, Supplement 1, 125049 (2012). https://doi.org/10.1142/S021952591250049X

15. Lave, C. L., March, J. G.: An Introduction to Models in the Social Sciences. University Press of America, Lahnalam, ML (1993).

16. Sakai, K., Managi, S., Demura, K., Vitanov, N. K.: Transition of Chaotic Motion to a Limit Cycle by Intervention of Economic Policy: An Empirical Analysis in Agriculture. Nonlinear Dynamics, Psyhology and Life Sciences 11, 253–265 (2007).

17. Forester, J. W.: Counterintuitive Behavior of Social Systems. Technology Review 73, 52–68 (1971). https://doi.org/10.1007/BF00148991

18. Vitanov, N. K., Ausloos, M.: Test of two hypotheses explaining the size of populations in a system of cities. Journal of Applied Statistics 42, 2686–2693 (2015). https://doi.org/10.1080/02664763.2015.1047744

19. Lorenz, E. N.: The statistical prediction of solutions of dynamical equations. Proceedings of the International Symposium on Numerical Weather Prediction (Tokyo, Japan, November 1960), 629–635 (1962).

20. Shilnikov, L. P.: A case of the existence of a denumerable set of periodic motions. Sov. Math. Dokl. 6, 163–166 (1965).

21. Rössler, O. E.: An equation for continuous chaos. Chaos, Phys. Lett. A 57 (5), 397–398 (1976). https://doi.org/10.1016/0375-9601(76)90101-8

22. Arnédo, A., Coullet, P., Tresser, C.: Oscillators with chaotic behavior: An illustration of a theorem by Shil'nikov. J. Statist. Phys. 27 (1), 171–182 (1982). https://doi.org/10.1007/BF01011745

23. Arecchi, F., Meucci, R., Gadomski, W.: Laser Dynamics with Competing Instabilities. Physical review letters 58 (21), 2205 (1987). https://doi.org/10.1103/PhysRevLett.58.2205

24. de Tomasi, F., Hennequin, D., Zambon, B., Arimondo, E.: Instabilities and chaos in an infrared laser with saturable absorber: experiments and vibrorotational model. JOSA B 6 (1), 45–57 (1989). https://doi.org/10.1364/JOSAB.6.000045

25. Braun, T., Lisboa, J. A., Gallas, J. A. C.: Evidence of homoclinic chaos in the plasma of a glow discharge. Physical review letters 68 (18), 27–70 (1992). https://doi.org/10.1103/PhysRevLett.68.2770

26. Pisarchik, A., Meucci, R., Arecchi, F.: Theoretical and experimental study of discrete behavior of Shilnikov chaos in a CO 2 laser. The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics 13 (3), 385–391 (2001). https://doi.org/10.1007/s100530170257

27. Viktorov, E. A., Klemer, D. R., Karim, M. A.: Shil'nikov case of antiphase dynamics in a multimode laser. Optics communications 113 (4-6), 441–448 (1995). https://doi.org/10.1016/0030-4018(94)00533-Z

28. Deng, B., Hines, G.: Food chain chaos due to Shilnikov's orbit. Chaos 12 (3), 533–538 (2002). https://doi.org/10.1063/1.1482255

29. Dimitrova, Z. I., Vitanov, N. K.: Influence of adaptation on the nonlinear dynamics of a system of competing populations. Physics Letters A 272, 368–380 (2000). https://doi.org/10.1016/S0375-9601(00)00455-2

30. Dimitrova, Z. I., Vitanov, N. K.: Adaptation and its impact on the dynamics of a system of three competing populations. Physica A 300, 91–115 (2001). https://doi.org/10.1016/S0378-4371(01)00330-2

31. Dimitrova, Z. I., Vitanov, N. K.: Chaotic pairwise competition. Theoretical Population Biology 66, 1–12 (2004). https://doi.org/10.1016/j.tpb.2003.10.008

32. Dimitrova, Z. I., Vitanov, N. K.: Dynamical consequences of adaptation of the growth rates in a system of three competing populations. J. Phys. A: Math. Gen. 34, 7459–7473 (2001). https://doi.org/10.1088/0305-4470/34/37/303/meta

33. Dimitrova, Z. I., Vitanov, N. K.: Shilnikov chaos in a generalized system for modelling dynamics of competing populations. Comptes redus de l'Academie bulgares des Sciences 58, 257–264 (2005).

34. Nikolova, E. V.: Chaos in a generalized system for modeling dynamics of three interacting economic sectors. AIP Conference Proceedings 2321 (1), 030025 (2021). https://doi.org/10.1063/5.0040086

35. Nikolova, E. V., Vitanov, N. K.: On the Possibility of Chaos in a Generalized Model of Three Interacting Sectors. Entropy 22 (12), 1388 (2020). https://doi.org/10.3390/e22121388

36. Seifert, B., Adamski, D., Uhl, C.: Analytic Quantification of Shilnikov Chaos in Epileptic EEG Data. Frontiers Appl. Math. Stat. 4, 57 (2018). https://doi.org/10.3389/fams.2018.00057

37. Itik, M., Banks, S. P.: Chaos in a Three-Dimensional Cancer Model. International Journal of Bifurcation and Chaos 20 (1), 71–79 (2010). https://doi.org/10.1142/S0218127410025417

Mathematical Analysis of Hepatitis B Virus Combination Treatment

Irina Volinsky

Abstract In the current paper, we research the influence of IL-2 therapy and we introduce the regulation by distributed feedback control. The results of the stability analysis are presented. We use the Cauchy matrix $C(t, s)$, in order to study the behavior (stability) of corresponding system of integro-differential equations.

Keywords Functional differential equations · Exponential stability · Cauchy matrix · Integro-differential systems · Hepatitis B · Immune system

1 Introduction

The hepatitis B virus (HBV) represents a huge problem for public health, increasing the risk of cirrhosis and carcinoma in the population [1].

It needs to integrate its DNA with the host cell DNA to survive. Chronic HBV infection accounts for at least 50% of carcinoma cases [2], and have a high mortality rate.

Current therapy for chronic HBV infection is based on two strategies: interferons and nucleoside analog [3].

Five nucleoside analogs are approved in the United States: lamivudine, adefovir, entecavir, tenofovir disoproxil, and tenofovir alafenamide [1]. Their role is to block virus replication .

With all of these therapeutic options, therapeutic goals are rarely achieved [4–7]. Possible drug improvement could be obtained with the integration of interleukin-2 (IL-2) and new clinical trials are evaluating its impact on the combination therapy (NCT02360592, NCT00451984) help for viral clearing [8].

I. Volinsky (✉)
Ariel University, Ariel, Israel
e-mail: irinav@ariel.ac.il

Mathematical models are useful tool for representing molecular mechanisms and try to predict new biological insights [17–23].

Previous models have tried to predict the effectiveness of antiviral therapy, but without considering the important role of the immune system [9–11]. This approximation does not allow for realistic predictions.

A complete model was presented in [12], where immune system and drug response were both taken into account. In this study, we propose a mathematical model for HBV infection, using a system of functional differential equations. This approach has already proven its efficacy in biology and medicine [13, 14]. We have taken into account the role of the immune system and the two different treatments (interferon and nucleoside analogs), and we have included a control function to optimize the role of a possible IL-2 co-treatment. In [15], a global analysis of a hepatitis C virus (HCV) model is presented with CTL and antibody responses. The model consisted of five nonlinear equations as follows:

$$\begin{cases} X' = r - dX - \beta V X \\ Y' = \beta V X - aY - pYZ \\ V' = kY - uV - qVW \\ W' = -hW + gVW \\ Z' = cYZ - bZ \end{cases} \tag{1}$$

Variables are uninfected cells, infected cells, free virus numbers, antibody response, and CTL response. Those are denoted, respectively, by $X(cells/mL)$, $Y(cells/mL)$, $V(IU/mL)$, $W(IU/mL)$, and $Z(cells/mL)$. X cells are produced at a rate r, die at a rate dX, and are infected by virus at a rate $\beta V X$. The infected cells, Y, grow at the rate of $\beta V X$, die (naturally) at a rate aY, and are killed by the CTLs at a rate pYZ. Free viruses V are produced by infected cells Y at a rate kY, decay at a rate uV, and are neutralized by antibodies at a rate qVW. Antibodies W develop in response to free viruses at a rate gvW and decay at a rate hW. The number of CTLs, Z, expands in response to infected cells at a rate cYZ and decays in the absence of infection at a rate bZ.

The dynamics of both HBV and HCV are similar. Hence, we can use a similar model to describe each one of them, using the five variables described above.

2 Modified Model of HBV

Let us now discuss the novelty of our approach. In this paper, we consider the efficiency of the interferon therapy and the efficiency of therapy with nucleoside analogs. These are, respectively, denoted by η and ε, and their values are between 0 and 1. Due to their influence on the therapy, they directly affect the growth or the decay of the infected cells.

Additionally, we wish to introduce the distributed control function $U(t)$, which represents the impact of the IL-2 therapy, as the following integral form:

$$U(t) = \int\limits_{0}^{t} e^{-\alpha(t-s)} Z(s) ds. \tag{2}$$

Since the role of IL-2 is still not very well characterized and the literature is controversial, we have decided to consider that it is produced mainly by the CTLs, and we applied it to the Z equations due to its role in increasing CTL response and T lymphocyte duplication and activation. We wish to study and predict the patient's disease and advancement, and will allow us to properly adjust the therapeutic process to each subject.

The goal of this paper is to demonstrate a new approach of co-treatment for HBV using a distributed control in the model of the HBV disease. A model of infectious diseases is proposed based on the analysis of the integro-differential systems studied in [19].

Modified model is as follows:

$$\begin{cases} X' = r - dX - (1-\eta)\beta VX \\ Y' = (1-\eta)\beta VX - aY - pYZ \\ V' = (1-\varepsilon)kY - uV - qVW \\ W' = -hW + gVW \\ Z' = cYZ - bZ + DU \end{cases} \tag{3}$$

Using the reduction method we pass from a integro-differential system (3) to an ordinary differential system:

$$\begin{cases} X' = r - dX - (1-\eta)\beta VX \\ Y' = (1-\eta)\beta VX - aY - pYZ \\ V' = (1-\varepsilon)kY - uV - qVW \\ W' = -hW + gVW \\ Z' = cYZ - bZ + DU \\ U' = Z - \alpha U \end{cases} \tag{4}$$

3 Main Results

We wish to study the disease-free equilibrium, since it represents a patient who is either a healthy subject or a recovered one. From this stationary state, it is possible to deduce the others.

$$P = \{X, Y, V, W, Z, U\} = \left(\frac{r}{d}, 0, 0, 0, 0, 0\right). \tag{5}$$

Linearizing systems (1) and (4) in the neighborhood of the disease-free equilibrium, we obtain

$$\begin{cases} x_1' = -dx_1 - \frac{\beta r}{d} x_3 \\ x_2' = -ax_2 + \frac{\beta r}{d} x_3 \\ x_3' = kx_2 - ux_3 \\ x_4' = -hx_4 \\ x_5' = -bx_5 \end{cases} \tag{6}$$

$$\begin{cases} x_1' = -dx_1 - \frac{(1-\eta)\beta r}{d} x_3 \\ x_2' = -ax_2 + \frac{(1-\eta)\beta r}{d} x_3 \\ x_3' = (1-\varepsilon)kx_2 - ux_3 \\ x_4' = -hx_4 \\ x_5' = Dx_6 - bx_5 \\ x_6' = x_5 - \alpha x_6 \end{cases}, \tag{7}$$

where
$$x_1 = X - \frac{r}{d}, x_2 = Y, x_3 = V, x_4 = W, x_5 = Z, x_6 = U.$$

Denote the corresponding matrices of coefficients of system (6) and (7):

$$B = \begin{pmatrix} -d & 0 & -\frac{\beta r}{d} & 0 & 0 \\ 0 & -a & \frac{\beta r}{d} & 0 & 0 \\ 0 & k & -u & 0 & 0 \\ 0 & 0 & 0 & -h & 0 \\ 0 & 0 & 0 & 0 & -b \end{pmatrix}. \tag{8}$$

$$A = \begin{pmatrix} -d & 0 & -\frac{(1-\eta)\beta r}{d} & 0 & 0 & 0 \\ 0 & -a & \frac{(1-\eta)\beta r}{d} & 0 & 0 & 0 \\ 0 & (1-\varepsilon)k & -u & 0 & 0 & 0 \\ 0 & 0 & 0 & -h & 0 & 0 \\ 0 & 0 & 0 & 0 & -b & D \\ 0 & 0 & 0 & 0 & 1 & -\alpha \end{pmatrix} .. \tag{9}$$

The characteristic polynomial of system (6) has five roots:

$$\begin{cases} \lambda_1 = -h \\ \lambda_2 = -d \\ \lambda_3 = \frac{-d(a+u)+\sqrt{d^2(a-u)^2+4k\beta rd}}{2d} \\ \lambda_4 = \frac{-d(a+u)-\sqrt{d^2(a-u)^2+4k\beta rd}}{2d} \\ \lambda_5 = -b. \end{cases} \tag{10}$$

The characteristic polynomial of system (7) has six roots:

$$\begin{cases} \lambda_1^* = -h \\ \lambda_2^* = -d \\ \lambda_3^* = \frac{-d(a+u)+\sqrt{d^2(a-u)^2+4k\beta rd(\varepsilon-1)(\eta-1)}}{2d} \\ \lambda_4^* = \frac{-d(a+u)-\sqrt{d^2(a-u)^2+4k\beta rd(\varepsilon-1)(\eta-1)}}{2d} \\ \lambda_5^* = -(\frac{\alpha+b}{2}) + \frac{\sqrt{(\alpha-b)^2+4D}}{2} \\ \lambda_6^* = -(\frac{\alpha+b}{2}) - \frac{\sqrt{(\alpha-b)^2+4D}}{2}. \end{cases} \tag{11}$$

Theorem 1 *If all of the coefficients of system (1) are positive, then system (1) is exponentially stable in the neighborhood of disease-free equilibrium.*

Let us denote the corresponding spectral radius of systems (6) and (7) by

$$\rho = max_{1 \le i \le 5} |\lambda_i|, \quad \rho^* = max_{1 \le j \le 6} |\lambda_i^*|. \tag{12}$$

Theorem 2 *[25] If all of the coefficients of system (4) are positive, η and ε are parameters defined between 0 and 1, and inequalities*

$$(\varepsilon - 1)(\eta - 1) < \frac{aud}{\beta kr}, \quad D < \alpha b$$

are fulfilled, then system (4) is exponentially stable in the neighborhood of disease-free equilibrium and $\rho \le \rho^*$.

Consider the following system with an uncertain coefficient:

$$\begin{cases} X' = r - dX - (1-\eta)\beta \, V X \\ Y' = (1-\eta)\beta V X - aY - pYZ \\ V' = (1-\varepsilon)kY - uV - qVW \\ W' = -hW + gVW \\ Z' = cYZ - bZ + (D + \Delta D(t))U. \end{cases} \tag{13}$$

The uncertain coefficient $\Delta D(t)$ can be the result of individual conditions of the patient, because the assimilation of a drug in the body of different patients can have different rates of influence. We assumed that $\Delta D(t)$ is essentially a bounded function.

As before, this system of integro-differential equations can be reduced to the following system of homogeneous ordinary differential equations:

$$\begin{cases} X' = -dX - (1-\eta)\beta V X \\ Y' = (1-\eta)\beta V X - aY - pYZ \\ V' = (1-\varepsilon)kY - uV - qVW \\ W' = -hW + gVW \\ Z' = cYZ - bZ + (D + \Delta D(t))U \\ U' = Z - \alpha U. \end{cases} \tag{14}$$

Let us denote (see Appendix A):

$$\Delta D^* = {}_{t \geq 0}|\Delta D(t)|, \qquad Q_1^* = Q_2^* = Q_3^* = Q_4^* = 0,$$

$$Q_5^* = \Delta D^* \left[\left| \frac{c_{28}}{c_{13} - c_{14}} \right| + \left| \frac{c_{28}}{c_{13} + c_{14}} \right| \right], \quad Q_6^* = \Delta D^* \left[\left| \frac{c_{31}}{c_{13} - c_{14}} \right| + \left| \frac{c_{32}}{c_{13} + c_{14}} \right| \right].$$

Theorem 3 *[24] Let the assumption of Theorem 2 be fulfilled, $c_{13} \neq c_{14}$ and the inequality $max_{1 \leq i \leq 6}\{Q_i^*\} < 1$ be true, then system (14) is exponential stable in the neighborhood of disease-free equilibrium.*

Remark 1 From [16], the following values of the parameters in system (4) are used:

$$d = 0.00333, \ g = q = 5, \ b = 0.112, \ \beta = 7, \ h = 2, \ a = 0.56, \ u = 0.67,$$

$$p = c = 5.14, k = 20, r = 6.17 * 10^{-4}.$$

Example 1 Using the coefficient values from Remark 1, and taking the following values of parameters $\alpha = 0.5, \quad D = 0.04$, such that condition $D < \alpha b$ is fulfilled. From Theorem 3,

$$Q_1^* = Q_2^* = Q_3^* = Q_4^* = 0, \quad Q_5^* \approx 69 \Delta D^*, \quad Q_6^* \approx 7 \Delta D^*.$$

We obtain the following condition for uncertain coefficient: $\Delta D^* < 0.145$.

Consider system (3), with unbounded memory control function

$$U(t) = \int\limits_{t-\tau(t)}^{t} e^{-\alpha(t-s)} Z(s) ds. \tag{15}$$

Denote

$$\tilde{U}(t) = \int\limits_{0}^{t} e^{-\alpha(t-s)} Z(s) ds. \tag{16}$$

Let us write (15) in the following form:

$$U(t) = \int\limits_{0}^{t} e^{-\alpha(t-s)} Z(s) ds - e^{-\alpha \tau(t)} \int\limits_{0}^{t-\tau(t)} e^{-\alpha((t-\tau(t))-s)} Z(s) ds$$

$$= \tilde{U}(t) - e^{-\alpha \tau(t)} \tilde{U}(t - \tau(t)). \tag{17}$$

Reducing the integro-differential system (3), where $U(t)$ is defined by (15) to an ordinary differential system, we obtain the following:

$$\begin{cases} X' = r - dX - (1 - \eta)\beta V X \\ Y' = (1 - \eta)\beta V X - aY - pYZ \\ V' = (1 - \varepsilon)kY - uV - qVW \\ W' = -hW + gVW \\ Z' = cYZ - bZ + D\tilde{U} - De^{-\alpha\tau(t)}\tilde{U}(t - \tau(t)) \\ \tilde{U}' = Z - \alpha\tilde{U}. \end{cases} \quad (18)$$

Linearizing system (18) in the neighborhood of the disease-free equilibrium, we obtain the corresponding homogeneous linear systems:

$$\begin{cases} x_1'(t) = r - dx_1(t) - \frac{(1-\eta)\beta r}{d} x_3(t) \\ x_2'(t) = -ax_2(t) + \frac{(1-\eta)\beta r}{d} x_3(t) \\ x_3'(t) = (1 - \varepsilon)kx_2(t) - ux_3(t) \\ x_4'(t) = -hx_4(t) \\ x_5'(t) = Dx_6(t) - bx_5(t) - De^{-\alpha\tau(t)}x_6(t - \tau(t)) \\ x_6'(t) = x_5(t) - \alpha x_6(t), \end{cases} \quad (19)$$

where

$$x_1(t) = X - \frac{r}{d}, x_2(t) = Y, x_3(t) = V, x_4(t) = W, x_5(t) = Z, x_6(t) = \tilde{U}.$$

Let us denote $\tau_* = ess \inf_{t \geq 0} |\tau(t)|$ and

$$Q_1 = De^{\alpha\tau_*}\left(\left|\frac{c_{28}}{c_{14} - c_{13}}\right| + \left|\frac{c_{28}}{c_{13} + c_{14}}\right| + \left|\frac{c_{31}}{c_{14} - c_{13}}\right| + \left|\frac{c_{32}}{c_{13} + c_{14}}\right|\right). \quad (20)$$

Theorem 4 [25] *Let the assumption of Theorem 2 be fulfilled, $c_{13} \neq c_{14}$ and the inequality $Q_1 < 1$ be true, then system (18) is exponentially stable in the neighborhood of disease-free equilibrium.*

Example 2 Using the coefficient values from Remark 1, and taking the following values of parameters $\alpha = 0.5$, $D = 0.01$, such that condition $D < \alpha b$ is fulfilled. From Theorem 4, $Q_1 \approx 0.33e^{0.5\tau_*}$. We obtain the following condition for unbounded memory control function limit: $\tau_* < 2.2$.

Consider system (3), with delay in upper limit of control function

$$U(t) = \int_0^{t-\tau(t)} e^{-\alpha(t-s)} Z(s) ds. \quad (21)$$

Let us write (21) in the following form:

$$U(t) = e^{-\alpha\tau(t)} \int\limits_{0}^{t-\tau(t)} e^{-\alpha((t-\tau(t))-s)} Z(s)ds = e^{-\alpha\tau(t)} \tilde{U}(t - \tau(t)). \qquad (22)$$

Reducing integro-differential system (3), where $U(t)$ is defined by (21), to an ordinary differential system, we obtain

$$\begin{cases} X' = r - dX - (1-\eta)\beta V X \\ Y' = (1-\eta)\beta V X - aY - pYZ \\ V' = (1-\varepsilon)kY - uV - qVW \\ W' = -hW + gVW \\ Z' = cYZ - bZ + D\tilde{U} + De^{-\alpha\tau(t)}\tilde{U}(t - \tau(t)) - D\tilde{U} \\ \tilde{U}' = Z - \alpha\tilde{U}. \end{cases} \qquad (23)$$

Linearizing systems (23) in the neighborhood of the disease-free equilibrium, we obtain the corresponding homogeneous linear systems:

$$\begin{cases} x_1'(t) = r - dx_1(t) - \frac{(1-\eta)\beta r}{d}x_3(t) \\ x_2'(t) = -ax_2(t) + \frac{(1-\eta)\beta r}{d}x_3(t) \\ x_3'(t) = (1-\varepsilon)kx_2(t) - ux_3(t) \\ x_4'(t) = -hx_4(t) \\ x_5'(t) = Dx_6(t) - bx_5(t) + De^{-\alpha\tau(t)}x_6(t - \tau(t)) - Dx_6(t) \\ x_6'(t) = x_5(t) - \alpha x_6(t). \end{cases} \qquad (24)$$

Let us denote

$$Q_2 = De^{\alpha\tau_*} \left(\left| \frac{c_{28}}{c_{14} - c_{13}} \right| + \left| \frac{c_{28}}{c_{13} + c_{14}} \right| + \left| \frac{c_{31}}{c_{14} - c_{13}} \right| + \left| \frac{c_{32}}{c_{13} + c_{14}} \right| \right)$$

$$+ D \left(2 |c_{28}| + |c_{31}| + |c_{32}| \right).$$

Theorem 5 [25] *Let the assumption of Theorem 2 be fulfilled, $c_{13} \neq c_{14}$ and the inequality $Q_2 < 1$ be true, then system (24) is exponentially stable in the neighborhood of disease-free equilibrium.*

Example 3 Using the coefficient values from Remark 1, and taking the following values of parameters $\alpha = 0.5$, $D = 0.01$, such that condition $D < \alpha b$ is fulfilled. From Theorem 5, $Q_2 \approx 0.33e^{0.5\tau_*} + 0.06$. We obtain the following condition for unbounded memory control function limit: $\tau_* < 2$.

4 Appendix A

Let us denote the following:

$$c_{11} = \frac{d(a+u)}{2d}, c_{12} = \frac{\sqrt{4k\beta rd(\varepsilon-1)(\eta-1)+d^2(a-u)^2}}{2d}, c_{13} = \frac{\alpha+b}{2},$$

$$c_{14} = \frac{\sqrt{(\alpha-b)^2+4D}}{2}, c_{15} = \frac{\beta r(\eta-1)}{d(-c_{11}+c_{12}+d)}, c_{16} = \frac{\beta r(\eta-1)}{d(-c_{11}+c_{12}+a)},$$

$$c_{17} = \frac{\beta r(\eta-1)}{d(-c_{11}-c_{12}+d)}, c_{18} = \frac{\beta r(\eta-1)}{d(-c_{11}-c_{12}+a)}, c_{19} = \frac{D}{-\frac{\alpha}{2}+\frac{b}{2}+c_{14}},$$

$$c_{21} = \frac{D}{-\frac{\alpha}{2}+\frac{b}{2}-c_{14}}, c_{22} = \frac{c_{15}-c_{17}}{c_{16}-c_{18}}, c_{23} = \frac{1}{c_{16}-c_{18}}, c_{25} = \frac{c_{15}\cdot c_{18}-c_{16}\cdot c_{17}}{c_{16}-c_{18}},$$

$$c_{26} = \frac{c_{18}}{c_{16}-c_{18}}, c_{27} = \frac{c_{16}}{c_{16}-c_{18}}, c_{28} = \frac{1}{c_{19}-c_{21}}, c_{31} = \frac{c_{21}}{c_{19}-c_{21}}, c_{32} = \frac{c_{19}}{c_{19}-c_{21}}.$$

We assume that the denominators are not zero.

The Cauchy matrix of system (4) is given in the following (see [24]):

$$C_1(t,s) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} e^{-d(t-s)},$$

$$C_2(t,s) = \begin{pmatrix} c_{22} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} e^{-d(t-s)} + \begin{pmatrix} -c_{15}\cdot c_{23} \\ c_{16}\cdot c_{23} \\ -c_{23} \\ 0 \\ 0 \\ 0 \end{pmatrix} e^{(-c_{11}+c_{12})(t-s)} + \begin{pmatrix} c_{17}\cdot c_{23} \\ -c_{18}\cdot c_{23} \\ c_{23} \\ 0 \\ 0 \\ 0 \end{pmatrix} e^{(-c_{11}-c_{12})(t-s)},$$

$$C_3(t,s) = \begin{pmatrix} c_{25} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} e^{-d(t-s)} + \begin{pmatrix} -c_{15}\cdot c_{26} \\ c_{16}\cdot c_{26} \\ -c_{26} \\ 0 \\ 0 \\ 0 \end{pmatrix} e^{(-c_{11}+c_{12})(t-s)} + \begin{pmatrix} c_{17}\cdot c_{27} \\ -c_{18}\cdot c_{27} \\ c_{27} \\ 0 \\ 0 \\ 0 \end{pmatrix} e^{(-c_{11}-c_{12})(t-s)},$$

$$C_4(t,s) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} e^{-h(t-s)},$$

$$C_5(t,s) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ c_{19} \cdot c_{28} \\ c_{28} \end{pmatrix} e^{(-c_{13}+c_{14})(t-s)} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -c_{21} \cdot c_{28} \\ -c_{28} \end{pmatrix} e^{(-c_{13}-c_{14})(t-s)},$$

$$C_6(t,s) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -c_{19} \cdot c_{31} \\ -c_{31} \end{pmatrix} e^{(-c_{13}+c_{14})(t-s)} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ c_{21} \cdot c_{32} \\ c_{32} \end{pmatrix} e^{(-c_{13}-c_{14})(t-s)}.$$

References

1. Tang LSY, Covert E, Wilson E, Kottilil S. Chronic Hepatitis B Infection: A Review. JAMA. 2018 May 1;319(17):1802-1813. https://doi.org/10.1001/jama.2018.3795.Erratum in: JAMA. 2018 Sep 18;320(11):1202. PMID: 29715359.
2. Parkin DM. The global health burden of infection-associated cancers in the year 2002. Int J Cancer. 2006 Jun 15;118(12):3030-44. https://doi.org/10.1002/ijc.21731.PMID: 16404738.
3. European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu; European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. J Hepatol. 2017 Aug;67(2):370-398. https://doi.org/10.1016/j.jhep.2017.03.021.Epub 2017 Apr 18. PMID: 28427875.
4. Tenney DJ, Rose RE, Baldick CJ, Pokornowski KA, Eggers BJ, Fang J, Wichroski MJ, Xu D, Yang J, Wilber RB, Colonno RJ. Long-term monitoring shows hepatitis B virus resistance to entecavir in nucleoside-naïve patients is rare through 5 years of therapy. Hepatology. 2009 May;49(5):1503-14. https://doi.org/10.1002/hep.22841.PMID: 19280622.
5. Buster EH, Flink HJ, Cakaloglu Y, Simon K, Trojan J, Tabak F, So TM, Feinman SV, Mach T, Akarca US, Schutten M, Tielemans W, van Vuuren AJ, Hansen BE, Janssen HL. Sustained HBeAg and HBsAg loss after long-term follow-up of HBeAg-positive patients treated with peginterferon alpha-2b. Gastroenterology. 2008 Aug;135(2):459-67. https://doi.org/10.1053/j.gastro.2008.05.031.Epub 2008 May 15. PMID: 18585385.
6. Marcellin P, Lau GK, Bonino F, Farci P, Hadziyannis S, Jin R, Lu ZM, Piratvisuth T, Germanidis G, Yurdaydin C, Diago M, Gurel S, Lai MY, Button P, Pluck N; Peginterferon Alfa-2a HBeAg-Negative Chronic Hepatitis B Study Group. Peginterferon alfa-2a alone, lamivudine alone, and the two in combination in patients with HBeAg-negative chronic hepatitis B. N Engl J Med. 2004 Sep 16;351(12):1206-17. https://doi.org/10.1056/NEJMoa040431.PMID: 15371578.
7. Buti M, Tsai N, Petersen J, Flisiak R, Gurel S, Krastev Z, Aguilar Schall R, Flaherty JF, Martins EB, Charuworn P, Kitrinos KM, Subramanian GM, Gane E, Marcellin P. Seven-year efficacy and safety of treatment with tenofovir disoproxil fumarate for chronic hepatitis B virus infection. Dig Dis Sci. 2015 May;60(5):1457-64. https://doi.org/10.1007/s10620-014-3486-7.Epub 2014 Dec 23. PMID: 25532501; PMCID: PMC4427621.
8. Monsalve-De Castillo F, Romero TA, Estévez J, Costa LL, Atencio R, Porto L, Callejas D. Concentrations of cytokines, soluble interleukin-2 receptor, and soluble CD30 in sera of patients with hepatitis B virus infection during acute and convalescent phases. Clin Diagn Lab Immunol. 2002 Nov;9(6):1372-5. https://doi.org/10.1128/cdli.9.6.1372-1375.2002.PMID: 12414777; PMCID: PMC130099.

9. Colombatto P, Civitano L, Bizzarri R, Oliveri F, Choudhury S, Gieschke R, Bonino F, Brunetto MR; Peginterferon Alfa-2a HBeAg-Negative Chronic Hepatitis B Study Group. A multiphase model of the dynamics of HBV infection in HBeAg-negative patients during pegylated interferon-alpha2a, lamivudine and combination therapy. Antivir Ther. 2006;11(2):197-212. PMID: 16640101.

10. Nowak MA, Bonhoeffer S, Hill AM, Boehme R, Thomas HC, McDade H. Viral dynamics in hepatitis B virus infection. Proc Natl Acad Sci U S A. 1996 Apr 30;93(9):4398-402. https://doi.org/10.1073/pnas.93.9.4398.PMID: 8633078; PMCID: PMC39549.

11. Perelson AS, Ribeiro RM. Hepatitis B virus kinetics and mathematical modeling. Semin Liver Dis. 2004;24 Suppl 1:11-6. https://doi.org/10.1055/s-2004-828673.PMID: 15192796.

12. Fatehi Chenar F, Kyrychko YN, Blyuss KB. Mathematical model of immune response to hepatitis B. J Theor Biol. 2018 Jun 14;447:98-110. https://doi.org/10.1016/j.jtbi.2018.03.025.Epub 2018 Mar 21. PMID: 29574141.

13. A. Domoshnitsky, I.Volinsky, M.Bershadsky, Around the Model of Infection Disease: The Cauchy Matrix and Its Properties, *Symmetry* 2019, 11(8),1016; https://doi.org/10.3390/sym11081016

14. Alexander Domoshnitsky, Irina Volinsky and Olga Pinhasov, Some developments in the model of testosterone regulation, *AIP Conference Proceedings* **2159**, 030010, (2019); https://doi.org/10.1063/1.5127475

15. Yousdi, N.; Hattaf, K.; Rachik, M. Analysis of a HCV model with CTL and antibody responses. *Appl. Math. Sci.* 3, 2009, 2835–2845.

16. Chenar, F.F.; Kyrychko, Y.N.; Blyuss, K.B. Mathematical model of immune response to hepatitis B. *J. Theor. Biol.*, **2018**, *447*, 98–110.

17. Domoshnitsky, A.; Volinsky, I.; Polonsky, A.; Sitkin, A. Stabilization by delay distributed feedback control. *Math. Model. Nat. Phenom.* **2017**, *12*, 91–105 https://doi.org/10.1051/mmnp/2017067.

18. Domoshnitsky, A.; Volinsky, I.; Polonsky, A. Stabilization of third order differential equation by delay distributed feedback control with unbounded memory. *Math. Slovaca* **2019**, *69*, 1165–1176. https://doi.org/10.1515/ms-2017-0298.

19. Domoshnitsky, A.; Volinsky, I.; Bershadsky, M. Around the Model of Infection Disease: The Cauchy Matrix and Its Properties. *Symmetry* **2019**, *11*, 1016; https://doi.org/10.3390/sym11081016.

20. Bershadsky, M.; Chirkov, M.; Domoshnitsky, A.; Rusakov, S.; Volinsky, I. Distributed Control and the Lyapunov Characteristic Exponents in the Model of Infectious Diseases. *Complexity* **2019**, *2019*, 5234854, https://doi.org/10.1155/2019/5234854.

21. Domoshnitsky, A.; Volinsky, I.; Levi, S.; Shemesh, S. Stability of third order neutral delay differential equations. *AIP Conf. Proc.* **2019**, *2159*, 020002, https://doi.org/10.1063/1.5127464.

22. Domoshnitsky, A.; olinsky, I.; Pinhasov, O.; Bershadsky, M. Questions of Stability of Functional Differential Systems around the Model of Testosterone Regulation. *Bound. Value Probl.* **2019**, *1*, 1–13; https://doi.org/10.1186/s13661-019-01295-2.

23. Domoshnitsky, A.; Volinsky, I.; Polonsky, A. Practical constructing the Cauchy function of integro-differential equation, *Funct. Differ. Equations* **2016**, *23*, 109–118.

24. Volinsky I., Lombardo S.D., Cheredman P. Stability Analysis and Cauchy Matrix of a Mathematical Model of Hepatitis B Virus with Control on Immune System near Neighborhood of Equilibrium Free Point, Symmetry, 13 (2), 166, 2021.

25. Volinsky I. Stability Analysis of a Mathematical Model of Hepatitis B Virus with Unbounded Memory Control on the Immune System in the Neighborhood of the Equilibrium Free Point, Symmetry, 13 (8), 1437, 2021.

Oncolytic Virus Versus Cancer: Modeling and Simulation of Virotherapy with Differential Equations

Iordanka Panayotova and Maila Hallare

Abstract Mathematical models are useful analytical tools that can help medical practitioners in understanding cancer treatment options. Testing various treatment assumptions and scenarios by varying the parameters in a mathematical model can provide analysis that may help in predicting prognosis for patients, improving the effectiveness of treatment plans, and providing deeper insights into questions that cannot be addressed by clinical or experimental studies alone. In this research, we use differential equations to model and simulate virotherapy as a treatment approach for spherical tumors. Using a system of five nonlinear ordinary differential equations, we investigate the complex dynamics between cancer cells, infected cancer cells, oncolytic viruses, immune cells, and dead cancer cells using analytical methods and numerical simulations. The model is analyzed qualitatively by computing the equilibria and deriving conditions for their local stability. The rate of change of the tumor's radius was derived and used to provide the basis for numerical simulations that establish the effectiveness of virotherapy as a treatment. Numerical simulations and sensitivity analysis were performed to identify parameters with the greatest effect on the success of virotherapy which could be then used in the derivation of optimal therapeutic strategies.

Keywords Mathematical modeling · Cancer treatment · Numerical simulations · Ordinary differential equations · Equilibria · Stability · Virotherapy

I. Panayotova (✉)
Department of Mathematics, Christopher Newport University, 1 Avenue of Arts, Newport News, VA 23606, USA
e-mail: iordanka.panayotova@cnu.edu

M. Hallare
Department of Mathematical Sciences, United States Air Force Academy, 2354 Fairchild Dr Ste 6D124 USAF Academy, Colorado Springs, CO 80840, USA
e-mail: maila.hallare@afacademy.af.edu

1 Introduction

According to the World Health Organization, cancer is defined as "a generic term for a large group of diseases that can affect any part of the body" [1]. A defining characteristic of cancer lies in its rapid creation of abnormal cells beyond usual boundaries in the body which can, as malignant tumors grow, invade adjoining parts of the body and spread to other tissues. This process is called metastases, and it is a major cause of death for patients with cancer [1]. Cancer is one of the leading causes of death in the world, but there is still no functional cure for it. There are more than 9.8 million deaths related to cancer worldwide every year. Yet, according to the American Cancer Society, there will be 1,918,030 new cancer cases and more than 600 thousand cancer deaths in the US in 2022 alone [2].

Some of the traditional methods to treat cancer include surgery, chemotherapy, radiation, and stem cell transplants. A fairly new and very promising approach designed to fight cancer with less toxicity is virotherapy. This therapy uses a genetically modified virus which is injected into the cancer tumor directly [3]. The oncolytic virus attacks the cancer cells by introducing its own RNA, causing the cell to produce more viruses, and as a result dies by releasing the viruses in a process called lysis. A major advantage of virotherapy is that while the virus attacks cancer cells, it does not affect healthy cells; therefore; the virotherapy is safer and less invasive than other methods of cancer treatment. Even though virotherapy is a very promising technique, it is not without its challenges. When the immune system recognizes the virus as an invader, it attacks the virus and works to destroy it, making it impossible for the oncolytic virus to spread and infect other cancer cells [4]. Hence, for virotherapy to be effective, the immune system must be suppressed.

One way to study how the suppression of the immune system will affect the effectiveness of virotherapy is through mathematical modeling. Mathematical models and numerical simulations are powerful tools that can be used to reduce the experimental cost of virotherapy by simulating the dynamics outcome of a given setup of the system. It can also be used to test hypothesis, understand the effect of each parameter on the system's dynamics, and study the interactions with other variables. Here, we use five differential equations to study the short- and long-term effects of virotherapy as a treatment for cancerous tumors with different types of immune-suppression.

Previous studies of mathematical modeling of oncolytic virotherapy include treatment with time delay [5] and modeling dynamics of cancer virotherapy with immune response [6]. In Wang's paper [5], virotherapy was modeled focusing on the lytic cycle and virus-specific CTL response, which relates to the body's immune system. Wang's results supported the need to include an immune-suppressing drug in order to optimize virotherapy as a cancer treatment. Al-Tuwairqi's model included the stimulation of immune response due to the presence of tumors as well as the destruction of tumors due to the release of cytokines from natural killer cells [6].

In this work, we explore a mathematical model of virotherapy with the immune response which is the ODE version of the PDE model introduced in [7]. We extend the model to include dead cells in order to use the principle of conservation of mass in

deriving the equation for the rate of change of the tumor's radius. We use that equation to verify the efficacy of the therapy. A simplified version of this model, using three and four state variables, was developed by the authors in [8] particularly for use in a differential equations class with the goal of teaching undergraduate students direct applications of mathematical modeling to contemporary issues in biology. Mathematical models are useful tools that may help in understanding cancerous tumor growth; in this way, they can contribute to better predict the prognosis for patients and improve the effectiveness of the treatment plans. Mathematical models can be also used to test different assumptions by varying the models' parameters. In this way, they provide deeper insights into questions that cannot be addressed by clinical studies or experimental studies, which can be very expensive. Using sensitivity analysis by direct differential methods, we found the most sensitive parameters of this model which then were varied to gain more insight into these critical parameters. This may possibly lead to better efficacy of virotherapy as a cancer treatment.

The remainder of this paper is organized as follows. First, we introduce the mathematical model and its underlying assumptions. Next, we study the model qualitatively by showing the existence and uniqueness of the state variables, calculating the steady-state solutions, and finding their local stability. Then, we perform a sensitivity analysis on the system in order to investigate the behaviors of solutions with respect to small perturbations in the parameters' values.

2 Mathematical Model

We use a mathematical model to study the complex interactions between cancer cells, oncolytic virus particles, and immune cells during virotherapy as a cancer treatment and the effect of immuno-suppression on the effectiveness of the treatment. Suppose $x(t)$ is the number of uninfected cancer cells, $y(t)$ is the number of infected cancer cells, $z(t)$ is the number of immune cells, $v(t)$ is the number of oncolytic virus particles, and $n(t)$ is the number of dead cancer cells. We assume that the cancer cells grow exponentially with an intrinsic growth rate α. This assumption is reasonable in the initial stages of tumor growth, which is the goal of our investigation. Data collected over the span of 30 days from mice in Huang et al. [9] did not show any signs of attaining a carrying capacity, which also justifies the exponential growth assumption in our model. Moreover, this simplification makes the model analytically tractable. When the virus is injected, it infects the cancer cells at a rate β; as a result, the infected cancer cells die at a rate δ. Due to lysis, new virus particles are released with a burst size b. The free virus particles decay at a rate γ. Once infected, cancer cells trigger the response of the immune system, which kills the infected cancer cells at a rate of k. These cells then become dead cells and contribute to the rate of change of n with a clearance rate of μ. The immune response is stimulated by the infected cancer cells at a rate s and decays at a rate ω. The dynamics of the model is illustrated by the compartment diagram in which the solid lines represent stimulation or growth

and the dashed lines represent inhibition or killing/decay, and they are governed by the following system of five nonlinear ordinary differential equations:
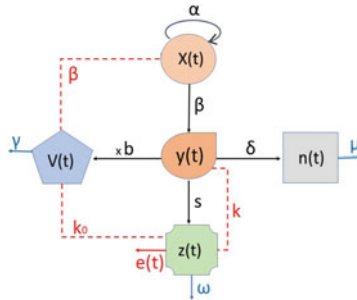
Compartmental model diagram.

$$\begin{cases} \frac{dx}{dt} = \alpha x - \beta x v \\ \frac{dy}{dt} = \beta x v - kyz - \delta y \\ \frac{dz}{dt} = syz - \omega z^2 - ez \\ \frac{dn}{dt} = kyz + \delta y - \mu n \\ \frac{dv}{dt} = b\delta y - k_0 vz - \gamma v \end{cases} \tag{1}$$

with initial conditions

$$x(0) = x_0 \geq 0, \quad y(0) = y_0 \geq 0, \quad z(0) = z_0 \geq 0, \quad n(0) = n_0 \geq 0, \quad v(0) = v_0 \geq 0.$$

We consider three kinds of delivery $e(t)$ of the immune-suppressor drug, namely a constant E, a decaying function with decay parameter $q > 0$ and initial amplitude E, that is, $e(t) = E \exp(-qt)$, and a periodic function with period $P > 0$ and amplitude $E > 0$, that is, $e(t) = E \sin\left(\frac{2\pi}{P}t\right)$. Hence, with t_0 as the initial time and t_F as the final time, the immune-suppressing term has the form:

$$e(t) = \begin{cases} E, & t_0 \leq t \leq t_F \\ E \exp(-qt), & t_0 \leq t \leq t_F \\ E \sin\left(\frac{2\pi}{P}t\right), & t_0 \leq t \leq t_F. \end{cases}$$

3   Qualitative Analysis

In this section, we investigate model (1) qualitatively. We first discuss the existence and uniqueness of the state variables, then we find the steady-state solutions of the model and define their stability.

3.1   Existence and Uniqueness of the Solutions

Theorem 1 *Given an initial condition* $(x_0, y_0, z_0, n_0, v_0)$ *in the region*

$$Q = \{(x, y, z, n, v) \in \mathbb{R}^5 : x \geq 0, y \geq 0, z \geq 0, n \geq 0, v \geq 0\}$$

the system (1) has a unique solution which lies in Q.

Proof The right-hand side of each differential equation is continuously differentiable in a neighborhood of Q and thus for initial conditions in Q, locally, there exists a unique solution. Now we will show that Q is positively invariant. Rewrite the first equation of (1) as

$$\frac{dx}{x} = \psi_1(v)dt$$

where $\psi_1(x, v) = \alpha - \beta v$. Integrating over $[0, t]$, we obtain

$$x(t) = x(0) \exp\left[\int_0^t \psi_1(v)ds\right].$$

From here, since $x(0) = x_0 \geq 0$ we conclude that $x(t) \geq 0$ for all $t \geq 0$. From the second equation of (1), we see that

$$\frac{dy}{dt} \geq -y(kz + \delta),$$

which can be rewritten as

$$\frac{dy}{y} \geq \psi_2(z)dt,$$

where $\psi_2(z) = -(kz + \delta)$. Now, integrating over $[0, t]$ gives

$$y(t) \geq y(0) \exp\left[\int_0^t \psi_2(z)ds\right]$$

from where we can conclude that if $y(0) = y_0 \geq 0$ then $y(t) \geq 0$ for all t. Similarly, from the third equation in (1), we can write it as the inequality

$$\frac{dz}{z} \geq \psi_3(z)dt,$$

where $\psi_3(z) = -(\omega z + e(t))$, and after integrating over $[0, t]$, we get the inequality

$$z(t) \geq z(0) \exp\left[\int_0^t \psi_3(z)ds\right].$$

Again, if $z(0) \geq 0$, then $z(t) \geq 0$ for all t. Similarly, from the fourth equation, rewritten as the differential inequality $\frac{dn}{n} \geq -\mu$ and after integrating over $[0, t]$, we obtain $n(t) \geq n(0)e^{-\mu t}$, which is always non-negative for a non-negative initial condition. Finally, from the last equation, we obtain

$$\frac{dv}{v} \geq \psi_4(z),$$

where $\psi_4(z) = -(k_0 z + \gamma)$, and integrating over $[0, t]$ gives

$$v(t) \geq v(0) \exp\left[\int_0^t \phi_4(z)ds\right],$$

which is non-negative as long as the initial condition is non-negative. This concludes the proof that all solutions are non-negative and Q is positively invariant, i.e., solutions that start in Q will remain in Q throughout their interval of existence.

3.2 Equilibrium Solutions

The equilibria of the system are the steady-state solutions obtained by setting the rates of change of equations in (1) to zero and solving for x, y, z, n, and v. We write equilibria in the form $E(x, y, z, n, v)$. Thus, we consider the system

$$\begin{cases} x(\alpha - \beta v) & = 0 \\ \beta xv - kyz - \delta y & = 0 \\ z(sy - \omega z - e) & = 0 \\ kyz + \delta y - \mu n & = 0 \\ b\delta y - k_0 vz - \gamma v = 0. \end{cases} \quad (2)$$

Case 1. Suppose $x = 0$ and $z = 0$. Then the second, fourth, and fifth equations in (2) tell us that $y = 0$, $n = 0$, $v = 0$, respectively. We label this equilibrium as $E_0 = (0, 0, 0, 0, 0)$. This equilibrium E_0 represents the free equilibrium where all cancer cells (uninfected, infected, and dead), immune cells, and viruses are zero.

Case 2. Suppose $x = 0$ and $z \neq 0$. Then the system (2) reduces to

$$\begin{cases} x & = 0 \\ y(kz + \delta) & = 0 \\ sy - \omega z - e & = 0 \\ kyz + \delta y - \mu n & = 0 \\ b\delta y - k_0 vz - \gamma v = 0, \end{cases} \quad (3)$$

which tells us that there are two sub-cases: $y = 0$ or $z = -\frac{\delta}{k}$. Since we do not consider negative solutions z, it follows that y must be zero. Now, assuming $y = 0$, the third equation in the reduced system shows that $z = \frac{-e(t)}{w}$.

Note that this state is possible in case the delivery of the immune-suppressor drug has the form $e(t) = E \sin\left(\frac{2\pi}{P}t\right)$. Next, with $y = 0$, the fourth equation implies that $n = 0$. Finally, with $y = 0$ and $z = \frac{-e(t)}{\omega}$, we have that $v = 0$.

Summarizing, in case $e(t)$ is either a constant or an exponentially decaying function, then there is no equilibrium with $x = 0$, $z \neq 0$. But with $z = -\frac{e(t)}{\omega}$ over some t where $e(t)$ is negative, our second equilibrium is defined by

$$E_1 = \left(0, 0, \frac{-1}{\omega}e(t), 0, 0\right), \text{ provided, } e(t) = \begin{cases} E \sin\left(\frac{2\pi}{P}t\right), & \frac{P}{2} < t < P \\ 0, & \text{otherwise} \end{cases}.$$

Case 3. Suppose $x \neq 0$ and $z = 0$. Now, the system (2) becomes

$$\begin{cases} \alpha - \beta v = 0 \\ \beta xv - \delta y = 0 \\ z = 0 \\ \delta y - \mu n = 0 \\ b\delta y - \gamma v = 0, \end{cases} \implies \begin{cases} v = \frac{\alpha}{\beta} \\ x = \frac{\delta}{\beta}\frac{y}{v} \\ z = 0 \\ n = \frac{\delta}{\mu}y \\ y = \frac{\gamma}{b\delta}v. \end{cases}$$

With $v = \frac{\alpha}{\beta}$ from the first equation, we now have a third equilibria E_2:

$$E_2\left(\frac{\gamma}{b\beta}, \frac{\alpha\gamma}{b\beta\delta}, 0, \frac{\alpha\gamma}{b\beta\mu}, \frac{\alpha}{\beta}\right).$$

Now, we see that when z is zero, or there are no immune cells to attack the virus, there is a partial victory for virotherapy as there is a coexistence of uninfected and infected cells as well as the viruses, as described in the equilibrium E_2. In addition, increasing the bursting parameter b and the infection rate β decreases the steady state for the cancer cells, and hence the virotherapy is more efficient.

Case 4.   Finally, suppose that $x \neq 0$ and $z \neq 0$. Then the system (2) reduces to

$$\begin{cases} \alpha - \beta v = 0 \\ \beta xv - kyz - \delta y = 0 \\ sy - \omega z - e = 0 \\ kyz + \delta y - \mu n = 0 \\ b\delta y - k_0 vz - \gamma v = 0, \end{cases} \implies \begin{cases} v = \frac{\alpha}{\beta} \\ \alpha x - kyz - \delta y = 0 \\ sy - \omega z - e = 0 \\ kyz + \delta y - \mu n = 0 \\ b\delta y - \frac{k_0\alpha}{\beta}z - \frac{\alpha\gamma}{\beta} = 0. \end{cases}$$

Observe that we can solve for y and z by looking at the third and fifth equations to get the smaller system:

$$\begin{cases} sy - \omega z - e = 0 \\ b\delta y - \frac{k_0\alpha}{\beta}z - \frac{\alpha\gamma}{\beta} = 0. \end{cases}$$

Using Cramer's Rule, for example, we obtain

$$y = \frac{\begin{vmatrix} e & -\omega \\ \frac{\alpha\gamma}{\beta} & \frac{-k_0\alpha}{\beta} \end{vmatrix}}{\begin{vmatrix} s & -\omega \\ b\delta & \frac{-k_0\alpha}{\beta} \end{vmatrix}} = \frac{\frac{-ek_0\alpha}{\beta} + \frac{\alpha\gamma\omega}{\beta}}{\frac{-sk_0\alpha}{\beta} + b\delta\omega} = \frac{\alpha\gamma\omega - ek_0\alpha}{b\beta\delta\omega - sk_0\alpha}$$

$$z = \frac{\begin{vmatrix} s & e \\ b\delta & \frac{\alpha\gamma}{\beta} \end{vmatrix}}{\begin{vmatrix} s & -\omega \\ b\delta & \frac{-k_0\alpha}{\beta} \end{vmatrix}} = \frac{\frac{s\alpha\gamma}{\beta} - be\delta}{\frac{-sk_0\alpha}{\beta} + b\delta\omega} = \frac{s\alpha\gamma - be\beta\delta}{b\beta\delta\omega - sk_0\alpha}.$$

To ensure existence of z, we must require that the above denominator $b\beta\delta\omega - sk_0\alpha$ is nonzero. Then, to solve for x and n, we have

$$x = \frac{1}{\alpha}(kyz + \delta y) \quad \text{and} \quad n = \frac{1}{\mu}(kyz + \delta y),$$

and so we get

$$x = \frac{\gamma\omega - ek_0}{b\beta\delta\omega - sk_0\alpha}\left(\frac{ks\alpha\gamma - kb\beta\delta e}{-sk_0\alpha - b\beta\delta\omega} + \delta\right) \quad \text{and} \quad n = \frac{\alpha\gamma\omega - ek_0\alpha}{\mu b\beta\delta\omega - \mu sk_0\alpha}\left(\frac{ks\alpha\gamma - kb\beta\delta e}{-sk_0\alpha - b\beta\delta\omega} + \delta\right)$$

Thus, we have a fourth equilibrium point $E_3(x_3, y_3, z_3, n_3, v_3)$ where

$$
\begin{aligned}
x_3 &= \frac{\gamma\omega - ek_0}{b\beta\delta\omega - sk_0\alpha}\left(\frac{ks\alpha\gamma - kb\beta\delta e}{b\beta\delta\omega - sk_0\alpha} + \delta\right) \\
y_3 &= \frac{\alpha\gamma\omega - ek_0\alpha}{b\beta\delta\omega - sk_0\alpha} \\
z_3 &= \frac{s\alpha\gamma - be\beta\delta}{b\beta\delta\omega - sk_0\alpha} \\
n_3 &= \frac{\alpha\gamma\omega - ek_0\alpha}{\mu b\beta\delta\omega - \mu sk_0\alpha}\left(\frac{ks\alpha\gamma - kb\beta\delta e}{b\beta\delta\omega - sk_0\alpha} + \delta\right) \\
v_3 &= \frac{\alpha}{\beta}.
\end{aligned}
$$

Observe that the equilibrium E_3 is the so-called coexistence equilibrium, where all cells and viruses are present. We get the following statement for the existence of equilibrium E_3:

Theorem 2 *If the following conditions are true, E_3 is biologically relevant:*

$$\min\left\{\frac{k_0}{\omega}, \frac{\gamma}{e}\right\} < \frac{b\beta\delta}{\alpha s} < \max\left\{\frac{k_0}{\omega}, \frac{\gamma}{e}\right\}.$$

3.3 Stability of Equilibria

Here, we investigate the local stability of the equilibrium points using the linearization method. The Jacobian $J(E)$ at the point $E(x, y, z, n, v)$ associated with the system (1) is given by

$$J(E) = \begin{pmatrix} \alpha - \beta v & 0 & 0 & 0 & -\beta x \\ \beta v & -kz - \delta & -ky & 0 & \beta x \\ 0 & sz & sy - 2\omega z - e & 0 & 0 \\ 0 & kz + \delta & ky & -\mu & 0 \\ 0 & b\delta & -k_0 v & 0 & -k_0 z - \gamma \end{pmatrix}. \tag{4}$$

3.3.1 Stability of the Equilibrium E_0

The stability of the equilibrium E_0 can be obtained by analyzing the eigenvalues of $J(E_0)$:

$$J(E_0) = \begin{pmatrix} \alpha & 0 & 0 & 0 & 0 \\ 0 & -\delta & 0 & 0 & 0 \\ 0 & 0 & -e & 0 & 0 \\ 0 & \delta & 0 & -\mu & 0 \\ 0 & b\delta & 0 & 0 & -\gamma \end{pmatrix}.$$

Since this matrix is lower-triangular, its eigenvalues are the diagonal entries. Hence, the eigenvalues are $\lambda_1 = \alpha$, $\lambda_2 = -\delta$, $\lambda_3 = -e$, $\lambda_4 = -\mu$, and $\lambda_5 = -\gamma$. Four of the eigenvalues are negative while one of them is positive; thus, E_0 is an unstable equilibrium.

Theorem 3 *The zero equilibrium $E_0 = (0, 0, 0, 0, 0)$ has one positive and four negative eigenvalues for any parameter values and hence is always unstable.*

3.3.2 Stability of the Equilibrium E_1

To analyze the stability of $E_1 = (0, 0, \frac{-1}{\omega} e(t), 0, 0)$, let us find the Jacobian at E_1:

$$J(E_1) = \begin{pmatrix} \alpha & 0 & 0 & 0 & 0 \\ 0 & -\delta & 0 & 0 & 0 \\ 0 & -\frac{se}{\omega} & e & 0 & 0 \\ 0 & \delta - \frac{ke}{\omega} & 0 & -\mu & 0 \\ 0 & b\delta & 0 & 0 & \frac{k_0 e}{\omega} - \gamma \end{pmatrix}.$$

Since the matrix is lower-triangular, its eigenvalues are the elements on its diagonal. Hence, the eigenvalues are $\lambda_1 = \alpha, \lambda_2 = -\delta, \lambda_3 = -e, \lambda_4 = -\mu$, and $\lambda_5 = \frac{k_0 e}{\omega} - \gamma$. As at least one eigenvalue is always positive, this equilibrium is always unstable.

Theorem 4 *The equilibrium $E_1 = (0, 0, \frac{-1}{\omega} e(t), 0, 0)$ has at least one positive eigenvalue for any parameter values and hence is always unstable.*

3.3.3 Stability of the Equilibrium E_2

The stability of the equilibrium E_2 can be obtained by analyzing the eigenvalues of $J(E_2)$:

$$J(E_2) = \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{-\gamma}{b} \\ \alpha & -\delta & \frac{-k\alpha\gamma}{b\beta\delta} & 0 & \frac{\gamma}{b} \\ 0 & 0 & \frac{s\alpha\gamma}{b\beta\delta} - e & 0 & 0 \\ 0 & \delta & \frac{k\alpha\gamma}{b\beta\delta} & -\mu & 0 \\ 0 & b\delta & \frac{-k_0\alpha}{\beta} & 0 & -\gamma \end{pmatrix}. \tag{5}$$

The eigenvalues λ can be determined by solving the characteristic equation $\det(\lambda I - J(E_1)) = 0$, where

$$\lambda I - J(E_2) = \begin{pmatrix} \lambda & 0 & 0 & 0 & \frac{\gamma}{b} \\ -\alpha & \lambda + \delta & \frac{k\alpha\gamma}{b\beta\delta} & 0 & -\frac{\gamma}{b} \\ 0 & 0 & \lambda - \frac{s\alpha\gamma}{b\beta\delta} + e & 0 & 0 \\ 0 & -\delta & -\frac{k\alpha\gamma}{b\beta\delta} & \lambda + \mu & 0 \\ 0 & -b\delta & \frac{k_0\alpha}{\beta} & 0 & \lambda + \gamma \end{pmatrix}.$$

Computing the determinant by expanding about the fourth column to get

$$\det(\lambda I - J(E_2)) = (\lambda + \mu) \det \begin{pmatrix} \lambda & 0 & 0 & \frac{\gamma}{b} \\ -\alpha & \lambda + \delta & \frac{k\alpha\gamma}{b\beta\delta} & -\frac{\gamma}{b} \\ 0 & 0 & \lambda - \frac{s\alpha\gamma}{b\beta\delta} + e & 0 \\ 0 & -b\delta & \frac{k_0\alpha}{\beta} & \lambda + \gamma \end{pmatrix},$$

then expanding about the third row:

$$\det(\lambda I - J(E_2)) = (\lambda + \mu) \left(\lambda - \frac{s\alpha\gamma}{b\beta\delta} + e \right) \det \begin{pmatrix} \lambda & 0 & \frac{\gamma}{b} \\ -\alpha & \lambda + \delta & -\frac{\gamma}{b} \\ 0 & -b\delta & \lambda + \gamma \end{pmatrix} =$$
$$= (\lambda + \mu)(\lambda - \frac{s\alpha\gamma}{b\beta\delta} + e)(\lambda(\lambda + \gamma)(\lambda + \delta) + \alpha\delta\gamma - \delta\gamma\lambda)$$
$$= (\lambda + \mu)(\lambda - \frac{s\alpha\gamma}{b\beta\delta} + e)(\lambda^3 + (\delta + \gamma)\lambda^2 + \alpha\delta\gamma).$$

The eigenvalues so far are $\lambda_1 = -\mu$ and $\lambda_2 = \frac{s\alpha\gamma}{b\beta\delta} - e$. Next, we apply the Routh–Hurwitz Criteria on $\lambda^3 + A\lambda^2 + B\lambda + C = 0$, where $A = \delta + \gamma$, $B = 0$, $C = \alpha\delta\gamma$. Since A and C are positive, we need to look at the sign of $AB - C$:

$$AB - C = -\alpha\delta\gamma < 0,$$

which is always negative. According to the Routh–Hurwitz Criteria, there will be a root with a positive real part, which means that this equilibrium will be always unstable. Let us now state our theorem:

Theorem 5 *The equilibrium* $E_2 \left(\frac{\gamma}{b\beta}, \frac{\alpha\gamma}{b\beta\delta}, 0, \frac{\alpha\gamma}{b\beta\mu}, \frac{\alpha}{\beta} \right)$ *is always unstable.*

3.3.4 Stability of the Equilibrium E_3

The stability of the equilibrium E_3 can be obtained by analyzing the eigenvalues of $J(E_3)$:

$$J(E_3) = \begin{pmatrix} 0 & 0 & 0 & 0 & -\beta x \\ \alpha & -kz - \delta & -ky & 0 & \beta x \\ 0 & sz & -\omega z & 0 & 0 \\ 0 & kz + \delta & ky & -\mu & 0 \\ 0 & b\delta & \frac{-k_0\alpha}{\beta} & 0 & -k_0 z - \gamma \end{pmatrix}. \tag{6}$$

The eigenvalues λ can be determined by solving the characteristic equation $\det(\lambda I - J(E_3)) = 0$, where

$$\lambda I - J(E_3) = \begin{pmatrix} \lambda & 0 & 0 & 0 & \beta x \\ -\alpha & \lambda + kz + \delta & ky & 0 & -\beta x \\ 0 & -sz & \lambda + \omega z & 0 & 0 \\ 0 & -kz - \delta & -ky & \lambda + \mu & 0 \\ 0 & -b\delta & \frac{k_0\alpha}{\beta} & 0 & \lambda + k_0 z + \gamma \end{pmatrix}.$$

Computing the determinant by expanding about the fourth column to get

$$\det(\lambda I - J(E_3)) = (\lambda + \mu) \det \begin{pmatrix} \lambda & 0 & 0 & \beta x \\ -\alpha & \lambda + kz + \delta & ky & -\beta x \\ 0 & -sz & \lambda + \omega z & 0 \\ 0 & -b\delta & \frac{k_0\alpha}{\beta} & \lambda + k_0 z + \gamma \end{pmatrix},$$

then expanding about the first row:

$$\det(\lambda I - D(E_3)) = (\lambda + \mu) \left[\lambda \det \begin{pmatrix} \lambda + kz + \delta & ky & -\beta x \\ -sz & \lambda + wz & 0 \\ -b\delta & \frac{k_0\alpha}{\beta} & \lambda + k_0 z + \gamma \end{pmatrix} - \beta x \det \begin{pmatrix} -\alpha & \lambda + kz + \delta & ky \\ 0 & -sz & \lambda + \omega z \\ 0 & -b\delta & \frac{k_0\alpha}{\beta} \end{pmatrix} \right].$$

Hence, one eigenvalue is $\lambda_1 = -\mu$. To find the other four eigenvalues, consider first each determinant from above:

$$\det \begin{pmatrix} \lambda + kz + \delta & ky & -\beta x \\ -sz & \lambda + wz & 0 \\ -b\delta & \frac{k_0\alpha}{\beta} & \lambda + k_0 z + \gamma \end{pmatrix} =$$

$$= (\lambda + kz + \delta)(\lambda + \omega z)(\lambda + k_0 z + \gamma) + sk_0\alpha xz - b\delta\beta x(\lambda + \omega z) + ksyz(\lambda + k_0 z + \gamma),$$

and

$$\det \begin{pmatrix} -\alpha & \lambda + kz + \delta & ky \\ 0 & -sz & \lambda + \omega z \\ 0 & -b\delta & \frac{k_0\alpha}{\beta} \end{pmatrix} = \alpha sz\frac{k_0\alpha}{\beta} - \alpha b\delta(\lambda + \omega z).$$

Combining all together, we find that the other four eigenvalues of $\det(\lambda I - J(E_3))$ are the solutions to the fourth-order polynomial:

$$P(\lambda) = \lambda^4 + A_1\lambda^3 + B_1\lambda^2 + C_1\lambda + D_1 = 0, \quad \text{where}$$

$A_1 = \omega z + kz + \delta + k_0 z + \gamma > 0,$
$B_1 = z^2(k_0\omega + k\omega + kk_0) + z(\omega\gamma + k\gamma + \delta\omega + k_0\delta) + \delta\gamma + ksyz - b\delta\beta x,$
$C_1 = (k_0 z + \gamma)(k\omega z^2 + \delta\omega z + ksyz) + \alpha b\delta\beta x + k_0 s\alpha xz - b\delta\beta\omega xz,$
$D_1 = \alpha b\delta\beta\omega xz - \alpha^2 k_0 sxz = \alpha xz(b\beta\delta\omega - sk_0\alpha).$

Observe that the first coefficient is always positive $A_1 > 0$. For stability, we need all eigenvalues to be either negative real numbers or have a negative real part if they are complex numbers. Using Routh–Hurwitz Criteria, we can find conditions on the coefficients of $P(\lambda)$ which will guarantee negative real parts for the eigenvalues. These results are summarized in the following theorem:

Theorem 6 *For the coexistence equilibrium E_3, the following is true:*

1. *if $C_1 > 0$ and $A_1 B_1 C_1 > C_1^2 + A_1^2 D_1$, then the coexistence equilibrium E_3 is stable;*
2. *if either one of the above conditions is not satisfied, then the coexistence equilibrium E_3 is unstable.*

4   Rate of Change of the Tumor's Radius

We want to find an equation describing the rate of change of the tumor's radius $R(t)$ if we assume that the tumor has a spherical form. This will allow us to gauge the effectiveness of the virotherapy treatment to shrink the tumor. Let us assume that

$P(t) = x(t) + y(t) + z(t) + n(t)$ is the total cell population of the tumor. From the corresponding differential equations in (1), the derivative of P is given by

$$\frac{dP}{dt} = \frac{d}{dt}(x + y + z + n) = \alpha x + syz - \omega z^2 - ez - \mu n. \tag{7}$$

Using the assumption that the tumor is a sphere with radius R, we can find its volume

$$V(t) = \frac{4}{3}\pi R^3(t), \tag{8}$$

and differentiating with respect to t, we get the equation for its rate of change

$$\frac{dV}{dt} = 4\pi R^2 \frac{dR}{dt}.$$

Dividing both sides of the above equation by V and then substituting $V = \frac{4}{3}\pi R^3$, we get the following equation for the volume's rate of change and the radius's rate of change:

$$\frac{1}{V}\frac{dV}{dt} = \frac{3}{R}\frac{dR}{dt}. \tag{9}$$

When P changes, the tumor's volume will also change; however, we can assume that their ratio remains a constant such as $P = \theta_0 V$, where $\theta_0 = x_0 + y_0 + z_0 + n_0$, and without loss of generality, we can assume that x_0, y_0, z_0, and n_0 are the initial conditions for these variables. Hence, differentiating $\frac{P}{V} = \theta_0$ gives the following relationship between their derivatives:

$$\frac{1}{P}\frac{dP}{dt} = \frac{1}{V}\frac{dV}{dt}. \tag{10}$$

Then combining the relationships (7), (9), and (10) gives an equation for the rate of change of R

$$\frac{3}{R}\frac{dR}{dt} = \frac{1}{V}\frac{dV}{dt} = \frac{1}{P}\frac{dP}{dt} = \frac{1}{P}\left(\alpha x + syz - \omega z^2 - ez - \mu n\right)$$

or equivalently

$$\frac{3}{R}\frac{dR}{dt} = \frac{1}{\theta_0 V}\left(\alpha x + syz - \omega z^2 - ez - \mu n\right).$$

Finally, using Eq. (8) for the volume, we get the following equation describing the rate of change of tumor's radius $R(t)$ in time:

$$\frac{dR}{dt} = \frac{1}{4\pi\theta_0 R^2}(\alpha x + syz - \omega z^2 - ez - \mu n) \tag{11}$$

where $\theta_0 = x_0 + y_0 + z_0 + n_0$ is the sum of the initial conditions of x, y, z, n.

5 Sensitivity Analysis

Sensitivity analysis is the systematic investigation of a system to small perturbations of its parameters. Here, we employ a local method of sensitivity analysis. Each parameter is varied one at a time by a small amount around some fixed point, and the effect of individual perturbations on the solutions is calculated. In particular, we use the so-called direct differential method. In this method, suppose that a mathematical model assumes the form $\dot{y}_i = f_i$, where $f_i = f_i(y_i, p_j)$ with parameters p_j, then we can analyze the rate of change of y_i with respect to p_j via the equations

$$\frac{d}{dt}\left(\frac{\partial y_i}{\partial p_j}\right) = \frac{\partial f_i}{\partial y_i}\frac{\partial y_i}{\partial p_j} + \frac{\partial f_i}{\partial p_j} = J \cdot S_j + F_j, \tag{12}$$

where matrix J, the Jacobian, is given by

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \frac{\partial f_1}{\partial y_2} & \cdots & \frac{\partial f_1}{\partial y_k} \\ \frac{\partial f_2}{\partial y_1} & \frac{\partial f_2}{\partial y_2} & \cdots & \frac{\partial f_2}{\partial y_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_k}{\partial y_1} & \frac{\partial f_k}{\partial y_2} & \cdots & \frac{\partial f_k}{\partial y_k} \end{pmatrix},$$

vector S_j, the sensitivity vector for parameter p_j, and vector F_j, the derivative of the right-hand side of the system with respect to the parameter p_j, are given by the

vectors: $S_j = \left(\frac{\partial y_i}{\partial p_j}\right) = \begin{pmatrix} \frac{\partial y_1}{\partial p_j} \\ \frac{\partial y_2}{\partial p_j} \\ \cdots \\ \frac{\partial y_k}{\partial p_j} \end{pmatrix}$, $F_j = \begin{pmatrix} \frac{\partial f_1}{\partial p_j} \\ \frac{\partial f_2}{\partial p_j} \\ \cdots \\ \frac{\partial f_k}{\partial p_j} \end{pmatrix}$. Looking at our system (1), we have

five dependent variables and 11 parameters; hence, $k = 5$ and $j = 11$. In particular, we have our Jacobian J (computed in a previous section)

$$J = \begin{pmatrix} \alpha - \beta v & 0 & 0 & 0 & -\beta x \\ \beta v & -kz - \delta & -ky & 0 & \beta x \\ 0 & sz & sy - 2\omega z - e & 0 & 0 \\ 0 & kz + \delta & ky & -\mu & 0 \\ 0 & b\delta & -k_0 v & 0 & -k_0 z - \gamma \end{pmatrix}. \tag{13}$$

Let us arrange all parameters as the vector $p = (\alpha \ \ \beta \ \ k \ \ \delta \ \ s \ \ \omega \ \ e \ \ \mu \ \ b \ \ k_0 \ \ \gamma)$. The full sensitivity matrix S and the full derivative of the right-hand side with respect to the parameters F are given correspondingly by

$$S = \begin{pmatrix} \frac{\partial x}{\partial \alpha} & \frac{\partial x}{\partial \beta} & \frac{\partial x}{\partial k} & \cdots & \frac{\partial x}{\partial \gamma} \\ \frac{\partial y}{\partial \alpha} & \frac{\partial y}{\partial \beta} & \frac{\partial x}{\partial k} & \cdots & \frac{\partial y}{\partial \gamma} \\ \frac{\partial z}{\partial \alpha} & \frac{\partial z}{\partial \beta} & \frac{\partial x}{\partial k} & \cdots & \frac{\partial z}{\partial \gamma} \\ \frac{\partial n}{\partial \alpha} & \frac{\partial n}{\partial \beta} & \frac{\partial x}{\partial k} & \cdots & \frac{\partial n}{\partial \gamma} \\ \frac{\partial v}{\partial \alpha} & \frac{\partial v}{\partial \beta} & \frac{\partial x}{\partial k} & \cdots & \frac{\partial v}{\partial \gamma} \end{pmatrix}, \quad F = \begin{pmatrix} x & -xv & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & xv & -yz & -y & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & yz & -z^2 & -z & 0 & 0 & 0 & 0 \\ 0 & 0 & yz & y & 0 & 0 & 0 & -n & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \delta y & -vz & -v \end{pmatrix}. \quad (14)$$

Hence, for example, the system of differential equations that investigates the rate of change of the dependent variables with respect to the parameter α is given by ten differential equations:

$$\begin{cases} \frac{dx}{dt} = \alpha x - \beta xv \\ \frac{dy}{dt} = \beta xv - kyz - \delta y \\ \frac{dz}{dt} = syz - \omega z^2 - ez \\ \frac{dn}{dt} = kyz + \delta y - \mu n \\ \frac{dv}{dt} = b\delta y - k_0 vz - \gamma v \\ \frac{ds_{11}}{dt} = (\alpha - \beta v)s_{11} - \beta xs_{51} + x \\ \frac{ds_{21}}{dt} = \beta vs_{11} - (kz + \delta)s_{21} - kys_{31} + \beta xs_{51} \\ \frac{ds_{31}}{dt} = szs_{21} + (sy - 2\omega z - e)s_{31} \\ \frac{ds_{41}}{dt} = (kz + \delta)s_{21} + kys_{31} - \mu s_{41} \\ \frac{ds_{51}}{dt} = b\delta s_{21} - k_0 vs_{31} - (\gamma + k_0 z)s_{51}. \end{cases}$$

Observe that the first five differential equations are exactly the differential equations in the system while the bottom five come from (12) by using information from (13) and the first column of F from (14).

Since different parameters can have different units and thus can have different orders of magnitude difference in size, we calculate the relative sensitivity $\bar{S}_{ij} = \frac{\partial y_i}{\partial p_j} \frac{p_j}{y_i}$. Then the sensitivity of each parameter p_j is given by the sensitivity index defined as the magnitude of the corresponding norm of the i, j-th element in $\bar{S}$:

$$\| \bar{S}_{ij} \|_2 = \sqrt{\sum_{k=1}^{N} \left(\frac{\partial y_i}{\partial p_j}(t_k) \frac{p_j}{y_i(t_k)} \right)^2}.$$

For a given parameter set, we computed the relative sensitivity matrix which is given in Table 1. The cancer is most sensitive to the death rate parameter δ of the infected cancer cells, the bursting size b, and the death rate parameter γ of the virus. Infected cancer cells are also most sensitive to perturbations in their death rate parameter δ, and second most sensitive to perturbations in the cancer growth rate parameter α as well as in the infection rate parameter β. The immune cells are most sensitive to the immune-suppression parameter e and to the immune stimulation rate s, both of which in general could be expected. The virus is most sensitive to the bursting size b as well as to the death rate parameter δ of the infected cancer cells and its death rate

Table 1 L_2 norms of the rate of change of each variable with respect to the parameters, including the maximum value for each variable and each parameter. The Table also includes the second and the third most sensitive parameters for each of the variables

p	$\left\|\frac{\partial x}{\partial p}\frac{p}{x}\right\|_2$	$\left\|\frac{\partial y}{\partial p}\frac{p}{y}\right\|_2$	$\left\|\frac{\partial z}{\partial p}\frac{p}{z}\right\|_2$	$\left\|\frac{\partial n}{\partial p}\frac{p}{n}\right\|_2$	$\left\|\frac{\partial v}{\partial p}\frac{p}{v}\right\|_2$	$\times$	max
α	3.8148	0.0136	0.1945	0.0161	0.0136	10^3	x
β	3.7848	0.0131	0.1213	0.0099	0.0091	10^3	x
k	1.1340	0.0068	0.0481	0.0024	0.0036	10^2	x
δ	5.3606	0.0565	0.2301	0.0107	0.0185	10^3	x
s	1.4652	0.0058	0.2930	0.0016	0.0044	10^3	x
ω	0.5609	0.0014	0.0698	0.0003	0.001	10^{-3}	x
e	1.4485	0.0056	0.5366	0.0014	0.0042	10^3	x
μ	0.0000	0.0000	0.0000	0.1528	0.0000	10^2	n
b	5.2434	0.0081	0.0787	0.0062	0.0192	10^3	x
k_0	5.8636	0.0005	0.0046	0.0004	0.0169	10^1	x
γ	4.2068	0.0004	0.0036	0.0003	0.0172	10^3	x
max	δ	δ	e	α	b		
2nd, 3rd	b, γ	α, β	s, δ	μ, δ	δ, γ		

parameter γ. And finally, the dead cells n are most sensitive to the cancer growth rate parameter α and to the clearance rate parameter μ. One interesting result from our sensitivity analysis is that the cancer cells turned out to be most sensitive to the parameter δ even though that parameter does not participate directly in the right-hand side of the equation for the rate of change of the uninfected cancer cells x.

6   Numerical Simulations

In this section, we utilize the results from sensitivity analysis of the system to explore the dynamics when the most sensitive parameters are varied to gain more insight into those critical parameters that may lead to better efficacy of the virotherapy treatment. As the cancer cells are most sensitive to the death rate of the infected cancer cells δ, we start by analyzing the model's outputs when it is varied on the interval $1/48 < \delta < 1/18$ [7, 10]. All parameters used in the simulations are given in Table 2. Figure 1 shows the cells' population over time when a constant immune-suppression $e(t) = 0.5$ is applied. Increasing the δ supports the immune-suppression drug's control of the tumor's growth. Even though the tumor's radius does not decrease significantly, the virotherapy works by suppressing tumor's growth and keeping it under its initial

Table 2 The parameter values used in the simulations

Parameter	Description	Value	Units	References
α	Tumor growth rate	$2 \times 1/10$	day^{-1}	[7]
β	Infection rate of the virus	7×10^{-8}	$\text{virions}^{-1}\,\text{hr}^{-1}$	[7]
k	Immune killing rate of infected tumor cells	2×10^{-8}	$\text{mm}^3\,\text{hr}^{-1}\,\text{cell}^{-1}$	[7]
δ	Death rate of the infected tumor cells	$1/18$	$1/h$	[7, 10]
s	Stimulation rate of immune cells by infected cells	5.6×10^{-7}	hr^{-1}	[7]
ω	Death rate of the immune cells	2×10^{-12}	hr^{-1}	[7]
e	Immune-suppressive drug	0.5	hr^{-1}	[7]
μ	Clearing rate of the dead cells	$1/48$	hr^{-1}	[7]
b	Burst size of free virus	200	viruses cell^{-1}	[7, 11]
k_0	Immune killing rate of the virus	10^{-8}	$\text{mm}^3\,\text{hr}^{-1}\text{cell}^{-1}$	[7]
γ	Clearance rate of viruses	2.5×10^{-2}	$\text{virions}^{-1}\,\text{hr}^{-1}$	[7, 12]

size. Interestingly, as seen in Fig. 2 when δ is at its lowest value, the exponentially decaying suppression is not sufficient to control the radius' growth. However, as δ increases, the same exponential decay suppression is able to control the tumor's growth and the radius even slightly decreases in time. With the periodic application of the immune-suppression drug, as can be seen in Fig. 3, the tumor could not be taken under control and on average its radius increases even with the highest death rate of the infected tumor cells ($\delta = 1/2$), which is outside of the natural range of this parameter.

Figure 4 shows the dynamics of the state variables over time when the bursting size b of the viruses released after the lysing of the infected cancer cells was varied over its range of variability. Starting from $b = 100$ [7] and doubling the size up to $b = 800$ [11], we can see that in the case with no immune-suppression– which is illustrated with the solid lines– as higher b is, as lower the tumor's radius gets. However, the radius stabilizes at a value that is higher than its original size. When a strong constant suppression is applied, illustrated with the dashed lines, with the increasing parameter b the viruses increase as well and the radius of the tumor decreases to a lower level than its original size. However, even the strong immune-suppression and the highest bursting size are not sufficient to eradicate the tumor. The periodic type of immune-suppression with varying bursting size b is shown in Fig. 5. As the bursting size increases, the tumor's radius decreases on average; however, it varies at a level higher than its original size and hence is not sufficient to suppress tumor growth.

The next most sensitive parameter is the death rate of the virus γ. When it was varied within the range [0.0025, 0.25] shown with solid lines in Fig. 6, the exponentially decaying immune-suppression $e(t) = 0.2e^{-0.5t}$ was not sufficient to control the tumor growth when γ is at the upper end of its range. However, even a weak but constant immune-suppression $e = 0.2$ (see the dashed lines in Fig. 6) is sufficient to control cancer growth and the radius of the tumor decreases even below its initial value.
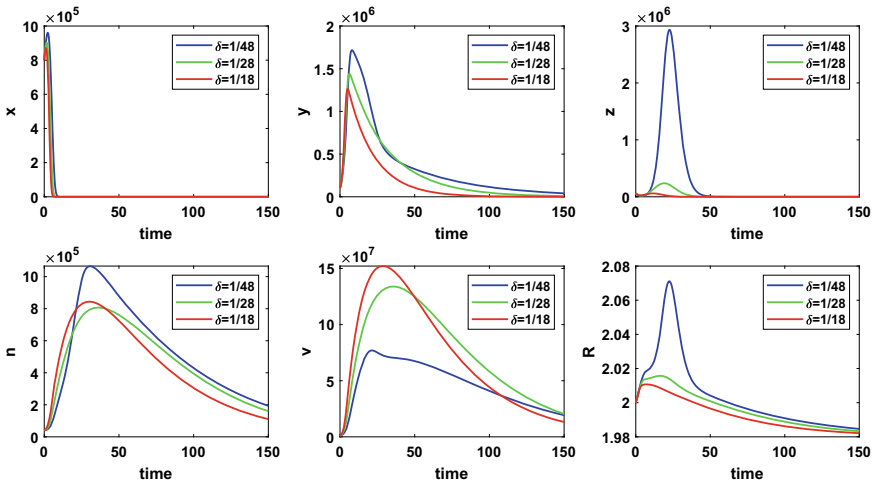
Fig. 1 The populations of cancer cells x, infected cancer cells y, immune cells z, dead cells n, viruses v, and the change of the radius R in time when the death rate of the infected tumor cells δ is varied over its range of variability and a constant immune-suppression $e(t) = 0.5$ is applied. $\delta = \frac{1}{48}$ is given in blue, $\delta = \frac{1}{28}$ is given in green, and $\delta = \frac{1}{18}$ is in red. All other parameters are given in Table 2
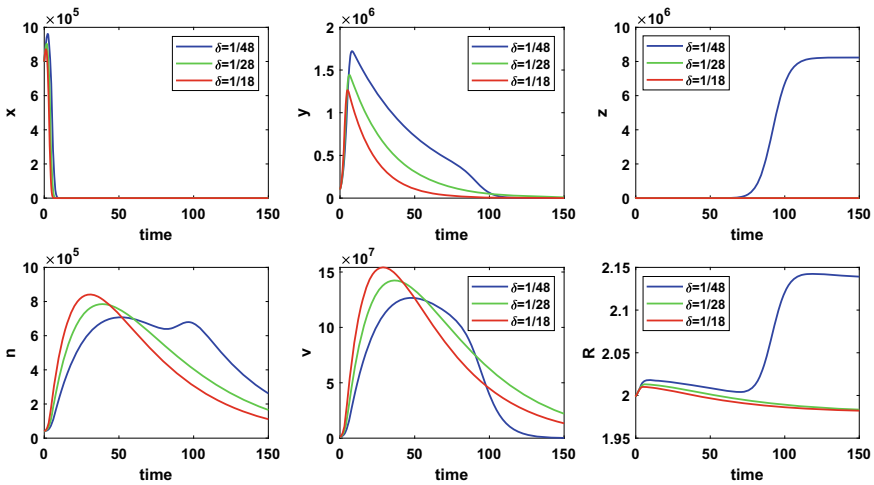
Fig. 2 The populations of cancer cells x, infected cancer cells y, immune cells z, dead cells n, viruses v, and the change of the radius R in time when the death rate of the infected tumor cells δ is varied over its range of variability and exponentially decaying immune-suppression $e(t) = 0.2e^{-0.5t}$ is applied. $\delta = \frac{1}{48}$ is given in blue, $\delta = \frac{1}{28}$ is given in green, and $\delta = \frac{1}{18}$ is in red. All other parameters are given in Table 2
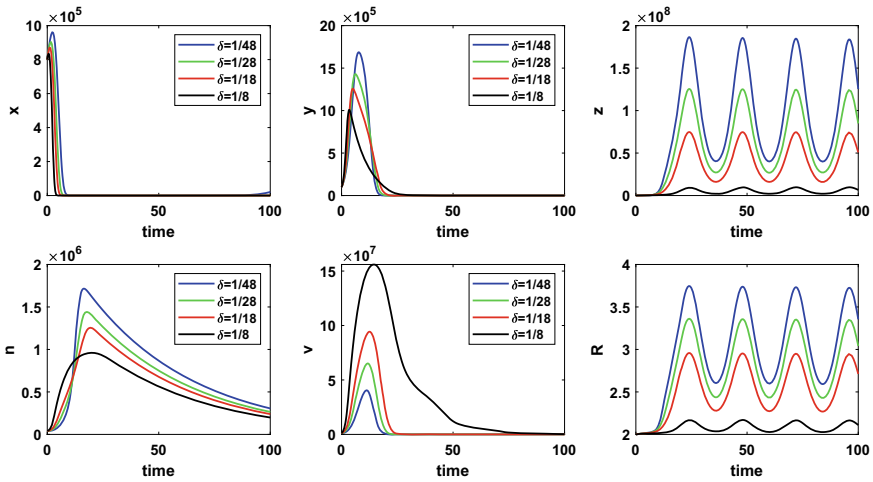
Fig. 3 The populations of cancer cells x, infected cancer cells y, immune cells z, dead cells n, viruses v, and the change of the radius R in time when the death rate of the infected tumor cells δ is varied over its range of variability and a periodic immune-suppression $e(t) = 0.2\sin(\pi/12t)$ is applied. $\delta = \frac{1}{48}$ is given in blue, $\delta = \frac{1}{28}$ is given in green, $\delta = \frac{1}{18}$ is in red, and $\delta = \frac{1}{8}$ in black. All other parameters are given in Table 2
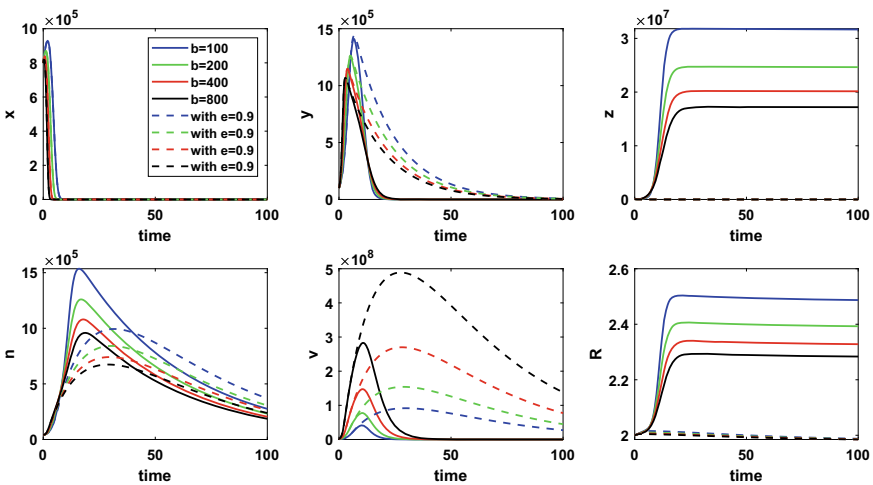
Fig. 4 The populations of cancer cells x, infected cancer cells y, immune cells z, dead cells n, viruses v, and the change of the radius R in time when the bursting size b is varied. Two cases were considered: without immune-suppression and with strong constant immune-suppression $e(t) = 0.9$ applied (given by the dashed lines). $b = 100$ is given in blue, $b = 200$ is given in green, $b = 400$ is in red, and $b = 800$ in black. All other parameters are given in Table 2
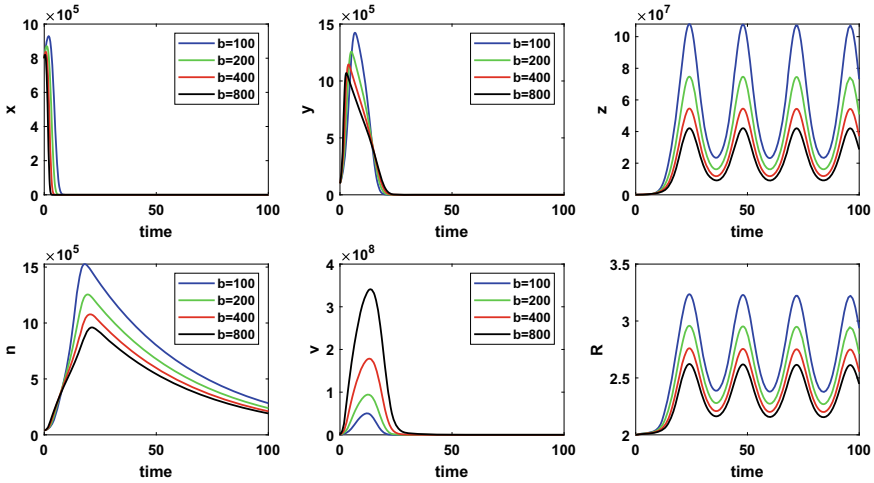
Fig. 5 The populations of cancer cells x, infected cancer cells y, immune cells z, dead cells n, viruses v, and the change of the radius R in time when the bursting size b is varied and a periodic immune-suppression $e(t) = 0.2 \sin(\pi/12t)$ is applied. $b = 100$ is given in blue, $b = 200$ is given in green, $b = 400$ is in red, and $b = 800$ in black. All other parameters are given in Table 2
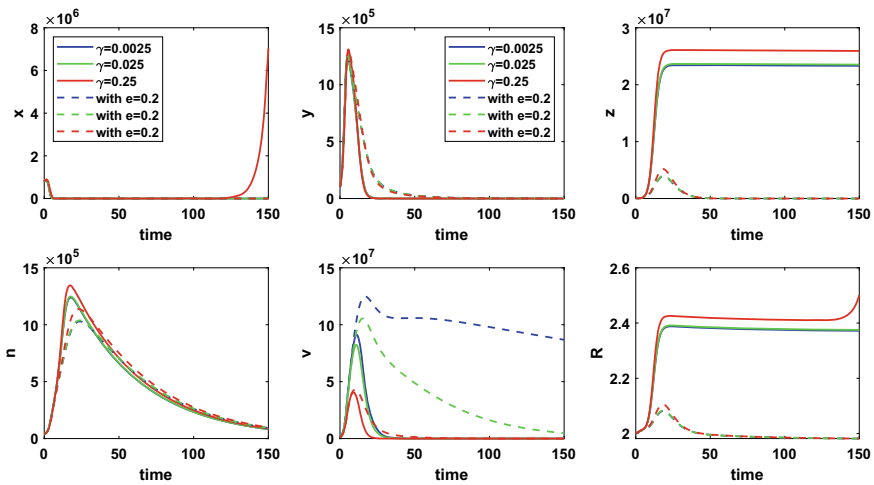
Fig. 6 The populations of cancer cells x, infected cancer cells y, immune cells z, dead cells n, viruses v, and the change of the radius R in time when the the death rate of the virus γ is varied over its range of variability. Two cases are considered, with exponentially decaying immune-suppression $e(t) = 0.2e^{-0.5t}$, given in solid lines, and with a weak constant immune-suppression $e(t) = 0.2$, given by the dashed lines. $\gamma = 0.0025$ is given in blue, $\gamma = 0.025$ is given in green, and $\gamma = 0.25$ is in red. All other parameters are given in Table 2

7 Conclusions

With this study, we have shown that mathematics can be a powerful tool in furthering biological understanding of virotherapy as a possible cancer treatment. In this work, we explored a mathematical model of virotherapy with the immune response which is the ODE version of the PDE model introduced in [7]. We extended the model to include dead cells, which allowed us to use conservation of mass to derive the equation for the rate of change of the tumor's radius. We use that equation to gauge the efficacy of the therapy. Using mathematical modeling, we studied the complex interplay among tumor cells, oncolytic virus particles, and immune response.

Qualitative analysis found four equilibria, namely a trivial equilibrium, a disease-free equilibrium, an equilibrium containing no immune cells, and a coexistence equilibrium. The conditions for the stability of each equilibrium were derived.

Studying the sensitivity analysis of the model is very important as it allows us to identify the most sensitive parameters of the model which in turn indicate the vulnerability of the model pathways. This model was most sensitive to the death rate δ of the infected cancer cells, the bursting size b, and the death rate γ of the virus. The state variables are also sensitive to perturbations in the cancer growth rate α and the infection rate β. The immune cells dynamics is mostly affected by the immune-suppression parameter e and the immune stimulation rate s. As the virus is most sensitive to the bursting size b, this means that the oncolytic viruses must be chosen so that they have as large as possible lysis production.

Using numerical simulations, we showed that immune system suppression is necessary for virotherapy to be effective. Additionally, simulations also show that even though the cancer cells decrease significantly with immune control, the tumor's radius decreases only to a limiting value and the tumor does not get eradicated. Among the three different types of immune-suppression delivery, the constant and exponential decay showed to be the most effective in controlling the tumor growth. Periodic type of immune-suppression may control the tumor growth only with a very high death rate δ of the infected cancer cells. Hence, to increase the efficacy of virotherapy, it needs to be supplemented with another treatment. These results can guide medical professionals in exploring the choice of additional therapeutic interventions to try in combination with virotherapy to eradicate the tumor.

References

1. World Health Organization. 2022. *"Cancer."* Retrieved February 19, 2022, from www.who.int/news-room/fact-sheets/detail/cancer

2. American Society of Clinical Oncology (ASCO). (n.d.). *Cancer terms.* Cancer.Net. Retrieved February 19, 2022, from https://www.cancer.net/navigating-cancer-care/cancer-basics/

3. Bell, J. C. (Ed.). 2020. *How Oncolytic Virus Therapy is Changing Cancer Treatment.* Cancer Research Institute. Retrieved February 19, 2022, from https://www.cancerresearch.org/immunotherapy/treatment-types/oncolytic-virus-therapy

4. Johns Hopkins Medicine. (n.d.). *The immune system.* Johns Hopkins Medicine. Retrieved February 19, 2022, from https://www.hopkinsmedicine.org/health/conditions-and-diseases/the-immune-system

5. Wang ZZ, Guo ZM, Smith H. A mathematical model of oncolytic virotherapy with time delay. Math Biosci Eng. 2019 Mar 6;16(4):1836–1860. https://doi.org/10.3934/mbe.2019089.PMID: 31137188.

6. Al-Tuwairqi, S.M., Al-Johani, N.O. and Simbawa, 2020. E.A. Modeling dynamics of cancer virotherapy with immune response. Adv Differ Equ 2020, 438. https://doi.org/10.1186/s13662-020-02893-6

7. Friedman A, Tian JP, Fulci G, Chiocca EA, Wang J. Glioma virotherapy: effects of innate immune suppression and increased viral replication capacity. Cancer Res. 2006 Feb 15;66(4):2314–9. https://doi.org/10.1158/0008-5472.CAN-05-2661. PMID: 16489036.

8. Panayotova, I., Brucal-Hallare, M. (2022). 6-017-OncolyticViruses-ModelingScenario. SIMIODE, QUBES Educational Resources. https://doi.org/10.25334/6J2B-1E35

9. Huang, J. H., Zhang, S. N., Choi, K. J., Choi, I. K., Kim, J. H., Lee, M. G., Lee, M., Kim, H., & Yun, C. O. (2010). Therapeutic and tumor-specific immunity induced by combination of dendritic cells and oncolytic adenovirus expressing IL-12 and 4-1BBL. Molecular therapy: the journal of the American Society of Gene Therapy, 18(2), 264–274. https://doi.org/10.1038/mt.2009.205

10. Camara,B.I., Mokrani H., Afenya Evans K. Mathematical modeling of glioma therapy using oncolytic viruses[J]. Mathematical Biosciences and Engineering, 2013, 10(3): 565–578. https://doi.org/10.3934/mbe.2013.10.565

11. Urenda-Cazares E., Gallegos A., Macias-Dias J.E.(2020) A mathematical model that combines chemotherapy and oncolytic virotherapy as an alternative treatment against a glioma, Journal of Mathematical Chemistry 58:544–554 https://doi.org/10.1007/s10910-019-01084-3

12. Timalsina A, Tian JP, Wang J. Mathematical and Computational Modeling for Tumor Virotherapy with Mediated Immunity. Bull Math Biol. 2017 Aug;79(8):1736–1758. https://doi.org/10.1007/s11538-017-0304-3. Epub 2017 Jun 7. PMID: 28593497.

Numerical Analysis of Thermoregulation in Honey Bee Colonies in Winter Based on Sign-Changing Chemotactic Coefficient Model

Atanas Z. Atanasov, Miglena N. Koleva, and Lubin G. Vulkov

Abstract The present work is inspired by beekeeper practice and the recent developments in laboratory experiments made on honey bee colonies, where the thermoregulation in the hive is studied. We consider the modification of the Keller–Segel model, including chemotactic coefficient of which the sign can change as honey bees move to a preferred temperature, and mortality of the organisms as well. In this case, the known studies on Keller–Segel-type models are not applicable. We construct a second order in space positivity preserving for the bee density, difference scheme. This allows us to perform numerical parameter model calibration to real data, obtained from temperature measurements.

Keywords Keller–Segel model · Honeybee thermoregulation · Finite difference scheme · Flux limiter · Positivity preserving

1   Introduction

During winter, honey bee colonies have high mortality of individual bees and this results in the extinction of the colony before the next spring season [1, 3–6, 15]. Honey bee colonies are not dormant during the winter: they remain active and maintain the hive temperature. Honey bees form combs for brood and storage of honey and pollen and bees are found to thermoregulating cluster in between the combs with the highest temperature in the center [3, 12]. The core temperature maintained between 18°C and 32°C if the ambient temperature range of −15–10°C. The thermoregulation is a result of individual bees attempting to regulate their own body

A. Z. Atanasov · M. N. Koleva (✉) · L. G. Vulkov
Ruse University, 8 Studentska Street, 7017 Ruse, Bulgaria
e-mail: mkoleva@uni-ruse.bg

A. Z. Atanasov
e-mail: aatanasov@uni-ruse.bg

L. G. Vulkov
e-mail: lvalkov@uni-ruse.bg

temperatures. This enables them to survive long periods of cold. During the winter the colony is dependent on the survival of a long-lived cohort of bees that is produced in the autumn [3, 4, 15].

The most important factor for the survival of bees in winter is the generation and retention of heat. Bees produce heat through flight muscle activities [3, 7]. Below a certain temperature the bee starts shivering with her flight muscles, whereas above this temperature she remains at rest. Also, honey bees have a thermotactic movement which is based on temperature differences in their local neighborhood [3].

Variety mathematical models using non-linear ordinary differential equations have been proposed to predict and analyze the main factors in the honey bee colony dynamics under specific conditions, see [2, 5, 13] and reference therein. In [15], coupled thermotaxis-diffusion equation for the cluster density with a heat equation with a temperature- and bee-density-dependent source is used to describe the self-organized thermoregulation of honeybee clusters. In [3], a Keller–Segel model with a sign-changing chemotactic coefficient is used to model the honey bee colonies in winter. This model is the extension of the one in [15] by including the mortality of individual bees.

In this work, we develop and examine positivity-preserving numerical method for solving the problem, introduced in [3]. To this aim we apply flux-limiter approach and extend the method, proposed in [10] for more complicated problem with sign-changing chemotactic coefficient and discontinuous functions.

The remaining part of the paper is organized as follows. In the next section, we present the model problem. In Sect. 3, we develop the positivity-preserving numerical method. In Sect. 4, we present and discuss numerical results. We give the results from computations with real data, measured in winter of 2017 in the village of *Brestovitca*, *Bulgaria*. The paper is closed by concluding remarks.

2 Model Description

In [3], the thermotactic movement of honey bees is described by the generalized Keller–Segel model [9, 14]

$$\frac{\partial T}{\partial t} = \triangle T + f(T)\rho, \quad \frac{\partial \rho}{\partial t} = \nabla[\nabla \rho - \mathcal{X}(T)\rho \nabla T] - \theta(\rho, T)\rho.$$

Here $\rho \geq 0$ is the bee density and T is the local temperature, $f(T)\rho$ models the heat generation by bees, and $\mathcal{X}$ changes sign form positive to negative and generates very different dynamics compared to the models considered in the mathematical literature. Namely, when the local temperature T is too low, $T > T_{\mathcal{X}}$, bees move toward higher temperature; when $T < T_{\mathcal{X}}$, they move away to lower temperature. This means that the chemotactic coefficient $\mathcal{X}(T)$ changes sign at $T_{\mathcal{X}}$, and hence it can become negative, which is very different from the generalized Keller–Segel model, where $\mathcal{X}$ has a fixed positive sign.

In [3], the individual mortality bees, described by the function $\theta(\rho, T) > 0$, is incorporated and the functions f and $\mathcal{X}$ are simplified, based on those in [15]. Similarly, based on the data in [15] f is chosen to be a step function as well

$$f(T) = \begin{cases} f_{\text{low}}, & T < T_f, \\ f_{\text{high}}, & T > T_f, \end{cases} \qquad \mathcal{X}(T) = \begin{cases} +\chi_1, & T < T_{\mathcal{X}}, \\ -\chi_2, & T > T_{\mathcal{X}}, \end{cases} \tag{1}$$

where $f_{\text{low}}, f_{\text{high}}, \chi_1, \chi_2 > 0$, $T_f < T_{\mathcal{X}}$. The temperature $T_{\mathcal{X}}$ can thought as the preferred temperature for the bees, as bees prefer to move toward locations with this temperature.

Further, in [3] is taken a cross section of this cluster from the center to the edge, and study the cluster in the spatial dimension $[0, L]$ where $x = 0$ is the center of the colony and $x = L$ is the edge of it. To construct the set of four boundary conditions, the following arguments are applied. Bees do not leave the colony in winter, and temperature at the edge of the colony at $x = L$ is equal to the ambient temperature T_a, which we assume to be fixed and to below the preferred temperature $T_{\mathcal{X}}$ in winter and because of the assumed symmetry boundary conditions both for T and ρ at the center of colony are imposed. Finally, for $x \in [0, L], t \in [0, t_f]$, the model becomes

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial x^2} + f(T)\rho, \quad \frac{\partial \rho}{\partial t} = \frac{\partial^2 \rho}{\partial x^2} - \frac{\partial}{\partial x}\left(\mathcal{X}(T)\rho\frac{\partial T}{\partial x}\right) - \theta(\rho, T)\rho, \tag{2}$$

$$\frac{\partial T}{\partial x}(0, t) = 0; \quad T(L, t) = T_a < T_{\mathcal{X}}, \tag{3}$$

$$\frac{\partial \rho}{\partial x}(0, t) = 0; \quad \left(\frac{\partial \rho}{\partial x} - \mathcal{X}(T)\rho\frac{\partial T}{\partial x}\right)(L, t) = 0, \tag{4}$$

$$\rho(x, 0) = \rho^0(x), \quad T(x, 0) = T^0(x). \tag{5}$$

The mortality function is constructed as a product of the following distinctive effects: **(i)** the effect of the local temperature (θ_T); **(ii)** the effective refresh rates of heat generating bees (θ_D); **(iii)** the effect of parasitic mites in the colony (θ_M)

$$\theta(T, \rho) = \theta_0\theta_T(T)\theta_D(\rho)\theta_M(\rho), \tag{6}$$

$$\theta_T(T) = \begin{cases} 1, & T < T_\theta, \\ 0, & T \geq T_\theta, \end{cases} \quad \theta_D(\rho) = \frac{\rho}{(\rho_{\text{tot}})^\gamma}, \quad \theta_M(\rho) = 1 + \frac{m}{\rho_{\text{tot}}}, \quad \rho_{\text{tot}} = \int_0^L \rho(x)dx,$$

where θ_0 is a constant that must fit to the observations, $\gamma > 0$ is some unknown exponent, $T_\theta > T_a$, and m is the amount of mites in the colony.

In [3], authors study the model (2)–(5). First, they analyze system (2) in the absence of bee mortality, $\theta = 0$ and find two types of steady-state configurations, *type I*: the temperature stays below $T_{\mathcal{X}}$ in the whole colony, i.e., $T(x) < T_{\mathcal{X}}$ for all $x \in [0, L]$ and *type II*: the temperature is larger than $T_{\mathcal{X}}$ at $x = 0$, and hence there exists a point x^* such that $T(x) > T_{\mathcal{X}}$ for $x < x^*$ and $T(x) < T_{\mathcal{X}}$ for $x > x^*$, distinguishable by the colony size ρ_{tot}.

3 Numerical Method

In this section, we develop bee density positivity-preserving numerical method for solving problem (1)–(6). We unfold the flux-limiter approach [10].

Since the winter period is very large, we consider the rescaled model. Let $\overline{x} = x/x_r, \overline{t} = t/t_r, \overline{T} = T/T_r, \overline{\rho} = \rho/\rho_r, \overline{\theta}_0 = \theta_0/\theta_{0,r}, \overline{\theta} = \theta/\theta_r, \overline{f} = f/f_r, \overline{\mathcal{X}} = \mathcal{X}/\mathcal{X}_r$. Then, if $x_r = L, t_r = L^2, \theta_r = \mathcal{X}_r = T_r = 1, \rho_r = 1/L, \theta_{0,r} = 1/L, \ f_r = 1/L$ and keeping the notation x, t, T, ρ instead of $\overline{x}, \overline{t}, \overline{T}, \overline{\rho}, \overline{\theta}_0, \overline{\theta}, \overline{f}, \overline{\mathcal{X}}$, the rescaled problem is (2)–(5), $x \in [0, 1], t \in [0, t_f/L^2]$ with

$$f(T) = L \begin{cases} f_{\text{low}}, \ T < T_f, \\ f_{\text{high}}, \ T > T_f \end{cases}, \quad \theta_0 := L\theta_0, \ \rho_{\text{tot}} = \int\limits_0^1 \rho(x, t)dx. \quad (7)$$

First we construct the spatial semidiscretization. Consider uniform meshes

$$\overline{\omega}_h = \{x_i : x_i = ih, \ i = 0, 1, \dots, N, \ x_N = L\},$$
$$\mathring{\omega}_h = \{x_{i+1/2} = x_i + h/2, \ i = 0, 1, \dots, N, \ x_{-1/2} = x_0, \ x_{N+1/2} = x_N\}$$

and denote by $u_i(t)$ and $u_{i+1/2}(t)$ the mesh function u at point (x_i, t) and $(x_{i+1/2}, t)$, respectively. We will also use the notations

$$u^{\pm} = \max\{0, \pm u\}, \ u_{x,i+1/2} = \frac{u_{i+1} - u_i}{h}, \ u_{\overline{x}x,i} = \frac{u_{x,i+1/2} - u_{x,i-1/2}}{h}.$$

The first equation in (2),(7) and the corresponding boundary conditions (3),(7) are approximated by the standard second-order finite difference scheme

$$\begin{aligned}
\frac{dT_i}{dt} &= T_{\overline{x}x,i} + f(T_i)\rho_i, \ i = 1, 2, \dots, N - 1, \\
\frac{dT_0}{dt} &= \frac{2}{h}T_{x,1/2} + f(T_0)\rho_0, \ T_N = T_a.
\end{aligned} \quad (8)$$

Consider the second equation in (2),(7). We apply finite volume method. Integrating over the intervals $[x_{i-1/2}, x_{i-1/2}], i = 0, 1, \dots, N$, we derive

$$\int\limits_{x_{i-1/2}}^{x_{i-1/2}} \frac{\partial \rho}{\partial t}dx = \left(\frac{\partial \rho}{\partial x} - \mathcal{X}(T)\rho\frac{\partial T}{\partial x}\right)\Bigg|_{x_{i-1/2}}^{x_{i-1/2}} - \int\limits_{x_{i-1/2}}^{x_{i-1/2}} \theta(\rho, T)\rho dx. \quad (9)$$

Denote by $F_{i+1/2} = F_{i+1/2}(\rho_i, T_i)$ the approximation of the flux $\mathcal{X}(T)\rho\frac{\partial T}{\partial x}$ at point $(x_{i+1/2}, t)$. Approximating the integrals in (9) by mid-point rule, we get

$$\frac{d\rho_i}{dt} = \rho_{\overline{x}x,i} - \frac{F_{i+1/2} - F_{i-1/2}}{h} - \theta(\rho_i, T_i)\rho_i, \quad i = 1, 2, \ldots, N-1, \quad (10)$$

$$\frac{d\rho_N}{dt} + 2\frac{\rho_N - \rho_{N-1}}{h^2} + \theta(\rho_N, T_N)\rho_N = \frac{2F_{N-1/2}}{h}, \quad (11)$$

$$\frac{d\rho_0}{dt} + 2\frac{\rho_0 - \rho_1}{h^2} + \theta(\rho_0, T_0)\rho_0 = -\frac{2F_{1/2}}{h}, \quad (12)$$

where ρ_{tot} is discretized by the trapezoidal rule quadrature and for the approximation of (9) at boundary nodes we use (4),(7).

For the terms $F_{i\pm1/2}$, we use van Leer flux-limiter approach, see, e.g., [8]

$$\Phi(r) = \frac{|r| + r}{1 + |r|}, \quad r_{i+1/2} = \frac{\rho_{i+1} - \rho_i}{\rho_i - \rho_{i-1}}, \quad (13)$$

and the technique presented in [10]. Observe that $\Phi(r)$ is Lipschitz continuous, continuously differentiable for all $r \neq 0$, and

$$\Phi(r) = 0, \text{ if } r \leq 0 \text{ and } \Phi(r) \leq 2\min(1, r).$$

Let $\mathcal{X}_{i+1/2} := \mathcal{X}(T_{i+1/2})T_{x,i+1/2}$, $\mathcal{X}(T_{i+1/2}) = 0.5(\mathcal{X}(T_{i+1}) + \mathcal{X}(T_i))$. The discrete flux is expressed in two forms, depending on the sign of $\mathcal{X}(T)\frac{\partial T}{\partial x}$ [8, 10]

$$F_{i+1/2} = \begin{cases} \mathcal{X}_{i+1/2}\left(\rho_i + \frac{1}{2}\Phi(r_{i+1/2})(\rho_i - \rho_{i-1})\right), & \mathcal{X}_{i+1/2} \geq 0 \\ \mathcal{X}_{i+1/2}\left(\rho_{i+1} + \frac{1}{2}\Phi(r_{i+3/2}^{-1})(\rho_{i+1} - \rho_{i+2})\right), & \mathcal{X}_{i+1/2} < 0. \end{cases} \quad (14)$$

Further, due to the symmetry property of the flux limiter $\Phi(r) = r\Phi(r^{-1})$ and in view of (13) we derive an equivalent form of $F_{i+1/2}$

$$\begin{aligned} F_{i+1/2} &= \mathcal{X}_{i+1/2}\left(\rho_{i+1} + \frac{1}{2}\Phi(r_{i+3/2}^{-1})(\rho_{i+1} - \rho_{i+2})\right) \\ &= \mathcal{X}_{i+1/2}\left(\rho_{i+1} + \frac{1}{2}\Phi(r_{i+3/2}^{-1})\frac{\rho_{i+1} - \rho_{i+2}}{\rho_{i+1} - \rho_i}(\rho_{i+1} - \rho_i)\right) \\ &= \mathcal{X}_{i+1/2}\left(\rho_i + \left(1 - \frac{1}{2}\Phi(r_{i+3/2})\right)(\rho_{i+1} - \rho_i)\right). \end{aligned} \quad (15)$$

Similarly, the flux $F_{i-1/2}$ is defined by shifting the index i and applying the symmetry property of the flux limiter

$$F_{i-1/2} = \begin{cases} \mathcal{X}_{i-1/2}\left(\rho_i + \left(1 - \frac{1}{2}\Phi(r_{i-1/2}^{-1})\right)(\rho_{i-1} - \rho_i)\right), & \mathcal{X}_{i-1/2} \geq 0, \\ \mathcal{X}_{i-1/2}\left(\rho_i + \frac{1}{2}\Phi(r_{i+1/2}^{-1})(\rho_i - \rho_{i+1})\right), & \mathcal{X}_{i-1/2} < 0. \end{cases} \quad (16)$$

Note that, for example, $F_{N-1/2}$ involves the values $\Phi(r_{N+1/2})$ and for $i = 0$ in (16), ρ_{-1} must be determined. For the value of ρ_i at the outer grid nodes x_{N+1} and x_{-1} a second-order extrapolation formula [11] can be used or second-order finite difference discretization of the boundary conditions (4)

$$\rho_{N+1} = 3\rho_N - 3\rho_{N-1} + \rho_{N-2}, \tag{17}$$

$$\rho_{-1} = \rho_1, \ \rho_{N+1} = \rho_{N-1} - 2h\mathcal{X}(T_N)\rho_N T_{\bar{x},N}^<, \ T_{\bar{x},N}^< = \frac{T_N - 3T_{N-1} + 2T_{N-2}}{h}. \tag{18}$$

From (10) and in view of (16), (15), we get

$$\frac{d\rho_i}{dt} = \rho_{\bar{x}x,i} - \frac{F_{i+1/2} - F_{i-1/2}}{h} - \theta(\rho_i, T_i)\rho_i, \ i = 1, 2, \dots, N-1, \tag{19}$$

$$F_{i+1/2} - F_{i-1/2} = \left[\mathcal{X}_{i+1/2}\right]^+ \left(\rho_i + \frac{1}{2}\Phi(r_{i+1/2})(\rho_i - \rho_{i-1})\right)$$
$$- \left[\mathcal{X}_{i+1/2}\right]^- \left(\rho_i + \left(1 - \frac{1}{2}\Phi(r_{i+3/2})\right)(\rho_{i+1} - \rho_i)\right)$$
$$- \left[\mathcal{X}_{i-1/2}\right]^+ \left(\rho_i + \left(1 - \frac{1}{2}\Phi(r_{i-1/2}^{-1})\right)(\rho_{i-1} - \rho_i)\right)$$
$$+ \left[\mathcal{X}_{i-1/2}\right]^- \left(\rho_i + \frac{1}{2}\Phi(r_{i+1/2}^{-1})(\rho_i - \rho_{i+1})\right).$$

To compute the slope $r_{1/2}$ we use (18), while for $r_{N+1/2}$, the extrapolation formula (17) is applied.

From (11), applying (16), (15), we get

$$\frac{d\rho_N}{dt} + 2\frac{\rho_N - \rho_{N-1}}{h^2} + \theta(\rho_N, T_N)\rho_N = \frac{2F_{N-1/2}}{h}, \tag{20}$$

$$F_{N-1/2} = \left[\mathcal{X}_{N-1/2}\right]^+ \left(\rho_N + \left(1 - \frac{1}{2}\Phi(r_{N-1/2}^{-1})\right)(\rho_{N-1} - \rho_N)\right)$$
$$- \left[\mathcal{X}_{N-1/2}\right]^- \left(\rho_N + \frac{1}{2}\Phi(r_{N+1/2}^{-1})(\rho_N - \rho_{N+1})\right).$$

Here ρ_{N+1} is eliminated from (18). As before, for the slope $r_{N+1/2}^{-1}$ we use (17).

In the same fashion, from (12), (16), (15), we derive

$$\frac{d\rho_0}{dt} + 2\frac{\rho_0 - \rho_1}{h^2} + \theta(\rho_0, T_0)\rho_0 = -\frac{2F_{1/2}}{h}, \tag{21}$$

$$F_{1/2} = \left[\mathcal{X}_{1/2}\right]^+ \left(\rho_0 + \frac{1}{2}\Phi(r_{1/2})(\rho_0 - \rho_{-1})\right) - \left[\mathcal{X}_{1/2}\right]^- \left(\rho_0 + \left(1 - \frac{1}{2}\Phi(r_{3/2})\right)(\rho_1 - \rho_0)\right),$$

where quantity ρ_{-1} is eliminated from (18).

In the time interval $[0, t_f]$, we introduce the non-uniform temporal mesh $\bar{\omega}_\tau = \{t_n : t_0 = 0, \ t_n = t_{n-1} + \tau_n, \ n = 1, 2, \dots, M, \ t_M = t_f\}$ and denote by u_i^n the mesh function u at grid node point (x_i, t_n). Using implicit–explicit time stepping, the full discretization of (8) is

$$\frac{T_i^{n+1} - T_i^n}{\tau_n} - T_{\bar{x}x,i}^{n+1} = f(T_i^n)\rho_i^n, \quad i = 1, 2, \ldots, N-1,$$

$$\frac{T_0^{n+1} - T_0^n}{\tau_n} - \frac{2}{h}T_{x,1/2}^{n+1} = f(T_0^n)\rho_0^n, \quad T_N^{n+1} = T_a. \tag{22}$$

The ODEs (19), (20), (21) are approximated as follows:

$$\frac{\rho_i^{n+1} - \rho_i^n}{\tau_n} - \rho_{\bar{x}x,i}^{n+1} + \theta(\rho_i^n, T_i^{n+1})\rho_i^{n+1}$$

$$= -\frac{F_{i+1/2}(\rho_i^n, T_i^{n+1}) - F_{i-1/2}(\rho_i^n, T_i^{n+1})}{h}, \quad i = 1, 2, \ldots, N-1, \tag{23}$$

$$\frac{\rho_N^{n+1} - \rho_N^n}{\tau_n} + 2\frac{\rho_N^{n+1} - \rho_{N-1}^{n+1}}{h^2} + \theta(\rho_N^n, T_N^{n+1})\rho_N^{n+1} = \frac{2F_{N-1/2}(\rho_N^n, T_N^{n+1})}{h}, \tag{24}$$

$$\frac{\rho_0^{n+1} - \rho_0^n}{\tau_n} + 2\frac{\rho_0^{n+1} - \rho_1^{n+1}}{h^2} + \theta(\rho_0^n, T_0^{n+1})\rho_0^{n+1} = -\frac{2F_{1/2}(\rho_0^n, T_0^{n+1})}{h}. \tag{25}$$

We implement the scheme (22)–(25), such that to avoid iteration process. At each time level we perform the next two steps:

Decoupling Algorithm (DA)

(1) Solve the linear system (22) to find T^{n+1}.
(2) Solve the linear system (23)–(25) for known T^{n+1} and find ρ^{n+1}.

Theorem 1 *Let* $\|u\| = \max_{0 \le i \le N} |u_i|$ *and* $\rho^0 \ge 0$. *If* $\tau_n \le h\left(4 \left\| \mathcal{X}(T^{n+1})T_x^{n+1} \right\|\right)^{-1}$, *then at each time level, the solution* ρ^n *of* (22)–(25)*, realized by the two-step Decoupling Algorithm (DA), is non-negative.*

Proof *(outline)* The coefficient matrix of (23)–(25) is an M-matrix and since $0 \le 1 - \frac{1}{2}\left(\Phi(r_{i+1/2}^{\pm 1})\right) \le 1$, $1 \le 1 + \frac{1}{2}\left(\Phi(r_{i+1/2}^{\pm 1})\right) \le 2$, $\theta(\rho_i^n, T_i^{n+1}) > 0$, $i = 0, 1, \ldots, N$, conditions of the theorem guarantee the non-negativity of the right-hand side.

4 Numerical Simulations

In this section, we test the accuracy, the order of convergence, and the relevance of the proposed scheme (22)–(25), implemented by DA. *All data and results are given for the problem* (1)–(6) and the computations are performed with the rescaled problem (1)–(7). We consider the following set of parameters [3]: $T_\theta = 21$, $T_\mathcal{X} = 25$, $\chi_1 = \chi_2 = 1$, $f_{\text{high}} = 0.6$, $f_{\text{low}} = 3$, $T_f = 15$.

Example 1 (Convergence test) First, we test the algorithm DA for the problem with exact solution. We add residual function $f_R(x, t)$ in the right-hand side of Eqs. (2), such that $T(t, x) = T^0(x)e^{-t/(3t_f)}$, $\rho(t, x) = \rho^0(x)e^{-t/t_f}$ to be the exact solution of

Table 1 Errors and convergence rates, test with exact solution

	Non-smooth $\mathcal{X}(T)$				Smooth $\mathcal{X}(T)$			
N	$\mathcal{E}_T^N$	CR_T	$\mathcal{E}_\rho^N$	CR_ρ	$\mathcal{E}_T^N$	CR_T	$\mathcal{E}_\rho^N$	CR_ρ
80	1.92201e-2		2.46059e-3		1.50392e-2		1.89418e-3	
160	4.42138e-3	2.1200	5.73106e-4	2.1021	2.17345e-3	2.7906	3.18897e-4	2.5704
320	1.20971e-3	1.8698	1.54615e-4	1.8901	5.42347e-4	2.0026	7.98900e-5	1.9970
640	2.95805e-4	2.0319	3.79081e-5	2.0281	1.35910e-4	1.9966	1.99727e-5	2.0000
1280	7.36004e-5	2.0068	9.44061e-6	2.0055	3.39614e-5	2.0007	4.99473e-6	1.9996

(1)–(7). We take $T^0(x) = T_a + 40(1 - (x/L)^4)$ and for $\rho^0(x)$ we use steady-state solution, obtained in [3].

We test the convergence both for original (the derivative of the flux is computed in left and right sense in order to obtain $f_R(x, t)$) and smoothed function $\mathcal{X}(T) = \tanh[(T_\mathcal{X} - T)/\epsilon]$ [15] with $\gamma = 1, m = 10, \theta_0 = 4.10^{-3}$ [3]. Errors and convergence rates at time t_f are given by $\mathcal{E}_T^N = \max_{0 \leq i \leq N} |T(t_f, x_i) - T_i^M|$, $CR_T = \log_2(\mathcal{E}_T^N/\mathcal{E}_T^{2N})$, $\mathcal{E}_\rho^N = \mathcal{E}_T^N = \max_{0 \leq i \leq N} |\rho(t_f, x_i) - \rho_i^M|$, $CR_\rho = \log_2(\mathcal{E}_\rho^N/\mathcal{E}_\rho^{2N})$.

In Table 1, we present the results with $T_a = -9°C$, $L = 10$ cm at $t_f = 100$ min for uniform temporal mesh with step size τ and fixed ratio $\tau = h^2$ with non-smooth $\mathcal{X}(T)$ (only a simple averaging is used, see (16)) and smooth $\mathcal{X}(T)$, $\epsilon = 0.5$. We observe second order of spatial convergence in both cases. Further, our computations are performed without smoothing the function $\mathcal{X}(T)$.

In Fig. 1, we plot solution profiles of the numerical density and temperature on different scales at fixed time, obtained by DA for the original problem ($f_R(x, t) = 0$). Just as in [3], the two type solutions are clearly distinguished—*type I*: the temperature is less than $T_\mathcal{X}$ in the whole colony and *type II*: the temperature is larger than $T_\mathcal{X}$ at $x = 0$ and the distribution of bees has a peak at the location with preferred temperature $T_\mathcal{X}$.

Example 2 (Practical example) In practice the temperature in the beehive can be measured, but on a certain, not enough small for the numerical computations, distance in space. We apply the model (1)–(7) to compute the density, using the known values of the temperature and to turn into align the model parameters with the measurements, often called model calibration.

For illustration, we use the data, obtained in the winter of 2017 by the measuring system, placed in a hive, located in the village of *Brestovitca*, in the *Northeastern part of Bulgaria*. Geographical location of experimental apiary in Brestovica is $43°32'4.02''C$ N, $25°45'14.1''$ E and at an altitudinal range of 222 m. The experimental hive is a 12-frame Dadant-Blatt. Six temperature sensors—DS18B20—are placed at a distance of 5 cm from each other and the temperature is measured in intervals of 5, 10, or 15 min. We use the data from the first four sensors ($N_{\text{meas}} = 3$), which are in a line (Ox-axis) and located in the cluster. The first sensor is in the cluster core ($x = 0$). Instead of the ambient temperature T_a, we use the values of the
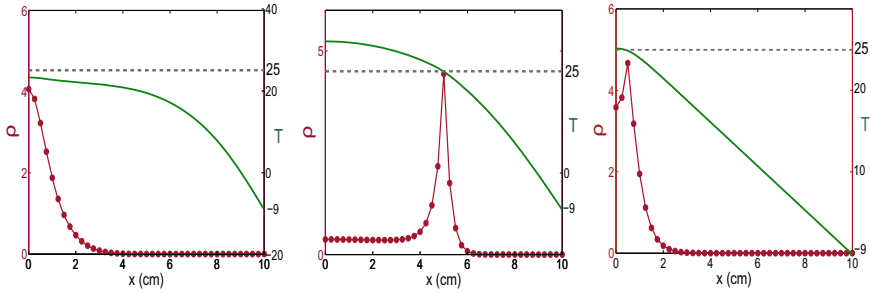
Fig. 1 Solution profiles of the numerical density and temperature, solution *type I* (*left*) and *type II* (*center and right*)

fourth sensor, situated close to the edge. For the particular experiment $L = 15$cm and for the simulations we use the measured temperature from 5.01.2017, 9:55 pm up to 7.01.2017, 1:20 pm, at 308 time points ($M_{\text{meas}} = 307$), $t_f = 2000$ min, see Fig. 2 (*left*). We denote by $T_{\text{meas}}(\widetilde{x}_i, \widetilde{t}_n)$, the measured values of the temperature at points $(\widetilde{x}_i, \widetilde{t}_n)$, $i = 0, 1, \ldots, N_{\text{meas}}$, $n = 0, 1, \ldots, M_{\text{meas}}$.

The values $T_{\text{meas}}(\widetilde{x}_i, \widetilde{t}_0)$ at the initial time are interpolated in order to fit the data to the spatial mesh $\overline{\omega}_h$. We find numerically by multiple runs, the optimal set of parameters $\theta_0 = 2.10^{-3}$, $\gamma = 1$, $m = 25$, such that the error $\mathcal{E}_{\text{meas}} = \|T(\widetilde{x}_i, \widetilde{t}_n) - T_{\text{meas}}(\widetilde{x}_i, \widetilde{t}_n)\| / \|T(\widetilde{x}_i, \widetilde{t}_n)\|$, $\|v\| = \max\limits_{0 \le i \le N_{\text{meas}}, \, 0 \le n \le M_{\text{meas}}} |v(\widetilde{x}_i, \widetilde{t}_n)|$, to be as small as possible. Here $T(\widetilde{x}_i, \widetilde{t}_n)$ is the solution, computed by DA and then interpolated on points $(\widetilde{x}_i, \widetilde{t}_n)$, $i = 0, 1, \ldots, N_{\text{meas}}$, $n = 0, 1, \ldots, M_{\text{meas}}$.

In Fig. 2 (*right*), we plot the computed temperature. The error $\epsilon_{\text{meas}} = |T(\widetilde{x}_i, \widetilde{t}_n) - T_{\text{meas}}(\widetilde{x}_i, \widetilde{t}_n)| / \|T(\widetilde{x}_i, \widetilde{t}_n)\|$ and the density are depicted in Fig. 3. More precise results are obtained at the colony's core. The two types of density distribution, depending on the values of the temperature, are clearly distinguishable. Although the measured temperature corresponds to the solution of *Type I*, while the
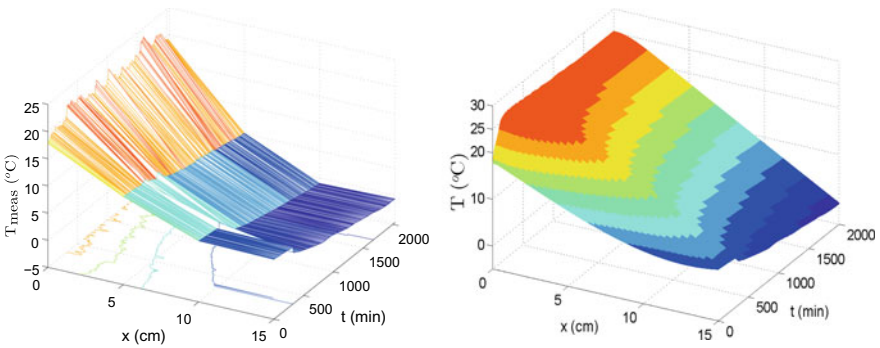
Fig. 2 Measured temperature (*left*) and computed temperature (*right*)
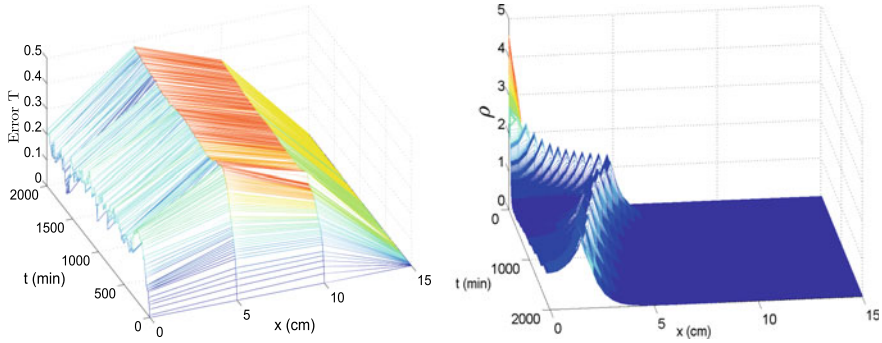
Fig. 3 The error ϵ_{meas} (*left*) of the temperature and the computed density (*right*)

computed temperature is referred to *Type 2* solution, the relative error is optimal and the fitting between computed and measured temperature is acceptable, taking into account that the considered model is simplified to 1D case with constant diffusion coefficient and parameters θ_0, γ, and m are not precisely determined.

5 Conclusions

In this paper, we constructed positivity preserving, second order in space numerical scheme for solving a Keller–Segel model with sign-changing chemotactic coefficient, describing the thermoregulation in honey bee colonies during the winter. The experiment with real data showed that the method can produce relevant results.

In our future work, we will make additional efforts in order to identify parameters θ_0, γ, and m. The engineers are able to use the temperature sensors, while the density observations are very difficult to perform. The temperature readings from the sensor at arbitrary specified times allow us to consider inverse problem of identifying $\rho = \rho(x, t)$ as an unknown reaction coefficient of the first equation in (2). Next, on the base of the second equation in (2), the inverse problem of numerical determination of θ_0, γ, and m will be solved.

References

1. Atanasov, A.Z.,. Georgiev, S.G., Vulkov, L.G.:Reconstruction analysis of honeybee colony collapse disorder modeling. Optim. Eng. 22, 2481–2503 (2021).
2. Bagheri, S., Mirzaie, M.: A mathematical model of honey bee colony dynamics to predict the effect of pollen on colony failure. PLoS ONE 14(11): e0225632 (2019).
3. Bastaansen, R., Doelman, A., van Langevede, F., Rottschafer, V: Modeling honey bee colonies in winter using a Keller-Segel model with a sign-changing chemotectic coefficient. SIAM J. Appl. Math. 80(20), 839–863 (2020).
4. Calovi, M., Grozinger, C.M., Miller, D.A., Goslee, S.C.: Summer weather conditions influence winter survival of honey bees (Apis mellifera) in the northeastern United States. Sci. Rep. 11, 1553 (2021).
5. Chen, J., DeGrandi-Hoffman, G., Ratti, V., Kang, Y.: Review on mathematical modeling of honeybee population dynamics. Math. Biosci. Eng. 18(6), 9606–9650 (2021).
6. van Dooremalen, C., Gerritsen, L., Cornelissen, B., van der Steen, J. JM, van Langevelde, F., Blacquiere, T.: Winter survival of individual honey bees and honey bee colonies depends on level of varroa destructor infestation. PloS One 7(4), e36285 (2012).
7. Esch, H.: Beitrfige zum Problem der Entfernungsweisung in den Schwfinzeltfinzen der Honigbienen. Z. Vergleich. Physiol. 48, 534–546 (1964).
8. Gerisch, A., Griffiths, D.F., Weiner, R., Chaplain, M.: A Positive splitting method for mixed hyperbolic-parabolic systems. Num. Meth. PDEs 17 152–168 (2001).
9. Hillen, T., Painter, K.J.: A user guide to pde models for chemotaxis, J. Math. Biology 58(1-2), 183 (2009).
10. Kolev, M., Koleva, M., Vulkov, L.: Two positivity preserving flux limited, second-order numerical methods for a haptotaxis model. Num. Meth. PDEs 29, 1121–1148 (2013).
11. Samarskii, A.A.: The Theory of Difference Schemes, Marcel Dekker Inc, 2001.
12. Ocko, S.A., Mahadevan, L.: Collective thermoregulation in bee clusters. J. R. Soc. Interface 11(91), 20131033 (2014).
13. Stabentheiner, A., Kovac, H., Brodschneider, R.: Honeybee colony thermoregulation regulatory mechanisms and contribution of individuals in dependence on age, location and thermal stress. PLoS One 5(1), e8967 (2010).
14. Tindall, M.J., Maini, P.K., Porter, S.L. Armitage, J.P.: Overview of mathematical approaches used to model bactertial chemotaxis ii: bacterial populations. Bull. Math. Biol. 70(6), 1570 (2008).
15. Watmough, J., Camazine, S.: Self-organized thermoregulation of honeybee clusters. J. Theor. Biology 176(3), 391–402 (1995).

On Jansen–Rit System Modeling Epilepsy Phenomena

Anna Coletti

Abstract Our aim was to describe a generative model of EEG rhythms based on the interaction of anatomically interconnected neuronal populations. In our study we have seen how the description of oscillatory behaviors in the Jansen model is therefore closely related to the study of its bifurcations. They often identify a change in the performance of the system. Even if the behavior of even small systems $d \geq 3$ can be complex, the central manifold theorem allows us to study the behavior of the system in critical points of particular interest, reducing their size, referring to the simpler normal forms that can be easily analyzed. The numerical simulation of mathematical models allows to experiment their dynamics, often confirms mathematical hypotheses, or suggests new possible variations that can be carried out on it.

Keywords Jansen–Rit system · Epilepsy · Center manifold theorem · Andronov–Hopf theorem

1 The Jansen–Rit System

1.1 The Model's Equations

The Jansen and Rit system [1, 2] studies the temporal dynamics of a cortical column through the interaction of three populations of interconnected neurons: pyramidal neurons, excitatory interneurons, and inhibitors interneurons. The population of pyramidal cells receives inhibitory or excitatory feedbacks from local interneurons that are other stellate pyramidal cells that reside in the same column, and excitatory input from nearby cortical units and subcortical structures such as the thalamus.

The excitatory input is represented by an arbitrary mean firing rate $p(t)$ which can be random (representing non-specific background activity) or explicit, taking into account some specific activity in other cortical units [3, 4]. The three families

A. Coletti (✉)
University of Pisa, Pisa, Italy
e-mail: a.coletti@studenti.unipi.it

and the synaptic interactions between them are shaped by different systems [5]. Each of the neuronal populations is modeled by two blocks.

The first types of blocks transform the average frequency of activation into an EPSP or IPSP, that is, into an average excitatory or inhibitory post-synaptic membrane potential, respectively [6, 7].

Both pyramidal neurons and excitatory interneurons use glutamate as a neurotransmitter, and therefore the impulse response of the respective synapses is the same for which it is indicated with the function $h_e(t)$.

On the contrary, inhibitory interneurons make use of the neurotransmitter GABA and the impulse response is indicated with $h_i(t)$.

From the point of view of signal analysis, the representation of the PSP block is a linear transformation given by a convolution with an impulse response function or, equivalently, by a linear differential equation of the second order. The excitatory or inhibitory impulse response function is, respectively, of the form

$$h_i(t) = \begin{cases} Bbte^{-bt} & \text{if } t \geq 0, \\ 0 & \text{if } t < 0. \end{cases}$$

$$h_i(t) = \begin{cases} Bbte^{-bt} & \text{if } t \geq 0, \\ 0 & \text{if } t < 0. \end{cases}$$

In other words, if $x(t)$ is the input to the system, its output $y(t)$ is the convolution product $h_e * x(t)$ or $h_i * x(t)$. The constants A and B, expressed in millivolts, determine the maximal amplitude of the post-synaptic potentials while a and b, expressed in s^{-1}, lumps together characteristic delays of the synaptic transmission. We can express this convolution product also as a second-order differential equation, or equivalently as a system of two first-order differential equations. The second block transforms the average membrane potential of a population of neurons into an average pulse density of action potentials fired by the neurons.

This transformation is described by a sigmoid function of the form

$$Sigm(v) = \frac{2e_0}{1 + e^{r(v_0 - v)}}, \tag{1}$$

where e_0 defines the maximum activation speed of the neural population, v_0 the PSP for which 50% of the discharge rate is reached, and r the slope of the sigmoidal transformation at v_0 which can be viewed either as a firing threshold or as the excitability of the populations. The Sigmoid systems introduce a nonlinear component in the model.

The connectivity constants C_1 C_2 C_3 C_4 account for the number of synapses established between two neuron populations; they can be reduced to a single parameter C.

There are three main variables in the model, the outputs of the three post-synaptic boxes noted y_0, y_1, and y_2; we also consider their derivatives $\dot{y}_0$ $\dot{y}_1$ $\dot{y}_2$ noted y_3, y_4, and y_5, respectively.

If we write two first-order differential equations for each post-synaptic system we obtain a system of six first-order differential equations that describes *Jansen's neural mass model* [1, 2]:

$$
\begin{cases}
\dot{y}_0(t) = y_3(t) \\
\dot{y}_1(t) = y_4(t) \\
\dot{y}_2(t) = y_5(t) \\
\dot{y}_3(t) = AaSigm[y_1(t) - y_2(t)] - 2ay_3(t) - a^2 y_0(t) \\
\dot{y}_4(t) = Aa\{p(t) + C_2 Sigm[C_1 y_0(t)]\} - 2ay_4(t) - a^2 y_1(t) \\
\dot{y}_5(t) = BbC_4 Sigm[C_3 y_0(t)] - 2by_5(t) - b^2 y_2(t).
\end{cases}
\tag{2}
$$

We focus on studying the variable $y = y_1 - y_2$, the membrane potential of the main family of neurons.

We think of this quantity as the output of the unit because in the cortex, pyramidal cells are the main carriers of long-range cortico-cortical connections. Furthermore, their electrical activity corresponds to the EEG signal: pyramidal neurons launch their apical dendrites into the superficial layers of the cortex where the post-synaptic potentials are summed, representing the essential part of the EEG activity.

To allow the model to produce a set of EEG-like signals, based on different neuroanatomical studies, we used the parameter values as established in the work by Jansen and Rit [2].

1.2 Geometric Study

In this section, we consider p as a parameter of the system and we propose to study the behavior of a unit as p varies.

We study the dynamic system with p kept constant and the other parameters set with the values given by Jansen and Rit [2]. We are interested in computing the fixed points and periodic orbits of the system as function of p because they will allow us to account for the appearance of alpha-like and spike-like epileptic activity. Equilibrium point satisfies $\dot{Y} = f(Y, p) = 0$ with $y \in \mathbb{R}^n$ and $p \in \mathbb{R}$, that is, a system of n scalar equations in $\mathbb{R}^{n+1}$ endowed with the coordinates (y, p). Generically $f(y, p) = 0$ defines a smooth one-dimensional manifold (curve) M in $\mathbb{R}^{n+1}$.

The problem of computing the curve M is a specific case of the general (finite-dimensional) continuation problem, which means finding a curve in $\mathbb{R}^{n+1}$ defined by n equations: $F(y) = 0$, $F : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$. By the Implicit Function Theorem [8], the previous system locally defines a smooth curve M, provided that rank $J = n$, where $J = F_y(\bar{y})$ is the Jacobian matrix of $F(y)$ calculated at the equilibrium points $\bar{y}$, which we noted as $S(y)$. Writing $\dot{Y} = 0$ we obtain the system of equations

$$\begin{cases} y_0 = \frac{A}{a} Sigm[y_1 - y_2] \\ y_1 = \frac{A}{a}(p + C_2 Sigm[C_1 y_0]) \\ y_2 = \frac{B}{b} C_4 Sigm[C_3 y_0] \\ y_3 = 0 \\ y_4 = 0 \\ y_5 = 0, \end{cases} \tag{3}$$

which implies the only family of equilibrium points in the (p, y) plane:

$$y = \frac{A}{a} p + \frac{A}{a} C_2 Sigm \left[\frac{A}{a} C_1 Sigm(y) \right] - \frac{B}{b} C_4 Sigm \left[\frac{A}{a} C_3 Sigm(y) \right], \tag{4}$$

where y indicates $y = y_1 - y_2$ which can be thought of as the representative of the EEG activity and p our parameter of interest, labeled by PP.

Number of intersections between the curve and the vertical line $p = constant$ are the equilibrium points for that particular value of p. We draw the curve also for negative p values, for which there is no biological value as p as at firing rate but which play a fundamental role in the mathematical description of the model.

To study stability without solving the system of nonlinear differential equations, we use the indirect Lyapunov method or linearization method [9, 10].

$J(S(y))$ is the Jacobian matrix of $f(y, p)$ respect of y; it is the matrix of the linearized dynamic system in a neighborhood of the equilibrium point $S(y)$. So we have

$$J(S(y)) = \begin{pmatrix} 0_3 & I_3 \\ KM(y) & -K \end{pmatrix},$$

where

$$K = 2 \, diag(a, a, b), \quad M(y) = \begin{pmatrix} \frac{-a}{2} & \gamma(y) & -\gamma(y) \\ \delta(y) & \frac{-a}{2} & \\ \theta(y) & 0 & \frac{-b}{2}. \end{pmatrix}$$

I_3 is the three-dimensional identity matrix and 0_3 the three-dimensional null matrix.

The three functions $\gamma(y)$, $\delta(y)$ and $\theta(y)$ are defined by

$$\gamma(y) = \frac{A}{2} Sigm'(y), \quad \delta(y) = \frac{AC_1 C_2}{2} Sigm'(C_1 y_0(y)),$$

$$\theta(y) = \frac{BC_3 C_4}{2} Sigm'(C_3 y_0(y)),$$

Fig. 1 A branch of equilibria in the (PP, y0)-plane differentiating stable and unstable fixed points and the bifurcation points
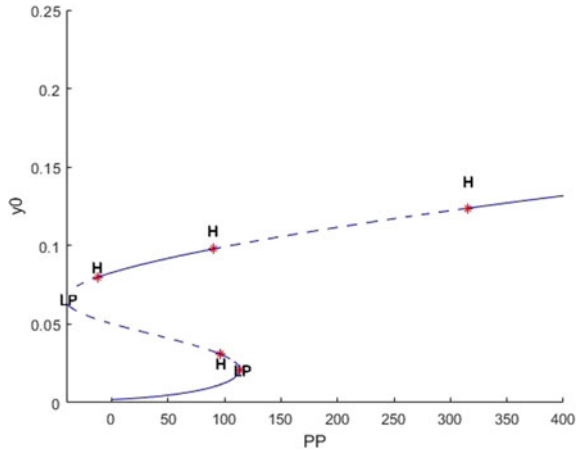
where $y_0(y)$ is the first coordinate of $S(y)$ and $Sigm'(x)$ is the derivative of the function $Sigm(x)$.

In dynamic models or physical systems, the mechanism underlying a transition in behavior is often a local bifurcation of equilibria [11]. The bifurcation analysis of the system means building its one-parameter bifurcation diagram. In particular, study the dependence of equilibria and limit cycles on the parameter, as well as locate and analyze their bifurcations. The complexity of the system given by the sigmoidal function makes the research and study of the latter analytically impractical.

In order to identify and study the bifurcations, we used the MatCont software based on the interpreted language MATLAB [12]. MatCont is a graphical MATLAB software package for the interactive numerical study of dynamical systems. Starting from an orbit converging to an equilibrium point, we have calculated by continuation the entire curve of fixed points. Subsequently we compute the eigenvalues of J along the curve to analyze the stability of the family of equilibrium points. In particular, the fixed points for which the eigenvalues have a real negative part are stable, defined by a continuous line, while the fixed points for which at least one eigenvalue has a real positive part are unstable, defined by a dashed line. For the purpose of the geometric study of the system, it is important to identify the points where at least one eigenvalue has no real part, called the bifurcation points, where a drastic and sudden change in the behavior of the system takes place (Fig. 1).

The first limit point appears for $p = 113.58624$. Here one eigenvalue has zero real part; the curve of equilibria has a turning point, on one side the equilibria are stable on the other unstable. Here there is a saddle-node bifurcation with homoclinic orbits.

In the meantime MatCont finds a special point also in $p = 96$, here a test function is zero, but that is due to two real eigenvalues, one positive and one negative, summing up to zero, not a complex pair. This is mentioned in the main MatCont window as a neutral saddle equilibrium. At $p = -41.30141$ another Limit point is found. Here

the branch does not become stable, just one more eigenvalue has now positive real part.

In the model for $p = 89.829112$ and for $p = 315.6964$ two complex conjugate eigenvalues cross the imaginary axis. The system undergoes what is called a Hopf bifurcation. In particular, a theorem due to Hopf which we will see in the following section shows that for $p = 89.83$ a one-parameter family of stable periodic orbits appears at the fixed point that has two complex conjugate eigenvalues crossing the imaginary axis toward positive real parts.

These periodic orbits persist till $p = 315.70$ where a second Hopf bifurcation occurs: the two eigenvalues whose real parts became positive for $p = 89.83$ see them become negative again, corresponding to the (re)creation of a simple attractive fixed point. For p between 89.83 and 315.70, there is a family of periodic orbits that we call Hopf cycles (Fig. 2). Numerically if we consider the difference function $y = y_1 - y_2$ this oscillates around 10 Hz, witch correspond to alpha activity, as can also be seen in the simulations of $y(t)$ for $p = 200$ (Fig. 3).

We have experimentally calculated the period of the orbits in the Hopf cycle and we have noticed how this remains constant as p varies, as shown in the following figure (Fig. 4).

We do not mention the bifurcations obtained for negative p-values, in particular, for $p = -12.147468$ we have another Hopf bifurcation, because only the positive p-values have biological relevance, while the study for each p value will be necessary for mathematical evaluations. Nevertheless, this last bifurcation will be important in the model to explain phenomena of spike activity.

The other branch of limit cycles lies in the domain between the star labeled LP, where there is a saddle-node bifurcation and the red line LPC representing a fold bifurcation of limit cycles. We will call these orbits spike cycles. The branch of the spike cycles begins at $p = 113.58$, thanks to a bifurcation of the saddle node and ends at $p = 137.38$ due to a fold bifurcation of the limit cycles that we have identified with

Fig. 2 Hopf cycle
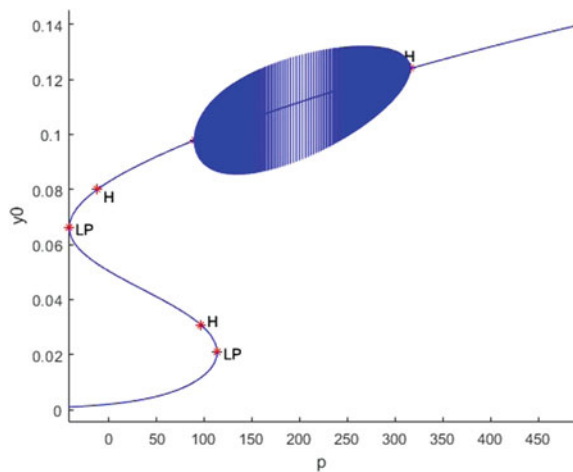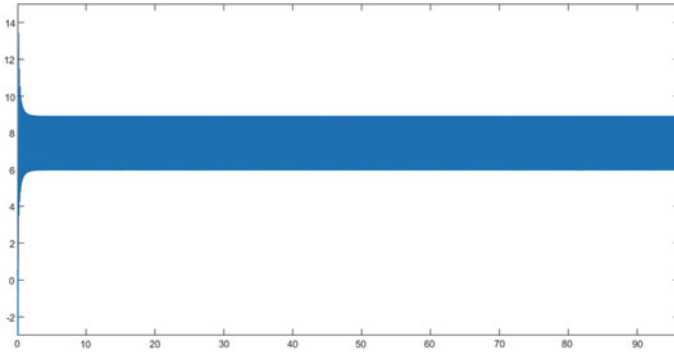
Fig. 3 $y(t)$ for $p = 200$

Fig. 4 Period of Hopf cycle
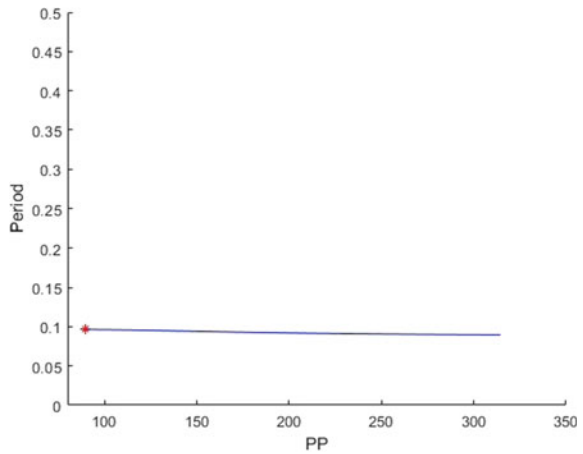
MatCont and appears in the window as the red line labeled LPC. This bifurcation results from the fusion of a stable and an unstable family of periodic orbits.

The unstable family of orbits originates in the Hopf bifurcation obtained for $p = -12.147468$.

In fact, we note that for a value belonging to such an interval such as $p = 135$, the system numerically simulates an activity similar to that recorded by epileptic EEGs.
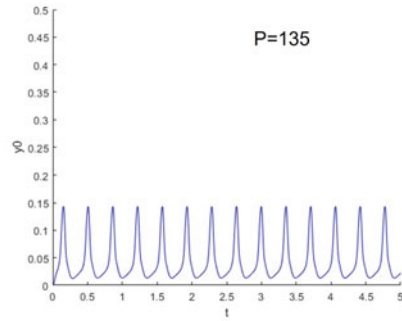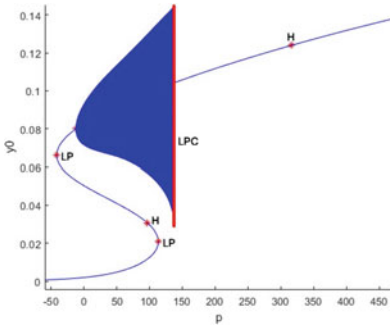
1.3 The Central Manifold Theorem and the Andronov–Hopf Bifurcation

The term Hopf bifurcation (or Poincaré-Andronov–Hopf bifurcation) refers to local birth or death of a periodic solution from an equilibrium when a parameter exceeds a critical value. It is the simplest bifurcation that does not involve only the equilibria and therefore belongs to what is sometimes called dynamic (as opposed to static) bifurcation theory. In a differential equation a Hopf bifurcation typically occurs when a complex conjugate pair of linearized flow eigenvalues at a fixed point becomes purely imaginary. This implies that a Hopf bifurcation can only occur in systems of dimension two or greater.

The behavior of even small systems $d \geq 3$ can be complex. The central manifold theorem is a mathematical tool that allows to study the behavior of the system near critical points of particular interest, reducing their size. We say that a subspace $V \subset \mathbb{R}^n$ is invariant for the system if $x \in V$ implies $e^{At}x \in V$ for all $t \in \mathbb{R}$.

The space $\mathbb{R}^n$ can therefore be decomposed as the sum of a stable, unstable, and central subspace, crossed, respectively, by the (generalized) eigenvectors corresponding to eigenvalues of A with negative, positive, and zero real parts. We have so $\mathbb{R}^n = V^s \oplus V^u \oplus V^c$. In particular these sub-spaces are invariant for the flow of the system.

The Center Manifold Theorem applies to the case where V^c is nontrivial. It says that near the origin all the interesting dynamics takes place on an invariant manifold $\mathcal{M}$, tangent to the center subspace V^c.

Theorem 1 *Center Manifold Theorem Let $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ be a vector field in $\mathbb{C}^{k+1}$ (here $k \leq 1$), with $f(0) = 0$. Consider the matrix $A = Df(0)$, and let V^s, V^u, V^c be the corresponding stable, unstable, and center sub-spaces. Then there exists $\delta > 0$ and a local center manifold $\mathcal{M}$ with the following properties:*

– There exists a $\mathbb{C}^k$ function $\phi : V^c \mapsto \mathbb{R}^n$ with $\pi_c \phi(x_c) = x_c$ such that

$$\mathcal{M} = \{\phi(x_c); x_c \in V^c, |x_c| < \delta\}.$$

- *The manifold $\mathcal{M}$ is locally invariant for the flow of the system, i.e., $x \in \mathcal{M}$ implies $\tilde{x}(t, x) \in \mathcal{M}$ for $|t|$ small.*
- *$\mathcal{M}$ is tangent to V^c at the origin.*
- *Every globally bounded orbit remaining in a suitably small neighborhood of the origin is entirely contained inside M.*
- *Given any trajectory such that $x(t) \mapsto 0$ as $t \mapsto \infty$, there exists $\nu > 0$ and a trajectory $t \mapsto y(t) \in \mathcal{M}$ on the center manifold such that $e^{\nu t}|x(t) - y(t)| \mapsto 0$, as $t \mapsto \infty$.*

1.3.1 Andronov–Hopf Theorem

In an n-dimensional system, the Center Manifold Theorem guarantees that the oscillatory dynamics will be qualitatively similar to the two-dimensional case, there is therefore a two-dimensional surface on which the limit cycle will emerge.

Theorem 2 *Andronov–Hopf theorem In $n \geq 3$ assume that its Jacobian matrix $A(\alpha) = f_x(x^0(\alpha), \alpha)$ has*

- *one pair of pure complex eigenvalues $\lambda_{1,2} = +- i\omega(\alpha)$;*
- *n_s eigenvalues with $Re\lambda_j < 0$;*
- *n_u eigenvalues with $Re\lambda_j > 0$;*

with $n_s + n_u + 2 = n$. According to the Center Manifold Theorem, there is a family of smooth two-dimensional invariant manifolds W_α^c near the origin. The n-dimensional system restricted on W_α^c is two dimensional, hence has the normal form above. Moreover, under the non-degeneracy conditions, the n-dimensional system is locally topologically equivalent near the origin to the suspension of the normal form by the standard saddle,

$$\dot{y}_1 = \beta y_1 - y_2 + \sigma y_1(y_1^2 + y_2^2) \ \dot{y}_2 = y_1 + \beta y_2 + \sigma y_2(y_1^2 + y_2^2)$$

$$\dot{y}^s = -y^s \ \dot{y}^u = +y^u,$$

where $y = (y_1, y_2)^T \in \mathbb{R}^2$, $y^s \in \mathbb{R}^{n_s}$ and $y^u \in \mathbb{R}^{n_u}$.

In particular $sign(l_1(0)) = \sigma = \pm 1$ determines if there is a supercritical or a subcritical Andronov–Hopf bifurcation.

If $\sigma = -1$, the normal form has an equilibrium at the origin, which is asymptotically stable for $\beta \leq 0$ (weakly at $\beta = 0$) and unstable for $\beta < 0$. Moreover, there is a unique and stable circular limit cycle that exists for $\beta > 0$ and has radius $\sqrt{\beta}$. This is a supercritical Andronov–Hopf bifurcation.

If $\sigma = +1$, the origin in the normal form is asymptotically stable for $\beta < 0$ and unstable for $\beta \leq 0$ (weakly at $\beta = 0$), while a unique and unstable limit cycle exists for $\beta < 0$. This is a subcritical Andronov–Hopf bifurcation.

1.3.2   Conditions Necessary for the Hopf Bifurcation

Consider now an equilibrium solution for the second-order differential equation of one population $y = Y(p)$, such that $h(0, Y, p) = 0$, so that $y = Y$ is a solution for any fixed p. There is no loss of generality in assuming $Y(p) = 0$ for all values of p, since we can always change variables $y_{old} = Y(p) + y_{new}$. The linearized equation near the equilibrium solution $Y(p) \equiv 0$ is $\ddot{y} - 2\alpha\dot{y} + \beta y = 0$, where $\alpha = \alpha(p) = -\frac{1}{2}h_{\dot{y}}(0, 0, p)$ and $\beta = \beta(p) = h_y(0, 0, p)$.

The critical point is a spiral point if $\beta > \alpha^2$.

If we assume now that at $p = 0$ the critical point changes from a stable to an unstable spiral point (if the change occurs for some other $p = p_c$, one can always redefine $p_{old} = p_c + p_{new}$). Thus $\alpha < 0$ for $p < 0$ and $\alpha > 0$ for $p > 0$, with $\beta > 0$ for p small.

In fact, assume h is smooth and $\alpha(0) = 0$, $\beta(0) > 0$ and $\frac{d}{dp}\alpha(0) > 0$.

This last is known as the transversality condition. It guarantees that the eigenvalues cross the imaginary axis as p varies.

In our case, the equations are

$$\begin{cases} \ddot{y}_0 - Aa\,Sigm[y_1 - y_2] + 2a\dot{y}_0 - a^2 y_0 = 0 \\ \ddot{y}_1 - Aa\{p + C_2 Sigm[C_1 y_0]\} + 2a\dot{y}_1 + a^2 y_1 = 0 \\ \ddot{y}_2 - BbC_4 Sigm[C_3 y_=] + 2b\dot{y}_2 + b^2 y_2 = 0. \end{cases}$$

The Hopf bifurcation appears for $\bar{p} = 89.829112$, and in this point we have the following phase space coordinates $Y(\bar{p}) = (0.097839\ 20.164937\ 13.425371\ 0.000000\ 0.000000\ 0.000000)$.

The system must therefore be translated into $y_{old} = Y(\bar{p}) + y_{new}$, where $y = (y_0, \ldots, y_5)$. We omit the development in the translated point and the verification as it is pure algebra.

2   Simulate Signals

The bifurcation diagram is a valid tool for describing the behaviors of the Jansen–Rit model when the afferent input from the adjacent regions is modeled by a white noise or a Gaussian noise, whose input frequency is different.

In fact, the geometric study suggests the values of the parameter p for which an interesting behavior of the system is expected.

Setting the input $p(t)$ in the equations of the model as a random input uniformly distributed between 120 and 320 pulses per second, an alpha-like activity was observed for $C = 135$, which we can see in Fig. 5.

On the other hand, if we use as input $p(t)$ in the model a Gaussian white noise with mean 90 and standard deviation 30, corresponding to a rate of 30–150 pulse/s

Fig. 5 Alpha rhythm with $p(t)$ as a random input uniformly distributed between 120 and 320 pulses per second
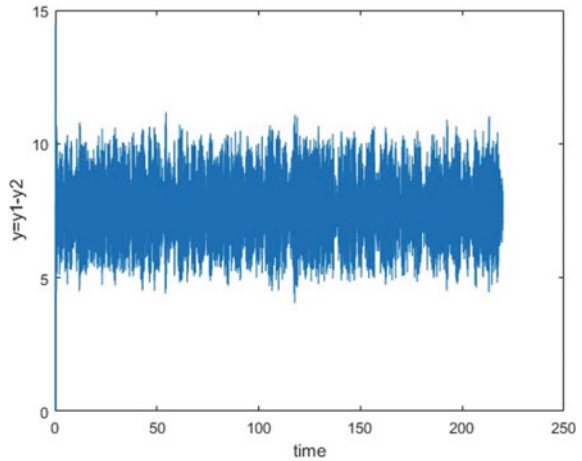
Fig. 6 Spike activity with $p(t)$ as a Gaussian white noise corresponding to a rate of 30–150 pulse per second
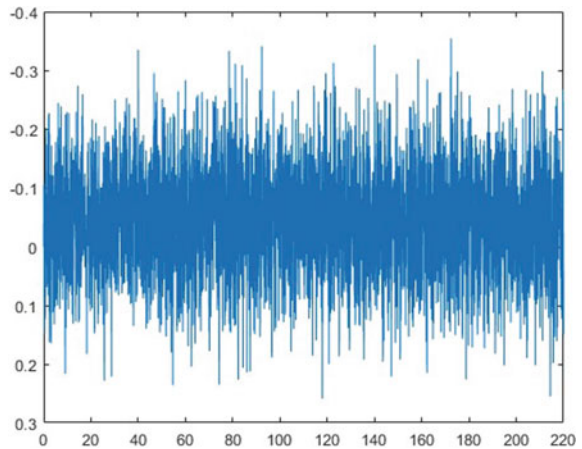
we can see that the model produces a signal similar to a spontaneous EEG recorded from neocortical structure electrodes during interictal periods (Fig. 6).

Notes and Comments All stochastic simulations were performed using the Milstein method, a variation of the Euler–Maruyama method while the deterministic simulations are performed using the Runge–Kutta 4 method.

References

1. Ben H. Jansen, George Zouridakis, Michael E. Brandt, *A neurophysiologically-based mathematical model of flash visual evoked potentials*, Biol. Cybern. 68, 275–283 (1993)
2. Ben H. Jansen, Vincent G. Rit, *Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns*, Biol. Cybern. 73, 357–366 (1995)

 3. Hugh R. Wilson, Jack D. Cowan, *Excitatory and inhibitory interactions in localized populations of model neurons*, Biophysical Journal Volume 12, 1972
 4. H. R. Wilson and J. D. Cowan, *A Mathematical Theory of the Functional Dynamics of Cortical and Thalamic Nervous Tissue*, Kybernetik 13, 55–80, Springer-Verlag 1973
 5. F. H. Lopes da Silva, A. Hocks*, H. Smits, and L. H. Zetterberg, *Model of Brain Rhythmic Activity The Alpha-Rhythm of the Thalamus*, Kybernetik 15, 27–37 (1974), Springer-Verlag 1974
 6. F. Wendling, J.J. Bellanger, F. Bartolomei, P. Chauvel, *Relevance of nonlinear lumped-parameter models in the analysis of depth-EEG epileptic signals*, Biological Cybernetics, November 2000
 7. R. C. Sotero, N. J. Trujillo-Barreto, *Realistically Coupled Neural Mass Models can generate EEG rhythms*, Neural computation 19, 4788–512 (2007)
 8. Lawrence Perko, *Differential Equations and Dynamical Systems*, Springer, 2001
 9. Yuri A. Kuznetsov, *Elements of Applied Bifurcation Theory, Second Edition*, Springer, 1998
10. The MIT press, *Hopf Bifurcations*, Department of Mathematics, Massachusetts Institute of Technology Cambridge, Massachusetts MA 02139
11. Francois Grimbert, Olivier Faugeras, *Bifurcation analysis of Jansens neural mass model*, Neural Computation 18, 3052–3068 (2006)
12. W. Govaerts, Yu. A. Kuznetsov V. De Witte, A. Dhooge, H.G.E. Meijer, W. Mestrom, A.M. Riet and B. Sautoi, *MATCONT and CL MATCONT: Continuation toolboxes in matlab*, 2011, University Gent, Utrecht University The Netherlands

Applications in Financial Mathematics

Forex Time Series Forecasting Using Hybrid Convolutional Neural Network/Long Short-Term Memory Network Model

Maya Markova (ORCID)

Abstract The deep learning approach plays a meaningful role in predicting financial time series data. This research proposes a time series deep learning hybrid model based on the convolutional neural network and long short-term memory (CNN-LSTM) framework for predicting EUR/USD exchange rate. The CNN-LSTM Encoder–Decoder model for multivariate multi-step time series forecasting is developed and evaluated with the 5-min time interval foreign exchange rate of EUR/USD data. The historical data are transformed into a three-dimensional structure to prepare the data for fitting the model. Dataset preparation and CNN-LSTM model are made using Python. To solve a problem with dying ReLU in 1D CNN layers, a parametric leaky ReLU activation function with He kernel initialization is used. The model takes six timesteps of four features as input—open, high, low, and close prices in 5-min intervals of EUR/USD price and predicts the following three timesteps for the same four features.

Keywords Forex · Deep neural network · CNN-LSTM model

1 Introduction

The foreign exchange market (Forex) is an intercontinental market where you can trade any currency of any country in the world. The foreign exchange market opened up in the 1970s by introducing floating exchange rates. Since then, individuals and companies that have participated in the foreign exchange market have always determined the comparison between currency prices based on the law of supply and demand. The Forex market has always been considered a free market as it is free from any external control. All those wishing to engage in forex trading are free to compete with each other and can choose to trade or not at any time.

The sheer number of forex trades taking place in a continuous motion every day makes the forex market one of the most liquid money markets in the world.

M. Markova (✉)
Angel Kanchev University of Rousse, Rousse 7017, Bulgaria
e-mail: maya.markova@gmail.com

According to various research studies, the volume of money traded in the foreign exchange market amounts to almost five trillion dollars daily. The highest transaction volume recorded is almost $6 trillion in just one day. The average daily amount of total transactions is around $3 trillion. The exact amount cannot really be determined since not all foreign exchange transactions are processed at a central exchange office.

Foreign exchange transactions could be conducted from anywhere in the world using telecommunications networks and the Internet. Unlike the stock market, which only trades during traditional business hours, the forex market trades 24 h a day, 5 days a week. Forex trading starts at 00:00 GMT on Mondays and ends at 22:00 GMT on Fridays. Always a dealer could be found in all time zones. Major currency exchanges are located in the following major markets: New York in North America, Frankfurt and London in Europe, Tokyo and Hong Kong in Asia, and Australia and New Zealand in the Pacific [1].

There are two different methods of analyzing the financial markets: fundamental analysis and technical analysis. Fundamental analysis is principally based on economics, while technical analysis uses historical prices to try to predict future movements. Short-term traders focus their strategies mostly on price movements and prefer to use technical analysis. Fundamental analysis focuses on the social, economic, and political forces that drive supply and demand. Those who use fundamental analysis as a trading tool look at various macroeconomic indicators such as interest rates, growth rates, inflation, and unemployment.

With the growing popularity of technical analysis and the emergence of new technologies, the influence of technical trading on the forex market has increased significantly. Technical analysis emphases more on reviewing price movements. It uses historical currency data to predict future price developments. Technical analysis supposes that all information about the current market is already revealed in the currency price. Therefore, studying price action is all that is required to make informed trading decisions. Furthermore, technical analysis assumes that history may repeat itself [2].

According to investors lately, foreign exchange has surpassed other investment instruments like futures, bonds, and stock exchange [3–5]. When the exchange rate oscillates often, investors may make substantial gains at a minimal cost. As a result, there has been a significant increase in interest in forecasting exchange rates using artificial neural networks, in particular time series forecasting. The researchers found that exchange rate fluctuations behave as a nonlinear dynamic system affected by many complex variables. It was evident that using linear thinking cannot predict the performance of a nonlinear system. To meet the market demand for higher predictability it is crucial to use a nonlinear exchange rate forecasting method like deep learning. Deep learning is also proving effective in the field of currency forecasting, in addition to its already well-known application areas such as image and video processing, voice recognition, natural language processing, and others. As a consequence, deep learning algorithms are attracting increasing interest [6].

In recent years, deep neural network studies have made significant progress in data science. On many prediction tasks, models like CNN, LSTM, and transformers perform better than traditional machine learning. However, due to the uncertain, irregular, and unstable nature of financial markets and the complexity of financial time series data, research based on deep neural network financial forecasts still has room for improvement [7]. Machine learning is more effective in short-term prediction (minutes or hours) than long-term prediction because economic news is less likely to influence the price movement.

This project aims to investigate combining CNN and LSTM layers in one hybrid CNN-LSTM model for Forex time series forecasting with 5-min EUR/USD exchange data.

2   Convolutional Neural Networks

Convolutional neural networks (CNN) capacity to learn automatically from raw data may be used for time series forecasting. The data could be seen as a one-dimensional picture that a CNN model could interpret and find its most important features. CNNs support multivariate input and output, as well as the ability to learn arbitrary but complicated functional relationships [8].

The theme of sequence modeling in deep learning was mainly associated with recurrent neural network designs such as LSTM and GRU, until recently. Bai et al. [9] guide that this is an old-fashioned way of thinking, so the convolutional networks must be considered the primary option for representing sequential data. They demonstrated that convolutional networks beat RNNs in a wide variety of tasks while avoiding typical disadvantages allied with recurrent models, including the exploding/vanishing gradient issue or a lack of memory retention. By combining multiple Conv1D layers, a CNN may successfully study long-term relationships in time series [8].

3   Long Sort-Term Memory Neural Networks

The long short-term memory (LSTM) network is a recurrent neural network (RNN), which was first proposed by Sepp Hochreiter and Jurgen Schmidhuber [10] and later improved by Alex Graves [11].

LSTMs were created to overcome the issue of vanishing and exploding gradients that can be seen during the training of conventional RNNs. LSTM unit includes a memory cell that stores information updated by three special gates: the forget, the input, and the output gate. The cell remembers values over random time intervals, and the three gates regulate the stream of data into and out of the cell (see Fig. 1).

Fig. 1 LSTM architecture
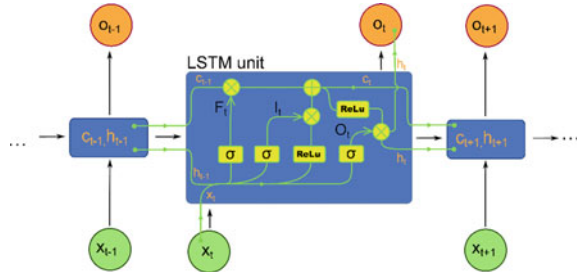
LSTM internal state variable is passed from cell to cell and is modified by Operation Gates.

1. Forget Gate

$$F_t = \sigma(W_f.\big[h_{t-1}, x_t\big] + b_f).\qquad(1)$$

The output at $t - 1$ is concatenated with the current input at time t into a single tensor. Then is applied a linear transformation afterward, a sigmoid. The output of this gate is between zero and one because of the sigmoid. The results number is multiplied by the internal state, and if $f_t = 0$, the previous internal state is forgotten, where the name forget gate is come. If $f_t = 1$, the previous state will be passed through unchanged.

2. Input Gate

$$I_t = \sigma(W_i.\big[h_{t-1}, x_t\big] + b_i).\qquad(2)$$

The input gate passes the previous output and the new input through one more sigmoid layer. The output of the gate is between zero and one, and the output of the candidate layer is multiplied by the input gate value.

$$C_t = ReLU(W_C.\big[h_{t-1}, x_t\big] + b_C).\qquad(3)$$

This layer applies a Rectified Linear Unit (ReLU) function to the input and previous output and returns a candidate vector which is summed with the internal state.

The following rule is used to update the internal state:

$$C_t = f_t * C_{t-1} + i_t * C_t.\qquad(4)$$

The forget gate and the previous state are multiplied and added to the segment of the new candidate allowed by the output gate.

3. Output Gate

$$O_t = \sigma\left(W_O.\left[h_{t-1}, x_t\right] + b_O\right), \tag{5}$$

$$h_t = O_t * ReLUC_t. \tag{6}$$

This gate works similarly and controls how much of the internal state is sent to the output.

The described three gates have separate weights and biases. The network finds out which part of the current input and past output to save and which part of the internal state to send to output [12].

4 Activation Function Parametric ReLU

An essential component of artificial neural networks (ANN) is activation functions. They are mathematical equations that define the outputs of neural nodes. The activation functions permit the network to learn complex data features and add nonlinearity to a neural network. The selection of activation function has a sustainable influence on the ANNs performance, and one of the trendy choices recently is the ReLU [13].

Unfortunately, the ReLU activation function suffers from a so-called dying ReLUs problem. During training, part of the neurons starts to output only zeros. Sometimes, this could happen with half of the neurons in the network, especially if it is trained with a significant learning rate. The gradient of the ReLU function is zero when its input is negative or zero. Because of zeros, output gradient descent does not influence anymore. To solve this problem, variants of the ReLU function could be used, like leaky ReLU or parametric ReLU [14]. Adding a slight slope in the negative area of the function that causes small negative outputs for smaller than zero inputs is an effective way to solve a dying ReLU problem [13].

The parametric leaky ReLU (PReLU) function is a variant of leaky ReLU where the parameter a is not a hyperparameter but is learned during the training. The parametric leaky ReLU function in [15] is defined as

$$f(y_i) = \begin{cases} y_i, & if\ y_i > 0 \\ a_i y_i, & if\ y_i \leq 0 \end{cases}, \tag{7}$$

where y_i is the input of the activation function and the coefficient a_i controlling the slope of function in the negative part. When $a_i = 0$, it became a ReLU function and, when a_i is hyperparameter with the small and fixed value, it became a Leaky ReLU function. When a_i is learned during the training function, it is called Parametric ReLU. Equation (7) is equivalent to

$$f(y_i) = \max(0, y_i) + a_i \min(0, y_i). \tag{8}$$

As described in [15], PReLU could be trained simultaneously with other layers during backpropagation. The gradient of a_i for one layer is

$$\frac{\partial \varepsilon}{\partial a_i} = \sum_{y_i} \frac{\partial \varepsilon}{\partial f(y_i)} \frac{\partial f(y_i)}{\partial a_i}, \tag{9}$$

where ε represents the objective function. The term $\frac{\partial \varepsilon}{\partial f(y_i)}$ is the deeper layer gradient. The gradient of the activation is assumed by

$$\frac{\partial f(y_i)}{\partial a_i} = \begin{cases} 0, & if\ y_i > 0 \\ y_i, & if\ y_i \leq 0 \end{cases}. \tag{10}$$

5 CNN-LSTM Encoder–Decoder Model with Multivariate Input

In an encoder–decoder architecture, a convolutional neural network could be adopted as the encoder. The 1D CNN can read across sequence input and automatically learn the relevant features. Then the data could be interpreted by an LSTM decoder. We discuss hybrid CNN-LSTM models that use CNN and LSTM layers, forming an encoder–decoder architecture. The CNN layer expects as input the same 3D structure data as the LSTM layers, though multiple features are read like different channels of an image that finally have the same effect [16].

5.1 The Data

Data from www.finam.ru [17] for 5-min exchange rates of EUR/USD from 1 June 2017 to 27 May 2022 (total 379 325 timesteps) were selected for the study. The open, high, low, and close prices were preprocessed in the shape of (samples, timesteps, and features). We have four features: an open, a high, a low, and a close price. The timesteps are six (a total of half an hour) at a 5-min interval. The input and output data samples are made by dividing the multivariate input data sequence in the following manner. The first six timesteps of data are input, while the following three timesteps of each of the fourth features are used as output (target) data for the model. The second sample takes the input data between 2 and 7 timesteps and output between 8 and 10 timesteps, and so on.

5.2 The Model

We made multivariate multi-step input data for the first 1D CNN layer to forecast the Forex time series. In a 1D CNN, the kernel travels in a single direction. Input data has four features—time series for open, high, low, and close prices in 5-min intervals of EUR/USD. The model predicts the following three timesteps for the same four features. The convolution is like a sliding window function applied in one direction to a input data matrix. This sliding window in a convolutional layer is named a filter or feature detector. The size/height of this filter, called kernel size, in our model is equal to the number of input timesteps.

We created many models, and our final model consists of five layers after refining.

The parameter settings of the constructed CNN-LSTM model are shown in Table 1.

The model consists of two 1D convolutional layers, a max pooling layer, one LSTM layer, and a Dense output layer. Input data are in a three-dimensional data vector (None, 6, 4), in which 6 is the size of the timestep and 4 is the four features of the input data. To solve a problem with dying ReLU, in 1D CNN layers parametric leaky ReLU activation function is used.

Table 1 Parameter settings of the CNN-LSTM model

Parameters	Value
Convolution layer1 filters	2048
Convolution layer2 filters	2048
Convolution layer kernel_size	6
Convolution layer activation function	PReLU
Convolution layer padding	Causal
Pooling layer pool_size	2
Pooling layer padding	Same
Number of hidden units in LSTM layer	256
LSTM layer activation function	ReLU
Time_step	6
Batch_size	2048
Learning rate	0.00004
Optimizer	Adam
Loss function	mean_squared_error
Metrics	RootMeanSquaredError
Epochs	320

6 Construction of CNN-LSTM Model in Python

The Google Colab Pro environment and Python's Keras deep learning library are used for building the model. The Python code of the model construction is shown in Fig. 2.

The training data are separated into 70% for training and 30% for validation. To prevent overfitting, which often occurs in more complex models, we use regularization L2 for kernel_regularizer, bias_regularizer, and activity_regularizer (see Fig. 2).

We first made a model with a ReLU activation function for 1D CNN and LSTM layers. However, at some point, the model stopped to learn. To solve this problem with dying ReLU, we use a parametric leaky ReLU activation function in 1D CNN layers and a lower learning rate. The model results (Fig. 3) show that the loss function for the training data is 1.7E-04. The validation data loss function measured by MSE is about 1.8E-04, measured by RMSE is 0.0036, which is an outstanding result.

After fitting, the model is evaluated on the test dataset for 5-min exchange rates of EUR/USD from 27 May 2022 to 22 June 2022, and the loss function for the test data set measured with mean squared error is 2.12E-04 (Fig. 4).

```
n_steps_in,n_steps_out=6,3
n_features=X_train.shape[2]

kernel_reg_bias=regularizers.l2(l2=0.1)
kernel_reg_act=regularizers.l2(l2=0.0001)
kernel_reg=regularizers.l2(l2=0.0000001)
kernel_reg_act_LSTM=regularizers.l2(l2=0.0000006)
opt = Adam(learning_rate=0.00004)
pleaky_relu = keras.layers.PReLU()

model = Sequential()
model.add(Conv1D(2048, 6, padding='causal', activation=pleaky_relu, input_shape=(n_steps_in, n_features), kernel_initializer='he_normal',
                 kernel_regularizer=kernel_reg, bias_regularizer=kernel_reg_bias, activity_regularizer=kernel_reg_act))
model.add(Conv1D(2048, 6, padding='causal', activation=pleaky_relu, kernel_initializer='he_normal', kernel_regularizer=kernel_reg,
                 bias_regularizer=kernel_reg_bias, activity_regularizer=kernel_reg_act))
model.add(MaxPooling1D(pool_size=2, strides=2, padding='same'))
model.add(LSTM(256, activation='relu', return_sequences=True, kernel_initializer='he_normal', kernel_regularizer=kernel_reg_act_LSTM))
model.add(TimeDistributed(Dense(4)))
model.compile(optimizer=opt, loss='mse', metrics=[tf.keras.metrics.RootMeanSquaredError()])
print(model.summary())
```

Fig. 2 The Python code of the model construction

```
Epoch 311/320
130/130 [=====] - loss: 1.8079e-04  root_mean_squared_error: 0.0024 - val_loss: 1.8749e-04 val_root_mean_squared_error: 0.0034
Epoch 312/320
130/130 [=====] - loss: 1.7990e-04  root_mean_squared_error: 0.0024 - val_loss: 1.8673e-04 val_root_mean_squared_error: 0.0034
Epoch 313/320
130/130 [=====] - loss: 1.7943e-04  root_mean_squared_error: 0.0026 - val_loss: 1.8876e-04 val_root_mean_squared_error: 0.0039
Epoch 314/320
130/130 [=====] - loss: 1.7918e-04  root_mean_squared_error: 0.0027 - val_loss: 2.0099e-04 val_root_mean_squared_error: 0.0053
Epoch 315/320
130/130 [=====] - loss: 1.7822e-04  root_mean_squared_error: 0.0027 - val_loss: 2.0643e-04 val_root_mean_squared_error: 0.0059
Epoch 316/320
130/130 [=====] - loss: 1.7740e-04  root_mean_squared_error: 0.0028 - val_loss: 1.9841e-04 val_root_mean_squared_error: 0.0052
Epoch 317/320
130/130 [=====] - loss: 1.7643e-04  root_mean_squared_error: 0.0028 - val_loss: 1.8822e-04 val_root_mean_squared_error: 0.0043
Epoch 318/320
130/130 [=====] - loss: 1.7512e-04  root_mean_squared_error: 0.0027 - val_loss: 1.8305e-04 val_root_mean_squared_error: 0.0038
Epoch 319/320
130/130 [=====] - loss: 1.7394e-04  root_mean_squared_error: 0.0027 - val_loss: 1.8094e-04 val_root_mean_squared_error: 0.0036
Epoch 320/320
130/130 [=====] - loss: 1.7290e-04  root_mean_squared_error: 0.0026 - val_loss: 1.7966e-04 val_root_mean_squared_error: 0.0036
```

Fig. 3 Results for loss function measured by mean square error and root mean squared error

```
dataset_test=read_csv('/content/drive/MyDrive/Python/EURUSD_220527_220609.csv', header=0, infer_datetime_format

history=model.fit(X_train, y_train,batch_size=2048, validation_split=0.3, shuffle=False, epochs=320, verbose=1)

model.save('/content/drive/MyDrive/Python/model_CNN_LSTM_2022_0307_03.h5',overwrite=True)
print("Saved model to disk")

test_mse = model.evaluate(X_test, y_test, batch_size=2048, verbose=1)

None
2/2 [==============================] - 1s 722ms/step - loss: 2.1224e-04 - root_mean_squared_error: 0.0074
1/1 [==============================] - 0s 319ms/step
```

Fig. 4 Fit and evaluate model python code

To predict new data, we loaded the model and inserted new data for six timesteps of open, high, low, and close prices. The result is a prediction of the following three 5-min interval timesteps for four features open, high, low, and close price (Fig. 5). The result shows that the error of our model prediction with fresh, unseen data is from -0.00054 to 0.017336 and is smaller for first of the three steps.

```
# load and evaluate a saved model
from keras.models import load_model
from numpy import array

# load model
model = load_model('/content/drive/MyDrive/Python/model_CNN-LSTM_2022_0307_03.h5')
# summarize model.
model.summary()

n_steps_in,n_steps_out=6,3
n_features=4

x_input=array([[1.0190100,1.0196800,1.0188000,1.0191700],
               [1.0192000,1.0197800,1.0189000,1.0195000],
               [1.0195100,1.0213300,1.0193000,1.0210400],
               [1.0210200,1.0211000,1.0200000,1.0204100],
               [1.0204000,1.0208000,1.0199000,1.0207300],
               [1.0207600,1.0211100,1.0203000,1.0206100]])
x_input=x_input.reshape((1,n_steps_in,n_features))
yhat=model.predict(x_input,verbose=1)
print(yhat)

1/1 [==============================] - 0s 221ms/step
[[[1.0199577 1.0192828 1.0189508 1.0203512]
  [1.0306733 1.0299082 1.0295185 1.0309415]
  [1.0375664 1.0366689 1.036367  1.0376991]]]
```

Fig. 5 Making predictions with new data of the following six timesteps

7   Conclusion

In this paper, we make a hybrid CNN-LSTM model for time series prediction of EUR/USD exchange rate using historical intraday data for the open, high, low, and close price of 5-min interval. The Google Colab Pro environment is used for creating and testing the Python model. The model has five layers—two 1D convolutional layers with 2048 filters and kernel size 6, one max pooling layer with pool size 2, one LSTM with 256 hidden units, and one fully connected layer. To solve a problem with dying ReLU in 1D CNN layers, a parametric leaky ReLU activation function with He kernel initialization is used. The model was compiled with loss function mean squared error and optimizer Adam with a learning rate of 0.00004. The results show that the loss function for the training data is 1.7E-04. The validation data loss function measured by MSE is about 1.8E-04, measured by RMSE is 0.0036, which is an outstanding result. Future work will be to train the model with more data and receive better accuracy. After training, the model was saved to file for further use and predictions with new data. The model is fed with new data for six timesteps of open, high, low, and close prices to show prediction. The outcome predicts the following three 5-min interval timesteps for four features open, high, low, and close prices (Fig. 5). The result shows that the error of our model prediction with fresh, unseen data is from -0.00054 to 0.017336 and is smaller for the first of the three steps.

References

1. Harris, T.: Forex The Simple Strategy on Trading Currency Successfully. First Class Publishing, 2016.
2. Davison, N.: Forex trading 2020. Independently published, 2019.
3. Raeva, I.: Computation of risk in pricing of investment projects. AIP Conference Proceedings 2164, 060016. 2019.
4. Centeno, V., Georgiev, I., Mihova, V., Pavlov V.: Price forecasting and risk portfolio. AIP Conference Proceedings 2164, 060006. 2019.
5. Georgiev, I., Cenetno, V., Mihova, V., Pavlov, V.: A Modified Ordinary Differential Equation Approach in Price Forecasting. New Trends in the Applications of Differential Equations in Sciences. AIP Publishing, 2459, 2022. 030008-1–030008-7.
6. Lina, N., Li, Y., Wang, X., Zhang, J., Yu, J., Qi. C.: Forecasting of Forex Time Series Data Based on Deep Learning. 2018 International Conference on Identification, Information and Knowledge in the Internet of Things, IIKI 2018. 2019. 647–652.
7. Li, Y., Xie, Y., Yu, C., Yu, F., Jiang, B., Khushi. M.: Feature importance recap and stacking models for forex price prediction. 2021.
8. Markova, M. Convolutional neural networks for forex time series forecasting. New Trends in the Applications of Differential Equations in Sciences. AIP Publishing, 2459, 2022, 030024-9–030024-9. https://doi.org/10.1063/5.0083533

9. Shaojie, B., Kolter, J., Koltun, V.: An Empirical Evaluation of Generic Convolutional and Recurrent Networks. arXiv:1803.01271, 2018.
10. Hochreiter, J.,: Schmidhuber Long short-term memory. Neural Computation vol. 9, no. 8 (1997): 1735–1780.
11. Graves, A., Mohamed, A., Hinton. G.: Speech recognition with deep recurrent neural networks. IEEE International Conference on Acoustics, Speech and Signal Processing, May 2013. 2013. 6645–6649.
12. Pant, N.: A Guide For Time Series Prediction Using Recurrent Neural Networks (LSTMs). 7 9 2017. https://medium.com/cube-dev/time-series-prediction-using-recurrent-neural-networks-lstms-807fa6ca7f, last accessed 2022/07/07.
13. Leung, K.: The Dying ReLU Problem, Clearly Explained. https://towardsdatascience.com/. 31 03 2021. https://towardsdatascience.com/the-dying-relu-problem-clearly-explained-42d 0c54e0d24, last accessed 2022/07/07.
14. Geron, A.: Hand-on machine learning with Scikit-Learn, Keras & TensorFlow. Sebastopol, CA 95472: O'Reilly, 2019.
15. He, K., Zhang, X., Ren, S., Sun. J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. IEEE International Conference on Computer Vision. 2015. 1026–1034.
16. Brownlee, J.: Deep Learning for Time Series Forecasting Predict the Future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery, 2018.
17. Finam.ru, https://www.finam.ru/profile/forex/eur-usd/export/, last accessed 2022/07/07.

Diversification and Optimization of the Financial Portfolio in Times of Uncertainty

Ivan Georgiev, Victoria Deninska, Vesela Mihova, and Velizar Pavlov

Abstract In conditions of pandemic, economic crisis, or other force majeure circumstances that can lead to dynamics: increase or decrease in the prices of certain financial instruments, it is good to organize the investor's portfolio in many different types of financial instruments—shares, bonds, money-market funds, indices, crypto and fiat currencies, oil, and other commodities. Such a mixed portfolio is discussed in the presented paper, combining crypto and fiat currencies, indices, and fuels. The choice of instruments takes into account the instability of the markets and how these instruments behave against the background of world events. Weekly data for the period July 2020–December 2021 incl. has been used. With the help of ARIMA models, estimated asset prices for one period ahead have been derived. They have been used to calculate the estimated rates of return of the assets. The standard deviations of the rates of return and the correlation matrix between the rates of return of the considered financial instruments have been estimated using a historical data. The derived estimates have been mixed in an optimal risk portfolio, based on multi-objective optimization problem, which maximizes return and minimizes risk. Different scenarios for the full portfolio are considered, based on the coefficient of risk aversion.

Keywords Price forecasting · Risk diversification · Portfolio optimization

1 Introduction

Let us consider uncertainty as the inability to predict future events. Uncertainty about economic processes development is of a major concern for agents in the economy. Households and firms base their actions on how likely they think the economy is to grow, stagnate, or be in recession [1]. Moreover, economies and global events appear increasingly interconnected. Globalization definitely has its positive aspects, but when events such as pandemic, war, financial crisis, global recession, etc. break out, it often leads to increased government deficits and attempts by investors to move

I. Georgiev · V. Deninska · V. Mihova (✉) · V. Pavlov
University of Ruse, 8 Studentska Street, 7017 Ruse, Bulgaria
e-mail: vmicheva@uni-ruse.bg

their money out of equities to safer assets, such as precious metals, government bonds, and money-market instruments [2]. The transfer of investor capital from equities to safer financial instruments leads to stock market depreciation. The likelihood that an area of an investment portfolio will suffer serious losses is significant, so the key to a successful portfolio is that it is well diversified.

There are several major events that have put investors in times of uncertainty in recent months (and even years). On the one hand, in recent years, COVID-19 is causing disruption around the world, bringing unprecedented changes to daily life, local infrastructure, and the global economy. Some industries are highly vulnerable to the resulting economic changes, others not so much, but constant changes in restrictions, prevention, etc. bring uncertainty to us all. On the other hand, Russia's military invasion of Ukraine since 24 February 2022 has led to international sanctions on Russia, a spike in fuel prices, waves of refugees, and a complication of the international economic situation. The mentioned and some other events (that do not enjoy such media attention) affect financial markets, increasing their volatility. Under these circumstances, the present work proposes a portfolio organization consisting of eight different financial instruments, including crude oil, cryptocurrencies, fiat currencies, indices, and government bonds. The choice of instruments takes into account the instability of the markets and how these instruments behave against the background of world events. The estimated prices of the instruments for one period ahead are calculated using Autoregressive Integrated Moving Averages (ARIMA) models. Predicted rates of return on the assets are derived. Their standard deviations and the covariance matrix of the standard deviations are estimated using historical data. Based on this information, an optimal risk portfolio is constructed. The structure of this portfolio has been found using a code, derived for the purpose by the authors, that solves a multi-objective optimization problem.

2 Diversification

Given the international political and economic environment, the choice of financial instruments for the investment portfolio as well as its diversification is essential. Let us consider risk as the deviation in one or more returns of one or more future events from their expected value [3]. It is important that a portfolio contains many assets so that the negative deviations in some of them are offset by the positive ones in others. Alternative investments have proven to be an interesting hedge for diversification during recessions [4]. Such an alternative instrument is cryptocurrency, a form of payment that can circulate without the need for a central monetary authority such as a government or bank. It is naturally diversified—it represents wealth across all borders and is not tied to any one country's economy. In this paper, two cryptocurrencies have been used for investment portfolio compilation: CMC Crypto and Bitcoin USD. Another instruments, added to the constructed portfolio, are the fiat currencies TRY/USD, EUR/USD, and GBP/USD. Fiat currencies give central banks more control over the economy because they can control how much money is printed,

which in turn is used to deal with inflation and to influence aggregate demand and economic activity.

In the present work, due attention is also given to crude oil. Oil is the engine of the world economy and perhaps the most important commodity on the planet. Real oil prices can have an effect on all goods and services and on the general purchasing power of the population. It is influenced by geopolitical uncertainty and global economic data: volatile economic data can support oil and weak data can have a negative impact. Two indices, which could be used as indicators of how the U.S. economy is doing, are added to the portfolio. The first one is the Dow Jones Industrial Average (DJIA), which tracks the performance of 30 of the largest publicly traded companies on the New York Stock Exchange and Nasdaq. DJIA is price-weighted, meaning that stocks with a higher price have a greater weight in the index. The second one is the Standard & Poor's 500 (S&P 500), which is weighted by the market capitalization of the top 500 U.S. companies. The price behavior of these eight financial instruments is examined for the period July 2020–December 2021 inclusive, taking into account weekly closing price data. Instruments with different trends have been deliberately selected—this allows for short and long selling. In times of deep uncertainty about the direction of the economy and world politics, most investors are looking for the predictability of the Treasury market. Treasury bills are taken in this work as a non-risky asset. It is a matter of personal preference (to the level of risk aversion) how much of an investor's capital will be invested in the risky portfolio and how much in the risk-free asset [5]. Therefore, the focus of this work is on constructing an optimal risk portfolio, which can then be mixed with the risk-free asset, depending on the investor's individual preferences.

3   Price Modeling

Using the data described above, ARIMA models are built that give forecasts for the prices of financial instruments 1 period ahead (in this case, 1 week or for 03.01.2022). IBM SPSS software has been used for the purpose, and the selection of the models has been made from the Expert Modeler option of the software. Thus, for all instruments, the ARIMA (0, 1, 0) model is applied.

The graphs of auto-correlation functions (ACFs) and partially auto-correlation functions (PACFs) are examined for each model. For the instrument's currencies TRY/USD, EUR/USD GBP/USD, and S&P 500, there are no jumps outside the 95% confidence intervals for ACF and PCF, while for other assets there are 1 or at most 2 small jumps outside the confidence intervals for ACF or PACF. For the purposes of the study, these are good enough forecasts.

Table 1 shows the prices, predicted from the ARIMA (0, 1, 0) models for each financial instrument, the known prices for the previous period, and the expected rates of return (RoR), calculated as the difference between these prices. In the last column, the historical data for the standard deviations of RoRs is presented.

Table 1 Predictions, last known values, expected RoRs, and standard deviation of RoRs

Financial instrument	Estimated value	Last known value	Expected RoR (%)	Std. dev. RoR (%)
Crude oil	75.6531	75.2100	0.5891	6.1028
CMC crypto	1196.5174	1166.2900	2.5918	18.3224
Bitcoin USD	42,954.3994	41,911.6000	2.4881	10.2676
TRY/USD	0.0712	0.0721	−1.3087	6.9267
EUR/USD	1.1365	1.1364	0.0088	0.9472
GBP/USD	1.3602	1.3590	0.0883	1.1714
S&P 500	4787.1565	4766.1800	0.4401	3.1124
DIJA	36,473.0556	36,338.3000	0.3708	2.1133

4 Optimization

Consider a portfolio with n assets, where

r_i rate of return of the ith asset,
$E(r_i)$ expected rate of return of the ith asset, and
σ_i^2 dispersion of the return of the ith asset

Then the risk portfolio will have an expected rate of return corresponding to the mean value of the expected rates of return of the separate stocks, weighted with the corresponding proportion each of them takes in the portfolio [6]: $E(r_p) = \sum_{i=1}^{n} w_i E(r_i)$, where w_i is the weight of the ith stock.

The standard deviation of the risk portfolio is given by $\sigma_P = \sqrt{\sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1, i \neq j}^{n} \sum_{j=1}^{n} w_i w_j \sigma_i \sigma_j \rho(r_i, r_j)}$,

where $\rho(r_i, r_j)$ is the correlation coefficient between the ith and the jth stock rates of return.

Investors seek to both maximize expected return and minimize risk. This is in fact a difficult task, as often maximizing return comes at the expense of greater risk.

The P_1 portfolio dominates the P_2 portfolio if $E(r_{P_1}) \geq E(r_{P_2})$ and $\sigma_{P_1} \leq \sigma_{P_2}$.

A portfolio is called efficient if it dominates all possible portfolios for a given level of one of these metrics [7]. The assumption that investors are risk averse means that they prefer as little risk as possible for a given level of expected return and, out of two portfolios with identical returns, they will choose the one with less risk. The intersection of the set of portfolios with the maximum return and the set of portfolios with the minimum risk are called the efficient frontier.

The structure of an optimal portfolio of financial instruments could be computed based on any of the following optimization problems [8]:

1. Select a portfolio structure such that its risk is minimal with an expected return not lower than a given level E_c, ($E_c = const$):

$$\min \sigma_p = \sqrt{\sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1, i \neq j}^{n} \sum_{j=1}^{n} w_i w_j \sigma_i \sigma_j \rho(r_i, r_j)},$$

$$E(r_p) \geq E_c, \sum_{i=1}^{n} w_i = 1, w_i \geq 0.$$

2. Select a portfolio structure such that its expected rate of return is maximal with risk not greater than a given level σ_c, ($\sigma_c = const$):

$$\max E(r_p) = \sum_{i=1}^{n} w_i E(r_i), \sigma_p \leq \sigma_c, \sum_{i=1}^{n} w_i = 1, w_i = 0.$$

3. Select a portfolio structure that maximizes the risk premium at the best reward to risk ratio (S—the maximum slope of the capital allocation line CAL): $\max S = \frac{E(r_p) - r_f}{\sigma_P}$, $\sum_{i=1}^{n} w_i = 1, w_i \geq 0$.

Which problem the investor prefers depends on his objectives. In this paper, the three problems are combined into a multi-objective optimization problem. Several objective functions (private criteria), which are usually written as elements of a vector [9], have to be minimized:

$$\min_{w \in W} Z(W), \tag{1}$$

$$\sum_{i=1}^{n} w_i = 1, \tag{2}$$

$$w_i \geq 0. \tag{3}$$

In the considered problem $n = 3$ and:

$$Z(W) = [Z_1(W), -Z_2(W), -Z_3(W)]^T, \ W = [w_1, w_2, \dots w_n]^T,$$

$$Z_1 = \sigma_p, Z_2 = E(r_p), Z_3 = \frac{E(r_p) - r_f}{\sigma_p}.$$

A minimum is chased over the criteria $Z_1(W)$, while on $Z_2(W)$ and $Z_3(W)$ a maximum is required, which explains the negative sign in front of these functions in the vector criterion. A characteristic feature of multi-objective problems is the absence of a unique optimal solution. Usually solving them results in a set of so-called Pareto-optimal solutions obtained using the principle of optimality proposed by Vilfredo Pareto [10]. Since none of them is better than the others (if no additional considerations are taken into account), it is appropriate to find as many Pareto-optimal solutions as possible.

The goal in this work is to find the Pareto front of optimal solutions to problems (1)–(3) and possibly to isolate one particular solution that reflects to some extent the authors' subjective view of the trade-off. The vector criterion (1) is linear and the constraints (2)–(3) are linear implying that the admissible set is convex. From the fact that the admissible set (2)–(3) is convex, the weighting method proposed by Zadeh [11] will be applied to find the Pareto-optimal solutions. A generalized criterion is constructed as a linear combination of the partial criteria with weights λ_j, $j = 1, 2, 3$ and a condition on the weights

$$\sum_{j=1}^{3} \lambda_j = 1, \lambda_j \geq 0, j = 1, 2, 3. \tag{4}$$

Uniformly distributed random numbers satisfying condition (4) have been assigned to the weights λ_k. For each random set of weights, the single-criterion integer problem (5) is solved:

$$\min Z = \lambda_1 Z_1 + \lambda_2(-Z_2) + \lambda_3(-Z_3) \tag{5}$$

with constraints (2) and (3).

Different solutions are selected. They form the searched Pareto front.

The choice of a particular solution is subjective and depends very much on the expertise of the particular investor or portfolio manager (the person who decides on the risk-reward trade-off). In the present work, the global criterion method [9] is applied to select a particular solution from the Pareto front. In this method, the distance between some given point $v^\circ = \left[v_1^\circ, v_2^\circ, \ldots, v_n^\circ\right]^T$ in the criterion space and the points in the reachable set is minimized. Usually, the utopian point U is chosen as the point v°. In this case, problem (1) is transformed into form (6):

$$\min_{w \in W} d(W), \tag{6}$$

where

$$d(W) = \left(\sum_{j=1}^{3} k_j \left[Z_j(W) - v_j^\circ\right]^p\right)^{\frac{1}{p}}. \tag{7}$$

If all criteria are equally important, the weighting factors are assumed to be the same: $k_j = 1$. When $p = 1$, the generalized criterion (7) is a linear combination of the components of the vector $|Z_j(W) - v_j^\circ|$. The magnitude d at $p = 2$ coincides with the distance between two points of the three-dimensional Euclidean space. The solution obtained by this scheme is Pareto optimal and uniformly approximates all the partial criteria $Z_j(W)$ to their ideal values v_j°.

In this paper, all weights of the private criteria are equal $\left(k_j = 1\right)$ and $p = 2$. The problem is solved using the MATLAB product. The input data consist of: vector with the predicted RoRs of the financial instruments; vector with the standard deviations of the RoRs; correlation matrix of assets' RoRs; RoR on a risk-free asset; and number of points on the Pareto front we want to see. The output data are: matrix with the weight distributions for the portfolios, where each row corresponds to one point on the Pareto front; matrix with values for the three criteria, where each row corresponds to one point on the Pareto front; the closest to the utopian point that lies on the Pareto front, whose coordinates appear to be the values of the private criteria, as well as the portfolio distribution for which these criteria values are obtained. The problem is solved under the following assumptions: risk-free stocks exist; borrowing is possible at a risk-free rate; and short sales of risky stocks are allowed. Treasury bills are used as a risk-free asset with an annual yield of 1.512%, which equates to a 0.0291% (or 0.000291) weekly yield (1.512/52).

The expected RoRs and the historical data for the standard deviations (Table 1) are used along with the correlation matrix (Table 2) in order to compile an optimal risk portfolio of the eight assets. The instruments with negative expected rate of return are sold in a short position. Figure 1 shows some points, part of the Pareto frontier, the utopian point, the closest point to the utopian, and the distance between them.

The results for the optimal risk portfolio obtained using the closest point to the utopian are presented in Fig. 2. The shares of TRY/USD are hold in a short position, while the shares of the rest financial instruments are taken in a long position. This distribution leads to a risk portfolio with an expected rate of return of 0.4605%, and a standard deviation estimate of 1.0371%. Slope of capital allocation line in this case is S = 0.4160.

Table 2 Correlation matrix

Instrument	Crude oil	CMC crypto	Bitcoin USD	TRY/USD	EUR/ USD	GBP/USD	S&P 500	DIJA
Crude oil	1.00	−0.23	−0.09	0.03	−0.08	0.19	−0.18	0.07
CMC crypto	−0.23	1.00	0.05	−0.10	0.02	−0.09	0.12	0.03
Bitcoin USD	−0.09	0.05	1.00	−0.05	−0.04	−0.10	−0.07	−0.02
TRY/ USD	0.03	−0.10	−0.05	1.00	−0.05	−0.07	−0.05	0.00
EUR/ USD	−0.08	0.02	−0.04	−0.05	1.00	0.24	−0.10	0.06
GBP/ USD	0.19	−0.09	−0.10	−0.07	0.24	1.00	−0.14	−0.13
S&P 500	−0.18	0.12	−0.07	−0.05	−0.10	−0.14	1.00	0.78
DIJA	0.07	0.03	−0.02	0.00	−0.06	−0.13	0.78	1.00

Fig. 1 Pareto frontier, utopian solution, and closest point to the utopian
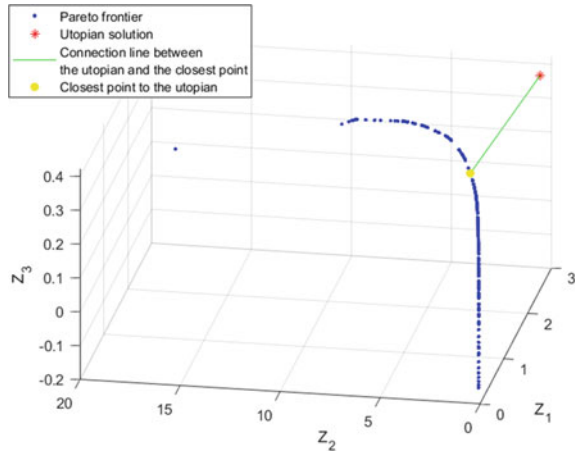
Fig. 2 Shares in the optimal risk portfolio obtained using the closest point to the utopian
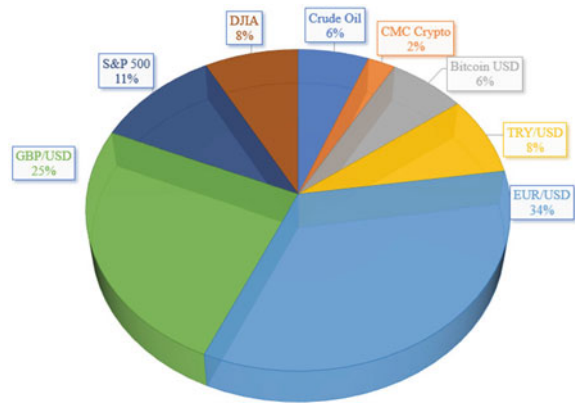
Column 3 from Table 3 presents the share and the position (long (+) and short (−)) for each financial instrument in the organized portfolio. Last column, which also shows (+) and (−), presents the real movement of the closing prices. The difference between the last known price and actual price for each financial instrument has been calculated. Taking into account the positions (long, short) and the percentage distribution of the portfolio, it is possible to calculate what would be the real percentage return. The trends of Crude Oil, Bitcoin USD, EUR/USD, and GBP/USD are correctly predicted and they bring profit, while the trend of the rest instruments are not correctly predicted and they bring loss. The actual percentage return, obtained with the presented approach, is calculated as follows:

$$\sum_{i=1}^{8} \varepsilon_i w_i \frac{|P_{i,n+1} - P_{i,n}|}{P_{i,n+1}} * 100$$

$$= \left[0.0593 * \frac{|78.900 - 75.2100|}{78.900} + \cdots + (-1) * 0.0789 * \frac{|36231.6600 - 36338.3000|}{36231.6600} \right] * 100 = 0.1049\%$$

Table 3 Portfolio weights, estimated, last known, and actual values

Financial instrument	Estimated value—ARIMA	Portfolio weights	Last known price	Actual price
Crude Oil	75.6531	5.93 (+)	75.2100	78.9000 (−)
CMC Crypto	1196.5174	2.30 (+)	1166.2900	1039.2100 (−)
Bitcoin USD	42,954.3994	6.49 (+)	41,911.6000	43,113.8800 (+)
TRY/USD	0.0712	7.92 (+)	0.0721	0.0736 (+)
EUR/USD	1.1365	33.98 (+)	1.1364	1.1416 (+)
GBP/USD	1.3602	24.94 (+)	1.3590	1.3677 (+)
S&P 500	4787.1565	10.54 (+)	4766.1800	4677.0300 (−)
DIJA	36,473.0556	7.89 (+)	36,338.3000	36,231.6600 (−)

where:

$P_{i,n}$ last known price for instrument i (opening price for the position),
$P_{i,n+1}$ actual price for instrument i (closing price for the position), and
ε_i takes values 1 or −1 depending on whether the direction of the trend is correctly predicted.

The resulting return is on a weekly basis. If the investor manages to keep his portfolio return at this level within 1 year, it will equate to an annual return of 5.457%. For comparison, the annual return on the risk-free asset is 1.512%, indicating that the annual risk premium would be 5.457% − 1.512% = 3.945%.

5   Conclusions

This paper examines different types of financial instruments, their volatility, and their combination in a well-diversified portfolio. The aim is to reduce risk given the complicated international economic environment. ARIMA models are used to forecast the closing prices of the assets. Expected rates of return on assets for one period ahead are derived from the models. These were used as inputs to construct an optimal risk portfolio, together with historical data on standard deviations and correlation matrices of instrument rates of return. To find the portfolio, three of the most common optimization criteria are assumed and combined in a multi-objective optimization problem. This problem involves finding a Pareto front and selecting a specific point on that front. A set of portfolios corresponding to the Pareto front is found and a specific point among them is selected with the help of a code developed for this purpose in MATLAB. Using this point, the structure of the optimal portfolio is obtained. Taking into account the positions (long, short), the percentage distribution of the portfolio, the last known, and the actual prices, the real percentage of the return is calculated. On an annualized basis, this percentage equates to 5.457%, resulting in a risk premium of 3.945%.

References

1. Erdoğan, A., Hülagü, T.: Measuring Uncertainty In The Financial Sector. In: ISI REGIONAL STATISTICS CONFERENCE 2014.
2. Investopedia Homepage, https://www.investopedia.com/, last accessed 2022/06/10.
3. Raeva, I.: Computation of risk in pricing of investment projects. In: Application of Mathematics in Technical and Natural Sciences, vol. 2164, p. 060016, AIP Publishing (2019).
4. Milinchuk, A.: Investing During Times Of Uncertainty. Forbes Finance Council, (12, 2021).
5. Mihova, V., Centeno, V., Georgiev, I., Pavlov, V.: An Application of Modified Ordinary Differential Equation Approach for Successful Trading on the Bulgarian Stock Exchange. In: New Trends in the Applications of Differential Equations in Sciences, vol. 2459, pp. 030025–1–9, AIP Publishing (2022).
6. Centeno, V., Georgiev, I., Mihova, V., Pavlov, V.: Price forecasting and risk portfolio optimization. In: Application of Mathematics in Technical and Natural Science, vol. 2164, pp. 060006–1–15, AIP Publishing (2019).
7. Bodie, Z., Kane, A., Marcus, A.: Investments. 10th global ed. McGraw-Hill Education, Berkshire (2014).
8. Dochev, D., Petkov, I.: Theory of Decision Making. Science and Economics, Economic University – Varna, Varna (2008).
9. Vitliemov, V.: Optimization Methods.2nd ed. University of Ruse, Ruse (2004).
10. Pareto, V.: Notice biographique. Centre d'Etudes Interdisciplinires Walras (2003).
11. Miettinen, K.: Nonlinear Multiobjective Optimization. Kluwer Academic Publishers, Boston, MA (1999).

Simulating Stochastic Differential Equations in Option Pricing

Velizar Pavlov and **Teodora Klimenko**

Abstract Investors use financial derivatives to increase the expected return and reduce the risk associated with an investment. Options are very attractive since they offer limited risk. Due to their intensive use in risk management and portfolio hedging, many researchers develop models that accurately price options. In this paper, we consider the best known option pricing model, the Black–Scholes model, and then apply numerical modeling techniques to stochastic differential equations, namely, the Euler–Maruyama and Milstein methods, and further use Monte Carlo simulations. Next, we prove that the numerical approximations do converge in a strong sense to the price of European call options, and we show that the empirically computed convergence rate approaches the theoretical convergence rate for both the Euler–Maruyama and Milstein methods.

Keywords Stochastic differential equations · Monte Carlo · Option pricing

1 Introduction

Over the past few years, derivative securities (options, futures and forward contracts, swaps, warrants) have become essential instruments for both speculators and investors, regardless of their corporate profile. Derivatives facilitate the transfer of financial risks. As such, they can be used to hedge risky positions or to take risks in anticipation of gains [1].

Options are essentially contracts between two parties which give their holders the right, but not the obligation to buy or sell a prescribed amount of an underlying asset at a specified price within a specified time period. The value of the option is tied to the value of the underlying asset, which can be stocks, bonds, currencies, interest

V. Pavlov (✉) · T. Klimenko
Department of Applied Mathematics and Statistics, University of Ruse, 8 Studentska Str, 7004 Ruse, Bulgaria
e-mail: vpavlov@uni-ruse.bg

T. Klimenko
e-mail: tediklimenko@abv.bg

rates, market indices, exchange traded funds, futures contracts, other derivatives, and so on. Options themselves are securities, like stocks or bonds, and it is because they derive their value from something else, they are called derivatives [2].

The price of an option is constituted by two components: The first one is the internal value, which is defined as the difference between the market price of the underlying asset and the strike price of an option. The second one is the time value, which through a varying, nonlinear relationship reflects the expected value at the maturity date discounted to the present [3].

The pricing of options and other financial instruments is one of the most important problems in finance, since its purpose is to determine the fair price of a security with respect to more liquid securities whose price is determined by the law of supply and demand (most commonly priced are "vanilla" and "exotic" options, convertible bonds, etc.) [4]. To determine the option price, one must first choose a model of the price dynamics of the underlying asset, usually in the form of ordinary differential equations [5, 6] or stochastic differential equation depending on market parameters such as interest rate, volatility of the asset, etc., then find a solution to the chosen model in some form, and finally obtain the required option price. The volatility of the asset can be constant (Black–Scholes model) [7], may depend on the value of the underlying asset (volatility smile), the time (volatility term structure) [8], on both (local volatility model) or may satisfy some other stochastic differential equation (stochastic volatility model) [9]. An exact analytical formula for the European vanilla option price has been derived from the Black–Scholes model and several other local and stochastic volatility models, but in general, approximations or numerical methods must be used to estimate options (when it is difficult or impossible to use other approaches) [10]. To solve the problem, we use the Monte Carlo method to construct the empirical density of the underlying asset distribution and then determine the option value [11–13]. Possible applications of local volatility models with parallel computation include modeling and managing large equity portfolios [14, 15], assessing and managing market and credit risks [16, 17].

Option pricing via Monte Carlo simulation is among the most popular ways to value certain types of financial options (e.g., exotic path-dependent derivatives, such as Asian options) by directly simulating several thousand possible (but random) price paths of the underlying asset with the corresponding option strike price for each path and then calculating the average outcome of the process discounted to the present day and that is the option value today [18]. This is the first from the two main Monte Carlo approaches [19], while the other is based on Monte Carlo approximation of multidimensional integrals. Monte Carlo simulation is essentially useful in evaluating derivative contracts [20].

The method can guarantee that with a certain probability the Monte Carlo approximation error is less than a given value (probabilistic error). Monte Carlo methods approach the solution faster than numerical integration methods, require less memory, have fewer data dependencies, and are easier to program [21].

In this paper, various simulations of Geometric Brownian Motion paths are considered, the approximation error of the Euler–Maruyama and Milstein methods is

evaluated, and the convergence rate of the two methods is empirically determined. Employing these simulations, the value of European and Asian options is obtained.

2 Simulations by the Euler–Maruyama and Milstein Methods

In order to simulate the price of a European call option, one must first determine the process that the price of the underlying asset follows over the life of the option t ∈ [0, T]. The underlying asset is known to follow the Geometric Brownian Motion (GBM) given by the stochastic differential Eq. (1), widely used in the study and modeling of dynamical systems that are subject to various random disturbances. The price of the underlying asset S_t is assumed to have an expected return μ and a constant volatility σ:

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \tag{1}$$

The Euler–Maruyama (2) and Milstein (3) techniques are commonly used to approximate the solution of stochastic differential equations, where μ represents the mean return, σ is the volatility, h is the size of the discretization time step, and ΔW_i is the increment of the Brownian motion at each time step. The latter is simulated by generating random numbers with a standard normal distribution (4):

$$S_{i+1} = S_i + \mu S_i h + \sigma S_i \Delta W_i, \tag{2}$$

$$S_{i+1} = S_i + \mu S_i h + \sigma S_i \Delta W_i + \frac{\sigma^2 S_i}{2}\left(\Delta W_i{}^2 - h\right), \tag{3}$$

$$\Delta W_j = W_{t_{j+1}} - W_{t_j} = Z\sqrt{\Delta t}\,\text{with}\,Z \sim \text{N}(0, 1). \tag{4}$$

As can be seen from the formulas above (Formulas (2) and (3)), the Milstein method is almost the same as the Euler–Maruyama method, but here in order to obtain a higher accuracy of approximation, an additional term is added by including one of the double integrals from the Itô–Taylor expansion for Itô processes.

$$X_t = X_{t_0} + \int_{t_0}^{t} \{a(X_{t_0}) + \int_{t_0}^{s} [a'(X_z)a(X_z) + \frac{1}{2}a''(X_z)b^2(X_z)]dz$$

$$+ \int_{t_0}^{s} a'(X_z)b(X_z)dW_z\}ds + \int_{t_0}^{t} \{b(X_{t_0})$$

$$+ \int_{t_0}^{s} [b'(X_z)a(X_z) + \frac{1}{2}b''(X_z)b^2(X_z)]dz$$

$$+ \int_{t_0}^{s} b'(X_z)b(X_z)dW_z\}dW_s$$

$$=> X_t - X_{t_0} = a(X_{t_0})(t - t_0) + b(X_{t_0})\int_{t_0}^{t} dW_s + R.$$

If R is neglected, the Euler–Maruyama method is obtained (2).
It can be found that the dominant term in the residual R is

$$\int_{t_0}^{t_0+h} \int_{t_0}^{s} dW_z dW_s \approx \mathcal{O}(h)$$

Then:

$$\int_{t_0}^{t} \int_{t_0}^{s} b'(X_z)b(X_z)dW_z dW_z \approx b'(X_z)b(X_z)\int_{t_0}^{t} \int_{t_0}^{s} dW_z dW_s = \frac{b'b}{2}((\Delta W)^2 - h)$$

$$=> X_t - X_{t_0} = a(X_{t_0})(t - t_0) + b(X_{t_0})\int_{t_0}^{t} dW_s + \frac{b'b}{2}\int_{t_0}^{t}\int_{t_0}^{s} dW_z dW_s + \tilde{R}.$$

If $\tilde{R}$ is neglected, the Milstein method (3) is obtained.

Geometric Brownian Motion.

Let the following parameter values be given as input to the simulations:

$$\mu = 0.06; \sigma = 0.3; S_0 = 50; h = 10^{-2}; t \in [0, 1],$$

where S_0 is the price of the underlying asset today; and the output is $S(t)$, which represents the dynamics of the underlying asset price for the given realization of the Brownian motion in the interval $[0, T]$, where T is the maturity date of the option.

Fifty paths are plotted using the Milstein (Fig. 1) method. On the graphs, the abscissa represents the time points $[t]$ and the ordinate $[S]$ is the realization of the price of the underlying asset.

Let us calculate the exact analytical solution of the Geometric Brownian Motion by considering the step size $h = 10^{-i}$ for $i = 1, 2, 3, 4, 5$ using formula (5), where t is a specified time moment in the future and W is the Wiener process itself, which
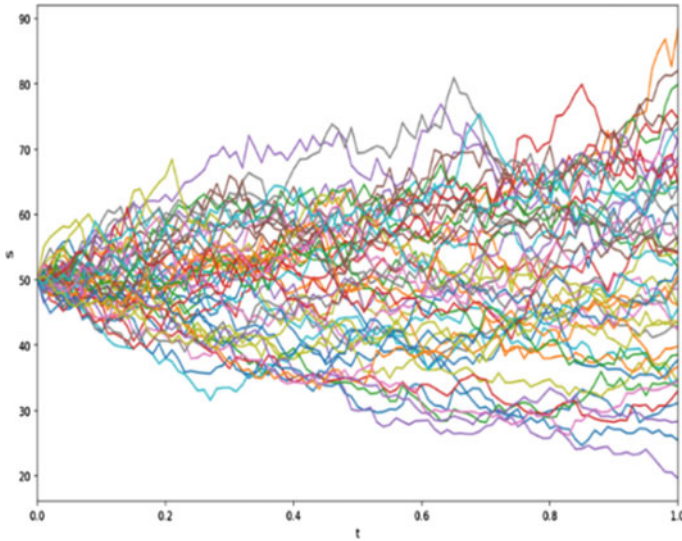
Fig. 1 Stock price dynamics for $h = 0.01$ using Milstein method

is the accumulated sum of the increments dW up to the time instant t.

$$S(t) = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W}. \tag{5}$$

For each step size h, this exact solution $(S_{T,k})$ is compared with the approxima-tions $(\hat{S}_{T,k})$ from the Euler–Maruyama and Milstein methods, and the error $\hat{\epsilon}(h)$ is computed using the following formula:

$$\hat{\epsilon}(h) := \frac{1}{M} \sum_{k=1}^{M} \left| S_{T,k} - \hat{S}_{T,k} \right|.$$

Here, we only need the price of the underlying asset at the maturity date T, which is why only the last calculated value from the exact solution and the approximation is taken into account to estimate the error. So, to get the average error, we first take the sum of the absolute errors for each path and then divide this sum by M, the number of paths we have simulated.

Table 1 shows the approximation errors for the price of the underlying asset at maturity date T using the Euler–Maruyama (**Err_EM**) and Milstein (**Err_M**) methods, where h is the time step size.

As can be seen from the table, when the step size h decreases, the errors of both methods also decrease. At the largest step size $h = 10^{-1}$, the errors are also the largest, and vice versa. It can be noted that the Milstein method is more accurate and

Table 1 Approximation error by the Euler–Maruyama and Milstein methods

h	Err_EM	Err_M
0.10000	0.814502	0.077256
0.01000	0.288945	0.009201
0.00100	0.080778	0.001012
0.00010	0.026781	0.000088
0.00001	0.008354	0.000009

the errors are smaller than those of the Euler–Maruyama method, just because it has a higher order of convergence.

The quality of any algorithm that approximates the true value of the solution depends heavily on the convergence rate. It is necessary to evaluate how fast the approximation converges to the true solution. To evaluate the strong convergence rate of each method, we take the decimal logarithm of the step sizes $h = 10^{-i}$ for $i = 1,2,3,4,5$ and the logarithm of the errors computed for the two methods, Euler–Maruyama and Milstein.

As can be seen from Fig. 2, the estimated error for the Milstein method is not only smaller than the error for the Euler–Maruyama method, but also decreases much faster (the slope is steeper). Therefore, at $h = 10^{-1}$, it can be seen that the difference between the two methods is smaller than the difference for $h = 10^{-5}$, because the convergence order of the Milstein method is higher than that of the Euler–Maruyama method. Applying linear regression, we find the strong convergence rate for both methods. It can be seen that the results converge to the theoretical values of the convergence orders: ½ for the Euler–Maruyama method and 1 for the Milstein method.

3 Valuation of European and Asian Options

In order to find the fair value of the European call option, in addition to the above values of the Geometric Brownian Motion parameters, we set the strike price of the option $K = 50$ and the interest rate $r = 0.05$. We apply the Black–Scholes formula:

$$V^{\text{EC}} = S_t N(d_1) - K e^{-r(T-t)} N(d_2),$$

where

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{s^2}{2}} \, ds,$$

$$d_1 = \frac{\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}},$$
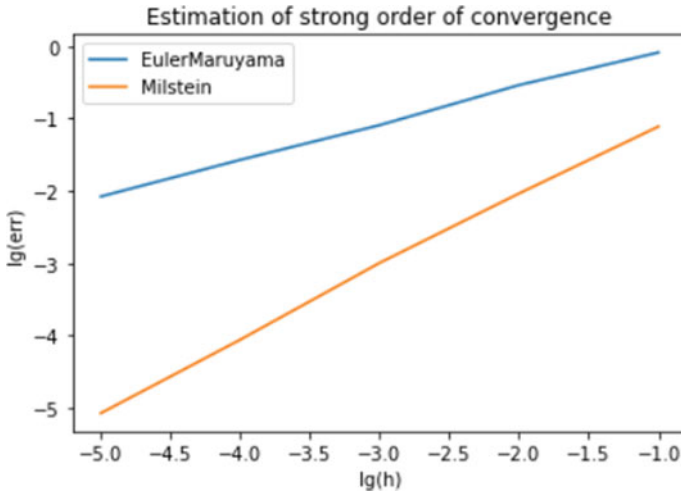
Fig. 2 Empirical estimation of the strong convergence rate of the Euler–Maruyama and Milstein methods

$$d_2 = \frac{\ln\left(\frac{S_t}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)(T - t)}{\sigma\sqrt{T - t}}.$$

To approximate the price of the European call option, the Euler–Maruyama and Milstein methods are applied, and for this purpose, firstly μ must be replaced by r, i.e., the price of the underlying asset S must have an expected return μ equal to the risk-free rate r. So, it is obtained the *risk-neutral* stochastic differential equation:

$$dS_t = r S_t dt + \sigma S_t dW_t.$$

To estimate today's option price V_0, we take the expected value of the payoff function under a risk-neutral measure $\mathbb{Q}$ discounted to today's point in time:

$$V_0^{EC} = e^{-rT} * \mathbb{E}^{\mathbb{Q}}[S_T - K]^+.$$

To approximate the expected value of the payoff function under the risk-neutral measure, the Monte Carlo simulation method is used, where M paths are simulated and, using the index i, a certain path from 1 to M is fixed. The payoff function $[S_T^i - K]^+$ is calculated for each simulation by taking the last estimated price of the underlying asset for each path, i.e., at the maturity date T. Then, the sum of the payoff functions for all paths is divided by the number of simulations M to obtain the average value of the payoff function, which is multiplied by e^{-rT} to discount to today's point in time.

Thus, the fair value of the option today is found:

$$\tilde{V}_0^{EC} = e^{-rT} * \frac{1}{M} * \sum_{i=1}^{M} [S_T^i - K]^+ \tag{6}$$

Finally, the results obtained by Euler–Maruyama and Milstein methods are compared with the Black–Scholes solution for different number of simulated paths $M = \{10,100,1000,10,000\}$ and fixed time step size $h = 10^{-4}$. For this purpose, the absolute error of the estimated option price (6) and the exact Black–Scholes solution is computed for both methods:

$$\left| \tilde{V}_0^{EC} - V_0^{EC} \right|. \tag{7}$$

Similarly, the values of the Asian call option are calculated. It is known that the Asian arithmetic option does not possess an exact solution. It differs from its European counterpart in the way its payoff is calculated. The Asian option belongs to the class of the *path-dependent* options since its payoff depends on the realization of the underlying price S_t for the whole life of the option $t \in [0, T]$:

$$V_0^{AC} = e^{-rT} * \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{J} \sum_{j=1}^{J} S_{t_j} - K \right]^+, \tag{8}$$

where $\{t_j\}_{j=1}^{J}$ is a discrete set with observation times (for a discretely monitored Asian option).

By the same argument, the Monte Carlo approximation to the Asian call premium is given by averaging the simulations, M in number:

$$\tilde{V}_0^{AC} = e^{-rT} * \frac{1}{M} * \sum_{i=1}^{M} \left[\frac{1}{J} \sum_{j=1}^{J} S_{t_j}^i - K \right]^+. \tag{9}$$

Table 2 provides an estimate of the fair price of a European call option using the Black–Scholes method, an approximation of the option price using the Euler–Maruyama and Milstein methods, and an estimate of the approximation error. The Euler–Maruyama and Milstein approximations (9) of the fair value of an Asian call option are given as well, but since there is no exact formula for (8), the approximation error cannot be obtained.

The first column of Table 2 is the number of simulated paths M, but as can be seen, the exact option price (in the **C_BS** column) does not depend on the number of simulations M, so here in the table it is the same for each $M = [10,100,1000,10000]$.

C_EM is the approximated European call option price calculated using the Euler–Maruyama method, and **Err_EM** is the error of this approximation. **C_M** is the approximated price of the European call option calculated using the Milstein

Table 2 Estimation of the fair value of the European and Asian call option

M	C_BS	C_EM	Err_EM	C_M	Err_M	A_C_EM	A_C_M
10	7.115627	1.249625	5.866003	1.243205	5.872422	0.565354	0.566168
100	7.115627	4.558837	2.556790	4.561021	2.554606	2.456645	2.458124
1000	7.115627	6.867675	0.247952	6.867469	0.248158	3.822855	3.822648
10,000	7.115627	6.988541	0.127087	6.988551	0.127076	3.942129	3.941976

method and **Err_M** is the error of this approximation. **A_C_EM** and **A_C_M** are the approximated prices of the Asian option by the both methods, respectively.

As can be seen from Table 2, when only 10 paths are simulated, the approximated price is far from the theoretical price, but as M increases, the approximation results get closer and closer to the exact European call option price. The error decreases as M increases, and, for example, by the Euler–Maruyama method from 5.87 it drops to 0.13. The error from the Milstein method is plotted on Fig. 3a. Also, here it can be seen that there is no significant difference between the results obtained by the Euler–Maruyama and the Milstein methods since the number of simulated paths M is too small. Therefore, to obtain better results when using the Monte Carlo method, possibly many more paths should be simulated.

The same results apply for the put option valuation (Table 3). We apply the Black–Scholes formula to find the exact European put option value:

$$V^{EP} = Ke^{-r(T-t)}N(-d_2) - S_tN(-d_1),$$

where $N(x)$, d_1 and d_2 are the same as in V^{EC}. Its exact value is $V^{EP} = 4.677099$.

The put premium approximations and the difference between them are calculated analogously to (6) and (7), but only the call payoff is replaced by the put payoff $[K - S_T]^+$:

(a)

(b)
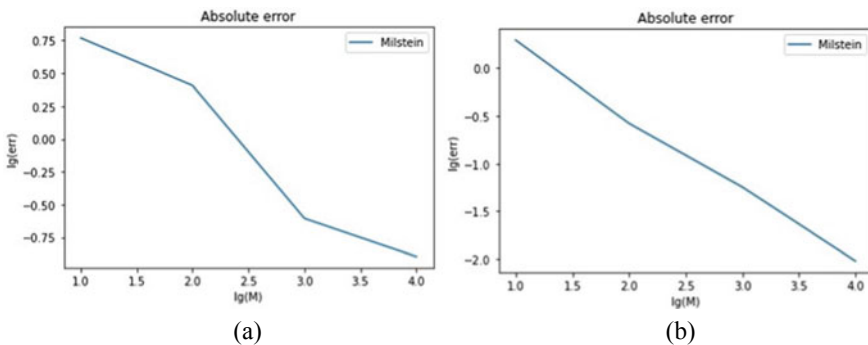
Fig. 3 Absolute error in the option valuation, associated with the Milstein method. Call option error (**a**) and put option error (**b**)

Table 3 Estimation of the fair value of the European and Asian put option

M	P_EM	Err_EM	P_M	Err_M	A_P_EM	A_P_M
10	6.648960	1.971861	6.647441	1.970342	5.153849	5.154948
100	4.412524	0.264574	4.412199	0.264900	2.756057	2.756767
1000	4.734758	0.057659	4.733905	0.056806	2.902844	2.902490
10,000	4.667667	0.009432	4.667551	0.009547	2.768098	2.768045

$$\widetilde{V}_0^{\ EP} = e^{-rT} * \frac{1}{M} * \sum_{i=1}^{M} \left[K - S_T^i \right]^+.$$

In similar manner, for the valuation of the Asian put option, the payoff is replaced by $\left[K - \frac{1}{J} \sum_{j=1}^{J} S_{t_j} \right]^+$:

$$\widetilde{V}_0^{\ AP} = e^{-rT} * \frac{1}{M} * \sum_{i=1}^{M} \left[K - \frac{1}{J} \sum_{j=1}^{J} S_{t_j}^i \right]^+.$$

The same implications could be drawn for the pricing of the put options. In that case, the results seem to be more precise, but this phenomenon is due to the randomness of the simulations. It is also visible that the more simulations are performed, the more accurate the results are (Fig. 3b).

4 Conclusion

In this study, we have shown the advantages of applying the Monte Carlo simulation approach for valuation of European and Asian options. Firstly, the Euler–Maruyama and Milstein discretization methods are introduced, and they are employed in simulating trajectories of the Geometric Brownian Motion. Then, we proposed a simple way to empirically estimate the strong convergence rates for the both methods, which fully coincide with the theoretical ones. This is done by calculating the approximation error of the simulation of the underlying asset price development. Furthermore, the Black–Scholes formula is given in brief and it is explained how to estimate the approximation error of the option valuation through Monte Carlo simulations. Next, the fair values of European and Asian options are calculated and the accuracy dependence on the number of simulations is commented. For the European options, the error could be computed, but unfortunately this is not the case with the Asian options, since they do not experience an exact formula. The aforementioned calculations are done for call and put options. Finally, some implications are outlined which are of help when estimating vanilla and exotic options via Monte Carlo simulations for a wide variety of purposes from hedging to portfolio management.

References

1. Seydel, R.: Tools for Computational Finance. Springer-Verlag London Limited (2012).
2. Øksendal, B.: Stochastic Differential Equations: An introduction with Applications. Springer-Verlag Berlin-Heidelberg-New York (2003).
3. Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus. Springer Science+Business Media, Inc. (1991).
4. Hirsa, A.: Computational Methods in Finance. CRC Press (2013).
5. Mihova, V., Centeno, V., Georgiev, I., Pavlov, V.: An application of modified ordinary differential equation approach for successful trading on the Bulgarian Stock Exchange. In New Trends in the Applications of Differential Equations in Sciences, vol. 2459, pp. 030025–1–9, AIP Publishing (2022).
6. Georgiev, I., Cenetno, V., Mihova, V., Pavlov, V.: A modified ordinary differential equation approach in price forecasting. In New Trends in the Applications of Differential Equations in Sciences, vol. 2459, pp. 030008–1–7, AIP Publishing (2022).
7. Black, F., Scholes, M.: The pricing of options and corporate liabilities. Journal of Political Economy 81, 637-659 (1973).
8. Georgiev, S., Vulkov, L.: Computational recovery of time-dependent volatility from integral observations in option pricing. Journal of Computational Science 39, 101054 (2020).
9. Günther, M., Jüngel, A.: Finanzderivative mit MATLAB. Vieweg+Teubner, 2nd ed (2010).
10. Föllmer, H., Schied, A.: Stochastic Finance: An introduction in Discrete Time. De Gruyter Studies in Mathematics (2011).
11. Andersen, L., Piterbarg, V.: Interest Rate Modeling. Volume 1: Foundations and Vanilla Models. Atlantic Financial Press (2010).
12. Petters, A., Dong, X.: An Introduction to Mathematical Finance with Applications. Understanding and Building Financial Intuition, Springer Undergraduate Texts in Mathematics and Technology (2018).
13. Teng, L.: Lecture Notes on Computational Finance 1. Bergische Universität Wuppertal (2021).
14. Kostadinova, V, Georgiev, I., Mihova, V., Pavlov, V.: An application of Markov chains in stock price prediction and risk portfolio optimization. In New Trends in the Applications of Differential Equations in Sciences, vol. 2321, pp. 030018–1–11, AIP Publishing (2021).
15. Centeno, V., Georgiev, I., Mihova, V., Pavlov, V.: Price forecasting and risk portfolio optimization. In Application of Mathematics in Technical and Natural Sciences, vol. 2164, pp. 060006–1–15, AIP Publishing (2019).
16. Mihova, V., Pavlov, V: A customer segmentation approach in commercial banks. In Application of Mathematics in Technical and Natural Sciences, vol. 2025, pp. 03003–1–9, AIP Publishing (2018).
17. Mihova, V., Pavlov, V.: Comparative analysis on the probability of being a good payer. In Application of Mathematics in Technical and Natural Sciences, vol. 1895, pp. 050006–1–15, AIP Publishing (2017).
18. Minkova, L.: Stochastic Analysis and Applications. Lecture notes, University of Ruse (2015).
19. Dimov, I.: Monte Carlo Methods for Applied Scientists. World Scientific, New Jersey, London, Singapore (2008).
20. Todorov, V., Georgiev, S.: A stochastic optimization method for European option pricing. Annals of Computer Science and Information Systems (2022).
21. Shreve, S.: Stochastic Calculus for Finance II: Continuous-Time Models. Springer Finance (2013).

Comparison of the Growth Between the Number and the Payments of IBNR Claims with Chain-Ladder Method

Elitsa Raeva and Velizar Pavlov

Abstract The problem of estimation of the claims which incur but are not reported (IBNR) in the same calendar year is one of the toughest tasks, which the actuaries are dealing with. The main tool that is used in practice is the Chain-Ladder Method (CLM). It is often preferred because there are only two propositions which are necessary and because it presents an algorithm, which could be easily implemented in different software. Another advantage to use CLM is the suitable way of representing the sample data. Using this advantage, a comparison between the development growth of the number of the claims and their cost value is made. To make this comparison, the data is collected using a sequence of development triangles for longer period. In this paper, an idea for analyzing the behavior of the data inside the CLM and how it depends or changes over the development year of the claim settlement is given.

Keywords Insurance · IBNR claims · Chain-ladder method · Run-off triangle

1 Introduction

The Chain-Ladder method is one of the most useful tools in the actuary problems, which are related to the outstanding reserves. One of the main purposes of the current work is to analyze the robustness of the Chain-Ladder methodology over the comparison of the cost and the number of claims due to the Motor Third Party Liability Insurance type. Firstly, the classical Chain-Ladder method [1, 2] and the inflation adjusted Chain-Ladder method [3, 4] are considered. Then, a methodology for analyzing the results as a time series data, using series of development triangles [2, 5], is applied to compare both methods. The methods, combined in the current work, allow the insurer

E. Raeva (✉) · V. Pavlov

Department of Applied Mathematics and Statistics, University of Ruse, 8 Studentska Str., 7004 Ruse, Bulgaria

e-mail: eraeva@uni-ruse.bg

to different conclusions about the price estimations, even to construct different portfolios of insurance risks [6, 7]. It could be in use for managing different projects of insurance products and risk in pricing the investment resources [8].

2 Chain-Ladder Methodology

2.1 Basic Chain-Ladder Method for Number of Claims

As a start point of the analysis, the basic Chain-Ladder method is presented. For illustrating the steps of the algorithm, the method is applied to data corresponding to the number of claims for the period of years 2000–2010. The data is taken from the official site of the Financial Supervision Commission [9]. Following the methodology from [5], firstly, the cumulative run-off triangle is considered. Using the values from the up-triangle part of Table 1, the development factors are calculated with Formula (1), and the results are shown in the last row of Table 1.

$$f_k = \frac{E\{CS_k^{n-k}\}}{E\{CS_{k-1}^{n-k}\}} = \frac{\sum_{j=1}^{n-k} CS_{j;k}}{\sum_{j=1}^{n-k} CS_{jk-1}}, k = 1, 2, \ldots, n-1, \tag{1}$$

where $E\{CS_k^i\} = \frac{\sum_{j=1}^{i} CS_{jk}}{i}, i = 1, 2, \ldots, n; k = 0, 1, \ldots, n-1, i + k \leq n$. For the concrete data $n = 11$, because in Bulgaria, the maximum number of which the insurer payment for insurance of car driver liability could be delayed is 10 years.

Table 1 Cumulative values of the run-off triangle

CS_{ij}											
Occur Year	0	1	2	3	4	5	6	7	8	9	10
2000	14097	17913	18677	18954	19185	19455	19487	19528	19545	19549	19556
2001	27479	35727	37203	38166	38588	38846	39026	39077	39107	39130	39144
2002	24380	32640	34812	35879	36207	36542	36679	36718	36749	36766	36779
2003	28564	38920	41484	42781	43317	43707	43845	43890	43926	43946	43962
2004	30742	44531	49545	51268	51877	52320	52491	52557	52600	52625	52644
2005	32539	47199	55553	58032	59262	59606	59811	59887	59936	59964	59985
2006	34917	54543	61662	63856	64559	65089	65314	65396	65450	65480	65503
2007	38375	62080	67167	68709	69612	70183	70425	70514	70572	70605	70630
2008	42546	68301	72940	75240	76228	76854	77119	77217	77280	77315	77343
2009	43392	67254	73478	75795	76790	77421	77688	77786	77850	77886	77913
2010	45450	67252	73476	75793	76788	77419	77686	77784	77848	77884	77911
f_k		1.480	1.093	1.032	1.013	1.008	1.003	1.001	1.001	1.000	1.000

Table 2 Estimates of the run-of triangle of the number of claims

CS_{ij}											
Occur Year	0	1	2	3	4	5	6	7	8	9	10
2000	12097	3816	764	277	231	270	32	41	17	4	7
2001	25478	8248	1476	963	422	258	180	51	30	23	14
2002	22378	8260	2172	1067	328	335	137	39	31	17	13
2003	26561	10356	2564	1297	536	390	138	45	36	20	16
2004	28738	13789	5014	1723	609	443	171	66	43	24	19
2005	30534	14661	8354	2479	1230	344	205	76	49	28	21
2006	32911	19626	7119	2194	703	530	224	83	54	30	23
2007	36368	23706	5086	1542	903	572	242	89	58	32	25
2008	40538	25755	4639	2300	988	626	265	98	63	36	28
2009	41383	23862	6224	2317	996	631	267	98	64	36	28
2010	43440	21802	6224	2317	996	631	267	98	64	36	28

The estimations of the cumulative values, which are fulfilled in the down-triangle part of Table 1 are achieved with the following expression

$$\widehat{CS}_{ik} = CS_{i;k-1} * f_k, i = 2, 3, \ldots, n; k = 1, 2, \ldots, n-1; n < i + k \leq 2n - 1. \tag{2}$$

The predicted values shown in Table 2 are obtained after moving to a development triangle using a formula for the differences from the values in Table 1. $\widehat{X}_{ij} = \widehat{CS}_{ij} - \widehat{CS}_{i;j-1}, i = 2, 3, \ldots, n; j = 1, 3, \ldots, n-1; n < i + j \leq 2n - 1$.

There are different criterions for estimations and forecasts of the expected values of the free reserves by Chain-Ladder method in practice. The use of one or another type of result depends on the necessary information for the next year. Estimates of expected values, which will have to be paid by damages arising from previous years (PYR—previous year reserves), are considered in the current work [4]. These values are distributed by years for the period of the next n years. These results are obtained after summing the values of the diagonals of the down-triangle part of Table 2, using the formula:

$$PYR_{Y_n+i} = \sum_{j=0}^{n-i-1} \widehat{X}_{n-j;i+j}, i = 1, 2, \ldots, n-1. \tag{3}$$

Summing the values on the diagonals, the realized and expected values for reserves pending payments are obtained. Table 3 shows the estimated forecasts for the next 9 years.

The first number in Table 3 is obtained from the sum of the numbers below the main diagonal in Table 2, which are marked in dark color.

Table 3 Estimates of the number of claims

Year	2011	2012	2013	2014	2015	2016	2017	2018	2019
PYR_{Y_i}	32097	10477	4352	1104	484	223	127	64	28

2.2 Inflation Adjusted Chain-Ladder Method for the Cost of Claims

The inflation rate plays an important factor in estimating the expected values of claims that the insurers pay [4]. Inflation Adjusted Chain Ladder [3] was used for this purpose. The method is applied on data on the gross amount of paid claims under the Motor Third Party Liability Insurance for the same period considered in the previous paragraph, i.e., from 2000 to 2010. The data are presented by a development triangle in Table 4.

Data on inflation for the considered period were taken from the official website of the National Statistical Institute of Bulgaria [4]. Table 5 presents the inflation values for the respective years, as well as discount factors, $DF_i = \prod_{j=i}^{n}(IR_j + 1), i = 1, 2, \ldots, n - 1$. From the last row of Table 5, Table 6 is obtained. At the next

Table 4 Run-off triangle of the claims cost (in millions)

CS_{ij}	Development year										
Occur Year	0	1	2	3	4	5	6	7	8	9	10
2000	9.33	5.95	1.95	0.79	0.62	0.43	0.18	0.20	0.16	0.08	0.54
2001	19.25	9.95	3.24	2.01	1.12	0.53	0.45	0.27	0.26	0.22	
2002	14.26	10.91	4.26	2.1	1.44	0.69	0.51	0.23	0.23		
2003	1.61	13.96	4.55	2.12	1.21	1.13	0.51	0.55			
2004	17.92	17.69	7.86	3.16	2.35	2.3	0.48				
2005	18.98	20.09	11.92	3.76	2.16	1.96					
2006	22.78	28.46	11.59	5.17	1.94						
2007	29.64	41.65	13.39	5.34							
2008	40.7	55.66	14.35								
2009	43.57	54.08									
2010	43.8										

Table 5 Discount factor table

Year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
IR_i	0.086	0.042	0.011	0.050	0.026	0.057	0.056	0.109	0.063	−0.002	0.00
$IR_i + 1$	1.086	1.042	1.011	1.050	1.026	1.057	1.056	1.109	1.063	0.998	1.00
DF	1.619	1.490	1.430	1.415	1.347	1.313	1.242	1.177	1.061	0.998	1.00

Table 6 Discount triangle tables

CS_{ij}	Development year										
Occur Year	0	1	2	3	4	5	6	7	8	9	10
2000	1.619	1.490	1.430	1.415	1.347	1.313	1.242	1.177	1.061	0.998	1.00
2001	1.490	1.430	1.415	1.347	1.313	1.242	1.177	1.061	0.998	1.00	1.021
2002	1.430	1.415	1.347	1.313	1.242	1.177	1.061	0.998	1.00	1.021	1.062
2003	1.415	1.347	1.313	1.242	1.177	1.061	0.998	1.00	1.021	1.062	1.083
2004	1.347	1.313	1.242	1.177	1.061	0.998	1.00	1.021	1.062	1.083	1.075
2005	1.313	1.242	1.177	1.061	0.998	1.00	1.021	1.062	1.083	1.075	1.075
2006	1.242	1.177	1.061	0.998	1.00	1.021	1.062	1.083	1.075	1.075	1.077
2007	1.177	1.061	0.998	1.00	1.021	1.062	1.083	1.075	1.075	1.077	1.093
2008	1.061	0.998	1.00	1.021	1.062	1.083	1.075	1.075	1.077	1.093	1.119
2009	0.998	1.00	1.021	1.062	1.083	1.075	1.075	1.077	1.093	1.119	1.155
2010	1.00	1.021	1.062	1.083	1.075	1.075	1.077	1.093	1.119	1.155	1.146

stage, each element of Table 7 is multiplied by each element of Table 4 to obtain an upper triangular table with the cumulative values, which is shown in Table 7. The development factors and estimations should be calculated analogously to the basic Chain-Ladder method using Formulas (1) and (2), respectively. Sample of the results is presented in Table 7.

Then converting again to a triangle of development should be done (Table 8). In order to bring the estimates into forecast values, it is necessary to make an inflation forecast for the next period. The length of the estimated values of the inflation factor is equal to the number of rows (or columns) of the development triangle. In the current case, a forecast for 10 future observations is needed, i.e., from 2011 to 2020 inclusive. The forecast is made using the statistical software SPSS. Forecasting time series is a

Table 7 Cumulative estimations of the claims cost

CS_{ij}	Development year							
Occur Year	0	1	2	…	…	8	9	10
2000	15092197	23964414	26749830			29905697	29981020	30517164
2001	28684685	42917991	47493929			53408350	53631779	54590864
2002	20402145	35830317	41564064	…	…	47921926	48093766	48953817
2003	22787148	41594454	47566147			54147685	54341850	55313633
…				…	…			
2007	34876950	79059174	92422429			105520420	105898799	107792564
2008	43176528	98721356	113066604			129467576	129931826	132255369
2009	43485466	97562930	114369780	…	…	130959785	131429385	133779708
2010	43802789	88669424	103944218			119021934	119448728	121584803
f_k		2.0243	1.1723	…	…	1.0050	1.0036	1.0179

Table 8 Run-off triangle of the claims cost

CS_{ij}				Development year				
Occur Year	0	1	2	...	7	8	9	10
2000	15092197	8872218	2785415		238776	167795	75323	536144
2001	28684685	14233306	4575938		290658	260039	223428	959 085
2002	20402145	15428172	5733747	...	231740	225365	171 840	860 050
2003	22787148	18807306	5971692		553778	269 508	194 165	971 783
2004	24145600	23230305	9783934		474 939	333 266	240 099	1 201 679
2005	24923052	24964061	14027115		521 887	366 210	263 833	1 320 467
2006	28302950	33479971	12296989		602 389	422 699	304 530	1 524 151
2007	34876950	44182224	13363255		748 471	525 205	378 379	1 893 765
2008	43176528	55544828	14345248		918 332	644 397	464 250	2 323 542
2009	43485466	54077464	16806850	...	928 916	651 824	469 601	2 350 323
2010	43802789	44866635	15274794		844 239	592 406	426 794	2 136 076

whole subject of mathematics and the accuracy of the forecast is not a main focus in the current work. For this reason, the given option for modeling ARIMA time series in SPSS is used only. The results could be seen in Fig. 1. Since the development triangle under consideration covers previous periods, for which past inflation data are already available, the real data are taken. The forecast seen in Fig. 1 (the blue line) is needed for the analyses later in the paper. With these data, the values from the future inflation discount factor are calculated and the results are presented in Table 9. The last row of Table 9 is calculated by the formula $FDF_i = \prod_{j=1}^{i}(IR_{n+j+1} + 1)$, $i = 2, 3, \ldots, n$.
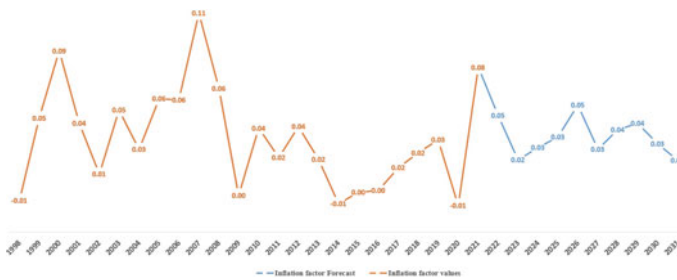
Fig. 1 Inflation factor forecast

Table 9 Future discount factor table

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
IR_i	0.000	0.021	0.040	0.020	−0.007	0.000	0.001	0.015	0.024	0.032	−0.008
$IR_i + 1$	1.000	1.021	1.040	1.020	0.993	1.000	1.001	1.015	1.024	1.032	0.992
FDF_i	1.000	1.021	1.062	1.083	1.075	1.075	1.077	1.093	1.119	1.155	1.146

Table 10 Estimates claims cost

Year	2011	2012	2013	2014	...	2016	2017	2018	2019
PYR_{Y_i}	77751246	33074 098	17644002	...	...	4211426	3788242	3206878	2446902

Multiplying the values from the down-triangle part of Table 6 with Table 8, the final estimations are obtained and summarized in Table 10.

After considering the highlights in both cases, similar calculations should be carried out for a number of development triangles and the results should be summarized as a time trend for the period 2000–2021.

3   Chain Ladder for a Number of Development Triangles

Following the calculations on the algorithm described above, the results of 12 triangles were obtained and summarized, which include the data on the number and value of the paid civil liability insurance benefits for Bulgaria. The analysis was made in two directions: the rate of growth of development factors during the considered period was tracked; a validation was made between the obtained estimates from the Chain-Ladder Method and the realized values.

3.1   Growth Rate of the Development Factors

After the calculations made according to the algorithm described above, generalized results for the development factors were also obtained. In Table 11, the growth rate for the development factors is calculated according to the respective year of development and the period to which the assessment refers.

The last row of Table 12 is calculated by

$$G_i = \sqrt[11]{\prod_{j=2}^{11}\left(\frac{f_i^j}{f_i^{j-1}}\right)}, i = 1, 2, \ldots 10. \tag{4}$$

Analogous results from the data of the claims cost are shown in Table 12.

Table 11 Average growth rate of the development factors according to the number of claims

Year	1	2	3	4	5	6	7	8	9	10
2000–2010	1.480	1.093	1.032	1.013	1.008	1.003	1.001	1.001	1.001	1.000
2001–2011	1.502	1.093	1.031	1.013	1.008	1.004	1.001	1.001	1.000	1.000
2002–2012	1.511	1.096	1.032	1.014	1.008	1.003	1.001	1.001	1.000	1.000
2003–2013	1.534	1.095	1.034	1.018	1.009	1.003	1.001	1.000	1.000	1.000
2004–2014	1.539	1.093	1.034	1.020	1.012	1.004	1.002	1.001	1.000	1.000
2005–2015	1.550	1.088	1.033	1.018	1.012	1.005	1.002	1.001	1.000	1.000
2006–2016	1.545	1.076	1.030	1.019	1.012	1.005	1.002	1.001	1.000	1.000
2007–2017	1.466	1.068	1.029	1.020	1.015	1.008	1.007	1.001	1.001	1.000
2008–2018	1.476	1.072	1.030	1.019	1.013	1.006	1.003	0.997	1.001	1.000
2009–2019	1.497	1.078	1.032	1.020	1.011	1.005	1.003	0.997	1.001	1.000
2010–2020	1.495	1.079	1.033	1.018	1.009	1.005	1.003	0.997	1.001	1.001
2011–2021	1,496	1.078	1.031	1.017	1.010	1.006	1.000	1.002	1.001	1.001
G_i in %	0.10	−0.12	−0.01	0.04	0.01	0.02	−0.01	0.01	0.00	0.00

Table 12 Average growth rate of the development factors from the claims' cost

| Year	1	2	3	4	5	6	7	8	9	10
2000–2010	2.024	1.172	1.061	1.032	1.024	1.009	1.007	1.005	1.004	1.018
2001–2011	2.091	1.174	1.065	1.034	1.022	1.009	1.006	1.004	1.002	1.002
2002–2012	2.125	1.177	1.066	1.035	1.024	1.009	1.006	1.004	1.002	1.003
2003–2013	2.202	1.173	1.074	1.036	1.025	1.012	1.006	1.002	1.002	1.001
.	…	…	…	…	…	…	…	…	…	…
2009–2019	2.492	1.185	1.081	1.046	1.025	1.013	1.007	1.006	1.001	1.001
2010–2020	2.484	1.187	1.081	1.045	1.028	1.013	1.008	1.005	1.003	1.002
2011–2021	2.453	1.184	1.079	1.047	1.029	1.015	1.013	1.005	1.004	1.002
G_i in %	1.76	0.09	0.16	0.13	0.04	0.05	0.05	0.00	0.00	−0.14

3.2 Validation of the Estimations

For the 12 development triangles, there are 12 estimated values for the PYR estimation and their corresponding 11 values of the already realized data. The results for the number of claims and the claims cost values are given in Table 13.

The graphics of the forecast validation are presented in Fig. 2. There are no jumps observed on the blue line, which corresponds to the PYR forecast. This shows the robust of the Chain-Ladder method over extreme periods. The inflation adjusted method (on the right-sided graph) estimates strongly positive trend due to the inflation factor. Both realized values have jumps in after 2018 in 2019, which most probably is a result of the pandemic condition at this time. Using the same formula (4), one

Table 13 PYR estimations for the number of claims and the claims cost values

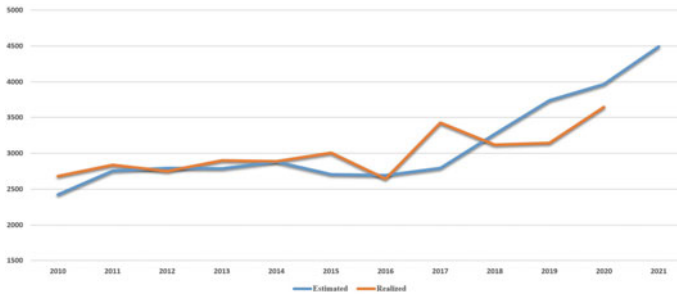| Year	Number of claims		Claims cost values	
	PYR	Realized values	PYR	Realized values
2000–2010	32,097	35,950	77,751,246	96,275,255
2001–2011	33,064	33,064	91,213,766	93,716,086
2002–2012	35,286	39,533	98,582,000	108,744,941
2003–2013	36,070	35,253	100,562,823	102,179,028
2004–2014	36,229	36,483	104,277,365	105,360,841
2005–2015	39,060	38,092	105,625,711	114,587,644
2006–2016	42,325	50,994	113,900,593	134,682,148
2007–2017	40,553	48,439	113,372,166	165,961,501
2008–2018	40,458	58,750	132,206,718	183,341,060
2009–2019	48,123	49,925	180,088,245	157,139,850
2010–2020	46,211	41,580	183,284,837	151,520,067
2011–2021	51,818		232,896,898	
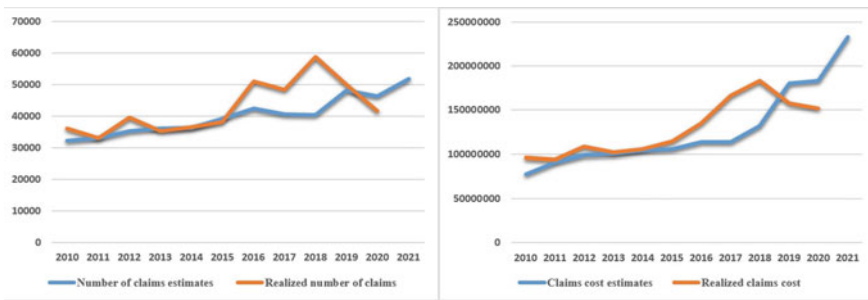
Fig. 2 PYR and realized values of the number of claims (on the left graph) and claims cost (on the right-side graph)

can look at the growth rate of the estimated values. For the number of claims, there is 4% growth rate against 9% growth rate for the forecast of the claims cost. While the realized values show 1% growth rate for the number of claims and only 5% for the claims cost. In Fig. 3, the results of the average values are illustrated, which the insurer could estimate for a single claim, calculated for each of the considered periods.

4 Conclusions

Basic Chain-Ladder method and Chain-Ladder method with inflation factor were considered in the current work. Twelve development triangles corresponding to the

Fig. 3 Average cost per claim

number of claims and the cost value, which have been paid over a 12 years' time horizon were analyzed and estimated. Over the calculation the result could be summarized in two directions—empirical results and model specifications. The growth rate of the development factors with 1-year delay corresponding to the number of claims is 0.1% versus 1.76% for the paid. This may be interpreted as a tendency for that, cases closed in the second year, to lead more expenses than these, which are paid during the current year of occurrence. This is also confirmed by the values of the development factor with 1-year delay of the claims cost, which are more than 2. Because of the inflation factor, the forecast for the payments has obvious positive trend in the last 3 years, which is not that clearly expressed by the realized values. On the other hand, the realized number of claims has a tendency to decrease during the last 3 years and the calculated estimations keep the low trend in long-term period. The estimated growth rate of the payments is 9% and the realized is just 5%. For the number of claims there is realized 1% growth rate and 4% expected. This is a result of the inflation growth during the crisis last 3 years. The considered Chain-Ladder method with analyzing a series of development triangles gives the possibility to follow the tendency for delay of the insurance claims. The Chain-Ladder estimation is robust to unexpected fluctuations and it is sensitive about the inflation factor. Combination of the results from the number and the claim cost values could be used for estimating the average value, which the insurer should be prepared to reserve for individual claims.

References

1. Weindorfer, B.: A Practical Guide to Use of the Chain-ladder method for determining technical provisions for outstanding reported claims in non-life insurance, University of Applied Science, bfi Vienne, October 2012, (2012).
2. Raeva E., Pavlov, V., Georgieva, S.: Claim Reserving Estimation by Using the Chain Ladder Method, In: AIP Publishing, New Trends in the Applications of Differential Equations in Sciences 2021, vol. 2321, pp. 030029–1–10, (2021).
3. Bente C.: Basic Chain Ladder Method Versus Inflation Adjusted Chain Ladder, Annals of Faculty of Economics, University of Oradea, Faculty of Economics, vol. 1(2), pages 343-354, (2016).

4. Raeva, E., Pavlov, V.: Inflation as a Factor in the Chain Ladder Method for Estimating Outstanding Claims Rese, In: AIP Publishing, New Trends in the Applications of Differential Equations in Sciences 2021, vol. 2459, pp. 030031–1–11, (2022).
5. Raeva E., Pavlov V.: Planning Outstanding Reserves in General Insurance, In: AIP Publishing, Application of Mathematics in Technical and Natural Sciences, vol 1895, pp. 050009–1–10, (2017).
6. Centeno, V., Georgiev, I., Mihova, V., Pavlov, V.: Price forecasting and risk portfolio optimization. In: Application of Mathematics in Technical and Natural Sciences 2019, vol. 2164, pp. 060006–1–15, AIP Publishing (2019).
7. Pavlov V., Mihova V.: An Application of Survival Model in Insurance, In: AIP Publishing, Application of Mathematics in Technical and Natural Sciences 2018, vol. 2025, pp. 030005–1–12, (2018).
8. Raeva I.: Computation of risk in pricing of investment projects, In: AIP Publishing, Application of Mathematics in Technical and Natural Sciences2019, vol. 2164, pp. 060016-1–10, (2019)
9. Official site of Financial Supervision Commission, http://www.fsc.bg/bg, accessed 3rd of June 2022

Comparative Analysis of ARIMA and Modified Differential Equation Approaches in Stock Price Prediction and Portfolio Formation

Vesela Mihova, Virginia Centeno, Ivan Georgiev, and Velizar Pavlov

Abstract In portfolio management theory, the principle of separation states that with the same input, all investors will have the same optimal risk portfolio. Whether the portfolio will actually be optimal depends on how accurate the results of the technical analysis conducted by the portfolio manager, or the investor is in order to predict the rate of return on the financial assets included in the portfolio. In this article, Autoregressive Integrated Moving Average (ARIMA) models have been used to predict assets' prices of four Bulgarian companies. Estimated rates of return have been calculated from the models. An optimal risk portfolio has been organized based on the Markowitz model. The resulting portfolio has been compared with a similar one obtained on the same data, using Modified Ordinary Differential Equations (ODE) to derive the forecast rates of return of the assets.

1 Introduction

Nowadays, financial markets are very important as they propose six main functions to each economy: Defining the price level, assets' liquidity, efficiency of purchases and sales, loans, and credits, up-to-date information for the cash flow, and risk management. As there are a lot of investors with different preferences, the financial markets offer different financial instruments, some of which are focused on short terms while others are more appropriate for long terms. The stock markets require that the investors must have better knowledge of the processes there because the market terms, factors, and circumstances suggest strong and sustained movements in traded instruments. The past publicly available information has some predictive relationship to the future stock returns, and one should know how to determine it and how to implement it. In portfolio management theory, the principle of separation states that

V. Mihova (✉) · V. Centeno · I. Georgiev · V. Pavlov
University of Ruse, 8 Studentska Street, 7017 Ruse, Bulgaria
e-mail: vmicheva@uni-ruse.bg

with the same input, all investors will have the same optimal risk portfolio. Therefore, the problem of choosing a portfolio can be divided into two separate unrelated problems [1]. The first problem is clearly a technical problem. An optimal portfolio should be defined. In fact, a portfolio manager offers the same portfolio P for all his clients regardless of their degree of risk aversion. This portfolio, obtained in response to the problem of maximizing the return/variability ratio, is the optimal risk portfolio that one could reach, taking into account the available input data, such as rates of return, standard deviations of rates of return, and the correlation matrix. The instruments used by the portfolio manager to predict the expected rates of return of the assets are significant for determining the risk portfolio. The second problem that should be solved is to allocate the treasury bonds and the risk portfolio in between the full investment portfolio in accordance with the personal preferences of each investor and how much he is willing to take risks. There is a wild variety of different approaches to time series forecasting and defining an optimal investment portfolio (see [2–5]). Some of them are technical approaches, while others are more easily applicable. The most widely used approaches to time series forecasting and that are providing complementary approaches to the problem are Autoregressive Integrated Moving Average (ARIMA) models [3]. They have been used to predict assets' prices of four Bulgarian companies. Estimated rates of return have been obtained from the models. Further, an optimal risk portfolio has been organized based on the Markowitz model. The portfolio has been compared with a similar one obtained on the same data, using modified ODE [6] to derive the forecast rates of return of the assets. The comparison is done under equal other conditions: The same choice of financial instruments, respectively, standard deviations and correlation matrices, the same program code for optimization and the same non-risky asset. The investment portfolio mixes shares of four Bulgarian companies in different work areas—technology, real estate, courier/transport services, and finances. These are the technology holding Allterco JSC, the joint stock company Elana AgroCredit JSC, the courier company Speedy JSC, and the holding Chimimport JSC. A daily data for the period 01.06.2020–29.10.2020 is used for the purposes of the study. The financial results of the four companies to the end of 2020 are presented in Table 1.

Table 1 Financial results of the four companies to the end of 2020

Company	Market capital (million BGN)	Turnover for the previous year (BGN)	Total assets (pieces)	Price per share (BGN)
Alterco	194.400	11, 617, 459.00	17, 999, 999	6.77
Elana Agrocredit	38.828	3, 114, 040.94	36, 629, 925	1.02
Speedy	408.699	2, 884, 193.70	5, 377, 619	58.39
Chimimport	225.267	9, 080, 952.02	239, 646, 267	0.94

2 ARIMA Approach

ARIMA models can predict a stock's future prices based on its past performance or forecast a company's earnings based on past periods. The parameters p, d, and q [7] are used to build ARIMA models. The autoregressive element p is the impact of the data from p previous moments in the model. The integrated element d is the trend in the data while the element q shows how many members are used to smooth small fluctuations with the help of a moving average. Generally, one ARIMA model with parameters p, d, and q [7], can be shown by: $Y_t = C + \varphi_1 \Delta^d Y_{t-1} + \cdots +$ $\varphi_p \Delta^d Y_{t-p} - \theta_1 \varepsilon_{t-1} - \ldots \theta_q \varepsilon_{t-q} + \varepsilon_t$, where C, φ_i, $i = \overline{1, p}$, and θ_j, $j = \overline{1, q}$ are the parameters sought; ε_j is a randomly distributed value with a zero mathematical expectation and dispersion σ^2. If there is no information about the distribution, then it is assumed by default to be normal. Δ is the difference operator, which is defined as: $\Delta^0 Y_t = Y_t$, $\Delta^1 Y_t = Y_t - Y_{t-1}, \ldots, \Delta^k Y_t = \Delta^{k-1} Y_t - \Delta^{k-1} Y_{t-1}$. Different combinations of parameters (p, d, q) are tested in this work in order to identify the time series. The graphs of auto-correlation functions (ACFs) and partially auto-correlation functions (PACFs, see [8]) are examined for each combination. These functions depend on a fixed number of lags and are calculated for each moment t, exception some end ones, where they cannot be calculated. Taking into consideration the jumps in both functions, the appropriate parameters for each model are obtained. If for Y_t ACF and PACF have no jumps or one small jump outside the 95% confidential intervals, then the model is taken as good enough for the purposes of the study.

3 Modified ODE Approach

To predict the prices of financial instruments, the authors propose in [6], the following approach based on modified ordinary differential equations. Let the chronologically arranged moments in ascending order and the observed values for an asset considered as a time series be, respectively, $t_0, t_2, \ldots, t_n$ and $y_0, y_2, \ldots, y_n$. An ODE could be fitted on these time series [9, 10], describing the values at the given discrete time instants: $y'(t) = g(t, y)$, $y(t_0) = y_0$. The function $g(t, y)$ can be largely arbitrary. Let it has the following form, as described in [6, 9]: $g(t, y) = \left(\sum_{i=0}^{M} a_i t^i\right) y + b_0 + \sum_{j=1}^{N} b_j \sin\left(\frac{2\pi j}{\theta} y + c_j\right)$, i.e., $y'(t) = \left(\sum_{i=0}^{M} a_i t^i\right) y + b_0 + \sum_{j=1}^{N} b_j \sin(\frac{2\pi j}{\theta} y + c_j)$. Here the coefficients (1) are unknown.

$$a_0, a_1, \ldots a_M, b_0, b_1, \ldots, b_N, c_1, c_2, \ldots, c_N, \theta. \tag{1}$$

These coefficients could be found by solving an inverse problem using a numerical one-step method for ODE. After finding these coefficients, the unknown

$n + 1$ value can be predicted, using the following equation: $y_{n+1} = y_n + h\left[\left(\sum_{i=0}^{M} a_i t_n^i\right) y_n + b_0 + \sum_{j=1}^{N} b_j \sin\left(\frac{2\pi j}{\theta} y_n + c_j\right)\right]$.

In [11], this approach is applied by the authors to the instruments considered here for the same time period. The results obtained in [11] are used to make a comparison with those obtained in the current paper.

4 Empirical Analysis

Using the daily closing prices of the instruments Alterco (A4L), Chimimport (CHIM), Elana AgroCredit (EAC), and Speedy (SPDY) from the period 01.06.2020 to 29.10.2020 (151 periods—from 0 to 150, including weekends and missing data), different predictions with ARIMA approach are made for the value of the closing price on the next day, i.e., for 30.10.2020 (period 151). The last known $l = 6$ periods are kept for validation. The development sample for each instrument consists of the data that excludes the periods kept for validation for the corresponding instrument. When a model works well on the development sample, it is tested by adding step by step the validation periods. The errors from the predictions on each step are calculated. They are used to calculate the weighted average error, applying the rule that the more recent the observations, the bigger the weight. Checks are made whether the models on development and validation data have jumps beyond the 95% confidence intervals in the ACF and PACF residual graphs. If there are no jumps or they are small, the model is applied to the entire data set for the corresponding instrument, and the predicted value for one period ahead is derived. It has been found empirically that the most suitable models for the A4L instrument are ARIMA (1, 0, 10) and ARIMA (2,1,6). There is a lack of any jumps outside the 95% confidence intervals in the ARIMA (1, 0, 10) model on all available data in both ACF and PACF residuals. The model predicts almost no change in the company's equity prices for 30.10.2020. For ARIMA (2, 1, 6), it has been checked that there is a lack of any jumps outside the 95% confidence intervals on all available data in both ACF and PACF residuals. This model predicts a growth in the company's equity prices for 30.10.2020. Similarly, to the described procedure with Alterco, ARIMA (3, 1, 10), and ARIMA (4,1,7) models are built on the relevant development sample for the CHIM instrument. It is verified that there are no jumps in ACF and there is a small jump outside PACF's confidence intervals for both models on the total population. Both models predict an increase in CHIM's prices for 30.10.2020. ARIMA (1, 0 ,6) and ARIMA (2,1,6) models are considered for the EAC instrument. There are no jumps outside ACF's and PACF's confidence intervals for both models. The model ARIMA (1, 0, 6) predicts a decrease in EAC's prices for 30.10.2020, while ARIMA (2,1,6) predicts an increase. The models, chosen for the last instrument—SPDY, are ARIMA (2, 1, 9) and ARIMA (3,1,10). There are no jumps outside ACF's and PACF's confidence intervals for both models. Both models predict a decrease in SPDY's prices for 30.10.2020. For each period over which the respective models

are validated, the forecast errors are derived. The results are presented in Tables 2, 5. The data for periods 145 and 146 (24.10.2020 and 25.10.2020) is missing as they are Saturdays and Sundays (days when the financial exchange is closed). For some of the instruments, data are also missing for working days as described below. For A4L (Table 2), data are missing for period 150 (29.10.2020).

For CHIM (Table 3), data are missing for period 144 (23.10.2020).

For EAC (Table 4), data are missing for period 147 (26.10.2020).

For SPDY (Table 5), there were no missing data for working days in the considered validation period.

The aim is to get a forecast for the period 151 (30.10.2020).

Table 2 Absolute errors of the predictions of the closing prices of A4L instrument under different model selection

Error	$l = 142$	$l = 143$	$l = 144$	$l = 147$	$l = 148$	$l = 149$
ARIMA (2, 1, 6)	0.04	0.18	0.07	0.04	0.10	0.14
ARIMA (1, 0, 10)	0.08	0.10	0.01	0.25	0.07	0.08

Table 3 Absolute errors of the predictions of the closing prices of CHIM

Error	$l = 142$	$l = 143$	$l = 147$	$l = 148$	$l = 149$	$l = 150$
ARIMA (4, 1, 7)	0.01	0.01	0.01	0.04	0.01	0.01
ARIMA (3, 1, 10)	0.01	0.01	0.01	0.03	0.01	0.01

Table 4 Absolute errors of the predictions of the closing prices of EAC

Error	$l = 142$	$l = 143$	$l = 144$	$l = 148$	$l = 149$	$l = 150$
ARIMA (2, 1, 6)	0.01	0.00	0.02	0.01	0.01	0.01
ARIMA (1, 0, 6)	0.00	0.01	0.01	0.00	0.00	0.00

Table 5 Absolute errors of the predictions of the closing prices of SPDY

Error	$l = 143$	$l = 144$	$l = 147$	$l = 148$	$l = 149$	$l = 150$
ARIMA (2, 1, 9)	1.17	0.38	0.06	0.62	0.33	0.66
ARIMA (3, 1, 10)	1.02	0.26	0.21	0.80	0.37	0.93

Table 6 Predicted value and weighted average error on the chosen models for the instruments

Instrument—model	Predicted value	Weighted average error
A4L—ARIMA (1, 0, 10)	5.0102	0.0993
A4L—ARIMA (2, 1, 6)	5.1525	0.0965
CHIM—ARIMA (3, 1, 10)	0.8758	0.0135
CHIM—ARIMA (4, 1, 7)	0.8826	0.0152
EAC—ARIMA (1, 0, 6)	1.0394	0.0031
EAC—ARIMA (2, 1, 6)	1.0477	0.0101
SPDY—ARIMA (2, 1, 9)	56.3522	0.5224
SPDY—ARIMA (3, 1, 10)	56.0732	0.6015

The weighted average error is derived as follows. The weights for calculating the weighted average error were given, respectively, $\frac{1}{7.5}$; $\frac{1.1}{7.5}$; $\frac{1.2}{7.5}$; $\frac{1.3}{7.5}$; $\frac{1.4}{7.5}$; $\frac{1.5}{7.5}$. They are applied to the error in the cases retained for validation. The more recent the observations, the bigger the weight. The choice of weights is somewhat subjective. In this case, a linear function was chosen based on empirical experience. The latter error has 1.5 times the weight of the former. Tables 6 shows the predicted values and weighted average errors on the chosen models for each of the four financial instruments, respectively.

For example, the weighted average error for the model A4L - ARIMA (2,1,6) from Table 6 is derived from Table 2 as follows: $\frac{1}{7.5} * 0.04 + \frac{1.1}{7.5} * 0.18 + \frac{1.2}{7.5} * 0.07 + \frac{1.3}{7.5} * 0.04 + \frac{1.4}{7.5} * 0.1 + \frac{1.5}{7.5} * 0.14 = 0.0965$.

The rest weighted average errors are calculated analogically.

A linear combination of the forecasts for the two models for each instrument is used to obtain the final predicted value for the corresponding instrument. The coefficients in this linear combination are inversely proportional to weighted average errors from the given models. The results are presented in Table 7.

For example, the estimated value for A4L is obtained from Table 6 as follows: $\frac{\frac{1}{0.0965} * 5.1525 + \frac{1}{0.0993} * 5.0102}{\frac{1}{0.0965} + \frac{1}{0.0993}} = 5.0824$.

Table 7 Predictions, actual values, and relative error in % between predicted and actual closing price values

Financial instrument	Estimated value	Actual value	Relative error in %
A4L	5.0824	4.98	2.0562
CHIM	0.8790	0.87	1.0343
EAC	1.0413	1.04	0.1297
SPDY	56.2225	57.00	1.3640

5 Optimal Risk Portfolio

The forecasts, obtained so far (Table 7), are used to organize an optimal risk portfolio via a MATLAB programming code, previously developed by the authors [5]. The code is based on the Markowitz model [1] under the following constraints: risk-free instruments exist; borrowing is possible at a risk-free rate; short sales of risky instruments are allowed. The purpose of the code is to maximize the slope of Capital Allocation Line (CAL) for each eligible risk portfolio p. The code deals with the following optimization problem: $\min F = -\max S_P = -\frac{E(r_p)-r_f}{\sigma_P}$ $\sum_{i=1}^{n} w_i = 1$, where:

- w_i—weight of the i-th stock,
- $E(r_p)$—the expected rate of return of the risk portfolio; it is calculated as follows: $E(r_p) = \sum_{i=1}^{n} w_i E(r_i)$.
- σ_P—the standard deviation of the risk portfolio: $\sigma_P = \sqrt{\sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1,i\neq j}^{n} \sum_{j=1}^{n} w_i w_j \sigma_i \sigma_j \rho(r_i, r_j)}$.
- $\rho(r_i, r_j)$—the correlation coefficient between the rates of return of the i-th and the j-th instrument.

The following data is used as an input for the programming code:

- The estimates of the expected rates of return (RoRs—see Table 8) are obtained as a relative difference between the estimated value for 30.10.2020 and the real value for 29.10.2020.
- The standard deviations (Table 8) and the correlation matrix (Table 9) are calculated in [11] using the historical data.

Table 8 Expected RoRs from the ARIMA models and standard deviations of RoRs

Financial instrument	Expected RoR (%)	Std. deviation (%)
A4L	1.6480	3.3276
CHIM	−0.1136	1.7765
EAC	0.1250	1.2198
SPDY	−1.3640	1.5851

Table 9 Correlation matrix

Instrument	A4L	CHIM	EAC	SPDY
A4L	1.0000	0.0430	0.0572	−0.1008
CHIM	0.0430	1.0000	0.0820	−0.1137
EAC	0.0572	0.0820	1.0000	−0.0294
SPDY	−0.1008	−0.1137	−0.0294	1.0000

Table 10 Estimated values, portfolio weights, last known price, and actual price

Instrument	Est. value ARIMA	Weights ARIMA, %	Est. value ODE	Weights ODE, %	Last known price	Actual price
A4L	5.0824	19.23 (+)	4.9846	9.18 (−)	5.00	4.98 (−)
CHIM	0.8790	8.37 (−)	0.8752	57.77 (−)	0.88	0.87 (−)
EAC	1.0413	7.20 (+)	1.0404	0.06 (+)	1.04	1.04 (0)
SPDY	56.2225	65.21 (−)	57.1085	32.99 (+)	57.00	57.00 (0)

– The return on the risk-free asset. A government security with an annual yield of 3%, which is equivalent to 0.0083% daily yield (3% / 360), has been used for the purpose.

The results obtained in this paper are as follows. The optimal risk portfolio consists of: 19.23% shares in A4L in a long position, 8.37% shares in CHIM in a short position, 7.20% shares in EAC in a long position, and 65.21% shares in SPDY in a short position. The expected rate of return on the risk portfolio is 1.2248%, and the standard deviation estimate is 1.1622%. In comparison, the risk portfolio for which the forecast prices obtained under the modified ODE approach was used which has an expected rate of return of 0.4031%, and a standard deviation estimate of 1.1371%. A comparison of the risk portfolio structure of the two approaches is shown in Table 10.

Table 10 presents long (+) and short (-) positions of each asset in the organized portfolio. Last column, which shows (+), (-), (0), presents the real movement of the closing prices, where 0 means that there is no change in the price. The difference between the last known price and actual price for each financial instrument has been calculated. Taking into account the positions (long, short) and the percentage distribution of the portfolio in both methods, it is possible to calculate what would be the real percentage return in each of the two approaches. The actual percentage return on the ARIMA approach is calculated as follows: $\left(-(5 - 4.98) * \frac{100}{4.98}\right) * 0.1923 + \left((0.88 - 0.87) * \frac{100}{0.87}\right) * 0.0837 = 0.0190\%$. The EAC and SPDY instruments have no change in the closing price and actually adding them to the portfolio brings neither profit nor loss, but they are involved in risk diversification. Similarly, for the real percentage return in the modified differential equation approach, we get $\left((5 - 4.98) * \frac{100}{4.98}\right) * 0.0918 + \left((0.88 - 0.87) * \frac{100}{0.87}\right) * 0.5777 = 0.7009\%$.

6 Conclusions

The present work examines an ARIMA approach, to make a forecast of the future prices of four Bulgarian financial instruments: A4L, CHIM, EAC, and SPDY, based on the daily close prices of the instruments for the period 01.06.2020–29.10.2020. The derived results are used to obtain a structure of the optimal risk portfolio (the percentage share of each asset), for a day ahead. The structure of this portfolio is

compared with a similar one obtained on the same data, where the forecasts are made using modified ODE and the portfolio is organized under equal other conditions. The following conclusions could be drowned: (1) Alterco (A4L) notes a real decline in the closing price. The ARIMA approach is predicting a slight increase, while the modified ODE approach is predicting a slight decline. Obviously, the forecast of the modified ODE approach is closer to the reality. (2) CHIM shows a real decline in the closure price, which is correctly predicted by both approaches. (3) The EAC and SPDY instruments have no change in the closing price and actually adding them to the portfolio brings neither profit nor loss. In fact, the forecasts from both approaches give a small increase or decrease in these instruments, and it turns out that the forecasts made with the modified ODE approach are closer. (4) In both approaches, a real positive return on the risk portfolio is reported. With the ARIMA approach, it is approximately 0.02%, while with the modified differential equation approach it is about 0.70%. It could be seen that in this case the modified differential equation approach is the better choice to predict the change in the prices of the selected financial instruments for the period under review, as well as to organize the investment portfolio. (5) It is interesting to note that the expected rate of return for the ARIMA approach (nearly 1.23%) is higher than that of the modified ODE approach (0.40%), and it is combined with a slightly higher standard deviation. This effect can be explained by the greater discrepancy between the estimated and the actual value of the ARIMA approach, which artificially increases the expected return.

References

1. Bodie, Z., Kane, A., Marcus, A.: Investments. 10th global ed. McGraw-Hill Education, Berkshire (2014).
2. Markova, M.: Convolutional neural networks for forex time series forecasting. In: New Trends in the Applications of Differential Equations in Sciences, vol. 2459, pp. 030024–1–9, AIP Publishing (2022).
3. Hyndman, R. J., Athanasopoulos, G.: Forecasting: principles and practice. 2nd ed. OTexts (2018).
4. Raeva, E., Nikolaev, I.: Retrospective Review of the Bulgarian Insurance Market Using Time Series Analysis. In: Application of Mathematics in Technical and Natural Sciences, vol. 2522, pp. 1–10, AIP Publishing (2022).
5. Centeno, V., Georgiev, I., Mihova, V., Pavlov, V.: Price forecasting and risk portfolio optimization. In: Application of Mathematics in Technical and Natural Sciences, vol. 2164, pp. 060006–1–15, AIP Publishing (2019).
6. Georgiev, I., Centeno, V., Mihova, V., Pavlov, V.: A Modified Ordinary Differential Equation Approach in Price Forecasting. In: New Trends in the Applications of Differential Equations in Sciences, vol. 2459, pp. 030008–1–7, AIP Publishing (2022).
7. Ngo, T. H. D., Bros, W.: The Box-Jenkins methodology for time series models. In: Proceedings Of The Sas Global Forum 2013 Conference, vol. 6, pp. 1–11, (2013).

8. Tabachnick, B., Fidell, L., Ullman, J.: Using multivariate statistics (Vol. 5, pp. 481–498). Boston, MA: Pearson (2007).
9. Xue, M., Lai, C. H.: From time series analysis to a modified ordinary differential equation. Journal of Algorithms & Computational Technology 12(2), pp. 85-90 (2018).
10. Lascsáková, M. The analysis of the numerical price forecasting success considering the modification of the initial condition value by the commodity stock exchanges. Acta Mechanica Slovaca 22(3), pp. 12-19 (2018).
11. Mihova, V., Centeno, V., Georgiev, I., Pavlov, V.: An Application of Modified Ordinary Differential Equation Approach for Successful Trading on the Bulgarian Stock Exchange. In: New Trends in the Applications of Differential Equations in Sciences, vol. 2459, pp. 030025–1–9, AIP Publishing (2022).

Models for Measuring and Forecasting the Inferred Rate of Default

Vilislav Boutchaktchiev

Abstract The rate of default is a measure of the credit risk of a portfolio of loans and is generally considered confidential information. However, the *inferred rate of default, (IRD)* is an estimate based on publicly reported information. We use the Bulgarian National Bank's quarterly reports on the credit quality to measure IRD for the major bank groups in Bulgaria. Our estimation is based on current regulations enforcing the IFRS 9 accounting standard in the Bulgarian bank system. Furthermore, focusing on banks of Group 2, we suggest an original methodology for forecasting IRD based on macroeconomic indicators. We report and compare the result of two approaches of estimation: a hybrid ARIMA regression, and an Asymptotic Single Risk Factor model of the Vasicek–Merton type. The general conclusion is that IRD resembles the known characteristics of the confidential rate of default and can be useful for credit risk analysis. In addition, it has the advantage of allowing estimation based on stale financial information and fresh macroeconomic forecasts in an intuitive and manageable way.

Keywords Probability of default · ARIMA estimation · Vasicek model · Random walk · IFRS9 · Expected credit loss

2010 Mathematics Subject Classification 62M05 · 62N02 · 91B70

JEL Classification G21 · M41

V. Boutchaktchiev (✉)
University of National and World Economy, Sofia, Bulgaria
e-mail: vboutcha@unwe.bg

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria

1 Introduction

For any given loan, the *probability of default* is the probability of it reaching a state where the borrower will be unlikely to repay the amount due. The *rate of default* at the end of period t, RD_t, furthermore, is constructed as an estimator of the probability of default, by setting up a portfolio of loans with the same risk profile and observing the amounts transitioning between states over a given period

$$RD_t = \frac{N_t}{P_{t-1}}. \tag{1}$$

Here P_{t-1} is the amount of all performing loans at the beginning of the period t, and N_t—the amount of those performing at the beginning but defaulting at the end of period t. In this study, we consider exclusively rates of default over 12 months; this horizon is methodologically useful for the estimation of expected credit loss according to the International Financial Reporting Standard, IFRS 9.

We define the *inferred rate of default* as a ratio derived from publicly available data concerning the quality of banks' assets with the purpose of approximate estimation of the rate of default. Next, we provide evidence that IRD can be utilized to analyze the risk tendencies in banks of Group 2 using time-series analysis. We show that the IRD is subject to similar time-series forecasting models as the ones previously developed with private data for the confidential rates of default in [2].

Additionally, in this study, we develop a Vasicek-type Asymptotic Single Risk Factor model for forecasting IRD and transitioning back and forth between trough-the-cycle and point-in-time probability of inferred impairment.

We hypothesize, that the IRD would prove useful for comparing credit risk cross-sectionally, between banks and bank groups. (This analysis, is left for the future.)

With the introduction of the IFRS 9 accounting standard, Bulgarian banks were required to employ models to produce economically justified forecasts of the probabilities of default for the estimation of their expected credit loss. (Cf., e.g., [1].) The development of such forecasting models for point-in-time PD is a subtle task because the results needn't be perceived as overly optimistic, or excessively conservative since both could bring in additional accounting risks. (Cf., e.g., [6].) For this reason, the banks are encouraged to build their own forecasting policies using available private and public data. A bank's private series of rates of default, however, is often found in need of corrections, due to errors, omissions, or corporate events, which contaminate the data. In such cases, the respective inferred rates of default can be instrumental in the computation of the necessary adjustments of RD distribution.

Further steps in the IRD study would involve the application of the techniques of event study and principal component analysis (similarly to, e.g., [10]) to analyze the effect of various system-wide events and regulatory changes.

2 Inferred Rate of Default

Since 2007, the banks in Bulgaria have been consistently classified into three categories in the following way: Group 1 consists of the five largest banks, Group 3 are the branches of foreign banks, and Group 2—all the rest. The Bulgarian National Bank (BNB) publishes quarterly reports on credit quality by bank groups. These reports have been issued under different titles through the years: *Information about credit quality and impairments* till the end of 2014, and since 2015 as *Information on nonperforming loans and impairments*. Despite minor variances in the classification, we were able to identify the necessary quantities during the entire period 2007–2022.[1]

Each quarterly report classifies the loans of the respective bank group into several portfolios, based on the type of borrowers. Furthermore, for each portfolio are identified the amounts that, at the end of the reporting period, t, fall into the following four categories:

– **Regular Performing loans**. These are loans that continue without any violation of the contract. We call this state $S1$ and denote the total amount of the loans in it by P_t.
– **Loans in breach of contract**. These are further sub-classified into three groups based on the length of the delinquency period:

  • Loans past due no more than 90 days, denoted by $S2$ with total P_t'. These loans were considered *under review* for most of the period of available data, using the previous accounting regulations, and currently, most of them fall into Stage 2 of impairment under IFRS 9.
  • Loans past due more than 90 but less than 181 days, *nonperforming*, denoted by $S3$ with total N_t.
  • Loans past due more than 180 days, *lost*, denoted by $S4$ with a total N_t^0.

We construct the *inferred rate of default* as follows:

$$IRD_t = \frac{N_t + \max\{0, N_t^0 - N_{t-1}^0\}}{P_{t-1} + P_{t-1}'} \tag{2}$$

We concentrate on two types of borrowers only:

1. **Non-financial corporations**. These are corporate credits, excluding loans extended to any state and local governments and their subsidiaries, as well, as banks and other financial institutions
2. **Households**. These are retail credits, including all consumer and mortgage loans.

---

[1] In 2019, with the oncoming of the COVID-2019 pandemic, the Bulgarian government introduced a series of measures that gave the banks the option to postpone the recognition of failing loans as "defaulted". This has had the effect of "smoothing", the rates of default, and inferred rates of default during that period. To avoid contamination, we have omitted from consideration all data after December 31, 2019.

In the remaining sections of the paper, we focus on the IRD of the portfolios of Group 2, which as of March 31, 2022, hold about 30% of all loans in the system. We suggest, that a comparison between the IRD series of the three bank groups would be an interesting topic for future analysis, which might turn out to be revealing of the portfolio quality of the different types of banks.

2.1 Economic Interpretation of the Inferred Rate of Default

The IRD only approximately resembles the usual RD.

1. In constructing the IRD, the loans past due less than 90 days are considered as performing. Most of them, very likely, would have been considered as such in the computation of ECL, even with an elevated risk (i.e., classified in Stage 2 of impairment, Cf. [6]). Some of them, however, should have been regarded as defaulting, according to the individual bank's policies.
2. While it is, in most situations, fair to consider loans less than 6 months past due as fresh NPE for the given quarterly report, some of them will probably have been classified as defaulting by the banks even earlier. The IRD catches such cases with eventual delay.
3. We consider in the IRD the positive growth in stale NPE (past due more than 180 days) as recent defaults. Our experience indicates that banks often manage the riskier loans in their portfolio in a way that allows them to record newly defaulting loans directly into this group. This is, in part, due to their policies being formulated during an earlier regulatory period which allowed loans to be considered as "exposure under review" up until the moment when it was declared "lost". On the other hand, excluding these exposures from consideration would have made IRD uncharacteristically small at this time. Perhaps, as regulatory supervision becomes more diligent in the future this component would have to be modified or even excluded from consideration.

Remark 1 As the rate of default measures the relative frequency of loans transitioning from the regular state to default, the *inferred rate of default* is designed to estimate the probability of transitioning from states $S1$ or $S2$ to state $S3$. (The positive growth in state $S4$ is added to the numerator, as explained above, only as a correction and because most of the available reports date back to the period before the introduction of IFRS9.) Such transition is costly for the bank and, hence, for the contract holder, as well.

Definition 1 The event of a loan transitioning from states $S1$ or $S2$ to $S3$ or $S4$ is referred to as *inferred impairment* and the *probability of inferred impairment* (PII for short) is its expected relative frequency using the information set S_{t-1}, available 1 year back

$$PII_t = E[IRD_t | S_{t-1}].$$

2.2 Other Data Used in the Analysis

In the building of forecasting models for IRD, we use the reports of the National Statistical Institute (NSI) of Bulgaria for the Gross Domestic Product (GDP) and the Unemployment Rate. These are the most common macroeconomic factors used for forecasting RD. In addition, this choice allows us to compare IRD with the results in [2]. For further details on variable construction we refer to [7].
Gross Domestic Product. We use Real GDP quarterly reports to compute, at the end of each quarter the one-year growth rate of annual GDP.

$$G(t) = \frac{RGDP[t-3, t]}{RDGP[t-4, t-7]} - 1.$$

Here, $RGDP[t-3, t]$ denotes the cumulative real GDP at average 2015 prices as reported in the four quarters $t-3, ..., t$.
Unemployment Rate. The unemployment rate is reported by NSI at the end of each quarter and it is measured as the number of unemployed persons as a proportion of the labor force. The variable used for this analysis is the annual growth or the rate of default

$$U(t) = \frac{UR[t]}{UR[t-4]} - 1.$$

Here $UR[t]$ denotes the Unemployment Rate measured at the end of quarter t.

3 ARMA Regressions

In this section, we develop forecasting models for IRD using the methodology of ARMA regressions.

3.1 Explanatory Variables

In a sequence of linear regressions, we test the dependence between the macroeconomic indicators and the IRD.

Table 1 show the results of OLS regression estimation of the following models:

$$y_t = \alpha + \beta x_{t-\tau} + \epsilon_t \tag{3}$$

Here

$$y_t = \ln(IRD_t) - \ln(IRD_{t-4})$$

Table 1 Comparison between OLS regression with different lags (3). The coefficient α is not reported. Reported are the R^2 and β coefficients are reported along with the p values of their individual t-tests of significance. For the Corporate portfolio Group 2, the best model is with Lag 0. Similarly, in the Retail segment, the portfolio exhibits a strong positive relationship with the current unemployment rate. Notice, that the reversal of the relationship with stale data may be evidence of compensating for an overreaction

		Lag				
		$\tau = 0$	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$
Corporate	Coef.	-7.82	-7.33	-5.65	-4.25	-2.78
	p-value	(0.008)	(0.014)	(0.058)	(0.136)	(0.313)
	R^2	0.16	0.14	0.09	0.06	0.03
Retail	Coef.	1.71	0.79	0.41	-0.39	-1.49
	p-value	(0.009)	(0.240)	(0.547)	(0.577)	(0.028)
	R^2	0.16	0.04	0.01	0.01	0.13

is always the annual logarithmic growth in IRD, $\tau = 0, 1, ..., 4$ is a time lag in quarters, and the explanatory variables are defined as follows:

– For the corporate portfolio:

$$x_t = G(t) - G(t - 4)$$

– For the retail portfolio

$$x_t = U(t) - U(t - 4)$$

As a result, we observe in the Corporate portfolio of Group 2 a quick reaction (after only one quarter) to changes in GDP. Similarly, in the Retail portfolio, IRD shows an immediate response to changes in the dynamics of the Unemployment Rate.

3.2 Model Estimation and Comparison with the Rate of Default

The stationarity of all the time series, $x(t)$ and $y(t)$ is confirmed using the Dickey–Fuller test.

For the **Corporate portfolio** in Group 2, using the Box–Jenkins method, we identify the ARMA(1,1) model as best-fitting.

$$y_t = \beta x_t + \rho y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t \tag{4}$$

The results of the estimation are shown in Table 2.

Table 2 ARMA regressions for the Corporate IRD. The models for the portfolio in Group 2 is according to specification (4). All coefficients are significant to 95% level, as shown by p-values. The Wald χ^2 of the overall significance of the model passes at the 99% level

	β	θ	ρ	χ^2
Coef.	-5.89	0.87	-0.52	49.92
p-value	(0.029)	(0.0001)	(0.023)	(0.0001)

Table 3 ARMA regressions for the Retail IRD. Two models are reported. The models for the portfolio in Group 2: the best fitting model (5) (lines 1, 2) and specification (6) which better resembles the RD model (lines 3, 4). All coefficients are significant at the 95% level as seen from the p-values. The Walt χ^2 test overall significance of the model passes at 99%

		β	ρ_1	ρ_4	χ^2
(5)	Coef.	1.73	–	-0.44	20.38
	p-value	(0.023)		(0.0001)	(0.0001)
(6)	Coef.	2.02	0.44	–	10.78
	p-value	(0.045)	(0.006)		(0.004)

Similarly, we apply the Box–Jenkins to select a AR(4) model as best fitting in the **Retail portfolio** of Group 2

$$y_t = \beta x_{t-\tau} + \rho_4 y_{t-4} + \epsilon_t \tag{5}$$

The results of the estimation are shown in Table 3. We have reported the result of the model AR(1) which fits reasonably well and has a similar form to the one reported in [2]:

$$y_t = \beta x_{t-\tau} + \rho_1 y_{t-1} + \epsilon_t \tag{6}$$

Comparison with Rate of Default. In comparison, similar forecasting models for RD in [2]. For the Corporate portfolio

$$y_t = -6.96x_t + 0.86y_{t-1} - 0.99\epsilon_{t-1} + \epsilon_t.$$

and for the Retail portfolio

$$y_t = 0.86x_{t-1} + 0.697y_{t-1} + \epsilon_t.$$

The similarity of models for IRD and RD with data for banks of Group 2 suggests that the IRD distribution can be used to build models for the replacement of missing and faulty data.

4 A Modified Vasicek Model

In this section, the 12-month probability of inferred impairment of a portfolio, PII, is considered to have an underlying value, the *trough-the-cycle PII* (TTC-PII), and at any time t there is a actual, *point-in-time PII* (PID-PII).

Vasicek's Asymptotic Single Risk Factor model (Cf., e.g., [9]) applies a geometric Brownian motion equation to a firm's asset value

$$dA_i = \mu_i A_i dt + \sigma_i A_i dx_i,$$

where A_i is the individual borrower's asset value, μ_i and σ_i are the drift rate and volatility of that value and x_i is a Wiener process. The value A_i at the loan's maturity time T can be represented as

$$\ln A_i(T) = \ln A_i(0) + \mu_i T - \frac{1}{2}\sigma_i^2 T + \sigma_i \sqrt{T} X_i,$$

where X_i has a standard normal distribution.

Although the transition of a given exposure to the state of inferred impairment does not necessarily lead to the total loss of the firm, it is costly for the borrower, and we assume, that it will be allowed to happen only after the value of the asset drops below a certain level, $\tilde{B}_i$.

This assumption (Cf. [5]) allows us to conclude

$$PII_i = P[A_i(T) < \tilde{B}_i] = P[X_i < D_i] = \Phi(D_i).$$

Here Φ is the CDF of the standard normal distribution and

$$D_i = \frac{\ln \tilde{B}_i - \ln A_i(0) - \mu_i T + \frac{\sigma_i^2 T_i}{2}}{\sigma_i \sqrt{T_i}}$$

the *inferred impairment threshold* of the asset.

These observations justify the assumption that the values $\Phi^{-1}(IRD_t)$ (similarly to $\Phi^{-1}(RD_t)$) are drawn from a standard normal distribution.

Furthermore, we make a standard (cf. [4, 9]) assumption, that a bank's portfolio is built of assets with a similar risk profile resulting in the constraint that the respective variables X_i are jointly standard normal and of equal pairwise correlation ρ. We have

$$X_i = Z\sqrt{\rho} + Y_i\sqrt{1 - \rho},$$

where $Z, Y_1, \dots Y_n$ are mutually independent standard normal variables. In this setup, Z is considered as a *common factor* which has an effect on the portfolio (e.g., the overall macroeconomic and business cycle over the interval $[0, T]$), Y_i represents the borrowers' respective *idiosyncratic risk* and $\rho = \mathrm{corr}(Z, X_i)$ is the *asset correlation*.

Although originally developed for an individual asset, the Vasicek setup applies to the entire portfolio of loans, instead.

Definition 2 The *point-in-time PII* of a portfolio is the function

$$p(z) = P[X < D|Z = z] = \Phi\left(\frac{D - z\sqrt{\rho}}{\sqrt{1 - \rho}}\right),$$

i.e., the probability of inferred impairment given the systemic factor has a specific value $Z = z$.

Here D can be interpreted as the inferred impairment threshold of the portfolio's "average client", or, the "central tendency client".

Definition 3 The *through-the-cycle PII* of a portfolio is obtained as the marginal probability by averaging through all possible macroeconomic conditions (i.e., values of Z)

$$p = E_Z[p(Z)] = \int_{-\infty}^{\infty} p(z)\phi(z)\,dz = \Phi(D).$$

The parameters D and ρ can be estimated using the observed data of default rates. This type of estimation is sometimes done using a more complex methodology involving Kaplan filters and Monte Carlo integration (Cf., e.g., [3, 8]. Implementation of this technique is left for the future.) Here, we suggest a simpler application of a method of moments that produces a satisfactory approximate result.

Proposition 1 *In the setup of the Vasicek model with the additional assumption that both the asset correlation (i.e., the parameter ρ) and the factor Z remain invariant over a prolonged period, we denote with $\{IRD_t : t \in A\}$ the time series of observed inferred rates of default (i.e., the measurements of PIT-PI). Denote by $\hat{\mu}$ and $\hat{\sigma}^2$ the sample mean and variance of the time series $\{\Phi^{-1}(IRD_t) : t \in A\}$. Then, assuming that the available data describes sufficiently well the full possible variety of systemic macroeconomic conditions the following estimates of D, R, and Z_t are valid:*

$$\hat{D}\frac{\hat{\mu}}{\sqrt{1 + \hat{\sigma}^2}}, \quad \hat{\rho} = \frac{\hat{\sigma}^2}{1 + \hat{\sigma}^2}, \quad and \quad \hat{Z}_t = \frac{\hat{\mu} - \Phi^{-1}(IRD_t)}{\hat{\sigma}}.$$

The proof will be reported elsewhere.

4.1 Model Estimation

Setting

$$y_t = Z_t - Z_{t-4},$$

Table 4 ARMA regressions for the Z-score of the IRD. The model for Corporate Group 2 is according to specification (4), and for Retail—according to (5), The results reveal that corporate Z-scores positively correlate to GDP, and Z-scores for the retail portfolio negatively correlate to the unemployment rate, as expected

Corporate		β	ρ	θ	χ^2
	Coef.	5.62	0.86	−0.45	46.13
	p-value	(0.068)	(0.0001)	(0.048)	(0.0001)
Retail		β	ρ_4		χ^2
	Coef.	−1.88	−0.45		21.40
	p-value	(0.011)	(0.0001)		(0.0001)

we confirm that the forecasting models (4) and (5) are valid. The results of the estimation are shown in Table 4.

5 Conclusions

We define the *inferred rate of default* to approximate the rate of default using available systemic data. The resulting data series exhibits the following characteristics:

1. The IRD is a measure of the credit riskiness of a portfolio, reflecting the relative frequency of transition of the exposure of the portfolio from a regular state to the state of implied impairment.
2. The IRD similarly depends on macroeconomic factors as RD. The IRD for corporate loans is explained by the growth of the GDP, and for retail—by the growth of the Unemployment Rate.
3. The IRD can be modeled using an Asymptotic Single Risk Factor, similar to the Vasicek model, and can be foretasted similarly. This type of model offers a slight advantage over less complex time series models and provides a clear and sturdy theoretical frame of all underlying assumptions of the forecasting process. Our observations, however, indicate that ASFR models, are found by smaller banks less intuitive and their results are harder to manage.

References

1. Basel Committee on Banking Supervision *Prudential treatment of problem assets—definitions of non-performing exposures and forbearance*, Consultative Document. BIS, July 15, 2016.
2. Boutchaktchiev, V. *On the use of macroeconomic factors to forecast the probability of default.* Proc. of the Fprthy-seventh Conf. of UMB, 2018, pp. 95–101.

3. Dimov I., Georgieva R., Todorov V. Balancing of Systematic and Stochastic Errors in Monte Carlo Algorithms for Integral Equations. In: Dimov I., Fidanova S., Lirkov I. (eds) Numerical Methods and Applications. NMA 2014. Lecture Notes in Computer Science, vol 8962. Springer, (2015)
4. Gordy, M. B., *A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules* (2002). FEDS Working Paper No. 2002-55, Available at SSRN: https://ssrn.com/abstract=361302
5. Merton, R.C., *On the Pricing of Corporate Debt: The Risk Structure of Interest Rates.* The J. of Finance, 29: 449–470, (1974)
6. Milanova, E. Compatibility Between IFRS 9 Financial Instruments and the Basel Capital Requirements Framework. ICPA Articles, Inst. of CPA 2016, pp 1–48. (in Bulgarian.)
7. National Statistical Institute: Statistical Reference Book 2020, Sofia, Bulgaria 2020.
8. Todorov, V. and Dimov, I. T., *Monte Carlo methods for multidimensional integration for European option pricing*, AIP Conference Proceedings, 1773:1, 100009, (2016), https://doi.org/10.1063/1.4965003.
9. Vasicek, O., *The Distribution of Loan Portfolio Value* RISK, Vol. 15, No. 12, (2002), pp. 160–162.
10. Vitanov, N. K., Sakai, K, Jordanov, I.P., Managi, S., Demura K., *Analysis of a Japan government intervention on the domestic agriculture market,* Physica A: Statistical Mechanics and its Applications, 382:1, 2007, pp 330–335.

Applications in Neuroscience

Application of the Wavelet Data Transformation for the Time Series Forecasting by the Artificial Neural Network

Anastasia Butorova ⓘ **, Elena Baglaeva** ⓘ **, Irina Subbotina,**
Marina Sergeeva ⓘ **, Aleksandr Sergeev** ⓘ **, Andrey Shichkin** ⓘ **,**
Alexander Buevich ⓘ **, and Pavel Petrov** ⓘ

Abstract   The study tested how the wavelet transform of the data affects the accuracy of an artificial neural network model for forecasting surface methane concentration. A model based on the nonlinear autoregressive neural network with external input (NARX) was used. For comparison, we used the base NARX model and the hybrid model. The hybrid model was created based on the data to which the discrete wavelet transform (DWT) was applied. For DWT, the Daubechies wavelet of the fourth level was used. The initial data for the study were collected on the measurements of the concentration of greenhouse gases in the Russian Arctic zone. We evaluated the accuracy of the models by the following indicators: Mean absolute error, root mean square error, and the index of agreement. The proposed approach has improved the accuracy of the forecast. The accuracy of the hybrid model has increased by more than 10%.

Keywords   Wavelet transform · Artificial neural networks · Hybrid models

1   Introduction

More and more researchers are using the artificial neural network (ANN) approaches to predict time series in different fields of knowledge. To obtain adequate forecasting results, it is necessary to take into account many ANN parameters. In addition, the preliminary preparation of the initial data for training the ANN can also increase

A. Butorova (✉) · E. Baglaeva · I. Subbotina · M. Sergeeva · A. Sergeev · A. Shichkin ·
A. Buevich
Institute of Industrial Ecology UB RAS, S. Kovalevskoy Str., 20, Ekaterinburg, Russia 620990
e-mail: a.s.butorova@urfu.ru

A. Butorova
Ural Federal University, Mira Str., 19, Ekaterinburg, Russia 620002

P. Petrov
University of Economics, 77 Knyaz Boris I Blvd, 9002 Varna, Bulgaria

forecasting efficiency. One of these methods is the discrete wavelet transform (DWT), which is successfully used in the analysis of time series [1–12].

In this paper, we compared the performance of a model based on an ANN-type NARX and a hybrid model in which the wavelet-transformed data were used to train NARX.

2 Materials and Methods

For the study, a time series was taken, which represented the average hourly measurements of the concentration of one of the main greenhouse gases—methane. These data were obtained during the monitoring of greenhouse gases in the Arctic part of the Russian Federation (Bely Island, Yamalo-Nenets Autonomous District). Measurements were carried out by a cavity ring-down spectrometer Picarro G2401. In addition, the main meteorological parameters (temperature, atmospheric pressure, and air humidity) were measured and synchronized. These measurements were taken by the Vaisala Automatic Weather Station AWS310. The location of the measurement site is shown in Fig. 1.

As a result, 1175 time points were obtained for each of the four scalar time series: Temperature, atmospheric pressure, humidity, and CH4 concentrations. Of which, 1103 time points were a training time series that was used for training, and the
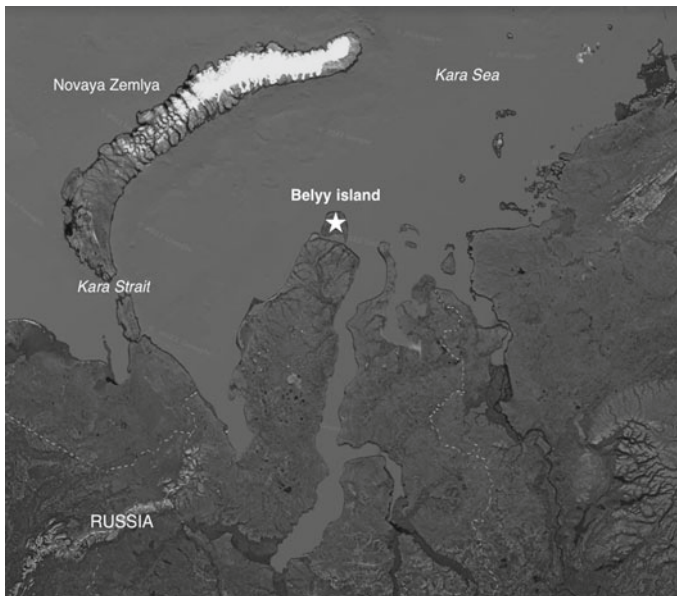
Fig. 1 Place of measurements (Google Earth)

remaining 72 time points were used for testing and did not participate in the training process.

ANN NARX, which is one of the best for forecasting time series [13–17], was used as a base model. The NARX network is an iterative, dynamic multi-layer feedback network. The use of feedbacks allows using the description of recurrent neural networks NARX in the form of a set of states, which makes them convenient devices for nonlinear forecasting and modeling.

The hybrid model (DWTNARX) was also built on the basis of the NARX network for training which DWT transformed time series was used.

Before DWT, the time series of CH4 concentrations was scaled from 0 to 1. DWT was used to represent the scaled training series of CH4 concentrations as a set of successive approximations of the approximating and detailing components:

$$S(t) = A_J(t) + \sum_{m=1}^{J} D_m(t),$$ (1)

where $S(t)$ is the original time series, $A_J(t)$ and $D_m(t)$, „ (m = 1,…,J) are approximating and detailing components, J is the DWT decomposition level.

For DWT, the Daubechies wavelet of the fourth level was used. Daubechies wavelets are orthogonal, asymmetric, and have no analytical form. In the Matlab package, Daubechies are given by a set of weighting factors.

In the training procedure, the input data of the artificial neural network NARX were meteorological time series: Temperature, humidity, and pressure, which were scaled from 0 to 1. The output data was scaled from 0 to 1 values of the approximating and detailing components.

For each component, the neural network was trained and the values of the test time series of CH4 concentrations were forecasted. Each network was trained 500 times. The network with the smallest root mean square error (RMSE) was selected. The final forecasting was calculated as the rescaled sum of the forecasting of the approximating and detailing components for CH4 concentrations.

To train the ANNs, the Levenberg–Marquardt Learning Algorithm was applied.

To assess the prediction accuracy, we used the following indices: Mean absolute error (MAE) (2), RMSE (3), and the index of agreement (d), (a standardized measure of the degree of model prediction error and varies between 0 and 1, where a value of 1 indicates a perfect match, and 0 indicates no agreement at all [18] (4).
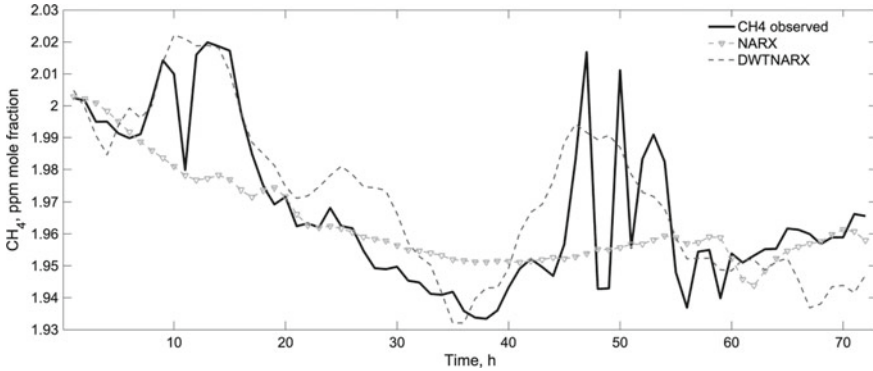
$$MAE = \frac{\sum_{i=1}^{n} |P(x_i) - M(x_i)|}{n},$$ (2)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P(x_i) - M(x_i))^2}{n}},$$ (3)

$$d = 1 - \frac{\sum |P(x_i) - M(x_i)|}{\sum (|P(x_i) - \overline{m}| + |M(x_i) - \overline{m}|)},$$ (4)

Table 1 The results of prediction of the methane concentration

Model	MAE, ppm	RMSE, ppm	d
NARX	0.0119	0.0180	0.775
DWTNARX	0.0124	0.0161	0.886

Fig. 2 Comparison of different prediction approaches

where $P(x_i)$ is a predicted concentration in location x_i, $M(x_i)$ is a measured concentration, $\overline{m}$ is a mean concentration, and n is a number of points.

3 Results and Discussion

The final neuron number in the hidden layer was 20. Table 1 shows the parameters used to compare the performance of the different methods (the best values for the test interval are in bold).

Figure 2 shows the prediction made by the NARX and hybrid DWTNARX models. The forecast for the hybrid DWTNARX turned out to be more accurate. On the index of agreement, the accuracy improved by 13%.

4 Conclusion

The paper shows that the use of data preliminarily prepared for the training time series made it possible to improve the accuracy of forecasting changes in the concentration of greenhouse gas CH4. The discrete wavelet transform (the Daubechies wavelet) method was used to transform the data. The ANN NARX trained on these data improved the quality of the model by 13%.

References

1. Adamowski, J., Chan, H. F.: A wavelet neural network conjunction model for groundwater level forecasting. Journal of Hydrology 407, 28–40 (2011).
2. Ebrahimi, H., Rajaee, T.: Simulation of groundwater level variations using wavelet combined with neural network, linear regression and support vector machine. Global and Planetary Change 148, 181–191 (2017).
3. Kalteh, A. M.: Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. Computers & Geosciences 54, 1–8 (2013).
4. Graf, R., Zhu, S., Sivakumar, B.: Forecasting river water temperature time series using a wavelet–neural network hybrid modelling approach. Journal of Hydrology 578, 124115 (2019).
5. Siwek, K., Osowski, S.: Improving the accuracy of prediction of PM10 pollution by the wavelet transformation and an ensemble of neural predictors. Engineering Applications of Artificial Intelligence 25(6), 1246–1258 (2012).
6. Chen, Y., Shi, R., Shu, S., Gao, W.: Ensemble and enhanced PM10 concentration forecast model based on stepwise regression and wavelet analysis. Atmospheric Environment 74, 346–359 (2013).
7. Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J.: Artificial neural networks forecasting of PM 2.5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmospheric Environment 107, 118–128 (2015).
8. Dunea, D., Pohoata, A., Iordache, S.: Using wavelet–feedforward neural networks to improve air pollution forecasting in urban environments. Environmental Monitoring and Assessment 187(7), 477 (2015).
9. Bai, Y., Li, Y., Wang, X., Xie, J., Li, C.: Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. Atmospheric Pollution Research 7(3), 557–566 (2016).
10. Cabaneros, S. M., Calautit, J. K., Hughes, B.: Spatial estimation of outdoor NO2 levels in Central London using deep neural networks and a wavelet decomposition technique. Ecological Modelling 424, 109017 (2020).
11. Osowski, S., Garanty, K.: Forecasting of the daily meteorological pollution using wavelets and support vector machine. Engineering Applications of Artificial Intelligence 20(6), 745–755 (2007).
12. Su, X., An, J., Zhang, Y., Zhu, P., Zhu, B.: Prediction of ozone hourly concentrations by support vector machine and kernel extreme learning machine using wavelet transformation and partial least squares methods. Atmospheric Pollution Research 11(6), 51–60 (2020).
13. Zhou, Q., Jiang, H., Wang, J., Zhou, J.: A hybrid model for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network. Science of the Total Environment 496, 264–274 (2014).
14. Zhang, G.: Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 5, 159–175 (2003).
15. Russo, A., Raischel, F., Lind, P.: Air quality prediction using optimal neural networks with stochastic variables. Atmospheric Environment 79, 822–830 (2013).
16. Strnad, D., Nerat, A., Kohek, Z.: Neural network models for group behavior prediction: a case of soccer match attendance. Neural Computing and Applications, 28(2), 287–300 (2017).

17. Mohanty, K., Majumdar, T. J.: Using artificial neural networks for synthetic surface fitting and the classification of remotely sensed data. International Journal of Applied Earth Observation and Geoinformation 1(1), 78–84 (1999).
18. Willmott, C. J.: On the Validation of Models. Physical Geography 2, 184–194 (1981).

Discrete Neural Networks with Maximum and Non-instantaneous Impulses with Computer Simulation

Snezhana Hristova and Kremena Stefanova

Abstract A nonlinear non-instantaneous impulsive difference equations with maximum of the state variable over a past time interval is investigated. The exponential stability concept is studied and some criteria are derived. These results are also applied for a neural networks with switching topology at certain moments and long time lasting impulses. It is considered the general case of time varying connection weights. The equilibrium is defined and exponential stability is studied. The obtained results are illustrated on examples.

Keywords Difference equations · Non-instantaneous impulses · Maximum · Exponential stability · Discrete neural networks · Switching topology

1 Introduction

One of the most important problems in the theory and application of difference equations is stability (see, for example, [3, 6, 7, 9, 11, 12]). At the same time impulses are a very useful mathematical apparatus to model some instantaneous perturbations in the process. In the case when the acting time of the impulses is not possible to be neglected, these impulses are called non-instantaneous impulses (for continuous case, see, [4]).

In this paper we study nonlinear difference equations with a special type of delay in the case there are some impulses starting at initially given points and acting on finite time intervals. The delay is presented as the maximum value of the unknown function over a past time discrete interval. By utilizing the Lyapunov stability theory

S. Hristova (✉)
Faculty of Mathematics and Informatics, Plovdiv University P. Hilendarski, Plovdiv 4000, Bulgaria
e-mail: snehri@gmail.com

K. Stefanova
Department of Computer Technologies, Plovdiv University P. Hilendarski, Plovdiv 4000, Bulgaria

and discrete-time Gronwall inequality, we establish some sufficient conditions for exponential stability of the zero solution.

Neural networks have received extensive interests in recent years in connection with their potential applications in signal processing, content addressable memory, pattern recognition, combinatorial optimization. It is well known that the existence of delays in neural networks causes undesirable complex dynamical behaviors such as instability, oscillation and chaotic phenomena. In practice, for computation convenience, continuous-time neural networks are often discretized to generate discrete-time neural networks. Thus, the study of discrete-time neural networks attracts more and more interests.

In this paper, we deal with a class of discrete-time neural networks with a special type of delay subject to long time lasting impulsive perturbations. The delay is presented by the maximum value of the state variables ove a past time interval with fixed interval. The basic characteristic of these perturbations is that the time of their action is not negligible small, so we consider the so called non-instantaneous impulses. We consider the general case when the connection weights between neurons are changeable in time. We apply the obtained theoretical results to obtain exponential stability criteria and new exponential convergence rate for non-instantaneous impulsive discrete-time neural networks with delays and variable connection weights.

Some discrete neural networks are considered and the theoretical results are applied. The example is computer realized by the help of Wolfram Mathematica. Following the theoretical schemes for solving the problems, the corresponding algorithms are coded to calculate the values of the solution for each step. The graphs are generated by CAS Wolfram Mathematica.

2　Statement of the Problem and Definition of Solution

We will introduce basic notations used in this paper. Most of them are well known and used in the literature. Let $\mathbb{Z}_+$ be the set of all nonnegative integers; the increasing sequence $\{n_i\}_{i=0}^{\infty}:\ n_0 = 0,\ n_i \in \mathbb{Z}_+,\ n_i \geq n_{i-1} + 3, i = 1, 2, \ldots$ and the sequence $\{d_i\}_{i=1}^{\infty}:\ d_i \in \mathbb{Z}_+,\ 1 \leq d_i \leq n_{i+1} - n_i - 2, i = 1, 2, \ldots$ be given; $\mathbb{Z}[a, b] = \{z \in \mathbb{Z}_+ : a \leq z \leq b\}, a,\ b \in \mathbb{Z}_+,\ a < b,\ \mathbb{Z}_a = \{z \in \mathbb{Z}_+ : z \geq a\}$ and $I_k = \mathbb{Z}[n_k + d_k, n_{k+1} - 1],\ k \in \mathbb{Z}_+,$ and $J_k = \mathbb{Z}[n_k + 1, n_k + d_k],\ k \in \mathbb{Z}_1,$ where $d_0 = 0$.

Let $\phi \in \mathbb{Z}(-h, 0) \to \mathbb{R}^N$ with $\|\phi\|_0 = \max\limits_{\sigma \in [-h, 0]} ||\phi(\sigma)||$, where $||.||$ is a norm in $\mathbb{R}^N$.

Consider the *initial value problem (IVP)* for the system of nonlinear *difference equation with non-instantaneous impulses*

$$x(n + 1) = Ax(n) + F\left(n, \max_{\xi \in \mathbb{Z}[n-h,n]} x(\xi)\right) \quad \text{for } n \in \bigcup_{k=0}^{\infty} I_k,$$

$$x(n) = P_k(n, x(n_k)), \quad \text{for } n \in J_k, \ k \in \mathbb{Z}_1, \tag{1}$$

$$x(n) = \phi(n), \quad n \in \mathbb{Z}[-h, 0],$$

where $x \in \mathbb{R}^n$, $x = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$, A is $N \times N$ square matrix, $F = (F_1, F_2, \ldots, F_N)$, $F_i : \bigcup_{k=0}^{\infty} I_k \times \mathbb{R}^N \to \mathbb{R}$, $P_k = (P_{k,1}, P_{k,2}, \ldots, P_{k,N})$, $P_{k,i} : J_k \times \mathbb{R}^N \to \mathbb{R}$, $i = 1, 2, \ldots, N$, $k = 1, 2, \ldots$, and h is a natural number.

Denote by $\mathbb{M}_N$ the set of all quadratic $N \times N$ dimensional matrices with the spectral norm $|A| = \sqrt{\lambda_{max}(A^T A)}$, and for any vector $x \in \mathbb{R}^N$ we will use the norm $|x| = \sqrt{\sum_{i=1}^{N} x_i^2}$. Moreover, denote by $\lambda_{min}(A)$ and $\lambda_{max}(A)$ the minimum and the maximum eigenvalue of a positive definite symmetric matrix A and $\Phi(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}$.

Usually, the difference equation describes the development of a certain phenomenon by recursively defining a sequence, each of whose terms is defined as a function of the preceding terms, once one or more initial terms are known (see, for example, [10]). Differently than that, we consider a difference equation in which the present state is also involved nonlinearly in the right side part. It makes the answer of the question about the existence of the solution more complicated.

Definition 1 The trivial solution of the system (1) is called *globally exponentially stable*, if there exist constants $L > 0$ and $\alpha \in (0, 1)$ such that for any initial function ϕ the inequality $|x(n)| \leq L\alpha^n ||\phi||_0$, $n = 1, 2, \ldots$ holds.

The constant α is called the *exponential convergence rate*.

Consider the Lyapunov equation

$$A^T H A - H = -C, \tag{2}$$

where $A, \ H, \ C \in \mathcal{M}_N$.

3   Exponential Stability of Linear Delay Discrete Equations

We will study the exponential stability of the linear system (1).

Theorem 1 (Exponential stability results). *Let*

1. *The matrix $A \in \mathbb{M}_N$ and $C \in \mathbb{M}_N$ be a positive definite matrix.*
2. *The function $F \in C(\mathbb{Z}_+ \times \mathbb{R}^N, \mathbb{R}^N)$, $F(n, 0) = 0$ for any $n \in \mathbb{Z}_+$ and there exists a constant $K > 0$ such that $|F(n, u)| \leq \sqrt{K}|u|$ for $u \in \mathbb{R}^N$, $n \in \mathbb{Z}_+$.*
4. *The functions $P_k \in C(\mathbb{Z}_+ \times \mathbb{R}^N, \mathbb{R}^N)$, $P_k(n, 0) = 0$, $k = 1, 2, \ldots$ and there exist constants $M_k > 0$ such that $|P_k(n, u)| \leq \sqrt{M_k}|u|$ for any $u \in \mathbb{R}^N$, $n \in J_k$ and $k = 1, 2, \ldots$.*

5. *There exists a solution $H \in \mathbb{M}_n$ of ([2](#)) such that $|H| M_k < 1$, $k = 1, 2, \ldots,$*
    $L_1(H) - L_2(H) < \lambda_{max}(H) - \lambda_{min}(H)$, and

$$\Phi(H)\Big(|A^T H| + |HA| + K\lambda_{max}(H)\Big)$$
$$min(C) < \lambda_{max}(H) + 0.5\Phi(H)(|A^T H| + |HA|)$$

where $L_1(H) = \lambda_{max}(H) - \lambda_{min}(C) + 0.5\Phi(H)(|A^T H| + |HA|)$, and $L_2(H)$
$= \lambda_{min}(H) - \Phi(H)K\Big(\lambda_{max}(H) + 0.5|HA| + 0.5|A^T H|\Big)$.

Then the zero solution of ([1](#)) is exponentially stable.

Proof Denote $\Theta = \max\left\{\Lambda, \frac{L_1(H)}{\lambda_{max}(H)} - L_2 + \lambda_{min}(H)\right\} < 1$,   where $\Lambda = \sup_{k \geq 1} |M_k^T H M_k|$.

Consider the function $V(x) = x^T H x$ for $x \in \mathbb{R}^N$. Then $\lambda_{min}(H)|x|^2 \leq V(x) \leq \lambda_{max}(H)|x|^2$.

Let $x(n)$, $n \in \mathbb{Z}[-h + 1, \infty)$, be a solution of the IVP ([1](#)) with the initial function ϕ.

Let $n \in \bigcup_{k=0}^{p} I_k$. Then we have

$$
\begin{aligned}
V(x(n+1)) &- V(x(n)) \\
&\leq \left(-\lambda_{min}(C) + 0.5|A^T H| + 0.5|HA|\right)|x(n)|^2 \\
&\quad + \left(\lambda_{max}(H) + 0.5|HA| + 0.5|A^T H|\right)K|\max_{\xi \in \mathbb{Z}[n-h,n]} x(\xi)|^2 \\
&\leq -\lambda_{min}(C)|x(n)|^2 + |A^T H B||x(n)|^2 \\
&\quad + \left(|A^T H B| + |B^T H B|\right)|\max_{\xi \in \mathbb{Z}[n-h,n]} x(\xi)|^2.
\end{aligned}
\tag{3}
$$

Apply the inequalities $-|x(n)|^2 \leq -\frac{V(x(n))}{\lambda_{max}(H)}$, and

$$\left|\max_{\xi \in \mathbb{Z}[n-h,n]} x(\xi)\right|^2 = |x(\eta)|^2 \leq \frac{V(x(\eta))}{\lambda_{min}(H)} \leq \frac{\lambda_{max}(H)}{\lambda_{min}(H)}\|\phi\|_0^2$$

to ([3](#)) and obtain

$$
\begin{aligned}
V(x(n+1)) &\leq \left(1 - \frac{\lambda_{min}(C)}{\lambda_{max}(H)} + \frac{|A^T H B|}{\lambda_{min}(H)}\right)V(x(n)) \\
&\quad + \left(|A^T H B| + |B^T H B|\right)\frac{\lambda_{max}(H)}{\lambda_{min}(H)}\|\phi\|_0^2.
\end{aligned}
\tag{4}
$$

From equalities

$$1 - \frac{\lambda_{min}(C)}{\lambda_{max}(H)} + 0.5 \frac{|A^T H| + |HA|}{\lambda_{min}(H)}$$

$$= \frac{1}{\lambda_{max}(H)} \left(\lambda_{max}(H) - \lambda_{min}(C) + 0.5\Phi(H)(|A^T H| + |HA|) \right)$$

and

$$\left(\lambda_{max}(H) + 0.5|HA| + 0.5|A^T H| \right) \frac{K}{\lambda_{min}(H)}$$

$$= \Phi(H)\left(\lambda_{max}(H) + 0.5|HA| + 0.5|A^T H| \right) \frac{K}{\lambda_{max}(H)}$$

and inequality (4) we get

$$V(x(n+1)) \leq \frac{L_1(H)}{\lambda_{max}(H)} V(x(n-1)) + \left(\lambda_{min}(H) - L_2(H) \right) \|\phi\|_0^2. \qquad (5)$$

Let $n = 0$. Then from inequality (5) we obtain

$$V(x(1)) \leq \left(L_1(H) - L_2(H) + \lambda_{min}(H) \right) \|\phi\|_0^2 < \sqrt[1+h]{\Theta(H)} \|\phi\|_0^2. \qquad (6)$$

Let $n = 1$. Then from inequalities (5) and (6) we get

$$V(x(2)) \leq \frac{L_1(H)}{\lambda_{max}(H)} V(x(1)) + \left(\lambda_{min}(H) - L_2(H) \right) \|\phi\|_0^2$$

$$\leq \Theta(H)\|\phi\|_0^2 < \sqrt[1+h]{\Theta^2(H)} \|\phi\|_0^2. \qquad (7)$$

Consider the following two possible cases:

Case 1. Let $m \geq n_1 - 1$. Then using induction, the inequalities $\Theta < \sqrt[p]{\Theta}$ for $p > 1$, $n \leq n_1 - 1 < m + 1$, i.e. $\frac{m+1}{n} > 1$ for $n \in I_0$ and inequality (5), we prove that

$$V(x(n+1)) \leq \sqrt[1+h]{\Theta^n(H)} \|\phi\|_0^2, \quad \text{for } n \in \mathbb{Z}[0, n_1 - 1].$$

Case 2. Let $h < n_1 - 1$. Then using induction and inequality (5) we prove that

$$V(x(n+1)) \leq \sqrt[1+h]{\Theta^n(H)} \|\phi\|_0^2, \text{ for } n = 1, 2, \ldots, h.$$

Then for $k = 1, 2, \ldots, n_1 - h$ we get

$$V(x(h+k)) \leq \frac{L_1(H)}{\lambda_{max}(H)} V(x(h)) + \left(\lambda_{min}(H) - L_2(H) \right) \|\phi\|_0^2$$

$$\leq \Theta(H) \sqrt[1+h]{\Theta^k(H)} \|\phi\|_0^2 < \sqrt[1+h]{\Theta^{h+k}(H)} \|\phi\|_0^2. \qquad (8)$$

By induction we prove that

$$V(x(n_1 + k)) < \sqrt[1+h]{\Theta^{n_1+k-1}} v_0, \quad k = 0, 1, \ldots, d_1.$$

Let $n = n_1 + d_1$. Then using the inequalities $n_1 + d_1 + 1 - h > 0$ and (5) we get

$$
\begin{aligned}
V(x(n_1 + d_1 + 1)) &\leq \frac{1}{\lambda_{max}(H)} L_1(H) \sqrt[1+h]{\Theta^{n_1+d_1}} \|\phi\|_0^2 \\
&+ \left(-L_2(H) + \lambda_{min}(H) \right) \sqrt[1+h]{\Theta^{n_1+d_1+1-h}} \|\phi\|_0^2 \\
&< \Theta^{\frac{n_1+d_1+1}{h+1}} \|\phi\|_0^2.
\end{aligned}
\tag{9}
$$

Similarly, $V(x(n_1 + d_1 + 2)) < \Theta^{\frac{n_1+d_1+2}{h+1}} \|\phi\|_0^2$.
By induction process we prove the validity of the inequality

$$V(x(n)) < \Theta^{\frac{n}{h+1}} \|\phi\|_0^2 \quad \text{for all } n \in \mathbb{Z}_1. \tag{10}$$

Therefore, we get $|x(n)| < L\|\phi\|_0 \alpha^n$, for all $n \in \mathbb{Z}_1$ with $\alpha = \sqrt[2(h+1)]{\Theta} < 1$, and $L = \sqrt{\Phi(H)} = \sqrt{\frac{\lambda_{max}(H)}{\lambda_{min}(H)}}$.

$\square$

4   Exponential Stability of Discrete Neural Networks with Maximum, Non-instantaneous Impulses and Time Variable Connection Weights

Consider the following neural network modeled by discrete system with maximum, non-instantaneous impulses and time variable connection weights

$$
\begin{aligned}
u_i(n+1) &= a_i u_i(n) + \sum_{j=1}^{n} \Psi_{ij}(n) f_j \left(\max_{\xi \in \mathbb{Z}[n-h,n]} u_j(\xi) \right) + G_i \\
&\quad \text{for } n \in \bigcup_{k=0}^{\infty} I_k, \ i \in \mathbb{Z}[1, N], \\
u_i(n) &= M_{ik} u_i(n_k) + \sum_{j=1}^{n} \Phi_{ij}^k(n) S_j(u_j(n_k)) + Q_{ik}, \\
&\quad \text{for } n \in J_k, \ k \in \mathbb{Z}_1, \ i \in \mathbb{Z}[1, N], \\
u_i(n) &= \phi_i(n), \ n \in \mathbb{Z}[-h, 0], \ i \in \mathbb{Z}[1, N],
\end{aligned}
\tag{11}
$$

where $u_i(n)$, $i \in \mathbb{Z}[1, N]$, denotes the state of the i-th neuron at discrete time n, a_i, $i \in \mathbb{Z}[1, N]$, represents the passive decay rate, f_j is the neuron activation function with $f_j(0) = 0$, G_i is the exogenous input, P_j is the neuron output signal function

which is a continuous function, $\Psi_{ij}(n)$ and $\Phi_{ij}^k(n)$ denote the connection weight from the neuron j to the neuron i at time n, $m \in \mathbb{Z}(1)$ is the transmission delay, $\phi_i(n), n \in \mathbb{Z}[-h, 0]$ is the initial function for the i-th neuron.

For any $u = (u_1, u_2, \ldots, u_N)$ we denote $f(u) = \big(f_1(u_1), f_2(u_2), \ldots, f_N(u_N)\big)$ and $S(u) = \big(S_1(u_1), S_2(u_2), \ldots, S_N(u_N)\big)$.

We will introduce the following assumptions:

A1. The functions $f_i \in C(\mathbb{R}, \mathbb{R})$, $i \in \mathbb{Z}[1, N]$, and there exist positive constants $L_i, i \in \mathbb{Z}[1, N]$, such that $|f_i(u) - f_i(v)| \leq L_i|u - v|, u, v \in \mathbb{R}$.

A2. The functions $S_i \in C(\mathbb{R}, \mathbb{R})$, $i \in \mathbb{Z}[1, N]$, and there exist positive constants $K_i, i \in \mathbb{Z}[1, N]$, such that $|S_i(u) - S_i(v)| \leq K_i|u - v|, u, v \in \mathbb{R}$.

A3. The functions $\Psi_{ij} : \bigcup_{k=0}^{\infty} I_k \to \mathbb{R}, i, j \in \mathbb{Z}[1, N]$, and $\Phi_{ij}^k : J_k \to \mathbb{R}, i, j \in \mathbb{Z}[1, N], k \in \mathbb{Z}_+$, are bounded, i.e. there exists constants $\beta_{ij}^k > 0, \gamma_{ij} > 0$ such that $|\Psi_{ij}(n)| \leq \beta_{ij}$ for $n \in \bigcup_{k=0}^{\infty} I_k$ and $|\Phi_{ij}^k(n)| \leq \gamma_{ij}^k$ for $n \in J_k, k \in \mathbb{Z}_+, i, j \in \mathbb{Z}[1, N]$.

In the non-homogeneous case we will define an equilibrium of the model (11):

Definition 2 ([2]) A vector $u^* \in \mathbb{R}^N : u^* = (u_1^*, u_2^*, \ldots, u_N^*)$ is said to be an equilibrium point of the impulsive discrete-time neural network (11) if it satisfies the equalities

$$
\begin{aligned}
u_i^* &= a_i u_i^* + \sum_{j=1}^{N} \Psi_{ij}(n) f_j(u_j^*) + G_i \quad \text{for } n \in \bigcup_{k=0}^{\infty} I_k, \ i \in \mathbb{Z}[1, N], \\
u_i^* &= M_{ik} u_i^* + \sum_{j=1}^{N} \Phi_{ij}^k(n) S_j(u_j^*) + Q_{ik} \quad \text{for } n \in J_k, \ k \in \mathbb{Z}_1.
\end{aligned}
\tag{12}
$$

Let (11) has an equilibrium $u^* \in \mathbb{R}^N$. Substitute $x = u - u^* \in \mathbb{R}^N$ in (11) and obtain

$$
\begin{aligned}
x_i(n+1) &= a_i x_i(n) + \sum_{j=1}^{N} \Psi_{ij}(n) \mathcal{F}_j \left(\max_{\xi \in \mathbb{Z}[n-h, n]} x_j(\xi) \right), \ n \in \bigcup_{k=0}^{\infty} I_k, \\
x_i(n) &= M_{ik} x_i(n_k)) + \sum_{j=1}^{N} \Phi_{ij}^k(n) r_j(x_j(n_k)), \ n \in J_k, \ k \in \mathbb{Z}_1,
\end{aligned}
\tag{13}
$$

where $\mathcal{F}_i(y) = f_i(y - u_i^*) - f_i(u_i^*)$ and $r_i(y) = S_i(y - u_i^*) - S_i(u_i^*)$, $i = 1, 2, \ldots, N$, for $y \in \mathbb{R}$.

The stability behavior of the equilibrium of (11) is equivalent to the stability behavior of zero solution of (13).

The system (13) could be written in the matrix form (1) where

$$A = \begin{bmatrix} a_1 & 0 & 0 & \ldots\ldots & 0 \\ 0 & a_2 & 0 & \ldots\ldots & 0 \\ \ldots\ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots\ldots & a_N \end{bmatrix}, \quad \mathcal{B}(n) = \begin{bmatrix} \Psi_{11}(n) & \Psi_{12}(n) & \Psi_{13}(n) & \ldots\ldots & \Psi_{1N}(n) \\ \Psi_{21}(n) & \Psi_{22}(n) & \Psi_{23}(n) & \ldots\ldots & \Psi_{2N}(n) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \Psi_{N1}(n) & \Psi_{N2}(n) & \Psi_{N3}(n) & \ldots\ldots & \Psi_{NN}(n) \end{bmatrix},$$

$$\mathcal{F}(u) = \begin{bmatrix} \mathcal{F}_1(u_1) \\ \mathcal{F}_2(u_2) \\ \ldots \\ \mathcal{F}_N(u_N) \end{bmatrix}, \quad r(u) = \begin{bmatrix} r_1(u_1) \\ r_2(u_2) \\ \ldots \\ r_N(u_N) \end{bmatrix}, \quad \mathcal{M}_k = \begin{bmatrix} M_{1k} & 0 & 0 & \ldots\ldots & 0 \\ 0 & M_{2k} & 0 & \ldots\ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots\ldots & M_{Nk} \end{bmatrix},$$

$$\mathcal{G}_k(n) = \begin{bmatrix} \Phi_{11}^k(n) & \Phi_{12}^k(n) & \gamma_{13}^k(n) & \ldots\ldots & \Phi_{1N}^k(n) \\ \Phi_{21}^k(n) & \Phi_{22}^k(n) & \Phi_{23}^k(n) & \ldots\ldots & \Phi_{2N}^k(n) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \Phi_{N1}^k(n) & \Phi_{N2}^k(n) & \Phi_{N3}^k(n) & \ldots\ldots & \Phi_{NN}^k(n) \end{bmatrix},$$

$u = (u_1, u_2, \ldots, u_N)$, $F = (F_1, F_2, \ldots, F_N)$, $F(n, u) = \mathcal{B}(n)\mathcal{F}(u)$, $P_k = (P_{k,1}, P_{k,2}, \ldots, P_{k,N})$, $P_k(n, u) = \mathcal{G}_k(n)r(u) + \mathcal{M}_k u^T$.

From assumption (A1) and the inequality $\left(\sum_{j=1}^{N} \gamma_j u_j \right)^2 \le N \sum_{j=1}^{N} \left(\gamma_j u_j \right)^2$ we have

$|F(n, u)| \le \sqrt{N \sum_{i=1}^{N} \max_j (L_j \beta_{ij})^2} \sqrt{\sum_{j=1}^{N} u_j^2}$, i.e. the condition 3 of Theorem 1 is satisfied with $K = N \sum_{i=1}^{N} \max_j (L_j \beta_{ij})^2$.

From assumption (A2) and the inequality $\left(\sum_{j=1}^{N} \gamma_j u_j \right)^2 \le N \sum_{j=1}^{N} \left(\gamma_j u_j \right)^2$ we have

$|P_k(n, u)| \le \sqrt{2 \left(\max_i M_{ik}^2 + \sum_{i=1}^{N} N \max_j \left(\gamma_{ij}^k K_j \right)^2 \right) \sum_{i=1}^{N} u_i^2}$, i.e. the condition 4 of Theorem 1 is satisfied with $M_k = 2 \left(\max_i M_{ik}^2 + \sum_{i=1}^{N} N \max_j \left(\gamma_{ij}^k K_j \right)^2 \right)$.

Theorem 2 *Let the conditions (A1)–(A3) be satisfied and:*

1. *The discrete model (11) has an equilibrium u^*.*
2. *The constants a_i, $M_{ik} \in \mathbb{R}$, β_{ij}, $\gamma_{ij}^k > 0$, G_i, $Q_{ik} \in \mathbb{R}$, i, $j \in \mathbb{Z}[1, N$ $k \in \mathbb{Z}_1$.*
3. *The inequalities $\max_i M_{ik}^2 + N \sum_{i=1}^{N} \max_j \left(\gamma_{ij}^k K_j \right)^2 < 1$ for $k \in \mathbb{Z}_1$, $\max_i a_i^2 + \max_i |a_i| + N(1 - \max_i |a_i|) \sum_{i=1}^{N} \max_j (L_j \beta_{ij})^2) < 1$, and $\max_i a_i^2 + \max_i |a_i| + K \left(1 + \max_i |a_i| \right) < 1$ hold.*

Then the equilibrium point of the difference neural network with non-instantaneous impulses (11) is exponentially stable with a rate $\alpha = {}^{2(h+1)}\!\sqrt{\Theta} < 1$. where $\Theta = \max \left\{ \Lambda, \max_i a_i^2 + \max_i |a_i| + K \left(1 + \max_i |a_i| \right) \right\} < 1$ with $\Lambda = \sup_{k \ge 1} \left(\max_i M_{ik}^2 + \sum_{i=1}^{N} N \max_j \left(\gamma_{ij}^k K_j \right)^2 \right)$.

Proof Let $H = E$, where $E \in \mathbb{M}_n$ is the unit matrix. Then (2) is satisfied with $C \in \mathbb{M}_n : c_{ii} = 1 - a_i^2$ and $c_{ij} = 0$ for $i \neq j$. Then $\lambda_{max}(H) = \lambda_{min}(H) = \phi(H) = 1$ and $\lambda_{min}(C) = 1 - \max_i a_i^2$, $|A| = \max_i |a_i|$. According to Theorem 1 the zero solution of (13) is exponentially stable.

5 Application

Consider a system with three agent with constant connection weights modeled by the following discrete model of neural network

$$
\begin{aligned}
u_1(n+1) &= \frac{1}{2}u_1(n) + \frac{1}{8}\sin\left(\max_{\xi\in\mathbb{Z}(n-3,n)} u_1(\xi)\right) - \frac{1}{4}\sin\left(\max_{\xi\in\mathbb{Z}(n-3,n)} u_2(\xi)\right) \\
&\quad + \frac{1}{16}\max_{\xi\in\mathbb{Z}(n-3,n)} u_3(\xi) + 1 \\
u_2(n+1) &= \frac{1}{3}u_2(n) + \frac{1}{4}\sin\left(\max_{\xi\in\mathbb{Z}(n-3,n)} u_1(\xi)\right) + \frac{1}{8}\sin\left(\max_{\xi\in\mathbb{Z}(n-3,n)} u_2(\xi)\right) + 2 \\
u_3(n+1) &= \frac{1}{4}u_3(n) + \frac{1}{16}\sin\left(\max_{\xi\in\mathbb{Z}(n-3,n)} u_1(\xi)\right) - \frac{1}{8}\sin\left(\max_{\xi\in\mathbb{Z}(n-3,n)} u_2(\xi)\right) \\
&\quad + \frac{1}{16}\max_{\xi\in\mathbb{Z}(n-3,n)} u_3(\xi)) + 1 \quad \text{for } n \in \bigcup_{k=0}^{\infty} I_k,
\end{aligned}
\tag{14}
$$

with non-instantaneous impulses for $n \in J_k, \ k \in \mathbb{Z}_1$

$$
\begin{aligned}
u_1(n) &= \frac{1}{2}u_1(n_k) + \frac{1}{8}\sin(u_1(n_k)) - \frac{1}{4}\sin(u_2(n_k)) + \frac{1}{16}u_3(n_k) + 1 \\
u_2(n) &= \frac{1}{3}u_2(n_k) + \frac{1}{4}\sin(u_1(n_k)) + \frac{1}{8}\sin(u_2(n_k)) + 2 \\
u_3(n) &= \frac{1}{4}u_3(n_k) + \frac{1}{16}\sin(u_1(n_k)) - \frac{1}{8}\sin(u_2(n_k)) + \frac{1}{16}u_3(n_k) + 1,
\end{aligned}
\tag{15}
$$

and initial conditions

$$
u_i(n) = \phi_i(n), \quad n = -3, \ -2, \ -1, \ 0,
\tag{16}
$$

where $n_0 = 0$, $d_0 = 0$, $n_1 = 4$, $d_1 = 6$, $n_2 = 18$, $d_2 = 5$, $n_3 = 33$, $d_3 = 7$, $n_4 = 45$. In this particular case $I_0 = \mathbb{Z}[0, 3]$, $I_1 = \mathbb{Z}[10, 17]$, $I_2 = \mathbb{Z}[23, 32]$, $I_3 = \mathbb{Z}[40, 44]$ $J_1 = \mathbb{Z}[5, 10]$, $J_2 = \mathbb{Z}[19, 23]$, $J_3 = \mathbb{Z}[34, 40]$. (The whole interval is $\mathbb{Z}[-3,45]$).

The point $u^* = (2.40568, 3.23436, 1.53241)$ is the equilibrium point of (14), (15). The conditions of Theorem 2 are reduced to $0.25 + 3(\frac{1}{16} + \frac{1}{16} + \frac{1}{64}) = 0.680556 < 1$, and $0.25 + 0.5 + 3(1 - 0.5)(\frac{1}{16} + \frac{1}{16} + \frac{1}{64}) = 0.965278 < 1$. Therefore, the equilibrium u^* is exponentially stable with a rate $\alpha = \sqrt[8]{0.965278} \approx$
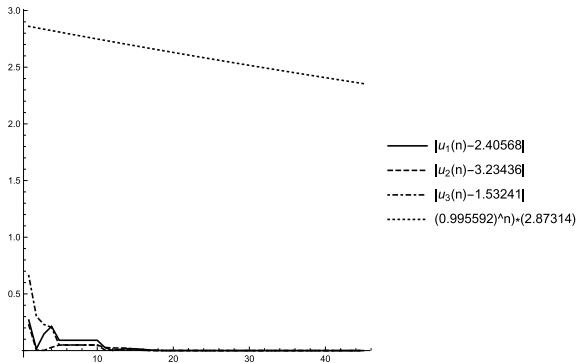
Fig. 1 Graphs of the bound $\alpha^n \|\phi - u^*\|_0 = 2.87314 * 0.995592^n$ and the differences $|u_1(n) - 2.40568|$, $|u_2(n) - 3.23436|$, $|u_3(n) - 1.53241|$ for $n \in [1, 45]$

Table 1 Values of the bound $\alpha^n \|\phi - u^*\|_0 = 2.87314 * 0.995592^n$ and the differences $|u_1(n) - 2.40568|$, $|u_2(n) - 3.23436|$, $|u_3(n) - 1.53241|$ for $n = 1, \ldots, 45$

n	$2.87314 * 0.995592^n$	$\|u_1(n) - 2.40568\|$	$\|u_2(n) - 3.23436\|$	$\|u_3(n) - 1.53241\|$
1	2.86048	0.269342	0.226707	0.660759
2	2.84787	0.0153586	0.00404188	0.301958
3	2.83531	0.146954	0.00197866	0.229919
4	2.82281	0.213725	0.0285321	0.204716
5	2.81037	0.0909878	0.0490802	0.0496394
…	…	…	…	…
38	2.42913	0.000020443	$1.71883 * 10^{-6}$	$3.53387 * 10^{-6}$
39	2.41842	0.000020443	$1.71883 * 10^{-6}$	$3.53387 * 10^{-6}$
40	2.40776	0.000020443	$1.71883 * 10^{-6}$	$3.53387 * 10^{-6}$
41	2.39714	$6.77539 * 10^{-6}$	$1.13363 * 10^{-7}$	$1.04167 * 10^{-6}$
42	2.38658	$4.65835 * 10^{-7}$	$2.03659 * 10^{-6}$	$2.0664 * 10^{-6}$
43	2.37606	$2.16839 * 10^{-6}$	$3.92243 * 10^{-6}$	$2.09433 * 10^{-6}$
44	2.36558	$3.24319 * 10^{-6}$	$5.03918 * 10^{-6}$	$1.98103 * 10^{-6}$
45	2.35516	$3.681 * 10^{-6}$	$5.61059 * 10^{-6}$	$1.90291 * 10^{-6}$

0.995592 and $L = \sqrt{\Phi(H)}$, $\Phi(H) = \frac{\lambda_{max}(H)}{\lambda_{min}(H)}$, $H = E$, where $E \in \mathbb{M}_n$ is the unit matrix.

Consider the solution $\tilde{u}(n)$ of the discrete model (14), (15), (16) with initial functions $\phi_1(n) = 3n + 2$, $\phi_2(n) = 2n + 2$, $\phi_3(n) = n + 4$, $n = -3, -2, -1, 0$. In this case $\|\phi - u^*\|_0 = \max_{\sigma \in [-h, 0]} \|\phi(\sigma) - u^*\| = 2.87314$. The graphs of the differences $|u_1(n) - 2.40568|$, $|u_2(n) - 3.23436|$, $|u_3(n) - 1.53241|$ and the bound $\alpha^n \|\phi - u^*\|_0 = 2.87314 * 0.995592^n$ are given on Fig. 1 and Table 1. From both it could be seen the equilibrium point u^* is exponentially stable.

References

1. Hritova, S., Stefanova, K., Golev, A.: Exponential stability for difference equations with maximum:theoretical study and computer simulations, Intern. J. Diff. Eq. Appl. 20 (2), 169–178, (2021).
2. Hritova, S., Stefanova, K,: Exponential stability of discrete neural networks with non-instantaneous impulses, delays and variable connection weights with computer simulations, Intern. J. Appl. Math. 33, (2), 187–209, (2020).
3. Agarwal, R.: Difference Equations and Inequalities. Theory, Methods and Applications, in: Monographs and Textbooks in Pure and Applied Mathematics, second ed., Marcel Dekker, Inc., New York, (2000).
4. Agarwal, R., Hristova, S., O'Regan, R.: Non-instantaneous impulses in Caputo fractional differential equations, Frac. Calc. Appl. Anal. 20, (3), 595–622, (2017).
5. Agarwal, R., O'Regan, D., Hristova, S.: Monotone iterative technique for the initial value problem for differential equations with non-instantaneous impulses, Appl. Math. Comput. 298, (1), 45–56, (2017).
6. Alshammari, M., Raffoul, Y.: Exponential stability and instability in multiple delays difference equations, Khayyam J. Math. 1, (2), 174–184, (2015).
7. Barreira, L., Valls, C.: Stability in delay difference equations with nonuniform exponential behavior, J. Diff. Eq. 238, (2), 470–490, (2007).
8. Danca, M., Feckan, M., Pospsil, M. :Difference equations with impulses, Opuscula Math. 39 (1), 5–22, (2019).
9. Diblk, J., Khusainov, D., Bastinec, J., Sirenko, A.: Exponential stability of linear discrete systems with constant coefficients and single delay, Appl. Math. Lett. 51, 68–73, (2016).
10. Elaydi, S.: An Introduction to Difference Equations, Springer, New York, NY, USA, 3rd edition, (2005).
11. Laksmikantham, V., Trigiante, D.: Theory of Difference Equations (Numerical Methods and Applications), second ed., Marcel Dekker, (2002).
12. Li, H., Li, C., Huang, T,: Comparison principle for difference equations with variable-time impulses, Modern Physics Letters B 32 (2) (2018), https://doi.org/10.1142/S0217984918500136.

Prediction of the Time Series by the Various Types of Artificial Neural Networks by the Example of Different Time Intervals of the Content of Methane in the Atmosphere

Anastasia Butorova ⬡**, Alexander Buevich** ⬡**, Andrey Shichkin** ⬡**,**
Aleksandr Sergeev ⬡**, Elena Baglaeva** ⬡**, Marina Sergeeva** ⬡**,**
Irina Subbotina ⬡**, and Julian Vasilev** ⬡

Abstract The paper presents a forecast of changes in the methane content in the air in surface layer of the atmosphere. The forecast was made by the models based on the two most common types of artificial neural networks (ANN): A nonlinear autoregressive neural networks with exogenous inputs (NARX) and Elman neural network (ENN). For training, we used the Levenberg–Marquardt learning algorithm. The data were collected upon monitoring the greenhouse gases on Bely Island, Yamal-Nenets Autonomous Okrug, Russia. For the comparison, the three time intervals with the different patterns of changes in methane content were chosen. To assess the prediction accuracy of the models, we used the mean absolute error, mean square error, and the standardized measure of the model prediction error degree—the index of agreement. The model based on the artificial neural network NARX for all simulated intervals was the most accurate.

Keywords Time series · Artificial neural networks · Forecasting

A. Butorova (✉) · A. Buevich · A. Shichkin · A. Sergeev · E. Baglaeva · M. Sergeeva · I. Subbotina
Institute of Industrial Ecology UB RAS, S. Kovalevskoy Str., 20, Ekaterinburg, Russia 620990
e-mail: a.s.butorova@urfu.ru

A. Butorova
Ural Federal University, Mira Str., 19, Ekaterinburg, Russia 620002

J. Vasilev
University of Economics, 77 Knyaz Boris I Blvd, 9002 Varna, Bulgaria

1 Introduction

Time series forecasting is becoming more and more popular in various fields of science. This is largely due to the rapid development of the methods and algorithms based on artificial neural networks. Such methods make it possible to predict complex systems with high accuracy [1–9].

In this paper, we compared the prediction results of the two most efficient ANNs using the example of predicting changes in the surface methane concentration for three different time intervals.

2 Materials and Methods

The main greenhouse gases were monitored on the territory of Bely Island (Russia), located in the southern part of the Arctic Ocean (Fig. 1). The island is located far from industrial sites and is an excellent natural testing ground for such research.

The concentrations of the greenhouse gases were measured using a Picarro G2401 cavity ring-down spectrometer.
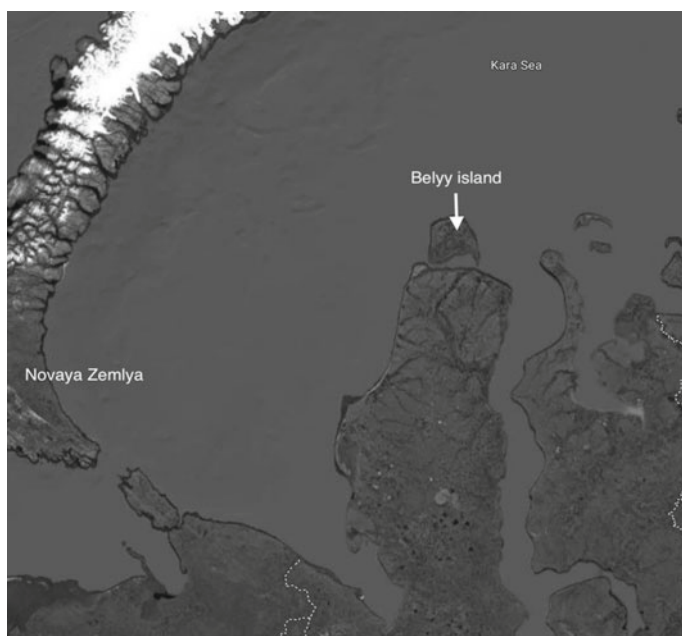
Fig. 1 Place of measurements (Google Earth)

For forecasting, three time intervals were selected, which represented the average hourly data on the methane content. Each time interval consisted of 168 measurements of methane concentration. For prediction, the intervals were divided into two parts. The first part of the 144-long dimension was a training time series. The second part of 24 dimensions was the predicted time series.

Each time interval was characterized by its own characteristics in the dynamics of changes in methane concentration. So, the interval 1 did not have a pronounced dependence on the time of day, it was more chaotic. Interval 2 had a noticeable cyclicality depending on the time of day. During the day, the concentration of methane increased, and fell at night. Interval 3 also had a similar relationship, with the average methane concentration increasing from the beginning to the end of this interval.

The two ANNs were chosen for forecasting: ENN and NARX. In such studies, it is these types of ANNs that demonstrate the most accurate results [10–14].

Elman's artificial neural network is a simple recurrent neural network, which consists of three layers—the input (distribution), the hidden, and output (processing) layers. The hidden layer has feedback on itself. Unlike the usual feedforward network, the input image of the recurrent network is not one vector, but a sequence of vectors, the vectors of the input image are fed to the input in a given order, while the new state of the hidden layer depends on its previous states.

The NARX network is a recurrent dynamic network with multi-level feedback. The architectural structure of recurrent neural networks can take many different forms, but the simplest ones use architectural neural networks with feedback. This model has a single input that applies to the memory on the delay lines and a single output shorted to the input through the memory on the delay lines. The contents of these two memory blocks are used to power the input layer of the perceptron. In addition, the use of feedbacks makes it possible to use the description of recurrent neural networks NARX in the form of a set of states, which makes them convenient devices in nonlinear forecasting and modeling.

The structure of ANNs was determined by computer simulation. In the hidden layer, the number of neurons was chosen using the minimum RMSE—root mean square error. The range of variation of neurons was between 5 and 25. The training of each network was carried out 500 times. After that, the best networks were selected. The input was a time interval. The number of neurons that corresponded to the smallest error formed the hidden layer. The concentration of methane corresponding to the time interval was presented in the output layer.

To train the ANNs, the Levenberg–Marquardt learning algorithm was applied.

The following indices were used to assess the accuracy of the prediction: mean absolute error (MAE) (1), root mean square error (RMSE) (2), and index of agreement (d), which is the standardized measure of the model prediction error degree and varies from 0 to 1, where 1 indicates a complete match and 0 indicates a complete lack of agreement [15].

$$MAE = \frac{\sum_{i=1}^{n} |P(\boldsymbol{x}_i) - M(\boldsymbol{x}_i)|}{n}, \tag{1}$$

Table 1 The results of prediction of the methane concentration for a 24-h time series

Method	MAE, ppm	RMSE, ppm	d
Interval 1			
Elman	0.032	0.039	0.47
NARX	0.025	0.032	0.62
Interval 2			
Elman	0.018	0.021	0.40
NARX	0.010	0.011	0.52
Interval 3			
Elman	0.007	0.011	0.56
NARX	0.004	0.005	0.63

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P(x_i) - M(x_i))^2}{n}}, \qquad (2)$$

$$d = 1 - \frac{\sum |P(x_i) - M(x_i)|}{\sum (|P(x_i) - \overline{m}| + |M(x_i) - \overline{m}|)}, \qquad (3)$$

where $P(x_i)$ is the concentration in the location x_i that was predicted, $M(x_i)$ is the concentration in the location x_i that was measured, $\overline{m}$ is the average concentration, and n is the number of points.

3 Results and Discussion

For Elman and NARX networks in the hidden layer, the final number of neurons was 20. The parameters used for comparison of the performance of different methods are presented in Table 1. The best values for the test interval for NARX are in bold.

Both networks accurately predicted changes in methane concentration for all time intervals. The NARX network showed the best accuracy based on all indices (Fig. 2).

4 Conclusion

Both ANN-based models demonstrated a sufficiently accurate forecast for changes in the methane concentration at three 24-h time intervals. A more accurate prediction was obtained for intervals 2 and 3 with a pronounced dependence on the time of day. It can be argued that the approach used is suitable for predicting changes in the concentration of greenhouse gases.
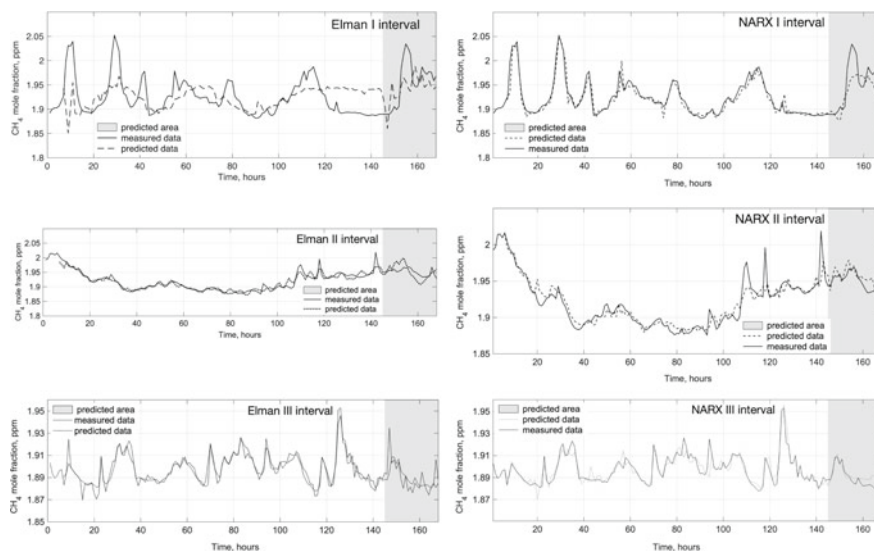
Fig. 2 Comparison of different prediction approaches

References

1. Dmitriev, A. V.: Time series prediction of morbidity using artificial neural networks. Biomedical Engineering 47(1), 43–45 (2013).
2. Stojadinovic, S.: Prediction of fly rock launch velocity using artificial neural networks. Neural Computing & Application 27, 515–524 (2016).
3. Zhang, G.: Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 5, 159–175 (2003).
4. Russo, A., Raischel, F., Lind, P.: Air quality prediction using optimal neural networks with stochastic variables. Atmospheric Environment 79, 822–830 (2013).
5. Erdil, A., Arcaklioglu, E.: The prediction of meteorological variables using artificial neural network. Neural Computing & Application 22, 1677–1683 (2013).
6. Zhou, Q., Jiang, H., Wang, J., Zhou, J.: A hybrid model for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network. Science of the Total Environment 496, 264–274 (2014).
7. Strnad, D., Nerat, A., Kohek, Z.: Neural network models for group behavior prediction: a case of soccer match attendance. Neural Computing and Applications 28(2), 287–300 (2017).
8. Dimitrov, G., Panayotova, G., Garvanov, I., Bychkov, O., Petrov, P., Angelov, A.: Performance analysis of the method for social search of information in university information systems. In 3rd International Conference on Artificial Intelligence and Pattern Recognition (AIPR), pp. 149–153. IEEE, Lodz, Poland (2016).

9. Mohanty, K., Majumdar, T. J.: Using artificial neural networks for synthetic surface fitting and the classification of remotely sensed data. International Journal of Applied Earth Observation and Geoinformation 1(1), 78–84 (1999).

10. Vasilev, J., Turygina, V. F., Kosarev, A. I., Nazarova, Y. Y.: Mathematical optimization in environmental economics. Algorithm of gradient projection method. In International Multi-disciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management (SGEM), pp. 349–356. STEF92 TECHNOLOGY LTD, Albena, Bulgaria (2016).

11. Tarasov, D. A., Buevich, A. G., Sergeev, A. P., Shichkin, A. V.: High Variation Topsoil Pollution Forecasting in the Russian Subarctic: Using Artificial Neural Networks Combined with Residual Kriging, Applied Geochemistry 88(B), 188–197, (2018).

12. Menezes, J. M. Jr., Barreto, G.A.: Long-term time series prediction with the NARX network: An empirical evaluation. Neurocomputing 71, 3335–3343 (2008).

13. Ardalani-Farsa M., Zolfaghari, S.: Chaotic time series prediction with residual analysis method using hybrid Elman–NARX neural networks. Neurocomputing 73, 2540–2553 (2010).

14. Zemouri, R., Gouriveau, R., Zerhouni, N.: Defining and applying prediction performance metrics on a recurrent NARX time series model. Neurocomputing 73, 2506–2521 (2010).

15. Willmott, C.J.: On the Validation of Models. Physical Geography 2, 184–194 (1981).

Highly Accurate Scrambled Stochastic Approaches for Multidimensional Sensitivity Analysis in Air Pollution Modeling

Venelin Todorov and Slavi Georgiev

Abstract A thorough experimental investigation of Monte Carlo algorithms based on Halton and Sobol scrambling algorithms has been performed for the first time for the model under consideration. For scrambling Halton sequence, we use a permutation of the radical inverse coefficients derived by applying a reverse-radix operation to all of the possible coefficient values. For scrambling Sobol sequence, we use random linear scramble combined with a random digital shift. These methods have never been applied and compared before for the Unified Danish Eulerian model and this motivates our study. The novelty of the proposed approaches is that Halton scrambling algorithm has been applied for the first time to sensitivity studies of the particular air pollution model. The computational experiments demonstrate that the proposed stochastic approaches are efficient for the considered multidimensional integrals evaluation and especially for computing small by value sensitivity indices which are very important for the reliability of the mathematical model.

Keywords Monte carlo · Halton · Sobol · Sensitivity analysis · Air pollution

V. Todorov · S. Georgiev
Department of Information Modeling, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. Georgi Bonchev Str., Block 8, 1113 Sofia, Bulgaria
e-mail: sggeorgiev@math.bas.bg

V. Todorov (✉)
Department of Parallel Algorithms, Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Acad. Georgi Bonchev Str., Block 25A, 1113, Sofia, Bulgaria
e-mail: vtodorov@math.bas.bg

S. Georgiev
Department of Applied Mathematics and Statistics, Faculty of Natural Sciences and Education, University of Ruse, 8 Studentska Str., 7004 Ruse, Bulgaria

1   Introduction

The reliability of the large-scale mathematical models when used to make decisions, appears to be a significant problem [8–10, 18]. In improving the reliability, the sensitivity analysis of the model outputs with respect to artificial or natural perturbation of the model inputs plays a central role. By definition [4, 17], sensitivity analysis is a procedure for studying how sensitive are the output results of large-scale mathematical models to some uncertainties of the input data. A large-scale mathematical model, describing remote transport of air pollutants—**Uni**fied **D**anish **E**ulerian **M**odel (UNI-DEM), is used to derive the input data for sensitivity analysis. This model was developed at the Danish National Environmental Research Institute (http://www2.dmu.dk/AtmosphericEnvironment/DEM/, [19–21]). The results described here can be used for increasing the reliability of the mathematical model results, and identifying input parameters that should be measured more precisely. UNI-DEM is employed in the present study since it is one of the most advanced large-scale mathematical models that appropriately describe the physical and chemical processes in full. One of the most appealing features of UNI-DEM is its advanced chemical scheme, the Condensed CBM IV, which consider a large number of chemical species and numerous reactions between them, of which the ozone is one of the most important pollutants for its central role in many practical applications of the results. The computational domain [3] is large enough to completely covers the European region and the Mediterranean, and the calculations are done for a certain time period. It is appropriate to explore the relationships between input parameters and model outputs to identify the reliability of the model. Sensitivity analysis is most often deduced by variance-based methods. Sobol variance-based approach for global sensitivity analysis has been applied to compute the corresponding sensitivity measures which leads to multidimensional integrals [15].

2   Sensitivity Analysis and Stochastic Methods

2.1   Sobol Approach for Global Sensitivity Analysis

It is assumed that the mathematical model can be presented by a model function

$$u = f(x), \quad \text{where} \quad x = (x_1, x_2, \ldots, x_d) \in U^d \equiv [0; 1]^d \tag{1}$$

is a vector of input parameters with a joint **p**robability **d**ensity **f**unction (p.d.f.) $p(x) = p(x_1, \ldots, x_d)$.

The concept of Sobol approach is based on a decomposition of an integrable model function f into terms of increasing dimensionality [15, 17]:

$$f(\mathbf{x}) = f_0 + \sum_{\nu=1}^{d} \sum_{l_1 < \ldots < l_\nu} f_{l_1 \ldots l_\nu}(x_{l_1}, x_{l_2}, \ldots, x_{l_\nu}), \tag{2}$$

where f_0 is a constant. The representation (2) is referred to as the ANOVA-representation of the model function $f(\mathbf{x})$ if each term is chosen to satisfy the following condition [16]:

$$\int_0^1 f_{l_1 \ldots l_\nu}(x_{l_1}, x_{l_2}, \ldots, x_{l_\nu}) \mathrm{d}x_{l_k} = 0, \quad 1 \le k \le \nu, \quad \nu = 1, \ldots, d.$$

It guarantees that the functions in the right-hand side of (2) are defined in a unique way, where $f_0 = \int_{U^d} f(\mathbf{x}) \mathrm{d}x$. The quantities

$$\mathbf{D} = \int_{U^d} f^2(\mathbf{x}) \mathrm{d}x - f_0^2, \quad \mathbf{D}_{l_1 \ldots l_\nu} = \int f_{l_1 \ldots l_\nu}^2 \mathrm{d}x_{l_1} \ldots \mathrm{d}x_{l_\nu} \tag{3}$$

are the so-called total and partial variances, respectively. A similar decomposition holds for the total variance that is represented by the corresponding partial variances: $\mathbf{D} = \sum_{\nu=1}^{d} \sum_{l_1 < \ldots < l_\nu} \mathbf{D}_{l_1 \ldots l_\nu}$.

The Sobol global sensitivity indices [14, 16] are the main sensitivity measures that follow the Sobol approach and are defined as

$$S_{l_1 \ldots l_\nu} = \frac{\mathbf{D}_{l_1 \ldots l_\nu}}{\mathbf{D}}, \quad \nu \in \{1, \ldots, d\}. \tag{4}$$

and the **t**otal **s**ensitivity **i**ndex (TSI) of an input parameter x_i, $i \in \{1, \ldots, d\}$ defined by [14, 16]:

$$S_i^{\text{tot}} = S_i + \sum_{l_1 \ne i} S_{il_1} + \sum_{l_1, l_2 \ne i, l_1 < l_2} S_{il_1 l_2} + \ldots + S_{il_1 \ldots l_{d-1}}, \tag{5}$$

where S_i is called *the main effect (first-order sensitivity index)* of x_i and $S_{il_1 \ldots l_{j-1}}$ is the jth order sensitivity index. The higher-order terms describe the interaction effects between the unknown input parameters $x_{i_1}, \ldots, x_{i_\nu}$, $\nu \in \{2, \ldots, d\}$ on the output variance.

The proper mathematical approach to the problem of conducting global sensitivity analysis is constituted by computing the total sensitivity indices (5) of corresponding order, following formulae (3)–(4), which in turn leads to evaluating multidimensional integrals.

2.2 Sobol and Halton Sequences

The distinctive feature of a quasirandom or low discrepancy sequence (for instance Halton and Sobol sequences) is the lower degree of randomness compared to a pseudorandom number sequence. Nevertheless, this is a desired property which makes the quasirandom sequence more advantageous for numerical integration purposes in higher dimensions, since the low discrepancy sequences incline to sample the space in a "more uniform" manner that the pseudorandom sequences [2].

Let $x_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(s)})$ for $i = 1, 2, \ldots$ and $n = \ldots a_3(n), a_2(n), a_1(n)$ be the representation of n in base b. The multidimensional quasi-random sequence is defined as $X_n = (\phi_{b_1}(n), \phi_{b_2}(n), \ldots, \phi_{b_s}(n))$, where the base b_i numbers are relatively prime.

Halton sequence [5, 6] is defined as:

$$s_n^{(k)} = \sum_{i=0}^{\infty} \sigma_{i+1}^{(k)} a_{i+1}^{(k)}(n) b_k^{-(i+1)},$$

where $(b_1, b_2, \ldots, b_s) \equiv (2, 3, 5, \ldots, p_s)$, and p_i denotes the ith prime, and $\sigma_i^{(k)}$, $i \geq 1$—set of permutations on $(0, 1, 2, \ldots, p_k - 1)$.

Sobol sequence [1, 7] is defined by:

$$x_k \in \overline{\sigma}_i^{(k)}, k = 0, 1, 2, \ldots$$

where $\overline{\sigma}_i^{(k)}$, $i \geq 1$ - set of permutations on every $2^k, k = 0, 1, 2, \ldots$ subsequent points of the Van der Corput sequence, defined by $n = \sum_{i=0}^{\infty} a_{i+1}(n) b^i$, $\phi_b(n) = \sum_{i=0}^{\infty} a_{i+1}(n) b^{-(i+1)}$ when $b = 2$.

In binary for the Sobol sequence we have that: $x_n^{(k)} = \bigoplus_{i \geq 0} a_{i+1}(n) v_i$, where v_i, $i = 1, \ldots, s$ is a set of direction numbers [7].

2.3 Scrambling

The original motivation of scrambling aims toward obtaining more uniformity for quasi-random sequences in high dimensions. The proved convergence rate for the Scrambling Algorithms improves significantly the rate for the unscrambled nets [13], which is $n^{-1}(\log n)^{d-1}$. That is why it is important to compare numerically our algorithms with their scrambled versions. The idea of scrambling is based on randomization of a single digit at each iteration. Let

$$x^{(i)} = (x_{i,1}, x_{i,2}, \ldots, x_{i,s}), \ i = 1, \ldots, n \tag{6}$$

be quasi-random numbers in $[0, 1)^s$, and let

$$z^{(i)} = (z_{i,1}, z_{i,2}, \dots, z_{i,s}) \qquad (7)$$

be the scrambled version of the point $x^{(i)}$. Suppose that each $x_{i,j}$ can be represented in base b as

$$x_{i,j} = (0.x_{i1,j} \, x_{i2,j} \dots x_{iK,j} \dots)_b \qquad (8)$$

with K being the number of digits to be scrambled. For scrambling Halton sequence we use a permutation of the radical inverse coefficients derived by applying a reverse-radix operation to all of the possible coefficient values [11]. For scrambling Sobol sequence we use random linear scramble combined with a random digital shift [12].

3 Case Study and Numerical Results

The present sensitivity analysis uses data originating from runs of a large large-scale mathematical model describing remote transport of air pollutants (UNI-DEM, [19, 21]). The geographical area into consideration (4800×4800 km^2) includes wholly Europe and the Mediterranean and partially Asia and Africa. It includes all main physical, chemical and photochemical processes occurring between the studied species as well as the emissions and the rapidly changing meteorological conditions. The reason to choose this model as a case study is that it is among the atmospheric chemistry models which consider the chemical processes in great detail and in accurate manner.

3.1 Sensitivity Studies with Respect to Emission Levels

In this section we will present the results for the sensitivity of UNI-DEM output (in particular, the ammonia mean monthly concentrations) with respect to the anthropogenic emissions input data variation, shown and discussed in this section. The anthropogenic emissions input consists of four different components
$\mathbf{E} = (\mathbf{E^A}, \mathbf{E^N}, \mathbf{E^S}, \mathbf{E^C})$:

$\mathbf{E^A}$ − ammonia (NH_3); $\qquad\qquad\qquad$ $\mathbf{E^S}$ − sulphur dioxide (SO_2);
$\mathbf{E^N}$ − nitrogen oxides ($NO + NO_2$); $\qquad$ $\mathbf{E^C}$ − anthropogenic hydrocarbons.

Results of the relative error estimation for the quantities f_0, the total variance $\mathbf{D}$, first-order (S_i) and total (S_i^{tot}) sensitivity indices are given in Tables 1, 2, 3 and Fig. 1, respectively. f_0 is presented by a 4-dimensional integral, while the rest of the above quantities are presented by 8-dimensional integrals, following the ideas of *correlated sampling* technique to compute sensitivity measures in a reliable way (see [8, 17]).

Table 1 Relative error for the evaluation of $f_0 \approx 0.048$

# of samples	CRU	SOBOL	HALTON	SOBOLSCR	HALTONSCR
n	Relative error	Relative error	Relative error	Relative error	Relative error
2^{10}	1.0203e−02	3.4504e−04	2.5491e−04	**1.8415e−05**	1.5307e−04
2^{12}	3.4246e−03	6.9277e−05	1.4346e−04	1.8209e−05	**5.4839e−06**
2^{14}	2.5052e−03	7.8022e−06	2.8407e−05	**4.8534e−06**	7.0189e−06
2^{16}	1.7268e−03	4.6585e−06	9.6538e−06	**3.6422e−07**	2.3992e−06
2^{18}	4.3151e−04	1.7393e−06	3.4711e−06	**1.5539e−07**	1.8033e−07
2^{20}	6.7226e−05	2.5234e−07	1.1020e−06	1.1501e−07	**4.7965e−08**
2^{22}	6.4635e−05	8.3823e−08	2.8645e−07	**3.3934e−09**	3.2964e−08
2^{24}	1.6251e−05	1.5669e−08	9.0096e−08	4.4868e−09	**2.8637e−09**

The five different stochastic approaches used for numerical integration are presented in separate columns of the tables.

For $n = 2^{24}$ for the model function f_0 the best algorithm is the Halton scrambled sequence, followed by the Halton sequence – see the results in Table 1 for the maximum number of samples. For number of samples $n = 2^{24}$ for the total variance D the best algorithm is the Sobol sequence, followed by the Halton scrambled sequence – see the results in Table 2 for the maximum number of samples.

For number of samples $n = 2^{24}$ in Table 3 Sobol gives the best relative errors for S_1, S_3, S_1^{tot}, S_2^{tot}, S_3^{tot}, while Halton scrambled sequence gives the best results for S_2, S_4 and S_4^{tot}. However, for the most important and smallest in value sensitivity index S_4 and S_4^{tot} Halton scrambled sequence gives the most reliable results. The numerical results suggest that the best algorithm is the Halton scrambled sequence, followed by the Sobol sequence (for small in value sensitivity indices Halton scrambled sequence gives better results), and after that follows Sobol scrambled sequence and Halton sequence. In most of the cases the scrambled and non scrambled sequences produced results of the same order. In all experiments as expected the Crude Monte Carlo approach gives the worst results.

## 3.2 Sensitivity Studies with Respect to Chemical Reactions Rates

In this section we will study the sensitivity of the ozone concentration values in the air over Genova with respect to the rate variation of some chemical reactions of the condensed CBM-IV scheme [19], namely: # 1, 3, 7, 22 (time-dependent) and # 27, 28 (time independent). The simplified chemical equations of those reactions are:

Table 2 Relative error for the evaluation of the total variance $\mathbf{D} \approx 0.0002$

# of samples	CRU	SOBOL	HALTON	SOBOLSCR	HALTONSCR
n	Relative error	Relative error	Relative error	Relative error	Relative error
2^{10}	1.1512e−01	1.4797e−03	1.5217e−02	**2.6245e−04**	7.1057e−03
2^{12}	2.8713e−02	1.5736e−03	8.1209e−03	**1.2114e−03**	1.5066e−03
2^{14}	4.3025e−02	4.7282e−04	1.8880e−03	**5.7270e−05**	5.9235e−04
2^{16}	1.7631e−02	1.1726e−04	6.8346e−04	**3.3306e−06**	1.2015e−04
2^{18}	1.1619e−02	1.3197e−05	1.5618e−04	**7.2002e−06**	4.1675e−05
2^{20}	5.7971e−03	**8.4017e−06**	4.7374e−05	1.7242e−05	8.9747e−06
2^{22}	7.3641e−04	1.8001e−06	1.5029e−05	**1.1211e−06**	4.3661e−06
2^{24}	1.9965e−03	**3.2922e−08**	3.1611e−06	8.2382e−07	1.5148e−07

Table 3 Relative error for estimation of sensitivity indices of input parameters using various Monte Carlo and quasi-Monte Carlo approaches ($n \approx 2^{24}$)
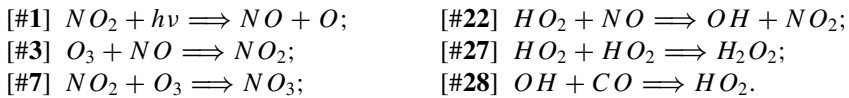
EQ	RV	CRU	SOBOL	HALTON	SOBOLSCR	HALTONSCR
S_1	9e−01	2.0736e−04	**3.6546e−07**	1.1917e−06	6.0288e−07	6.2995e−07
S_2	2e−04	3.3428e−03	1.7825e−05	2.4151e−04	4.4273e−05	**2.3471e−06**
S_3	1e−01	1.3562e−03	**1.8064e−06**	9.1091e−06	3.4129e−06	4.0456e−06
S_4	4e−05	1.6781e−01	1.4025e−04	4.0279e−04	1.7591e−04	**8.4522e−05**
S_1^{tot}	9e−01	1.6667e−04	**2.1422e−07**	1.1955e−06	4.2602e−07	4.9628e−07
S_2^{tot}	2e−04	6.3335e−02	**2.9270e−05**	2.5980e−04	4.3709e−05	5.6982e−05
S_3^{tot}	1e−01	1.7633e−03	**3.0504e−06**	8.8825e−06	4.7584e−06	5.1651e−06
S_4^{tot}	5e−05	1.5280e−02	1.9103e−04	3.9080e−04	9.5501e−05	**1.7606e−05**
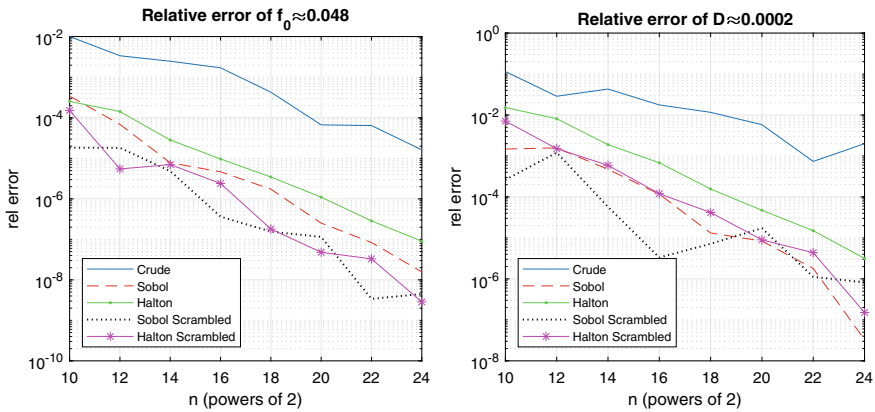
Fig. 1 Relative errors for the calculation of $f_0 \approx 0.048$ (left) and $\mathbf{D} \approx 0.0002$ (right)

Table 4 Relative error for the evaluation of $f_0 \approx 0.27$

# of samples	CRU	SOBOL	HALTON	SOBOLSCR	HALTONSCR
n	Relative error	Relative error	Relative error	Relative error	Relative error
2^{10}	6.9655e−04	2.4527e−04	3.6545e−04	3.9428e−04	**1.7657e−04**
2^{12}	8.0553e−04	1.0223e−04	8.6328e−05	4.0466e−05	**2.3302e−05**
2^{14}	2.8035e−03	2.2274e−05	3.4302e−05	**5.2466e−06**	2.4262e−05
2^{16}	5.9493e−04	2.2604e−06	3.6220e−06	**1.2498e−06**	5.3505e−06
2^{18}	7.6591e−04	1.8400e−06	1.7421e−06	**6.0489e−07**	8.4729e−07
2^{20}	3.4494e−04	3.5561e−07	6.0821e−07	**1.0645e−07**	3.2548e−07
2^{22}	4.6980e−05	9.3845e−08	1.6051e−07	**1.6511e−08**	6.0443e−08
2^{24}	8.6192e−06	**1.8639e−09**	4.9903e−08	1.1468e−08	1.8102e−08

[**#1**] $NO_2 + h\nu \Longrightarrow NO + O$; 　　　[**#22**] $HO_2 + NO \Longrightarrow OH + NO_2$;
[**#3**] $O_3 + NO \Longrightarrow NO_2$; 　　　　　　[**#27**] $HO_2 + HO_2 \Longrightarrow H_2O_2$;
[**#7**] $NO_2 + O_3 \Longrightarrow NO_3$; 　　　　　[**#28**] $OH + CO \Longrightarrow HO_2$.

The domain under consideration is the 6-dimensional hypercube $[0.6, 1.4]^6$.

The relative error estimation for the quantities f_0, the total variance **D** and some sensitivity indices are given in Tables 4, 5, 6 and Fig. 2, respectively.

The value of f_0 is calculated from a 6-dimensional integral, while the rest quantities are calculated from 12-dimensional integrals, following the paradigm of *correlated sampling* [8]. For $n = 2^{24}$ for the model function f_0 the best algorithm is the Sobol sequence, followed by the Sobol scrambled sequence—see the results in Table 4 for the maximum number of samples. For number of samples $n = 2^{24}$ for the total variance D the best algorithm is the Sobol scrambled sequence, followed again by the Sobol sequence—see the results in Table 5 for the maximum number of samples.

For number of samples $n = 2^{24}$ in Table 6 Sobol gives the best relative errors for S_1, S_2, S_6, S_1^{tot}, S_6^{tot}, S_{12}, S_{14}, S_{24} and S_{45}. Halton scrambled sequence gives the best results for S_3, S_4, S_5, S_3^{tot} and interestingly Halton sequence gives the best relative errors for S_2^{tot}, S_4^{tot} and S_5^{tot}. It can be seen that for the small in value sensitivity index S_{45} Sobol sequence gives also reliable results. However, for the most important and smallest in value sensitivity index S_5 the Halton scrambled sequence produce the best results. The algorithm performance demonstrate that the best approach is to apply the Sobol sequence, followed by the Halton scrambled sequence (better results are associated with the latter for small in value sensitivity indices). The Sobol scrambled sequence follows. It appears again that the Crude Monte Carlo method produces the worst results.

Table 5 Relative error for the evaluation of the total variance $\mathbf{D} \approx 0.0025$

# of samples	CRU	SOBOL	HALTON	SOBOLSCR	HALTONSCR
n	Relative error	Relative error	Relative error	Relative error	Relative error
2^{10}	7.2185e−02	**2.1299e−03**	4.9601e−02	7.4282e−03	3.3594e−02
2^{12}	9.5413e−02	1.8037e−03	2.7994e−03	**5.3106e−04**	4.8055e−03
2^{14}	6.3987e−02	**1.1021e−04**	2.3975e−03	5.3462e−04	1.1157e−04
2^{16}	2.9741e−02	1.2418e−04	3.4838e−04	**1.0328e−04**	7.2396e−04
2^{18}	7.4173e−03	2.2215e−05	4.0181e−04	**1.9470e−06**	5.3410e−05
2^{20}	8.9182e−03	**3.3461e−06**	4.6222e−05	6.2125e−06	3.2061e−05
2^{22}	2.2089e−03	5.2062e−06	2.8782e−06	**2.0482e−06**	6.0767e−06
2^{24}	1.2915e−03	1.5526e−06	2.5678e−06	**7.9422e−07**	2.7586e−06

Table 6 Relative error for estimation of sensitivity indices of input parameters using various QMC approaches ($n \approx 2^{24}$)

EQ	RV	CRU	SOBOL	HALTON	SOBOLSCR	HALTONSCR
S_1	4e−01	2.4330e−03	**4.5630e−07**	1.0773e−05	1.6426e−06	3.4398e−06
S_2	3e−01	1.8600e−03	**9.9892e−07**	1.7563e−05	1.0094e−06	1.1826e−06
S_3	5e−02	3.9300e−03	5.2706e−06	1.3779e−05	4.9645e−06	**1.7893e−06**
S_4	3e−01	1.7259e−03	2.5748e−06	8.7453e−06	7.4544e−07	**5.9071e−07**
S_5	4e−07	6.2991e+01	7.5741e−02	2.0971e−01	9.4203e−02	**6.4810e−02**
S_6	2e−02	1.3933e−02	**3.5457e−06**	3.6907e−05	4.1779e−06	9.6398e−06
S_1^{tot}	4e−01	1.9688e−03	**6.6739e−07**	1.0799e−05	1.7121e−06	2.9868e−06
S_2^{tot}	3e−01	2.6026e−03	1.1905e−06	1.4399e−05	**8.4968e−07**	2.7793e−06
S_3^{tot}	5e−02	2.0711e−03	3.0697e−06	1.4532e−05	2.8127e−06	**1.5540e−07**
S_4^{tot}	3e−01	1.7469e−03	3.5570e−06	8.3469e−06	**2.3770e−06**	5.2178e−06
S_5^{tot}	2e−04	4.2987e−02	1.0014e−04	8.5496e−05	**7.3245e−06**	2.0781e−04
S_6^{tot}	2e−02	5.3593e−03	**4.0620e−06**	3.8012e−05	7.3849e−06	9.0746e−06
S_{12}	6e−03	2.4931e−02	**1.2872e−05**	5.9774e−05	4.9952e−05	7.3132e−05
S_{14}	5e−03	1.2130e−02	**1.4795e−05**	3.1874e−05	2.7816e−05	9.9435e−05
S_{24}	3e−03	2.6639e−02	**1.9557e−05**	7.0366e−05	7.1769e−05	2.3935e−04
S_{45}	1e−05	1.5179e−01	**1.0902e−04**	3.2605e−03	2.2443e−04	2.9941e−03

4 Conclusion

A number of stochastic algorithms for multidimensional integral evaluation are examined to analyze their computational efficiency in terms of relative error. The UNI-DEM model is considered and the output sensitivity with respect to variation in the input emissions of the anthropogenic pollutants and in rates of several chemical reactions is studied. The influence of emission levels over very important air pollutants (ammonia, ozone, and ammonium sulphate and ammonium nitrate) is considered. The numerical experiments show that Halton and Sobol sequences and their scrambled versions constitute part of the best available stochastic methods to evaluate
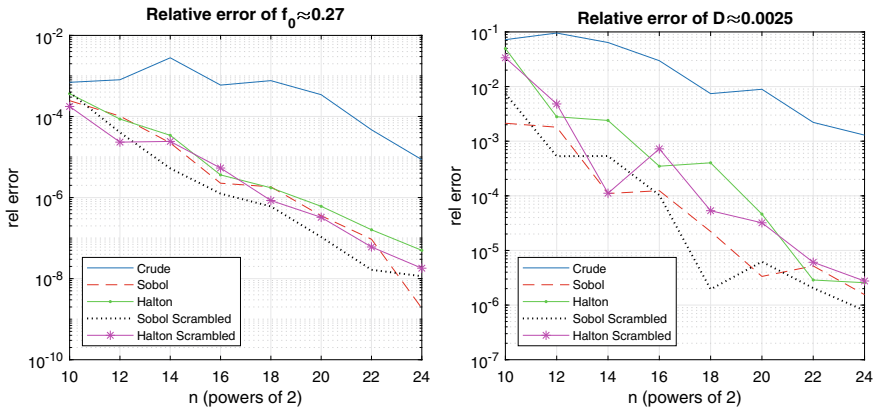
Fig. 2 Relative errors for the calculation of $f_0 \approx 0.27$ (left) and $\mathbf{D} \approx 0.0025$ (right)

sensitivity indices. In particular, they cope with the smallest in value sensitivity indices, which are vital for the reliability of the results as well as being the most difficult problem. Experiments show that scrambling is useful for obtaining more uniformity for quasi-random sequences in high dimensions which is definitely the case in the Halton scrambled sequence. This paper shows that essential improvement is obtained with Halton scrambled sequence over the original Halton sequence and this is the best available approach for most of the small in value sensitivity indices which are very important for the reliability of the model results.

References

1. Antonov I., Saleev V.: An economic method of computing LP_τ-sequences, USSR Comput. Math. Phy. 19 (1979) 252–256.
2. Dimov I.: Monte Carlo Methods for Applied Scientists, New Jersey, London, Singapore, World Scientific, (2008), 291p.
3. Dimov I.T. , Georgieva R., Ostromsky Tz, Zlatev Z.: Advanced Algorithms for Multidimensional Sensitivity Studies of Large-scale Air Pollution Models based on Sobol Sequences. Special issue of Computers and Mathematics with Applications 65 (3), "Efficient Numerical Methods for Scientific Applications". Elsevier, (2013), 338 - 351.

4. Ferretti, F., Saltelli A., Tarantola, S.: Trends in Sensitivity Analysis Practice in the Last Decade Journal, Science of the Total Environment, Special issue on Human and Biota Exposure, Vol. 568, (2016), 666–670, Elsevier.

5. Halton J.: On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals, Numerische Mathematik, Vol. 2, (1960), 84–90.

6. Halton J., Smith GB: Algorithm 247: Radical-Inverse Quasi-Random Point Sequence, Communications of the ACM, Vol. 7, (1964), 701–702.

7. Joe S., Kuo F.:, Remark on Algorithm 659: Implementing Sobol's Quasirandom Sequence Generator, ACM Transactions on Mathematical Software, Vol. 29, No 1, (March 2003), pp. 49–57.

8. Homma T., Saltelli A.: Importance Measures in Global Sensitivity Analysis of Nonlinear Models, Reliability Engineering and System Safety 52, (1996), 1–17.

9. Karaivanova, A., Atanassov, E., Gurov, T., Stevanovic, R., & Skala, K.: Variance Reduction MCMs with Application in Environmental Studies: Sensitivity Analysis. In AIP Conference Proceedings (Vol. 1067, No. 1, pp. 549–558). AIP, (2008, October).

10. Karaivanova, A., Dimov, I., & Ivanovska, S.: A quasi-Monte Carlo method for integration with improved convergence. In International Conference on Large-Scale Scientific Computing (pp. 158–165). Springer, Berlin, Heidelberg, (2001, June).

11. Kocis, L., and Whiten W. J.: "Computational Investigations of Low-Discrepancy Sequences". ACM Transactions on Mathematical Software. Vol. 23, No. 2, (1997), pp. 266-294.

12. Matousek, J: "On the L2-Discrepancy for Anchored Boxes". Journal of Complexity. Vol. 14, No. 4, (1998), pp. 527–556.

13. Ökten G., Göncüb A.: Generating low-discrepancy sequences from the normal distribution: Box-Muller or inverse transform?, Mathematical and Computer Modelling 53 (2011) 1268–1281.

14. Saltelli, A., Tarantola, S., Campolongo, F., Ratto,M.: Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. Halsted Press, New York (2004).

15. Sobol I.: Numerical methods Monte Carlo. Nauka, Moscow, (1973).

16. Sobol I.M.: Sensitivity estimates for nonlinear mathematical models, Mathematical Modeling and Computational Experiment 1 (4) (1993) 407–414.

17. Sobol I.M., Tarantola S., Gatelli D., Kucherenko S., Mauntz W.: Estimating the approximation error when fixing unessential factors in global sensitivity analysis, Reliability Engineering & System Safety, (2007), 92, 957–960

18. Zaharieva, S. L., Georgiev, I. R., Mutkov, V. A., & Neikov, Y. B.: Arima Approach For Forecasting Temperature In A Residential Premises Part 2. In 2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH) (2021), (pp. 1–5), IEEE.

19. Zlatev Z., Computer treatment of large air pollution models, KLUWER Academic Publishers, Dorsrecht-Boston-London, (1995).

20. Zlatev, Z., Dimov, I.T., Georgiev, K.: Three-dimensional version of the Danish Eulerian model. Z. Angew. Math. Mech. 76(S4), 473–476 (1996).

21. Zlatev Z., Dimov I.T.: Computational and Numerical Challenges in Environmental Modelling, Elsevier, Amsterdam, (2006).

Advanced Biased Stochastic Approaches Based on Modified Sobol Sequences for the Fredholm Integral Equation

Venelin Todorov, Ivan Dimov, and Rayna Georgieva

Abstract In this paper we propose and analyse advanced biased stochastic methods for solving a class of integral equations—the second kind Fredholm integral equations. We study and compare innovative possible approaches to compute linear functionals of the integral under consideration: biased Monte Carlo method based on evaluation of truncated Liouville-Neumann series and transformation of this problem into the problem of computing a finite number of integrals. Advanced Monte Carlo algorithms for numerical integration based on modified Sobol sequence have been applied for computing linear functionals. Error balancing of both stochastic and systematic errors has been discussed and applied during the numerical implementation of the biased algorithms. We compare the results obtained by some of the best biased stochastic approaches with the results obtained by the standard Monte Carlo method for integral equations.

V. Todorov
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 8 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria

V. Todorov (✉) · I. Dimov · R. Georgieva
Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, 25A Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
e-mail: vtodorov@math.bas.bg

I. Dimov
e-mail: ivdimov@bas.bg

R. Georgieva
e-mail: rayna@paralel.bas.bg

1 Introduction

The existing MC methods for integral equations (MCM-IE) are based on probabilistic representations of the Liouville-Neumann (LN) series for the second kind Fredholm integral equation [2, 4, 8]. The possible unbiased approaches deal with infinite series, while the biased MC approaches use probabilistic representations of truncated LN series. A well-known and widely used biased method is the Markov chain MC (see, for example [12]). Usually, the Markov chain stops after a fixed number of steps. Integral equations are very important for areas such as mechanics, geophysics, electricity and magnetism, kinetic theory of gases, quantum mechanics, mathematical economics, and queuing theory [10, 11, 13].

A possible approach to deal with the problem of approximation of linear functionals of the solution of an integral equation is to transform it into approximate evaluation of finite number of integrals (FNI) (linear functionals of iterative functions) [3, 4]. Here we extend the study of the properties to four different MC algorithms for multidimensional numerical integration for solving integral equations. These algorithms are Crude MC algorithm [17], based on SIMD-oriented fast Mersenne Twister pseudorandom number generator [15, 20], quasi-Monte Carlo (QMC) algorithm based on $\Lambda \Pi_\tau$ Sobol quasirandom sequences and MC algorithms (MCA-MSS-1 and MCA-MSS-2) based on modified Sobol quasirandom sequences [5–7].

Consider the second kind Fredholm integral equation:

$$u(x) = \int_\Omega k(x, x')u(x')dx' + f(x) \tag{1}$$

or

$$u = \mathcal{K}u + f \quad (\mathcal{K} \text{ is an integral operator}), \quad \text{where}$$

$k(x, x') \in L_2(\Omega \times \Omega)$, $f(x) \in L_2(\Omega)$ are given functions and $u(x) \in L_2(\Omega)$ is an unknown function, x, $x' \in \Omega \subset R^n$ (Ω is a bounded domain).

We are interested in Monte Carlo methods for evaluation of linear functionals of the solution of the following type:

$$J(u) = \int \varphi(x)u(x)dx = (\varphi, u), \tag{2}$$

where $\varphi(x) \in L_2(\Omega)$ is a given function.

We can apply successive approximation method for solving integral equations:

$$u^{(i)} = \sum_{j=0}^{i} \mathcal{K}^{(j)} f = f + \mathcal{K}f + \ldots + \mathcal{K}^{(i-1)} f + \mathcal{K}^{(i)} f, \quad i = 1, 2, \ldots \tag{3}$$

where $u^{(0)}(x) \equiv f(x)$.

An approximation of the unknown value (φ, u) can be obtained using a truncated LN series (3) for a sufficiently large i:

$$(\varphi, u^{(i)}) = (\varphi, f) + (\varphi, \mathcal{K}f) + \ldots + (\varphi, \mathcal{K}^{(i-1)}f) + (\varphi, \mathcal{K}^{(i)}f).$$

So, we transform the problem for solving integral equations into a problem for approximate evaluation of a finite number of multidimensional integrals. We will use the following denotation $(\varphi, \mathcal{K}^{(j)}f) = I(j)$, where $I(j)$ is a value, obtained after integration over $\Omega^{j+1} = \Omega \times \ldots \times \Omega$, $j = 0, \ldots, i$. It is obvious that the calculation of the estimate $(\varphi, u^{(i)})$ can be replaced by an evaluation of a sum of linear functionals of iterative functions of the type $(\varphi, \mathcal{K}^{(j)}f)$, $j = 0, \ldots, i$, which can be presented as:

$$(\varphi, \mathcal{K}^{(j)}f) = \int_{\Omega} \varphi(t_0)\mathcal{K}^{(j)}f(t_0)dt_0 =$$

$$= \int_{G} \varphi(t_0)k(t_0, t_1)\ldots k(t_{j-1}, t_j)f(t_j)dt_0 \ldots dt_j, \tag{4}$$

where $t = (t_0, \ldots, t_j) \in G \equiv \Omega^{j+1} \subset R^{n(j+1)}$. If we denote by $F_j(t)$ the integrand function

$$F(t) = \varphi(t_0)k(t_0, t_1)\ldots k(t_{j-1}, t_j)f(t_j), \quad t \in \Omega^{j+1},$$

then we obtain the following expression for (4):

$$I(j) = (\varphi, \mathcal{K}^{(j)}f) = \int_{G} F_j(t)dt, \quad t \in G \subset R^{n(j+1)}. \tag{5}$$

Thus, we consider the problem for an approximate calculation of multiple integrals of the type (5). The definition of domain G given above is important for the further presented method of solving the integral equation as a set of FNI. Actually, both representations (4) and (5) allow to define a biased algorithm. In this case the approximation $(\varphi, \mathcal{K}^{(j)}f)$ to the true inner product (φ, u) is presented as a set of l integrals $I(j)$ (see, (5)) with integrands $F_j(t)$, where t is $N(j + 1)$ dimensional point from the newly defined domain G.

2 Biased Stochastic Approaches

2.1 Crude Monte Carlo for Integrals

The simplest stochastic approach known as Crude Monte Carlo method (CMCM) for evaluating integrals reduces the problem to the approximate calculation of a mathematical expectation which coincides with the unknown functional defined by

[4]. The Monte Carlo quadrature formula is based on the probabilistic interpretation of an integral. If $x_l, l = 1, \ldots, N$ is a uniformly distributed sequence in G, then the Crude Monte Carlo approximation to the integral $I(j)$ is

$$I(j) \approx I_N(j) = \frac{V(G)}{N} \sum_{l=1}^{N} F_j(x_l)$$

with integration error $\varepsilon_N(j) = |I(j) - I_N(j)| \approx \sqrt{\dfrac{Var(F)}{N}}$, where $V(G)$ is the volume of G and $Var(F_j)$ is the variance of the corresponding random variable whose mathematical expectation coincides with the unknown functional. This random variable depends on the integrand $F(t)$.

2.2 Monte Carlo Method for Integral Equations

It is known (see [4, 17]) that the approximate calculation of linear functionals of the solution of an integral equation (φ, u) brings to the calculation of a finite sum of linear functionals of iterative functions $(\varphi, \mathcal{K}^j f)$, $j = 0, \ldots, i$. First, we construct a random trajectory (Markov chain) T_i of length i starting from state x_0 in the domain Ω:

$$T_i : x_0 \longrightarrow x_1 \longrightarrow \ldots \longrightarrow x_i$$

according to the initial $\pi(x)$ and transition $p(x, x')$ probabilities. The functions $\pi(x)$ and $p(x, x')$ satisfy the requirements for non-negativeness, to be acceptable to function $\varphi(x)$ and the kernel $k(x, x')$ respectively and $\int_\Omega \pi(x)dx = 1$, $\int_\Omega p(x, x')dx' = 1$ for any $x \in \Omega \subset R^n$. The Markov chain transition probability $p(x, x')$ is chosen to be proportional to $|k(x, x')|$ following [4]. It means that $p(x, x') = c|k(x, x')|$, and the constant c is computed such that

$$c = \left(\int_\Omega |k(x, x')|dx' \right)^{-1} = 1 \ \text{ for any } \ x \in \Omega.$$

From the above supposition on the kernel $k(x, x')$ and the well-known facts that

$$E\theta_i[\varphi] = (\varphi, u^{(i)}), \quad \text{where} \quad \theta_i[\varphi] = \frac{\varphi(x_0)}{\pi(x_0)} \sum_{j=0}^{i} W_j f(x_j),$$

$$\text{and} \quad W_0 = 1, \quad W_j = W_{j-1} \frac{k(x_{j-1}, x_j)}{p(x_{j-1}, x_j)}, \quad j = 1, \ldots, i$$

it follows that the corresponding Monte Carlo estimation of $(\varphi, u^{(i)})$ can be presented in the following form:

$$(\varphi, u^{(i)}) \approx \frac{1}{N} \sum_{l=1}^{N} \theta_i[\varphi]_l.$$

Therefore, the random variable $\theta_i[\varphi]$ can be considered as a biased estimate of the desired value (φ, u) for i sufficiently large with a statistical error of order $\mathcal{O}(N^{-1/2})$, where N is the number of chains and $\theta_i[\varphi]_l$ is the value of $\theta_i[\varphi]$ taken over the l-th chain. The same trajectories of the type T_i can be used for a biased approximate evaluation of $(\varphi, u^{(i)})$ for various functions $\varphi(\mathrm{x})$. Furthermore, they can be used for various integral equations with the same kernel $k(\mathrm{x}, \mathrm{x}')$, but with different right-hand sides $f(\mathrm{x})$.

2.3 Monte Carlo Algorithms Based on Modified Sobol $\Lambda\Pi_\tau$ Sequences

$\Lambda\Pi_\tau$ sequences are *uniformly distributed sequences* (u.d.s.). The term *u.d.s.* was introduced by Hermann Weyl [19]. For practical purposes an u.d.s. should satisfy the following requirements [17, 18]: (i) the best asymptote as $N \to \infty$, (ii) well distributed points for small N, and (iii) a computationally inexpensive algorithm. Suitable distributions such as $\Lambda\Pi_\tau$ sequences are also called (t, m, s)-nets and (t, s)-sequences in base $b \geq 2$ [14]. Sobol [17] defines his Π_τ-meshes and $\Lambda\Pi_\tau$ sequences, which are (t, m, s)-nets and (t, s)-sequences in base 2, respectively. Subroutines to compute these points can be found in [1], and with more details in [16]. Here we consider two randomized quasi-Monte Carlo algorithms based on Sobol $\Lambda\Pi_\tau$ sequences, namely MCA-MSS-1 and MCA-MSS-2.

2.4 Summary of the MCM-MSS-1

One of the algorithms (based on a procedure of *shaking*) was proposed recently in [6]. The idea is that we take a Sobol $\Lambda\Pi_\tau$ d-dimensional point (vector) x. Then x is considered as a centrum of a sphere with a radius $\rho << 1$. A random point $\xi \in U^d$ uniformly distributed on the sphere is taken. Consider a random variable θ defined as a value of the integrand at that random point, i.e., $\theta_j = F_j(\xi)$. Consider random points $\xi^{(l)}(\rho) \in U^d, l = 1, \ldots, N$. Assume $\xi^{(l)}(\rho) = \mathrm{x}^{(l)} + \rho\omega^{(l)}$, where $\omega^{(l)}$ is a unique uniformly distributed vector in U^d. The radius ρ is relatively small $\rho \ll \frac{1}{2^{d_j}}$, such that $\xi^{(l)}(\rho)$ is still in the same elementary lth interval $E_l^d = \prod_{\nu=1}^{d} \left[\frac{a_j^{(l)}}{2^{d_\nu}}, \frac{a_\nu^{(l)} + 1}{2^{d_\nu}} \right]$,

where the pattern $\Lambda\Pi_\tau$ point $\mathrm{x}^{(l)}$ is. We use a subscript i in E_i^d to indicate that the l-th $\Lambda\Pi_\tau$ point $\mathrm{x}^{(i)}$ is in it. So, we assume that if $\mathrm{x}^{(l)} \in E_l^d$, then $\xi^{(l)}(\rho) \in E_l^d$ too. It was proven in [6] that the mathematical expectation of the random variable $\theta_j = F_j(\xi)$ is equal to the value of the integral (5), that is $\mathrm{E}\theta_j = \int_{U^d} F_j(\mathrm{t})d\mathrm{t}$. This result allows to define a randomized algorithm called Monte Carlo Algorithm based on Modified Sobol Sequences (MCA-MSS-1).

In [7] the probability error of the algorithm MCA-MSS-1 is analysed. Denote by $\| F_j \|$ the following expression:

$$\| F_j \| = \sup \left\{ \left| \frac{\partial^r F_j}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} \right|, |\alpha_1 + \ldots + \alpha_d| = k \right\}.$$

It is proved that for integrands with continuous and bounded first derivatives, i.e. $F_j \in \mathbf{W}^1(L; U^d)$, where $L = \|F_j\|$, it holds $err(F_j, d) \leq c_d' \|F_j\| N^{-\frac{1}{2}-\frac{1}{d}}$ and $r(F_j, d) \leq c_d'' \|F_j\| N^{-\frac{1}{2}-\frac{1}{d}}$, where the constants c_d' and c_d'' do not depend on N.

2.5 Summary of the MCM-MSS-2

The second algorithm called MCA-MSS-2 is a modification of MCA-MSS-1 algorithm. It is proposed and analysed in [5]. It is assumed that $n = m^d, m \geq 1$. The unit cube U^d is divided into m^d disjoint sub-domains, such that they coincide with the elementary d-dimensional subintervals $U^d = \bigcup_{l=1}^{m^d} K_l$, where $K_l = \prod_{\nu=1}^d [a_\nu^{(l)}, b_\nu^{(l)})$, with $b_\nu^{(l)} - a_\nu^{(l)} = \frac{1}{m}$ for all $l = 1, \ldots, d$. In such a way in each d-dimensional sub-domain K_l there is exactly one $\Lambda\Pi_\tau$ point $\mathrm{x}^{(l)}$. Assuming that after shaking, the random point stays inside K_l, i.e., $\xi^{(l)}(\rho) = \mathrm{x}^{(l)} + \rho\omega^{(l)} \in K_l$ one may try to exploit the smoothness of the integrand in case if the integrand F_j belongs to $\mathbf{W}^2(L; U^d)$. In each sub-domain K_j the central point is denoted by $\mathrm{s}^{(j)}$, where $\mathrm{s}^{(j)} = (s_1^{(j)}, s_2^{(j)}, \ldots, s_d^{(j)})$.

Suppose two random points $\xi^{(l)}$ and $\xi^{(l)'}$ are chosen, such that $\xi^{(l)}$ is selected during our procedure used in MCA-MSS-1. The second point $\xi^{(l)'}$ is chosen to be symmetric to $\xi^{(l)}$ according to the central point $s^{(l)}$ in each cube K_l. In such a way the number of random points is $2m^d$. One may calculate all function values $f(\xi^{(l)})$ and $f(\xi^{(l)'})$, for $l = 1, \ldots, m^d$ and approximate the value of the integral in the following way:

$$I(f) \approx \frac{1}{2m^d} \sum_{l=1}^{2N} \left[f(\xi^{(l)}) + f(\xi^{(l)'}) \right]. \tag{6}$$

This estimate corresponds to MCA-MSS-2.

In [5] it is proved that the Monte Carlo algorithm MCA-MSS-2 has an optimal rate of convergence for functions with continuous and bounded second derivative [4]. This means that if $F_j \in \mathbf{W}^2(L; U^d)$, where $L = \|F_j\|$, it holds $err(F_j, d) \leq$

$\widetilde{c}_d \left\| F_j \right\| N^{-\frac{1}{2}-\frac{2}{d}}$ and $r(F_j, d) \leq \widetilde{c}_d'' \left\| F_j \right\| N^{-\frac{1}{2}-\frac{2}{d}}$, where the constants c_d' and c_d'' do not depend on N.

Note that both MCA-MSS-1 and MCA-MSS-2 have one control parameter, that is the radius ρ of the *sphere of shaking*. At the same time, to be able to efficiently use this control parameter one should increase the computational complexity. It happens because after *shaking* the random point may leave the multidimensional sub-domain, and one needs to check if the random point is still in the same sub-domain. It is clear that the procedure of checking if a random point is inside the given domain is a computationally expensive procedure when one has to deal with a large number of points.

2.6 Error Analysis of Biased Stochastic Approaches

Let us deal with the problem for approximate calculation of linear functional of the solution of integral equation (φ, u). Let us denote the i-th iterative approximation of u ($i \geq 0$) by $u^{(i)}$ and the Monte Carlo approximation by $\tilde{u}$. It is known that two errors systematic one (a truncation error) R_{sys} and stochastic (a probabilistic) one, R_N appear. Actually, we approximate the truncated LN series. If ε is a given sufficiently small positive parameter then

$$|(\varphi, u) - (\varphi, \tilde{u})| \leq |(\varphi, u) - (\varphi, u^{(i)})| + |(\varphi, u^{(i)}) - (\varphi, \tilde{u})| = \varepsilon_i + \varepsilon_N < \varepsilon,$$

where ε_i is the truncation error, ε_N is the probability error. We can obtain a lower bound for the number i of iterations using [2]:

$$i > \frac{\ln \dfrac{\varepsilon}{2|(\varphi, u^{(0)}) - (\varphi, u)|}}{\ln \|\mathcal{K}\|_{L_2}},$$

where $u^{(0)}$ is the initial approximation of u. Using the Cauchy-Bunyakowski inequality it is easy to show that

$$|(\varphi, u^{(0)}) - (\varphi, u)| \leq \|\varphi\|_{L_2(\Omega)} \|u^{(0)} - u\|_{L_2(\Omega)}.$$

The second multiplier can be estimated:

$$\|u^{(0)} - u\|_{L_2(\Omega)} \leq (\|I\|_{L_2} + \|\mathcal{K}\|_{L_2} + \ldots + \|\mathcal{K}\|_{L_2}^{i} + \ldots)\|r^{(0)}\|_{L_2(\Omega)},$$

$$\text{where} \quad r^{(j)} = f - u^{(j)} - \mathcal{K}u^{(j)}, \quad j = 0, 1, \ldots.$$

Taking the limit for $i \to \infty$ one can obtain

$$\|u^{(0)} - u\|_{L_2} \leq \frac{\|r^{(0)}\|_{L_2(\Omega)}}{1 - \|\mathcal{K}\|_{L_2}}.$$

It holds if the condition $\|\mathcal{K}\|_{L_2} < 1$ is fulfilled.

3 Numerical Results

The numerical algorithms are tested on the following example [8, 9]:

$$u(x) = \int_{\Omega} k(x, x')u(x')dx' + f(x), \qquad \Omega \equiv [0; 1] \tag{7}$$

where

$$k(x, x') = x^2 e^{x'(x-1)}, \; f(x) = x + (1-x)e^x, \; \varphi(x) = \delta(x - x_0). \tag{8}$$

The solution of this test problem is $u(x) = e^x$. We are interested in an approximate calculation of (φ, u), where $\varphi(x) = \delta(x - x_0)$, $x_0 = 0.5$.

We have performed the following biased stochastic algorithms: MC algorithm and QMC based on Sobol sequences for integral equations and several algorithms for integrals into which the problem of solving integral equations is transformed. For the above set of algorithms we use Crude MCA, quasi-Monte Carlo algorithm based on $\Lambda \Pi_\tau$ Sobol quasirandom sequences, MCA-MSS-1 and MCA-MSS-2. The algorithms have been studied after 10 runs to average the final approximation and for various sample sizes chosen according to proper sample size for Sobol quasirandom sequences. The number of iterations i/d is fixed, but it is chosen according to the L_2-norm of the kernel $k(x, x')$. For an approximate computation of any integral $I(j), \; j = 0, \ldots, i$ different number of samples are chosen to satisfy the error balancing requirements (Table 1).

Table 1 Relative errors and computational time for computing $u(x_0)$ at point $x_0 = 0.5$ using MCM for integral equations

i	ε	N	Crude MC		Sobol seq.	
			Rel. err.	Time (s)	Rel. err.	Time (s)
1	0.4	128	0.05960	<0.0001	0.05916	<0.0001
2	0.14	1,024	0.00566	<0.0001	0.00670	<0.0001
2	0.08	4,096	0.00294	0.01	0.00712	<0.0001
3	0.02	32,768	0.00092	0.17	0.00204	0.01
4	0.018	65,536	0.00076	0.49	0.00096	0.03

Table 2 Relative errors and computational time for computing $u(x_0)$ as a finite number (i) of integrals from the LN series at $x_0 = 0.5$ using Crude MCM

i	N	Integral equation				FNI	
		Sobol seq.		Crude MC		Sobol seq.	Crude MC
		Rel. err.	Time (s)	Rel. err.	Time (s)	Rel. R_N	Rel. R_N
1	128	0.05158	<0.0001	0.05197	<0.0001	0.00024	0.00016
2	1,024	0.01439	<0.0001	0.01405	<0.0001	2e-05	0.00033
2	4,096	0.01438	<0.0001	0.01422	<0.0001	5e-06	0.00015
3	32,768	0.00399	0.01	0.00404	0.11	1e-06	4e-05
4	65,536	0.00111	0.03	0.00108	0.63	1e-06	3e-05

Table 3 Relative errors and computational time for computing $u(x_0)$ as a finite number of integrals from the LN series at $x_0 = 0.5$ using MCA-MSS-1 and MCA-MSS-2 algorithms for integrals

i	N	ρ	MCA-MSS-1					MCA-MSS-2		
			IE		Time	FNI		IE		FNI
			Sobol	CMCM		Sobol	CMCM	CMCM		
			Rel. err.	Rel. err.	(s)	Rel. R_N	Rel. R_N	Rel. R_N	Time (s)	Rel. R_N
1	128	2e-03	0.0516	0.0515	<1e-04	0.0003	0.0004	0.0514	<1e-04	0.0005
2	1017	2e-04	0.0144	0.0144	0.02	2e-06	2e-06	0.0144	0.02	2e-05
2	4,081	6e-05	0.0144	0.0144	0.14	2e-06	3e-06	0.0144	0.16	7e-06
3	32,749	8e-06	0.0040	0.0040	10.4	8e-07	8e-07	0.0040	10.73	1e-06
4	65,521	4e-06	0.0011	0.0011	54.9	1e-07	1e-07	0.0011	55.4	2e-07

Because of the bias of the first two classes of algorithms there is a reason to present the relative errors, as well as, the approximation of the corresponding LN series with a fixed length in Tables 2 and 3, as well. We have used symbolic computations to determine the values of the systematic error R_{sys} for number of iterations up to 3.

In Table 4 we present the computed values for the systematic error R_{sys} at point $x_0 = 0.1$ and $x_0 = 0.5$ just to have an idea about the magnitude of this kind of error. One can clearly see that the systematic error decreases when the number of iterations increase. At the same time, for such small number of iteration i the systematic error dominates in almost all cases. The reason for that is that the analysis of balancing of stochastic and systematic errors are done following the assumption that Crude MC method is used. Actually, the applied algorithms, especially randomized quasi-Monte Carlo algorithms MCA-MSS-1 and MCA-MSS-2 based on *shaking* of Sobol points are of much higher quality and their stochastic errors R_N are much smaller than their systematic errors R_{sys}. Some typical values of R_N are, say 10^{-6} for relatively small number of random trajectories. These algorithms are very efficient when the norm of the kernel is relatively small. Then one may increase (but not too much) the number of iterations i, which will allow for a small number of N (and low computational complexity) to have a relatively high accuracy.

Table 4 Relative systematic error R_{sys} computed for three number of iterations $i = 1, 2, 3$ at two points x_0

x_0	$i = 1$	$i = 2$	$i = 3$
$x = 0.1$	2.28 e-03	0.635 e-03	0.176 e-03
$x = 0.5$	51.8 e-03	14.3 e-03	3.99 e-03

4 Conclusion

In our investigation four biased Monte Carlo algorithms for numerical integration have been applied. MCA-MSS-2 algorithm gives a slightly larger relative stochastic error R_N in comparison with MCA-MSS-1 related to the linear functional under consideration, and smaller relative stochastic error (almost 10 times smaller for the most cases) related to the finite number of integrals. The main disadvantage of MCA-MSS-1 and MCA-MSS-2 algorithms is the high computational complexity due to computing the minimal distance between the generated original Sobol sequences. MCM-IE is reliable enough approach, but the technique of FNI leads to smaller relative errors and give an opportunity to apply various algorithms for numerical integration. A crucial step for the quality of the biased algorithm is balancing of both stochastic R_N and systematic R_{sys} errors. The algorithms based on evaluation of FNI will suffer more from the effect of high dimensionality, because they are based on quadrature points. The developed reliable approaches for solving integral equations will be important in different areas of applied mathematics, physics, and engineering.

References

1. Bradley, P., Fox, B.: Algorithm 659: Implementing Sobol's quasi random sequence generator, ACM Trans. Math. Software 14(1) (1988) 88-100.
2. Curtiss, J. H.: Monte Carlo methods for the iteration of linear operators, J. Math Phys., vol. 32, 209-232, (1954).
3. Dimov, I.: Efficient and overconvergent Monte Carlo Methods. Parallel algorithms., *Advances in Parallel Algorithms, I.Dimov, O.Tonev (Eds.), Amsterdam, IOS Press*, 100–111, (1994).
4. Dimov, I.: Monte Carlo methods for applied scientists., *World Scientific*, New Jersey, London, Singapore, World Scientific, (2008), 291p.
5. Dimov, I. T., Georgieva, R.: Multidimensional Sensitivity Analysis of Large-scale Mathematical Models. O.P. Iliev et al. (eds.), Numerical Solution of Partial Differential Equations: Theory,

Algorithms, and Their Applications, *Springer Proceedings in Mathematics and Statistics*, 45, New York, 137–156, (2013).

6. Dimov, I.T. , Georgieva, R. : Monte Carlo method for numerical integration based on Sobol' sequences, in: LNCS 6046, Springer, (2011), 50-59.

7. Dimov, I. T., Georgieva, R., Ostromsky, Tz., Zlatev, Z.: Advanced algorithms for multidimensional sensitivity studies of large-scale air pollution models based on Sobol sequences, Computers and Mathematics with Applications 65 (3), "Efficient Numerical Methods for Scientific Applications", Elsevier, (2013), pp. 338-351.

8. Dimov, I.T., Maire, S.: A new unbiased stochastic algorithm for solving linear Fredholm equations of the second kind. Adv Comput Math 45, 1499-1519 (2019).

9. Farnoosh, R., Ebrahimi, M.: Monte Carlo Method for Solving Fredholm Integral Equations of the Second Kind, *Appl. Math. and Comp.*, 195 309–315, (2008).

10. Grozev, D., Milchev, M., Georgiev, I.: Study the work of specialized car service as queue theory. Mathematical Modeling, 4(1), 31-34, (2020).

11. Grozev, D., Milchev, M., Georgiev, I: Analysis of the load on the taxi system in a medium-sized city. In IOP Conference Series: Materials Science and Engineering (Vol. 664, No. 1, p. 012035). IOP Publishing, (2019, October).

12. Kalos, Malvin H.; Whitlock, Paula A.: Monte Carlo Methods. Wiley-VCH. ISBN 978-3-527-40760-6, (2008).

13. Kress R.: Linear integral equations, Springer, 2nd ed. DOI 10.1007/978-1-4612-0559-3 1. Integral equations. 1. Title. II. Series: Applied mathematical sciences (Springer-Verlag New York Inc.), (1999).

14. Niederreiter, H.: Low-discrepancy and low-dispersion sequences, Journal of Number Theory 30 (1988) 51-70.

15. Saito, M., Matsumoto, M.: SIMD-oriented fast Mersenne Twister: a 128-bit pseudorandom number generator, in: Keller, A., Heinrich, S., Niederreiter, H. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2006, Springer (2008) 607-622.

16. Sobol, I., Asotsky, D., Kreinin, A., Kucherenko, S.: Construction and comparison of high-dimensional Sobol' generators, Wilmott Journal (2011) 67-79.

17. Sobol, I. M.: *Monte Carlo Numerical Methods*, Nauka, Moscow, (1973), (in Russian).

18. Sobol, I. M.: On quadratic formulas for functions of several variables satisfying a general Lipschitz condition, USSR Comput. Math. 29(6) (1989) 936-941.

19. Weyl, H.: Ueber die Gleichverteilung von Zahlen mod Eins. Math. Ann. 77(3) (1916) 313-352.

20. www.math.sci.hiroshima-u.ac.jp/∼ m-mat/MT/SFMT/index.html.

Applications in Fractional Analysis

Conditional Boundedness of Generalized Proportional Caputo Fractional Differential Equations

Snezhana Hristova and Krasimira Ivanova

Abstract In this paper, we study generalized proportional Caputo fractional differential equations via Lyapunov functions. We define conditional boundedness and obtain some sufficient conditions. Several examples are provided to illustrate the application of the proved conditions.

Keywords Generalized proportional caputo fractional derivative · Differential equations · Lyapunov functions · Conditional boundedness

1   Introduction

In [7], Jarad, Abdeljawad, and Alzabut introduced a new type of fractional derivative, the so-called generalized proportional fractional derivative. This type of derivative preserves the semigroup property, possesses nonlocal character and, upon limiting cases, it converges to the original function and its derivative [12]. Some stability properties of Ulam type for generalized proportional fractional differential equations were studied in [14]. We emphasize that the boundedness was not investigated yet. In this paper, we study an initial value problem (IVP) for a nonlinear system of generalized proportional Caputo fractional differential equations. We define conditional boundedness and obtain some sufficient conditions. The base of these conditions is the application of Lyapunov functions. Several examples are provided to illustrate the application of the proven conditions.

S. Hristova (✉) · K. Ivanova
Plovdiv University P. Hilendarski, Plovdiv 4000, Bulgaria
e-mail: snehri@gmail.com

2 Notes on Fractional Calculus

We recall that the generalized proportional fractional integral and the generalized Caputo proportional fractional derivative of a function $u : [a, \infty) \to \mathbb{R}$ are defined respectively by (as long as all integrals are well defined, see [7])

$$(_a\mathcal{I}^{\alpha,\rho}u)(t) = \frac{1}{\rho^\alpha \Gamma(\alpha)} \int_a^t e^{\frac{\rho-1}{\rho}(t-s)} (t-s)^{\alpha-1} u(s)\, ds, \tag{1}$$
$$t \in (a, b], \quad \alpha \geq 0, \quad \rho \in (0, 1],$$

and

$$(_a^C\mathcal{D}^{\alpha,\rho}u)(t) = (_a\mathcal{I}^{1-\alpha,\rho}(\mathcal{D}^{1,\rho}u))(t)$$
$$= \frac{1}{\rho^{1-\alpha}\Gamma(1-\alpha)} \int_a^t e^{\frac{\rho-1}{\rho}(t-s)} (t-s)^{-\alpha} (\mathcal{D}^{1,\rho}u)(s)\, ds, \tag{2}$$
$$\text{for } t \in (a, b], \quad \alpha \in (0, 1), \quad \rho \in (0, 1],$$

where $(\mathcal{D}^{1,\rho}u)(t) = (\mathcal{D}^\rho u)(t) = (1-\rho)u(t) + \rho u'(t)$.

Remark 1 If $\rho = 1$, then the generalized Caputo proportional fractional derivative is reduced to the classical Caputo fractional derivative of order $\alpha \in (0, 1)$: $_a^C\mathcal{D}^\alpha$.

Remark 2 Note the generalized proportional Caputo fractional derivative easily could be generalized for $f \in C([a, b], \mathbb{R}^n)$ via component-wise.

Remark 3 The generalized proportional Caputo fractional derivative of a constants is not zero for $\rho \in (0, 1)$.

Remark 4 The relation

$$(_a^C\mathcal{D}^{\alpha,\rho}e^{\frac{\rho-1}{\rho}(.)})(t) = 0 \quad \text{for } t > a \tag{3}$$

is known from [7, Remark 3.2].

Lemma 1 (Proposition 5.2 [7]) *For $\rho \in (0, 1]$ and $q \in (0, 1)$, we have*

$$(_a^C\mathcal{D}^{q,\rho}(e^{\frac{\rho-1}{\rho}t}(t-a)^{\beta-1})(t) = \frac{\rho^q \Gamma(\beta)}{\Gamma(\beta-q)} e^{\frac{\rho-1}{\rho}t}(t-a)^{\beta-1-q}, \quad \beta > 0. \tag{4}$$

We will use the explicit form of the solution of the initial value problem for the scalar linear generalized proportional Caputo fractional differential equation which is given in Example 5.7 [7] and which is (with necessary slight corrections) given in the following result.

Lemma 2 *The solution of the scalar linear generalized proportional Caputo fractional initial value problem*

$$({}_a^C\mathcal{D}^{\alpha,\rho}u)(t) = \lambda u(t), \quad u(a) = u_0, \quad \alpha \in (0, 1), \ \rho \in (0, 1] \tag{5}$$

has a solution

$$u(t) = u_0 e^{\frac{\rho-1}{\rho}(t-a)} E_\alpha(\lambda(\frac{t-a}{\rho})^\alpha), \tag{6}$$

where $E_\alpha(t)$ is the Mittag–Leffler function of one parameter.

Lemma 3 ([2]) *Let the function $u \in C^1([a, b], \mathbb{R})$ with $a, b \in \mathbb{R}$, $b \leq \infty$ (if $b = \infty$ then the interval is half open), and $q \in (0, 1)$, $\rho \in (0, 1]$ be two reals. Then,*

$$({}_a^C\mathcal{D}^{q,\rho}u^2)(t) \leq 2u(t)({}_a^C\mathcal{D}^{q,\rho}u)(t), \quad t \in (a, b]. \tag{7}$$

The fractional derivatives for scalar functions could be easily generalized to the vector case by taking fractional derivatives with the same fractional order for all components.

3 Statement of the Problem and Basic Definitions

Consider the initial value problem (IVP) for a nonlinear system of generalized proportional Caputo fractional differential equations with $q \in (0, 1)$, $\rho \in [0, 1]$

$$\begin{aligned}
{}_{t_0}^C D_t^{q,\rho}y(t) &= f(t, y(t)) \ \text{ for } t > t_0 \\
y(t_0) &= y_0,
\end{aligned} \tag{8}$$

where $y_0 \in \mathbb{R}^n$, the function $f \in C([t_0, \infty) \times \mathbb{R}^n, \mathbb{R}^n)$, $f(t, 0) = 0$ for $t \geq 0$.

We will generalize Lipschitz stability for ordinary differential equations [6] to systems of generalized proportional Caputo fractional differential equations.

The IVP for the system of generalized proportional Caputo fractional differential equations (8) is said to be conditionally bounded if there exist constants $M = M(q, \rho, t_0) \geq 1$ and $\delta > 0$, such that for any initial value $y_0 \in \mathbb{R}^n : ||y_0|| < \delta$ the inequality $||y(t; t_0, y_0)|| \leq M$ holds for $t \geq t_0$.

Let $J \subset \mathbb{R}_+$, $0 \in J$, $\rho > 0$. Consider the following sets:

$$\mathcal{K}(J) = \{a \in C[J, \mathbb{R}^+] : a(0) = 0, \ \text{a(r) is strictly increasing in } J\},$$
$$K(J) = \{a \in C[J, \mathbb{R}^+] : a(0) = 0, \ \text{a(r) is strictly increasing in } J, \ \text{and}$$
$$a(r) \leq K_a r \ \text{for some constant} \ K_a > 0\},$$
$$S_\rho = \{x \in \mathbb{R}^n : ||x|| \leq \rho\}.$$

Remark 5 The function $a(u) = K_1 u$, $K_1 > 0$ is from the class $K(\mathbb{R}_+)$ with $K_a = K_1$. The function $a(u) = K_2 u^2$, $K_2 \in (0, 1]$ is from the class $M([1, \infty))$ with $q(u) = \sqrt{\frac{u}{K_2}} \geq 1$ for $u \geq 1$.

Lemma 4 ([8]). *Let* $v \in C([a, b], \mathbb{R})$ *be such that* $(t - a)^{1-q} v \in C([a, b], \mathbb{R})$ *and there exists a point* $\tau \in (a, b)$: $v(\tau) = 0$ *and* $v(t) \leq 0$ *for* $t \in [a, \tau]$. *Then* $_a^C D_\tau^q v(\tau) \geq 0$.

We will use the following scalar comparison scalar generalized proportional Caputo fractional differential equation:

$$_{t_0}^C D_t^q u(t) = g(t, u(t)) \text{ for } t > t_0, \qquad u(t_0) = u_0, \tag{9}$$

where $u, u_0 \in \mathbb{R}$, $g : [t_0, \infty) \times \mathbb{R} \to \mathbb{R}$.

We obtain some comparison results.

We introduce the following condition:

(A) The function $g(t, u) \in C([0, \infty) \times \mathbb{R}_+, \mathbb{R})$ is strictly decreasing w.r.t. its second argument and $g(t, 0) = 0$ for $t \geq t_0$.

In our main results, we will use the conditional boundedness of the zero solution of the scalar comparison generalized proportional Caputo fractional differential equation (9).

Example 1 Consider the scalar generalized proportional Caputo fractional differential equation

$$_0^C D_t^{0.25, 0.5} u(t) = \frac{0.5^{0.25} \Gamma(7)}{\Gamma(6.75)} t^{5.75} \text{ for } t > 0, \qquad u(0) = u_0. \tag{10}$$

From Lemma 1, we get $(_0^C \mathcal{D}^{0.25, 0.5}(e^{-t} t^6))(t) = \frac{0.5^{0.25} \Gamma(7)}{\Gamma(7-0.25)} e^{-t} t^{6-0.25}$ and by Remark 4.

Thus, the solution of (10) is given by $u(t) = (u_0 + t^6) e^{-t}$, $t > 0$, and it is conditionally bounded with $M = 120 > \sup_{t>0} e^{-t} t^6$ and $\delta = 100$ (see Fig. 1).

Example 2 Consider the scalar generalized proportional Caputo fractional differential equation

$$_0^C D_t^{0.4, 0.5} u(t) = 0.75 u(t) \text{ for } t > 0, \qquad u(0) = u_0. \tag{11}$$

From Lemma 2, the solution is $u_0 e^{-t} E_{0.4}(0.75(2t)^{0.4})$. The solution is conditionally bounded with $M = 2.4$ and $\delta = 1$ (see Fig. 2).

Lemma 5 *Assume the following conditions are satisfied:*

1. *The assumption (A) is satisfied.*
2. *The function* $x^*(t) = x(t; t_0, x_0) : [t_0, T) \to \Delta$, $x^*(.) \in C^{q,\rho}([t_0, T))$ *is a solution of (8), where* $\Delta \subset \mathbb{R}^n$, $0 \in \Delta$.
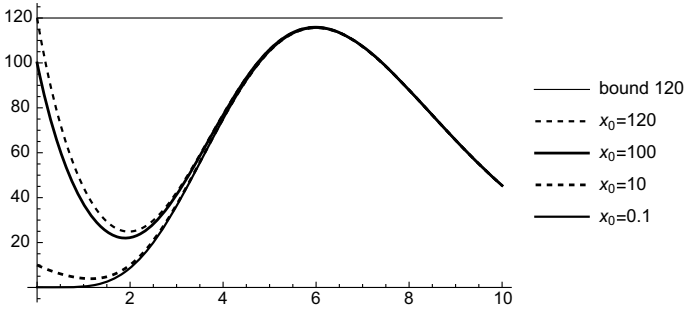
Fig. 1 Graphs of the solutions of (10) with various initial values
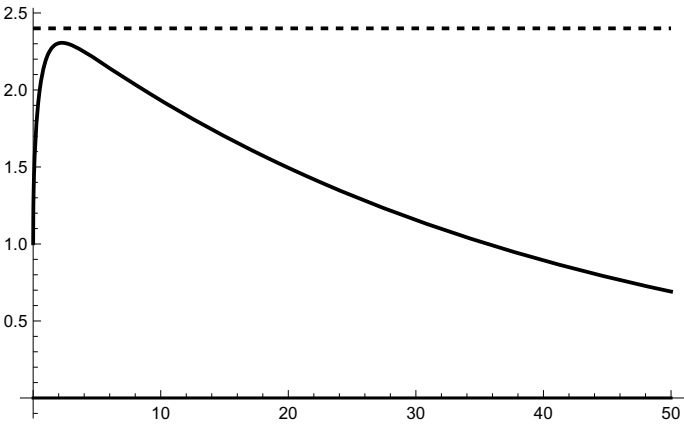

Fig. 2 Graph of the solution of (11) with $x_0 = 1$ and the bound $M = 2.4$

3. *The function $V \in \Lambda([t_0, T), \Delta)$ is such that for $t \in [t_0, T)$, the inequality*

$$ {}_{t_0}^{C} D_t^{q,\rho} V(t, x^*(t)) \le g(t, V(t, x^*(t))) $$

holds;

If $V(t_0, x_0) \le u_0$, then the inequality $V(t, x^*(t)) \le r(t)$ for $t \in [t_0, T)$ holds, where $r(t) = r(t; t_0, u_0)$ is the maximal solution on $[t_0, T)$ of (9) with $u_0 \ge 0$.

Proof Let $m(t) = V(t, x^*(t))$, $t \ge t_0$. We will prove

$$ m(t) \le u(t), \quad t \ge t_0. \tag{12} $$

Let $\varepsilon > 0$ be an arbitrary number. We will prove

$$ m(t) < u(t) + \varepsilon, \quad t \ge [t_0, T]. \tag{13} $$

Note $m(t_0) = V(t_0, x_0) \le u_0 < u_0 + \varepsilon$, i.e., inequality (13) holds for $t = t_0$. If the inequality (13) is not true, then there exists a point $t^* \in (t_0, s_1]$, such that $m(t^*) = u(t^*) + \varepsilon$, $m(t) < u(t) + \varepsilon$, $t \in [t_0, t^*)$.

From Lemma 4 with $a = t_0$, $b = s_1$, $\tau = t^*$ and $v(t) = m(t) - u(t) - \varepsilon$, the inequality ${}^{C}_{t_0}D^q_t m(t^*) \ge^C_{t_0} D^q_t u(t^*) = g(t^*, u(t^*))$ holds.

From conditions (A6) and 3(i), the inequality ${}^{}_{t_0}D^q_t m(t^*) \le g(t^*, m(t^*)) = g(t^*, u(t^*) + \varepsilon) < g(t^*, u(t^*))$ holds. The contradiction proves the validity of (13). Since ε is an arbitrary positive number, we obtain inequality (12) for $t \in [t_0, s_1]$.   □

4 Main Results

Theorem 1 *Let the following conditions be satisfied:*

1. *Assumption (A) is fulfilled.*
2. *There exist a function $V \in \Lambda(\mathbb{R}_+, \mathbb{R}^n)$ and*

    (i) *the inequalities*

    $$b(||x||) \le V(t, x) \le a(||x||), \quad x \in \mathbb{R}^n, t \in \mathbb{R}_+$$

    holds, where $a \in K([0, B])$, $b \in \mathcal{K}([0, B])$, $B > 0$;
    (ii) *for any initial data and any solution $x(t)$ of (8) defined on $[t_0, \infty)$, the inequality*

    $$ {}^{C}_{t_0}D^{q,\rho}_t V(t, x(t)) \le g(t, V(t, x(t))), \quad t > t_0 $$

    holds.

3. *The zero solution of (9) is conditionally bounded.*

Then the zero solution of (8) is conditionally bounded.

Proof Let the zero solution of (9) be conditionally bounded. From condition 3, there exist constants $M \ge 1$ and $\delta > 0$, such that for any $u_0 \in \mathbb{R} : |u_0| < \delta_1$ the inequality

$$|u(t; t_0, u_0)| \le M \text{ for } t \ge t_0 \tag{14}$$

holds, where $u(t; t_0, u_0)$ is a solution of (9) with the initial data (t_0, u_0).

From the inclusions $a \in K([0, B])$ and $b \in M([0, B])$, there exist a function $q_b(u)$ and a positive constant K_a. Without loss of generality, we can assume $K_a \ge 1$. Choose the constant M_1 such that $M_1 > \max\{1, b^{-1}(a(B)), b^{-1}(M)\}$ and $\delta_2 \le \frac{B}{2M_1} \le \frac{B}{2} \le B$. Therefore, $2M_1\delta_2 \le B$.

Let $\delta = \min\left\{\delta_1, \delta_2, \frac{\delta_1}{K_a}\right\}$. Choose the initial value $x_0 : ||x_0|| < \delta$. Therefore, $||x_0|| < \delta \le \delta_2 \le B$, i.e. $x_0 \in S_B$. Consider the solution $x(t) = x(t; t_0, x_0)$ of system (8) for the chosen initial data (t_0, x_0).

Let $u_0^* = V(t_0, x_0)$. From the choice of x_0 and the properties of the function $a(u)$ applying condition $2(i)$, we get $u_0^* = V(t_0, x_0) \leq a(||x_0||) \leq K_a||x_0|| < K_a\delta \leq \delta_1$. Therefore, the function $u^*(t)$ satisfies (14) for $t \geq t_0$ with $u_0 = u_0^*$, where $u^*(t) = u(t; t_0, u_0^*)$ is a solution of (9) with an initial data (t_0, u_0^*).

Let $\varepsilon \in (0, M_1]$ be an arbitrary number. We will prove

$$V(t, x(t)) < b(M_1 + \varepsilon), \quad t \geq t_0. \tag{15}$$

For $t = t_0$, we get $V(t_0, x_0) \leq a(||x_0||) \leq a(B) \leq b(M_1) \leq b(M_1 + \varepsilon)$, i.e., inequality (15) holds.

Assume (15) is not true. Therefore, there exists a point $T > t_0$, such that $V(t, x(t)) < b(M_1 + \varepsilon)$ for $t \in [t_0, T)$, $V(T, x(T)) = b(M_1 + \varepsilon)$. Then for $t \in [t_0, T]$ applying 2(i), we obtain the inequalities $||x(t)|| \leq b^{-1}(V(t, x(t)) \leq M_1 + \varepsilon \leq 2M_1$, i.e., $x(t) \in S_{2M_1}$ for $t \in [t_0, T]$, and according to condition 2(ii) of Theorem 1 with $\tau = T$, it follows that condition 3(i) of Lemma 5 is satisfied for the solution $x(t)$ on the interval $[t_0, T]$ and $\Delta = S_{2M_1}$.

According to Lemma 5, we get

$$V(t, x(t)) \leq u^*(t) \text{ for } t \in [t_0, T]. \tag{16}$$

From inequality (16) and condition $2(i)$, we obtain

$$M_1 < M_1 + \varepsilon = b^{-1}(V(T, x(T))) \leq b^{-1}(u^*(T)) \leq b^{-1}(M) \leq M_1. \tag{17}$$

The contradiction proves the validity of (15). From inequality (15) and condition 2(i), we have Theorem 1. $\square$

Theorem 2 *Let the conditions of Theorem 1 be satisfied where the condition 2(i) is replaced by*

$2^*(i)$ the inequalities $\lambda_1(t)||x||^2 \leq V \leq \lambda_2(t)||x||^2$, $x \in S_\rho$, $t \in \mathbb{R}^+$ holds, where $\lambda_1, \lambda_2 \in C(\mathbb{R}_+, (0, \infty))$ and there exists positive constant $A_1, A_2 : A_1 < A_2$, such that $\lambda_1(t) \geq A_1$, $\lambda_2(t) \leq A_2$ for $t \geq 0$, and $\rho > 0$.

If the zero solution of (9) is conditionally bounded, then the zero solution of (8) is conditionally bounded.

Proof The proof is similar to the one in Theorem 1, where $M_1 = \sqrt{M\frac{A_2}{A_1}}$.

Example 3 Consider the generalized proportional Caputo fractional differential equations

$$\begin{aligned}
{}_0^C D_t^{0.25} x_1(t) &= \frac{0.5^{0.25}\Gamma(7)}{0.45\Gamma(6.75)} t^{5.75} x_1(t) e^{-x_1^2(t)} - 0.5x_1(t) - x_2(t), \\
{}_{t_0}^C D_t^{0.25} x_2(t) &= -0.5x_2(t) + x_1(t) \qquad \text{for } t > t_0.
\end{aligned} \tag{18}$$

Let $V(t, x) = x_1^2 + x_2^2$, $x = (x_1, x_2)$. Then according to Lemma 3, for any $t > 0$, we get ${}_0^C D_t^{0.25, 0.5} V(t, x(t)) \leq 2x_1(t) {}_0^C D_t^{0.25, 0.5} x_1(t) + 2x_2(t) {}_0^C D_t^{0.25, 0.5} x_2(t) =$
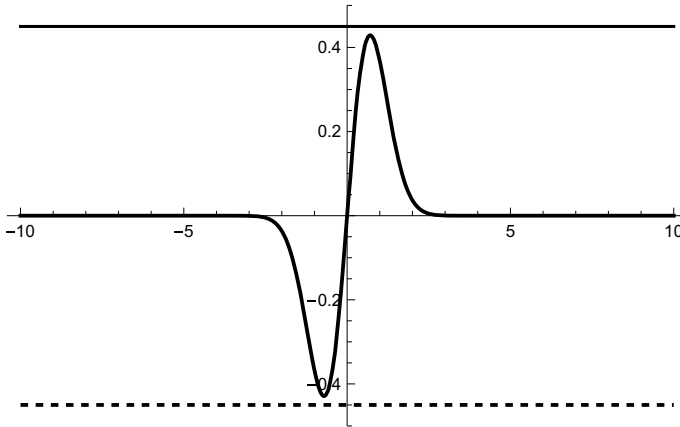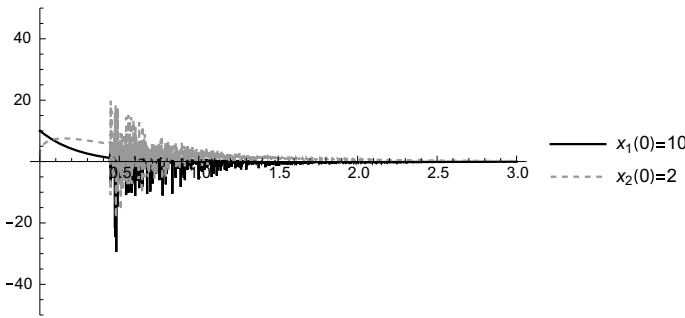
Fig. 3 Graph of the function xe^{-x^2}
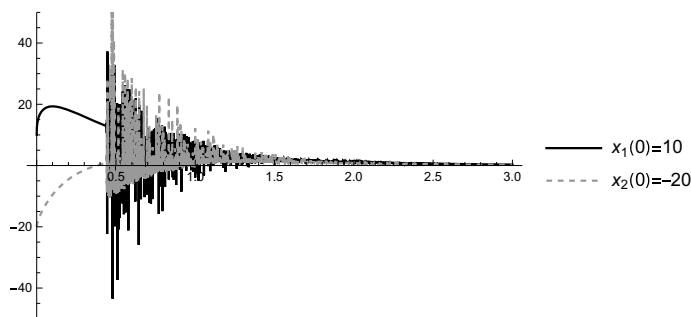

Fig. 4 Graph of the solution of (11)

$-x_1^2 - x_2^2 + \frac{0.5^{0.25}\Gamma(7)}{0.45\Gamma(6.75)} t^{5.75} e^{-x_1^2(t)} x_1(t) \leq \frac{0.5^{0.25}\Gamma(7)}{\Gamma(6.75)} t^{5.75}$ because $-0.45 < xe^{-x^2} < 0.45$ (see Fig. 3).

Consider the scalar equation (10) whose solution according to Example 1 is conditionally bounded. According to Theorem 1, the zero solution of (18) is conditionally bounded.

Example 4 Consider the generalized proportional Caputo fractional differential equations

$$\begin{aligned} {}^{C}_{0}D_t^{0.4,0.5} x_1(t) &= 0.325 x_1(t) - x_2(t), \\ {}^{C}_{t_0}D_t^{0.4,0.5} x_2(t) &= 0.325 x_2(t) + x_1(t) \text{ for } t > 0. \end{aligned} \tag{19}$$

Let $V(t, x) = x_1^2 + x_2^2$, $x = (x_1, x_2)$. Then according to Lemma 3, for any $t > 0$, we get ${}^{C}_{0}D_t^{0.4,0.5} V(t, x(t)) \leq 2x_1(t) {}^{C}_{0}D_t^{0.4,0.5} x_1(t) + 2x_2(t) {}^{C}_{0}D_t^{0.4,0.5} x_2(t) = 0.75(x_1^2 + x_2^2) = 0.75 V(t, x(t))$.

Fig. 5 Graph of the solution of (11)

Consider the scalar equation (11) whose solution according to Example 2 is conditionally bounded. According to Theorem 1, the zero solution of (19) is conditionally bounded (see Figs. 4 and 5).

References

1. Agarwal, R. P., O'Regan, D., Hristova, S.: Stability of Caputo fractional differential equations by Lyapunov functions, Appl. Math. 60, 6, 653-676 ,(2015).
2. Almeida, R., Agarwal, R.P., Hristova, S., O'Regan, D.: Quadratic Lyapunov functions for stability of generalized proportional fractional differential equations with applications to neural networks, Axioms 10(4), 322, (2021); https://doi.org/10.3390/axioms10040322
3. Das, Sh. : Functional Fractional Calculus, Springer-Verlag Berlin Heidelberg, 2011.
4. Diethelm, K.: The Analysis of Fractional Differential Equations, Springer-Verlag Berlin Heidelberg, 2010.
5. Alzabut, J., Abdeljawad, T., Jarad, F., Sudsutad, W.: A Gronwall inequality via the generalized proportional fractional derivative with applications, J. Ineq. Appl. 2019:101, (2019).
6. Dannan, F.M., Elaydi, S.: Lipschitz stability of nonlinear systems of differential equations, J. Math. Anal. Appl. 113, 562-577, (1986).
7. Jarad, F., Abdeljawad, T., Alzabut, J.: Generalized fractional derivatives generated by a class of local proportional derivatives, Eur. Phys. J. Special Topics 226, 3457-3471 (2017).
8. Lakshmikantham, V., Leela, S., Vasundhara Devi, J.: Theory of Fractional Dynamic Systems, Cambridge, Cambridge : CSP, 2009, 170 p.
9. Podlubny, I.: Fractional Differential Equatio, Academic Press: San Diego, 1999.
10. Li, Y., Chen, Y.Q., Podlubny, I.: Stability of fractional-order nonlinear dynamic systems: Lyapunov direct method and generalized Mittag-Leffler stability, Comput. Math. Appl. 59, (5), 1810–18, (2010).
11. Aguila-Camacho N., Duarte-Mermoud M.A., Gallegos J.A., Lyapunov functions for fractional order systems, Commun. Nonlinear Sci. Numer. Simul. 19 (9), 2951-2957, (2014). https://doi.org/10.1016/j.cnsns.2014.01.022
12. F. Jarad, T. Abdeljawad, J. Alzabut, Generalized fractional derivatives generated by a class of local proportional derivatives, *Eur. Phys. J. Spec. Top.* **226**, 3457–3471, (2017). https://doi.org/10.1140/epjst/e2018-00021-7

13. Sudsutad, W., Alzabut, J., Nontasawatsri, S., Thaiprayoon, C.: Stability analysis for a generalized proportional fractional Langevin equation with variable coefficient and mixed integro-differential boundary conditions, J. Nonlinear Funct. Anal. 2020, Article ID 23, (2020).
14. Khaminsou, B., Thaiprayoon, C., Sudsutad, W., Jose, S.A.: Qualitative analysis of a proportional Caputo fractional differential pantograph differential equations with mixed nonlocal conditions, Nonl. Funct. Anal. Appl. 26,(1) 197–223, (2021). https://doi.org/10.22771/nfaa.2021.26.01.14

Practical Stability of Generalized Proportional Caputo Fractional Differential Equations by Lyapunov Functions

Tzanko Donchev and Snezhana Hristova

Abstract The practical stability of a nonlinear nonautonomous Caputo fractional differential equation is studied using Lyapunov-like functions. The novelty of this paper is based on the new definition of the derivative of a Lyapunov like function along the given fractional differential equation. Comparison results using this definition for scalar fractional differential equations are presented. Several sufficient conditions for practical stability, practical quasi-stability, strongly practical stability of the zero solution, and the corresponding uniform types of practical stability are established.

Keywords Practical stability · Proportional caputo derivative · Lyapunov functions

1 Introduction

Many practical qualitative questions and investigations of the properties of the equilibrium of models are connected with stability. In the literature, there are various types of stability. One of them is practical stability [14]. Note that this type of stability is neither weaker nor stronger than the usual stability. Practical stability is studied for various types of differential equations (see, for example, [3–7, 9, 11–13, 16]).

The stability of fractional order systems with proportional Caputo fractional derivatives is quite recent (see, for example, [2]).

In this paper, the practical stability of nonlinear nonautonomous generalized proportional Caputo fractional differential equations is defined and studied using Lyapunov functions and comparison results. Sufficient conditions for practical stability,

T. Donchev (✉)
Department of Mathematics, UACG, 1 Hr. Smirnenski bvd, 1046 Sofia, Bulgaria
e-mail: tzankodd@gmail.com

S. Hristova
Faculty of Mathematics and Informatics, University of Plovdiv "Paisii Hilendarski", 4000 Plovdiv, Bulgaria

quasi-practical stability, and strong practical stability are obtained. Some examples illustrate the obtained results.

2   Notes on Fractional Calculus

In engineering, the fractional order is $q \in (0, 1)$ and we will restrict our considerations to this case.

1: The generalized proportional fractional integral of a function $u : [a, \infty) \to R$ is defined by (as long as all integrals are well defined, see [10])

$$
(_a\mathcal{I}^{q,\rho}u)(t) = \frac{1}{\rho^\alpha \Gamma(\alpha)} \int_a^t e^{\frac{\rho-1}{\rho}(t-s)} (t-s)^{q-1} u(s) \, ds,
$$
$$
t \in (a, b], \ q \geq 0, \ \rho \in (0, 1]. \tag{1}
$$

2: The the generalized Caputo proportional fractional derivative of a function $u : [a, \infty) \to R$ is defined by (as long as all integrals are well defined, see [10])

$$
(_a^C\mathcal{D}^{q,\rho}u)(t) = (_a\mathcal{I}^{1-q,\rho}(\mathcal{D}^{1,\rho}u))(t)
$$
$$
= \frac{1}{\rho^{1-q}\Gamma(1-q)} \int_a^t e^{\frac{\rho-1}{\rho}(t-s)} (t-s)^{-q} (\mathcal{D}^\rho u)(s) \, ds,
$$
$$
\text{for } t \in (a, b], \ q \in (0, 1), \ \rho \in (0, 1], \tag{2}
$$

where $(\mathcal{D}^\rho u)(t) = (1-\rho)u(t) + \rho u'(t)$.

Remark 1 If $\rho = 1$, then the generalized Caputo proportional fractional derivative is reduced to the classical Caputo fractional derivative of order $q \in (0, 1)$: $(_a^C\mathcal{D}^{q,\rho}u)(t) = {_a^C}\mathcal{D}^q u(t)$.

Remark 2 Note the generalized proportional Caputo fractional derivative is generalized for $u \in C([a, b], R^n)$ via component-wise.

Lemma 1 *The generalized proportional fractional derivative of a constant $K \in R$ is*

$$
(_a^C\mathcal{D}^{q,\rho}K)(t) = \frac{(1-\rho)K}{\rho^{1-q}\Gamma(1-q)} \int_a^t e^{\frac{\rho-1}{\rho}(t-s)} (t-s)^{-q} \, ds
$$
$$
= \frac{(1-\rho)K}{\rho^{1-q}\Gamma(1-q)} \frac{1}{(\frac{\rho-1}{\rho})^{1+q}} [\Gamma(1+q) - \Gamma(1+q, \frac{\rho-1}{\rho}(t-a))] \tag{3}
$$
$$
= \frac{K}{\rho^{1-q}\Gamma(1-q)} \frac{1}{(\rho-1)^q} [\Gamma(1+q, \frac{\rho-1}{\rho}(t-a)) - \Gamma(1+q)] \text{ for } t > a.
$$

Remark 3 The relation

$$
(_a^C\mathcal{D}^{q,\rho}e^{\frac{\rho-1}{\rho}(\cdot)})(t) = 0 \text{ for } t > a \tag{4}
$$

is known from [10, Remark 3.2].

Definition 1 We say $u \in C^{q,\rho}([t_0, T], R^n)$ if $u(t)$ is differentiable (i.e., $u'(t)$ exists) and the generalized proportional Caputo fractional derivative $({}^C_a \mathcal{D}^{q,\rho} u)(t)$ exists.

We will use the explicit form of the solution of the initial value problem for the scalar linear generalized proportional Caputo fractional differential equation which is given in Example 5.7 [10] and which is (with necessary slight corrections) given in the following result.

Lemma 2 *The scalar linear generalized proportional Caputo fractional initial value problem*

$$({}^C_a \mathcal{D}^{\alpha,\rho} u)(t) = \lambda u(t), \ u(a) = u_0, \ \alpha \in (0, 1), \ \rho \in (0, 1] \tag{5}$$

has a solution

$$u(t) = u_0 e^{\frac{\rho-1}{\rho}(t-a)} E_\alpha(\lambda(\frac{t-a}{\rho})^\alpha), \tag{6}$$

where $E_\alpha(t)$ is the Mittag–Leffler function of one parameter.

We will recall the following property of the Mittag–Leffler function.

Proposition 1 (Theorem 1.2 *[15]*) *For every $q \in (0, 1)$, the function $\frac{e^t}{q} - E_q(t^q)$ is completely monotonic.*

As a partial case of Proposition 1, it follows

Corollary 1 *Let $q \in (0, 1)$. Then $E_q(t^q) < \frac{e^t}{q}, \ t \geq 0$.*

In this paper, we will use the following result:

Lemma 3 ([2]) *Let $u \in C^{q,\rho}([t_0, \infty), R^n)$. Then, for any $t \geq t_0$, the inequality*

$$({}^C_{t_0} \mathcal{D}^{q,\rho} \left(u^T(.) u(.) \right)(t) \leq 2 \, x^T(t) \, ({}^C_{t_0} \mathcal{D}^{q,\rho} u)(t)$$

holds.

3　Preliminary Results

Consider the initial value problem (IVP) for the system of nonlinear fractional differential equations (FrDE) with a generalized proportional Caputo fractional derivative for $0 < q < 1, \rho \in (0, 1]$,

$$({}^C_{t_0} \mathcal{D}^{q,\rho} x)(t) = f(t, x(t)), \ x(t_0) = x_0 \tag{7}$$

where $x_0 \in R^n, \ f \in C([t_0, \infty) \times R^n, R^n), t_0 \geq 0$ is a given number.

In this paper, we will assume that, for any initial value $(x_0) \in R^n$, the IVP for FrDE (7) has a solution $x(t; t_0, x_0) \in C^{q,\rho}([t_0, \infty), R^n)$. Note some sufficient conditions for the existence of solutions of IVP for FrDE (7) are given in [8].

The goal of our paper is to study various types of practical stability of the zero solution of the IVP for FrDE (7). We now present some types of practical stability of the zero solution of fractional differential equations. In the definition below, we assume $x(t; t_0, x_0)$ is any solution of the FrDE (7).

Definition 2 Let positive constants λ, A, $\lambda \leq A$ be given. The nonlinear system of generalized proportional Caputo fractional differential equations (7) is said to be

(S1) (λ, A)—*practically stable* if, for any $x_0 \in R^n$, the inequality $||x_0|| < \lambda$ implies $||x(t; t_0, x_0)|| < A$ for $t \geq t_0$;

(S2) (λ, A, T)—*practically quasi-stable* if, for any $x_0 \in R^n$, the inequality $||x_0|| < \lambda$ implies $||x(t; t_0, x_0)|| < A$ for $t \geq t_0 + T$;

(S3) (λ, A, K, T)—*strongly practically stable* if, for any $x_0 \in R^n$, the inequality $||x_0|| < \lambda$ implies $||x(t; t_0, x_0)|| < A$ for $t \geq t_0$ and $||x(t; t_0, x_0)|| < K$ for $t \geq t_0 + T$, where the positive constants λ, A, K, T, $K < \lambda < A$ are given.

The stability of the zero solution does not imply practical stability.

In this paper, we will use the followings sets:

$$\mathcal{K} = \{a \in C[R_+, R_+] : a \text{ is strictly increasing and } a(0) = 0\},$$
$$B(\lambda) = \{x \in R^n : \|x\| < \lambda\},$$
$$\tilde{B}(\lambda) = \{x \in R^n : \|x\| \leq \lambda\} \; \lambda = const > 0.$$

In our results, we will use the initial value problem for scalar fractional differential equations of the following form:

$$_{t_0}^{c}D^q u(t) = g(t, u(t)), \;\; t > t_0, \; u(t_0) = u_0 \tag{8}$$

where $u_0 \in R$, $g : [t_0, \infty) \times R \to R$. We will assume in the paper for any initial value $u_0 \in R$ the IVP for the scalar FrDE (8) has a solution $u(t; t_0, u_0) \in C^{q,\rho}([t_0, \infty), R)$.

In this paper, we will study the connection between the practical stability of the system FrDE (7) and the practical stability of the scalar FrDE (8) by the help with Lyapunov functions $V(t, x) \in C(J \times \Delta, [0, \infty))$ which are locally Lipschitzian with respect to the second argument (here $J = [t_0, T]$, $T \leq \infty$, $\Delta \subset R^n$).

In our further investigations, we will use the following comparison result by Lyapunov functions.

Lemma 4 (Lemma 6 *[1]*) *Assume the following conditions are satisfied:*

1. *The function $x(t) = x(t; t_0, x_0) \in C^{q,\rho}([t_0, T], \Delta)$ is a solution of the FrDE (7), where $\Delta \subset R^n$, t_0, $T \in R_+$, $t_0 < T \leq \infty$ are given constants, $x_0 \in \Delta$.*
2. *The function $g \in C([t_0, T] \times R, R)$.*
3. *The Lyapunov function $V(t, x) \in C([t_0, T] \times \Delta, [0, \infty))$ and the inequality*

$$(_{t_0}^{C}\mathcal{D}^{q,\rho}V(\cdot, x(\cdot))(t) \leq g(t, V(t, x(t))), \; t \in (t_0, T) \tag{9}$$

holds.

4. *The function $u^*(t) = u(t; t_0, u_0) \in C^{q,\rho}([t_0, T], [0, \infty))$ is the maximal solution of the initial value problem (8).*

Then the inequality $V(t_0, x_0) \leq u_0$ implies $V(t, x(t)) \leq u^*(t)$ for $t \in [t_0, T]$.

In some partial cases of the function $g(t, x)$ in Lemma 4, we obtain the following results:

Corollary 2 *Assume the conditions of Lemma 4 are satisfied with $g(t, u) \equiv 0$. Then for $t \in [t_0, T]$ the inequality $V(t, x(t)) \leq V(t_0, x_0)e^{\frac{\rho-1}{\rho}(t-t_0)}$ holds.*

Corollary 3 *Assume the conditions of Lemma 4 are satisfied with $g(t, u) = Ku$, where $K \in (-\infty, \frac{1-\rho}{\rho^{1-q}})$, $K \neq 0$. Then the inequality*

$$V(t, x(t)) \leq V(t_0, x_0)e^{\frac{\rho-1}{\rho}(t-t_0)}E_q(K(\frac{t-t_0}{\rho})^q), \quad t \in [t_0, T]$$

holds.

4 Main Results

We obtain some sufficient conditions for the practical stability of the system FrDE (7).

Theorem 1 *Suppose the following conditions hold:*

1. *The function $g \in C([t_0, \infty) \times R, R)$.*
2. *There exists $\Delta \subset R^n$, such that for any $x_0 \in \Delta$ the FrDE (7) has a solution $x(t) = x(t; t_0, x_0) \in C^{q,\rho}([t_0, \infty), \Delta)$.*
3. *There exists a Lyapunov function $V \in C([t_0, \infty) \times \Delta, [0, \infty))$ and there exists a constant $A > 0$ such that $\tilde{B}(A) \subset \Delta$ such that*

   (i) for any solution of (7) $x(.) \in \tilde{B}(A)$, $t \geq t_0$, the inequality

   $$({}^C_{t_0}\mathcal{D}^{q,\rho}V(\cdot, x(\cdot))(t) \leq g(t, V(t, x(t))), \quad t > t_0 \tag{10}$$

   holds;

   (ii) $b(||x||) \leq V(t, x) \leq a(||x||)$ for $t \geq t_0$, $x \in \tilde{B}(A)$, where $a, b \in \mathcal{K}$.

4. *The scalar FrDE (8) is $(a(\lambda), b(A))$—practically stable, where the constant $\lambda > 0$ is given such that $\lambda \leq A$, $a(\lambda) \leq b(A)$.*

Then the zero solution of the system of FrDE (7) is (λ, A)—practically stable.

Proof Let the couple (λ, A) be as it is described in conditions 3(ii) and 4. According to condition 4, if the initial value $\overline{u}_0$ be such that $|\overline{u}_0| < a(\lambda)$, then the inequality

$$|\overline{u}(t; t_0, \overline{u}_0)| < b(A) \text{ for } t \geq t_0, \tag{11}$$

holds, where $\overline{u}(t; t_0, \overline{u}_0)$ is a solution of (8).

Choose the initial value $x_0 \in B(\lambda) \subset \Delta$. According to condition 2, the corresponding solution $x(t) = x(t; t_0, x_0) \in \Delta, \ t \geq t_0$. We will prove the inequality

$$\|x(t)\| < A \text{ for } t \geq t_0. \tag{12}$$

Since $||x_0|| < \lambda \leq A$ inequality (12) holds for $t = t_0$. Assume (12) is not true. Thus, there exists a point $\tau > t_0$ such that

$$\|x(t)\| < A \ \text{ for } t \in [t_0, \tau) \text{ and } \|x(\tau)\| = A. \tag{13}$$

Thus, $x(\cdot) \in \tilde{B}(A), \ t \in [t_0, \tau]$. Let $u_0 = V(t_0, x_0)$. According to condition 3(ii) and the choice of x_0, we obtain $u_0 < a(\lambda)$. From Lemma 3 with $\Delta = \tilde{B}(A)$ and $T = \tau$, we obtain

$$V(t, x(t)) \leq u^*(t; t_0, u_0) \text{ for } t \in [t_0, \tau]; \tag{14}$$

where $u^*(t; t_0, u_0)$ is the maximal solution of (8). From inequalities (14), (11) and condition 3(ii), we get

$$b(A) = b(\|x(\tau)\|) \leq V(\tau, x(\tau)) \leq u^*(\tau; t_0, u_0) < b(A). \tag{15}$$

The obtained contradiction proves inequality (12) is true. Thus, the zero solution of FrDE (7) is (λ, A)—practically stable.

Theorem 2 *Suppose the following conditions hold:*

1. *The conditions 1 and 2 of Theorem 1 are fulfilled.*
2. *There exists a Lyapunov function $V \in C([t_0, \infty) \times \Delta, [0, \infty))$ such that*

    (i) for any solution of (7) $x(\cdot) \in \Delta, \ t \geq t_0$, the inequality

    $$({}^C_{t_0}\mathcal{D}^{q,\rho}V(\cdot, x(\cdot))(t) \leq g(t, V(t, x(t))), \ \ t > t_0 \tag{16}$$

    holds;
    (ii) $b(\|x\|) \leq V(t, x) \leq a(\|x\|)$ for $t \geq t_0, \ x \in \Delta$, where $a, b \in \mathcal{K}$.

3. *The zero solution of the scalar FrDE (8) is $(a(\lambda), b(A), T)$—practically quasi-stable, where the positive constants T, λ, A are given such that $\lambda < A, \ a(\lambda) < b(A), \ B(A) \subset \Delta$.*

Then the zero solution of the system of FrDE (7) is (λ, A, T)—practically quasi-stable.

Proof Let the triple (λ, A, T) be determined as in condition 4. Then, from condition 4, for any initial time $|\overline{u}_0| < a(\lambda)$, the inequality

$$|\overline{u}(t; t_0, \overline{u}_0)| < b(A) \text{ for } t \geq t_0 + T, \tag{17}$$

holds where $\overline{u}(t; t_0, \overline{u}_0)$ is a solution of (8) (with $\overline{u}(t_0) = \overline{u}_0$).

Choose the initial value $x_0 \in B(\lambda) \subset \Delta$ and let $x(t) = x(t; t_0, x_0) \in \Delta$ be the corresponding solution of (7). We will prove the inequality

$$\|x(t)\| < A \text{ for } t \geq t_0 + T \tag{18}$$

is true. Assume the contrary, i.e., there exists a point $\tau \geq t_0 + T$ such that $||x(\tau)|| \geq A$.

Let $u_0 = V(t_0, x_0)$. According to condition 2(ii) and the choice of x_0, we obtain $u_0 < a(\lambda)$. From Lemma 3 with $T = \tau$, we obtain

$$V(t, x(t; t_0, x_0)) \leq u^*(t; t_0, u_0) \text{ for } t \in [t_0, \tau]; \tag{19}$$

where $u^*(t; t_0, u_0)$ is the maximal solution of (8).

From inequality (19) and condition 2(ii), we get

$$b(A) \leq b(||x(\tau)||) \leq V(\tau, x(\tau)) \leq u^*(\tau; t_0, u_0) < b(A). \tag{20}$$

The obtained contradiction proves that inequality (18) is true. Therefore, the zero solution of FrDE (7) is (λ, A, T)—practically quasi-stable.

Theorem 3 *Suppose the following conditions hold:*
1. The conditions 1 and 2 of Theorem 1 and condition 2 of Theorem 2 are fulfilled.
2. The zero solution of scalar FrDE (8) is $(a(\lambda), b(A), b(K), T)$—strongly practically stable, where the positive constants T, λ, A, K are given such that $K < \lambda < A$, $b(K) < a(\lambda) < b(A)$, $B(A) \subset \Delta$.

Then the zero solution of the system of FrDE (7) is (λ, A, K, T)—strongly practically stable.

The proof of Theorem 3 is similar to that in Theorems 1 and 2, so we omit it.

Remark 4 Note, in the conditions of Theorems 1, 2, and 3, we could have $\Delta \equiv R^n$.

References

1. Agarwal R., Hristova S., O'Regan D.: Agarwal, R.; Hristova, S.; O'Regan, D. Stability of Generalized Proportional Caputo Fractional Differential Equations by Lyapunov Functions. Fractal Fract. 6, 34, (2022). https://doi.org/10.3390/fractalfract6010034
2. Almeida R., Agarwal R. P., Hristova S., O'Regan D.: Quadratic Lyapunov functions for stability of generalized proportional fractional differential equations with applications to neural networks. Axioms 10 (4), (2021) 322; https://doi.org/10.3390/axioms10040322
3. Bainov D., Dishliev A., Stamova I.: Practical stability of the solutions of impulsive systems of differential difference equations via the method of comparison and some applications to population dynamics. ANZIAM J. 43, (2002), 525–539.
4. Bernfeld S., Lakshmikantham V.: Practical stability and Lyapunov functions. Tohoku Math. J. 32, (1980), 607–613.
5. Henderson J., Hristova S.: Eventual practical stability and cone valued Lyapunov functions for differential equations with "Maxima". Commun. Appl. Anal. 14 (2010) 515–524.
6. Hristova S. G.: Practical stability and cone valued Lyapunov functions for differential equations with "maxima", Int. J. Pure Appl. Math. 57, 3, (2009), 313–323.
7. Hristova S.: Generalization of practical stability for delay differential equations with respect to initial time difference. AIP Conf. Proc. 1570, 313 (2013).
8. Hristova S., Abbas M.: Explicit solutions of initial value problems for fractional generalized proportional differential equations with and without impulses. Symmetry 2021, 13, (2021), 996.
9. Hristova S., Stefanova K.: Practical stability of impulsive differential equations with "supremum" by integral inequalities. Eur. J. Pure Appl. Math. 5, 1, (2012), 30–44.
10. Jarad F., Abdeljawad T., Alzabut J.: Generalized fractional derivatives generated by a class of local proportional derivatives. Eur. Phys. J. Spec. Top. 226, (2017), 3457–3471 https://doi.org/10.1140/epjst/e2018-00021-7
11. Kaczorek T.: Practical stability and asymptotic stability of positive fractional 2D linear systems. Asian J. Control. 12, 2, (2010) 200–207.
12. Lakshmikantham V., Leela S., Martynyuk A. A.: Practical Stability of Nonlinear Systems, World Scientific, Singapore, 1990.
13. Lakshmikantham V., Zhang Y.: Strict practical stability of delay differential equation. Appl. Math. Comput. 118, 2–3, (2001) 275–285.
14. La Salle J., Lefschetz S.: Stability by Lyapunov's Direct Method and Applications, Academic Press, New York, 1961.
15. Simon T.: Mittag-Leffler functions and complete monotonicity. Integral Transf. Special Functions. 26, 1, (2015) 36–50, https://doi.org/10.1080/10652469.2014.965704
16. Wang P., Sun W.: Practical Stability in terms of Two Measures for Set Differential Equations on Time Scales. The Scientific World Journal. 2014 (2014), Article ID 241034, 7 pages.

Nonlinear Evolution Inclusions with Causal Operators

Tzanko Donchev, Nikolay Kitanov, Alina I. Lazu, and Stanislav Stefanov

Abstract In this paper, we study nonlocal problems described by multivalued perturbations of m-dissipative evolution equations with multivalued terms depending on causal operators. The existence of the solution and some qualitative properties of the solution set are considered using the measure of noncompactness or dissipative type conditions.

Keywords m-Dissipative evolution equations · Causal operators

1 Introduction and Preliminaries

Let X be a real Banach space endowed with the norm $|\cdot|$, $A : D(A) \subset X \rightrightarrows X$ an m-dissipative operator, $F : I \times X \rightrightarrows X$ a multifunction with nonempty closed bounded values, where $I = [-\tau, T], \tau, T > 0$, and $Q : C(I, X) \to C(I, X)$ a causal operator.

In this paper, we study the following problem with the nonlocal initial condition:

$$\begin{cases} \dot{x}(t) \in Ax(t) + f_x(t), \ t \in (0, T) \\ f_x(t) \in F(t, Q(x)(t)) \\ x(s) = g(x(\cdot))(s), \ s \in [-\tau, 0], \end{cases} \tag{1}$$

T. Donchev · S. Stefanov
Department of Mathematics, UACG, 1 Hr. Smirnenski bvd, 1046 Sofia, Bulgaria
e-mail: stanislav.toshkov@abv.bg

N. Kitanov (✉)
Department of Mathematics, South West University, Blagoevgrad, Bulgaria
e-mail: nkitanov@abv.bg

Institute of Mathematics and Informatics - Bulgaria Academy of Sciences, Sofia, Bulgaria

A. I. Lazu
Department of Mathematics, Gh. Asachi Technical University, 700506 Iasi, Romania

where $g : C(I, X) \to C([-\tau, 0], X)$ is a given function such that $g(x(\cdot))(0) \in \overline{D(A)}$ for any $x(\cdot) \in C(I, X)$.

The theory of causal differential equations has the powerful quality of unifying ordinary differential equations, integro-differential equations, differential equations with finite or infinite delay, Volterra integral equations, and neutral equations. Causal operators have been introduced in [9]. The problems with causal operators include time lag systems [4, 7, 8, 15, 19] in local or nonlocal case. Causal differential equations are studied in [16, 17] in Banach space. We refer the reader to [3, 6, 11] for the theory of m-dissipative systems. The nonlocal multivalued perturbations of m-dissipative operators are studied among others in [2, 5, 20, 21]. The case of multivalued nonlocal semilinear problems with causal operators is considered in [1, 12, 13].

To our knowledge, the present article is the first one devoted to fully nonlinear evolution systems involving causal operators. We study the existence of solutions and qualitative properties of the solution set under compactness or dissipative type assumptions.

The compactness type assumptions are mainly in two directions.

(Cs) One is to assume that A generates a compact semigroup. In this case, the state space is assumed to be separable and one has to use some conditions on F, such as $F(\cdot, \cdot)$ has weakly compact values or the Banach space X is assumed to be reflexive. This approach is not used in the present paper.

(Mnc) The second one is to assume that A generates an equicontinuous semigroup, while $F(\cdot, \cdot)$ satisfies some measure of noncompactness (MNC) type condition. In this case, the solution set is nonempty and compact; however, one has to use additional assumptions such as X has uniformly convex dual or A is a sum of linear and continuous dissipative operators.

Another type of assumption is of dissipative type.

(L) Assume that $F(t, \cdot)$, Q, and g are Lipschitz. Imposing some conditions on the Lipschitz constants, one can prove the existence of solutions and some qualitative properties as continuous dependence of the initial conditions and relaxation theorem when the dual operator is single valued, which is the case when X^* is locally smooth.

To prove the existence of solutions for the nonlocal problem, one has to investigate first the corresponding local problem.

Let $f(\cdot)$ be a Bochner integrable function. A continuous function $x : [t_0, T] \to \overline{D(A)}$ is said to be an integral solution of

$$\begin{cases} \dot{x}(t) \in Ax(t) + f(t), \\ x(t_0) = x_0 \in \overline{D(A)} \end{cases} \tag{2}$$

if $x(t_0) = x_0$ and for every $u \in D(A)$, $v \in Au$, and $t_0 \leq s \leq t \leq T$, the following inequality holds:

$$|x(t) - u| \leq |x(s) - u| + \int_s^t [x(\tau) - u, f(\tau) + v]_+ d\tau. \tag{3}$$

The function $f(\cdot)$ in (2) will be called pseudoderivative of $x(\cdot)$ and will be denoted by $f_x(\cdot)$. Notice that $f_x(\cdot)$ depends on both $x(\cdot)$ and A. It is well known (cf. [3, 6]) that if $x(\cdot)$ and $y(\cdot)$ are the solutions of (2) with pseudoderivatives $f_x(\cdot)$ and $f_y(\cdot)$ and initial conditions x_0 and y_0, then

$$|x(t) - y(t)| \le |x_0 - y_0| + \int_0^t |f_x(s) - f_y(s)|ds. \tag{4}$$

The Hausdorff metric in the space of the closed bounded subsets of X is $D_H(A, B) := \max\{Ex(A, B), Ex(B, A)\}$, where $Ex(A, B) = \sup_{a \in A} \inf_{b \in B} |a - b|$. If $a \in X$ and $B \subset X$, then the distance between a and B is $dist(a, B) = \inf_{b \in B} |a - b|$. We denote by $\mathbb{B}$ the closed unit ball in X. The multifunction $G : I \times X \rightrightarrows X$ is said to be upper semicontinuous (USC) at (t, x) if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $D_H(G(t, x)), G(s, y)) < \varepsilon$ for any (s, y) with $|s - t| + |x - y| < \delta$. It is called almost USC if for any $\varepsilon > 0$ there exists a compact set $I_\varepsilon \subset I$ with Lebesgue measure $meas(I \setminus I_\varepsilon) < \varepsilon$ such that $G|_{I_\varepsilon \times X}$ is USC. The multifunction $G : I \times E \rightrightarrows E$ is said to be continuous if it is continuous with respect to the Hausdorff metric. It is called almost continuous if for any $\varepsilon > 0$ there exists a compact set $I_\varepsilon \subset I$ with Lebesgue measure $meas(I \setminus I_\varepsilon) < \varepsilon$ such that $G|_{I_\varepsilon \times E}$ is continuous.

Definition 1 An operator $Q : C(I, X) \to C(I, X)$ is called causal if for every $s \in [0, T]$ and any $u(\cdot), v(\cdot) \in C(I, X)$ such that $u(t) = v(t)$ when $t \in [0, s]$ one has $(Qu)(t) = (Qv)(t)$ for $t \in [0, s]$.

Clearly, this definition implies that the casual operator is similar to the delay operator x_t. However, $(Qu)(t)$ depends on the values of $u(\cdot)$ in the whole interval $[-\tau, t]$ in contrast with x_t which depends on the values of $x(\cdot)$ only on $[t - \tau, t]$. Of course, x_t is also a casual operator.

The Hausdorff measure of noncompactness of a nonempty bounded subset A of a Banach space E is defined by

$$\chi(A) = \inf\{\varepsilon > 0; \ A \text{ admits a finite cover by balls of radius } \le \varepsilon\}.$$

We recall some properties of χ.

(1) $\chi(A) = 0$ if and only if $\overline{A}$ is compact.
(2) If $\{A_n\}_{n \ge 1}$ is a decreasing sequence of bounded closed nonempty subsets of X and $\lim_{n \to \infty} \chi(A_n) = 0$, then $\bigcap_{n=1}^{\infty} A_n$ is nonempty and compact.

We will use also sequential MNC, generated by $\chi(\cdot)$, defined by

$$\chi_0(A) = \sup\{\chi(\{x_n : \ n \ge 1\}); \ (x_n) \text{ is a sequence in } A\}.$$

If E is separable, then $\chi_0(\Omega) = \chi(\Omega)$. In arbitrary Banach space E, we have

$$\chi_0(\Omega) \leq \chi(\Omega) \leq 2\chi_0(\Omega), \tag{5}$$

for every bounded $\Omega \subset X$.

For more information about MNC, we refer the reader to [14, 18].

2 Compactness Type Conditions

In this section, X is a Banach space with uniformly convex dual X^*. We introduce the following hypotheses:

(A) A is m-dissipative and generates an equicontinuous semigroup $S(t) : \overline{D(A)} \to X, t \geq 0$.

Notice that we can assume, without loss of generality, that $0 \in D(A)$ and $0 \in A0$.

(F1) $F(\cdot, \cdot)$ is almost upper semicontinuous with nonempty closed convex bounded values.

(F2) There exist two Lebesgue integrable functions $a(\cdot)$ and $b(\cdot)$ with positive values such that $\|F(t, \alpha)\| := \max_{v \in F(t,\alpha)} |v| \leq a(t) + b(t)|\alpha|$ for a.a. $t \in I$ and any $\alpha \in X$.

(F3) There exists a Lebesgue integrable function $\lambda(\cdot)$ such that $\chi(F(t, B)) \leq \lambda(t)\chi(B)$, for any nonempty bounded set $B \subset X$ and for a.a. $t \in I$.

(g1) There exists a constant $K > 0$ such that $\|g(x(\cdot)) - g(y(\cdot))\| \leq K\|x(\cdot) - y(\cdot)\|$ for any $x(\cdot), y(\cdot) \in C([-\tau, T], \overline{D(A)})$.

(g2) $g(\cdot)$ is a continuous compact map and there exist two constants $p, q > 0$ such that $\|g(u(\cdot))\| \leq p\|u(\cdot)\| + q$ for any $u(\cdot) \in C([-\tau, T], \overline{D(A)})$.

(Q1) Q is a continuous causal operator and there exists $L > 0$ such that $\|Qu(\cdot)\| \leq L\|u(\cdot)\|$ for any $u(\cdot) \in C([-\tau, T], \overline{D(A)})$.

(Q2) There exists a constant $R > 0$ with $\chi(QV) \leq R\chi(V)$ for each bounded set $V \subset C([-\tau, T], X)$.

Denote by $P = \max_{t \in I} \|S(t)\|$. We need also the following assumption:

(L) $K + R \displaystyle\int_0^T \lambda(t)dt < 1 - PK$.

Lemma 1 ([20, Theorem 2.1]) *If A generates an equicontinuous semigroup, $B \subset L^1([0, T], X)$ is uniformly integrable, and $C \subset \overline{D(A)}$ is compact, then the set of integral solutions of (2) for some $f_x \in B$ and some $x_0 \in C$ is bounded and equicontinuous.*

Denote $\alpha = \displaystyle\int_0^T a(t)dt$ and $\beta = \displaystyle\int_0^T b(t)dt$. Further, in this paper, we assume that $p + L\beta < 1$.

Lemma 2 *There exists a constant $M > 0$ such that for any $z(\cdot) \in C([-\tau, T], X)$ with $\|z(\cdot)\| \leq M$ we have $\|x(\cdot)\| \leq M$ for any solution $x(\cdot)$ of*

$$\begin{cases} \dot{x}(t) \in Ax(t) + f_x(t), \ t \in (0, T) \\ f_x(t) \in F(t, Q(z)(t)) \\ x(s) = g(z(\cdot))(s), \ s \in [-\tau, 0]. \end{cases} \quad (6)$$

Proof Let $M := \dfrac{q + \alpha}{1 - p - L\beta}$. Take $z(\cdot) \in C([-\tau, T], X)$ with $\|z(\cdot)\| \leq M$ and let $x(\cdot)$ be a solution of (6). Then, due to **(g2)**, **(F2)**, and **(Q1)**,

$$|x(t) - 0| \leq p\|z(\cdot)\| + q + \int_0^t [a(s) + Lb(s)\|z(\cdot)\|]ds \quad (7)$$

$$\leq p\|z(\cdot)\| + q + \int_0^t a(s)ds + L\|z(\cdot)\| \int_0^t b(s)ds \quad (8)$$

$$= q + \alpha + (p + L\beta)\|z(\cdot)\|, \quad (9)$$

which leads to $\|x(\cdot)\| \leq M$.

We will use the following theorem proved in [10].

Theorem 1 *Let Θ be a convex compact subset of a Banach space. If $\Psi : \Theta \rightrightarrows \Theta$ is with closed graph and compact contractive values, then there exists a fixed point $z \in \Psi(z)$.*

The main existence results with MNC are the following two theorems.

Theorem 2 *Under the conditions (A), (F1)–(F3), (g2), (Q1), (Q2) there exists a solution of the nonlocal problem (1). Furthermore, the solution set is compact in $C(I, X)$.*

Proof Consider the map $sol : C(I, X) \rightrightarrows C(I, X)$, where $sol(z)$ is the solution set of the problem (6) for any $z(\cdot) \in C(I, X)$.

I. First, we prove that there exists a nonempty convex compact set $\mathcal{M} \subset C([-\tau, T], X)$ such that $sol : \mathcal{M} \rightrightarrows \mathcal{M}$.

To this end, let $\mathcal{M}_1 = \overline{co} \bigcup_{\substack{z(\cdot) \in C(I, X) \\ \|z(\cdot)\| \leq M}} \{x(\cdot); \ x(\cdot) \in sol(z)\}$.

We define inductively the sequence of sets $(\mathcal{M}_n)$ as follows:

$$\mathcal{M}_{n+1} = \overline{co} \bigcup_{z(\cdot) \in \mathcal{M}_n} \{x(\cdot); \ x(\cdot) \in sol(z)\}.$$

Then $\mathcal{M}_{n+1} \subset \mathcal{M}_n$, i.e., $\chi(\mathcal{M}_{n+1}) \leq \chi(\mathcal{M}_n)$. Let $\mathcal{M}_n(t) = \bigcup_{x(\cdot) \in \mathcal{M}_n} \{x(t)\}$, i.e., the reachable set. Thus, $\chi(\mathcal{M}_{n+1}(t)) \leq \chi(\mathcal{M}_n(t))$ for every $t \in I$.

From inequality (5), for any $\varepsilon > 0$ there exists a sequence $\{u_k\}_{k \geq 1} \subset \mathcal{M}_n$ and a sequence $f_k(\cdot) \in L^1([-\tau, T], X)$ such that $f_k(t) \in F(t, Q(u_k)(t))$ and, in view of [6, Lemma 3.7, p. 39] (see also [21, Lemma 3.3]), using **(F3)**, **(Q2)**, we have

$$\chi(\mathcal{M}_{n+1}(t)) \leq \chi(\{W(u_k, f_k)(t); \ k \geq 1\}) + \varepsilon \leq 2R \int_0^t \lambda(t)\chi(\mathcal{M}_n(s))ds + \varepsilon,$$

where $W(u_k, f_k)$ denotes the solution set of (6) with u_k instead of z and f_k instead of f_x.

Since $\varepsilon > 0$ is arbitrary, $\chi(\mathcal{M}_{n+1}(t) \leq 2R \int_0^t \lambda(t)\chi(\mathcal{M}_n(s))ds$. Let $\beta(t) :=$ $\lim_{n \to \infty} \chi(\mathcal{M}_n(t))$, then $\beta(t) \leq 2R \int_0^t \lambda(s)\beta(s)ds$. Due to Gronwal's inequality, $\beta(t) = 0$. Lemma 1 together with Arzela's theorem imply that $\bigcap_{n \geq 1} \mathcal{M}_n = \mathcal{M}$ is nonempty and compact.

II. We prove that the multifunction $z \mapsto sol(z)$ has a closed graph.

To this aim, let $y_n \to y$ uniformly on $[-\tau, T]$ and let $x_n(\cdot) \in sol(y_n)$. Since $\{x_n(\cdot)\}_{n=1}^\infty \subset \mathcal{M}$ and the latter is nonempty convex and compact, passing to subsequences, $x_n(\cdot) \to x(\cdot)$ uniformly on $[-\tau, T]$. If $f_n(\cdot)$ are the corresponding pseudoderivatives, then $f_n(t) \in \mathcal{M}_1$ for every $t \in [0, T]$. Furthermore, $\mathcal{M}_1$ is weakly compact and Diestel's criterion tells us that $\{f_n(\cdot)\}_{n \geq 1}$ is weakly $L^1([0, T], X)$-compact. Then, passing to subsequences, $f_n(\cdot) \rightharpoonup f(\cdot)$ L^1-weakly. Since $F(\cdot, \cdot)$ is USC with convex and compact values, we get that $x(\cdot)$ is a solution of (6).

III. Now, we prove that $sol(z)$ is compact contractible for any $z \in \mathcal{M}$.

Since $sol(z) \subset \mathcal{M}$, one has that it is compact when it is closed. Then $sol(z)$ is compact thanks to **II**.

Let $z(\cdot) \in \mathcal{M}$ and let $x(\cdot) \in sol(z)$ have pseudoderivative $f_x(\cdot)$. We define the homotopy $h : [0, 1] \times sol(z) \to sol(z)$ as follows:

$$h(s, x)(t) = \begin{cases} x(t), & t \in [0, sT] \\ y_x(t, sT, y(sT)), & t \in [sT, T], \end{cases}$$

where $\dot{y}_x(t) \in Ay_x(t) + \bar{f}_y(t)$, $y_0(s) = y(z(\cdot))(s)$. Moreover,

$$\bar{f}_y(t) = \begin{cases} f_x(t), & t \in [0, sT] \\ f_y(t), & t \in [sT, T] \end{cases}$$

for any $y \in sol(z)$. Clearly, $h(0, x)(t) = y(t)$ and $h(1, x)(t) = x(t)$. Notice that the solution set of (6) is continuous on the initial condition, because $g(\cdot)$ and $Q(\cdot)$ are continuous. Therefore, $h(\cdot, x)$ is continuous and hence sol has contractible values.

Finally, we apply Theorem 1 to complete the proof.

Theorem 3 *Let X be separable. Under the conditions (A), (F1)–(F3), (g1), (Q2), (L) there exists a solution of the nonlocal problem (1). Furthermore, the solution set is compact in $C(I, X)$.*

Proof The proof follows similar lines as the proof of the previous theorem. Denote by $Sol(z)$ the solution set of the evolution inclusion

$$\begin{cases} \dot{x} \in Ax + F(t, z(t)), \\ x_0(s) = g(z(\cdot))(s), \ s \in [-\tau, 0] \end{cases}$$

for any $z \in C([-\tau, T], X)$.

It follows from Lemma 2 that there exists a closed convex bounded subset $\mathbb{U}$ of $\overline{D(A)}$ such that $Sol : \mathbb{U} \rightrightarrows \mathbb{U}$.

Denote by $W_0 = Sol(z(\cdot))$, $z(\cdot) \in \mathbb{U}$. By induction, we define $W_{n+1} = \overline{co} \, Sol(z)$, $z \in W_n$. Clearly, $W_{n+1} \subset W_n$. Denote $W_n(t)$ the corresponding reachable set. Since X is separable, one has that $\chi_0(B) = \chi(B)$ for every bounded set B. Therefore, $\chi(W_{n+1}) = \chi_0(W_{n+1})$. Clearly,

$$\chi(W_{n+1}(t)) \leq K \sup_{t \in [0,T]} \chi(W_n(t)) + R \int_0^t \lambda(s)\chi(W_n(s))ds.$$

Thus,

$$\sup_{t \in [0,T]} \chi(W_{n+1}(t)) \leq K \sup_{t \in [0,T]} \chi(W_n) + R \sup_{t \in [0,T]} \chi(W_n(t)) \int_0^t \lambda(s)ds + \varepsilon$$

$$= \left(K + R \int_0^T \lambda(t)dt \right) \sup_{t \in [0,T]} \chi(W_n(t)).$$

Due to **(L)**, $\lim_{n \to \infty} \sup_{t \in [0,T]} \chi(W_n(t)) = 0$. The latter implies that, for any $t \in [0, T]$, there exists $\lim_{n \to \infty} W_n(t) = W(t)$. Due to Gronwall's inequality, $\chi(W(t)) = 0$ for any $t \in [0, T]$.

Therefore, the solution set of (1) $sol(z) : z_0 \in W(0)$ is equicontinuous. Therefore, $W = \lim_{n \to \infty} W_n$ exists and it is nonempty convex and compact set.

The rest of the proof is the same as in the proof of the previous theorem.

3 Lipschitz Type Conditions

In this section, X is an arbitrary Banach space and we assume that the right-hand side and the initial function are Lipschitz continuous. Moreover, we don't assume that F has convex strongly compact values, an assumption used in the MNC case. However, in this case, the continuity assumptions on F are stronger.

We need the following assumptions:

(FF1) $F(\cdot, \cdot)$ is almost continuous with nonempty closed bounded values. Moreover, $\|F(t, 0)\| \leq a(t)$ for some L^1 positive function $a(\cdot)$. If X is separable, then we can drop almost the continuity of F and replace it with $F(\cdot, \alpha)$ measurable.

(FF2) There exists a positive function $\mu(\cdot)$ such that

$$D_H(F(t, (Qx)(t)), F(t, (Qy)(t)) \leq \mu(t)\|(Qx)(t) - (Qy)(t)\|,$$

for any $x(\cdot), y(\cdot) \in C([-\tau, T], X)$.

(Q3) There exists a constant $N > 0$ such that $|(Qx)(t) - (Qy)(t)| \leq N|x(t) - y(t)|$.

(C) $K + \exp\left(N \int_0^T \mu(t)dt\right) := \alpha < 1$.

Then, the following lemma holds.

Lemma 3 *Let $x(\cdot) \in C([-\tau, T], X)$ and let $f_x(t) \in F(t, Q(x)(t))$ be measurable. Then, under (FF1), (FF2), (Q3), for any $y(\cdot) \in C([-\tau, T], X)$ and any $\varepsilon > 0$ there exists $f_y(t) \in F(t, (Qy)(t))$ such that $|f_x(t) - f_y(t)| < N\mu(t)\|x(\cdot) - y(\cdot)\| + \varepsilon$.*

The proof is almost the same as the proof of Proposition 2.1 of [2] and it is omitted.

Theorem 4 *Under the assumptions (FF1), (FF2), (g1), (Q3), and (C), the problem (1) has a solution.*

Proof We will use successive approximations. Let $x^0(\cdot) \in C(I, X)$. We define $x^1(\cdot)$ to be a solution of

$$\begin{cases} \dot{x}^1(t) \in Ax^1(t) + f_1(t) \\ f_1(t) \in F(t, (Qx^0)(t)), \\ x^1(s) = g(x^0(\cdot))(s), \ s \in [-\tau, 0]. \end{cases}$$

For $\delta_n > 0$, we define successively $x^{n+1}(\cdot)$ to be the solution of

$$\begin{cases} \dot{x}^{n+1}(t) \in Ax^{n+1}(t) + f_{n+1}(t) \\ x^{n+1}(s) = x^n(s), \ s \in [-\tau, 0]. \end{cases}$$

Due to Lemma 3, we can choose $f_{n+1}(t) \in F(t, (Qx^n)(t))$ to satisfy

$$|f_{n+1}(t) - f_n(t)| < N\mu(t)\|x^n(\cdot) - x^{n-1}(\cdot)\| + \delta_n. \tag{10}$$

It follows from (4) that

$$|x^{n+1}(t) - x^n(t)| \leq |x^{n+1}(0) - x^n(0)| + \int_0^t |f_{n+1}(s) - f_n(s)|ds \tag{11}$$

$$< |(g(x^n(\cdot))(0) - g(x^{n-1}(\cdot))(0)| + \int_0^t (N\mu(s)\|x^n(\cdot) - x^{n-1}(\cdot)\| + \delta_n)ds \tag{12}$$

$$\leq \left(K + N \int_0^T \mu(s)ds\right)\|x^n(\cdot) - x^{n-1}(\cdot)\| + \delta_n T. \tag{13}$$

Clearly, one can choose δ_n such that $\delta_n T < \dfrac{1-\alpha}{2}$. Consequently,

$$\|x^{n+1}(\cdot) - x^n(\cdot)\| \le \frac{1+\alpha}{2}\|x^n(\cdot) - x^{n-1}(\cdot)\|$$

and hence $x^n(\cdot) \to x(\cdot)$ uniformly on I. Further, from (10), we get that $f_n(t) \to f(t)$ in $L^1([0,T])$-strongly. It is easy to show that $\dot{x}(t) \in Ax(t) + f(t)$ and hence $x(\cdot)$ is a solution of (1).

References

1. Agarwal, R., Arshad, S., Lupulescu, V., O'Regan, D.: Evolution equations with causal operators. Differ. Equ. Appl. 7, 15–26 (2015).
2. Ahmed, R., Donchev, T., Lazu, A.I.: Nonlocal m-dissipative evolution inclusions in general Banach spaces. Mediterr. J. Math. 14:215, https://doi.org/10.1007/s00009-017-1016-5 (2017).
3. Barbu, V.: Nonlinear differential equations of monotone types in Banach spaces. Springer, New York (2010).
4. Benchohra, M., Abbas, S.: Advanced Functional Differential Equations and Inclusions. Springer (2015).
5. Bilal, S., Cârjă, O., Donchev, T., Javaid, N., Lazu A.I.: Nonlocal evolution inclusions under weak conditions. Advances in Difference Equations, 2018: 399, https://doi.org/10.1186/s13662-018-1858-6 (2018).
6. Bothe, D.: Nonlinear Evolutions in Banach Spaces. Habilitationsschritt, Paderborn (1999).
7. Burlica, M., Rosu, D.: A class of nonlocal evolution equations with nonlocal initial conditions. Proc. AMS 142, 2445–2458 (2014).
8. Burlică, M., Necula, M., Roşu, D., Vrabie, I.I.: Delay Differential Evolutions Subjected to Nonlocal Initial Conditions. Monographs and Research Notes in Mathematics. CRC Press, New York (2016).
9. Corduneanu, C.: Functional equations with causal operators
10. Gorniewicz, L., Granas, A., Kryszewski, M.: Sur la methode l'homotopie dans la theorie des points fixes pour applications multivoques, II L'indice dans les ANR-s compacts. C.R.. Acad. Sci. Paris 308(14), 449–452 (1989).
11. Ito, K., Kappel, F.: Evolution Equations and Approximations. World Scientific, Singapore (2002).
12. Jabeen, T., Lupulescu, V.: Existence of mild solutions for a class of non-autonomous evolution equations with nonlocal initial conditions. J. Nonlinear Sci. Appl. 10, 141–153 (2017).
13. Jabeen, T., Agarwal, R., Lupulescu, V., O'Regan, D.: Impulsive evolution equations with causal operators. Symmetry 12, no. 1: 48, (2020).
14. Kamenskii, M.I., Obukhovskii, V., Zecca, P.: Condensing Multivalued Maps and Semilinear Differential Inclusions in Banach spaces. Walter de Gruyter, Berlin, New York (2001).
15. Ke, T.: Cauchy problems for functional evolution inclusions involve accretive operators. EJQTDE 75, 1–13 (2013).
16. Lakshmikantham, V., Leela, S., Drici, Z., McRae, F.: Theory of causal differential equations. Atlantic Press, Paris (2009).
17. Lupulescu, V.: Causal functional differential equations in Banach spaces. Nonlinear Anal. 69, 4787–5795 (2008).

18. Toledano, J., Benavides, T., Azedo, G.: Measures of Noncompactness in Metric Fixed Point Theory, Birkhauser, Basel, (1997).
19. Vrabie, I.I.: Existence in the large for nonlinear delay evolution inclusions with nonlocal initial conditions. J. Funct. Anal. 262, 1363–1391 (2012).
20. Xue, X.: Nonlinear differential equations with nonlocal conditions in Banach spaces. Nonlinear Anal. 63, 2850–2853 (2009).
21. Zhu, L., Hiang, Q., Li, G.: Existence and asymptotic properties of solutions of nonlinear multivalued differential inclusions with nonlocal conditions. J. Math. Anal. Appl. 390, 523–534 (2012).

Non-positivity of Operators Gs,n+

Teodora Zapryanova

Abstract The paper gives an example of a function, which is positive on $[-1, 1]$, but it's image by the operator G^+s,n is not positive on $[-1, 1]$.

Keywords K-functional · Moduli of functions · Linear operators

1 Introduction

In a recent years, we studied the rate of approximation by the operators $\mathbf{G}_{s,n}$, $\mathbf{G}_{s,n}^*$, $\mathbf{G}_{s,n}^+$ (see [7–9]). We established equivalence between the error of approximation by these operators and an appropriate Peetre K-functionals. Let us briefly list some of these results.

In [7], the following K-functional was introduced:

$$K\left(f, t; C[-1, 1], C^2, H\right) = \inf\left\{\|f - g\| + t\|Hg\| : g \in C^2\right\},$$

where the differential operator

$$H_1(g(x)) := \left(1 - x^2\right)^{\frac{1}{2}} \frac{d}{dx}(g(x)),$$
$$H(f(x)) := (H_1)^2(f(x)) = \left(1 - x^2\right)f''(x) - xf'(x).$$

Let $f \in C[-1, 1], n \in N, s \in N$

$$G_{s,n}(f, x) := \int_{-\pi}^{\pi} f(\cos(\arccos x + v)) K_{s,n}(v) dv,$$

T. Zapryanova (✉)
University of Economics-Varna, 77, Kniaz Boris I Blvd, 9000 Varna, Bulgaria
e-mail: teodorazap@ue-varna.bg

$$K_{s,n}(v) = \mu_{s,n} \left(\frac{\sin(nv/2)}{\sin(v/2)} \right)^{2s} = \frac{1}{2} + \sum_{k=1}^{sn-s} \rho_{k,sn-s} \cos kv.$$

$K_{s,n}(v)$ is even, non-negative trigonometric polynomial of degree less or equal to $m(n) = sn - s$, where $\mu_{s,n}$ is chosen, such that $\int_{-\pi}^{\pi} K_{s,n}(v)dv = 1$ (see [3]).

It was proved in the following.

Theorem A (see [7]) *For $f \in \mathbf{C}[-1, 1]$ and $s \geq 3$, we have*

$$\left\| G_{s,n} f - f \right\| \sim K(f, \frac{1}{n^2}; C[-1, 1], C^2, H), \quad n \in \mathbf{N}.$$

The notation $\theta_1(f, t) \sim \theta_2(f, t)$ means that there exists a positive constant independent of f and t such that

$$C^{-1}\theta_1(f, t) \leq \theta_2(f, t) \leq C\theta_1(f, t).$$

Above and throughout C denotes a positive constant, not necessarily the same at each occurrence, which is independent of the function f and the parameter n (or t).

In [2], Badea generalized to a large class of continuous linear operators T on $C[0, 1]$, the convergence results of the iterates of Bernstein operator B_m toward L-operator of linear interpolation at the endpoints of the interval $[0, 1]$. The conditions required for T include preserving linear function. In contrast, our result in [10] gives similar method based on spectral theory for operators preserving only the constant functions. In this case, the limit of the iterates $G_{s,n}^m$ of the Jackson-type operator $G_{s,n}$ is toward a suitable linear functional $P : C[-1, 1] \rightarrow C[-1, 1]$, defined by

$$Pf = \frac{1}{\pi} \int_{-1}^{1} \frac{f(x)}{\sqrt{1-x^2}} dx = \frac{1}{\pi} \int_{0}^{\pi} f(\cos t)dt, \quad \text{where } x = \cos t.$$

We proved the following.

Theorem B (see [10]) *With $\gamma_0 := \max_{1 \leq j \leq m(n)} \left\{ \left| \rho_{j, m(n)} \right| \right\}$ we have*

$$\left\| G_{s,n}^m - P \right\| \leq C\gamma_0^m$$

for some suitable positive constant C and therefore

$$\lim_{m \to \infty} \left\| G_{s,n}^m - P \right\| = 0.$$

Theorem B gave an answer to the open problem of Gonska on the limit of iterates of Jackson-type operator proposed in 2013 during the conference Constructive Theory of Function.

"Korovkin-type theorems" furnish simple and useful tools for ascertaining whether a given sequence of positive linear operators converges to the identity operator. In 1953, Korovkin considered a test monomial functions $1, x, and x^2$ which guarantee that the approximation property holds on the whole space provided it holds on them (see, e.g., [1, p. 218]). Indeed let $L_n : C[0, 1] \to C[0, 1]$ be sequence of positive linear operators and $e_\lambda(x) = x^\lambda, x \in [0, 1], \lambda \geq 0 (e_0 \equiv 1)$. Then

$$L_n(e_0) = e_0, L_n(e_1) = e_1, L_n(e_2) \to e_2 \Rightarrow L_n(f) \to f, \forall f \in C[0, 1]$$

and

$$L_n(e_0) = e_0, L_n(e_2) = e_2, L_n(e_1) \to e_1 \Rightarrow L_n(f) \to f, \forall f \in C[0, 1].$$

However, attempts to find positive linear polynomial operators $L_n : C[0, 1] \to C[0, 1]$ such that

$$L_n(e_1) = e_1 \ and \ L_n(e_2) = e_2$$

failed. In this direction, Gavrea and Ivan [5] proved that there is no positive linear analytic operator $L : C[0, 1] \to C[0, 1]$ preserving $e_i \ and \ e_j, 0 < i < j$. It is the positivity property that keeps analytic operators away from possessing more monomial fixed points (see Remark 2 in [5]).

2   Operator $G^+_{s, n}$

Let Lf be a linear operator that interpolates f at -1 and 1

$$L(f, x) := \frac{1}{2} f(1)(x + 1) + \frac{1}{2} f(-1)(1 - x), \quad -1 \leq x \leq 1. \tag{1}$$

We note that the uniform norm of the operator L is 1. We consider the sequence of operators $G^+_{s, n} : C[-1, 1] \to \Pi_{sn-s}$, introduced in [3]

$$\begin{aligned} G^+_{s, n}(f, x) &= G_{s, n}(f, x) + L(f, x) - L(G_{s, n}f, x) \\ &= G_{s, n}(f, x) + L\big((I - G_{s, n})f, x\big). \end{aligned} \tag{2}$$

In contrast to the operator $G_{s, n}$ the operator $G^+_{s, n}$ is not positive. We give an example of a function that illustrates this. Let us recall that the operator Φ is said to be positive if $\Phi(f) \geq 0$ for any non-negative function f on $[a, b]$. Let $\| f \| := \max\{|f(x)| : -1 \leq x \leq 1\}$ for $f \in C[-1, 1]$. We consider for $f \in C[-1, 1]$ the K-functional

$$K^+\left(f, \frac{1}{n^2}; C[-1, 1], C^2, H(I-L)\right)$$

$$:= \inf\left\{\|f-g\| + \frac{1}{n^2}\|(I-L)Hg\| \; : \; g \in C^2\right\}.$$

In [8] we proved direct and strong converse inequality of type A in the terminology of [4]. More precisely, our result is given in the following

Theorem C *If* $f \in C[-1, 1]$ *and* $s \geq 3$, *then for the operator* $G_{s,n}^+$ *and* K-*functional* K^+ *given above, we have*

$$\|G_{s,n}^+ f - f\| \sim K^+\left(f, \tfrac{1}{n^2}; C[-1, 1], C^2, H(I-L)\right), \quad n \in \mathbf{N}, n \geq n_0.$$
We shall use

Lemma D $\mu_{s,n} = \left[\int\limits_{-\pi}^{\pi} \left(\frac{\sin(nv/2)}{\sin(v/2)}\right)^{2s} dv\right]^{-1} \leq c(s)\, n^{-2s+1}$

$c(s)$ is a constant depending on s.

$$\int\limits_{-\pi}^{\pi} \left(\frac{\sin(nv/2)}{\sin(v/2)}\right)^{2s} dv \geq \int\limits_{0}^{\pi/n} \left(\frac{\sin(nv/2)}{\sin(v/2)}\right)^{2s} dv$$

Proof: $\left(as \quad \frac{2}{\pi}v \leq \sin v \leq v \quad for \quad 0 \leq v \leq \frac{\pi}{2}\right)$

$$\geq \int\limits_{0}^{\pi/n} \left(\frac{\frac{2}{\pi}\frac{nv}{2}}{\frac{v}{2}}\right)^{2s} dv = \left(\frac{2}{\pi}\right)^{2s} n^{2s}\frac{\pi}{n}.$$

Hence $\mu_{s,n} \leq c(s)n^{-2s+1}$. The lemma is proved. $\qquad\square$

3   An Example of a Function

Assertion 3.1 *Let* s *and* n *be natural numbers,* $s \geq 2$. *There exists a function* $f \in C[-1, 1]$ *such that* $f(x) \geq 0$ *for every* $x \in [-1, 1]$, *but* $(G_{s,n}^+ f)(x_0) < 0$ *for some* $x_0 \in [-1, 1]$.

Proof Let $f(x)$ satisfy the following conditions:

(1) $f(x) \in C^\infty[-1, 1]$;
(2) $f(x) \geq 0$, for $x \in [-1, 1]$;
(3) $f(x) \equiv 0$ in some neighborhood of zero;

(4) $f'(-1) > 0$ *and* $f'(1) < 0$.

Recall that according to the definition $f'(-1) > 0$ *and* $f'(-1) < 0$ $\Leftrightarrow$ $(Hf)(\pm 1) > 0$.

$f(x) \geq 0$ for every $x \in [-1, 1]$, but we will show $(G_{s,n}^+ f)(0) < 0$.

We define

$$f(\cos(\arccos x + v)) := f(\cos(z + v)) = \tilde{f}(z + v), \, z = \arccos x.$$

We note

$$H_1 f(\cos(\arccos x + v)) = -\tilde{f}'(z + v)$$

$$Hf(\cos(\arccos x + v)) = \tilde{f}''(z + v).$$

Expanding $\tilde{f}(z + v)$ *by Taylor's formula with integral remainder*

$$f(\cos(\arccos x + v)) - f(x)$$

$$= \tilde{f}(z + v) - \tilde{f}(z) = \tilde{f}'(z)v + \int_z^{z+v} \tilde{f}''(t)(z + v - t)dt.$$

Then

$$\left(I - G_{s,n}\right)(f, x) = -\int_{-\pi}^{\pi} [f(\cos(\arccos x + v)) - f(x)]K_{s,n}(v)dv$$

$$= -\int_{-\pi}^{\pi} \left[\tilde{f}'(z)v + \int_z^{z+v} \tilde{f}''(t)(z + v - t)dt\right]K_{s,n}(v)dv \qquad (3)$$

$$= -\int_{-\pi}^{\pi} \int_z^{z+v} \tilde{f}''(t)(z + v - t)dt\, K_{s,n}(v)dv.$$

We used that

$$\int_{-\pi}^{\pi} vK_{s,n}(v)dv = 0.$$

Furthermore, for sufficiently small positive number δ, we have

$$\int\limits_{|v|>\delta} [f(\cos(\arccos x + v)) - f(x)]K_{s,n}(v)dv$$

$$\left(as \left(\frac{\sin(nv/2)}{\sin(v/2)} \right)^{2s} \ is\ bounded\ for\ |v| > \delta \right)$$

$$\leq c\mu_{s,n} = c \left[\int\limits_{-\pi}^{\pi} \left(\frac{\sin(nv/2)}{\sin(v/2)} \right)^{2s} dv \right]^{-1} = O\left(n^{-2s+1}\right)\ for\ s \geq 2.$$

Then by (3), we get

$$\left(I - G_{s,n}\right)(f, x) = - \int\limits_{-\delta}^{\delta} \int\limits_{z}^{z+v} \tilde{f}''(t)(z + v - t)dt\, K_{s,n}(v)dv + O\left(n^{-2s+1}\right). \quad (4)$$

Now we set $x = 1$ in (4) and obtain

$$\left(I - G_{s,n}\right)(f, 1) = - \int\limits_{-\delta}^{\delta} \int\limits_{0}^{v} \tilde{f}''(t)(v - t)dt\, K_{s,n}(v)dv + O\left(n^{-2s+1}\right) \quad (5)$$

$$As\ for\ t\ close\ to\ 0\ \ Hf(\cos t) = \tilde{f}''(t) > c > 0,$$

by (5) get

$$\left(I - G_{s,n}\right)(f, 1) \leq -c \int\limits_{-\delta}^{\delta} \int\limits_{0}^{v} (v - t)dt\, K_{s,n}(v)dv + O\left(n^{-2s+1}\right)$$

$$= -c \int\limits_{0}^{\delta} v^2 K_{s,n}(v)dv + O\left(n^{-2s+1}\right)$$

$$= -c \int\limits_{0}^{\pi} v^2 K_{s,n}(v)dv + O\left(n^{-2s+1}\right) \sim -n^{-2},\ \ for\ s \geq 2.$$

We used that (see [6, p. 57]) $\int\limits_{0}^{\pi} v^2 K_{s,n}(v)dv \sim n^{-2}\ for\ s \geq 2$.

Thus for

$$s \geq 2$$

$$\left(I - G_{s,n}\right)(f, 1) \leq -cn^{-2}. \quad (6)$$

Now we set $x = -1$ in (4) and obtain

$$\left(I - G_{s,n}\right)(f, -1) = -\int_{-\delta}^{\delta} \int_{\pi}^{\pi+v} \tilde{f}''(t)(\pi + v - t)dt\, K_{s,n}(v)dv + O\left(n^{-2s+1}\right) \quad (7)$$

$$Hf(\cos t) = \tilde{f}''(t) > c > 0.$$

As for t close to π similarly by (7), we get

$$\left(I - G_{s,n}\right)(f, -1) \leq -c \int_{-\delta}^{\delta} \int_{\pi}^{\pi+v} (\pi + v - t)dt\, K_{s,n}(v)dv + O\left(n^{-2s+1}\right)$$

$$= -c \int_{0}^{\delta} v^2 K_{s,n}(v)dv + O\left(n^{-2s+1}\right)$$

$$= -c \int_{0}^{\pi} v^2 K_{s,n}(v)dv + O\left(n^{-2s+1}\right) \sim -n^{-2}, \quad for\ s \geq 2.$$

$$We\ used\ again\ that\ \int_{0}^{\pi} v^2 K_{s,n}(v)dv \sim n^{-2} \quad for\ s \geq 2.$$

Thus for

$$s \geq 2,$$

we have

$$\left(I - G_{s,n}\right)(f, -1) \leq -cn^{-2}. \quad (8)$$

Using (1), (6) and (8), for $s \geq 2$ we have

$$L\left(\left(I - G_{s,n}\right)f, x\right) \leq -cn^{-2}. \quad (9)$$

We now use

$$G_{s,n}(f, 0) = \int_{-\pi}^{\pi} f(\cos(\arccos 0 + v)) K_{s,n}(v)$$

$$= \int_{-\pi}^{\pi} f\left(\cos\left(\frac{\pi}{2} + v\right)\right) K_{s,n}(v)dv = \int_{-\pi}^{\pi} f(-\sin v) K_{s,n}(v)dv. \quad (10)$$

$$f(-\sin v) \equiv 0,$$

by the choice of function f (see Condition 3)

for v close to 0. According to (10) and the definition of $K_{s,n}$ (as $\left(\frac{\sin(nv/2)}{\sin(v/2)}\right)^{2s}$ is bounded for $|v| > \delta$), we have

$$G_{s,n}(f, 0) \leq c\mu_{s,n} \int\limits_{-\pi}^{\pi} f(-\sin v)dv \leq c\mu_{s,n} \leq cn^{-2s+1}.$$

Using the above inequality and (9), we obtain

$$G_{s,n}^{+}(f, 0) = G_{s,n}(f, 0) + L\big((I - G_{s,n})f, 0\big) \leq cn^{-2s+1} - cn^{-2} < 0, \quad for \ s \geq 2$$

as the constants c are positive. We showed that the operator $G_{s,n}^{+}$ is not positive. $\quad\square$

References

1. Altomare, F., Campiti, M.: Korovkin-Type Approximation Theory and Its Applications. in: de Gruyter Studies in Mathematics, vol. 17, Walter de Gruyter & Co. Berlin, 1994.
2. Badea, C.: Bernstein polynomials and operator theory. Results Math. 53 (3-4), pp. 229–236, 2009).
3. Cao, J., Gonska, H.: Approximation by Boolean Sums of Positive Linear Operator. II Gopengauz Type Estimates. Journal of Approximation Theory, 57, pp. 77–89, (1989).
4. Ditzian, Z., Ivanov, K.: Strong converse inequalities. J. d'Analyse Mathematique, 61, 61–111, (1991).
5. Gavrea, I., Ivan, M.: A note on the fixed points of positive linear operators. Journal of Approximation Theory, 227, pp. 27–36, (2018).
6. Lorentz, G.: Approximation of Functions. Holt-Rinehart and Winston, New York, 1966.
7. Zapryanova, T.: Approximation by Operators of Cao-Gonska type and Direct and Converse Theorem. Mathematics and Education in Mathematics, pp. 189–194, (2004).
8. Zapryanova, T.: Approximation by operators of Cao-Gonska type . Direct and converse theorems. Mathematics and Education in Mathematics, Proceedings of the Thirty Seventh Spring Conference of the Union of Bulgarian Mathematicians, pp. 169–175, (2008).
9. Zapryanova, T.: A Characterization of the $K-$ Functional for the Algebraic Version of the Trigonometric Jackson Integrals $G_{s,n}$ and the K - functionals for Cao-Gonska Operators $G_{s,n}^{*}$ and $G_{s,n}^{+}$. Result. Math. 54, pp. 397–413, (2009).
10. Zapryanova, T., Souroujon, D.: On the Iterates of Jackson Type Operator $G_{s,n}$. Mediterr. J. Math. 13, pp. 5053–5061, (2016).

Some Notes on Arcsine Exponentiated-X Family

Maria Vasileva and Nikolay Kyurkchiev

Abstract In 2020, He, Ahmad, Afify and Goual proposed a new family based on classical Arcsine distribution and exponentiated family named Arcsine Exponentiated-X family. The main purpose of this work is to continue studying the intrinsic properties of these families, namely to study the "saturation"—d to the horizontal asymptote in the Hausdorff sense. We obtain precious estimates for the value of the Hausdorff distance that can be used as an additional criterion in practice. We consider Hausdorff approximation of two submodels from the proposed family with baseline distribution—some extensions of the classical Weibull distribution. In this study, we also define a new "*adaptive ASE-W model with polynomial variable transfer*". We develop several simple dynamic programming modules implemented within the programming environment *CAS Wolfram Mathematica*. All new results are illustrated with suitable numerical experiments with real cumulative data.

Keywords Generated family · Trigonometric distribution · Hausdorff distance · Heaviside step function

M. Vasileva (✉) · N. Kyurkchiev
Faculty of Mathematics and Informatics, University of Plovdiv Paisii Hilendarski, 24 Tzar Asen, 4000 Plovdiv, Bulgaria
e-mail: mariavasileva@uni-plovdiv.bg

N. Kyurkchiev
Institute of Mathematics and Informatics,Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 8, 1113 Sofia, Bulgaria

1 Introduction

Over the last few years, there are many modifications to classical models in order to achieve a better fit to the modeling data. The basic requirement is to have an adequate description and a high degree of flexibility. One of the effective techniques to extend the classical distributions without adding extra parameters is using trigonometric functions and their inverses. In the literature, there are many modifications of Sin-G, Cos-G, Tan-G and Sec-G (see related articles [1–6] and reference therein). Also, there are several new families based on inverse trigonometric functions (see [7–10]).

Some of these new families are based on classical Arcsine distribution. Let us recall that the Arcsine distribution is a symmetric distribution with a minimum at $1/2$ and vertical asymptotes at 0 and 1. It is a well-know fact that the Arcsine distribution is a special case of the Beta distribution, especially $Beta(1/2; 1/2)$.

He, Ahmad, Afify and Goual [11] proposed a new family of distributions named as *Arcsine exponentiated-X* (ASE-X). The cumulative distribution function of the ASE-X family is given by

$$F(t; \lambda, \xi) = \frac{2}{\pi} \arcsin \left[G(t; \xi)^\lambda \right], \tag{1}$$

where $G(t; \xi)$ is the cumulative distribution function of the baseline distribution and ξ represents the parameter vector of the baseline distribution and $\lambda > 0$ is an additional shape parameter.

Note that the ASE-X family contains some of the G-class distributions as its submodels. For example, if $\lambda = 1/2$, the ASE-X family reduces to Arcsine-G family (see [12]), if $\lambda = 1$, the ASE-X family reduces to the Arcsine-X family (see [13]).

This paper deals with the asymptotic behavior of some adaptive functions of the Hausdorff distance between the Heaviside function and distribution functions based on the arcsine function. Similar investigation on some generalized trigonometric distributions (Sin-G, Cos-G, Tan-G and ArcTan-G families) can be found in [14–20]. This study can be very useful for specialists that are working in several scientific fields like insurance, financial mathematics, analysis and approximation of data sets in a various modeling problems and others. One can see some modeling and approximation problems in related articles [21–25] and references therein.

Definition 1 The shifted Heaviside step function is defined by

$$h_{t_0}(t) = \begin{cases} 0, & \text{if } t < t_0, \\ [0, 1], & \text{if } t = t_0, \\ 1, & \text{if } t > t_0. \end{cases}$$

Definition 2 ([26, 27]) The Hausdorff distance (the H-distance) $\rho(f, g)$ between two interval functions f, g on $\Omega \subseteq \mathbb{R}$, is the distance between their completed graphs $F(f)$ and $F(g)$ considered as closed subsets of $\Omega \times \mathbb{R}$. More precisely,

$$\rho(f, g) = \max\{ \sup_{A \in F(f)} \inf_{B \in F(g)} ||A - B||, \sup_{B \in F(g)} \inf_{A \in F(f)} ||A - B||\}, \qquad (2)$$

wherein $||.||$ is any norm in $\mathbb{R}^2$, e.g., the maximum norm $||(t, x)|| = \max\{|t|, |x|\}$; hence, the distance between the points $A = (t_A, x_A)$, $B = (t_B, x_B)$ in $\mathbb{R}^2$ is $||A - B|| = \max(|t_A - t_B|, |x_A - x_B|)$.

The main purpose of this work is to continue studying the intrinsic properties of these families, namely in addition to the analysis of the important characteristic "confidential bounds", it is appropriate to study the "saturation"—d to the horizontal asymptote in the Hausdorff sense. Based on the proposed approach, some extensions of the Weibull distribution called the Arcsine Exponentiated-Weibull (ASE-W) and Arcsine modified-Weibull (ASM-W) distributions are studied in detail. In this study, we also define a new "*adaptive ASE-W model with polynomial variable transfer*." The applicability of the proposed models is proved in simulation study to some cumulative data as the earthquake insurance, mortality rate of COVID-19 patients in Canada, runoff amounts of Jug Bridge, Maryland. We develop several simple dynamic programming modules implemented within the programming environment *CAS Wolfram Mathematica* to shows the applicability of presented results.

2   Hausdorff Approximation

In this section, we investigate the "saturation"—d in the Hausdorff sense to the horizontal asymptote $a = 1$. For the function $F_{ASE-X}(t; \lambda, \xi)$ defined by (1), we have

$$F_{ASE-X}(t_0; \lambda, \xi) = \frac{1}{2} \quad with \quad t_0 = G^{(-1)}\left(2^{-(2\lambda)^{-1}}; \xi\right).$$

Then the Hausdorff distance d between $F_{ASE-X}(t; \lambda, \xi)$ and the Heaviside function $h_{t_0}(t)$ at the "median level" satisfies the following nonlinear equation:

$$F_{ASE-X}(t_0 + d; \lambda, \xi) = 1 - d. \qquad (3)$$

Next, theorem gives upper and lower estimates for the Hausdorff distance d.

Theorem 1 *Let*

$$A = 1 + \frac{2}{\pi} \lambda \, 2^{(2\lambda)^{-1}} G'\left(G^{(-1)}\left(2^{-(2\lambda)^{-1}}\right); \xi\right). \qquad (4)$$

For the Hausdorff distance d between shifted Heaviside function $h_{t_0}(t)$ and the cumulative distribution function $F_{ASE-X}(t; \lambda, \xi)$ defined by (1), the following inequalities hold true for $2.1A > e^{1.05}$:

$$d_l = \frac{1}{2.1A} < d < \frac{\ln(2.1A)}{2.1A} = d_r.$$

Proof We examine the following approximation of

$$H(d) = F_{ASE-X}(t_0 + d; \lambda, a, b) - 1 + d$$

as we use the function

$$T(d) = -\frac{1}{2} + \left(1 + \frac{2}{\pi} \lambda \, 2^{(2\lambda)^{-1}} G'\left(G^{(-1)}\left(2^{-(2\lambda)^{-1}}\right); \xi\right)\right) d.$$

Indeed, from Taylor expansion, we get $T(d) - H(d) = \mathcal{O}(d^2)$. The functions $T(d)$ and $H(d)$ are increasing. This means that $T(d)$ approximates $H(d)$ with $d \to 0$ as $\mathcal{O}(d^2)$. Note that, if $2.1A > e^{1.05}$ holds, then it is easy to show that

$$T(d_l) < 0 \quad and \quad T(d_r) > 0.$$

This completes the proof. □

3 Some Special Cases of Arcsine Exponentialed-X Family

In this section, we consider a special case of ASE-X family. List of some submodels based on the ASE-X family can be found in [11]. The reader can formulate other special cases of the proposed family using different baseline distributions with corresponding approximation problems.

3.1 Arcsine Exponentiated-Weibull (ASE-W) Distribution

Fisrt, we consider a special case of ASE-X family with baseline Weibull distribution called Arcsine exponentiated-Weibull (ASE-W) distribution.

The cumulative function of Weibull distribution is

$$G(t) = 1 - e^{-bt^a},$$

where $a, b > 0$ are the shape parameters.

Definition 3 Arcsine exponentiated-Weibull (ASE-W) distribution is associated with the cumulative distribution function given as

$$F_{ASE-W}(t; \lambda, a, b) = \frac{2}{\pi} \arcsin\left[\left(1 - e^{-bt^a}\right)^\lambda\right]. \tag{5}$$

For $a = 1$, the ASE-W distribution reduces to the ASE-exponential distribution with parameter b and for $a = 2$, it reduces to the ASE-Rayleigh distribution with parameter b.

The Hausdorff distance d between $F_{ASE-W}(t; \lambda, a, b)$ defined by (5) and the Heaviside function $h_{t_0}(t)$ satisfies the relation

$$F_{ASE-W}(t_0 + d; \lambda, a, b) = 1 - d, \tag{6}$$

where

$$F_{ASE-W}(t_0; \lambda, a, b) = \frac{1}{2} \quad with \quad t_0 = \left(-\frac{1}{b} \log \left(1 - 2^{-(2\lambda)^{-1}} \right) \right)^{1/a}.$$

Next, corollary of Theorem 1 gives useful estimates for Hausdorff approximation d.

Corollary 1 *Let*

$$B = 1 + \frac{2}{\pi} ab\lambda \ (2^{(2\lambda)^{-1}} - 1) \left(\frac{1}{b} \log \left(1 - 2^{-(2\lambda)^{-1}} \right) \right)^{\frac{a-1}{a}}. \tag{7}$$

For the Hausdorff distance d between shifted Heaviside function $h_{t_0}(t)$ and the cumulative distribution function $F_{ASE-W}(t; \lambda, a, b)$ defined by (5), the following inequalities hold true for $2.1B > e^{1.05}$:

$$d_l = \frac{1}{2.1B} < d < \frac{\ln (2.1B)}{2.1B} = d_r.$$

In Table 1, we present some computational examples for different values of parameters λ, a and b. We use Corollary 1 for computation of the values of upper and lower estimates d_l and d_r.

In [13], considered data set related to the earthquake insurance. They obtained that this data set can be approximated with the Arcsine exponentiated-Weibull

Table 1 Bounds for Hausdorff distance d computed by Corrollary 1

λ	a	b	d_l	d computed by (6)	d_r
17.25	5.54	0.36	0.114921	0.146136	0.248632
0.12	1.71	18.31	0.101311	0.176773	0.231958
0.63	7.05	0.71	0.174054	0.183464	0.304314
1.75	2.95	1.93	0.208334	0.233169	0.326795
2.03	0.25	3.61	0.182551	0.309569	0.310469
0.15	0.89	0.98	0.238425	0.341143	0.34183
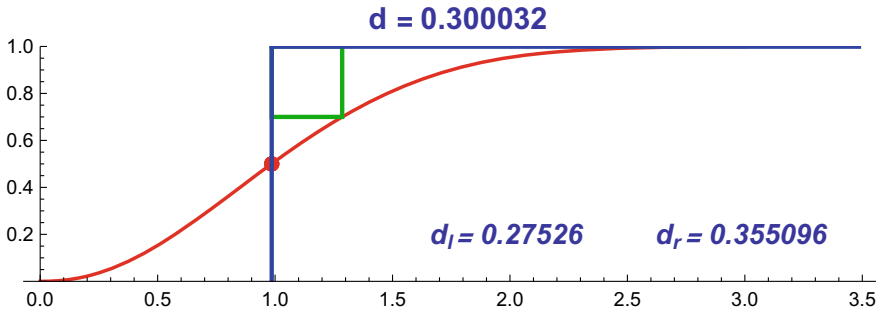
Fig. 1 Approximation of CDF function of ASE-W distribution to data set related to the earthquake insurance

(ASE-W) distribution with parameters $\lambda = 1$, $a = 2.222$ and $b = 1.268$. In Fig. 1, we present the results of the approximation of the Heaviside step function and cumulative distribution function of ASE-W distribution with corresponding parameters. From Corollary 1, we obtain the values of Hausdorff distance $d = 0.30032$ with estimates $d_l = 0.27526$ and $d_r = 0.355096$, respectively.

3.2 Arcsine-Modified Weibull (ASM-W) Distribution

There are many modifications of classical Weibull distribution in the literature. In 2014, Almalki and Nadarajah [29] present a review of some of them.

In 2021, Liu et al. [28] consider the ASE-X family with baseline modified Weibull distribution named Arcsine-modified Weibull model (ASM-W) (for $\lambda = 1$). Let us recall that the cumulative distribution function of modified Weibull distribution is defined by

$$G(t) = 1 - e^{-b_1 t^a - b_2 t},$$

where a, b_1, $b_2 > 0$ are the shape parameters. This modification improves the fitting power of the exponential, Rayleigh, linear failure rate and Weibull distributions (see [30]).

Definition 4 Arcsine modified-Weibull (ASM-W) distribution is associated with the cumulative distribution function given as

$$F_{ASM-W}(t; \lambda, a, b_1, b_2) = \frac{2}{\pi} \arcsin\left[\left(1 - e^{-b_1 t^a - b_2 t}\right)^\lambda\right]. \tag{8}$$

In [28], Liu et al. analyzed data set that represents the mortality rate of COVID-19 patients in Canada. They obtained that this data set can be approximated
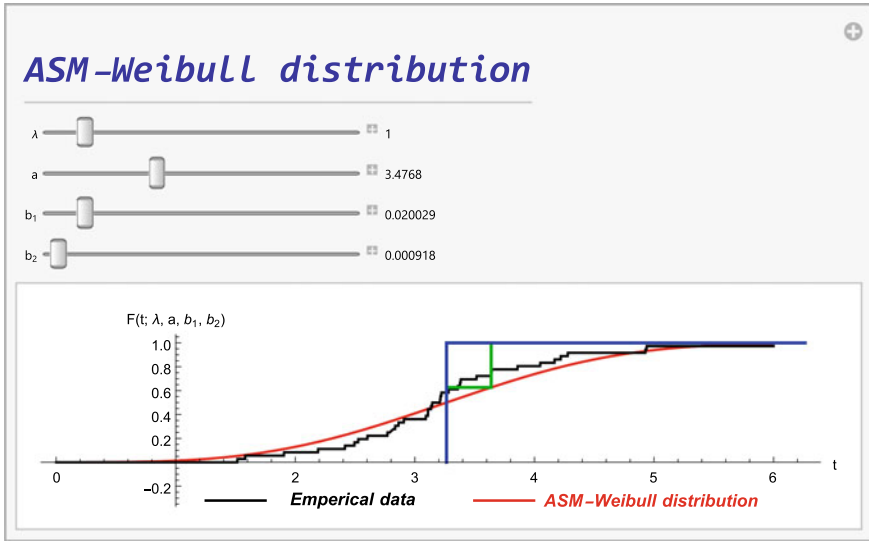
Fig. 2 Model (8) for data set (normalized) representing the mortality rate of the COVID-19 patients in Canada with Hausdorff distance $d = 0.373491$

with the Arcsine modified-Weibull (ASM-W) distribution with parameters $\lambda = 1$, $a = 3.4768$, $b_1 = 0.020029$ and $b_2 = 0.000918$. In Fig. 2, we present the results of approximation of the Heaviside step function and cumulative distribution function of ASM-W distribution with corresponding parameters. We obtain the Hausdorff distance $d = 0.373491$ from the relation

$$F_{ASM-W}(t_0 + d; \lambda, a, b_1, b_2) = 1 - d$$

for $t_0 = 3.26468$, that is the solution of $F_{ASM-W}(t_0; \lambda, a, b_1, b_2) = \dfrac{1}{2}$.

3.3 Adaptive ASE-W Model with Polynomial Variable Transfer

Definition 5 Consider the following new "*adaptive ASE-W model with polynomial variable transfer*":

$$F^*(t; \lambda, a, b) = \tfrac{2}{\pi} \arcsin\left[\left(1 - e^{-bf(t)^a}\right)^\lambda\right].$$

$$f(t) = \sum_{i=0}^{n} a_i t^i, \quad a_0 = 0. \tag{9}$$
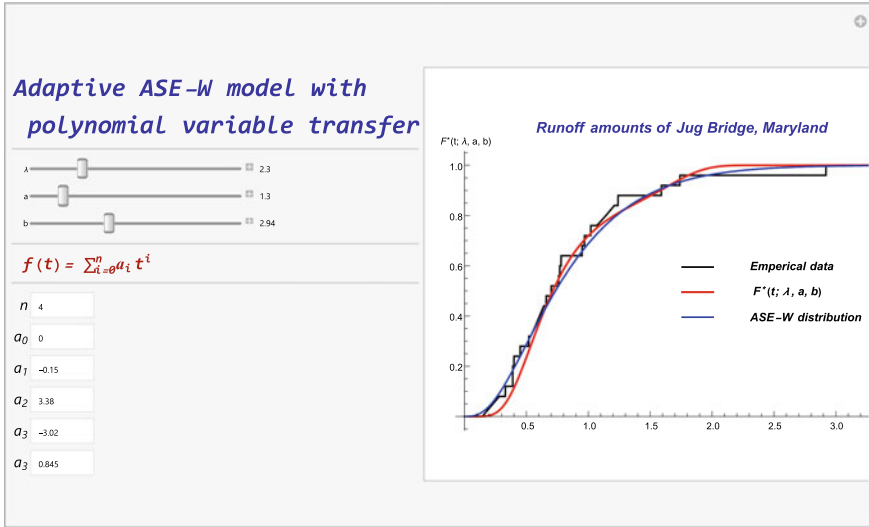
Fig. 3 The model (9) for runoff amounts data (normalized) of Jug Bridge, Maryland

Note that the model is highly sensitive to the type and location of the zeros of the polynomial $f(t)$ (see [31]).

The applicability of the model (9) is proved in the simulation study to the data represent runoff amounts of Jug Bridge, Maryland (see Chhikara and Folks [32]):

0.17, 1.19, 0.23, 0.33, 0.39, 0.39, 0.40, 0.45, 0.52, 0.56, 0.59, 0.64, 0.66

0.70, 0.76, 0.77, 0.78, 0.95, 0.97, 1.02, 1.12, 1.24, 1.59, 1.74, 2.92.

For the actual data in the specified period, our new model $F^*(t)$ for

$$\lambda = 2.3, \quad a = 1.3 \ b = 2.94,$$

$$n = 4, \quad a_0 = 0, \quad a_1 = -0.15, \quad a_2 = 3.4, \quad a_3 = -3.38, \quad a_4 = 0.845$$

is depicted on Fig. 3.

References

1. Souza, L., Junior, W., De Brito, C., Chesneau, C., Ferreira, T., Soares, L.: On the Sin-G class of distributions: theory, model and application. Journal of Mathematical Modeling 7(3), 357–379 (2019). https://doi.org/10.22124/jmm.2019.13502.1278

2. Al-Babtain, A. A., Elbatal, I., Chesneau, C., Elgarhy, M.: Sine Topp-Leone-G family of distributions: Theory and applications. Open Physics 18(1), 574–593 (2020). https://doi.org/10.1515/phys-2020-0180

3. Muhammad, M., Bantan, R.A.R., Liu, L., Chesneau, C., Tahir, M. H., Jamal, F., Elgarhy, M.: A new extended Cosine-G distributions for lifetime studies. Mathematics 9(21), 2758 (2021). https://doi.org/10.3390/math9212758

4. Souza, L., de Oliveira, W.R., de Brito, C.C.R., Chesneau, C., Fernandes, R., Ferreira, A. E.: Tan-G class of trigonometric distributions and its applications. CUBO 23(1), 01–20 (2021).

5. Souza, L., de Oliveira, W.R., de Brito, C.C.R., Chesneau, C., Fernandes, R., Ferreira, T.A.E.: Sec-G class of distributions: Properties and applications. Symmetry 14(2), 299 (2022). https://doi.org/10.3390/sym14020299

6. Mahmood, Z., Jawa, T., Sayed-Ahmed, N., Khalil, E. M., Muse, A. H., Tolba, A. H.: An extended cosine generalized family of distributions for reliability modeling: Characteristics and applications with simulation study. Math. Probl. Eng. 2022 Art. ID 3634698, (2022). https://doi.org/10.1155/2022/3634698

7. Chesneau, C., Tomy, L., Gillariose, J.: On a new distribution based on the arccosine function. Arabian Journal of Mathematics 10(3), 589–598 (2021). https://doi.org/10.1007/s40065-021-00337-x

8. Alkhairy, I., Nagy, M., Muse, A., Hussam, E.: The Arctan-X family of distributions: Properties, simulation, and applications to actuarial sciences. Complexity 2021 Art. ID 4689010, (2021). https://doi.org/10.1155/2021/4689010

9. Altun, E., Alizadeh, M., Ramires, T. G., Ortega, Edwin M. M.: Generalized Odd power Cauchy family and its associated heteroscedastic regression model. Stat. Optim. Inf. Comput. 9(3), 516–528 (2021). https://doi.org/10.19139/soic-2310-5070-765

10. Shrahili, M., Elbatal, I.: Truncated Cauchy power odd Fréchet-G family of distributions: Theory and applications. Complexity 2021 Art. ID 4256945, (2021). https://doi.org/10.1155/2021/4256945

11. He, W., Ahmad, Z., Afify, A. Z., Goual, H.: The arcsine exponentiated-X family: validation and insurance application. Complexity 2020 Art. ID 8394815, (2020). https://doi.org/10.1155/2020/8394815

12. Rahman, M.: Arcsine-G family of distributions. Journal of Statistics Applications Probability Letters 8(3), 169–179 (2021). https://doi.org/10.18576/jsapl/080303

13. Tung, Y. L., Ahmad, Z., Mahmoudi, E.: The arcsine-X family of distributions with applications to financial sciences. Computer Systems Science and Engineering 39(3), 351–363 (2021). https://doi.org/10.32604/csse.2021.014270

14. Kyurkchiev, N., Rahneva, O., Iliev, A., Malinova, A., Rahnev, A.: Investigations on some generalized trigonometric distributions. Properties and Applications. Plovdiv University Press, Bulgaria (2021).

15. Kyurkchiev, N., Iliev, A., Arnaudova, V., Rahnev, Asen: Investigations on some new cumulative distributions via cosine and sine functions. Applications. Int. J. Differ. Equ. 20(1), 75–88 (2021). https://doi.org/10.12732/ijdea.v20i1.6

16. Kyurkchiev, N., Iliev, A., Rahnev, Asen: A new Cos-G family with baseline cumulative function of Volmer-type. Applications. Int J Pure Appl Math 15, 55–65 (2021).

17. Kyurkchiev, N., Iliev, A., Rahnev, A.: Properties and applications of an Tan–G family of "adaptive functions". Int. J. Circuits, Syst. Signal Process. 15, 1292–1296 (2021). https://doi.org/10.46300/9106.2021.15.139

18. Kyurkchiev, N., Rahneva, O., Malinova, A., Iliev, A.: On some adaptive G-families. Applications. Int. J. Differ. Equ. 20(1), 89–101 (2021). https://doi.org/10.12732/ijdea.v20i1.7

19. Kyurkchiev, N., Iliev, A., Rahneva, O., Kyurkchiev, V.: A Look at Some Trigonometric-G Families with Baseline Inverted Exponential (CDF). Applications. Int. J. Differ. Equ. 20(1), 103–119 (2021). https://doi.org/10.12732/ijdea.v20i1.8

20. Vasileva, M.: Some Notes for Two Generalized Trigonometric Families of Distributions. Axioms 11(4), 149 (2022). https://doi.org/10.3390/axioms11040149

21. Vasileva, M.: Some Notes on the Omega Distribution and the Pliant Probability Distribution Family. Algorithms 13(12), 324 (2020). https://doi.org/10.3390/a13120324

22. Vasileva, M., Iliev, A., Rahnev, A., Kyurkchiev, N.: On the approximation of the Haar scaling function by sigmoidal scaling functions Int. J. Differ. Equ. 20(1), 1–13 (2021). https://doi.org/10.12732/ijdea.v20i1.932

23. Vasileva, M., Malinova, A., Rahneva, O., Angelova, E.: A note on the Unit–Rayleigh ”adaptive function“. AIP Conference Proceedings 2459(1), 030039 (2022). https://doi.org/10.1063/5.0083539

24. Kyurkchiev, V., Iliev, A., Rahnev, A., Terzieva, T., Angelova, E.: On Some Understudied Models with Applications in the Field of Debugging Theory. Commun. Appl. Anal. 26(1), 9–18 (2022). https://doi.org/10.12732/caa.v26i1.2

25. Kyurkchiev, V., Boyadjiev, G., Kyurkchiev, N.: A Software Tool for Simulating The Dynamics of a New Extended Family of Lotka–Volterra Competition Model. Int. J. Differ. Equ. 21(1), 33–46 (2022). https://doi.org/10.12732/ijdea.v21i1.3

26. Hausdorff, F., Set Theory. 2nd ed. Chelsea Publ., New York, NY, USA (1962).

27. Sendov, B.L., Hausdorff approximations. In *Mathematics and Its Applications*, Springer Science & Business Media: Berlin/Heidelberg, Germany, Vol. 50, 1–367 (1990).

28. Liu, X., Ahmad, Z., Gemeay, A. M., Abdulrahman A. T., Hafez E. H., Khalil N.: Modeling the survival times of the COVID-19 patients with a new statistical model: A case study from China. PLoS ONE 16(7), e0254999 (2021). https://doi.org/10.1371/journal.pone.0254999

29. Almalki. S. J., Nadarajah S.: Modifications of the Weibull distribution: A review. Reliab Eng Syst Saf 124, 32–55 (2014). https://doi.org/10.1016/j.ress.2013.11.010

30. Sarhan, A. M., Zaindin, M.: Modified Weibull distribution. Applied Sciences 11, 123–136 (2009).

31. Kyurkchiev, N.,: Initial approximations and root finding methods. 104, Wiley-VCh Berlin (1998).

32. Chhikara, R. S., Folks, J. L.: The inverse Gaussian distribution as a lifetime model. Technometrics 19 (4), 461–468 (1977).

On a Piecewise Smooth Gompertz Growth Function. Applications

Vesselin Kyurkchiev, Anton Iliev, Asen Rahnev, and Nikolay Kyurkchiev

Abstract Following the ideas given in (Kyurkchiev in Int J Differ Equ Appl 21(1):1–17, 2022 [5]), in this article, we study a hypothetical piecewise smooth Gompertz growth function $G(g_1(t), g_2(t))$. Precise bounds for the Hausdorff distance d between the Heaviside step function h_0 and the sigmoid G are given. The applicability of the new model is proved in simulation study to grouped data from $dataNASA = \{dataNasa1 \cup dataNasa2\}$ (Nagaraju et al. in SoftwareX 10:100357, 2019 [6]). Some numerical examples, using *CAS MATHEMATICA* are also given. Studies in this paper can also be applied to random shifted sigmoidal functions of Gompertz type.

Keywords Smooth Gompertz growth function · Hausdorff distance · Precise bounds

V. Kyurkchiev · A. Iliev (✉) · A. Rahnev · N. Kyurkchiev
Faculty of Mathematics and Informatics, University of Plovdiv Paisii Hilendarski, 24, Tzar Asen Str., 4000 Plovdiv, Bulgaria
e-mail: aii@uni-plovdiv.bg

V. Kyurkchiev
e-mail: vkurkchiev@uni-plovdiv.bg

A. Rahnev
e-mail: assen@uni-plovdiv.bg

N. Kyurkchiev
e-mail: nkyurk@uni-plovdiv.bg

A. Iliev · N. Kyurkchiev
Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Bl. 8, 1113, Sofia, Bulgaria

461

1 Introduction: The Basic Problem, Preliminaries Notations

Early studies in demography [1] gave rise to the familiar Gompertz and logistic functions [2, 3]. Both functions are sigmoid (S-shaped), more specifically they are monotone increasing on their definition domain. These sigmoid functions are used in biological applications, income and lifetime analysis, financial mathematics, fuzzy set theory, impulsive analysis, etc.

In [5], the author considers a hypothetical piecewise smooth sigmoidal growth function based on the Verhulst model. These ideas can be successfully continued.

Gompertz model makes an extensive use of the sigmoidal function:

$$g_1(t) = Ae^{-e^{-kt}}. \tag{1}$$

Without community restriction, we will consider model (1) in the constraint: $g_1(0) = Ae^{-1} = \frac{1}{2}$, i.e., $A = \frac{e}{2}$. For a depiction of this model at concrete value of the parameter k, see Fig. 1.

In the fields of Growth Theory and Theory of Computer Viruses Propagation, the stumper frequently arises how to develop a modified model at an already fixed value of parameter k and an amendment in the dynamics of the growth process for $t > t_0$, in which saturation to the horizontal asymptote at level $B = Ae^{-e^{-1}}$ is reached (in our model $A = \frac{e}{2} \approx 1.35914$, $B = Ae^{-e^{-1}} \approx 0.940798$). This can be performed as an example with $g_2(t)$ (for $t > 0$, see Fig. 2)

$$g_2(t) = Ae^{-e^{-\frac{kt}{1+kt}}}. \tag{2}$$
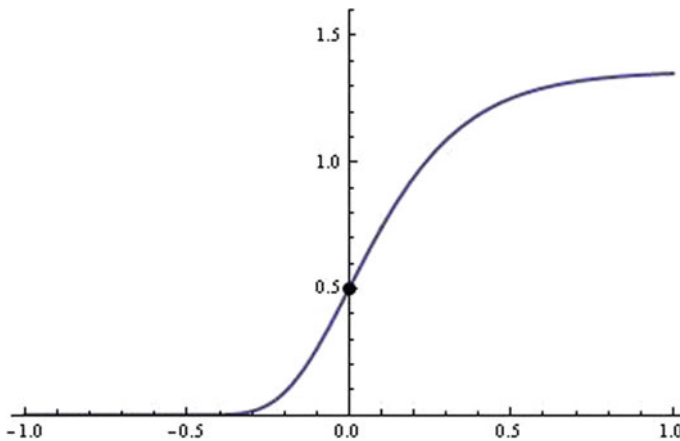
Fig. 1 The function $g_1(t)$ for fixed $k = 5$, $g_1(0) = \frac{1}{2}$)
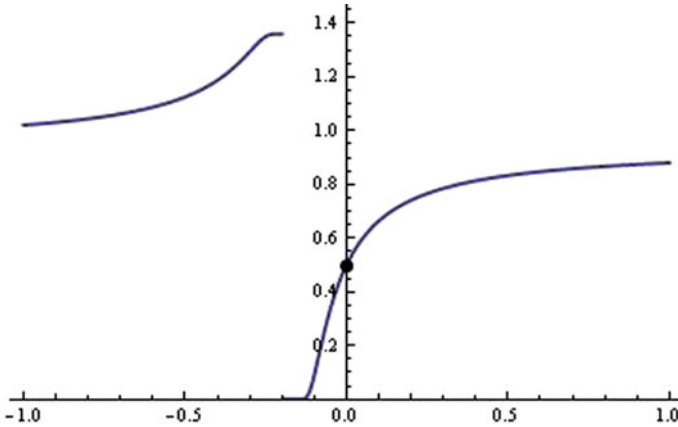
Fig. 2 The function $g_2(t)$ for fixed $k = 5$, $(g_2(0) = \frac{1}{2})$

To construct the function $g_2(t)$, we used a substantial "fractional linear correction".

Definition 1 This leads us to think of the following hypothetical piecewise smooth sigmoidal growth function:

$$
G(t) := \begin{cases} Ae^{-e^{-kt}} := g_1(t), & t < 0 \\ \dfrac{1}{2}, & t = 0 \\ Ae^{-e^{-\frac{kt}{1+kt}}} := g_2(t), & t > 0. \end{cases}
\tag{3}
$$

Evidently, from (3), we have

$$
g_1'(0) = g_2'(0).
$$

Definition 2 The modified Heaviside step function is defined by

$$
h_0(t) = \begin{cases} 0, & \text{if } t < 0, \\ [0, B], & \text{if } t = 0, \\ B, & \text{if } t > 0. \end{cases}
$$

In this paper, we will study some properties of this novel family. More exactly bounds for the H-distance [4] d between the Heaviside step function and the classical Gompertz function

$$
\sigma_{\alpha,\beta}(t) = e^{-\alpha e^{-\beta t}}
\tag{4}
$$

can be located in [7]:

Theorem 1 *The H-distance d between the step function h_0 (with $B = 1$) and the sigmoid Gompertz function $\sigma_{\alpha,\beta}(t)$ ($\alpha = 0.69314718...$) can be expressed in terms of the parameter β for any real $\beta \geq 2$ as follows:*

$$\frac{2\alpha - 1}{1 + \alpha\beta} < d < \frac{\ln(1 + \alpha\beta)}{1 + \alpha\beta}. \tag{5}$$

In supplement, we will view the provocative task of approximating the Heaviside step function with the new class of growth functions $G(t)$ with respect to the Hausdorff distance.

2 Main Results

2.1 The H-Distance Between Function h_0 and the Sigmoidal Function $G(t)$

For the H-distances—d_1 and d_2 through $g_1(t)$ and $g_2(t)$, we possess

$$g_1(-d_1) = d_1 \tag{6}$$

and

$$g_2(d_2) = B - d_2. \tag{7}$$

For illustration, for fixed $k = 10$, we find $d_1 = 0.0970532$ and $d_2 = 0.1534$. For $k = 20$, we have $d_1 = 0.0575575$ and $d_2 = 0.1149$ and, for $k = 15$, we find $d_1 = 0.0718839$ and $d_2 = 0.129876$ (see, Fig. 3). Obviously, for the H-distance $d = \rho(h_0, G)$ between the Heaviside step function h_0 and the sigmoidal function, G is satisfied
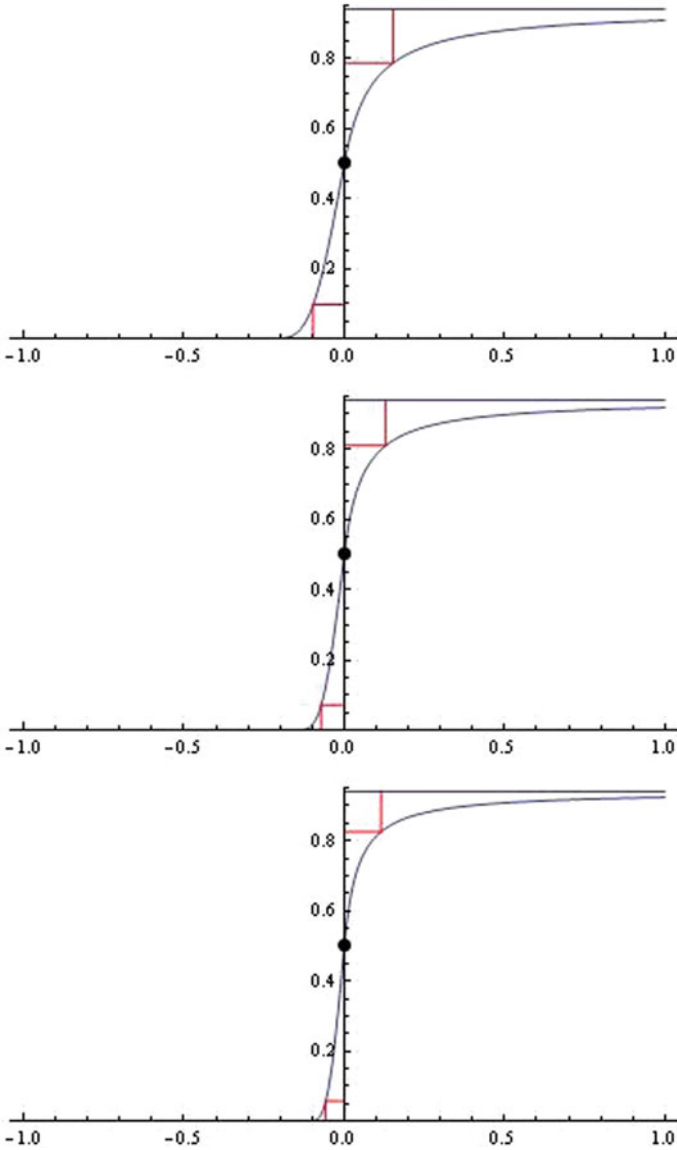
$$d = \max\{d_1, d_2\}. \tag{8}$$

The following is actual.

Theorem 2 *Let $M = 1 + \frac{k}{2}$ and $0 < d < \frac{1}{2}$. For large enough values of k, for the "saturation"-d, we have*
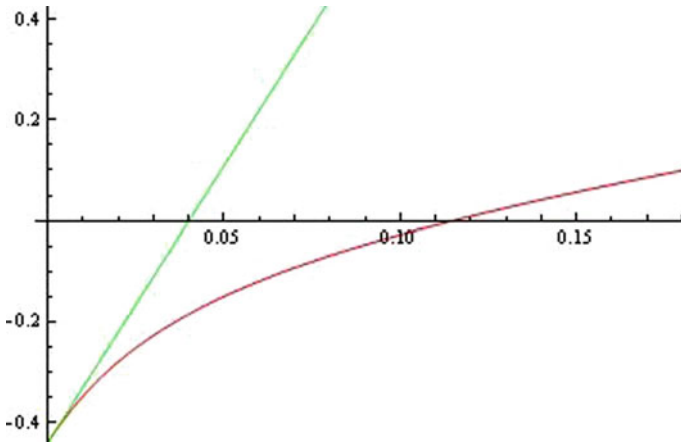
$$d \approx \frac{\ln(2.3M)}{2.3M} := d^*. \tag{9}$$

Draft of the proof Moreover, the proof is organized on a manner suggested in [9], we will accept that, from (7), it is can be concluded that $d(= d_2)$ is the only positive root of the equation:

$$F(d) := g_2(d) - B + d = 0. \tag{10}$$

Fig. 3 The function $G(t)$ for a) $k = 10$; b) $k = 15$; c) $k = 20$ (Asymptote at level $B = Ae^{-e^{-1}} \approx 0.940798$)

Fig. 4 The functions $F(d)$ (red) and $H(d)$ (green) for fixed $k = 20$; Hausdorff distance $d = 0.1149$

Table 1 Bounds for d computed by (7) and (9) for various values of k

k	d computed by (7)	d^* computed by (9)
10	0.1534	0.190193
15	0.129876	0.15207
20	0.1149	0.1277

Obviously, the function

$$H(d) := -0.440798 + \left(1 + \frac{k}{2}\right)d \tag{11}$$

approximates $F(d)$ with $d \to 0$ as $\mathcal{O}(d^2)$ (see, for example Fig. 4). Since an exact analysis, we obtain the estimate (10).

Numerical examples through (7) and (9) are depicted in Table 1.

2.2 Remarks

1. We will point out that the estimate (7) should be beneficial for consumers because of the circumstance that the adjustment of this model in every CAS supposes the information of a proper starting approximation for the root of the nonlinear equation (7).

2. The user can draw some possible piecewise smooth sigmoidal models through other "fractional rational corrections".
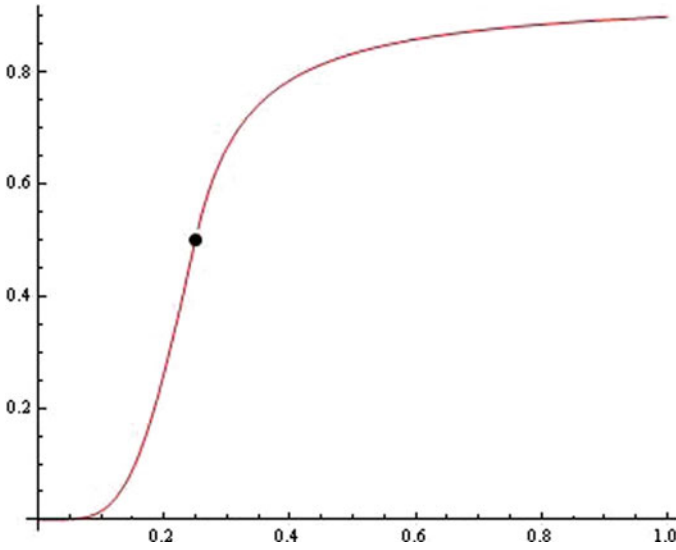
Fig. 5 The function $G^*(t)$ for fixed $k = 10$ and $r = 0.25$

3. Investigations in this paper can also be used for an arbitrarily shifted Gompertz function.

Definition 3 Define the following hypothetical piecewise smooth-shifted Gompertz growth function:

$$G^*(t) := \begin{cases} Ae^{-e^{-k(t-r)}} := g_1^*(t), & 0 < t < r \\ \dfrac{1}{2}, & t = r \\ Ae^{-e^{-\frac{k(t-r)}{1+k(t-r)}}} := g_2^*(t), & t > r. \end{cases} \tag{12}$$
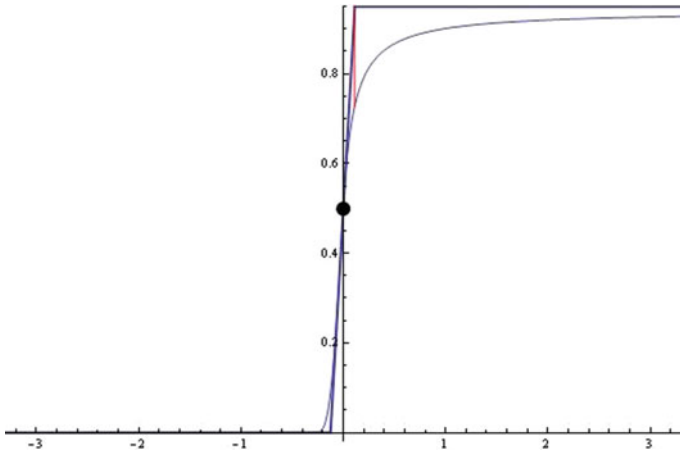
Obviously, from (12), we get

$$g_1^{*\prime}(r) = g_2^{*\prime}(r).$$

For example, for fixed $k = 10$, $r = 0.25$ see, Fig. 5.

4. Define the following piecewise cut function associated to the function $G(t)$ (3):

$$C(t) := \begin{cases} 0, & t < t_0 \\ \frac{k}{2}t + \frac{1}{2} := L(t), & -t_0 \le t \le t_1 \\ B, & t > t_1. \end{cases} \tag{13}$$

Fig. 6 The function $C(t)$ for fixed $k = 8$; The uniform distance $\rho(C, G) = 0.213968$

This function (see, for example, Fig. 6) can find a variety of applications in mathematics and engineering.

2.3 Applications

Frequently, investigators are in front of the following question—how to implement in practice a fixed sigmoidal function when approximating a dataset which is in a "normalized form". One chance is to use the methodology suggested in this paper for building piecewise smooth Gompertz growth function.
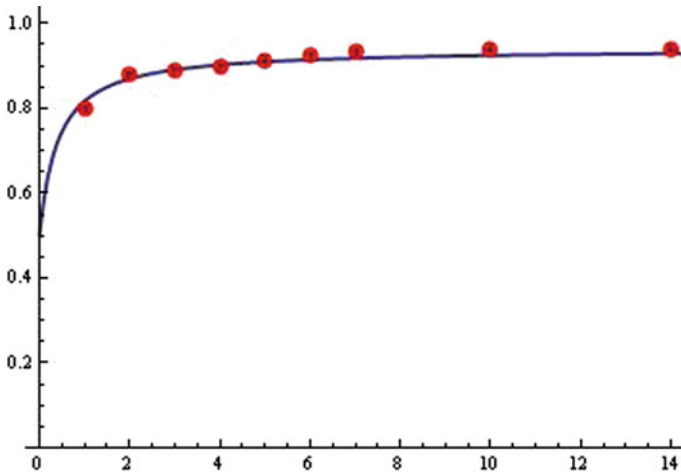
1. Storm worm propagation.

$dataStorm := \{\{1.8, 0.843\}, \{4, 0.926\}, \{5, 0.954\}, \{6, 0.967\},$

$\{7, 0.976\}, \{8, 0.981\}, \{9, 0.985\}, \{10, 0.991\}, \{22, 0.995\}, \{38, 0.997\},$

$\{51, 0.998\}, \{64, 0.9985\}, \{74, 0.999\}, \{83, 1\}, \{100, 1\}\}$

For the "data_Storm"—normalized (see, [10]), the fitted model

$$g_2(t) = Ae^{-e^{-\frac{kt}{1+kt}}}$$

for $k = 2.0749$ is presented on Fig. 7.

2. For the normalized cumulative failures
$dataNASA = \{dataNasa1 \cup dataNasa2\}$ [6],

**Fig. 7** Epidemic data of Storm Worm (normalized) fitted by $g_2(t)$ for $k = 2.07491$

$DataNASA1 := \{\{0, 0\}, \{0.13, 0.014285\}, \{0.24, 0.028571\},$
$\{0.31, 0.042857\}, \{0.38, 0.057142\}, \{0.39, 0.071428\}, \{0.42, 0.085714\},$
$\{0.43, 0.1\}, \{0.5, 0.128571\}, \{0.53, 0.157142\}, \{0.55, 0.185714\},$
$\{0.56, 0.2\}, \{0.57, 0.214285\}, \{0.6, 0.228571\}, \{0.72, 0.25\},$
$\{0.73, 0.264705\}, \{0.78, 0.308823\}, \{0.79, 0.323529\}, \{0.81, 0.338235\},$
$\{0.85, 0.397058\}, \{0.86, 0.411764\}, \{0.87, 0.426470\}, \{0.88, 0.428571\}\};$

$DataNASA2 := \{\{0.88, 0.428571\}, \{0.94, 0.442857\}, \{0.98, 0.457142\},$
$\{1, 0.542857\}, \{1.01, 0.557142\}, \{1.02, 0.571428\}, \{1.05, 0.585714\},$
$\{1.06, 0.6\}, \{1.09, 0.628571\}, \{1.14, 0.642857\}, \{1.15, 0.671428\},$
$\{1.19, 0.685714\}, \{1.2, 0.7\}, \{1.27, 0.714285\}, \{1.28, 0.757142\},$
$\{1.3, 0.8\}, \{1.34, 0.770270\}, \{1.41, 0.783783\}, \{1.49, 0.797297\},$
$\{1.68, 0.810810\}, \{1.73, 0.810810\}, \{1.9, 0.810810\}\};$

we will use model $G^*(g_1^*, g_2^*)$:

$$g_1^*(t) = Ae^{-e^{-k(t-r)}} \quad 0 < t < r \approx 0.9$$

to approximate $dataNASA1$ and

$$g_2^*(t) = Ae^{-e^{-\frac{k(t-r)}{1+k(t-r)}}}, \quad t > r$$
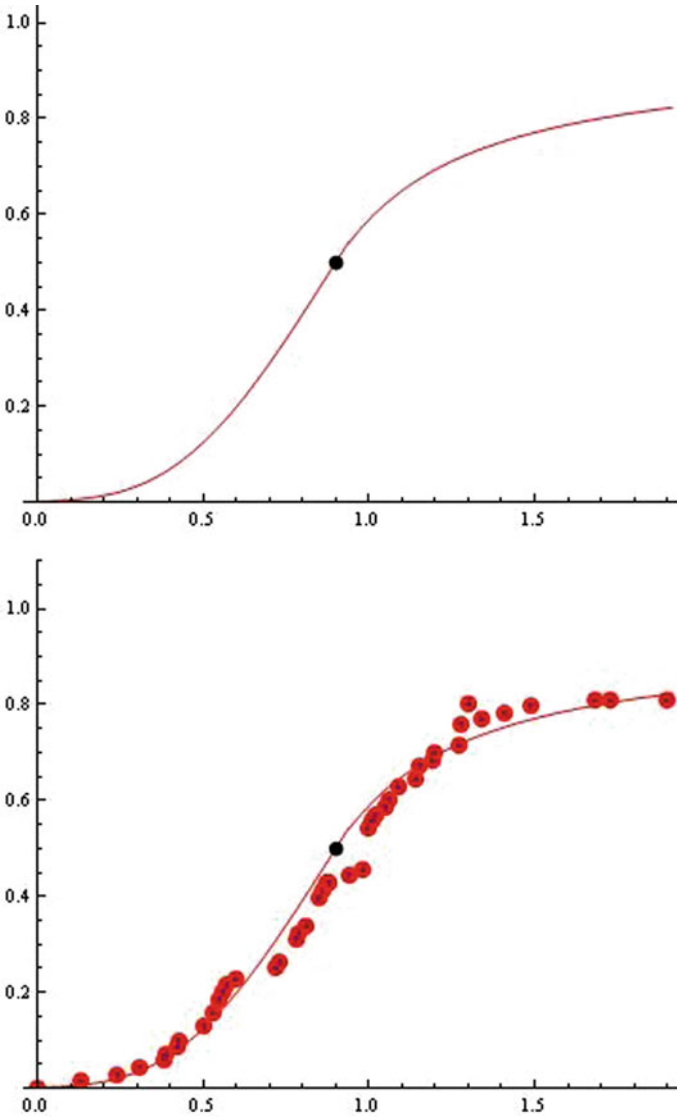
Fig. 8 The model (12) for $\{dataNasa1 \cup dataNasa2\}$

to approximate $dataNASA2$.

For the factual data in the concrete period, the new model for

$$A = 1.35914, \ k = 2.1841053, \ r \approx 0.9$$

is presented on Fig. 8.

The results are satisfactory. One of the advantages of the described model $G^*(g_1^*, g_2^*)$ over the classical Gompertz model (see Fig. 8) is that it better takes into account the real attenuation and saturation to the horizontal asymptote, so characteristic of the models in the field of Growth Theory and Population Dynamics.

Practically, when using such models—of the type (12), the Kolmogorov–Smirnov test should be implemented on the occasion when the parameters of the model are evaluated from "grouped data" (see, [8]).

With the methodological aspects discussed here, we aim to provide the researcher (who does not have to be a mathematician) with a reliable software tool for making adjustments to his chosen model—for the specific experiment.

References

1. Malthus, T. R.: An Essay on the Principle of Population. Oxford World's Classics reprint, (1798).
2. Verhulst, P. F.: Notice Sur la Loi Que la Population Poursuit dans Son Accroissement. Correspondance Mathematique et Physique, 10, (1838), pp. 113–121.
3. Gompertz, B.: On the nature of the function expressive of the law of human mortality. Philosophical Transactions, 27, (1825), pp. 513–519.
4. Sendov, B.: Hausdorff Approximations. Kluwer, Boston, (1990).
5. Kyurkchiev, N.: A note on a hypothetical piecewise smooth sigmoidal growth function: reaction network analysis, applications. International Journal of Differential Equations and Applications, 21, No. 1 (2022), pp. 1–17.
6. Nagaraju, V., Shekar, V., Steakelum, J., Luperon, M., Shi, Y., Fiondella, L.: Practical software reliability engineering with the Software Failure and Reliability Assessment Tool (SFRAT). SoftwareX, 10, (2019), p. 100357.
7. Kyurkchiev, N., Iliev, A.: Extension of Gompertz-type equation in modern science. 240 Anniversary of the birth of B. Gompertz. LAP LAMBERT Academic Publishing, (2018), ISBN: 978-613-9-90569-0.
8. Okamura, H., Dohi, T.: On Kolmogorov–Smirnov test for software reliability models with grouped data. Proc. IEEE 19th International Conference on Software Quality, Reliability and Security, (2019), pp. 77–82.
9. Kyurkchiev, N., Markov, S.: On the Hausdorff distance between the Heaviside step function and Verhulst logistic function. J. Math. Chem., 54, No. 1 (2016), pp. 109–119.
10. Iliev, A., Kyurkchiev, N., Rahnev, A., Terzieva, T.: Some models in the theory of computer viruses propagation. LAP LAMBERT Academic Publishing, Saarbrucken, (2019), ISBN: 978-620-0-00826-8.

11. Iliev, A., Kyurkchiev, N., Markov, S.: On the approximation of the step function by some sigmoid functions. Mathematics and Computers in Simulation, 133, No. 1 (2017), pp. 223–234.
12. Markov, S.: Reaction networks reveal new links between Gompertz and Verhulst growth functions. Biomath, 8, No. 1 (2019).
13. Markov, S.: The Gompertz model revisited and modified using reaction networks: Mathematical analysis. Biomath, 10 (2021), 2110023.
14. Iliev, A., Kyurkchiev, N., Markov, S.: A Note on the New Activation Function of Gompertz Type. Biomath Communications, 4, No. 2 (2017), 20 pp.
15. Anguelov, R., Borisov, M., Iliev, A., Kyurkchiev, N., Markov, S.: On the chemical meaning of some growth models possessing Gompertzian-type property. Math. Meth. Appl. Sci., (2017), pp. 1–12.
16. Markov, S., Iliev, A., Rahnev, A., Kyurkchiev, N.: On the exponential–generalized extended Compertz cumulative sigmoid. International Journal of Pure and Applied Mathematics, 120, No. 4 (2018), pp. 555–562.
17. Kyurkchiev, N., Iliev, A., Rahnev, A.: A new class of activation functions based on the correcting amendments of Gompertz-Makeham type. Dynamic Systems and Applications, 28, No. 2 (2019), pp. 243–257.
18. Kyurkchiev, V., Iliev, A., Rahnev, A., Kyurkchiev, N.: A look at the hypothetical piecewise smooth generalized sigmoidal growth function. Some applications. II. International Journal of Differential Equations and Applications, 21, No. 1 (2022), pp. 19–32.
19. Kyurkchiev, V., Iliev, A., Rahnev, A., Kyurkchiev, N.: Investigations on a hypothetical piecewise smooth log–logistic growth function. Some applications. III. International Electronic Journal of Pure and Applied Mathematics, 16, No. 1, (2022), pp. 1–12.