Springer Proceedings in Mathematics & Statistics

Saber Elaydi
Yoshihiro Hamaya
Hideaki Matsunaga
Christian Pötzsche *Editors*

Advances in Difference Equations and Discrete Dynamical Systems

ICDEA, Osaka, Japan, July 2016

Springer

Springer Proceedings in Mathematics & Statistics

Volume 212

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation 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.

More information about this series at http://www.springer.com/series/10533

Saber Elaydi · Yoshihiro Hamaya
Hideaki Matsunaga · Christian Pötzsche
Editors

Advances in Difference Equations and Discrete Dynamical Systems

ICDEA, Osaka, Japan, July 2016

Springer

Editors
Saber Elaydi
Department of Mathematics
Trinity University
San Antonio, TX
USA

Hideaki Matsunaga
Department of Mathematical Sciences
Osaka Prefecture University
Sakai, Osaka
Japan

Yoshihiro Hamaya
Department of Information Science
Okayama University of Science
Okayama
Japan

Christian Pötzsche
Institut für Mathematik
Alpen-Adria-Universität Klagenfurt
Klagenfurt
Austria

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-981-10-6408-1 ISBN 978-981-10-6409-8 (eBook)
https://doi.org/10.1007/978-981-10-6409-8

Library of Congress Control Number: 2017951059

Mathematics Subject Classification (2010): 37-XX, 39-XX, 92-XX

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This volume contains the proceedings of the 22nd International Conference on Difference Equations and Applications (ICDEA 2016), which was held in Osaka, Japan, from July 24–29, 2016. The conference was organized by Osaka Prefecture University (OPU), under the auspices of the International Society of Difference Equations (ISDE) and Okayama University of Science (OUS). There were more than 80 participants from 20 countries including Austria, Belarus, China, the Czech Republic, France, Hungary, Italy, Jamaica, Japan, Latvia, Malaysia, Norway, Poland, Portugal, Spain, Taiwan, Thailand, UAE, the United Kingdom, and the United States.

The main topics in ICDEA 2016 were difference equations and discrete dynamical systems with applications to mathematical biology and economics. The conference brought together both experts and novices in the theory and applications of difference equations and discrete dynamical systems.

The papers in the proceedings have been through a rigorous refereeing process to insure high scientific quality and standards. Four of the articles were written by the plenary speakers Ryusuke Kon, Christian Pötzsche, Sebastian J. Schreiber, and Petr Stehlík. This book will be of great value to researchers, scientists, and educators who work in the fields of difference equations, discrete dynamical systems, and their applications.

We would like to take this opportunity to express our gratitude to all the participants for making the conference a great success. We would also like to thank the organizing committee for their great efforts in organizing a successful and well-run conference, and the scientific committee who were instrumental in insuring the high scientific standards and quality of the conference. Finally, we are grateful to all the

authors for their contributions to these proceedings and to all the referees for their timely and valuable reviews of the manuscripts.

San Antonio, USA — Saber Elaydi
Okayama, Japan — Yoshihiro Hamaya
Osaka, Japan — Hideaki Matsunaga
Klagenfurt, Austria — Christian Pötzsche
June 2017

Organization Committees

Local Organizing Committees

Yoshihiro Hamaya, Okayama University of Science, Japan
Hideaki Matsunaga, Osaka Prefecture University, Japan
Masakazu Onitsuka, Okayama University of Science, Japan
Naoto Yamaoka, Osaka Prefecture University, Japan

Scientific Committees

Linda Allen, Texas Tech University, USA
Meng Fan, Northeast Normal University, China
Małgorzata Guzowska, University of Szczecin, Poland
Hideaki Matsunaga, Osaka Prefecture University, Japan
Stefan Siegmund, Technische Universität Dresden, Germany

Contents

Part I
Papers by Plenary Speakers

Stable Bifurcations in Multi-species Semelparous Population Models

Ryusuke Kon

Abstract It is known that the behavior of a nonlinear semelparous Leslie matrix model with the basic reproduction number close to one can be approximated by a solution of a Lotka-Volterra differential equation. Furthermore, even in multi-species cases, a similar approximation works as long as every species is semelparous. This paper gives a mathematical basis to this approximation and shows that Lotka-Volterra equations are helpful to study a certain bifurcation problem of multi-species semelparous population models. With the help of this approximation method, we find an example of coexistence of two biennial populations with temporal segregation. This example provides a new mechanism of producing population cycles.

Keywords Lotka-Volterra equations · Leslie matrix models · Bifurcation · Semelparity · Population cycles · Temporal segregation

1 Introduction

A species is said to be *semelparous* if it reproduces only once immediately before death. Semelparous species are often observed in insects. In order to reveal a mechanism of producing population cycles observed in insect populations, Bulmer [1] studied a nonlinear semelparous Leslie matrix model, which is an age-structured population model for a semelparous species. One of the important conclusions of this study is that population cycles occur if competition is more severe between than within age-classes. After Bulmer [1], several papers have studied the dynamics of nonlinear semelparous Leslie matrix models (e.g., see [2–6, 8, 11, 12, 15, 17]). In particular, the papers [2–5] focus on bifurcations that occur around the extinction (or population free) equilibrium and provide a clear mathematical formula expressing Bulmer's conclusion.

R. Kon (✉)
Faculty of Engineering, University of Miyazaki, Gakuen Kibanadai Nishi 1-1, Miyazaki 889-2192, Japan
e-mail: konr@cc.miyazaki-u.ac.jp

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_1

In such bifurcation studies, the basic reproduction number $\mathscr{R}_0$ is used as a bifurcation parameter. Since $\mathscr{R}_0$ represents the expected number of newborns reproduced by an individual over a lifetime, population persistence is unlikely if $\mathscr{R}_0 < 1$ and is likely if $\mathscr{R}_0 > 1$. In fact, the extinction equilibrium of a nonlinear semelparous Leslie matrix model is stable if $\mathscr{R}_0 < 1$ and is unstable if $\mathscr{R}_0 > 1$. Therefore, at the critical point $\mathscr{R}_0 = 1$, a branch of positive equilibria is expected to bifurcate from the extinction equilibrium. The papers [2–5] provide a condition for the existence and the stability of such a positive bifurcating branch. Cushing and Li [2] focus on a two-age-class semelparous Leslie matrix model and provide a condition for stable bifurcations of positive equilibria (see also Cushing [3]). Furthermore, it is also shown that if a branch of positive equilibria is unstable, a stable branch of 2-cycles bifurcates from the extinction equilibrium. Therefore, an occurrence of population cycles is predicted by the instability of bifurcating positive equilibria. These studies are extended to the case where the number of age-classes is more than two. Cushing [4] classifies the possible types of bifurcation in a three-age-class case. Furthermore, Cushing and Henson [5] provides a condition for stable bifurcations of positive equilibria that is applicable even if the number of age-classes is arbitrary large.

The purpose of this paper is to provide a simple method of dealing with such a bifurcation problem of nonlinear semelparous Leslie matrix models. The method is motivated by the study of Diekmann and van Gils [7], who showed that a solution of a nonlinear semelparous Leslie matrix model can be approximated by that of a Lotka-Volterra (differential) equation. Our method shows that the stability of bifurcating positive equilibria can be evaluated by that of positive equilibria of Lotka-Volterrra equations. That is, our bifurcation problem can be reduced to a stability problem of Lotka-Volterra equations. Since a solution of a multi-species semelparous population model can also be approximated by that of a Lotka-Volterra equation [13, 14], we develop our method in the form applicable to multi-species models. With this method, we rediscover the result of Cushing and Henson [5] on a nonlinear semelparous Leslie matrix model. Furthermore, our method allows us to study high dimensional multi-species semelparous population models and to construct an example of population cycles in a competitive system of two biennial populations without assuming severe between-age-class competition. The population cycle occurs as a result of temporal segregation caused by severe age-specific species competition. This example provides a new mechanism of population cycles.

This paper is organized as follows. Section 2 introduces a multi-species semelparous population model, which is constructed by coupling multiple semelparous Leslie matrix models. Section 3 develops a bifurcation theory for a Kolmogorov difference equation, and shows that a certain bifurcation problem of Kolmogorov difference equations can be reduced to a stability problem of Lotka-Volterra equations. In order to apply the bifurcation theory to our bifurcation problem, Sect. 4 shows that a multi-species semelparous population model can be transformed to a Kolmogorov difference equation, and Sect. 5 specifies the stability problem of Lotka-Volterra equations that we need to examine. Section 6 shows that the derived stability problem of Lotka-Volterra equations can be reduced to a stability problem of lower dimensional Lotka-Volterra equations if lifespans of species, which are positive integers, are pairwise coprime. Section 7 examines the case where such a reduction does not work and constructs an example that age-specific species interactions have an essential impact on the stability of population dynamics. The example provides a new mechanism of population cycles. Section 8 includes a concluding remark.

2 Multi-species Semelparous Population Models

Let $N \geq 1$ be the number of species. Suppose that species i has n_i (≥ 2) age-classes. Then there are $n_1 + n_2 + \cdots + n_N =: n$ age-classes in total. We consider the interaction among N species expressed by the following n-dimensional nonlinear difference equation

$$\begin{cases} u_{[i,1],k+1} &= f_i \sigma_{[i,n_i]}(\mathbf{u}_k) u_{[i,n_i],k} \\ u_{[i,2],k+1} &= s_{[i,1]} \sigma_{[i,1]}(\mathbf{u}_k) u_{[i,1],k} \\ &\vdots \\ u_{[i,n_i],k+1} &= s_{[i,n_i-1]} \sigma_{[i,n_i-1]}(\mathbf{u}_k) u_{[i,n_i-1],k} \end{cases} \qquad i = 1, 2, \ldots, N. \tag{1}$$

Here $\mathbf{u}_k = (u_{1,k}, u_{2,k}, \ldots, u_{n,k})^\top$ (the symbol $\top$ is used for vector or matrix transpose) and for $i \in \{1, 2, \ldots, N\}$ the following notation is used to simplify the expression:

$$[i, j] := n_0 + n_1 + \cdots + n_{i-1} + j,$$

where $n_0 = 0$ and $j \in \{1, 2, \ldots, n_i\}$. Therefore, for example, $\mathbf{u}_k$ is also written as

$$\mathbf{u}_k = (\underbrace{u_{[1,1],k}, \ldots, u_{[1,n_1],k}}_{n_1}, \underbrace{u_{[1,2],k}, \ldots, u_{[2,n_2],k}}_{n_2}, \ldots, \underbrace{u_{[N,1],k}, \ldots, u_{[N,n_N],k}}_{n_N})^\top.$$

The variable $u_{[i,j],k}$ denotes the number of individuals of age $j \in \{1, 2, \ldots, n_i\}$ of species $i \in \{1, 2, \ldots, N\}$ at time $k \in \{0, 1, 2, \ldots\}$. The vital rates $f_i \sigma_{[i,n_i]}$ and $s_{[i,j]} \sigma_{[i,j]}$ denote the number of newborns produced by an individual of age n_i of species i and the probability that an individual of age j of species i survives one unit of time, respectively. It is assumed that each species has a single reproductive age-class. Thus each species is assumed to be semelparous. The ability of each individual of age j of species i is characterized by a single vital rate, either $f_i \sigma_{[i,n_i]}$ or $s_{[i,j]} \sigma_{[i,j]}$. It is assumed that f_i and $s_{[i,j]}$ are positive constants and $\sigma_{[i,j]}$ is a positive function of the population vector $\mathbf{u}_k$. We normalize the functions $\sigma_{[i,j]}$ by $\sigma_{[i,j]}(\mathbf{0}) = 1$. This implies that the constants f_i and $s_{[i,j]}$ represent vital rates at low population sizes, and thus the functions $\sigma_{[i,j]}$ solely determine how the vital rates depend on (both conspecific and allospecific) population sizes. Under these assumptions, the nonnegative cone

$$\mathbb{R}^n_+ := \{(u_1, u_2, \ldots, u_n)^\top \in \mathbb{R}^n : u_i \geq 0 \text{ for all } i \in \{1, 2, \ldots, n\}\}$$

is forward invariant, i.e., $\mathbf{u}_k \in \mathbb{R}^n_+$ for all $k \geq 1$ if $\mathbf{u}_0 \in \mathbb{R}^n_+$.

If $N = 1$, then system (1) is reduced to a nonlinear semelparous Leslie matrix model, which is for instance studied in [1–6, 8, 11, 12, 15, 17].

3 Bifurcations in Kolmogorov Difference Equations

This section considers a bifurcation problem of the Kolmogorov difference equation

$$x_{i,k+1} = x_{i,k} g_i(\varepsilon, \mathbf{x}_k), \quad i = 1, 2, \ldots, n, \tag{2}$$

where $\mathbf{x}_k = (x_{1,k}, x_{2,k}, \ldots, x_{n,k})^\top$. This difference equation has a parameter $\varepsilon \in \mathbb{R}$. We assume that each g_i is a C^2 function defined in a neighborhood of $(0, \mathbf{0}) \in \mathbb{R} \times \mathbb{R}^n$ and satisfies $g_i(0, \mathbf{0}) = 1$. A vector is said to be *positive* (resp. *negative*) if all its components are positive (resp. negative). We are concerned with the positive equilibria of system (2), which are given by the positive vectors $\mathbf{x}$ satisfying the equation $\mathbf{g}(\varepsilon, \mathbf{x}) = \mathbf{1}$, where $\mathbf{1}$ is a column vector whose components are all 1 and $\mathbf{g}(\varepsilon, \mathbf{x}) := (g_1(\varepsilon, \mathbf{x}), g_2(\varepsilon, \mathbf{x}), \ldots, g_n(\varepsilon, \mathbf{x}))^\top$. We shall construct a positive equilibrium of system (2) near the origin $\mathbf{0}$ and show that such a positive equilibrium has the same stability property as a positive equilibrium of the Lotka-Volterra equation

$$\frac{dx_i}{dt} = x_i\left(r_i + \sum_{j=1}^{n} a_{ij} x_j\right), \quad i = 1, 2, \ldots, n, \tag{3}$$

where

$$r_i := \frac{\partial g_i}{\partial \varepsilon}(0, \mathbf{0}), \quad a_{ij} := \frac{\partial g_i}{\partial x_j}(0, \mathbf{0}).$$

The positive equilibria of system (3) are given by the positive vectors $\mathbf{x}$ satisfying the linear equation $\mathbf{r} + A\mathbf{x} = \mathbf{0}$, where $\mathbf{r} := (r_1, r_2, \ldots, r_n)^\top$ and $A := (a_{ij})$. If A is nonsingular, i.e., $\det A \neq 0$, then the equation has the unique solution $\mathbf{x}^* := -A^{-1}\mathbf{r}$, which might not be positive. In the following theorems, the matrix A is always assumed nonsingular. Since the situation that the equality $\det A = 0$ holds is negligible, the nonsingularity assumption does not impose significant restrictions on our results.

Theorem 1 *Suppose that A is nonsingular. Then there exists a constant $\varepsilon_0 > 0$ and a unique function $\hat{\mathbf{x}} : (-\varepsilon_0, \varepsilon_0) \to \mathbb{R}^n$ satisfying $\hat{\mathbf{x}}(0) = \mathbf{0}$ and $\mathbf{g}(\varepsilon, \hat{\mathbf{x}}(\varepsilon)) = \mathbf{1}$. Furthermore, if $\mathbf{g}$ is a C^d function ($d \geq 1$), then so is $\hat{\mathbf{x}}$.*

Proof By assumption, $\frac{\partial \mathbf{g}}{\partial \mathbf{x}}(0, \mathbf{0}) = A$ is nonsingular. Thus the conclusion of this theorem is an immediate consequence of the implicit function theorem. □

It is clear that $\hat{\mathbf{x}}$ is an equilibrium of system (2). Furthermore, around $\varepsilon = 0$, the function $\hat{\mathbf{x}}$ is written in the form

$$\begin{aligned}\hat{\mathbf{x}}(\varepsilon) &= \hat{\mathbf{x}}(0) + \varepsilon \frac{d\hat{\mathbf{x}}}{d\varepsilon}(0) + O(\varepsilon^2)\\ &= \varepsilon \mathbf{x}^* + O(\varepsilon^2)\end{aligned}$$

since $\hat{\mathbf{x}}(0) = \mathbf{0}$ and $\frac{d\hat{\mathbf{x}}}{d\varepsilon}(0) = -(\frac{\partial \mathbf{g}}{\partial \mathbf{x}}(0, \mathbf{0}))^{-1} \frac{\partial \mathbf{g}}{\partial \varepsilon}(0, \mathbf{0}) = -A^{-1}\mathbf{r}$. Therefore, a branch of positive equilibria of system (2) bifurcates from the origin as increasing (resp. decreasing) ε through $\varepsilon = 0$ if $\mathbf{x}^*$ is positive (resp. negative). That is, the bifurcation is to the right if $\mathbf{x}^* > \mathbf{0}$ and to the left if $\mathbf{x}^* < \mathbf{0}$.

The Jacobi matrix of system (3) evaluated at $\mathbf{x}^*$ is given by $\mathrm{diag}(\mathbf{x}^*)A$, where $\mathrm{diag}(\mathbf{x}^*)$ denotes the diagonal matrix

$$\begin{pmatrix} x_1^* & 0 & \cdots & 0 \\ 0 & x_2^* & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_n^* \end{pmatrix}.$$

The following theorem shows that the stability of $\hat{\mathbf{x}}$ constructed in Theorem 1 can be evaluated by the stability of $\mathrm{diag}(\mathbf{x}^*)A$. A matrix is said to be *stable* if all its eigenvalues have negative real part. For convenience, we denote the *stability modulus* of a matrix M by

$$s(M) := \max\{\mathfrak{Re}\,\lambda : \lambda \text{ is an eigenvalue of M}\},$$

where $\mathfrak{Re}\,\lambda$ denotes the real part of λ. Then M is stable if and only if $s(M) < 0$.

Theorem 2 *If $s(\mathrm{diag}(\mathbf{x}^*)A) < 0$ (resp. $s(\mathrm{diag}(\mathbf{x}^*)A) > 0$), then the equilibrium $\hat{\mathbf{x}}(\varepsilon)$ of system (2) is asymptotically stable (resp. unstable) for all sufficiently small $\varepsilon > 0$.*

Proof Since $\hat{\mathbf{x}}(\varepsilon)$ satisfies $\mathbf{g}(\varepsilon, \hat{\mathbf{x}}(\varepsilon)) = \mathbf{1}$, the Jacobi matrix of system (2) evaluated at $\hat{\mathbf{x}}(\varepsilon)$ is

$$\begin{aligned}J(\hat{\mathbf{x}}(\varepsilon)) &:= \frac{\partial}{\partial \mathbf{x}} \mathrm{diag}(\mathbf{x})\mathbf{g}(\varepsilon, \mathbf{x})\Big|_{\mathbf{x}=\hat{\mathbf{x}}(\varepsilon)}\\ &= \left(\mathrm{diag}(\mathbf{g}(\varepsilon, \mathbf{x})) + \mathrm{diag}(\mathbf{x})\frac{\partial \mathbf{g}}{\partial \mathbf{x}}(\varepsilon, \mathbf{x})\right)\Big|_{\mathbf{x}=\hat{\mathbf{x}}(\varepsilon)}\\ &= I + \mathrm{diag}\left(\hat{\mathbf{x}}(\varepsilon)\right)\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\left(\varepsilon, \hat{\mathbf{x}}(\varepsilon)\right),\end{aligned}$$

where I is the identity matrix. Around $\varepsilon = 0$, this is written in the form

$$J(\hat{\mathbf{x}}(\varepsilon)) = I + \varepsilon\ (\mathrm{diag}(\mathbf{x}^*)A + O(\varepsilon)).$$

Suppose that $s(\mathrm{diag}(\mathbf{x}^*)A) < 0$. Then because of continuous dependence of eigenvalues of a matrix on its entries, there exists a constant $\varepsilon_s \in (0, \varepsilon_0)$ such that every eigenvalue $\lambda(\varepsilon)$ of $\mathrm{diag}(\mathbf{x}^*)A + O(\varepsilon)$ satisfies

$$\left|\lambda(\varepsilon) + \frac{1}{\varepsilon}\right| < \frac{1}{\varepsilon}$$

for all $\varepsilon \in (0, \varepsilon_s)$. This inequality represents the situation that the disk centered at $-\frac{1}{\varepsilon}$ with radius $\frac{1}{\varepsilon}$ contains all eigenvalues of $\mathrm{diag}(\mathbf{x}^*)A + O(\varepsilon)$ on the complex plane. Since $\varepsilon > 0$, the inequality is reduced to $|1 + \varepsilon\lambda(\varepsilon)| < 1$, which implies that the spectral radius of $J(\hat{\mathbf{x}}(\varepsilon))$ is less than one. Therefore, $\hat{\mathbf{x}}(\varepsilon)$ is asymptotically stable for all $\varepsilon \in (0, \varepsilon_s)$.

Suppose that $s(\mathrm{diag}(\mathbf{x}^*)A) > 0$. Then $\mathrm{diag}(\mathbf{x}^*)A$ has an eigenvalue λ_u with positive real part and $\mathrm{diag}(\mathbf{x}^*)A + O(\varepsilon)$ has an eigenvalue $\lambda(\varepsilon)$ satisfying $\lambda(\varepsilon) \to \lambda_u$ as $\varepsilon \to 0$. Therefore, there exists a constant $\varepsilon_u \in (0, \varepsilon_0)$ such that

$$\left|\lambda(\varepsilon) + \frac{1}{\varepsilon}\right| > \frac{1}{\varepsilon}$$

holds for all $\varepsilon \in (0, \varepsilon_u)$. Since $\varepsilon > 0$, the inequality is equivalent to $|1 + \varepsilon\lambda(\varepsilon)| > 1$, which implies that the spectral radius of $J(\hat{\mathbf{x}}(\varepsilon))$ is larger than one. Therefore, $\hat{\mathbf{x}}(\varepsilon)$ is unstable for all $\varepsilon \in (0, \varepsilon_u)$. □

4 Derivation of Kolmogorov Difference Equations from System (1)

Define the *basic reproduction number* $\mathscr{R}_0^i$ for species i by $\mathscr{R}_0^i := s_{[i,1]}s_{[i,2]} \dots s_{[i,n_i-1]}f_i$. This number represents the expected number of newborns reproduced by an individual of species i per lifetime at low population sizes. We are concerned with a bifurcation that occurs in system (1) at $\mathscr{R}_0^1 = \mathscr{R}_0^2 = \cdots = \mathscr{R}_0^N = 1$. Since it is difficult to treat $\mathscr{R}_0^1, \mathscr{R}_0^2, \dots, \mathscr{R}_0^N$ as multiple independent bifurcation parameters, we only consider the case where the vector $(\mathscr{R}_0^1, \mathscr{R}_0^2, \dots, \mathscr{R}_0^N)^\top$ changes along a certain one-dimensional manifold and treat a one-parameter bifurcation problem. More precisely, in order to make the derived Lotka-Volterra equation simple, we choose an arbitrary fixed vector $\mathbf{c} = (c_1, c_2, \dots, c_N)^\top$ and consider bifurcations by changing the parameters $\mathscr{R}_0^1, \mathscr{R}_0^2, \dots, \mathscr{R}_0^N$ with maintaining the relation

$$\left(\log(\mathscr{R}_0^1)^{\frac{\nu}{n_1}}, \log(\mathscr{R}_0^2)^{\frac{\nu}{n_2}}, \dots, \log(\mathscr{R}_0^N)^{\frac{\nu}{n_N}}\right)^\top = \varepsilon(c_1, c_2, \dots, c_N)^\top, \quad \varepsilon \in \mathbb{R} \tag{4}$$

where ν is the least common multiple of $n_1, n_2, \dots, n_N$, and thus species i experiences $\frac{\nu}{n_i}$ generations within ν time steps and $(\mathscr{R}_0^i)^{\frac{\nu}{n_i}}$ represents the expected number of descendants of species i per individual per ν time step at low population sizes. Since the above relation is assumed to be always satisfied when we consider a bifurcation problem of system (1), the new parameter ε instead of $\mathscr{R}_0^1, \mathscr{R}_0^2, \dots, \mathscr{R}_0^N$ shall be used as a bifurcation parameter. Although our approach is practically sufficient to examine the dynamics of system (1) with the parameters around $\mathscr{R}_0^1 = \mathscr{R}_0^2 = \cdots = \mathscr{R}_0^N = 1$, there could exist exceptional cases that our approach is unable to treat (see Sect. 9). Note that increase of ε implies increase of $\mathscr{R}_0^i$ if $c_i > 0$ and decrease of $\mathscr{R}_0^i$ if $c_i < 0$. To include ε as an explicit parameter of system (1), we replace f_i by $\frac{e^{\frac{c_i n_i}{\nu}\varepsilon}}{s_{[i,1]}s_{[i,2]}\cdots s_{[i,n_i-1]}}$. Then system (1) becames

$$\begin{cases} u_{[i,1],k+1} = \frac{e^{\frac{c_i n_i}{\nu}\varepsilon}}{s_{[i,1]}s_{[i,2]}\cdots s_{[i,n_i-1]}}\sigma_{[i,n_i]}(\mathbf{u}_k)u_{[i,n_i],k} \\ u_{[i,2],k+1} = s_{[i,1]}\sigma_{[i,1]}(\mathbf{u}_k)u_{[i,1],k} \\ \qquad\vdots \\ u_{[i,n_i],k+1} = s_{[i,n_i-1]}\sigma_{[i,n_i-1]}(\mathbf{u}_k)u_{[i,n_i-1],k} \end{cases} \qquad i = 1, 2, \dots, N. \tag{5}$$

Define

$$D_i := \operatorname{diag}(1, s_{[i,1]}, s_{[i,1]}s_{[i,2]}, \dots s_{[i,1]}s_{[i,2]} \dots s_{[i,n_i-1]}), \quad i = 1, 2, \dots, N,$$

and $D := \operatorname{diag}(D_1, D_2, \dots, D_N)$. The rescaling of system (5) with $\mathbf{x} := D^{-1}\mathbf{u}$ gives

$$\begin{cases} x_{[i,1],k+1} = e^{\frac{c_i n_i}{\nu}\varepsilon}\sigma_{[i,n_i]}(D\mathbf{x}_k)x_{[i,n_i],k} \\ x_{[i,2],k+1} = \sigma_{[i,1]}(D\mathbf{x}_k)x_{[i,1],k} \\ \qquad\vdots \\ x_{[i,n_i],k+1} = \sigma_{[i,n_i-1]}(D\mathbf{x}_k)x_{[i,n_i-1],k} \end{cases} \qquad i = 1, 2, \dots, N.$$

Let π_i, $i = 1, 2, \dots, N$, be the cyclic permutation

$$\begin{pmatrix} [i,1] & [i,2] & \cdots & [i,n_i] \\ [i,n_i] & [i,1] & \cdots & [i,n_i-1] \end{pmatrix}$$

and P_{π_i} be its permutation matrix. The product of $\pi_1, \pi_2, \dots, \pi_N$ is denoted by π and its permutation matrix is denoted by P_π. Define

$$S_i(\varepsilon, \mathbf{x}) := P_{\pi_i}\operatorname{diag}(\sigma_{[i,1]}(D\mathbf{x}), \dots, \sigma_{[i,n_i-1]}(D\mathbf{x}), e^{\frac{c_i n_i}{\nu}\varepsilon}\sigma_{[i,n_i]}(D\mathbf{x})), \quad i = 1, 2, \dots, N,$$

and $S(\varepsilon, \mathbf{x}) := \operatorname{diag}(S_1(\varepsilon, \mathbf{x}), S_2(\varepsilon, \mathbf{x}), \dots, S_N(\varepsilon, \mathbf{x}))$. Then the rescaled equation is written as

$$\mathbf{x}_{k+1} = S(\varepsilon, \mathbf{x}_k)\mathbf{x}_k.$$

Let $\boldsymbol{\xi}(\mathbf{x}) := S(\varepsilon, \mathbf{x})\mathbf{x}$. Since ν is a common multiple of the periods of the cyclic permutations $\pi_1, \pi_2, \ldots, \pi_N$, the matrix $G(\varepsilon, \mathbf{x}) := S(\varepsilon, \boldsymbol{\xi}^{\nu-1}(\mathbf{x})) \ldots S(\varepsilon, \boldsymbol{\xi}(\mathbf{x}))S(\varepsilon, \mathbf{x})$ is diagonal. Thus the map $\boldsymbol{\xi}^{\nu}$, i.e.,

$$\mathbf{y}_{k+1} = G(\varepsilon, \mathbf{y}_k)\mathbf{y}_k \tag{6}$$

is a Kolmogorov difference equation. The behavior of $\mathbf{y}_k$ shows the stroboscopic behavior of $\mathbf{x}_k$ with period ν.

5 Lotka-Volterra Equations

In the previous two sections, it was shown that system (1) is reduced to a Kolmogorov difference equation and its bifurcation problem is reduced to a stability problem of a Lotka-Volterra equation. In this section, we shall identify the Lotka-Volterra equation that we need to study.

Define the $n \times n$ matrix $B = (b_{ij})$ by

$$b_{ij} := \frac{\partial \sigma_i}{\partial u_j}(\mathbf{0}),$$

i.e., $B = \frac{\partial \boldsymbol{\sigma}}{\partial \mathbf{u}}(\mathbf{0})$, where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n)^\top$. The parameter $b_{[i,k][j,l]}$ represents the intensity of density dependent effect from age-class l of species j to age-class k of species i at low population sizes. The interaction between age-class k of species i and age-class l of species j is competitive if $b_{[i,k][j,l]} < 0$ and $b_{[j,l][i,k]} < 0$, mutualistic if $b_{[i,k][j,l]} > 0$ and $b_{[j,l][i,k]} > 0$, and antagonistic if $b_{[i,k][j,l]}b_{[j,l][i,k]} < 0$ at low population sizes. Let $\mathbf{g} = (g_1, g_2, \ldots, g_n)^\top$ be the diagonal entries of G defined in the previous section, i.e., $\text{diag}(\mathbf{g}(\varepsilon, \mathbf{x})) = G(\varepsilon, \mathbf{x})$. Then it is clear that $\mathbf{g}(0, \mathbf{0}) = \mathbf{1}$ holds. Furthermore, we have

$$\frac{\partial g_{[i,1]}}{\partial \varepsilon}(0, \mathbf{0}) = \frac{\partial g_{[i,2]}}{\partial \varepsilon}(0, \mathbf{0}) = \cdots = \frac{\partial g_{[i,n_i]}}{\partial \varepsilon}(0, \mathbf{0}) = c_i, \quad i = 1, 2, \ldots, N.$$

Thus g_i can be written as

$$g_i(\varepsilon, \mathbf{x}) = \exp\left(\varepsilon \frac{\partial g_i}{\partial \varepsilon}(0, \mathbf{0})\right) \prod_{k=0}^{\nu-1} \sigma_{\pi^k(i)}(D\boldsymbol{\xi}^k(\mathbf{x})), \quad i = 1, 2, \ldots, n,$$

whose partial derivative with respect to x_j evaluated at $(\varepsilon, \mathbf{x}) = (0, \mathbf{0})$ is

$$\begin{aligned}
\frac{\partial g_i}{\partial x_j}(0, \mathbf{0}) &= \frac{\partial}{\partial x_j} \exp\left(\varepsilon \frac{\partial g_i}{\partial \varepsilon}(0, \mathbf{0})\right) \prod_{k=0}^{\nu-1} \sigma_{\pi^k(i)}(D\boldsymbol{\xi}^k(\mathbf{x}))\Big|_{(\varepsilon,\mathbf{x})=(0,\mathbf{0})} \\
&= \sum_{l=0}^{\nu-1} \prod_{\substack{k=0\\k\neq l}}^{\nu-1} \sigma_{\pi^k(i)}(D\boldsymbol{\xi}^k(\mathbf{x})) \frac{\partial}{\partial x_j} \sigma_{\pi^l(i)}(D\boldsymbol{\xi}^l(\mathbf{x}))\Big|_{(\varepsilon,\mathbf{x})=(0,\mathbf{0})} \\
&= \sum_{l=0}^{\nu-1} \prod_{\substack{k=0\\k\neq l}}^{\nu-1} \sigma_{\pi^k(i)}(D\boldsymbol{\xi}^k(\mathbf{x})) \frac{\partial}{\partial \mathbf{u}} \sigma_{\pi^l(i)}(D\boldsymbol{\xi}^l(\mathbf{x})) \frac{\partial}{\partial x_j} D\boldsymbol{\xi}^l(\mathbf{x})\Big|_{(\varepsilon,\mathbf{x})=(0,\mathbf{0})} \\
&= \sum_{l=0}^{\nu-1} \prod_{\substack{k=0\\k\neq l}}^{\nu-1} \sigma_{\pi^k(i)}(D\boldsymbol{\xi}^k(\mathbf{x})) \left(\frac{\partial \boldsymbol{\sigma}}{\partial \mathbf{u}}(D\boldsymbol{\xi}^l(\mathbf{x})) D \frac{\partial \boldsymbol{\xi}^l}{\partial \mathbf{x}}\right)_{\pi^l(i),j} \Big|_{(\varepsilon,\mathbf{x})=(0,\mathbf{0})} \\
&= \sum_{l=0}^{\nu-1} (P_\pi^{-l} BDP_\pi^l)_{ij}.
\end{aligned}$$

This implies

$$\frac{\partial \mathbf{g}}{\partial \mathbf{x}}(0, \mathbf{0}) = BD + P_\pi^{-1} BDP_\pi + \cdots + P_\pi^{-\nu+1} BDP_\pi^{\nu-1}.$$

Thus Theorems 1 and 2 suggest that the Lotka-Voterra equation (3) satisfying

$$\begin{aligned}
\mathbf{r} &= (\underbrace{c_1, c_1, \ldots, c_1}_{n_1}, \underbrace{c_2, c_2, \ldots, c_2}_{n_2}, \ldots, \underbrace{c_N, c_N, \ldots, c_N}_{n_N})^\top \\
A &= BD + P_\pi^{-1} BDP_\pi + \cdots + P_\pi^{-\nu+1} BDP_\pi^{\nu-1}
\end{aligned} \tag{7}$$

is helpful to study our bifurcation problem of system (5).

Each parameter in (7) has an important biological meaning. The parameters $c_1, c_2, \ldots, c_N$ represent the ratio of $\log(\mathscr{R}_0^1)^{\frac{\nu}{n_1}}, \log(\mathscr{R}_0^2)^{\frac{\nu}{n_2}}, \ldots, \log(\mathscr{R}_0^N)^{\frac{\nu}{n_N}}$, in which the basic reproduction numbers are compared with the same time scale. By definition, the $([i, k], [j, l])$-entry of A is written as

$$a_{[i,k][j,l]} = \sum_{\Delta=0}^{\nu-1} \big(BD\big)_{\pi^\Delta([i,k])\pi^\Delta([j,l])}.$$

Since π is the product of the cyclic permutations $\pi_1, \pi_2, \ldots, \pi_N$, every entry of B that appears in the right-hand side of this equation has the first subscript belonging to $\{[i, 1], [i, 2], \ldots, [i, n_i]\}$ and the second subscript belonging to $\{[j, 1], [j, 2], \ldots,$

$[j, n_j]\}$. Therefore, every $a_{[i,k][j,l]}$, $k = 1, 2, \ldots, n_i$, $l = 1, 2, \ldots, n_j$, indicates the intensity of an effect of species j on species i at low population sizes. In the subsequent sections, we shall see that age-specific effects of density dependence between species i and j intricately depend on n_i and n_j.

The rest of this section provides some basic properties of system (3) satisfying (7).

Lemma 1 *The vector* $\mathbf{r}$ *and the matrix* A *defined by* (7) *satisfy* $P_\pi \mathbf{r} = \mathbf{r}$ *and* $P_\pi A P_\pi^{-1} = A$.

Proof It is clear that the first equality holds. Since $P_\pi^\nu = P_\pi^{-\nu} = I$, we have $P_\pi BDP_\pi^{-1} = P_\pi^{-\nu+1} BDP_\pi^{\nu-1}$. Thus

$$\begin{aligned} P_\pi A P_\pi^{-1} &= P_\pi (BD + P_\pi^{-1} BDP_\pi + \cdots + P_\pi^{-\nu+1} BDP_\pi^{\nu-1}) P_\pi^{-1} \\ &= A, \end{aligned}$$

which shows that the second equality holds. □

Define the $N \times n$ matrix $T = (t_{ij})$ by

$$t_{ij} := \begin{cases} 1, & j \in \{[i,1], [i,2], \ldots, [i,n_i]\} \\ 0, & j \notin \{[i,1], [i,2], \ldots, [i,n_i]\}. \end{cases}$$

For an $n \times n$ matrix $M = (m_{ij})$, define the $N \times N$ matrix $\bar{M} = (\bar{m}_{ij})$ by

$$\bar{m}_{ij} := \frac{1}{n_i n_j} \sum_{k=1}^{n_i} \sum_{l=1}^{n_j} m_{[i,k][j,l]},$$

similarly, for an n-dimensional vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)^\top$, define the N-dimensional vector $\bar{\mathbf{v}} = (\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_N)^\top$ by

$$\bar{v}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} v_{[i,k]}.$$

Lemma 2 *Let* $\mathbf{r}$ *and* A *be the vector and the matrix defined by* (7). *Suppose that* A *is nonsingular. Then* $\mathbf{x}^* = -A^{-1}\mathbf{r}$ *satisfies*

$$x^*_{[i,1]} = x^*_{[i,2]} = \cdots = x^*_{[i,n_i]} = -\frac{1}{n_i} (\bar{A}^{-1} \bar{\mathbf{r}})_i, \quad i = 1, 2, \ldots, N.$$

Thus $\bar{\mathbf{r}} + \bar{A} T \mathbf{x}^* = \mathbf{0}$ *is fulfilled.*

Proof Multiplying the both sides of $\mathbf{r} + A\mathbf{x}^* = \mathbf{0}$ on the left by P_π, we have $P_\pi \mathbf{r} + P_\pi A P_\pi^{-1} P_\pi \mathbf{x}^* = \mathbf{0}$. Since $P_\pi \mathbf{r} = \mathbf{r}$ and $P_\pi A P_\pi^{-1} = A$ hold, we obtain $\mathbf{r} + A P_\pi \mathbf{x}^* = \mathbf{0}$. The same argument shows that $\mathbf{r} + A\mathbf{x}^* = \mathbf{r} + A P_\pi \mathbf{x}^* = \cdots = \mathbf{r} + A P_\pi^{\nu-1} \mathbf{x}^* = \mathbf{0}$, i.e., $A\mathbf{x}^* = A P_\pi \mathbf{x}^* = \cdots = A P_\pi^{\nu-1} \mathbf{x}^* = -\mathbf{r}$. Since A is nonsingular, $\mathbf{x}^* = P_\pi \mathbf{x}^* = \cdots = P_\pi^{\nu-1} \mathbf{x}^*$. This shows that $x^*_{[i,1]} = x^*_{[i,2]} = \cdots = x^*_{[i,n_i]}$ for each $i \in \{1, 2, \ldots, N\}$. Then we further obtain $\mathbf{0} = \overline{\mathbf{r} + A\mathbf{x}^*} = \bar{\mathbf{r}} + \overline{A\mathbf{x}^*} = \bar{\mathbf{r}} + \bar{A} T \mathbf{x}^*$. □

If $N = 1$, then π is a cyclic permutation of $\{1, 2, \ldots, n\}$. Thus all components of $\mathbf{r}$ are identical and A is a circulant matrix. In [10], the Lotka-Volterra equation with such $\mathbf{r}$ and A is studied. It is called the *May-Leonard system* [16] if $n = 3$.

6 Stable Bifurcations in Multi-species Semelparous Models

By combing the results of the previous sections, we can establish theorems on bifurcations of positive equilibria of system (5). In the theorems of this section, we focus on the case $\mathbf{x}^* > \mathbf{0}$ since the case $\mathbf{x}^* < \mathbf{0}$ can be examined by changing the signs of $c_1, c_2, \ldots, c_N$.

Theorem 3 *Assume that $\boldsymbol{\sigma}$ is a C^2 function. Let $\mathbf{r}$ and A be the vector and the matrix defined by (7). Suppose that A is nonsingular and $\mathbf{x}^* = -A^{-1}\mathbf{r} > \mathbf{0}$. Then system (5) has a unique branch of positive equilibria bifurcating from the origin as increasing ε through $\varepsilon = 0$. The bifurcation is stable if $s(\mathrm{diag}(\mathbf{x}^*)A) < 0$ and is unstable if $s(\mathrm{diag}(\mathbf{x}^*)A) > 0$.*

Proof By Theorem 1, the map $\boldsymbol{\xi}^\nu$ has a unique branch of positive equilibria written in the form $\hat{\mathbf{x}}(\varepsilon) = \varepsilon \mathbf{x}^* + O(\varepsilon^2)$. It is obvious that all of $\hat{\mathbf{x}}, \boldsymbol{\xi}(\hat{\mathbf{x}}), \ldots, \boldsymbol{\xi}^{\nu-1}(\hat{\mathbf{x}})$ are positive equilibria of $\boldsymbol{\xi}^\nu$ bifurcating from the origin. However, it is ensured that $\hat{\mathbf{x}}, \boldsymbol{\xi}(\hat{\mathbf{x}}), \ldots, \boldsymbol{\xi}^{\nu-1}(\hat{\mathbf{x}})$ are identical since a branch of positive equilibria of $\boldsymbol{\xi}^\nu$ bifurcating from the origin is unique. This implies that $\hat{\mathbf{x}}$ is a positive equilibrium of the map $\boldsymbol{\xi}$, i.e., system (5). The other statements follow from Theorem 2. □

In the rest of this section, we consider the sign of $s(\mathrm{diag}(\mathbf{x}^*)A)$. To derive the following results, a certain property of the integers $n_1, n_2, \ldots, n_N$ plays an important role. Two integers are said to be *coprime* if their greatest common divisor is 1. A set of integers is said to be *pairwise coprime* if every couple of different integers in this set is coprime.

Lemma 3 *Suppose that $M = (m_{ij})$ is an $n \times n$ matrix satisfying $P_\pi M P_\pi^{-1} = M$. If n_i and n_j are coprime for some disjoint $i, j \in \{1, 2, \ldots, N\}$, then there exists a constant μ such that $m_{[i,k][j,l]} = \mu$ for all $k \in \{1, 2, \ldots, n_i\}$ and $l \in \{1, 2, \ldots, n_j\}$.*

Proof Since π has the cycles visiting cyclically all elements of $\{[i, 1], [i, 2], \ldots, [i, n_i]\}$ and $\{[j, 1], [j, 2], \ldots, [j, n_j]\}$, respectively, it is sufficient to show that for every integer Δ there exists an integer k such that

$$m_{\pi^k([i,1])\pi^{k+\Delta}([j,1])} = m_{[i,1][j,1]}.$$

Let Δ be an arbitrary integer. Then there exists an integer s such that $\pi^{s+\Delta}([j, 1]) = [j, 1]$. Since n_i and n_j are coprime, $\{n_j, 2n_j, \ldots, n_i n_j\}$ is a complete system of incongruent residues of mod n_i (e.g., see [9, Theorem 56]). Therefore, there exists an integer t such that $\pi^{tn_j+s}([i, 1]) = [i, 1]$. For $k = tn_j + s$, the desired equation is satisfied as follows:

$$\begin{aligned} m_{\pi^k([i,1])\pi^{k+\Delta}([j,1])} &= m_{\pi^{tn_j+s}([i,1])\pi^{tn_j+s+\Delta}([j,1])} \\ &= m_{[i,1]\pi^{tn_j}([j,1])} \\ &= m_{[i,1][j,1]}. \end{aligned}$$

□

For an $n \times n$ matrix M, we denote by M_{ij} the $n_i \times n_j$ submatrix of M with $\{[i, 1], [i, 2], \ldots, [i, n_i]\}$ and $\{[j, 1], [j, 2], \ldots, [j, n_j]\}$ as the sets of row and column indices, respectively. Write an $n_i \times n_j$ matrix M_{ij} (possibly $i = j$) in partitioned form

$$M_{ij} = \begin{pmatrix} m_{[i,1][j,1]} & \mathbf{q}_1[M_{ij}]^\top \\ \mathbf{q}_2[M_{ij}] & Q[M_{ij}] \end{pmatrix}.$$

Then we obtain the following lemma.

Lemma 4 *Assume that $\{n_1, n_2, \ldots, n_N\}$ is pairwise coprime. Suppose that $M = (m_{ij})$ is an $n \times n$ matrix satisfying $P_\pi M P_\pi^{-1} = M$. Then the characteristic equation of M is given by*

$$\det\Big(\lambda I - \mathrm{diag}(n_1, n_2, \ldots, n_N)\bar{M}\Big) \prod_{i=1}^{N} \det\Big(\lambda I + \mathbf{q}_2[M_{ii}]\mathbf{1}^\top - Q[M_{ii}]\Big) = 0.$$

Proof Define the $n_i \times n_i$ matrix H_i by

$$H_i := \begin{pmatrix} 1 & -\mathbf{1}^\top \\ \mathbf{0} & I \end{pmatrix},$$

which is nonsingular and its inverse is

$$H_i^{-1} = \begin{pmatrix} 1 & \mathbf{1}^\top \\ \mathbf{0} & I \end{pmatrix}.$$

Then $H_i^{-1} M_{ij} H_j$ is equivalent to

$$\begin{aligned}
&\begin{pmatrix} 1 & \mathbf{1}^\top \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} m_{[i,1][j,1]} & \mathbf{q}_1[M_{ij}]^\top \\ \mathbf{q}_2[M_{ij}] & Q[M_{ij}] \end{pmatrix} \begin{pmatrix} 1 & -\mathbf{1}^\top \\ \mathbf{0} & I \end{pmatrix} \\
&= \begin{pmatrix} 1 & \mathbf{1}^\top \\ \mathbf{0} & I \end{pmatrix} \begin{pmatrix} m_{[i,1][j,1]} & -m_{[i,1][j,1]}\mathbf{1}^\top + \mathbf{q}_1[M_{ij}]^\top \\ \mathbf{q}_2[M_{ij}] & -\mathbf{q}_2[M_{ij}]\mathbf{1}^\top + Q[M_{ij}] \end{pmatrix} \\
&= \begin{pmatrix} m_{[i,1][j,1]} + \mathbf{1}^\top \mathbf{q}_2[M_{ij}] & -m_{[i,1][j,1]}\mathbf{1}^\top + \mathbf{q}_1[M_{ij}]^\top - \mathbf{1}^\top \mathbf{q}_2[M_{ij}]\mathbf{1}^\top + \mathbf{1}^\top Q[M_{ij}] \\ \mathbf{q}_2[M_{ij}] & -\mathbf{q}_2[M_{ij}]\mathbf{1}^\top + Q[M_{ij}] \end{pmatrix} \\
&= \begin{pmatrix} m_{[i,1][j,1]} + \mathbf{1}^\top \mathbf{q}_2[M_{ij}] & \mathbf{0}^\top \\ \mathbf{q}_2[M_{ij}] & -\mathbf{q}_2[M_{ij}]\mathbf{1}^\top + Q[M_{ij}] \end{pmatrix},
\end{aligned}$$

where we used the fact that each column sum of M_{ij} is identical to obtain the last equality. By Lemma 3, if $i \neq j$, then there exists a constant μ_{ij} such that all entries of M_{ij} are equal to μ_{ij}. Thus if $i \neq j$, then

$$H_i^{-1} M_{ij} H_j = \begin{pmatrix} n_i \mu_{ij} & \mathbf{0}^\top \\ \mu_{ij}\mathbf{1} & O \end{pmatrix},$$

where O denotes the zero matrix. We define the block diagonal matrix $H := \operatorname{diag}(H_1, H_2, \ldots, H_N)$, whose inverse is $H^{-1} = \operatorname{diag}(H_1^{-1}, H_2^{-1}, \ldots, H_N^{-1})$. Then we have

$$H^{-1} H M = \left(\begin{array}{cc|cc|c|cc}
\gamma_1 & \mathbf{0}^\top & n_1\mu_{12} & \mathbf{0}^\top & \cdots & n_1\mu_{1N} & \mathbf{0}^\top \\
\mathbf{q}_2(M_{11}) & \Gamma_1 & \mu_{12}\mathbf{1} & O & \cdots & \mu_{1N}\mathbf{1} & O \\ \hline
n_2\mu_{21} & \mathbf{0}^\top & \gamma_2 & \mathbf{0}^\top & \cdots & n_2\mu_{2N} & \mathbf{0}^\top \\
\mu_{21}\mathbf{1} & O & \mathbf{q}_2(M_{22}) & \Gamma_2 & \cdots & \mu_{2N}\mathbf{1} & O \\ \hline
\vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \hline
n_N\mu_{N1} & \mathbf{0}^\top & n_N\mu_{N2} & \mathbf{0}^\top & \cdots & \gamma_N & \mathbf{0}^\top \\
\mu_{N1}\mathbf{1} & O & \mu_{N2}\mathbf{1} & O & \cdots & \mathbf{q}_2(M_{NN}) & \Gamma_N
\end{array}\right),$$

where $\gamma_i := m_{[i,1][i,1]} + \mathbf{1}^\top \mathbf{q}_2[M_{ii}]$ and $\Gamma_i := -\mathbf{q}_2[M_{ii}]\mathbf{1}^\top + Q[M_{ii}]$. Thus it is straightforward to show that $\det(\lambda I - H^{-1} M H)$ is equivalent to

$$\det\left(\lambda I - \operatorname{diag}(n_1, n_2, \ldots, n_N)\bar{M}\right) \prod_{i=1}^{N} \det\left(\lambda I + \mathbf{q}_2[M_{ii}]\mathbf{1}^\top - Q[M_{ii}]\right)$$

where the fact that

$$\begin{pmatrix} \gamma_1 & n_1\mu_{12} & \cdots & n_1\mu_{1N} \\ n_2\mu_{21} & \gamma_2 & \cdots & n_2\mu_{2N} \\ \vdots & \vdots & & \vdots \\ n_N\mu_{N1} & n_N\mu_{N2} & \cdots & \gamma_N \end{pmatrix} = \operatorname{diag}(n_1, n_2, \ldots, n_N)\bar{M}$$

is used. Thus we obtain the desired form of the characteristic equation of M. □

Lemma 4 with $N = 1$ shows that the characteristic equation of M_{ii} is equivalent to

$$\det(\lambda I - n_i \bar{m}_{ii}) \det(\lambda I + \mathbf{q}_2[M_{ii}]\mathbf{1}^\top - Q[M_{ii}]) = 0.$$

Therefore, $s(M_{ii})<0$ (resp. $s(M_{ii})>0$) if and only if both $s(Q[M_{ii}] - \mathbf{q}_2[M_{ii}]\mathbf{1}^\top) < 0$ and $\bar{m}_{ii} < 0$ (resp. either $s(Q[M_{ii}] - \mathbf{q}_2[M_{ii}]\mathbf{1}^\top) > 0$ or $\bar{m}_{ii} > 0$).

The following theorem shows that the stability of $\mathrm{diag}(\mathbf{x}^*)A$ can be evaluated by the stability of some matrices whose sizes are smaller than that of $\mathrm{diag}(\mathbf{x}^*)A$.

Theorem 4 *Assume that* $\{n_1, n_2, \ldots, n_N\}$ *is pairwise coprime. Let* $\mathbf{r}$ *and* A *be the vector and the matrix defined by* (7). *Suppose that* A *is nonsingular,* $\mathbf{x}^* = -A^{-1}\mathbf{r} > \mathbf{0}$, *and* $\bar{a}_{ii} < 0$, $i = 1, 2, \ldots, N$. *Then* $s(\mathrm{diag}(\mathbf{x}^*)A) < 0$ *if and only if all of* $s(\mathrm{diag}(T\mathbf{x}^*)\bar{A})$ *and* $s(A_{ii})$, $i = 1, 2, \ldots, N$, *are negative, and* $s(\mathrm{diag}(\mathbf{x}^*)A) > 0$ *if and only if some of* $s(\mathrm{diag}(T\mathbf{x}^*)\bar{A})$ *and* $s(A_{ii})$, $i = 1, 2, \ldots, N$, *are positive.*

Proof Since $\mathrm{diag}(\mathbf{x}^*)A$ satisfies $P_\pi \mathrm{diag}(\mathbf{x}^*)AP_\pi^{-1} = \mathrm{diag}(\mathbf{x}^*)P_\pi AP_\pi^{-1} = \mathrm{diag}(\mathbf{x}^*)A$, we can apply Lemma 4 to $\mathrm{diag}(\mathbf{x}^*)A$. Then the characteristic equation of $\mathrm{diag}(\mathbf{x}^*)A$ is equivalent to

$$\begin{aligned}
&\det(\lambda I - \mathrm{diag}(n_1, n_2, \ldots, n_N)\overline{\mathrm{diag}(\mathbf{x}^*)A}) \\
&\quad \times \prod_{i=1}^{N} \det\left(\lambda I + \mathbf{q}_2[(\mathrm{diag}(\mathbf{x}^*)A)_{ii}]\mathbf{1}^\top - Q[(\mathrm{diag}(\mathbf{x}^*)A)_{ii}]\right) \\
&= \det\left(\lambda I - \mathrm{diag}(T\mathbf{x}^*)\bar{A}\right) \prod_{i=1}^{N} \bar{x}_i^* \det\left(\frac{\lambda}{\bar{x}_i^*} I + \mathbf{q}_2[A_{ii}]\mathbf{1}^\top - Q[A_{ii}]\right) = 0.
\end{aligned}$$

This characteristic equation and the remark after Lemma 4 completes the proof. □

As we shall see in Sect. 8, the assumption that $\{n_1, n_2, \ldots, n_N\}$ is pairwise coprime is essential to derive the conclusion of Theorem 4. It is known that the probability that two integers are coprime is $6/\pi^2 \approx 0.6$ (see [9, Theorem 332]). Therefore, if a community is composed of randomly chosen two semelparous species, then the assumption of Theorem 4 is satisfied with the probability $6/\pi^2$. However, if the number of species is large, the probability becomes very small. A natural situation that Theorem 4 can apply might be found when we consider evolution of lifespans. Since consecutive integers are coprime, Theorem 4 is applicable if $n_1, n_2, \ldots, n_N$ are consecutive integers. This situation might happen if we consider an interaction among allied species that are produced by gradual evolution of lifespans.

7 Interpretation of Stability Conditions

7.1 The Sign of $\bar{a}_{ii}$

By definition, we obtain

$$\bar{a}_{ii} = \frac{\nu}{n_i^2}\Big(b_{[i,1][i,1]} + b_{[i,1][i,2]}s_{[i,1]} + \cdots + b_{[i,1][i,n_i]}s_{[i,1]}s_{[i,2]}\ldots s_{[i,n_i-1]}$$
$$+b_{[i,2][i,1]} + b_{[i,2][i,2]}s_{[i,1]} + \cdots + b_{[i,2][i,n_i]}s_{[i,1]}s_{[i,2]}\ldots s_{[i,n_i-1]}$$
$$+\cdots + b_{[i,n_i][i,1]} + b_{[i,n_i][i,2]}s_{[i,1]} + \cdots + b_{[i,n_i][i,n_i]}s_{[i,1]}s_{[i,2]}\ldots s_{[i,n_i-1]}\Big).$$

Since only the entries of the diagonal block B_{ii} of the matrix B appear in this form, $\bar{a}_{ii}$ represents a gross effect of conspecific density dependence within species i at low population sizes. Thus the assumption $\bar{a}_{ii} < 0$ implies density-dependent self-inhibition in species i at low population sizes. On the other hand, the inequality $\bar{a}_{ii} > 0$ implies positive density dependence, i.e., Allee effect, in species i. Note that even if $\bar{a}_{ii} < 0$, some $b_{[i,j][i,k]}$, $j, k \in \{1, 2, \ldots, n_1\}$, could be positive. Therefore, $\bar{a}_{ii} < 0$ does not simply imply that all interaction within species i are competitive.

7.2 The Sign of $s(A_{ii})$

Suppose that $N = 1$ and A is nonsingular. Then $n_1 = n$. Choose $c_1 = 1$. Then Lemma 2 shows that $x_1^* = x_2^* = \cdots = x_{n_1}^* = -\frac{1}{n_1\bar{a}_{11}}$. Because of this property, if $\bar{a}_{11} < 0$ then $\mathbf{x}^* > \mathbf{0}$ and the sign of $s(\mathrm{diag}(\mathbf{x}^*)A)$ is equivalent to that of $s(A)$. Therefore, under the assumption $\bar{a}_{11} < 0$, system (5) with $N = 1$ has a branch of positive equilibria bifurcating from the origin as increasing $\mathscr{R}_0^1$ through $\mathscr{R}_0^1 = 1$ and the bifurcation is stable (resp. unstable) if $s(A) < 0$ (resp. $s(A) > 0$). An application of this result to multi-species cases shows that, under the assumption $\bar{a}_{ii} < 0$, $i = 1, 2, \ldots, N$, $s(A_{ii}) < 0$ implies that each single-species subsystem has a stable bifurcation of positive equilibria when all species are isolated from each other.

If all interactions within species i are competitive, i.e., $b_{[i,j][i,k]} < 0$ for every $j, k \in \{1, 2, \ldots, n_i\}$, then $s(A_{ii}) < 0$ implies that, within species i, competition is more severe within than between age-classes. In fact, since the matrix A_{ii} is circulant, its eigenvalues are

$$\lambda_k = \sum_{j=0}^{n_i-1} \kappa_j e^{\frac{2\pi\sqrt{-1}}{n_i}jk} = \kappa_0 + \sum_{j=1}^{n_i-1} \kappa_j e^{\frac{2\pi\sqrt{-1}}{n_i}jk}, \quad k = 0, 1, \ldots, n_i - 1,$$

where $\sqrt{-1}$ denotes the imaginary unit and $(\kappa_0, \kappa_1, \dots, \kappa_{n_i-1})$ is the first row of the matrix A_{ii}, i.e.,

$$\begin{aligned}\kappa_0 &:= \frac{\nu}{n_i}\left(b_{[i,1][i,1]} + b_{[i,2][i,2]}s_{[i,1]} + \dots + b_{[i,n_i][i,n_i]}s_{[i,1]}s_{[i,2]}\dots s_{[i,n_i-1]}\right)\\ \kappa_1 &:= \frac{\nu}{n_i}\left(b_{[i,n_i][i,1]} + b_{[i,1][i,2]}s_{[i,1]} + \dots + b_{[i,n_i-1][i,n_i]}s_{[i,1]}s_{[i,2]}\dots s_{[i,n_i-1]}\right)\\ &\ \vdots\\ \kappa_{n_{i-1}} &:= \frac{\nu}{n_i}\left(b_{[i,2][i,1]} + b_{[i,3][i,2]}s_{[i,1]} + \dots + b_{[i,1][i,n_i]}s_{[i,1]}s_{[i,2]}\dots s_{[i,n_i-1]}\right).\end{aligned}$$

By definition, $s(A_{ii}) < 0$ if and only if $\mathfrak{Re}\,\lambda_k < 0$ for all $k = 0, 1, \dots, n_i - 1$. These inequalities clearly hold if competition between age-classes is weak, i.e., all $b_{[i,j][i,k]}$, $j \neq k$, are sufficiently small since $\kappa_0 < 0$ holds when all interaction within species i are competitive and κ_0 is independent of $b_{[i,j][i,k]}$, $j \neq k$. The same conclusion is obtained in [5] and its Table 1 gives exact stability criteria for $n_i = 2, 3, \dots, 6$.

7.3 The Sign of $s(\mathbf{Diag}(T\mathbf{x}^*)\bar{A})$

We shall show that if $s(\mathrm{diag}(T\mathbf{x}^*)\bar{A}) < 0$ (resp. $s(\mathrm{diag}(T\mathbf{x}^*)\bar{A}) > 0$), then the N-species community in system (5) is evaluated as stable (resp. unstable) when each species is assumed to be fixed at a certain age-distribution. Define the vector $\mathbf{d}_i$, $i = 1, 2, \dots, N$, by $\mathbf{d}_i = (1, s_{[i,1]}, \dots, s_{[i,1]}s_{[i,2]}\dots s_{[i,n_i-1]})^\top$. Then $\mathbf{d}_i$ is an eigenvector of the matrix

$$L_i = \begin{pmatrix} 0 & 0 \cdots & 0 & \frac{1}{s_{[i,1]}s_{[i,2]}\cdots s_{[i,n_i-1]}} \\ s_{[i,1]} & 0 \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 \cdots & s_{[i,n_i-1]} & 0 \end{pmatrix}$$

associated with its dominant eigenvalue 1. This matrix is a Leslie matrix for a semelparous population whose basic reproduction number is 1. Let H be the plane spanned by $\mathbf{d}_i$, $i = 1, 2, \dots, N$. Then since each $\mathbf{d}_i$ is an eigenvector of L_i, the plane H is invariant under the linearized system of (5) at the origin when $\varepsilon = 0$. Suppose that $\mathbf{u}_k$ is on the plane H. Then each species i has the age-distribution parallel to $\mathbf{d}_i$ at time k. Define $\mathbf{w}_k = (w_{1,k}, w_{2,k}, \dots, w_{N,k})^\top$ by $\mathbf{w}_k := TD^{-1}\mathbf{u}_k$. The i-th component of $\mathbf{w}_k$ denotes a weighted total population size of species i at time k. Since $\mathbf{u}_k$ is on the plane H, we have

$$u_{[i,1],k} = \frac{u_{[i,2],k}}{s_{[i,1]}} = \dots = \frac{u_{[i,n_i],k}}{s_{[i,1]}s_{[i,2]}\dots s_{[i,n_i-1]}} = \frac{w_{i,k}}{n_i}, \quad i = 1, 2, \dots, N.$$

By Eq.(5), the weighted total population size of species i at time $k+1$ is given by

$$w_{i,k+1} = h_i(\varepsilon, \mathbf{w}_k) w_{i,k},$$

where

$$h_i(\varepsilon, \mathbf{w}_k) := \frac{1}{n_i}\left(\sigma_{[i,1]}(\mathbf{u}_k) + \sigma_{[i,2]}(\mathbf{u}_k) + \cdots + e^{\frac{c_i n_i}{\nu}\varepsilon}\sigma_{[i,n_i]}(\mathbf{u}_k)\right).$$

Since $\mathbf{u}_{k+1}$ might not be on the plane H, $\mathbf{w}_{k+2}$ is not given by iterating this Kolmogorov difference equation. However, it is used to estimate the average effect of species interactions when each species i has the age-distribution parallel to $\mathbf{d}_i$. In fact, we obtain

$$\frac{\partial h_i}{\partial \varepsilon}(0, \mathbf{0}) = \frac{c_i}{\nu}, \quad \frac{\partial h_i}{\partial w_j}(0, \mathbf{0}) = \frac{\bar{a}_{ij}}{\nu},$$

which shows that the species interactions can be modeled by the N-dimensional Lotka-Volterra equation

$$\nu \frac{dy_i}{dt} = y_i\left(c_i + \sum_{j=1}^{N} \bar{a}_{ij} y_j\right), \quad i = 1, 2, \ldots, N$$

as long as $\varepsilon > 0$ is very small and each species i has the age-distribution parallel to $\mathbf{d}_i$. In this unstructured model, the N species coexist (resp. cannot coexist) stably at a positive equilibrium if $s(\mathrm{diag}(T\mathbf{x}^*)\bar{A}) < 0$ (resp. $s(\mathrm{diag}(T\mathbf{x}^*)\bar{A}) > 0$). Therefore, roughly speaking, Theorem 4 shows that the unstructured model derived above under the assumption that each species i has the fixed age-distribution parallel to $\mathbf{d}_i$ correctly evaluates the stability of bifurcations in system (5) if all species have stable dynamics when they are isolated from each other (i.e., $s(A_{ii}) < 0, i = 1, 2, \ldots, N$) and $\{n_1, n_2, \ldots, n_N\}$ is pairwise coprime.

8 Examples of Instability

Theorem 4 shows that, under the condition that $\{n_1, n_2, \ldots, n_N\}$ is pairwise coprime, $\bar{a}_{ii} < 0, i = 1, 2 \ldots, N$, and $\mathbf{x}^* = -A^{-1}\mathbf{r} > \mathbf{0}$, the stability problem of the positive equilibrium of (5) bifurcating from the origin is reduced to that of $N+1$ matrices, A_{ii}, $i = 1, 2, \ldots, N$, and $\mathrm{diag}(T\mathbf{x}^*)\bar{A}$. Since their sizes are usually much smaller than that of $\mathrm{diag}(\mathbf{x}^*)A$, this reduction is useful. However, if $\{n_1, n_2, \ldots, n_N\}$ is not pairwise coprime, this simple reduction does not work. This section focuses on this point. We shall show that system (5) can posses an unstable branch of positive equilibria even if all of A_{ii}, $i = 1, 2, \dots, N$, and $\mathrm{diag}(T\mathbf{x}^*)\bar{A}$ are stable.

To this end, we consider the case where $N = 2$ and $n_1 = n_2 = 2$. Then n_1 and n_2 are not coprime and their least common multiple is $\nu = 2$. The vector $\mathbf{r}$ and the matrix A given by (7) can be rewritten as

$$\mathbf{r} = \begin{pmatrix} c_1 \\ c_1 \\ c_2 \\ c_2 \end{pmatrix}, \quad A = \begin{pmatrix} -k_1 & -k_2 & -\alpha_1 & -\alpha_2 \\ -k_2 & -k_1 & -\alpha_2 & -\alpha_1 \\ -\alpha_3 & -\alpha_4 & -k_3 & -k_4 \\ -\alpha_4 & -\alpha_3 & -k_4 & -k_3 \end{pmatrix},$$

where every constant is assumed to be positive. Then $\bar{a}_{11} = -\frac{k_1+k_2}{2} < 0$ and $\bar{a}_{22} = -\frac{k_3+k_4}{2} < 0$ are satisfied.

Suppose that A is nonsingular. Then the equation $\mathbf{r} + A\mathbf{x} = \mathbf{0}$ has a unique solution $\mathbf{x}^*$. Note that we can control the sign of $\mathbf{x}^*$ by choosing suitable signs of c_1 and c_2. By Lemma 2, $\mathbf{x}^*$ is written as $(\frac{w_1}{2}, \frac{w_1}{2}, \frac{w_2}{2}, \frac{w_2}{2})^\top$, where $(w_1, w_2)^\top = T\mathbf{x}^*$. We shall show that $\mathrm{diag}(\mathbf{x}^*)A$ can be destabilized under the following assumption:

(A): $A_{11} = \begin{pmatrix} -k_1 & -k_2 \\ -k_2 & -k_1 \end{pmatrix}$, $A_{22} = \begin{pmatrix} -k_3 & -k_4 \\ -k_4 & -k_3 \end{pmatrix}$,
and $\mathrm{diag}(T\mathbf{x}^*)\bar{A} = \begin{pmatrix} -\frac{k_1+k_2}{2} w_1 & -\frac{\alpha_1+\alpha_2}{2} w_1 \\ -\frac{\alpha_3+\alpha_4}{2} w_2 & -\frac{k_3+k_4}{2} w_2 \end{pmatrix}$ are stable.

Since $\mathrm{tr}\, A_{11} < 0$ and $\mathrm{tr}\, A_{22} < 0$ are satisfied, the stability conditions for A_{11} and A_{22} are reduced to

$$k_1 > k_2 \quad \text{and} \quad k_3 > k_4. \tag{8}$$

By the definition of A, the inequality implies that in each species competition between age-classes are more severe than within age-classes. Furthermore, since $\mathrm{tr}\, \mathrm{diag}(T\mathbf{x}^*)\bar{A} < 0$ is satisfied, the condition for $s(\mathrm{diag}(T\mathbf{x}^*)\bar{A}) < 0$ is reduced to

$$(k_1 + k_2)(k_3 + k_4) > (\alpha_1 + \alpha_2)(\alpha_3 + \alpha_4), \tag{9}$$

which shows that competition between species is more severe than within species (see Sect. 7.3). In order to destabilize $\mathrm{diag}(\mathbf{x}^*)A$, let us examine the characteristic polynomial $\det(\lambda I - \mathrm{diag}(\mathbf{x}^*)A)$, which is reduced to

$$\det(\lambda I - \mathrm{diag}(T\mathbf{x}^*)\bar{A}) \det(\lambda I - \tilde{A}),$$

where

$$\tilde{A} := \frac{1}{2}\mathrm{diag}\Big(T\mathbf{x}^*\Big) \begin{pmatrix} -k_1 + k_2 & -\alpha_1 + \alpha_2 \\ -\alpha_3 + \alpha_4 & -k_3 + k_4 \end{pmatrix}.$$

Since $\mathrm{diag}(T\mathbf{x}^*)\bar{A}$ is assumed to be stable, $\mathrm{diag}(\mathbf{x}^*)A$ can be destabilized if $\tilde{A}$ can be destabilized. By Eq. (8), $\mathrm{tr}\, \tilde{A} < 0$ holds, but the sign of $\det \tilde{A}$ is not determined.

Therefore, if the set of parameters satisfying (8), (9) and

$$(k_1 - k_2)(k_3 - k_4) < (\alpha_1 - \alpha_2)(\alpha_3 - \alpha_4) \tag{10}$$

is nonempty, then system (5) can possess an unstable branch of a positive equilibria even if A_{11}, A_{22} and $\mathrm{diag}(T\mathbf{x}^*)\bar{A}$ are stable.

Figures 1 and 2 give such examples. In the examples, it is assumed the nonlinearity is of Beverton-Holt type

$$\sigma_i(\mathbf{u}) = \frac{1}{1 + (B\mathbf{u})_i}, \quad i = 1, 2, \ldots, n.$$

To construct an example of system (5) satisfying (8), (9), and (10), we need to determine s_1, s_2, c_1, c_2, and B. We suppose that $s_1 = s_2 = 0.9$ and $c_1 = c_2 = 1$ (i.e., $\mathscr{R}_0^1 = \mathscr{R}_0^2 = e^{\varepsilon}$). Furthermore, we suppose that

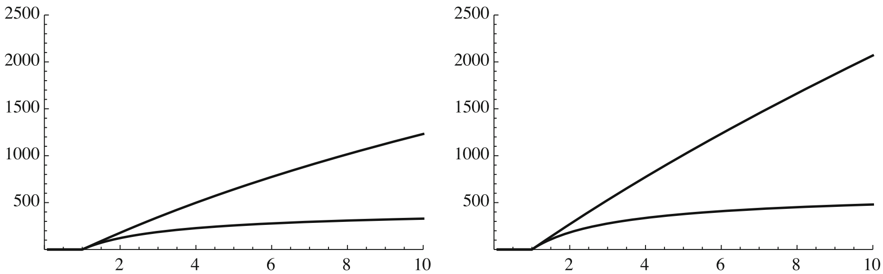


Fig. 1 Bifurcation diagram for system (5) with $N = 2$ and $n_1 = n_2 = 2$. In both panels, the horizontal axes denote e^{ε} $(= \mathscr{R}_0^1 = \mathscr{R}_0^2)$ and the vertical axes denote $u_1 + u_2$ and $u_3 + u_4$ in the left and right panels, respectively. The parameters are $s_1 = s_3 = 0.9$, $c_1 = c_2 = 1$, $k_1 = 4K$, $k_2 = 3K$, $k_3 = 3K$, $k_4 = 2K$, $\alpha_1 = 3K$, $\alpha_2 = K$, $\alpha_3 = 3K$, $\alpha_4 = K$, where $K = 10^{-3}$

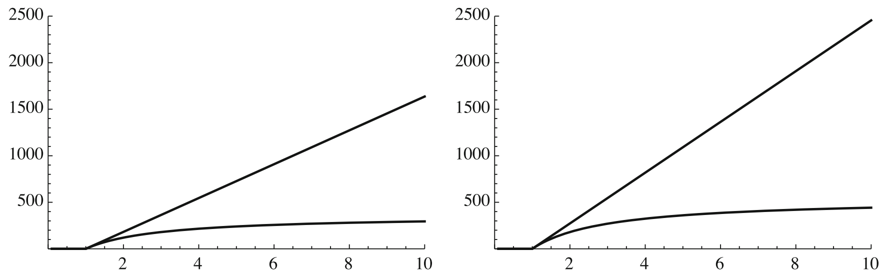


Fig. 2 Bifurcation diagram for system (5) with $N = 2$ and $n_1 = n_2 = 2$. In both panels, the horizontal axes denote e^{ε} $(= \mathscr{R}_0^1 = \mathscr{R}_0^2)$ and the vertical axes denote $u_1 + u_2$ and $u_3 + u_4$ in the left and right panels, respectively. The parameters are $s_1 = s_3 = 0.9$, $c_1 = c_2 = 1$, $k_1 = 4K$, $k_2 = 3K$, $k_3 = 3K$, $k_4 = 2K$, $\alpha_1 = K$, $\alpha_2 = 3K$, $\alpha_3 = K$, $\alpha_4 = 3K$, where $K = 10^{-3}$

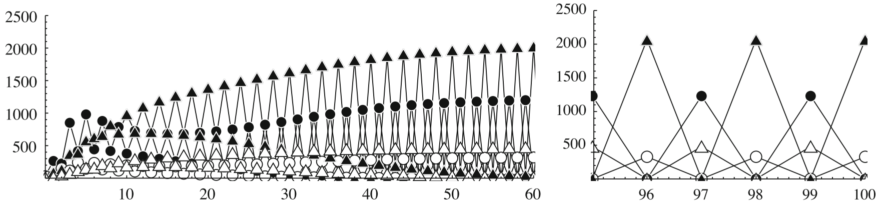


Fig. 3 Dynamics of system (5) with $N = 2$ and $n_1 = n_2 = 2$. The parameters are $s_1 = s_3 = 0.9$, $c_1 = c_2 = 1$, $e^{\varepsilon} = 10$ $(= \mathscr{R}_0^1 = \mathscr{R}_0^2)$, $k_1 = 4K$, $k_2 = 3K$, $k_3 = 3K$, $k_4 = 2K$, $\alpha_1 = 3K$, $\alpha_2 = K$, $\alpha_3 = 3K$, $\alpha_4 = K$, where $K = 10^{-3}$. The horizontal axes denote time k. The black and white circles denote $u_{1,k}$ and $u_{2,k}$, respectively. The black and white triangles denote $u_{3,k}$ and $u_{4,k}$, respectively. The left panel shows the transient dynamics and the right panel shows the ultimate dynamics

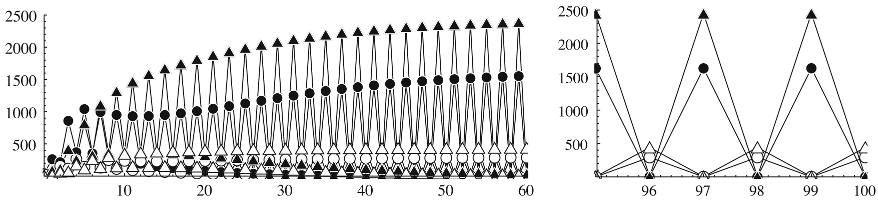


Fig. 4 Dynamics of system (5) with $N = 2$ and $n_1 = n_2 = 2$. The parameters are $s_1 = s_3 = 0.9$, $c_1 = c_2 = 1$, $e^{\varepsilon} = 10$ $(= \mathscr{R}_0^1 = \mathscr{R}_0^2)$, $k_1 = 4K$, $k_2 = 3K$, $k_3 = 3K$, $k_4 = 2K$, $\alpha_1 = K$, $\alpha_2 = 3K$, $\alpha_3 = K$, $\alpha_4 = 3K$, where $K = 10^{-3}$. The horizontal axes denote time k. The black and white circles denote $u_{1,k}$ and $u_{2,k}$, respectively. The black and white triangles denote $u_{3,k}$ and $u_{4,k}$, respectively. The left panel shows the transient dynamics and the right panel shows the ultimate dynamics

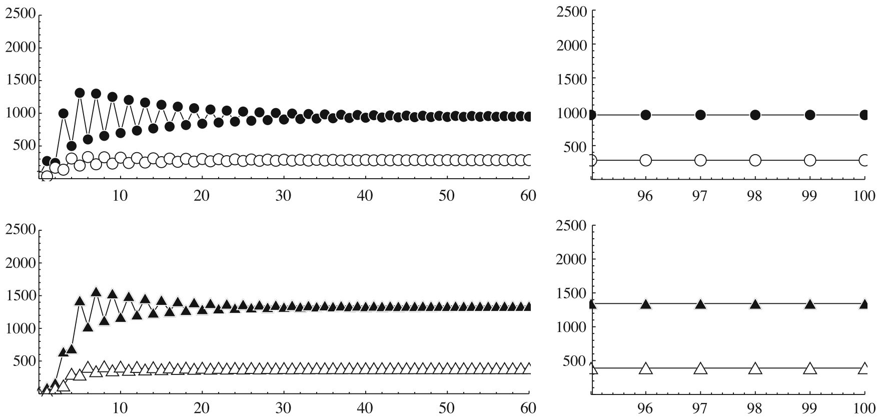


Fig. 5 Dynamics of system (5) with $N = 2$ and $n_1 = n_2 = 2$ when two species are isolated, i.e., $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$. All other parameters are the same as in Figs. 3 and 4. The horizontal axes denote time k. The black and white circles denote $u_{1,k}$ and $u_{2,k}$, respectively. The black and white triangles denote $u_{3,k}$ and $u_{4,k}$, respectively. The left panels show the transient dynamics and the right panels shows the ultimate dynamics

$$B = \frac{1}{2}\begin{pmatrix} -k_1 & -\frac{k_2}{s_1} & -\alpha_1 & -\frac{\alpha_2}{s_3} \\ -k_2 & -\frac{k_1}{s_1} & -\alpha_2 & -\frac{\alpha_1}{s_3} \\ -\alpha_3 & -\frac{\alpha_4}{s_1} & -k_3 & -\frac{k_4}{s_3} \\ -\alpha_4 & -\frac{\alpha_3}{s_1} & -k_4 & -\frac{k_3}{s_3} \end{pmatrix}.$$

Then Eq. (7) yields the matrix A shown above. The values of $\alpha_1, \ldots, \alpha_4, k_1, \ldots, k_4$ are given in the figure legends of Figs. 1 and 2. They show bifurcation diagrams for system (5). In each bifurcation diagram, system (5) does not have a stable positive equilibrium bifurcating from the origin and is settled in a 2-cycle. In Fig. 1, $\alpha_1 > \alpha_2$ and $\alpha_3 > \alpha_4$ are satisfied. This condition implies that two species compete severely between the same level of age-classes. As shown in Fig. 3, this case leads to coexistence of two species with temporal segregation between the same level of age-classes. In Fig. 2, $\alpha_1 < \alpha_2$ and $\alpha_3 < \alpha_4$ are satisfied. This condition implies that two species compete severely between the different level of age-classes. As shown in Fig. 4, this case leads to coexistence of two species with temporal segregation between the different level of age-classes. Figure 5 shows the dynamics of species 1 and 2, respectively, when they are isolated from each other. All parameters are the same as in Figs. 3 and 4 except $\alpha_1, \ldots, \alpha_4$. Thus this numerical simulation shows that age-specific species competition is an essential factor causing the population cycles observed in Figs. 3 and 4.

9 Concluding Remarks

This paper studied the dynamics of a multi-species semelparous population model, which is described by coupling multiple nonlinear semelparous Leslie matrix models. We focused on bifurcations of the extinction equilibrium and proposed a simple method of evaluating the stability of a branch of positive equilibria bifurcating from the extinction equilibrium. The method reduces the bifurcation problem into a stability problem of Lotka-Volterra equations. Using this reduction method, we found a population cycle in a competitive system composed of two biennial species. The mechanism of producing this population cycle is new in the sense that it is produced without either severe between-age-class competition or predator-prey like species interaction. It is a future problem to classify all possible dynamics of such a competitive system.

Our study provides a mathematical basis to some preceding studies. In [13, 14], the Lotka-Volterra equation with A and $\mathbf{r}$ given by (7) is derived from system (1). Our study was motivated by the study by Diekmann and van Gils [7], who derived a Lotka-Volterra equation with cyclic symmetry from a nonlinear semelparous Leslie matrix model. The three preceding studies do not show how the derived Lotka-Volterra equation reflects the dynamical behavior of the original single- or multi-species semelparous population model. However our study revealed that the derived Lotka-Volterra equation can be used to examine the stability of a branch of positive

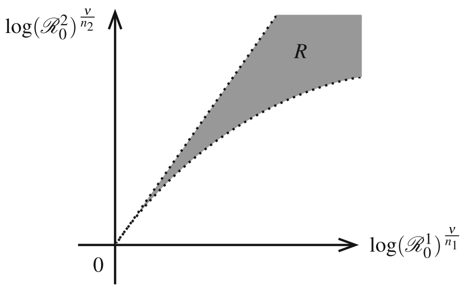


Fig. 6 The $(\log(\mathscr{R}_0^1)^{\frac{\nu}{n_1}}, \log(\mathscr{R}_0^2)^{\frac{\nu}{n_2}})$-parameter plane with an open parameter region R with a cusp at the origin

equilibria of the original model bifurcating the extinction equilibrium. Furthermore, our result rediscovered the result by Cushing and Henson [5], who obtained a condition for stable bifurcation of positive equilibria in nonlinear semelparous Leslie matrix models (see Sect. 7.2).

In our bifurcation study, we focused on a bifurcation that occurs at the critical point $\mathscr{R}_0^1 = \mathscr{R}_0^2 = \cdots = \mathscr{R}_0^N = 1$. In order to avoid treating a multi-parameter bifurcation problem, we perturb the parameters $\mathscr{R}_0^1, \mathscr{R}_0^2, \ldots, \mathscr{R}_0^N$ with maintaining the relation (4). This approach is practically sufficient to examine the dynamics of system (1) with the parameter around the critical point. However there could exist exceptional cases that our approach is unable to treat. Figure 6 shows the $(\log(\mathscr{R}_0^1)^{\frac{\nu}{n_1}}, \log(\mathscr{R}_0^2)^{\frac{\nu}{n_2}})$-parameter plane with an open parameter region R with a cusp at the origin. It is clear that any neighborhood of the origin intersects with R. However, for any vector $\mathbf{c}$, there exists a constant $\varepsilon_0 > 0$ such that $\varepsilon\mathbf{c} \notin R$ for all $\varepsilon \in (0, \varepsilon_0)$. This implies that our approach cannot detect the dynamics in such a region. Therefore, in order to reveal the dynamics of system (1) in a neighborhood of the origin of the parameter plane, we need to consider a multi-parameter bifurcation problem. Whether or not the region that our approach cannot detect exists remains an open question.

Acknowledgments This work was supported by KAKENHI Grant Number JP16K05279.

References

1. Bulmer, M.G.: Periodical insects. Am. Nat. **111**, 1099–1117 (1977)
2. Cushing, J.M., Li, J.: On Ebenman's model for the dynamics of a population with competing juveniles and adults. Bull. Math. Biol. **51**(6), 687–713 (1989)
3. Cushing, J.M.: Nonlinear semelparous Leslie models. Math. Biosci. Eng. **3**, 17–36 (2006)
4. Cushing, J.M.: Three stage semelparous Leslie models. J. Math. Biol. **59**(1), 75–104 (2009)
5. Cushing, J.M., Henson, S.M.: Stable bifurcations in semelparous Leslie models. J. Biol. Dyn. **6**, 80–102 (2012)
6. Davydova, N.V., Diekmann, O., van Gils, S.A.: Year class coexistence or competitive exclusion for strict biennials? J. Math. Biol. **46**(2), 95–131 (2003)
7. Diekmann, O., van Gils, S.A.: On the cyclic replicator equation and the dynamics of semelparous populations. SIAM J. Appl. Dyn. Syst. **8**, 1160–1189 (2009)

8. Ebenman, B.: Competition between age classes and population dynamics. J. Theor. Biol. **131**(4), 389–400 (1988)
9. Hardy, G.H., Wright, E.M.: An Introduction to the Theory of Numbers. Oxford University Press, Oxford, sixth edition (2008)
10. Hofbauer, J.: On the occurrence of limit cycles in the Volterra-Lotka equation. Nonlinear Anal. Theory Methods Appl. **5**(9), 1003–1007 (1981)
11. Kon, R.: Invasibility of missing year-classes in leslie matrix models for a semelparous biennial population. In: Proceedings of Czech-Japanese Seminar in Applied Mathematics 2005 of COE Lecturer Note, vol. 3, pp. 77–87. Kyushu Univ. The 21 Century COE Program, Fukuoka (2006)
12. Kon, R.: Competitive exclusion between year-classes in a semelparous biennial population. In: Deutsch, A., de la Parra, R.B., de Boer, R.J., Diekmann, O., Jagers, P., Kisdi, E., Kretzschmar, M., Lansky, P., Metz, H. (eds.) Mathematical Modeling of Biological Systems. Volume II, pp. 79–90. Birkhäuser, Boston (2007)
13. Kon, R.: Age-structured Lotka-Volterra equations for multiple semelparous populations. SIAM J. Appl. Math. **71**(3), 694–713 (2011)
14. Kon, R.: Permanence induced by life-cycle resonances: the periodical cicada problem. J. Biol. Dyn. **6**(2), 855–890 (2012)
15. Kon, R., Iwasa, Y.: Single-class orbits in nonlinear Leslie matrix models for semelparous populations. J. Math. Biol. **5**(5–6), 781–802 (2007)
16. May, R.M., Leonard, W.J.: Nonlinear aspects of competition between three species. SIAM J. Appl. Math. **29**(2), 243–253 (1975)
17. Mjølhus, E., Wikan, A., Solberg, T.: On synchronization in semelparous populations. J. Math. Biol. **50**(1), 1–21 (2005)

Dichotomy Spectra of Nonautonomous Linear Integrodifference Equations

Christian Pötzsche

Abstract We give examples of dichotomy spectra for nonautonomous linear difference equations in infinite-dimensional spaces. Particular focus is on the spectrum of integrodifference equations having compact coefficients. Concrete systems with explicitly known spectra are discussed for several purposes: (1) They yield reference examples for numerical approximation schemes. (2) The asymptotic behavior of spectral intervals is tackled illustrating their merging.

Keywords Integrodifference equations · Dichotomy spectrum · Sacker–Sell spectrum

1 Motivation and Introduction

Over the last decades, integrodifference equations (IDEs, for short) became popular models in theoretical ecology, since they provide a flexible tool to describe the growth and dispersal of populations with discrete nonoverlapping generations. In the simplest case, where growth precedes dispersal, they are of Hammerstein type

$$u_{t+1}(x) = \int_{\Omega} k_t(x, y) f_t(y, u_t(y))\, \mathrm{d}y \quad \text{for all } t \in \mathbb{Z},\ x \in \Omega \tag{1}$$

(see [17]). Here, the real-valued function u_t represents the density of a population at discrete time t over some spatial habitat $\Omega \subseteq \mathbb{R}^{\kappa}$, the kernels k_t are probability density functions describing the dispersal and f_t is a growth function of e.g. Beverton–Holt or Ricker type. Both functions k_t and f_t are allowed to depend on time in order to include temporally changing environments into our analysis; we refer to [16] for a concrete application. Typical state spaces for (1) are the continuous or the p-integrable functions over Ω.

C. Pötzsche (✉)
Institut für Mathematik, Universität Klagenfurt, 9020 Klagenfurt, Austria
e-mail: christian.poetzsche@aau.at

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_2

Apparently, linear IDEs are of fundamental nature. First, they describe Malthusian growth $f_t(y, u) = c_t(y)u$ with ambient growth functions c_t. Second, and more importantly, when linearizing (1) along a reference solution $(\phi_t^*)_{t\in\mathbb{Z}}$, one arrives at a linear variational equation

$$v_{t+1}(x) = \int_\Omega k_t(x, y) D_2 f_t(y, \phi_t^*(y)) v_t(y)\,\mathrm{d}y \quad \text{for all } t \in \mathbb{Z},\ x \in \Omega. \tag{2}$$

This is a nonautonomous linear difference equation in the infinite-dimensional state space of (1) and alone a local analysis near ϕ^* requires a thorough insight into the dynamical behavior of (2). Theoretically the *dichotomy spectrum* $\Sigma \subseteq (0, \infty)$ (also denoted as *dynamical* or *Sacker–Sell spectrum*) of (2) provides such an insight and hence an adequate "linear algebra" well-suited to establish a geometric theory of nonautonomous difference equation (cf. [20]) and particularly (1). In terms of *spectral intervals* it indeed gives nonautonomous counterparts to eigenvalue moduli, while the *spectral bundles* extend (generalized) eigenspaces to a time-variant setting. Specific applications of the dichotomy spectrum are as follows:

- The solution ϕ^* is uniformly asymptotically stable, if and only if $\Sigma \subseteq (0, 1)$ holds, while a spectral interval in $(1, \infty)$ implies instability.
- If $1 \notin \Sigma$, then the solution ϕ^* is robust and persists locally as unique bounded entire solution to (1) under variation of the system.
- For each gap in Σ one can construct a pair of invariant fiber bundles, which generalize the classical hierarchy of invariant manifolds to a nonautonomous setting. In case $1 \in \Sigma$ stability is determined by the behavior on such a center fiber bundle. Hence, the gaps determine the number of invariant fiber bundles corresponding to an entire solution ϕ^* to (1).

While the dichotomy spectrum dates back to [4, 25], a detailed analysis of its structure for difference equations in infinite-dimensional spaces is of more recent origin [24]. Nevertheless the motivation for this text is two-fold: First, already in finite dimensions only numerical methods allow an approximation of the spectrum (see [15]). It is thus handy to have a class of reference examples with explicitly known spectra available in order to verify computational methods. Second, we illustrate the structure of several spectra arising for nonautonomous IDEs and investigate the asymptotics of their spectral intervals.

The organization of this paper is as follows: We begin reviewing the dichotomy spectrum and some of its central properties for difference equations in infinite-dimensional state spaces. Particular focus is on the situation of compact operators, which was established in [24]. We then concentrate on operators having a discrete spectrum and provide the spectra for associate systems with multiplicative time-varying perturbations. As concrete application we consider IDEs. Sufficient criteria for their well-definedness in L^p- and C-spaces are quoted, we address the asymptotic behavior of the spectral intervals accumulating at 0, and finally present operators with explicitly known spectra or at least explicitly known asymptotics. The latter case applies to various equations relevant in applications.

As reference for difference equations in Banach spaces we mention [11, 20]. Corresponding results for nonautonomous parabolic evolutionary equations were obtained in [23].

Notation

Let $\mathbb{K}$ be one of the fields $\mathbb{R}$ or $\mathbb{C}$. The Kronecker symbol is denoted by δ_{kl}. A *discrete interval* $\mathbb{I}$ is the intersection of a real interval with $\mathbb{Z}$, i.e. a set of consecutive integers. We write $\mathbb{I}' := \{t \in \mathbb{I} : t+1 \in \mathbb{I}\}$ and suppose throughout that $\mathbb{I}$ is unbounded. For nonempty subsets $A, B \subseteq \mathbb{R}$ and $\lambda \in \mathbb{R}$ let us abbreviate

$$AB := \{ab \in \mathbb{R} : a \in A,\ b \in B\}, \qquad \lambda A := \{\lambda a \in \mathbb{R} : a \in A\}.$$

Unless further noted, X, Y are Banach spaces, resp. their complexification, if spectral theoretical matters are addressed. Let X' be the dual space of X with duality pairing $\langle \cdot, \cdot \rangle$. The bounded linear maps from X to Y are denoted by $L(X, Y)$, $L(X) := L(X, X)$ and I_X is the identity mapping on X. We write $N(T) := T^{-1}(\{0\})$ for the *kernel* and $R(T) := TX$ for the *range* of $T \in L(X, Y)$. The *spectrum* of $S \in L(X)$ is $\sigma(S) \subset \mathbb{C}$.

A subset $\mathscr{A} \subseteq \mathbb{I} \times X$ is called a *nonautonomous set*, if all *t-fibers*

$$\mathscr{A}(t) := \{x \in X : (t, x) \in \mathscr{A}\}, \ t \in \mathbb{I}$$

are nonempty. One speaks of a *vector bundle* $\mathscr{V} \subseteq \mathbb{I} \times X$, if every fiber $\mathscr{V}(t) \subseteq X$ is a linear subspace and in case all $\mathscr{V}(t)$ have the same dimension, it determines the *dimension* $\dim \mathscr{V}$ of $\mathscr{V}$. *Constant vector bundles* are of the form $\mathscr{V} = \mathbb{I} \times X_0$ with a subspace $X_0 \subseteq X$ and particular examples are

$$\mathscr{O} := \mathbb{I} \times \{0\}, \qquad \mathscr{X} := \mathbb{I} \times X.$$

2 Dichotomy Spectrum

Given a sequence $(\mathcal{K}_t)_{t \in \mathbb{I}'}$ of bounded linear operators in $L(X)$ as coefficients, we consider linear nonautonomous equations

$$\boxed{u_{t+1} = \mathcal{K}_t u_t} \qquad (L)$$

in an infinite-dimensional Banach space X. A vector bundle $\mathscr{V}$ is called *forward invariant* resp. *invariant*, provided $\mathcal{K}_t \mathscr{V}(t) \subseteq \mathscr{V}(t+1)$ or $\mathcal{K}_t \mathscr{V}(t) = \mathscr{V}(t+1)$ hold for all $t \in \mathbb{I}'$. Their *evolution operator* is the mapping

$$\Phi_{\mathcal{K}} : \{(t,s) \in \mathbb{I} \times \mathbb{I} : s \leq t\} \to L(X), \quad \Phi_{\mathcal{K}}(t,s) := \begin{cases} \mathcal{K}_{t-1} \cdots \mathcal{K}_s, & s < t, \\ I_X, & s = t. \end{cases}$$

For simplicity we suppose from now on that (L) has *bounded (forward) growth*, i.e.

$$\alpha_0 := \sup_{t \in \mathbb{I}'} \|\mathcal{K}_t\| < \infty. \tag{3}$$

One says a linear difference equation (L) has an *exponential dichotomy* (ED for short, cf. [14, p. 229, Definition 7.6.4]) on $\mathbb{I}$, if there exists a projector $P : \mathbb{I} \to L(X)$ and reals $K \geq 1$, $\alpha \in (0,1)$ such that

- $\mathcal{K}_t P(t) = P(t+1)\mathcal{K}_t$ for all $t \in \mathbb{I}'$ (P is an *invariant projector*)
- $\bar{\Phi}_{\mathcal{K}}(t,s) := \Phi_{\mathcal{K}}(t,s)|_{N(P(s))} : N(P(s)) \to N(P(t))$ is a topological isomorphism for $s < t$[1]
- $\|\Phi_{\mathcal{K}}(t,s)P(s)\| \leq K\alpha^{t-s}$ and $\left\|\bar{\Phi}_{\mathcal{K}}(s,t)\,[I_X - P(t)]\right\| \leq K\alpha^{t-s}$ for $s \leq t$.

The *dichotomy spectrum* of (L) is defined as

$$\Sigma_{\mathbb{I}}(\mathcal{K}) := \left\{\gamma > 0 : \ u_{t+1} = \gamma^{-1}\mathcal{K}_t u_t \text{ admits no ED on } \mathbb{I}\right\}$$

and $\rho_{\mathbb{I}}(\mathcal{K}) := (0,\infty) \setminus \Sigma_{\mathbb{I}}(\mathcal{K})$ denotes the *dichotomy resolvent*. If the discrete interval $\mathbb{I}$ is fixed, then we simply write $\Sigma(\mathcal{K})$ resp. $\rho(\mathcal{K})$.

Due to the bounded growth (3) one has $\Sigma(\mathcal{K}) \subseteq (0, \alpha_0]$. The components of $\Sigma(\mathcal{K})$ are called *spectral intervals* and the *dominant spectral interval* contains the largest elements. If $\Sigma(\mathcal{K})$ consists of isolated points, one speaks of a *discrete spectrum*.

Essential properties of the dichotomy spectrum can be summarized as follows:

- $\Sigma(\mathcal{K}) \cup \{0\}$ is compact, $\Sigma_{\mathbb{I}}(\mathcal{K}) \subseteq \Sigma_{\mathbb{Z}}(\mathcal{K})$ for unbounded subintervals $\mathbb{I} \subseteq \mathbb{Z}$ and

$$\Sigma(\lambda\mathcal{K}) = |\lambda|\,\Sigma(\mathcal{K}) \quad \text{for all } \lambda \in \mathbb{C} \setminus \{0\}$$

- It is upper-semicontinuous, i.e. for every $\varepsilon > 0$ there exists a $\delta > 0$ such that every sequence $(\bar{\mathcal{K}}_t)_{t \in \mathbb{I}'}$ in $L(X)$ fulfills

$$\sup_{t \in \mathbb{I}'} \left\|\bar{\mathcal{K}}_t - \mathcal{K}_t\right\| < \delta \quad \Rightarrow \quad \Sigma(\bar{\mathcal{K}}) \subseteq B_{\varepsilon}(\Sigma(\mathcal{K}))$$

- $\Sigma(\mathcal{K})$ is invariant under *kinematic similarity*, i.e. if there exists a sequence $(\mathcal{S}_t)_{t \in \mathbb{I}}$ of invertible operators $\mathcal{S}_t \in L(X,Y)$ with $\sup_{t \in \mathbb{I}} \max\left\{\|\mathcal{S}_t\|, \left\|\mathcal{S}_t^{-1}\right\|\right\} < \infty$, then (L) and $v_{t+1} = \mathcal{S}_{t+1}^{-1}\mathcal{K}_t\mathcal{S}_t v_t$ have the same dichotomy spectrum. The sequence $(\mathcal{S}_t)_{t \in \mathbb{I}}$ is called *Lyapunov transformation*.

[1] For this it suffices to assume that $\mathcal{K}_t|_{N(P(t))} : N(P(t)) \to N(P(t+1))$, $t \in \mathbb{I}'$, are isomorphisms.

Finally, for every $\gamma > 0$ we define the vector bundles

$$\mathcal{V}_\gamma^+ := \left\{(\tau, \xi) \in \mathcal{X} : \sup_{\tau \le t} \|\Phi_{\mathcal{K}}(t, \tau)\xi\| \gamma^{\tau - t} < \infty\right\},$$

$$\mathcal{V}_\gamma^- := \left\{(\tau, \xi) \in \mathcal{X} : \begin{array}{c}\text{there exists a solution } (\phi_t)_{t \in \mathbb{I}} \text{ of } (L) \\ \text{with } \phi_\tau = \xi \text{ and } \sup_{\tau \le t} \|\phi_t\| \gamma^{\tau - t} < \infty\end{array}\right\};$$

in case γ is chosen from the dichotomy resolvent $\rho(\mathcal{K})$, one denotes $\mathcal{V}_\gamma^+$ as a *pseudo-stable* and $\mathcal{V}_\gamma^-$ as a *pseudo-unstable* bundle of (L).

The subsequent classes of linear difference equations allow more detailed statements and insights into the structure of their dichotomy spectrum:

2.1 Periodic Difference Equations

Let (L) be p-periodic, i.e. there exists a $p \in \mathbb{N}$ such that $\mathcal{K}_t = \mathcal{K}_{t+p}$ for all $t \in \mathbb{Z}$.

Then the dichotomy spectrum reads as

$$\Sigma_{\mathbb{Z}}(\mathcal{K}) = |\{\lambda \in \mathbb{C} : \lambda \in \sigma(\Phi_{\mathcal{K}}(p, 0))\} \setminus \{0\}|^{1/p} \tag{4}$$

and in particular for autonomous equations ($p = 1$) it consists of the positive moduli of the spectral points for $\mathcal{K}$. The pseudo-stable and -unstable bundles of (L) can be characterized in terms of Riesz projections (see [8, p. 30, Theorem 1.5.4]) associated to the components of $\sigma(\Phi_{\mathcal{K}}(p, 0))$, but need not to be finite-dimensional.

Rather explicit information can be obtained in

Example 1 (*multiplication operator*) Suppose (Ω, Σ, μ) is a σ-finite measure space and $1 \le p < \infty$. For $\mathbb{K}$-valued functions $a_t \in L^\infty(\Omega, \mu)$ we define the *essential range*

$$\rho_{\text{ess}}(a_t) := \left\{\lambda \in \mathbb{C} : \mu\left(\left\{x \in \Omega : \left|a_t(x) - \lambda\right| < \varepsilon\right\}\right) \ne 0 \text{ for all } \varepsilon > 0\right\}$$

for all $t \in \mathbb{I}'$. On $X = L^p(\Omega, \mu)$ the multiplication operators

$$\mathcal{K}_t \in L(L^p(\Omega, \mu)), \qquad [\mathcal{K}_t v](x) := a_t(x)v(x) \quad \text{for all } t \in \mathbb{I}',\ x \in \Omega$$

are well-defined and yield an evolution operator of (L) given by

$$[\Phi_{\mathcal{K}}(t, \tau)v](x) = \left(\prod_{s=\tau}^{t-1} a_s(x)\right)v(x) \quad \text{for all } \tau \le t,\ v \in L^p(\Omega, \mu),$$

which is a multiplication operator again. In the periodic situation $a_t = a_{t+p}$, $t \in \mathbb{Z}$, the spectrum of $\Phi_{\mathcal{K}}(p, 0)$ is the essential range of the product $\prod_{s=0}^{p-1} a_s : \Omega \to \mathbb{K}$ (see [10, pp. 30ff]) and due to (4) we arrive at

$$\Sigma(\mathcal{K}) = \left|\rho_{\text{ess}}\left(\prod_{s=0}^{p-1} a_s\right) \setminus \{0\}\right|^{1/p}.$$

Example 2 (*shift operator*) Suppose that $(B_t)_{t\in\mathbb{Z}}$ is a bounded sequence in $L(Y)$ such that the difference equation $y_{t+1} = B_t y_t$ in Y has a nonempty dichotomy spectrum $\Sigma_{\mathbb{Z}}(B)$. Furthermore, let $X :=\ell^p(Y)$ be the space of p-summable sequences $(y_t)_{t\in\mathbb{Z}}$ in Y for $p \in [1, \infty]$ and define the shift

$$\mathcal{K} \in L(\ell^p(Y)), \qquad [\mathcal{K}v]_s :=B_{s-1}v_{s-1} \quad \text{for all } s \in \mathbb{Z},\ v \in \ell^p(Y).$$

In [21, Theorem 1] it is shown that $\sigma(\mathcal{K}) = \overline{\{\lambda \in \mathbb{C} : |\lambda| \in \Sigma_{\mathbb{Z}}(B)\}}$ and we hence obtain from (4) for $p = 1$ that $\Sigma_{\mathbb{I}}(\mathcal{K}) = \Sigma_{\mathbb{Z}}(B)$.

2.2 Compact Difference Equations

Let (L) be compact, i.e. the coefficients $\mathcal{K}_t \in L(X)$, $t \in \mathbb{I}'$, are compact operators.

Due to our global bounded growth assumption (3) the spectrum $\Sigma(\mathcal{K})$ is bounded above by α_0 and there exists a $\gamma_0 > 0$ such that $(\gamma_0, \infty) \subseteq \rho(\mathcal{K})$; we set

$$\mathcal{V}_{\gamma_0}^+ :=\mathcal{X}, \qquad \mathcal{V}_{\gamma_0}^- :=\mathcal{O}.$$

Furthermore, in [24, Corollary 4.13] it is shown that $\Sigma(\mathcal{K})$ is a union of at most countably many intervals which can only accumulate at a number $\bar{\mu} \geq 0$ and that the pseudo-unstable bundles $\mathcal{V}_\gamma^-$ are finite-dimensional. In detail, one of the cases holds:

($\mathfrak{S}_0$) $\Sigma(\mathcal{K}) = \emptyset$
($\mathfrak{S}_1$) $\Sigma(\mathcal{K})$ consists of finitely many closed spectral intervals:

($\mathfrak{S}_1^1$) There exists a $k \in \mathbb{N}$ and reals $0 < \alpha_k \leq \beta_k < \ldots < \alpha_1 \leq \beta_1 \leq \alpha_0$ with

$$\Sigma(\mathcal{K}) = \bigcup_{j=1}^{k} [\alpha_j, \beta_j]$$

and we choose reals $\gamma_j \in (\beta_{j+1}, \alpha_j)$, $1 \leq j < k$, and $\gamma_k \in (0, \alpha_k)$ (see Fig. 1)
($\mathfrak{S}_1^2$) There exists a $k \in \mathbb{N}_0$ and reals $0 < \beta_{k+1} < \alpha_k \leq \beta_k < \ldots < \alpha_1 \leq \beta_1 \leq \alpha_0$ with

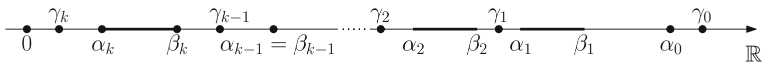


Fig. 1 Case $(\mathfrak{S}_1^1)$ with k compact spectral intervals

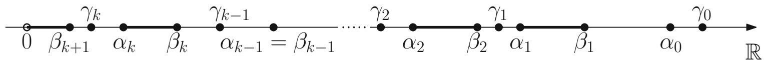


Fig. 2 Case $(\mathfrak{S}_1^2)$ with $k+1$ spectral intervals

$$\Sigma(\mathcal{K}) = (0, \beta_{k+1}] \cup \bigcup_{j=1}^{k} [\alpha_j, \beta_j]$$

and we choose reals $\gamma_j \in (\beta_{j+1}, \alpha_j)$, $1 \leq j \leq k$ (see Fig. 2).
In both cases the *spectral bundles*

$$\mathscr{X}_0 := \mathscr{V}_{\gamma_0}^-, \qquad \mathscr{X}_j := \mathscr{V}_{\gamma_{j-1}}^+ \cap \mathscr{V}_{\gamma_j}^- \neq \mathscr{O} \text{ for all } 1 \leq j \leq k$$

are finite-dimensional invariant vector bundles of (L) with the finite *Whitney sum*

$$\mathscr{X} = \bigoplus_{j=0}^{k} \mathscr{X}_j \oplus \mathscr{V}_{\gamma_k}^+$$

and the bundle $\mathscr{V}_{\gamma_k}^- = \bigoplus_{j=0}^{k} \mathscr{X}_j$ satisfying $k \leq \dim \mathscr{V}_{\gamma_k}^- = \sum_{j=0}^{k} \dim \mathscr{X}_j$

$(\mathfrak{S}_2)$ $\Sigma(\mathcal{K})$ consists of infinitely many spectral intervals: There exist strictly decreasing sequences $(\alpha_j)_{j\in\mathbb{N}}$, $(\beta_j)_{j\in\mathbb{N}}$ such that

$$\Sigma(\mathcal{K}) = \sigma_\infty \cup \bigcup_{j=1}^{\infty} [\alpha_j, \beta_j],$$

where $\bar{\mu} < \alpha_j \leq \beta_j$, $\lim_{j\to\infty} \alpha_j = \bar{\mu}$, $\sigma_\infty = \emptyset$ for $\bar{\mu} = 0$ and $\sigma_\infty = (0, \bar{\mu}]$ otherwise (see Fig. 3). If we choose reals $\gamma_j \in (\beta_{j+1}, \alpha_j)$, $j \in \mathbb{N}$, then the *spectral bundles*

$$\mathscr{X}_0 := \mathscr{V}_{\gamma_0}^-, \qquad \mathscr{X}_j := \mathscr{V}_{\gamma_{j-1}}^+ \cap \mathscr{V}_{\gamma_j}^- \neq \mathscr{O} \quad \text{for all } j \in \mathbb{N}$$

are finite-dimensional invariant vector bundles of (L) and for every $k \in \mathbb{N}$ one has the finite *Whitney sum*

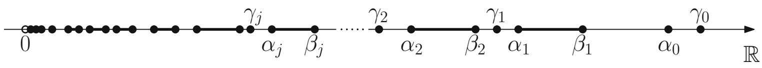


Fig. 3 Case $(\mathfrak{S}_2)$ with infinitely many spectral intervals $[\alpha_j, \beta_j]$ accumulating at $\bar{\mu} = 0$ i.e. $\sigma_\infty = \emptyset$

$$\mathscr{X} = \bigoplus_{j=0}^{k} \mathscr{X}_j \oplus \mathscr{V}_{\gamma_k}^+$$

and the bundle $\mathscr{V}_{\gamma_k}^- = \bigoplus_{j=0}^k \mathscr{X}_j$ satisfying $k \le \dim \mathscr{V}_{\gamma_k}^- = \sum_{j=0}^k \dim \mathscr{X}_j$.

By construction, the dominant interval is $[\alpha_1, \beta_1]$. The *order* of a spectral interval with maximum β_j is the dimension of the associate spectral bundle $\mathscr{X}_j$; a *simple* spectral interval has order 1.

2.3 Finite-Rank Difference Equations

Let (L) be of finite rank, i.e. there exists a finite-dimensional subspace $X_0 \subset X$ such that $R(\mathcal{K}_t) = X_0$ for all $t \in \mathbb{I}'$. In particular, every $\mathcal{K}_t$ is compact and (L) essentially behave like finite-dimensional equations.

If $d := \dim X_0$, then $\Sigma(\mathcal{K})$ is a union of at most d intervals (cf. [24, Theorem 4.14]), i.e. either $(\mathfrak{S}_0)$ holds or $\Sigma(\mathcal{K})$ consists of $k \in \{1, \dots, d\}$ spectral intervals: There exist reals $0 < \alpha_k \le \beta_k < \dots < \alpha_1 \le \beta_1 \le \alpha_0$ with closed spectral intervals:

$$\Sigma(\mathcal{K}) = \begin{cases} [\alpha_k, \beta_k] \\ (0, \beta_k] \end{cases} \cup \bigcup_{j=1}^{k-1} [\alpha_j, \beta_j]. \tag{5}$$

If possible, we choose $\gamma_k \in \rho(\mathcal{K})$ such that $(0, \gamma_k) \subseteq \rho(\mathcal{K})$ and otherwise, we define $\mathscr{V}_{\gamma_k}^+ = \mathscr{O}$ and $\mathscr{V}_{\gamma_k}^- = \mathscr{X}$. Then $\mathscr{X}_{k+1} = \mathscr{V}_{\gamma_k}^+$ and $\mathscr{X}_0 = \mathscr{V}_{\gamma_0}^-$ are invariant vector bundles of (L). For $k > 1$ we choose reals $\gamma_j \in (\beta_{j+1}, a_j)$, $1 \le j < k$. Then the sets

$$\mathscr{X}_j := \mathscr{V}_{\gamma_{j-1}}^+ \cap \mathscr{V}_{\gamma_j}^- \ne \mathscr{O} \quad \text{for all } 1 \le j \le k$$

are finite-dimensional invariant vector bundles of (L) with the *Whitney sum*

$$\mathscr{X} = \bigoplus_{j=0}^{k+1} \mathscr{X}_j.$$

Remark 1 Note that the above situation differs from the dichotomy spectrum introduced in [4] for finite-dimensional equations. Indeed, [4] work with the dichotomy concept from [3], which is not ℓ^∞-robust and yields a finer spectrum than ours.

2.4 Finite-Dimensional and Difference Equations

Suppose that $(B_t)_{t\in\mathbb{I}'}$ is a bounded sequence in $\mathbb{K}^{n\times n}$ and consider a linear equation

$$\boxed{y_{t+1} = B_t y_t} \tag{6}$$

with evolution operator $\Phi_B(t,s) \in \mathbb{K}^{n\times n}$, $s \le t$. Its dichotomy spectrum $\Sigma(B)$ fits in the above framework of Sect. 2.3. Each spectral interval in (5) corresponds to an invariant vector bundle

$$\mathscr{Y}_j := \left\{(t,x) \in \mathbb{I} \times \mathbb{K}^n : x \in R(p_j(t))\right\} \quad \text{for all } 1 \le j \le k,$$

where $p_j : \mathbb{I} \to L(\mathbb{K}^n)$ is an invariant projector for (6), and $\mathbb{I} \times \mathbb{K}^n = \bigoplus_{j=1}^k \mathscr{Y}_j$.

For scalar difference equations the following notion of Bohl exponents is central. Assume $(a_t)_{t\in\mathbb{I}'}$ is a *tempered* sequence in $\mathbb{K}$, i.e. it satisfies $a_t \neq 0$ for all $t \in \mathbb{I}'$ and

$$\sup_{t\in\mathbb{I}'} \max\left\{|a_t|, \left|a_t^{-1}\right|\right\} < \infty.$$

Let $I_T(\mathbb{I}) := \{\mathbb{J} \subseteq \mathbb{I} : \mathbb{J} \text{ is a discrete interval with } \#\mathbb{J} = T\}$ denote the family of all discrete subintervals of $\mathbb{I}$ with $T \in \mathbb{N}$ elements. The *upper* resp. *lower Bohl exponent* of a are given by

$$\overline{\beta}(a) := \lim_{T\to\infty} \sup_{\mathbb{J}\in I_T(\mathbb{I})} \sqrt[T]{\left|\prod_{s\in\mathbb{J}} a_s\right|}, \qquad \underline{\beta}(a) := \lim_{T\to\infty} \inf_{\mathbb{J}\in I_T(\mathbb{I})} \sqrt[T]{\left|\prod_{s\in\mathbb{J}} a_s\right|}$$

and one clearly has the homogeneity relations

$$\underline{\beta}(\lambda a) = |\lambda|\, \underline{\beta}(a), \qquad \overline{\beta}(\lambda a) = |\lambda|\, \overline{\beta}(a) \quad \text{for all } \lambda \in \mathbb{C} \setminus \{0\}.$$

Especially for $\mathcal{K}_t := a_t I_X$, $t \in \mathbb{I}$, one has the spectrum

$$\Sigma(\mathcal{K}) = [\underline{\beta}(a), \overline{\beta}(a)]$$

and we refer to [22] for further properties of Bohl exponents.

3 Operators with Discrete Spectrum

Assume now that $\mathcal{K} \in L(X)$ is a single linear operator. Given an eigenvalue $\lambda \in \mathbb{C}$ of $\mathcal{K}$, we denote its *order* as

$$o_\lambda = \min\left\{o \in \mathbb{N} : N(\mathcal{K} - \lambda I_X)^o = N(\mathcal{K} - \lambda I_X)^{o+1}\right\}$$

and our future analysis is based on the following properties:

- (H_1) There exist nonempty discrete intervals $\mathbb{J}(\mathcal{K}) \subseteq I(\mathcal{K}) \subseteq \mathbb{N}$ such that
  - $\sigma(\mathcal{K}) \setminus \{0\} = \{\lambda_i : i \in I(\mathcal{K})\}$ consists of eigenvalues λ_i such that $(|\lambda_i|)_{i\in I(\mathcal{K})}$ is a decreasing sequence
  - $|\sigma(\mathcal{K}) \setminus \{0\}| = \left\{\rho_j : j \in \mathbb{J}(\mathcal{K})\right\}$ with a strictly deceasing sequence $(\rho_j)_{j\in\mathbb{J}(\mathcal{K})}$ of positive reals and $s_j :=\# \left\{\lambda \in \sigma(\mathcal{K}) : |\lambda| = \rho_j\right\} < \infty$ for $j \in \mathbb{J}(\mathcal{K})$
- (H_2) Given bases of generalized (and norm 1) eigenvectors such that

$$N(\mathcal{K} - \lambda I_X)^{o_\lambda} = \operatorname{span}\left\{e_\lambda^1, \dots, e_\lambda^{o_\lambda}\right\} \quad \text{for all } \lambda \in \sigma(\mathcal{K}) \setminus \{0\},$$

  the sequence $(e_n)_{n\in N} :=(e_{\lambda_1}^1, \dots, e_{\lambda_1}^{o_{\lambda_1}}, e_{\lambda_2}^1, \dots, e_{\lambda_2}^{o_{\lambda_2}}, \dots)$ is a basis of X.

According to [8, p. 80, Lemma 3.3.1] one can complement the basis $(e_n)_{n\in N}$ of X to a biorthonormal system $(e_n, f_n)_{n\in N}$, where $N \subseteq \mathbb{N}$ is a discrete interval. This means there exists a sequence $(f_n)_{n\in N} :=(f_{\lambda_1}^1, \dots, f_{\lambda_1}^{o_{\lambda_1}}, f_{\lambda_2}^1, \dots, f_{\lambda_2}^{o_{\lambda_2}}, \dots)$ of functionals $f_n \in X'$ satisfying $\langle e_n, f_m\rangle = \delta_{nm}$ for all $m, n \in N$. Then

$$\Pi(\lambda) :=\sum_{n=1}^{o_\lambda} \left\langle \cdot, f_\lambda^n\right\rangle e_\lambda^n \quad \text{for all } \lambda \in \sigma(\mathcal{K}) \setminus \{0\}$$

is a bounded projector onto $N(\mathcal{K} - \lambda I_X)^{o_\lambda}$ with

$$\Pi(\lambda_i)\Pi(\lambda_j) = \delta_{ij}\Pi(\lambda_i), \qquad \Pi(\lambda_i)\mathcal{K} = \mathcal{K}\Pi(\lambda_i) \quad \text{for all } i, j \in I(\mathcal{K}), \tag{7}$$

since $(e_n, f_n)_{n\in N}$ is a biorthonormal system. We next define the *spectral spaces*

$$X_j :=\bigoplus_{|\lambda|=\rho_j} N(\mathcal{K} - \lambda I_X)^{o_\lambda} \quad \text{for all } j \in \mathbb{J}(\mathcal{K}),$$

which are invariant and of dimension $\sum_{|\lambda|=\rho_j} o_\lambda$, as well as finite rank mappings

$$\Pi_j : X \to X_j, \qquad \Pi_j :=\sum_{|\lambda|=\rho_j} \Pi(\lambda) \quad \text{for all } j \in \mathbb{J}(\mathcal{K}).$$

From (7) we readily obtain the commutativity relations

$$\Pi_j \Pi_i = \delta_{ij} \Pi_j, \qquad \mathcal{K}\Pi_j = \Pi_j \mathcal{K} \quad \text{for all } i, j \in \mathbb{J}(\mathcal{K}).$$

Thus, Π_j, $j \in \mathbb{J}(\mathcal{K})$, are a family of complementary projections onto the spectral spaces X_j.

Example 3 (*normal compact operators*) If $\mathcal{K} \in L(X)$ is a compact operator with $I(\mathcal{K}) = \mathbb{J}(\mathcal{K}) = \mathbb{N}$, then $\lim_{i\to\infty} \lambda_i = \lim_{j\to\infty} \rho_j = 0$ holds. In case X is an infinite-dimensional Hilbert space and $\mathcal{K}$ is normal, we identify X' with X by means of the Riesz representation theorem. One chooses $f_n :=e_n, n \in N$, and the projections Π_j, as well as the eigenspaces X_j are pairwise orthonormal (see [18, p. 484ff, Sect. 6.7]).

Example 4 (*finite rank operators*) Suppose that $X_0 :=R(\mathcal{K})$ is finite-dimensional with a basis $(x_1, \ldots, x_d)$ and let $S : X_0 \to \mathbb{C}^d$ be an isomorphism. Following [1, p. 274, Theorem 7.4] and using the representation

$$\mathcal{K}v = \sum_{j=1}^{d} \langle v, x'_j \rangle x_j \quad \text{for all } v \in X$$

we define the matrix $K :=(x'_i(x_j))_{i,j=1}^d \in \mathbb{C}^{d\times d}$ and obtain $\sigma(\mathcal{K}) = \sigma(K) \cup \{0\}$. By means of e.g. the Jordan form there exists an invertible matrix $T \in \mathbb{C}^{d\times d}$ such that

$$T^{-1}KT = \begin{pmatrix} S_k & & \\ & \ddots & \\ & & S_1 \end{pmatrix} \quad \text{and} \quad k \le d.$$

The eigenvalues of each block matrix $S_j \in \mathbb{C}^{d_j \times d_j}$ have the same moduli and satisfy $|\sigma(S_{j+1})| < |\sigma(S_j)|$ for $1 \le j < k$. One obtains the spectral spaces

$$X_j :=ST(\{0\} \times \mathbb{C}^{d_j} \times \{0\}) \subset X \quad \text{for all } 1 \le j \le k$$

and $\Pi_j :=ST\,\mathrm{diag}(0, I_{\mathbb{C}^{d_j}}, 0)(ST)^{-1}$ as corresponding projections.

In conclusion, we arrive at a weighted sum

$$\mathcal{K}v = \sum_{j\in\mathbb{J}(\mathcal{K})} \sum_{|\lambda|=\rho_j} \lambda \Pi(\lambda)v \quad \text{for all } v \in X$$

and the discrete semigroup $(\mathcal{K}^t)_{t\ge 0}$ generated by $\mathcal{K}$ has the Fourier representation

$$\mathcal{K}^t v = \sum_{j\in\mathbb{J}(\mathcal{K})} \sum_{|\lambda|=\rho_j} \lambda^t \Pi(\lambda)v \quad \text{for all } t \ge 0,\ v \in X. \tag{8}$$

For autonomous difference equations

$$\boxed{u_{t+1} = \mathcal{K} u_t}$$

in X with coefficients $\mathcal{K} \in L(X)$ satisfying (H_1)–(H_2) the above notions translate into the language of Sect. 2.2 as follows: We obtain a discrete dichotomy spectrum

$$\Sigma(\mathcal{K}) = \bigcup_{j \in \mathbb{J}(\mathcal{K})} \{\rho_j\}$$

and constant spectral bundles $\mathscr{X}_j = \mathbb{I} \times X_j$, $j \in \mathbb{J}(\mathcal{K})$, from (4). An immediate nonautonomous generalization is treated in

Theorem 1 (multiplicative perturbation 1) *If a sequence $(a_t)_{t \in \mathbb{I}}$ is tempered, then the difference equation*

$$\boxed{u_{t+1} = a_t \mathcal{K} u_t} \tag{9}$$

has the dichotomy spectrum $\Sigma(a\mathcal{K}) = \left[\underline{\beta}(a), \overline{\beta}(a)\right] \bigcup_{j \in \mathbb{J}(\mathcal{K})} \{\rho_j\}$ and constant spectral bundles.

Proof Using the Fourier representation (8) we obtain that the evolution operator of (9) reads as

$$\Phi_{a\mathcal{K}}(t, s) = \sum_{j \in \mathbb{J}(\mathcal{K})} \left(\prod_{r=s}^{t-1} a_r\right) \sum_{|\lambda| = \rho_j} \lambda^{t-s} \Pi(\lambda) \quad \text{for all } s \le t.$$

If $\left\{\lambda_j^1, \dots, \lambda_j^{s_j}\right\} \subseteq \sigma(\mathcal{K})$ is the set of eigenvalues with absolute value ρ_j, we obtain

$$\begin{aligned} \Pi_j \Phi_{a\mathcal{K}}(t, s) &= \sum_{j \in \mathbb{J}(\mathcal{K})} \left(\prod_{r=s}^{t-1} a_r\right) \sum_{i=1}^{s_j} (\lambda_j^i)^{t-s} P_j \Pi(\lambda_j^i) \\ &= \Phi_{a\mathcal{K}}(t, s) \Pi_j \quad \text{for all } s \le t. \end{aligned}$$

Hence, the finite-dimensional vector bundles $\mathscr{P}_j := \left\{(t, v) \in \mathscr{X} : v \in R(\Pi_j)\right\}$ are invariant w.r.t. (9) for all $j \in \mathbb{J}(\mathcal{K})$. Inside of each $\mathscr{P}_j$ the dynamics is given by

$$u_{t+1} = a_t \sum_{i=1}^{s_j} \lambda_j^i \Pi(\lambda_j^i) u_t,$$

having an evolution operator $\Phi^j(t, s) := \Phi_{a\mathcal{K}}(t, s) \Pi_j$ and the spectrum $\rho_j \left[\underline{\beta}(a), \overline{\beta}(a)\right]$. Thanks to $\Phi_{a\mathcal{K}}(t, s) = \sum_{j \in \mathbb{J}(\mathcal{K})} \Phi^j(t, s)$ for all $s \le t$ we thus obtain the assertion. □

Corollary 1 *If a sequence $(a_t)_{t\in\mathbb{Z}}$ in $\mathbb{K}$ is p-periodic with nonzero values, then*

$$\Sigma(a\mathcal{K}) = \sqrt[p]{\prod_{s=0}^{p-1} |a_s|} \bigcup_{j\in\mathbb{J}(\mathcal{K})} \{\rho_j\}.$$

Proof The upper and lower Bohl exponents of a are given by $\sqrt[p]{\prod_{s=0}^{p-1}|a_s|}$.

In the following, we are interested in systems of difference equation

$$\boxed{U_{t+1} = \tilde{\mathcal{K}}_t U_t} \tag{10}$$

on the state space X^n for coefficient sequences $(\tilde{\mathcal{K}}_t)_{t\in\mathbb{I}'}$ in $L(X^n)$. We conveniently abbreviate $U = (u_1, \dots, u_n) \in X^n$ throughout. Suppose that $(B_t)_{t\in\mathbb{I}'}$ is a sequence of invertible matrices in $\mathbb{K}^{n\times n}$ satisfying

$$\sup_{t\in\mathbb{I}'} \|B_t\| < \infty, \qquad \sup_{t\in\mathbb{I}'} \left\|B_t^{-1}\right\| < \infty \tag{11}$$

and having the entries $b_{ij}(t)$, $1 \le i, j \le n$. In [4, Theorem 2.1] and Sect. 2.4 it is shown that $\Sigma(B)$ consists of compact intervals in $(0, \infty)$.

Theorem 2 (multiplicative perturbation 2) *Suppose that* (11) *holds. If* (6) *possesses* full spectrum*, i.e.*

$$\Sigma(B) = \bigcup_{i=1}^{n} \sigma_i \tag{12}$$

with compact, decreasing and disjoint spectral intervals $\sigma_i \subset (0, \infty)$, then the difference Eq. (10) *with*

$$\tilde{\mathcal{K}}_t U := \begin{pmatrix} b_{11}(t)\mathcal{K}u_1 + \dots + b_{1n}(t)\mathcal{K}u_n \\ \vdots \\ b_{n1}(t)\mathcal{K}u_1 + \dots + b_{nn}(t)\mathcal{K}u_n \end{pmatrix} \quad \text{for all } t \in \mathbb{I}',\ U \in X^n$$

has the dichotomy spectrum $\Sigma(\tilde{\mathcal{K}}) = \bigcup_{j\in\mathbb{J}(\mathcal{K})} \rho_j \bigcup_{i=1}^{n} \sigma_i = \bigcup_{j\in\mathbb{J}(\mathcal{K})} \rho_j \Sigma(B)$.

Remark 2 (*computation of* (12)) For general coefficient sequences in (6) the computation of the dichotomy spectrum $\Sigma(B)$ is only possible using numerical schemes, as developed in [9, 15].

For the remaining section it is convenient to define the operator

$$\hat{\mathcal{K}} := \begin{pmatrix} \mathcal{K} & & \\ & \ddots & \\ & & \mathcal{K} \end{pmatrix} \in L(X^n).$$

Proof First of all, we obtain from [26, Reduction Theorem] that (6) is kinematically similar to a diagonal system in $\mathbb{K}^n$. More precisely, there exists a Lyapunov transformation $(S_t)_{t\in\mathbb{I}}$ in $\mathbb{K}^{n\times n}$ such that $S_{t+1}^{-1}B_tS_t = \operatorname{diag}(b_t^1,\dots,b_t^n)$ with tempered sequences $(b_t^i)_{t\in\mathbb{I}'}$ such that $\sigma_i = [\underline{\beta}(b^i), \overline{\beta}(b^i)]$, $1 \le i \le n$. One has

$$\hat{\mathcal{K}}S_tU = S_t\hat{\mathcal{K}}U \quad \text{for all } t\in\mathbb{I},\ U\in X^n$$

and consequently we arrive at

$$S_{t+1}^{-1}\tilde{\mathcal{K}}_tS_t = S_{t+1}^{-1}B_t\hat{\mathcal{K}}S_t = S_{t+1}^{-1}B_tS_t\hat{\mathcal{K}} = \begin{pmatrix} b_t^1\mathcal{K} & & \\ & \ddots & \\ & & b_t^n\mathcal{K}\end{pmatrix} \quad \text{for all } t\in\mathbb{I}'.$$

Hence, (10) is kinematically similar to a diagonal difference system in X^n and therefore $\Sigma(\tilde{\mathcal{K}}) = \bigcup_{i=1}^n \Sigma(b^i\mathcal{K})$. Then the assertion follows from Theorem 1 yielding the spectra $\Sigma(b^i\mathcal{K})$. □

We next investigate scalar multiplicative and time-dependent perturbations. The situation is related to Theorem 2, but allows a different proof.

Theorem 3 (multiplicative perturbation 3) *Suppose $D \in \mathbb{K}^{n\times n}$ is diagonalizable and $\sigma(D) = \{d_1,\dots,d_n\}$. If $(a_t)_{t\in\mathbb{I}'}$ is tempered, then the difference Eq. (10) with*

$$\tilde{\mathcal{K}}_tU := a_t\begin{pmatrix} d_{11}\mathcal{K}u^1 + \ldots + d_{1n}\mathcal{K}u^n \\ \vdots \\ d_{n1}\mathcal{K}u^1 + \ldots + d_{nn}\mathcal{K}u^n\end{pmatrix} \quad \textit{for all } t\in\mathbb{I}',\ U\in X^n$$

has the dichotomy spectrum $\Sigma(a\tilde{\mathcal{K}}) = \left[\underline{\beta}(a), \overline{\beta}(a)\right]\bigcup_{j\in\mathbb{J}(\mathcal{K})}\rho_j\bigcup_{i=1}^n|d_i|$ and constant spectral bundles.

Proof First of all, one has the representation $\tilde{\mathcal{K}}_t = a_tD\hat{\mathcal{K}}$ and therefore

$$\Phi_{\tilde{\mathcal{K}}}(t,s) = \left(\prod_{r=s}^{t-1}a_r\right)(D\hat{\mathcal{K}})^{t-s} \quad \text{for all } s\le t.$$

Since D and $\hat{\mathcal{K}}$ commute, we arrive at

$$\Phi_{\tilde{\mathcal{K}}}(t,s) = \left(\prod_{r=s}^{t-1}a_r\right)D^{t-s}\hat{\mathcal{K}}^{t-s} \quad \text{for all } s\le t.$$

By assumption D is diagonalizable and hence there is an invertible $T\in\mathbb{K}^{n\times n}$ with $D = T\operatorname{diag}(d_1,\dots,d_n)T^{-1}$. From $\hat{\mathcal{K}}T^{-1} = T^{-1}\hat{\mathcal{K}}$ we get

$$\begin{aligned} T\Phi_{\tilde{\mathcal{K}}}(t,s)T^{-1} &= \left(\prod_{r=s}^{t-1} a_r\right) TD^{t-s}\hat{\mathcal{K}}^{t-s}T^{-1} = \left(\prod_{r=s}^{t-1} a_r\right) TD^{t-s}T^{-1}\hat{\mathcal{K}}^{t-s} \\ &= \left(\prod_{r=s}^{t-1} a_r\right) (TDT^{-1})^{t-s}\hat{\mathcal{K}}^{t-s} \\ &= \left(\prod_{r=s}^{t-1} a_r\right) \operatorname{diag}((d_1\mathcal{K})^{t-s}, \dots, (d_n\mathcal{K})^{t-s}) \quad \text{for all } s \le t. \end{aligned}$$

Thus, (10) is kinematically similar to the n systems $u_{t+1} = d_i a_t \mathcal{K} u_t$ for all $1 \le i \le n$ and therefore has the dichotomy spectrum $\Sigma(\tilde{\mathcal{K}}) = \bigcup_{i=1}^n \Sigma(d_i a \mathcal{K})$. Using Theorem 1 again, this implies the assertion. □

On the basis of Corollary 1 it is easy to conclude the special case of a periodic Eq. (10) in Theorem 3.

4 Linear Integrodifference Equations

Throughout this section, we suppose that (Ω, Σ, μ) is a measure space. From now on the coefficients in our difference Eq. (L) are assumed to be integral operators

$$\mathcal{K}_t v := \int_\Omega k_t(\cdot, y)v(y)\,\mathrm{d}\mu(y) : \Omega \to \mathbb{K} \quad \text{for all } t \in \mathbb{I}'$$

of Fredholm type with appropriate *kernels* $k_t : \Omega^2 \to \mathbb{K}$. Such equations for instance occur as right-hand sides of variational Eq. (2). Consequently, (L) is an IDE and well-definedness of the coefficients $\mathcal{K}_t$ on various function spaces will be tacked in Sect. 4.1. On a purely formal level, the evolution operator of (L) is again an integral operator

$$\Phi_{\mathcal{K}}(t,\tau) = \int_\Omega k_\tau^{t-1}(\cdot, y)v(y)\,\mathrm{d}\mu(y) : \Omega \to \mathbb{K} \quad \text{for all } \tau < t$$

with the *iterated kernels* for all $x, y \in \Omega$ and $\tau, \tau + n \in \mathbb{I}'$ given by

$$k_\tau^{\tau+n}(x,y) := \begin{cases} k_\tau(x,y), & n = 1, \\ \underbrace{\int_\Omega \cdots \int_\Omega}_{n-1 \text{ times}} k_{\tau+n-1}(x, y_{n-1}) \dots k_{\tau+1}(y_2, y_1) k_\tau(y_1, y) \cdot & \\ \qquad \cdot\, \mathrm{d}\mu(y_{n-1}) \dots \mathrm{d}\mu(y_2)\,\mathrm{d}\mu(y_1), & n > 1. \end{cases}$$

4.1 Integral Operators

We now summarize basic properties of the integral operators $\mathcal{K}_t$. For this purpose it suffices to focus on the time-invariant situation

$$\mathcal{K}v := \int_\Omega k(\cdot, y)v(y)\,\mathrm{d}\mu(y). \tag{13}$$

Theorem 4 ([1, p. 275, Theorem 7.7]) *Let Ω be a compact metric space, μ be the Borel measure and $p \in [1, \infty]$. If $k \in C(\Omega^2)$, then $\mathcal{K} \in L(L^p(\Omega, \mu))$ is well-defined and compact.*

The Hilbert space $L^2(\Omega) = L^2(\Omega, \mu)$ with the Lebesgue measure μ is tackled in

Theorem 5 ([12, p. 47, Theorem 3.2.7]) *Let $\Omega \subseteq \mathbb{R}^\kappa$ be measurable. If $k \in L^2(\Omega^2)$, then $\mathcal{K} \in L(L^2(\Omega))$ is well-defined and compact with*

$$\|\mathcal{K}\| \le \sqrt{\int_\Omega \int_\Omega |k(x, y)|^2\,\mathrm{d}y\,\mathrm{d}x}.$$

In the setting of Theorems 4 and 5 the adjoint operator $\mathcal{K}^* \in L(L^2(\Omega))$ of $\mathcal{K}$ becomes

$$\mathcal{K}^*v = \int_\Omega \overline{k(y, \cdot)}v(y)\,\mathrm{d}y$$

and consequently $\mathcal{K}$ is

- *self-adjoint*, if and only if $k(x, y) = \overline{k(y, x)}$ for μ-almost all $(x, y) \in \Omega^2$. In this case one denotes the kernel k as *symmetric* and it follows that $\sigma(\mathcal{K}) \subset \mathbb{R}$
- *normal*, if and only if $k(x, y)\overline{k(z, y)} = \overline{k(y, x)}k(y, z)$ for μ-almost all $x, y, z \in \Omega$.

On the continuous functions we eventually obtain

Theorem 6 ([12, p. 45, Theorem 3.2.6]) *Let $\Omega \subset \mathbb{R}^\kappa$ be compact. If $k : \Omega^2 \to \mathbb{K}$ satisfies*

(i) $\int_\Omega |k(x, y)|\,\mathrm{d}y < \infty$
(ii) $\lim_{\xi \to x} \int_\Omega |k(\xi, y) - k(x, y)|\,\mathrm{d}y = 0$ *for all* $x \in \Omega$,

then $\mathcal{K} \in L(C(\Omega))$ is well-defined and compact.

The following consequence of Theorems 4 and 6 ensures that the spectrum of an integral operator $\mathcal{K}$ is independent of the state space:

Corollary 2 *For $k \in C(\Omega^2)$ one has $\|\mathcal{K}\|_{L(C(\Omega))} = \max_{x\in\Omega} \int_\Omega |k(x, y)|\,\mathrm{d}y$ and the spectrum of $\mathcal{K}$ is independent whether $\mathcal{K}$ is considered in $L(L^2(\Omega))$ or $L(C(\Omega))$.*

Proof See [12, p. 45, Lemma 3.2.2] for the assertion on the norm and [8, p. 113, Theorem 4.2.20]) concerning the spectrum. □

4.2 Asymptotics of Spectral Intervals

It is not difficult to construct difference equations (L) having an empty dichotomy spectrum (e.g. $\mathcal{K}_t \equiv 0$). However, whether $\Sigma(\mathcal{K})$ consists of a finite (case $(\mathfrak{S}_1)$, see Figs. 1 and 2) or an infinite number of spectral intervals (case $(\mathfrak{S}_2)$, see Fig. 3) depends on various factors. The relevance of this question is due to the fact that the gaps in the dichotomy spectrum $\Sigma(\mathcal{K})$ of a variational equation determines the number of invariant fiber bundles associated to the entire solution along which e.g. (1) is linearized.

In the prototypical situation of a multiplicative perturbation

$$\boxed{u_{t+1} = a_t \mathcal{K} u_t}$$

with a tempered sequence $(a_t)_{t\in\mathbb{I}'}$ in $\mathbb{K}$ it results from Theorem 3 that

$$\Sigma(a\mathcal{K}) = \bigcup_{j\in\mathbb{J}(\mathcal{K})} \sigma_j, \qquad \sigma_j := \left[\left|\lambda_j\right| \underline{\beta}(a), \left|\lambda_j\right| \overline{\beta}(a)\right].$$

Even for $\mathbb{J}(\mathcal{K}) = \mathbb{N}$ it is possible that consecutive intervals σ_j eventually overlap and yield a finite number of components and hence spectral intervals in $\Sigma(a\mathcal{K})$. Since the eigenvalues λ_j are ordered as in (H_1) we obtain: The intervals σ_j, σ_{j+1}

- merge in case $\max \sigma_{j+1} \geq \min \sigma_j$, which is equivalent to

$$\left|\lambda_j\right| \leq \frac{\overline{\beta}(a)}{\underline{\beta}(a)} \left|\lambda_{j+1}\right| \tag{14}$$

- stay apart for $\max \sigma_{j+1} < \min \sigma_j$, which holds if and only if

$$\left|\lambda_{j+1}\right| < \frac{\underline{\beta}(a)}{\overline{\beta}(a)} \left|\lambda_j\right|. \tag{15}$$

Hence, in order to have an infinite number of spectral intervals, one needs exponentially decaying eigenvalues of $\mathcal{K}$ with a suitable decay rate. This property depends on the smoothness of the kernel, as the following results illustrate:

- Let the compact set $\Omega \subset \mathbb{R}^\kappa$ be equipped with the Borel measure. If a continuous kernel $k : \Omega^2 \to \mathbb{K}$ satisfies a Hölder condition in the second variable with

$$\int_\Omega \|k(x, \cdot)\|_{C^\gamma} \, \mathrm{d}x < \infty$$

  for some exponent $\gamma \in (0, 1]$, then the eigenvalues of $\mathcal{K} \in L(L^2(\Omega, \mu))$ behave asymptotically like $\lambda_i = O(i^{-1/2-\gamma/\kappa})$ as $i \to \infty$ (see [13, Theorem 3]). For such positively definite kernels this can be improved to $\lambda_i = O(i^{-1-\gamma/\kappa})$ (see [7, Theorem 4]), which still cannot guarantee (15)

- Let $\Omega = [-1,1]$ and $k : \Omega^2 \to \mathbb{R}$ be of class C^1. If k is symmetric, $k(\cdot, y)$ has an analytic extension from $[-1,1]$ to the ellipse (foci ± 1, axis sum $R > 1$)

$$E_R := \left\{ z \in \mathbb{C} : \tfrac{(\Re z)^2}{a^2} + \tfrac{(\Im z)^2}{b^2} < 1 \right\}, \quad a := \tfrac{1}{2}(R + \tfrac{1}{R}), \quad b := \tfrac{1}{2}(R - \tfrac{1}{R})$$

and k is bounded on $E_R \times [-1,1]$, then $\lambda_i = O(R^{-i})$ (see [5, p. 68, Theorem 4.22]). An analytic extension to every such set thus yields super-exponential decay.

Further information on the asymptotic behavior of eigenvalues to integral operators can be found in the monograph [6].

4.3 Examples

In this section, we first collect miscellaneous examples of time-invariant integral operators (13) resp. corresponding kernel functions, for which both eigenvalues and -functions are explicitly known. Then several convolution kernels relevant for applications are discussed, which also allow to obtain information on the asymptotics of their spectrum. These operators fulfill the properties (H_1)–(H_2) from Sect. 3 and consequently the dichotomy spectra of the nonautonomous Eqs. (9) and (10) tackled in Theorems 1, 2 resp. 3 — which are now linear IDEs — can be determined.

By means of the following remark these results extend to wider classes of IDEs:

Remark 3 (*kinematic similarity*) Let $1 \le p < \infty$ and $\mathcal{K}_t \in L(L^p(\Omega, \mu))$. Suppose that $m_t \in L^\infty(\Omega, \mu)$ are $\mathbb{K}$-valued functions with $0 \notin \rho_{\mathrm{ess}}(m_t)$ for all $t \in \mathbb{I}'$ and

$$\sup_{t \in \mathbb{I}'} \rho_{\mathrm{ess}}(m_t) < \infty, \qquad \sup_{t \in \mathbb{I}'} \rho_{\mathrm{ess}}(m_t^{-1}) < \infty.$$

According to [10, pp. 30ff] the multiplication operators

$$\mathcal{M}_t \in L(L^p(\Omega, \mu)), \qquad [\mathcal{M}_t v](x) := m_t(x) v(x) \quad \text{for all } t \in \mathbb{I}', \ x \in \Omega$$

are well-defined and invertible. Consequently, due to

$$[\mathcal{M}_{t+1}^{-1} \mathcal{K}_t \mathcal{M}_t v](x) = \int_\Omega k_t(x, y) \frac{m_t(y)}{m_{t+1}(x)} v(y) \, \mathrm{d}\mu(y) \quad \text{for all } t \in \mathbb{I}', \ x \in \Omega$$

the linear IDE (L) and

$$u_{t+1} = \int_\Omega \frac{k_t(\cdot, y)}{m_{t+1}(\cdot)} m_t(y) u_t(y) \, \mathrm{d}\mu(y)$$

are kinematically similar and thus have the same dichotomy spectrum.

4.3.1 Explicitly Known Spectra

Assume that $(a_t)_{t\in\mathbb{I}'}$ is a tempered sequence in $\mathbb{K}$ with $\underline{\beta}(a) < \overline{\beta}(a)$.

Example 5 The Sturm–Liouville problem $-u'' = \lambda u$, $u(\alpha) = u(\beta) = 0$ leads to a continuous, symmetric Green's function (see Fig. 4 (left))

$$k(x, y) := \begin{cases} (y-\alpha)(\beta-x), & \alpha \le y \le x \le \beta, \\ (x-\alpha)(\beta-y), & \alpha \le x < y \le \beta. \end{cases}$$

Thanks to Theorem 5, on the interval $\Omega := (\alpha, \beta)$ the operator $\mathcal{K} \in L(L^2(\alpha, \beta))$ is compact with real eigenvalues $\lambda_j := \frac{(\beta-\alpha)^3}{\pi^2 j^2}$ of order $o_j = 1$ and normed eigenfunctions $e_j(x) := \sqrt{\frac{2}{\beta-\alpha}} \sin(\frac{\pi j}{\beta-\alpha}(x-\alpha))$, $j \in \mathbb{N}$. From (4) we obtain a discrete spectrum

$$\Sigma(\mathcal{K}) = \left\{ \frac{(\beta-\alpha)^3}{\pi^2 j^2} : j \in \mathbb{N} \right\}, \qquad \mathcal{X}_j := \mathbb{I} \times \operatorname{span}\left\{e_j\right\}$$

with simple spectral intervals. Moreover, (14) shows that $\Sigma(a\mathcal{K})$ is of the form $(\mathfrak{S}_1^2)$.

Example 6 On $\Omega := (\alpha, \beta)$ the analytical function (see Fig. 4 (right))

$$k(x, y) := \frac{1-\gamma^2}{1+\gamma^2 - 2\gamma\cos(\frac{2\pi}{\beta-\alpha}(x+y-2\alpha))} \quad \text{for all } \gamma \in (0, 1)$$

defines a symmetric kernel. By Theorem 5 the operator $\mathcal{K} \in L(L^2(\alpha, \beta))$ is compact, has real eigenvalues (of order $o_j = 1$) and eigenfunctions (cf. [2, pp. 254–255])

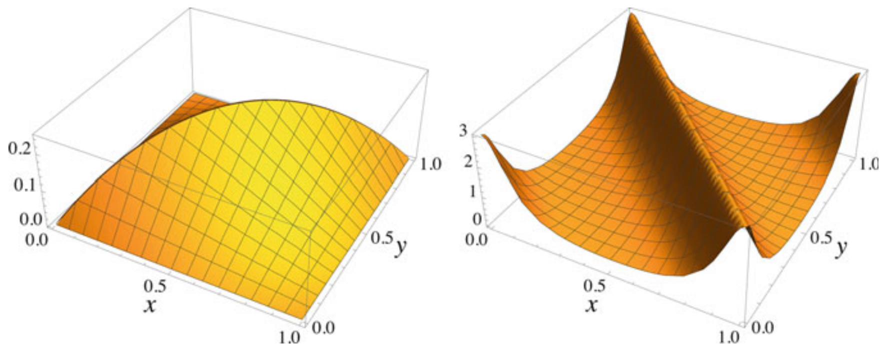


Fig. 4 The symmetric kernels $k : (0, 1)^2 \to \mathbb{R}$ from Example 5 (left) and Example 6 (right, for $\gamma = \frac{1}{2}$)

$$\lambda_j :=(\beta-\alpha)\begin{cases}\gamma^j, & j\geq 0,\\ -\gamma^{-j}, & j<0.\end{cases}, \qquad e_j(x):=\begin{cases}\sqrt{\frac{2}{\beta-\alpha}}\cos(\frac{2\pi j}{\beta-\alpha}(x-\alpha)), & j>0,\\ \sqrt{\frac{1}{\beta-\alpha}}, & j=0,\\ \sqrt{\frac{2}{\beta-\alpha}}\sin(\frac{2\pi j}{\beta-\alpha}(\alpha-x)), & j<0.\end{cases}$$

Note that the reals λ_j are exponentially decaying and symmetrically distributed around 0. It follows from $|\lambda_j| = |\lambda_{-j}|$ and (4) that

$$\Sigma(\mathcal{K}) = \left\{(\beta-\alpha)\gamma^j : j \in \mathbb{N}_0\right\}, \qquad \mathcal{X}_j = \mathbb{I}\times\begin{cases}\operatorname{span}\{e_0\}, & j=0,\\ \operatorname{span}\left\{e_j, e_{-j}\right\}, & j>0;\end{cases}$$

the dominant interval $\{\beta-\alpha\}$ is simple, while the other intervals have order 2. Furthermore, the concrete structure of $\Sigma(a\mathcal{K})$ depends on the ratio of the Bohl exponents. In case $\frac{\underline{\beta}(a)}{\overline{\beta}(a)} \leq \gamma$ it follows from (14) that $\Sigma(a\mathcal{K})$ is of the form $(\mathfrak{S}_1^2)$. For $\gamma < \frac{\underline{\beta}(a)}{\overline{\beta}(a)}$ however, (15) implies a countably infinite number of spectral intervals, where the dominant one $(\beta-\alpha)\left[\underline{\beta}(a), \overline{\beta}(a)\right]$ is simple, while the remaining ones are of order 2.

Example 7 On $\Omega :=(-\pi,\pi)$ consider the discontinuous kernel (see Fig. 5 (left))

$$k(x,y) :=\begin{cases}2, & -\pi \leq y \leq x \leq \pi,\\ 1, & -\pi \leq x < y \leq \pi,\end{cases}$$

which fails to be symmetric. It has the complex eigenvalues and -functions

$$\lambda_j = \frac{2\pi}{\ln 2 + 2\pi\imath j}, \qquad e_j(x) = \exp\left((\tfrac{\ln 2}{2\pi} + \imath j)x\right) \quad \text{for all } j \in \mathbb{Z}.$$

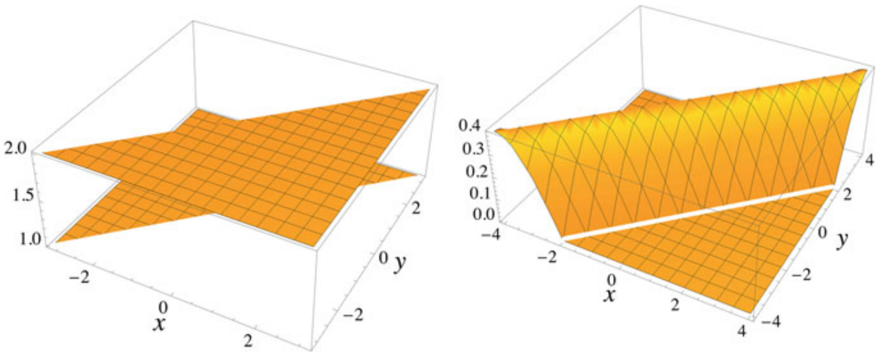


Fig. 5 The asymmetric kernel $k : (-\pi,\pi)^2 \to \mathbb{R}$ from Example 7 with $\alpha = 1$, $\beta = 2$ (left) and symmetric finite radius dispersal kernel from Example 8 (right) for $\alpha = 2$

Due to [8, p. 89, Theorem 3.3.15] the set $\{e_j\}_{j\in\mathbb{Z}}$ is a minimal complete set in $L^2(\Omega)$. Moreover, $|\lambda_j| = |\lambda_{-j}|$ and (4) imply

$$\Sigma(\mathcal{K}) = \left\{\frac{2\pi}{\sqrt{(\ln 2)^2+(2\pi j)^2}} : j \in \mathbb{N}_0\right\}, \quad \mathcal{X}_j = \mathbb{I} \times \begin{cases} \operatorname{span}\{e_0\}, & j = 0, \\ \operatorname{span}\{e_j, e_{-j}\}, & j > 0; \end{cases}$$

consequently, the dominant spectral interval $\{\frac{2\pi}{\ln 2}\}$ is simple, while the other spectral intervals have order 2. Moreover, since the eigenvalues decay merely linearly, it results that $\Sigma(a\mathcal{K})$ is of the form $(\mathfrak{S}_1^2)$.

We next discuss a class of kernels, where also a spectrum of the form $(\mathfrak{S}_1^1)$ (see Fig. 1) can be realized. Thereto, a kernel $k : \Omega^2 \to \mathbb{K}$ is denoted as *degenerate*, if it can be written as

$$k(x, y) := \sum_{j=1}^{d} a_j(y)x_j(x) \quad \text{for all } x, y \in \Omega$$

with linearly independent functions $x_1, \ldots, x_d : \Omega \to \mathbb{K}$. This brings us into the framework of finite rank operators discussed in Sect. 2.3 and Example 4 with

$$\mathcal{K}v = \int_\Omega \sum_{j=1}^{d} a_j(y)v(y)\,\mathrm{d}\mu(y)x_j = \sum_{j=1}^{d} \langle v, x_j'\rangle x_j : \Omega \to \mathbb{K}$$

and functionals $\langle v, x_j'\rangle := \int_\Omega a_j(y)v(v)\,\mathrm{d}\mu(y)$. The entries of the matrix $K \in \mathbb{K}^{d\times d}$ from Example 4 are $k_{ij} := \int_\Omega a_i(y)x_j(y)\,\mathrm{d}\mu(y)$, $1 \le i, j \le d$, yield the discrete spectrum

$$\Sigma(\mathcal{K}) = |\{\lambda \in \mathbb{C} : \det(\lambda I_{\mathbb{C}^d} - K) = 0\} \setminus \{0\}| .$$

Example 8 (*finite radius dispersal kernel*) Let $\Omega = (-1, 1)$. The kernel

$$k(x, y) := \begin{cases} \frac{\pi}{4\alpha}\cos\left(\frac{\pi(x-y)}{2\alpha}\right), & |x-y| \le \alpha, \\ 0, & |x-y| > \alpha \end{cases}$$

(cf. [17], see Fig. 5 (right)) is continuous and symmetric. Moreover, due to

$$k(x, y) = \begin{cases} \frac{\pi}{4\alpha}\left(\cos\frac{\pi x}{2\alpha}\cos\frac{\pi y}{2\alpha} + \sin\frac{\pi x}{2\alpha}\sin\frac{\pi y}{2\alpha}\right), & |x-y| \le \alpha, \\ 0, & |x-y| > \alpha \end{cases}$$

it is degenerate. Hence, for $\alpha \geq 2$ the integral operator $\mathcal{K}$ allows the representation

$$\mathcal{K}v = \sum_{j=1}^{2} \left(\int_{-1}^{1} a_j(y)v(y)\,\mathrm{d}y \right) x_j : \Omega \to \mathbb{K}$$

with $a_1(x) :=\cos \frac{\pi x}{2\alpha}$, $a_2(x) = \sin \frac{\pi x}{2\alpha}$ and the linearly independent functions

$$x_1(x) :=\tfrac{\pi}{4\alpha} \cos \tfrac{\pi x}{2\alpha}, \qquad\qquad x_2(x) :=\tfrac{\pi}{4\alpha} \sin \tfrac{\pi x}{2\alpha}.$$

Therefore, $\mathcal{K}$ is a rank 2 operator and its eigenvalues λ are the roots of the equation

$$\det \begin{pmatrix} \lambda - \int_{-1}^{1} a_1(y)x_1(y)\,\mathrm{d}y & -\int_{-1}^{1} a_1(y)x_2(y)\,\mathrm{d}y \\ -\int_{-1}^{1} a_2(y)x_1(y)\,\mathrm{d}y & \lambda - \int_{-1}^{1} a_2(y)x_2(y)\,\mathrm{d}y \end{pmatrix} = 0.$$

In the following example the eigenvalues are not explicitly known, but can be obtained as solutions of a transcendental equation in $\mathbb{R}$ yielding also their asymptotics.

Example 9 On $\Omega :=(0, 1)$ the continuous kernel

$$k(x, y) :=\tfrac{1}{2} \min\{x, y\}\,(2 - \max\{x, y\})$$

is symmetric. Suppose that $(\nu_j)_{j\in\mathbb{N}}$ denotes the strictly increasing sequence of positive real solutions to the transcendental equation $\nu + \tan \nu = 0$ (see Fig. 6). The associate integral operator $\mathcal{K}$ has the eigenvalues $\lambda_j :=\frac{1}{\nu_j^2}$ of order $o_j = 1$ with normed eigenfunctions $e_j(x) = 2\sqrt{\frac{\nu_j}{(2\nu_j - \sin(2\nu_j))}} \sin(\nu_j x)$, $j \in \mathbb{N}$ (see [19, p. 438]). This yields a discrete dichotomy spectrum with simple spectral intervals

$$\Sigma(\mathcal{K}) = \left\{ \nu_j^{-2} : j \in \mathbb{N} \right\}, \qquad\qquad \mathcal{X}_j :=\mathbb{I} \times \operatorname{span}\left\{e_j\right\}.$$

In addition, (14) implies that $\Sigma(a\mathcal{K})$ is of the form $(\mathfrak{S}_1^2)$.

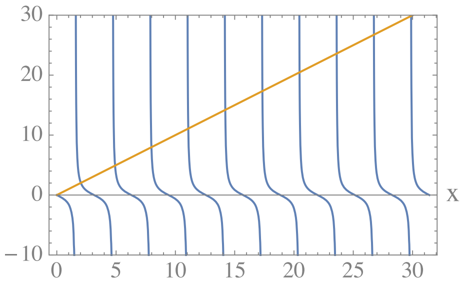


Fig. 6 The points of intersection $\nu_j > 0$ of the graphs to $x \mapsto x$ and $x \mapsto -\tan x$ yield the eigenvalues in Example 9

4.3.2 Spectra of Convolutive Operators

In the remaining, we suppose $\Omega = (-1, 1)$ and consider kernels of convolution type

$$[\mathcal{K}v](x) := \int_{-1}^{1} k_0(x - y)v(y)\,\mathrm{d}y \quad \text{for all } x \in (-1, 1)$$

with a real, even and integrable function $k_0 : \mathbb{R} \to \mathbb{R}$. These kernels frequently arise in applications [17] from theoretical ecology and have a real spectrum. In addition, we approximate their (largest) eigenvalues numerically using a Nyström method with the rectangular rule as quadrature and 1000 nodes.

Following [27], the (scaled) *Fourier transformation* of k_0 becomes

$$\tilde{k}_0(\xi) := \int_{\mathbb{R}} e^{\imath \xi x} k_0(x)\,\mathrm{d}x$$

and provided it is positive, we define $\Gamma(\xi) := -\ln \tilde{k}_0(\xi)$.

Example 10 (*Gauß kernel*) As archetypical mesokurtic distribution consider

$$k_0(x) := \frac{1}{\sqrt{2\pi\alpha^2}} \exp\left(-\frac{x^2}{2\alpha^2}\right) \quad \text{for all } \alpha > 0 \tag{16}$$

(see Fig. 7) with standard deviation $\alpha > 0$. It is real analytical with $\operatorname{lip} k_0 \le \frac{1}{\sqrt{2e\pi\alpha^2}}$, the Fourier transformation $\tilde{k}_0(\xi) = e^{-\frac{\alpha^2}{2}\xi^2}$ is bounded, even and positive, whence it is $\Gamma(\xi) = \frac{\alpha^2}{2}\xi^2$. Since Γ is convex and satisfies $\lim_{\xi\to\infty} \frac{\Gamma(\xi)}{\xi} = \infty$, it follows from [27, Corollary 1] that $\ln \lambda_j \sim -j \ln j$ as $j \to \infty$. Consequently, $\Sigma(\mathcal{K})$ and $\Sigma(a\mathcal{K})$ consists of an infinite number of spectral intervals accumulating at 0, i.e. both dichotomy spectra are of the form $(\mathfrak{S}_2)$ with $\bar{\mu} = 0$.

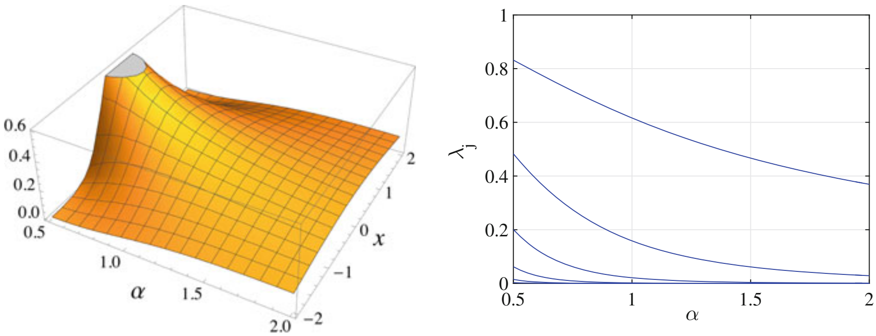


Fig. 7 The Gaussian convolution kernel $k_0 : \mathbb{R} \to \mathbb{R}$ from Example 10 (left) and the super-exponentially decaying largest eigenvalues λ_j depending on $\alpha \in [\frac{1}{2}, 2]$

Example 11 (*Cauchy kernel*) Another smooth kernel is the *Cauchy kernel*

$$k_0(x) :=\frac{\alpha}{\pi(\alpha^2+x^2)} \quad \text{for all } \alpha > 0$$

(see Fig. 8) resembling the Gauß kernel (16). The Fourier transform $\tilde{k}_0(\xi) = e^{-\alpha|\xi|}$ is bounded, even and positive with $\Gamma(\xi) = \alpha\,|\xi|$. From [27, Theorem 2] we hence obtain $\ln\lambda_j \sim -j\psi(\alpha)$ as $j \to \infty$ with the function $\psi(\alpha) :=\pi\frac{E(\operatorname{sech}(\pi/\alpha))}{E(\tanh(\pi/\alpha))} > 0$, where E stands for the complete elliptic integral of first kind. It results from (14) that $\Sigma(a\mathcal{K})$ is of the form $(\mathfrak{S}_1^2)$ for $e^{\psi(\alpha)} \le \frac{\overline{\beta}(a)}{\underline{\beta}(a)}$, while (15) and $\frac{\overline{\beta}(a)}{\underline{\beta}(a)} < e^{\psi(\alpha)}$ guarantee $(\mathfrak{S}_2)$, i.e. an infinite number of spectral intervals.

Example 12 (*Laplace kernel*) The *Laplace kernel* is given by the function

$$k_0(x) :=\tfrac{1}{2\alpha}\exp\left(-\tfrac{|x|}{\alpha}\right) \quad \text{for all } \alpha > 0$$

(see Fig. 9), which is continuous with $\operatorname{lip} k_0 \le \frac{1}{2\alpha^2}$. If $(\nu_j)_{j\in\mathbb{N}}$ denotes the strictly increasing sequence of positive solutions to the transcendental equation $\tan\frac{\nu}{\alpha} = \pm\nu$, then $\mathcal{K}$ possesses the eigenvalues $\lambda_j :=\frac{1}{1+\nu_j^2}$, $j \in \mathbb{N}$ (see [17]). On the one hand, this shows that λ_j decays quadratically to 0. On the other hand, the Fourier transform of k_0 is $\tilde{k}_0(\xi) = \frac{1}{1+\alpha^2\xi^2}$ and hence $\Gamma(\xi) = \ln(1+\alpha^2\xi^2)$. Referring to [27, Theorem I] it results that $\lambda_j \sim \tilde{k}_0(\frac{\pi}{2}j + o(j))$ as $j \to \infty$, which confirms the quadratic decay. Due to (14) this results in a dichotomy spectrum $\Sigma(a\mathcal{K})$ of the form $(\mathfrak{S}_1^2)$.

Example 13 (*exponential square root kernel*) For the kernel

$$k_0(x) :=\tfrac{1}{4\alpha}\exp\left(-\sqrt{\tfrac{|x|}{\alpha}}\right) \quad \text{for all } \alpha > 0$$

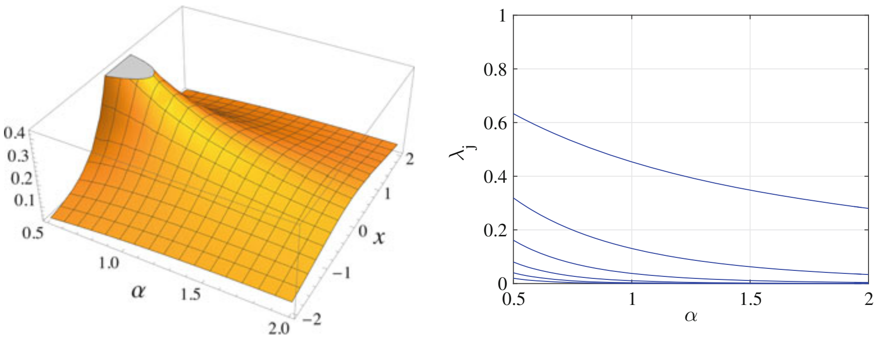


Fig. 8 The Cauchy convolution kernel $k_0 : \mathbb{R} \to \mathbb{R}$ from Example 11 (left) and the exponentially decaying largest eigenvalues λ_j depending on $\alpha \in [\frac{1}{2}, 2]$

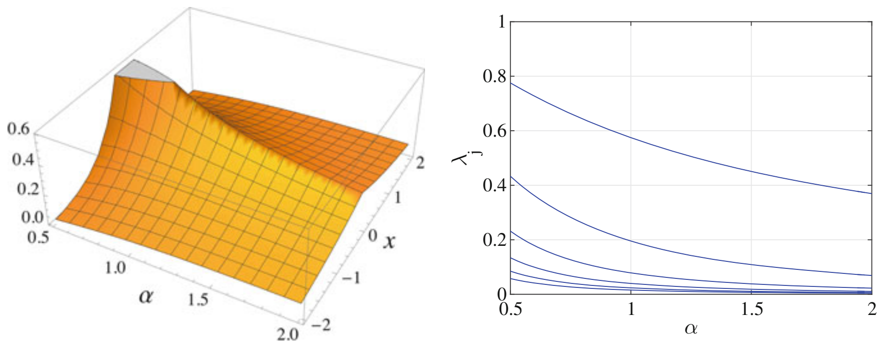


Fig. 9 The Laplacian convolution kernel $k_0 : \mathbb{R} \to \mathbb{R}$ from Example 12 (left) and the quadratically decaying largest eigenvalues λ_j depending on $\alpha \in [\frac{1}{2}, 2]$

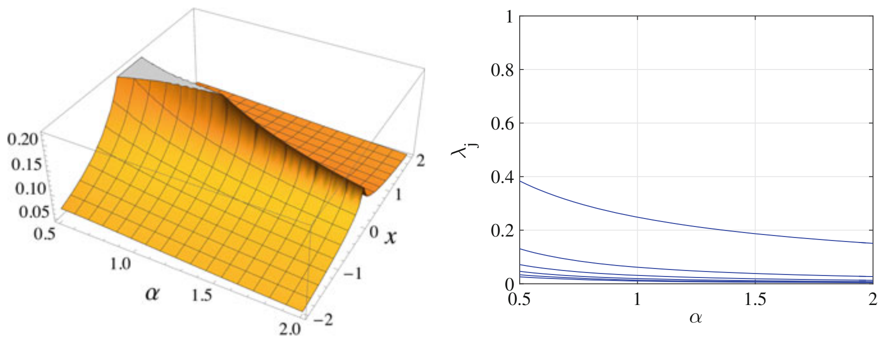


Fig. 10 The exponential square root convolution kernel $k_0 : \mathbb{R} \to \mathbb{R}$ from Example 13 (left) and the six largest eigenvalues λ_j depending on $\alpha \in [\frac{1}{2}, 2]$

(see Fig. 10) the tails are not exponentially bounded. It is continuous with a Hölder condition $\operatorname{hol}_{1/2} k_0 \le \frac{1}{4\alpha^{3/2}}$, but not differentiable in 0. The Fourier transformation

$$\tilde{k}_0(\xi) = \sqrt{2\pi}\,\frac{\sin\left(\frac{1}{4\alpha|\xi|}\right)\left(1 - 2S\left(\frac{1}{\sqrt{2\pi\alpha|\xi|}}\right)\right) + \cos\left(\frac{1}{4\alpha|\xi|}\right)\left(1 - 2C\left(\frac{1}{\sqrt{2\pi\alpha|\xi|}}\right)\right)}{|\alpha\xi|^{3/2}}$$

is bounded, even and positive, where S, C denote the Fresnel integrals. In this setting, [27, Theorem I] leads to $\lambda_j \sim \tilde{k}_0(\frac{\pi}{2} j + o(j))$ as $j \to \infty$.

Example 14 (*top hat kernel*) Let $\alpha \in (0, 1]$. The *top hat kernel* is defined as

$$k_0(x) := \frac{1}{2\alpha}\left(\theta(x+\alpha) - \theta(x-\alpha)\right) = \frac{1}{2\alpha}\chi_{[-\alpha,\alpha]}(x) \quad \text{for all } \alpha > 0$$

(see Fig. 11) and has the Fourier transform $\tilde{k}_0(\xi) = \frac{\sin(\alpha\xi)}{\alpha\xi}$, which is bounded, even, but fails to be positive. Hence, the results from [27] do not apply.

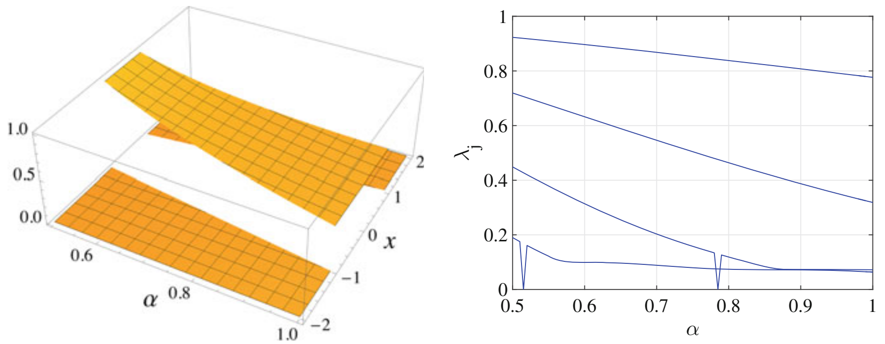


Fig. 11 The top hat convolution kernel $k_0 : \mathbb{R} \to \mathbb{R}$ from Example 14 (left) and the six largest eigenvalues λ_j depending on $\alpha \in [\frac{1}{2}, 1]$. The spikes appear to be due to numerical inaccuracies

References

1. Abramovich, Y., Aliprantis, C.: An Invitation to Operator Theory. Graduate Studies in Mathematics, vol. 50. AMS, Providence (2002)
2. Atkinson, K.: The Numerical Solution of Integral Equations of the Second Kind. Monographs on Applied and Computational Mathematics, vol. 4. University Press, Cambridge (1997)
3. Aulbach, B., Kalkbrenner, J.: Exponential forward splitting for noninvertible difference equations. Comput. Math. Appl. **42**, 743–754 (2001)
4. Aulbach, B., Siegmund, S.: The dichotomy spectrum for noninvertible systems of linear difference equations. J. Differ. Equ. Appl. **7**(6), 895–913 (2001)
5. Baumeister, J.: Stable Solution of Inverse Problems. Advanced Lectures in Mathematics. Vieweg and Sohn, Braunschweig (1987)
6. Birman, M.S., Solomjak, M.Z.: Quantitative Analysis in Sobolev Imbedding Theorems and Applications to Spectral Theory. AMS Translations, vol. 114. AMS, Providence (1980)
7. Cobos, F., Kühn, T.: Eigenvalues of integral operators with positive definite kernels satisfying integrated Hölder conditions over metric compacta. J. Approx. Theory **63**, 39–55 (1990)
8. Davies, E.: Linear Operators and Their Spectra. University Press, Cambridge (2007)
9. Dieci, L., Elia, C., Van Vleck, E.: Detecting exponential dichotomy on the real line: SVD and QR algorithms. BIT Numer. Math. **51**(3), 555–579 (2011)
10. Engel, K., Nagel, R.: One-Parameter Semigroups for Linear Evolution Equations. Graduate Texts in Mathematics, vol. 194. Springer, Berlin (2000)
11. Gil', M.: Difference Equations in Normed Spaces – Stability and Oscillation. Mathematics Studies, vol. 206. North-Holland, Amsterdam (2007)
12. Hackbusch, W.: Integral Equations - Theory and Numerical Treatment. Birkhäuser, Basel (1995)
13. Heinrich, S., Kühn, T.: Embedding maps between Hölder spaces over metric compacta and eigenvalues of integral operators. Indagationes Mathematicae **88**(1), 47–62 (1982)
14. Henry, D.: Geometric Theory of Semilinear Parabolic Equations. Lecture Notes in Mathematics. vol. 840, Springer, Berlin (1981)
15. Hüls, T.: Computing Sacker–Sell spectra in discrete time dynamical systems. SIAM J. Numer. Anal. **48**(6), 2043–2064 (2010)
16. Jacobsen, J., Jin, Y., Lewis, M.: Integrodifference models for persistence in temporally varying river environments. J. Math. Biol. **70**, 549–590 (2015)
17. Kot, M., Schaffer, W.: Discrete-time growth-dispersal models. Math. Biosci. **80**, 109–136 (1986)

18. Kubrusly, C.: The Elements of Operator Theory, 2nd edn. Birkhäuser, New York (2011)
19. Kythe, P., Schäferkotter, M.: Handbook of Computational Methods for Integration. Chapman and Hall/CRC, Boca Raton (2006)
20. Pötzsche, C.: Geometric Theory of Discrete Nonautonomous Dynamical Systems. Lecture Notes in Mathematics, 2002. Springer, Berlin (2010)
21. Pötzsche, C.: A note on the dichotomy spectrum. J. Differ. Equ. Appl. **15**(10), 1021–1025 (2009) (see also the corrigendum in J. Differ. Equ. Appl. **18**(7), 1257–1261 (2012))
22. Pötzsche, C.: Dichotomy spectra of triangular equations. Discret. Contin. Dyn. Syst. (Ser. A) **36**(1), 423–450 (2016)
23. Pötzsche, C., Ruß, E.: Notes on spectrum and exponential decay in non autonomous evolutionary equations. In: Krisztin, T. (ed) Proceedings of the 10th Coll. Qualitative Theory of Differential Equations, p. 15. (Szeged, Hungary, 2015) (2016)
24. Ruß, E.: Dichotomy spectrum for difference equations in Banach spaces. J. Differ. Equ. Appl. **23**(3), 576–617 (2016)
25. Sacker, R., Sell, G.: A spectral theory for linear differential systems. J. Differ. Equ. **27**, 320–358 (1978)
26. Siegmund, S.: Block diagonalization of linear difference equations. J. Differ. Equ. Appl. **8**(2), 177–189 (2002)
27. Widom, H.: Asymptotic behavior of eigenvalues of certain integral equations II. Arch. Ration. Mech. Anal. **17**(3), 215–229 (1964)

A Dynamical Trichotomy for Structured Populations Experiencing Positive Density-Dependence in Stochastic Environments

Sebastian J. Schreiber

Abstract Positive density-dependence occurs when individuals experience increased survivorship, growth, or reproduction with increased population densities. Mechanisms leading to these positive relationships include mate limitation, saturating predation risk, and cooperative breeding and foraging. Individuals within these populations may differ in age, size, or geographic location and thereby structure these populations. Here, I study structured population models accounting for positive density-dependence and environmental stochasticity i.e. random fluctuations in the demographic rates of the population. Under an accessibility assumption (roughly, stochastic fluctuations can lead to populations getting small and large), these models are shown to exhibit a dynamical trichotomy: (i) for all initial conditions, the population goes asymptotically extinct with probability one, (ii) for all positive initial conditions, the population persists and asymptotically exhibits unbounded growth, and (iii) for all positive initial conditions, there is a positive probability of asymptotic extinction and a complementary positive probability of unbounded growth. The main results are illustrated with applications to spatially structured populations with an Allee effect and age-structured populations experiencing mate limitation.

Keywords Structured populations · Environmental stochasticity · Allee effects · Positive density-dependence

1 Introduction

Higher population densities can increase the chance of mating success, reduce the risk of predation, and increase the frequency of cooperative behavior [5]. Hence, survivorship, growth, and reproductive rates of individuals can exhibit a positive

S.J. Schreiber (✉)
Department of Evolution and Ecology, University of California, One Shields Avenue, Davis, CA 95616, USA
e-mail: sschreiber@ucdavis.edu

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_3

relationship with density i.e. positive density-dependence. In single species models, positive density-dependence can lead to an Allee effect: the existence of a critical density below which the population tends toward extinction and above which the population persists [7, 12, 18, 22, 23]. Consequently, the importance of Allee effects have been widely recognized for conservation of at risk populations and the management of invasive species [5]. Populations experiencing environmental stochasticity and a strong Allee effect are widely believed to be especially vulnerable to extinction as the fluctuations may drive their densities below the critical threshold [6]. When population densities lie above the critical threshold for the unperturbed system, analyses and simulations of stochastic models support this conclusion [2, 7–9, 17, 21]. However, these studies also show that when population densities lie below the critical threshold, stochastic fluctuations can rescue the population from the deterministic vortex of extinction.

Individuals within populations often differ in diversity of attributes including age, size, gender, and geographic location [4]. Positive density-dependence may differentially impact individuals in populations structured by these attributes [5, 11]. This positive density-dependence can lead to an Allee threshold surface (usually a co-dimension one stable manifold of an unstable equilibrium) that separates population states that lead to extinction from those that lead to persistence [24].

While several studies have examined how environmental stochasticity and population structure interact to influence persistence of populations experiencing negative-density dependence [3, 13, 14, 20], I know of no studies that examine this issue for populations experiencing positive density-dependence. To address this gap, this paper examines stochastic, single species models of the form

$$X_{t+1} = A(X_t, \xi_{t+1})X_t \tag{1}$$

where $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{n,t}) \in [0, \infty)^n$ is a column-vector of population densities, $A(X_t, \xi_{t+1})$ is a $n \times n$ non-negative matrix that determines the population densities in the next year as a function of the current densities X_t and the environmental state ξ_{t+1} over the time interval $[t, t+1)$. To focus on the effects of positive density-dependence, I assume that the entries of A are non-decreasing functions of the population densities. Under additional suitable assumptions described in Sects. 2 and 3, this paper shows that there is a dynamical trichotomy for (1): (i) asymptotic extinction occurs with probability one for all initial conditions, (ii) long-term persistence occurs with probability one for all positive initial conditions, or (iii) long-term persistence and asymptotic extinction occur with complementary positive probabilities for all positive initial conditions. The model assumptions and definitions are presented in Sect. 2. Exemplar models of spatially-structured populations and age-structured populations are also presented in this section. The main results and applications to the exemplar models appear in Sects. 3 and 4. Proofs of the main results are relegated to Sect. 5.

2 Models, Assumptions, and Definitions

Throughout this paper, I consider stochastic difference equations of the form given by Eq. (1). The state space for these equations is the non-negative cone $C = [0, \infty)^n$. Define the standard ordering on this cone by $x \geq y$ for $x, y \in C$ if $x_i \geq y_i$ for all i. Furthermore, $x > y$ if $x \geq y$ but $x \neq y$ and $x \gg y$ if $x_i > y_i$ for all i. Throughout, I will use $\|x\| = \max_i |x_i|$ to denote the sup norm and $\|A\| = \max_{\|x\|=1} \|Ax\|$ to denote the associated operator norm. Define the co-norm of a matrix A by $\text{co}(A) = \min_{\|x\|=1} \|Ax\|$. The co-norm is the minimal amount that the matrix A stretches a vector. Define $\log^+ x = \max\{\log x, 0\}$ to be the non-negative component of $\log x$.

For (1), there are *five standing assumptions*

A1 Uncorrelated environmental fluctuations: $\{\xi_t\}_{t=0}^{\infty}$ is a sequence of independent and identically distributed (i.i.d) random variables taking values in a separable metric space E (such as $\mathbb{R}^k$).

A2 Feedbacks depend continuously on population and environmental state: the entries of the matrix function $A_{ij} : C \times E \to [0, \infty)$ are continuous functions of population state x and the environmental state ξ.

A3 The population only experiences positive feedbacks: For all i, j and $\xi \in E$, $A_{ij}(x, \xi) \geq A_{ij}(y, \xi)$ whenever $x \geq y$.

A4 Primitivity: There exists $\tau \geq 0$ such that $A(x, \xi)^\tau \gg 0$ for all $x \gg 0$ and $\xi \in E$.

A5 Finite logarithmic moments: For all $c \geq 0$, $\mathbb{E}[\log^+ \|A(c\mathbf{1}, \xi_t)\|] < \infty$ where $\mathbf{1} = (1, 1, \ldots, 1)$ is the vector of ones. There exists $c^* > 0$ such that $\mathbb{E}[\log^+(1/\text{co}\left(\prod_{t=1}^{\tau} A(c\mathbf{1}, \xi_t)\right))] < \infty$ for all $c \geq c^*$.

The first assumption implies that $(X_t)_{t\geq 0}$ is a Markov chain on C and the second assumption ensures this stochastic process is Feller. The third assumption is consistent with the intent of understanding how non-negative feedbacks, in and of themselves, influence structured population dynamics. An important implication of this assumption is that the system is monotone i.e. if $X_0 > \widetilde{X}_0 > 0$, then $X_t \geq \widetilde{X}_t$ for all $t \geq \tau$ where X_t, $\widetilde{X}_t$ are solutions to (1) with initial conditions X_0 and $\widetilde{X}_0$, respectively. The fourth assumption ensures that all states in the population contribute to all other population states after τ time steps. The final assumption is met for most models and ensures that Kingman's subadditive ergodic theorem![16] and the random Perron–Frobenius theorem [1] are applicable.

To see that these assumption include models of biological interest, here are a few examples.

Example 1 (*Scalar models*) Considered an unstructured population with $n = 1$ in which case $x \in [0, \infty)$. To model mate limitation, McCarthy et al. [18] considered a model where x corresponds to the density of females and, with the assumption of a 1:1 sex ratio, also equals the density of males. The probability of a female successfully mating is given by $ax/(1 + ax)$ where x is the male density and $a > 0$ determines how effectively individuals find mates. If a mated individual produces on average ξ daughters, then the population density in the next year is $\xi ax^2/(1 + ax)$.

If we allow ξ to be stochastic, then (1) is determined by $A(x,\xi) = \xi a x^2/(1+ax)$. Allowing the ξ_t to be a log-normal would satisfy assumptions **A1**–**A5**.

To model predator saturation [23], let $\exp(-M/(1+hx))$ be the probability that an individual escapes predation from a predator population with an "effective" attack rate of M and handling time h. If ξ is the number of offspring produced by an individual which escaped predation, then the population density in the next year is $\xi x \exp(-M/(1+hx))$. Letting ξ be stochastic yields $A(x,\xi) = \xi x \exp(-M/(1+hx))$. Allowing the ξ_t to be a log-normal would satisfy assumptions **A1**–**A5**.

Finally, Leibhold and Bascompte [17] used a more phenomenological model of the form $A(x,\xi) = \exp(x - C + \xi)$ where C is the critical threshold in the absence of stochasticity and ξ are normally distributed with mean zero. This model also satisfies all of the assumptions.

We can use these scalar models, which were studied by [21], to build structured models as the next two examples illustrate.

Example 2 (*Spatial models*) Consider a population that lives in n distinct patches. x_i is the population density in patch i. Let $C_i > 0$ be the critical threshold in patch i and ξ_i be the environmental state in patch i. Let d_{ij} be the fraction of individuals dispersing from patch j to patch i, and $D = (d_{ij})$ be the corresponding dispersal matrix. Then the spatial model is

$$A(x,\xi) = D\text{diag}(\exp(x_1 - C_i + \xi_1), \exp(x_2 - C_2 + \xi_2), \dots, \exp(x_n - C_n + \xi_n)) \tag{2}$$

where diag denotes a diagonal matrix with the indicated diagonal elements. If D is a primitive matrix and the $\xi_t = (\xi_{1,t}, \dots, \xi_{n,t})$ are a multivariate normals with zero means, then this model satisfies the assumptions.

Example 3 (*Age-structured models*) Consider a population with n age classes and x_i is the density of age i individuals. Assume that final ℓ age classes reproduce i.e. ages $n-\ell+1, n-\ell+2, \dots, n$ reproduce. If mate limitation causes positive density dependence (see Example 1) and reproductively mature individuals mate randomly, then the fecundity of individuals in age class $n-\ell+i$ equals $f_i(x,\xi) = \xi_i a \sum_{j=n-\ell+1}^{n} x_j/(1 + a\sum_{j=n-\ell+1}^{n} x_j)$ where ξ_i is the maximal fecundity of individuals of age i and $a > 0$. Let s_i be the probability an individual survives from age $i-1$ to age i. This yields the following nonlinear Leslie matix model

$$A(x,\xi) = \begin{pmatrix} 0 & \dots & 0 & f_1(x,\xi) & \dots & f_\ell(x,\xi) \\ s_2 & 0 & 0 & \dots & 0 & 0 \\ 0 & s_3 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & s_n & 0 \end{pmatrix}. \tag{3}$$

If $\ell \geq 2$ and $\xi_t = (\xi_{1,t}, \dots, \xi_{n,t})$ are multivariate log-normals, then this model satisfies the assumptions **A1**–**A5**.

3 Main Results

To state the main results, consider the linearization of (1) at the origin and near infinity. At the origin, the linearized dynamics are given by $X_{t+1} = A(0, \xi_{t+1})X_t$. Hence, the rate at which the population grows at low density is approximately given by the rate at which the random product of matrices, $A(0, \xi_t) \dots A(0, \xi_1)$, grows. Kingman's subadditive ergodic theorem [16] implies there exists r_0 (possibly $-\infty$) such that

$$\lim_{t\to\infty} \frac{1}{t} \log \|A(0, \xi_t) \dots A(0, \xi_1)\| = r_0 \text{ with probability one.}$$

To characterize population growth near infinity, for all $c > 0$ the subadditive ergodic theorem implies there exists an r_c such that

$$\lim_{t\to\infty} \frac{1}{t} \log \|A(c\mathbf{1}, \xi_t) \dots A(c\mathbf{1}, \xi_1)\| = r_c \text{ with probability one.}$$

Due to our assumption that the entries of $A(x, \xi)$ are non-decreasing with respect the entries of x, r_c is non-decreasing with respect to c. Hence, the following limit exists (possibly $+\infty$)

$$r_\infty = \lim_{c\to\infty} r_c.$$

With these definitions and assumptions, the following theorem is proven in Sect. 5.

Theorem 1 *Unconditional persistence* *If* $r_0 > 0$, *then*

$$\lim_{t\to\infty} \|X_t\| = \infty \textit{ with probability one whenever } X_0 \gg 0.$$

Unconditional extinction *If* $r_\infty < 0$, *then*

$$\lim_{t\to\infty} X_t = 0 \textit{ with probability one.}$$

Conditional persistence and extinction *If* $r_0 < 0 < r_\infty$, *then for all* $\varepsilon > 0$ *there exist* $c^* > c_* > 0$ *such that*

$$\mathbb{P}\left[\lim_{t\to\infty} X_t = 0 \middle| X_0 = x\right] \geq 1 - \varepsilon \textit{ whenever } x \leq c_*\mathbf{1}$$

and

$$\mathbb{P}\left[\lim_{t\to\infty} \|X_t\| = \infty \middle| X_0 = x\right] \geq 1 - \varepsilon \textit{ whenever } x \geq c^*\mathbf{1}.$$

To get statements about all initial conditions with probability one in the final case, an assumption that ensures that the environmental stochasticity can drive the population to low or high densities is needed. Define $\{0, \infty\}$ to be *accessible* if for all $c > 0$ there exists $\gamma > 0$ such that

$$\mathbb{P}\left[\{\text{ there is } t \geq 0 \text{ such that } X_t \gg c\mathbf{1} \text{ or } X_t \ll \mathbf{1}/c\} \middle| X_0 = x\right] \geq \gamma$$

for all $x \gg 0$. All of the examples in Sect. 2 satisfy this accessibility condition.

Theorem 2 *If $r_0 < 0 < r_\infty$ and $\{0, \infty\}$ is accessible, then*

$$\mathbb{P}\left[\lim_{t\to\infty} \|X_t\| = \infty \text{ or } \lim_{t\to\infty} X_t = 0 \middle| X_0 = x\right] = 1.$$

Proofs of both theorems are presented in Sect. 5. The scalar version of these theorems were proven in Theorem 3.2 of [21].

4 Applications

To illustrate the applicability of the two theorems, we consider the spatial structured and age structured models introduced in Sect. 3.

Example 2 (spatially structured populations) revisited Consider the spatial structured model described in Example 2 and characterized by (2). For this model,

$$A(c\mathbf{1}, \xi) = D\text{diag}(\exp(-C_1 + \xi_1), \exp(-C_2 + \xi_2), \dots, \exp(-C_n + \xi_n)) \exp(c).$$

For simplicity, let us assume that the fraction of individuals dispersing is d and dispersing individuals land with equal likelihood on any patch (including the possibility of returning to its original patch). Then $D = (d_{ij})$ is given by $d_{ij} = d/n$ for $i \neq j$ and $d_{ii} = (1 - d) + d/n$. Assume that $d \in (0, 1]$.

I claim that $r_\infty = \infty$. Indeed, let $b = \max\{1 - d, d/n\} > 0$. Then $D \geq b\text{Id}$ where Id denotes the identity matrix and

$$\begin{aligned}
\mathbb{E}[\log \| \prod_{s=1}^{t} A(c\mathbf{1}, \xi_s) \|] &\geq \mathbb{E}[\log \| \prod_{s=1}^{t} b\text{diag}(\exp(-C_1 + \xi_{1,s}), \exp(-C_2 \\
&\quad +\xi_{2,s}), \dots, \exp(-C_n + \xi_{n,s})) \exp(c) \|] \\
&\geq \mathbb{E}[\log \| \prod_{s=1}^{t} \text{diag}(\exp(\xi_{1,s}), \exp(\xi_{2,s}), \dots, \exp(\xi_{n,s})) \|] \\
&\quad + t(c + \log b - \max_i C_i) \\
&= \mathbb{E}[\max_i \sum_{s=1}^{t} \xi_{i,s}] + t(c + \log b - \max_i C_i) \\
&\geq t\left(\mathbb{E}[\xi_{1,1}] + c + \log b - \max_i C_i\right).
\end{aligned}$$

Dividing by t and taking the limit as $t \to \infty$, this inequality implies that $r_c \geq \mathbb{E}[\xi_{1,1}] + c + \log b - \max_i C_i$. Hence, $r_\infty = \lim_{c\to\infty} r_c = \infty$ as claimed. Theorem 2 implies that for all $x \gg 0$, $\|X_t\| \to \infty$ with positive probability whenever $X_0 = x$.

Understanding r_0 is more challenging. However, Proposition 3 of [3] implies that r_0 varies continuously as a function of d. In the limit of $d = 0$, $D = \text{Id}$ and $r_0 = \max_i \mathbb{E}[\xi_{i,1} - C_i]$. Hence, for populations where $d \approx 0$ but $d > 0$, there are two types of dynamics. If $\mathbb{E}[\xi_{i,1}] < C_i$ for all patches (i.e. populations are unable to persist in each patch at low density), then there is a positive probability of going either asymptotically extinct or a complementary positive probability of persistence. Alternatively, if $\mathbb{E}[\xi_{i,1}] > C_i$ for at least one patch, then the population persists with probability one whenever $X_0 \gg 0$.

Now consider the case that all individuals disperse i.e. $d = 1$. Then $r_0 = \mathbb{E}[\log \frac{1}{n} \sum_i \exp(\xi_{i,1} - C_i)]$ i.e. e^{r_0} is the geometric mean of the spatial average of the $\exp(\xi_{i,1} - C_i)$. By Jensen's inequality, r_0 when $d \approx 1$ is greater than r_0 when $d \approx 0$. Hence, one can get the scenario where increasing the dispersal fraction d shifts a population from experiencing asymptotic extinction with positive probability to a population that persists with probability one. This corresponds to a positive density-dependence analog of a phenomena observed in models with negative density-dependent feedbacks [3, 14] and density-independent feedbacks [10, 15, 19, 25]. However, in these models, the long-term outcome never exhibits a mixture of extinction and persistence.

Example 3 (age-structured populations) revisited Consider the age-structured model with mate-limitation in Example 3 where there are $\ell \geq 2$ reproductive stages. If ξ_t are multivariate log-normals, then $\{0, \infty\}$ is accessible. Define

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ s_2 & 0 & 0 & \dots & 0 & 0 \\ 0 & s_3 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & s_n & 0 \end{pmatrix}.$$

As $0 < s_i < 1$ for all i, the dominant eigenvalue λ of B is strictly less than one. Thus,

$$\begin{aligned} r_0 = \lim_{t\to\infty} \frac{1}{t} \mathbb{E}[\log \| \prod_{s=1}^{t} \|A(0, \xi_s)\|] &= \lim_{t\to\infty} \frac{1}{t} \log \|B^t\| \\ &= \log \lambda < 0. \end{aligned}$$

As $r_0 < 0$, it follows that for all positive initial conditions there is a positive probability of asymptotic extinction (in contrast the spatial model which always has a positive probability of persistence and unbounded growth.)

To say something about persistence, assume that $\xi_{1,t}, \dots, \xi_{\ell,t}$ have the same log mean μ and non-degenerate log-covariance matrix Σ^2. Then r_∞ is an increasing function of μ with $\lim_{\mu\to\infty} r_\infty = \infty$ and $\lim_{\mu\to-\infty} r_\infty < 0$. Hence, there is a critical μ, call it μ^*, such that the population goes asymptotically extinct with probability one whenever $\mu < \mu^*$ and the population persists with positive probability whenever $\mu > \mu^*$.

5 Proofs

First, I prove Theorem 1. Assume $r_0 > 0$ and $X_0 = x_0 \gg 0$. As the entries of $A(x, \xi)$ are non-decreasing functions of x,

$$\begin{aligned}\liminf_{t\to\infty} \frac{1}{t} \log \|X_t\| &= \liminf_{t\to\infty} \frac{1}{t} \log \|\prod_{s=1}^{t} A(X_{s-1}, \xi_s)x_0\| \\ &\geq \liminf_{t\to\infty} \frac{1}{t} \log \|\prod_{s=1}^{t} A(0, \xi_s)x_0\| \\ &= r_0 > 0 \text{ with probability one.}\end{aligned}$$

In particular, $\lim_{t\to\infty} \|X_t\| = \infty$ with probability one as claimed.

Next, assume that $r_\infty < 0$. Given any $X_0 = x_0 \gg 0$, choose $c > 0$ such that $c\mathbf{1} \geq x_0$ and

$$\lim_{t\to\infty} \frac{1}{t} \log \|\prod_{s=1}^{t} A(c\mathbf{1}, \xi_s)\| \leq r_\infty/2 \text{ with probability one.}$$

Then

$$\begin{aligned}\limsup_{t\to\infty} \frac{1}{t} \log \|X_t\| &\leq \limsup_{t\to\infty} \frac{1}{t} \log \|\prod_{s=1}^{t} A(c\mathbf{1}, \xi_s)x_0\| \\ &\leq r_\infty/2 < 0 \text{ with probability one.}\end{aligned}$$

In particular, $\lim_{t\to\infty} X_t = 0$ with probability one as claimed.

Finally, assume that $r_\infty > 0$ and $r_0 < 0$. As the entries of A are non-decreasing in x, there exists $c > 0$ such that $A(c\mathbf{1}, \xi) \leq A(0, \xi)\exp(-r_0/2)$ for $\xi \in E$. Hence,

$$\limsup_{t\to\infty} \frac{1}{t} \log \|\prod_{s=1}^{t} A(c\mathbf{1}, \xi_s)\| \leq r_0/2 < 0 \text{ with probability one.} \tag{4}$$

Define the random variable

$$R = \sup_{t\geq 1} \| \prod_{s=1}^{t} A(c\mathbf{1}, \xi_s)\|.$$

Equation (4) implies that $R < \infty$ with probability one. For all $k > 0$, define the event $\mathcal{E}_k = \{R \leq k\}$. For $x_0 \leq c\mathbf{1}/k$ and $X_0 = x_0$, I claim that $X_t \leq c\mathbf{1}$ for all $t \geq 0$ on the event $\mathcal{E}_k$. I prove this claim by induction. $X_0 \leq c\mathbf{1}$ by assumption. Suppose that $X_s \leq c\mathbf{1}$ for $0 \leq s \leq t-1$. Then

$$\begin{aligned}\|X_t\| &= \| \prod_{s=1}^{t} A(X_{s-1}, \xi_s)x_0\| \\ &\leq \| \prod_{s=1}^{t} A(c\mathbf{1}, \xi_s)c\mathbf{1}/k\| \text{ by induction and monotonicity} \\ &\leq \| \prod_{s=1}^{t} A(c\mathbf{1}, \xi_s)\|c/k \leq Rc/k \text{ by the definition of } R \text{ and } x \\ &\leq c \text{ on the event } \mathcal{E}_k.\end{aligned}$$

This completes the proof of the claim that $X_t \leq c\mathbf{1}$ for all $t \geq 0$ on the event $\mathcal{E}_k$. It follows that on the event $\mathcal{E}_k$ and $X_0 = x \leq c\mathbf{1}/k$ that

$$\begin{aligned}\limsup_{t\to\infty} \frac{1}{t} \log \|X_t\| &\leq \limsup_{t\to\infty} \frac{1}{t} \log \| \prod_{s=1}^{t} A(c\mathbf{1}, \xi_s)\|c \\ &\leq r_0/2 < 0 \text{ almost surely.}\end{aligned}$$

In particular, $\lim_{t\to\infty} X_t = 0$ almost sure on the event $\mathcal{E}_k$. As the events $\mathcal{E}_k$ are increasing with k, $\lim_{k\to\infty} \mathbb{P}[\mathcal{E}_k] = \mathbb{P}[\cup_k \mathcal{E}_k] = \mathbb{P}[R < \infty] = 1$. Therefore, given $\varepsilon > 0$, there exists k such that $\mathbb{P}[\mathcal{E}_k] > 1 - \varepsilon$. For this k, $x_0 \leq c\mathbf{1}/k$ and $X_0 = x_0$,

$$\mathbb{P}[\lim_{t\to\infty} X_t = 0 | X_0 = x_0] \geq \mathbb{P}[\mathcal{E}_k] \geq 1 - \varepsilon.$$

To show convergence to ∞ with positive probability when $r_\infty > 0$, choose $c \geq c^*$ sufficiently large so that

$$\lim_{t\to\infty} \frac{1}{t} \log \| \prod_{s=1}^{t} A(c\mathbf{1}, \xi_s)\| \geq r_\infty/2 > 0 \text{ with probability one.}$$

By the Random Perron–Frobenius theorem [1, Theorem 3.1 and Remark (ii) on pg. 878],

$$\lim_{t\to\infty} \frac{1}{t} \log\left(e_i^T \prod_{s=1}^{t} A(c\mathbf{1}, \xi_s) e_j \right) \geq r_\infty/2 > 0 \text{ with probability one.} \quad (5)$$

for all elements e_i, e_j of the standard basis of $\mathbb{R}^n$ and where ^{T} denotes the transpose of a vector. Equation (5) implies that all of the entries of $\prod_{s=1}^t A(c\mathbf{1}, \xi_s)$ grow exponentially in time at rate greater than $r_\infty/2$ with probability one.

Define

$$R_\infty = \inf_{t\geq 1, 1\leq i\leq n} e_i^T \prod_{s=1}^{t} A(c\mathbf{1}, \xi_s) c\mathbf{1}.$$

By (5) and the primitivity assumption **A4**, $R_\infty > 0$ with probability one. Define the events

$$\mathcal{F}_k = \{R_\infty > 1/k\} \text{ for } k \geq 1.$$

Now, suppose that $X_0 = x_0 \geq c\mathbf{1}k$. I claim that $X_t \geq c\mathbf{1}$ for all $t \geq 0$ on the event $\mathcal{F}_k$. $X_0 \geq c\mathbf{1}$ by the choice of x_0. Assume that $X_s \geq c\mathbf{1}$ for $0 \leq s \leq t-1$. Then

$$\begin{aligned} X_t &= \prod_{s=1}^{t} A(X_{s-1}, \xi_s) x_0 \\ &\geq \prod_{s=1}^{t} A(c\mathbf{1}, \xi_s) x_0 \text{ by inductive hypothesis} \\ &\geq R_\infty c\mathbf{1}k \text{ by definition of } R_\infty \text{ and } x_0 \\ &\geq c\mathbf{1} \text{ on the event } \mathcal{F}_k. \end{aligned}$$

Equation (5) implies that on the event $\mathcal{F}_k$

$$\liminf_{t\to\infty} \frac{1}{t} \log \|X_t\| \geq r_0/2 \text{ almost surely.}$$

Hence, $\lim_{t\to\infty} \|X_t\| = \infty$ almost surely on the event $\mathcal{F}_k$. As $\mathcal{F}_k$ are an increasing set of events, $\mathbb{P}[R_\infty > 0] = \mathbb{P}[\cup_{t\geq 1}\mathcal{F}_k] = 1$. For any $\varepsilon > 0$ there is $k \geq 1$ such that $\mathbb{P}[\mathcal{F}_k] \geq 1 - \varepsilon$. Hence, for this k and $X_0 = x \geq ck\mathbf{1}$,

$$\mathbb{P}[\lim_{t\to\infty} \|X_t\| = \infty | X_0 = x] \geq 1 - \varepsilon.$$

This completes the proof of Theorem 1.

The proof of Theorem 2 follows from Theorem 1 and the following proposition.

Proposition 1 *Assume* $\{0, \infty\}$ *is accessible. Let* $c > 0$ *and* $\delta \in [0, 1)$ *be such that*

$$\mathbb{P}\left[\lim_{t\to\infty} X_t = 0 | X_0 = x\right] \geq 1 - \delta \textit{ whenever } x \leq \mathbf{1}/c$$

and

$$\mathbb{P}\left[\lim_{t\to\infty} X_t = \infty | X_0 = x\right] \geq 1 - \delta \textit{ whenever } x \geq c\mathbf{1}.$$

Then

$$\mathbb{P}\left[\lim_{t\to\infty} X_t = \infty \textit{ or } \lim_{t\to\infty} X_t = 0 | X_0 = x\right] = 1 \textit{ whenever } x \gg 0.$$

Proof Define the event

$$\mathcal{C} = \left\{\lim_{t\to\infty} X_t = \infty \text{ or } \lim_{t\to\infty} X_t = 0\right\}.$$

For any $x \in C$, define $\mathbb{P}_x[\mathcal{E}] = \mathbb{P}[\mathcal{E}|X_0 = x]$ (respectively, $\mathbb{E}_x[Z] = \mathbb{E}[Z|X_0 = x]$) for any event $\mathcal{E}$ (respectively, random variable Z) in the σ-algebra generated by $\{X_0 = x, X_1, X_2, \ldots\}$. Furthermore, define $I_{\mathcal{E}}$ to be random variable that equals 1 on the event $\mathcal{E}$ and 0 otherwise.

Define the stopping time

$$S = \inf\{t \geq 0 \;:\; X_t \geq c\mathbf{1} \text{ or } X_t \leq \mathbf{1}/c\}.$$

Since $\{0, \infty\}$ is accessible, there exists $\gamma > 0$ such that $\mathbb{P}_x[S < \infty] > \gamma$ for all $x \gg 0$. Let $I_{\{S<\infty\}}$ equal 1 if $S < \infty$ and 0 otherwise. The strong Markov property implies that for all $x \gg 0$

$$\begin{aligned}
\mathbb{P}_x\left[\mathcal{C}\right] &= \mathbb{E}_x\left[\mathbb{P}_{X_S}\left[\mathcal{C}\right] I_{\{S<\infty\}}\right] + \mathbb{E}_x\left[\mathbb{P}_{X_S}\left[\mathcal{C}\right] I_{\{S=\infty\}}\right] \\
&= \mathbb{E}_x\left[\mathbb{P}_{X_S}\left[\mathcal{C}\right] I_{\{S<\infty\}}\right] \\
&\geq (1-\delta)\gamma.
\end{aligned}$$

Let $\mathcal{F}_t$ be the σ-algebra generated by $\{X_1, \ldots, X_t\}$. The Lévy zero-one law implies that for all $x \gg 0$, $\lim_{t\to\infty} \mathbb{E}_x\left[I_{\mathcal{C}}|\mathcal{F}_t\right] = I_{\mathcal{C}}$ almost surely. On the other hand, the Markov property implies that $\mathbb{E}_x\left[I_{\mathcal{C}}|\mathcal{F}_t\right] = \mathbb{P}_{X_t}[\mathcal{C}] \geq (1-\delta)\gamma$ for all $x \gg 0$. Hence $\mathbb{P}_x[\mathcal{C}] = 1$ for all $x \gg 0$. □

References

1. Arnold, L., Gundlach, V.M., Demetrius, L.: Evolutionary formalism for products of positive random matrices. Ann. Appl. Probab. **4**, 859–901 (1994)
2. Assas, L., Dennis, B., Elaydi, S., Kwessi, E., Livadiotis, G.: A stochastic modified beverton-holt model with the allee effect. J. Differ. Equ. Appl. **22**(1), 37–54 (2016)
3. Benaïm, M., Schreiber, S.J.: Persistence of structured populations in random environments. Theor. Popul. Biol. **76**, 19–34 (2009)
4. Caswell, H.: Matrix Population Models. Sinauer, Sunderland, Massachuesetts (2001)
5. Courchamp, F., Berec, L., Gascoigne, J.: Allee effects in ecology and conservation. Environ. Conserv. **36**(1), 80–85 (2008)

6. Courchamp, F., Clutton-Brock, T., Grenfell, B.: Inverse density dependence and the Allee effect. Trends Ecol. Evol. **14**, 405–410 (1999)
7. Dennis, B.: Allee effects: Population growth, critical density, and the chance of extinction. Natural Resour. Model. **3**, 481–538 (1989)
8. Dennis, B.: Allee effects in stochastic populations. Oikos **96**, 389–401 (2002)
9. Dennis, B., Assas, L., Elaydi, S., Kwessi, E., Livadiotis, G.: Allee effects and resilience in stochastic populations. Theor. Ecol. 1–13 (2015)
10. Evans, S.N., Ralph, P., Schreiber, S.J., Sen, A.: Stochastic growth rates in spatio-temporal heterogeneous environments. J. Math. Biol. **66**, 423–476 (2013)
11. Gascoigne, J., Lipcius, R.N.: Periodic dynamics in a two-stage allee effect model are driven by tension between stage equilibria. Theor. Popul. Biol. **68**(4), 237–241 (2005)
12. Gascoigne, J.C., Lipcius, R.N.: Allee effects driven by predation. J. Appl. Ecol. **41**, 801–810 (2004)
13. Hardin, D.P., Takáč, P., Webb, G.F.: Asymptotic properties of a continuous-space discrete-time population model in a random environment. J. Math. Biol. **26**, 361–374 (1988)
14. Hening, A., Nguyen, D.H., Yin, G.: Stochastic population growth in spatially heterogeneous environments: The density-dependent case. arXiv:1605.02027 (arXiv preprint) (2016)
15. Jansen, V.A.A., Yoshimura, J.: Populations can persist in an environment consisting of sink habitats only. Proc. Natl. Acad. Sci. USA **95**, 3696–3698 (1998)
16. Kingman, J.F.C.: Subaddit. Ergod. Theory. Ann. Prob. **1**, 883–909 (1973)
17. Liebhold, A., Bascompte, J.: The Allee effect, stochastic dynamics and the eradication of alien species. Ecol. Lett. 133–140 (2003)
18. McCarthy, M.A.: The Allee effect, finding mates and theoretical models. Ecol. Model. **103**, 99–102 (1997)
19. Metz, J.A.J., de Jong, T.J., Klinkhamer, P.G.L.: What are the advantages of dispersing; a paper by Kuno extended. Oecologia **57**, 166–169 (1983)
20. Roth, G., Schreiber, S.J.: Persistence in fluctuating environments for interacting structured populations. J. Math. Biol. **68**, 1267–1317 (2014)
21. Roth, G., Schreiber, S.J.: Pushed to brink: Allee effects, environmental stochasticity, and extinction, special issue on allee effects for. J. Biol. Dyn. **8**, 187–205 (2014)
22. Scheuring, I.: Allee effect increases dynamical stability in populations. J. Theor. Biol. **199**, 407–414 (1999)
23. Schreiber, S.J.: Allee effects, chaotic transients, and unexpected extinctions. Theor. Popul. Biol. (2003)
24. Schreiber, S.J.: On Allee effects in structured populations. Proc. Amer. Math. Soc. **132**(10):3047–3053 (electronic) (2004)
25. Schreiber, S.J.: Interactive effects of temporal correlations, spatial heterogeneity, and dispersal on population persistence. Proc. Royal Soc. Biol. Sci. **277**, 1907–1914 (2010)

Replicator Equations as Limits of Evolutionary Games on Complete Graphs

Petr Stehlík

Abstract In this paper we discuss connections between the evolutionary games on graphs and replicator equations. On the traditional examples of social dilemma games we introduce the basic ideas of replicator dynamics and the mathematical concepts behind evolutionary games on graphs. We show that the stability regions of evolutionary games on complete graphs with the sequential and synchronous updating with deterministic imitation dynamics converge to the stability regions of replicator equations. Finally, we show that by a finer choice of a time scale and a stochastic imitation dynamic update rule not only the stability regions but also the trajectories of evolutionary games on graphs converge to those of replicator equations.

Keywords Evolutionary games on graphs · Game theory · Replicator equations · Convergence

MSC2010: 05C90 · 37N25 · 37N40 · 91A22

1 Introduction

Standard evolutionary game theory [13, 28] considers infinite well-mixed populations. It enabled to describe the dynamics of such homogeneous populations via nonlinear differential equations, known as replicator equations [12, 13]. Evolution of cooperation, as one of the most interesting biological processes, has been one of the main questions which has been studied in this setting [3, 20]. In recent years, numerous authors considered its counterpart with discrete time, finite, heterogenous and spatially structured populations–evolutionary games on graphs [19, 21]. Researchers from various fields have shown that the finite spatial structure could extend or shrink

P. Stehlík (✉)
Faculty of Applied Sciences, Department of Mathematics and NTIS,
University of West Bohemia, Univerzitni 8, 30614 Pilsen, Czech Republic
e-mail: pstehlik@kma.zcu.cz

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_4

the regions of cooperation or coexistence of cooperation and defection [1, 2, 11, 21–23, 27].

Whereas in the prevalent biological applications, the stochastic update rules and large graphs occur naturally (given the stochastic nature of evolution and a connection between the game utilities and the biological fitness), in economic applications and mathematically, evolutionary games on graphs are an interesting (and still not very well described) concept even for small networks with deterministic update rules. The goal of our recent papers [5, 6] was to establish rigorous structures for evolutionary games on graphs (as discrete dynamical systems) and study some basic mathematical questions, especially for the deterministic imitation dynamics. Among other things, we showed complete description of dynamics on complete graphs, indicated why this is already a complicated question even for k-regular graphs [5]. Moreover, we proved that coexistence equilibria could exist on arbitrary graph or for arbitrary game-theoretical parameters [6].

The goal of this short note is to develop into detail the connection of evolutionary games on graphs and replicator equations which was indicated in [5]. We show that the stability regions of evolutionary games on complete graphs with the sequential and synchronous updating with deterministic imitation dynamics converge to the stability regions of replicator equations. Moreover, we go one step further and construct evolutionary games on graphs with a stochastic imitation dynamics and show that, when coupled with a finer choice of a time scale, then not only the stability regions but also the trajectories of evolutionary games on graphs converge to those of replicator equations.

Throughout the paper we consider only connected undirected graphs $G = (V, E)$, where V denotes the set of vertices and E the set of its edges. Moreover we also consider the 1-neighbourhood of a given vertex $i \in V$ to be the set

$$N_1(i) = \{j \in V : (i, j) \in E\},$$

see [9] for more details about graph theory.

2 Social Dilemma Games

Game theoretical parameters form an essential part of evolutionary games on graphs. In each time step graph neighbours play a game and are rewarded by given and known utilities. In this section we discuss in detail game theoretical fundamentals (see e.g. [4, 18] for more detailed introduction on game theory). For our purposes, we take into account only two-player two-strategies symmetric games. The two considered strategies will be called cooperation (C) and defection (D) and the utilities are given by the utility matrix with parameters $a, b, c, d \in \mathbb{R}$

	C	D
C	a	b
D	c	d

This matrix should be read in the following way. Any player gets the utility a if both he and his partner cooperate, he gets b if he cooperates and his partner defects. On the other hand, if the player defects and his partner cooperates he gets c, if both players defect, he gets d. We focus on the so-called social dilemma games (see e.g. [19, 27]). In this class of games, we make the following assumptions on the parameters a, b, c, d:

- $a > d$, i.e., it is always better if both players cooperate than if they both defect.
- $c > b$, i.e., if only one player cooperates, each player prefers to be the defector.
- $a > b$ and $c > d$, i.e., no matter what strategy a player chooses, it is always better for him if his opponent cooperates.

Additionally, it is sometimes assumed for simpler analysis that the parameters a, c are positive, i.e., there is a positive reward for cooperation. These assumptions could be summarized by

$$\min\{a, c\} > \max\{b, d\} \tag{1}$$

and they imply four different scenarios

- Prisoner's dilemma game (PD), if $c > a > d > b$,
- Stag hunt game (SH), if $a > c > d > b$,
- Hawk and dove game (HD), if $c > a > b > d$,
- Harmony game (HG), if $a > c > b > d$.

Note that the names differ from time to time (Hawk and dove game is, especially in economic applications, called the snowdrift game [11] and the not so frequently studied Harmony game appears, e.g., under the Full cooperation game name [5]). In the same spirit, especially in the Prisoner's dilemma scenario, the parameters a, b, c, d are being replaced by R, S, T, P for their specific meaning, R—reward, S—the sucker's pay-off, T—temptation and P—punishment, see e.g. [19, 24].

The key concept in static games is the Nash equilibrium [4, 17, 18]. Intuitively, it is the combination of (mixed) strategies (σ_1^*, σ_2^*) such that neither player could improve his utility by a change of his strategy only, i.e. there are two conditions in the case of two-player two-strategies games. First, $u_1(\sigma_1^*, \sigma_2^*) \geq u_1(\sigma_1, \sigma_2^*)$ for all admissible σ_1 (the former player cannot improve his utility by a change of his strategy). Secondly, $u_2(\sigma_1^*, \sigma_2^*) \geq u_2(\sigma_1^*, \sigma_2)$ for all admissible σ_2 (the corresponding condition for the latter player).

In our case, allowing for mixed strategies, we can study $\sigma_i \in [0, 1]$ where $\sigma_i = 1$ corresponds to the player i playing the pure strategy C and $\sigma_i = 0$ corresponds to the player i playing the pure strategy D. From a perspective of player 1 for a fixed σ_2 we get

$$u_1(\sigma_1, \sigma_2) = a\sigma_1\sigma_2 + b\sigma_1(1 - \sigma_2) + c(1 - \sigma_1)\sigma_2 + d(1 - \sigma_1)(1 - \sigma_2). \tag{2}$$

Table 1 Nash equilibria of social dilemma games

		Scenario	Nash equilibria	Symmetric nash eq
$c > a > d > b$	PD	Prisoner's dilemma	$\mathbf{N}_{PD} = \{(0,0)\}$	$\mathbf{S}_{PD} = \{0\}$
$a > c > d > b$	SH	Stag hunt	$\mathbf{N}_{SH} = \{(0,0),(1,1),(\sigma^*,\sigma^*)\}$	$\mathbf{S}_{SH} = \{0,1,\sigma^*\}$
$c > a > b > d$	HD	Hawk and dove	$\mathbf{N}_{HD} = \{(0,1),(1,0),(\sigma^*,\sigma^*)\}$	$\mathbf{S}_{HD} = \{\sigma^*\}$
$a > c > b > d$	HG	Harmony game	$\mathbf{N}_{HG} = \{(1,1)\}$	$\mathbf{S}_{HG} = \{1\}$

The player 1 maximizes his utility by $\sigma_1 = 1$ in HG scenario, $\sigma_1 = 0$ in PD scenario and in the HD and SH scenarios the situation depends on the exact value of σ_2. Similar analysis from the perspective of the player 2 leads to the Nash equilibria structure listed in Table 1, where $\sigma^* = \frac{d-b}{a-c+d-b}$.

We denote by **N** the set of all Nash equilibria for the given social dilemma scenario and by **S** the set of all symmetric Nash equilibria (represented by a single value). Note that the sets differ, asymmetric Nash equilibria occur only in the Hawk & Dove scenario.

3 Replicator Equations

The game, as introduced in Sect. 2, is in its nature static and in the case of *HD* and *SH* scenarios provides multiple equilibria. The beautifully elegant concept of replicator dynamics [12, 13] allows not only to introduce dynamics to games but also to select equilibria which are stable (in the case of games with two players and two strategies they correspond to the so-called evolutionary stable strategies [13, 28], which in turn represent one of many Nash equilibrium refinements [17]).

In our case with two strategies C and D, let us assume that in the population of $n = n_C + n_D$ individuals there are n_C cooperators and n_D defectors. Proportionally we can denote

$$x = x_C = \frac{n_C}{n}, \quad x_D = \frac{n_D}{n},$$

i.e., $x_C + x_D = 1$. For each strategy we denote by $u(C, x)$ and $u(D, x)$ the utilities which are obtained by the players playing C or D in the population with the fraction x of cooperators. If we consider the modification of the Malthusian growth with a growth factor r (combining birth rate minus the growth rate) adjusted by the utilities of each strategy we can introduce the dynamics for the number of cooperating and defecting individuals (we also simplify the situation by taking into account the continuous time model)

$$n_C'(t) = (r + u(C,x))n_C(t), \quad n_D'(t) = (r + u(D,x))n_D(t).$$

Easy calculation yields that for the number of individuals in the aggregate population we get the following modification of the Malthusian growth model

$$n'(t) = (r + \overline{u}(x))n(t),$$

where $\overline{u}(x) = xu(C, x) + (1 - x)u(D, x)$ denotes the average utility of the population. Being interested in the change of strategy ratios we observe that:

$$x'(t) = \frac{n'_C(t) - x(t)n'(t)}{n(t)} = \frac{(r + u(C, x))n_C(t) - x(t)(r + \overline{u}(x))n(t)}{n(t)}.$$

Employing $n_C/n = x$ we get the replicator equation for the proportion of cooperators

$$x'(t) = (u(C, x(t)) - \overline{u}(x(t)))x(t). \tag{3}$$

If we pass to infinite population, i.e. $n \to \infty$, we get

$$u(C, x) = ax + b(1 - x), \quad u(D, x) = cx + d(1 - x). \tag{4}$$

Similarly, one can obtain the average utility of the population:

$$\overline{u}(x) = xu(C, x) + (1 - x)u(D, x) = x(ax + b(1 - x)) + (1 - x)(cx + d(1 - x)). \tag{5}$$

Consequently, the replicator equation for social dilemma games becomes (starting from (3))

$$\begin{aligned} x'(t) = (ax(t) + b(1 - x(t)) - x(t)(ax(t) + b(1 - x(t))) \\ + (1 - x(t))(cx(t) + d(1 - x(t)))x(t). \end{aligned}$$

Factoring out, we can simplify it and rewrite it as

$$x'(t) = x(t)(1 - x(t)) \left(x(t)(a - c + d - b) + b - d\right). \tag{6}$$

We immediately observe that there are three stationary solutions of this ordinary differential equation $x_1^* = 0$, $x_2^* = 1$ and $x_3^* = \frac{d-b}{a-c+d-b}$.

Let us define two additional sets. Let $\mathbf{F}$ be the set of all fixed points of the replicator equation in [0, 1] and let $\mathbf{A}$ denote the set of all asymptotically stable fixed points of the replicator equation in [0, 1]. Then the analysis of (6) yields that we have the structure of fixed points for replicator equations of social dilemma games summed up in Table 2 (moreover, the trajectories for various initial fraction of cooperators are illustrated in Fig. 2).

Comparing Tables 1 and 2 we immediately observe the following relationship:

$$\mathbf{A} \subseteq \mathbf{S} \subseteq \mathbf{F} \tag{7}$$

Table 2 Fixed points of replicator dynamics of social dilemma games

		Scenario	Fixed points	Asymptotically stable fixed points
$c > a > d > b$	PD	Prisoner's dilemma	$\mathbf{F}_{PD} = \{0, 1\}$	$\mathbf{A}_{PD} = \{0\}$
$a > c > d > b$	SH	Stag hunt	$\mathbf{F}_{SH} = \{0, 1, \sigma^*\}$	$\mathbf{A}_{SH} = \{0, 1\}$
$c > a > b > d$	HD	Hawk and dove	$\mathbf{F}_{HD} = \{0, 1, \sigma^*\}$	$\mathbf{A}_{HD} = \{\sigma^*\}$
$a > c > b > d$	HG	Harmony game	$\mathbf{F}_{HG} = \{0, 1\}$	$\mathbf{A}_{HG} = \{1\}$

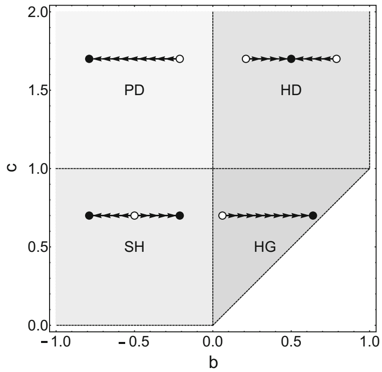


Fig. 1 Set of admissible parameters and four game-theoretic scenarios, $a = 1$ and $d = 0$

Remark 1 In fact, (7) (which we derived only for social dilemma games) is a special case of a more general result [13], which furthermore includes the set $\mathbf{E}$ of the so-called evolutionary stable strategies (ESS), which is one of equilibrium refinements defined by John Maynard Smith [28]. For general symmetric game we then have

$$\mathbf{E} \subseteq \mathbf{A} \subseteq \mathbf{S} \subseteq \mathbf{F},$$

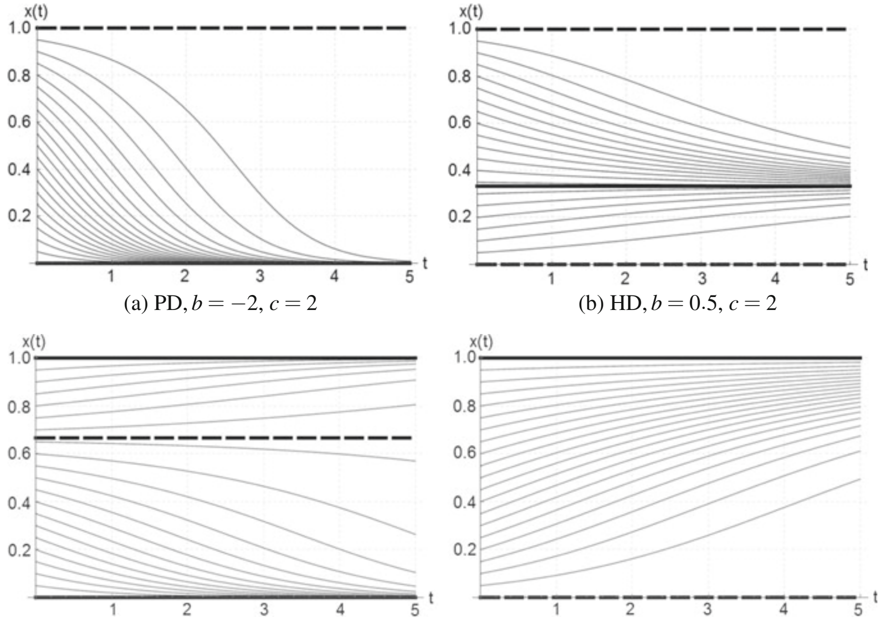
and $\mathbf{E} = \mathbf{A}$ in the case of two strategy games (for example, our social dilemma games).

Figures 1 and 2 indicate that for the sake of easier visualisation, we can reduce the dimension of parameter space from 4 to 2 by normalizing parameters a, b, c, d so that $\tilde{a} = 1$ and $\tilde{d} = 0$ by the following map

$$\tilde{x} = \frac{x - d}{a - d}, \quad x = a, b, c, d.$$

Similarly, given normalized values of parameters $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$, we could find for arbitrary a and d such that $a > d$ non-normalized values of parameters by

$$x = d + (a - d)\tilde{x}, \quad \tilde{x} = \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}.$$



(a) PD, $b = -2$, $c = 2$ (b) HD, $b = 0.5$, $c = 2$

(c) SH, $b = -1$, $c = 0.5$ (d) HG, $b = 0.7$, $c = 0.8$

Fig. 2 Replicator equation trajectories for social dillemma scenarios, $a = 1$ and $d = 0$. Light gray lines show trajectories for solutions with initial conditions $x(0) = i/20$, with $i = 0, 1, \ldots, 20$. Black solid lines represent asymptotically stable stationary solutions and black dashed lines correspond to unstable stationary solutions

This allows us to simplify conditions or plot regions corresponding to various scenarios, cf. Fig. 1 where the four social dilemma scenarios are depicted. For the sake of brevity, we will simplify our computations in the rest of the paper by assuming that $a = 1$ and $d = 0$.

4 Evolutionary Games on Graphs

There are two main disadvantages of replicator dynamics. First, the population is homogeneous and well-mixed, i.e. each player interacts with all other players. Secondly, the population size must be taken as infinite. These drawbacks have been fixed by the concept of evolutionary games on graphs [21]. The avalanche of ensuing papers studied these cellular automata[1] from numerous complex angles and showed

[1] In some cases one should rather speak about agent-based models, since the graph structure varies and agents are allowed to interact in a very complex fashion [24, 29], see also [10].

by more or less analytical methods that the finite and heterogeneous graph structure can either promote or inhibit cooperation.

In our previous papers [5, 6] we attempted to describe the simplest evolutionary games on graphs using discrete dynamical systems[2] and we showed that they could be defined as discrete dynamical systems in the following way.

Each vertex could either attain value 0 (defection) or 1 (cooperation). Evolutionary games on graphs consist of a non-directed graph G, game theoretical parameters p (see Table 1, an utility function u which assign to each vertex utility based on the graph structure, game theoretical parameters and the distribution of cooperation/defection. Consequently, there is an update order $\mathscr{T}$ assigning to each time instance t a set of vertices which could be update at t. The exact mechanism how the vertex is being updated is given by the an update rule φ.

Definition 1 An *evolutionary game on a graph* $\mathscr{E}$ is a quintuple $(G, p, u, \mathscr{T}, \varphi)$, where

(a) $G = (V, E)$ is a connected graph,
(b) $p = (a, b, c, d)$ are game-theoretical (social dilemma) parameters,
(c) $u : \{0, 1\}^V \to \mathbb{R}^V$ is a utility function,
(d) $\mathscr{T} : \mathbb{T} \to 2^V$ is an update order on an infinite discrete time scale $\mathbb{T} \subset [0, \infty)$,
(e) $\varphi : \mathbb{T}^2_{\geq} \times \{0, 1\}^V \to \{0, 1\}^V$ is a (generally nonautonomous) dynamical system (the so-called update rule).[3]

Remark 2 In contrast to our original definition [6, Definition 2.1] we allow for more general time $\mathbb{T}$ instead of $\mathbb{N}_0$. The only motivation for this slight modification is the study of the convergence to the replicator equations, for which the vanishing discrete time step is necessary. If not said otherwise (when dealing with convergence of trajectories) we still use $\mathbb{T} = \mathbb{N}_0$.

Remark 3 There could be several choices of utility functions. Since we study evolutionary games on complete graphs, the choice of a utility function does not play a key role (as is the case for irregular graphs [5, Sect. 8]). Therefore, we only consider the mean utility function given by

---

[2] Another motivation for our study had been the small size of cooperation macroeconomic networks, see e.g. [8, 14, 15]. Note that this is in contrast to the focus on large, often scale-free, networks in the physical and biological applications [27].

[3] We denote by $\mathbb{T}^2_{\geq}$ the set

$$\mathbb{T}^2_{\geq} := \{(t, s) \in \mathbb{T}^2 : t \geq s\}.$$

$$u_i^M(x) = \frac{1}{|N_1(i)|}\left(a\sum_{j\in N_1(i)} x_i x_j + b\sum_{j\in N_1(i)} x_i(1-x_j) + c\sum_{j\in N_1(i)}(1-x_i)x_j \right.$$
$$\left. + d\sum_{j\in N_1(i)}(1-x_i)(1-x_j),\right)$$

where $x \in \{0,1\}^V$ and its components are $x_i \in \{0,1\}$. Since we only consider the mean utility function, we often omit the upper index M and use u instead of u^M in the following.

Remark 4 There are two major deterministic examples of update orders $\mathscr{T} : \mathbb{T} \to 2^V$. Namely the synchronous update order

$$\mathscr{T}_{SYN}(t) = V, \text{ for each } t \in \mathbb{N}_0.$$

which mathematically leads to simpler autonomous evolutionary games $\mathscr{E}$ and the sequential update order, in which vertices can be ordered and numbered $1, \dots n$ and

$$\mathscr{T}_{SEQ}(t) = t(\text{mod } n) + 1, \text{ for each } t \in \mathbb{N}_0.$$

There are numerous other update orders, in the proof of convergence we consider an example of an update order which is (i) stochastic and (ii) defined on a finer time scale (with a smaller discretization step).

Note that both synchronous and sequential update orders are non-omitting and periodic (cf. [5, Definition 14]).

Remark 5 The major example of a deterministic update rule φ is the deterministic imitation dynamics φ^{ID} which we use in this paper. In this update rule, each vertex (if being updated) follows the strategy in its neighbourhood which, at a given time t, yields the highest utility. In other words, the update rule φ^{ID} is defined via its components $\varphi_i^{ID} := \text{proj}_i \circ \varphi : (\mathbb{N}_0)^2_{\geq} \times \{0,1\}^V \to \{0,1\}$ by

$$\varphi_i^{ID}(t+1,t,x) = \begin{cases} x_{\max} & \text{if } i \in \mathscr{T}(t), |A_i(x)| = 1 \text{ and } A_i(x) = \{x_{\max}\}, \\ x_i & \text{otherwise,} \end{cases} \tag{8}$$

where $A_i(x)$ is the set of strategies in the neighbourhood of x which yield the highest utility and is given by

$$A_i(x) = \left\{x_k : k \in \text{argmax}\left\{u_j(x) : j \in N_{\leq 1}(i)\right\}\right\}, \tag{9}$$

where $N_{\leq 1}(i) := \{j \in V : \text{dist}(i,j) \leq 1\}$ denotes the neighbourhood of i. The cardinality of $A_i(x)$ is used to ensure that all vertices with the highest utility have the same state. If that is not the case, the vertex preserves its current state (in order to keep the dynamics deterministic).

There are numerous other update rules, especially death-birth, birth-death and best response update rules, see [19, 24] or [5, Remark 4]. For the sake of convergence, we define later another stochastic update rule (20) which is a modification of imitation update rule with a probabilistic ingredient based on a difference between the utilities of more and less successful strategies.

Our main results which followed from this rigorous approach included the following universal existence theorems for coexistence equilibria[4] in the case of deterministic imitation dynamics φ^{ID}. First, we showed that for each social-dilemma parameters there is a graph so that there is a coexistence equilibrium of the corresponding evolutionary game (the proof is actually constructive [6, Theorem 4.1]).

Theorem 1 *For each $p = (a, b, c, d)$ and any update order $\mathscr{T}$ there exists a connected graph G such that the evolutionary game on a graph $(G, p, u^M, \mathscr{T}, \varphi^{ID})$ has a coexistence equilibrium.*

In a similar way, we showed that for each graph there are social-dilemma game theoretical parameters so that there is a coexistence equilibrium of the corresponding evolutionary game (again, the reader is invited to check the constructive proof [6, Theorem 4.2]).

Theorem 2 *For each connected graph G and any update order $\mathscr{T}$ there exists a parameter vector $p = (a, b, c, d)$ such that the evolutionary game on a graph $(G, p, u^M, \mathscr{T}, \varphi^{ID})$ has a coexistence equilibrium.*

In order to consider stability of configurations of evolutionary games on graphs, we also introduced the concept of attractivity. Roughly speaking, a state $x \in \{0, 1\}^V$ (or a set of such states) is attractive if a perturbation (represented by the change of exactly one x_i) eventually returns to the given state (or the set of states). The attractor $A \subset \{0, 1\}^n$ is called nontrivial if $A \nsubseteq \{(0, 0, \dots, 0), (1, 1, \dots, 1)\}$. For more details, see [5, Definition 5].

5 Evolutionary Games on Complete Graphs—Convergence of Stability Regions

In this section we consider evolutionary games on complete graphs K_n with synchronous and sequential updating and show that as $n \to \infty$ the stability regions of their stationary solutions coincide with those of replicator equations.

Let $m \in \{0, 1, \dots, n\}$ denote the number of cooperators. Then the mean utility of cooperators and defectors in the complete graph K_n could be evaluated as (if there

[4] We say that a state $x \in \{0, 1\}^V$ is a *coexistence equilibrium* (coexistence fixed point) of the evolutionary game on a graph $(G, p, u, \mathscr{T}, \varphi)$ if (a) it is a fixed point, i.e., $\varphi(t + 1, t, x) = x$ for all $t \in \mathbb{T}$, and (b) it is a coexistence state, i.e., $0 < \sum_{i \in V} x_i < |V|$, see [6] for more details.

are m cooperators then a cooperating vertex has $m-1$ cooperating neighbours and $n-m$ defecting neighbours)

$$u_n(C,m) = \frac{a(m-1)+b(n-m)}{n-1}, \quad u_n(D,m) = \frac{cm+d(n-m-1)}{n-1}, \tag{10}$$

and the average utility is given by

$$\begin{aligned}\overline{u}_n(m) &= \frac{mu_n(C,m)+(n-m)u_n(D,m)}{n}\\ &= \frac{m[a(m-1)+b(n-m)]+(n-m)[cm+d(n-m-1)]}{n(n-1)}.\end{aligned} \tag{11}$$

Applying few simple algebraic operations we can show that for each n we have

$$u_n(C,m^*) = \overline{u}_n(m^*) = u_n(D,m^*),$$

where m^* is a critical constant given by

$$m^* := \frac{n(b-d)-(a-d)}{(c-a)+(b-d)}, \tag{12}$$

which is not necessary integer.

Comparing the values of $u_n(C,m)$ and $u_n(D,m)$ we get immediately the complete characterization of the evolutionary game on complete graphs with synchronous updating

$$\mathcal{E}_{SYN} = \{K_n, p, u^M, \mathcal{T}_{SYN}, \varphi^{ID}\},$$

where $\mathcal{T}_{SYN}(t) = V$ for all $t \in \mathbb{N}_0$.

Theorem 3 *The synchronous evolutionary game* $\mathcal{E}_{SYN} = \{K_n, p, u^M, \mathcal{T}_{SYN}, \varphi^{ID}\}$ *has got at most three homogeneous (constant) solutions and*

(a) the state $(0,0,\ldots,0)$ *is a stationary solution which is attractive if and only if*

$$b < d, \text{ or } n < 1 + \frac{c-d}{b-d}. \tag{13}$$

(b) the state $(1,1,\ldots,1)$ *is a stationary solution which is attractive if and only if*

$$a > c, \text{ and } n > 1 + \frac{a-b}{a-c}. \tag{14}$$

(c) if $m^* \in \mathbb{N}$ *(see (12)) then there exists the set of stationary solutions* $M^* = \{x \in \{0,1\}^V : \sum_{i\in V} x_i = m^*\}$. *This set is never attractive.*

Proof The proof follows from [5, Theorems 9, 10].

Remark 6 Rather than repeating the individual steps of the proof (which can be found in [5, Theorems 9, 10]), we emphasize the differences between (13) and (14). It could seem counter-intuitive that one cannot transform one into another by a simple substitution of d, c, b, a for a, b, c, d. This could be easily seen from the assumptions on parameters (1). Indeed, the necessary and sufficient condition for $(0, 0, \dots, 0)$ being an attractive stationary solution is:

$$u_n(D, 1) - u_n(C, 1) > 0$$

which is equivalent with

$$(n-1)(d-b) + (c-d) > 0. \tag{15}$$

Similarly, the necessary and sufficient condition for $(1, 1, \dots, 1)$ being an attractive stationary solution is:

$$u_n(C, n-1) - u_n(D, n-1) > 0$$

which is equivalent with

$$(n-1)(a-c) + (b-a) > 0. \tag{16}$$

Note that (15) and (16) are mutually obtainable by the simple substitution of d, c, b, a for a, b, c, d. However, since (1) hold we observe that $c - d > 0$ but $b - a < 0$. Consequently, when we want to rewrite (15) and (16) in terms of dependence on n as in Theorem 3, we arrive to the conditions (13) and (14). In other words, these conditions imply that with the increasing n the basin of attraction of $(0, 0, \dots, 0)$ shrinks whereas the basin of attraction of $(1, 1, \dots, 1)$ expands.

One could repeat the reasoning for the evolutionary game on K_n with the sequential update order

$$\mathscr{E}_{SEQ} = \{K_n, p, u^M, \mathscr{T}_{SEQ}, \varphi^{ID}\},$$

where $T_{SEQ}(t) = \{i\}$ with $i = (t \mod n) + 1$ for all $t \in \mathbb{N}_0$.

Theorem 4 *The sequential evolutionary game* $\mathscr{E}_{SEQ} = \{K_n, p, u^M, \mathscr{T}_{SEQ}, \varphi^{ID}\}$ *has got at most three homogeneous (constant) solutions and*

(a) the state $(0, 0, \dots, 0)$ *is a stationary solution which is attractive if and only if*

$$b < d, \text{ or } n < 1 + \frac{c-d}{b-d}.$$

(b) the state $(1, 1, \dots, 1)$ *is a stationary solution which is attractive if and only if*

$$a > c, \text{ and } n > 1 + \frac{a-b}{a-c}.$$

(c) if $m^* \in \mathbb{N}$ *(see* (12)*) then there exists the set of stationary solutions* $M^* = \{x \in \{0,1\}^V : \sum_{i \in V} x_i = m^*\}$. *This set is attractive if and only if* $m^* \in \{2, 3, \dots, n-2\}$ *and* $c - a + b - d > 0$.

Proof The proof follows from [5, Theorems 26, 27].

Note that in the case of the sequential evolutionary game $\mathscr{E}_{SEQ}$ if $m^* \notin \mathbb{N}$ and $m^* \in (1, n-1)$ then there arise periodic solutions with the values of cooperators oscillating between $\lfloor m^* \rfloor$ and $\lceil m^* \rceil$. Consequently, we can claim that the set

$$M^* = \left\{ x \in \{0,1\}^V : \sum_{i \in V} x_i = \lfloor m^* \rfloor \text{ or } \sum_{i \in V} x_i = \lceil m^* \rceil \right\}$$

is also attractive if $m^* \in [2, n-2]$, cf. [5, Example 29]. Note that this set contains either periodic solutions (if m^* is not a natural number) or stationary solutions (if m^* is a natural number).

Theorems 3 and 4 immediately yield the following statement which could be interpreted as a convergence of regions of stability to those of replicator equations.

Theorem 5 *The following statements hold:*

- *Let* (a, b, c, d) *satisfy* $c > a > d > b$ *(PD region). Then* $(0, 0, \dots, 0)$ *is a unique non-trivial attractor of* $\mathscr{E}_{SEQ}$ *and* $\mathscr{E}_{SYN}$ *for all* $n \in \mathbb{N}$.
- *Let* (a, b, c, d) *satisfy* $a > c > d > b$ *(SH region). Then there exists* $n_0 \in \mathbb{N}$ *such that* $(0, 0, \dots, 0)$ *and* $(1, 1, \dots, 1)$ *are attractors of* $\mathscr{E}_{SEQ}$ *and* $\mathscr{E}_{SYN}$ *for all* $n > n_0$.
- *Let* (a, b, c, d) *satisfy* $c > a > b > d$ *(HD region). Then there exists* $n_0 \in \mathbb{N}$ *such that the set*

$$M^* = \left\{ x \in \{0,1\}^V : \sum_{i \in V} x_i = \lfloor m^* \rfloor \textit{ or } \sum_{i \in V} x_i = \lceil m^* \rceil \right\}$$

is an attractor of $\mathscr{E}_{SEQ}$. *Moreover, the stationary solution* $(0, 0, \dots, 0)$ *is not attractive for sufficiently large n and* $(1, 1, \dots, 1)$ *is never attractive.*

- *Let* (a, b, c, d) *satisfy* $a > c > b > d$ *(HG region). Then there exists* $n_0 \in \mathbb{N}$ *such that* $(1, 1, \dots, 1)$ *is a unique non-trivial attractor of* $\mathscr{E}_{SEQ}$ *and* $\mathscr{E}_{SYN}$ *for all* $n > n_0$.

The convergence of attractivity regions is illustrated in Fig. 3.

Remark 7 Note that necessary and sufficient attractivity conditions which appear in Theorems 3 and 4 could be easily modified for general k-regular graph. However, they are only sufficient conditions. Counterexamples to their necessity could be

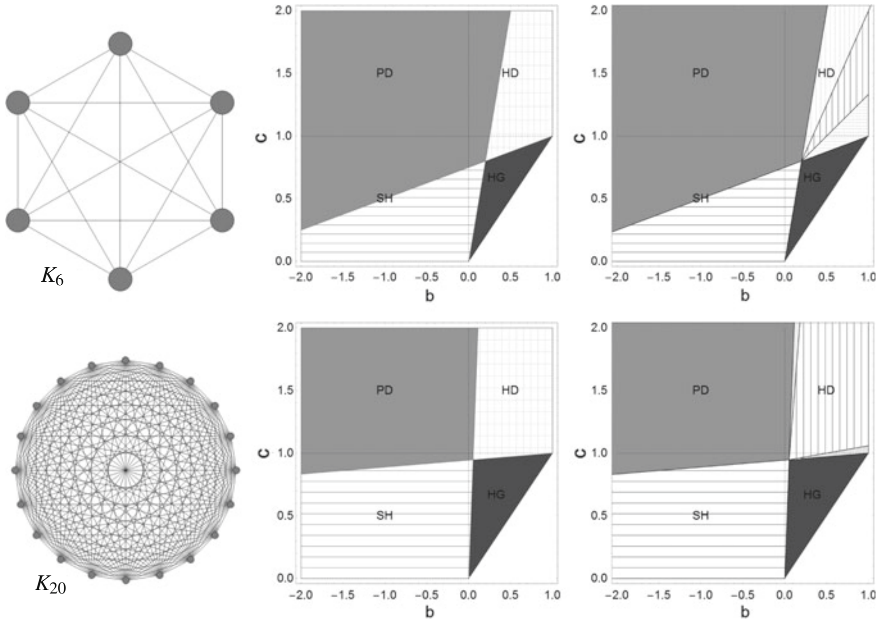


Fig. 3 Attractivity regions for various complete graphs, $a = 1$ and $d = 0$. The central graphs correspond to the synchronous updating, the rightmost ones to the sequential updating. In the light gray region, only $(0, 0, \ldots 0)$ is attractive, in the dark gray region only $(1, 1, \ldots 1)$ is attractive. In the horizontally hatched region, both $(0, 0, \ldots 0)$ and $(1, 1, \ldots 1)$ are attractive. In the *vertically hatched region* the coexistence attractors or attractive periodic solutions occur. In the *dotted regions*, there are no attractors. Compare those figures with the stability regions of the replicator equations, Fig. 1

constructed, see [5, Example 13]. So far, it is not clear, whether there are necessary and sufficient conditions for general k-regular graphs. Only special cases, e.g., cycles (i.e., 2-regular graphs) have been studied and fully described [22].

6 Evolutionary Games on Complete Graphs—Convergence of Trajectories

Apparently, Theorem 5 and Fig. 3 indicate that in the case of synchronous and sequential updating the dynamic properties of equilibria of evolutionary games on graphs converge to those of replicator equations. However, the trajectories differ significantly. For example, in the case of synchronous updating, the equilibria are reached in one step. The goal of this section is to discuss another update rule for which not only the dynamic properties coincide but the solution trajectories converge to the solutions of the replicator equations.

We modify the evolutionary game with the synchronous updating in two aspects. Firstly, since we want the convergence to functions defined on continuous sets, trajectories must be considered on a finer time scale, i.e., we consider time scale $\mathbb{T} = h\mathbb{N}_0 = \{0, h, 2h, \ldots\}$ with a time step h instead of $\mathbb{T} = \mathbb{N}_0$. Secondly, we modify the imitation dynamics update rule in the following way. As before, each player preserves his strategy if it yields better utility than the utilities of his neighbours. However, there will be a probabilistic rule which would describe whether the player with worse utility would switch or not. Roughly speaking, the more the difference in the utilities the higher probability of switching. This second modification ensures that we get different trajectories for various parameters in each social-dilemma region.

Since the utilities are bounded, we can make the difference in utilities arbitrarily small in the following way. Let us choose $k \in \mathbb{N}_0$ so that for each $x \in \{0, 1/n, 2/n, \ldots, 1\}$ and almost all $n \in \mathbb{N}$ we have

$$\frac{1}{n^k}(u_n(\chi, nx) - \overline{u}_n(nx)) \geq -1, \quad \chi \in \{C, D\}. \tag{17}$$

Note that this is straightforwardly possible because the quantities $(u_n(C, nx) - \overline{u}_n(nx))$, $(u_n(D, nx) - \overline{u}_n(nx))$ are bounded, see (10) and (11).

For such a k we define the sufficiently fine time scale

$$\mathbb{T} = h\mathbb{N}_0 = \{0, h, 2h, \ldots\}$$

where $h = \frac{1}{n^{k+1}}$.

Let us modify the sequential update order and consider the scaled sequential random update order $\mathscr{T}_{sc}(t) : h\mathbb{N}_0 \to \{0, 1\}^n$. $\mathscr{T}_{sc}(t)$ contains, for each $t \in h\mathbb{N}_0$ exactly one randomly selected vertex $i \in V$. Note that such an update order is no longer deterministic and it is almost surely non-omitting.

Let us define the scaled imitation dynamic rule φ_{sc}. Let $A_i(x)$ be the set of strategies in the neighbourhood of x_i which yield the highest utility at a given time t

$$A_i(x) = \left\{x_k : k \in \operatorname{argmax}\left\{u_j(x) : j \in n_{\leq 1}(i)\right\}\right\}. \tag{18}$$

Furthermore, let us denote

$$y_i^{\max} := \begin{cases} y & \text{if } |A_i(x)| = 1, \text{ and } A_i(x) = \{y\}, \\ x_i & \text{otherwise.} \end{cases}$$

The cardinality of $A_i(x)$ is used to ensure that all vertices with the highest utility have the same state. If that is not the case, the vertex preserves its current state. If that is the case, the vertex switches to the most successful strategy with probability

$$p^* = \begin{cases} -\frac{1}{n^k}(u_n(\chi, nx) - \overline{u}_n(nx)) & \text{if } u_n(\chi, nx) < \overline{u}_n(nx), \chi \in \{C, D\}, \\ 0 & \text{otherwise,} \end{cases} \tag{19}$$

(see (17) to check that this indeed defines a probability). Consequently, we can define the dynamical system

$$(\varphi_{sc})_i(t+h,t,x) = \begin{cases} y_i^{\max} & \text{with probability } p^* \text{ if } i \in \mathscr{T}_{sc}(t) \\ x_i & \text{otherwise.} \end{cases} \tag{20}$$

Now we are ready to consider the scaled evolutionary game $\mathscr{E}_{sc}$ on K_n

$$\mathscr{E}_{sc} = (K_n, p, u^M, \mathscr{T}_{sc}, \varphi_{sc}).$$

Let us analyze the mean trajectory (note that the trajectories are random since both $\mathscr{T}_{sc}$ and φ_{sc} are stochastic) of the scaled evolutionary game on K_n. For a given ratio of $x_n(t)$ cooperators we look for the ratio of cooperators $x_n(t+h)$ at time $t+h$.

Lemma 1 *The mean trajectory of the scaled evolutionary game* $\mathscr{E}_{sc} = (K_n, p, u^M, \mathscr{T}_{sc}, \varphi_{sc})$ *satisfies the difference equation*

$$x_n(t+h) = x_n(t) + hx_n(t)\left(u_n(C, nx_n(t)) - \overline{u}_n(nx_n(t))\right). \tag{21}$$

Proof Let us distinguish two cases, according to utility of cooperators and defectors:

- If $u_n(C, nx_n(t)) < \overline{u}_n(nx_n(t))$ (i.e., if defectors are doing better) we observe that exactly one player will be selected for possible update (ratio $1/n$), he will be cooperator with probability $x_n(t)$ and he will switch with probability $-\frac{1}{n^k}(u_n(C,nx) - \overline{u}_n(nx))$, see (19). Consequently,

$$\begin{aligned} x_n(t+h) &= x_n(t) + \frac{1}{n}x_n(t)\frac{1}{n^k}\left(u_n(C, nx_n(t)) - \overline{u}_n(nx_n(t))\right) \\ &= x_n(t) + hx_n(t)\left(u_n(C, nx_n(t)) - \overline{u}_n(nx_n(t))\right) \end{aligned}$$

- If $u_n(C, nx_n(t)) > \overline{u}_n(nx_n(t))$ (i.e., if cooperators are doing better) we observe that the cooperator keeps his strategy. Again, exactly one player is selected for possible update, he is a defector with probability $(1 - x_n(t))$ and he will switch with probability $-\frac{1}{n^k}(u_n(D,nx) - \overline{u}_n(nx))$, see (17). Since

$$\overline{u}_n(nx_n(t)) = x_n(t)u_n(C) + (1 - x_n(t))u_n(D),$$

we have

$$x_n(t)\left(u_n(C, nx_n(t)) - \overline{u}_n(nx_n(t))\right) = -(1 - x_n(t))\left(u_n(D, nx_n(t)) - \overline{u}_n(nx_n(t))\right),$$

and we observe that also in this case

$$\begin{aligned} x_n(t+h) &= x_n(t) + \frac{1}{n} x_n(t) \frac{1}{n^k} \left(u_n(C, n x_n(t)) - \overline{u}_n(n x_n(t)) \right) \\ &= x_n(t) - \frac{1}{n}(1 - x_n(t)) \frac{1}{n^k} (u_n(D, n x_n(t)) - \overline{u}_n(n x_n(t))) \\ &= x_n(t) + h x_n(t) \left(u_n(C, n x_n(t)) - \overline{u}_n(n x_n(t)) \right) \end{aligned}$$

We are ready to prove our main convergence theorem. Dealing with the convergence we extend $x_n(t)$, which is only defined on the discrete time scale $\mathbb{T}$, to a function defined on a continuous domain by defining

$$x_n^*(t) = x_n(t^*), \quad t^* := \max_{s \le t, s \in \mathbb{T}} s. \tag{22}$$

We use the following result as a key tool to prove the convergence of the scaled evolutionary game on complete graph to the replicator equations. This theorem can be seen as a variant of Euler method in which we not only have the convergence of discrete time scales but at the same time the convergence of nonlinear right hand sides. This auxiliary result follows from more general results for generalized ordinary differential equations [7, Theorem 2.5] and [26, Theorem 12].

Theorem 6 *Let*

- $f_n : [0, 1] \to \mathbb{R}$ *for each* $n \in \mathbb{N}_0$,
- *there exist* M *such that for each* $n \in \mathbb{N}_0$ *and for each* $x \in [0, 1]$ *we have*

$$|f_n(x)| < M,$$

- *there exist* L *such that for each* $n \in \mathbb{N}_0$ *and for each* $x, y \in [0, 1]$ *we have*

$$|f_n(x) - f_n(y)| \le L|x - y|,$$

- $f_n(x)$ *converge uniformly to* $f_0(x)$ *on* $[0, 1]$,
- $\mathbb{T}_n = h_n \mathbb{N}_0$ *be a sequence of discrete time scales such that* $T \in \mathbb{T}_n$ *such that* $h_n \to 0$,
- *for each* $n \in \mathbb{N}$ *let* $x_n(t)$ *be a solution of the recurrence equation*

$$\begin{cases} x_n(t+h) = x_n(t) + h_n f_n(x_n(t)), & t \in \mathbb{T}_n \cap [0, T], \\ x_n(0) = x_0^n \in [0, 1]. \end{cases}$$

- *there exist limits* $x_0 = \lim_{n\to\infty} x_0^n$ *and* $x_0(t) = \lim_{n\to\infty} x_n^*(t)$ *for all* $t \in [0, T]$.

Then the sequence $x_n^*(t)$ contains a subsequence $x_{n_k}^*(t)$ which converges uniformly to $x_0(t)$ and $x_0(t)$ solves

$$\begin{cases} x_0'(t) = f_0(x_0(t)), & t \in [0, T], \\ x_0(0) = x_0 \in [0, 1]. \end{cases} \tag{23}$$

Using this theorem we are ready to show that the solutions of the scaled evolutionary game on complete graph converge uniformly to the solutions of replicator equations.

Theorem 7 *Let $x_n : \frac{1}{n}\mathbb{N}_0 \to \mathbb{R}$ be the mean solution of the scaled evolutionary game $\mathcal{E}_{sc} = (K_n, p, u^M, \mathcal{T}_{sc}, \varphi_{sc})$ with $x_n(0) = \lfloor nx_0 \rfloor / n$ for some $x_0 \in [0, 1]$, $x_n^*(t)$ its extension to $[0, T]$ for some $T > 0$ given by (22). Let the limit $\lim_{n\to\infty} x_n^*(t)$ exist. Then*

$$\lim_{n\to\infty} x_n^*(t) = x_0(t),$$

for each $t \in [0, T]$ where $x_0 : [0, T] \to \mathbb{R}$ is the solution of the replicator equation (6) with $x_0(0) = x_0$ and the convergence is uniform.

Proof We only provide a sketch of the proof, we don't verify all assumptions of Theorem 6 in detail. The recursive scheme (21) could be written as

$$\frac{x_n(t+h) - x_n(t)}{h} = x_n(t)\,(u_n(C, nx_n(t)) - \overline{u}_n(nx_n(t))) =: f_n(x_n(t)) \tag{24}$$

Taking the definitions (10), (11) into account we observe that the functions $f_n(x)$ are polynomials defined on $[0, 1]$. Therefore the first three assumptions of Theorem 6 are satisfied. Moreover, a closer analysis of (4), (5), (10), (11) and the fact that the polynomials are defined on the compact interval $[0, 1]$ imply that the following convergences are uniform:

$$u_n(C, nx_n^*(t)) \rightrightarrows u(C, x_0(t)), \quad \overline{u}_n(nx_n^*(t)) \rightrightarrows \overline{u}(x_0(t)),$$

Consequently, we have

$$f_n(x_n(t)) \rightrightarrows f_0(x_0(t)) := x_0(t)(u(C, x_0(t)) - \overline{u}(x_0(t))).$$

Theorem 6 then implies that the sequence $x_n^*(t)$ converges uniformly to $x_0(t)$ which is the solution the replicator equation (cf. (6))

$$x_0'(t) = x_0(t)(u(C, x_0(t)) - \overline{u}(x_0(t))).$$

Note that the uniform convergence is an indirect consequence of Theorem 6. Theorem 6 only states that there is a uniformly convergent subsequence of $\{x_n^*(t)\}$. But we can reapply Theorem 6 to get that each subsequence of $\{x_n^*(t)\}$ has a uniformly convergent subsequence. However, all these subsequences must have the same uni-

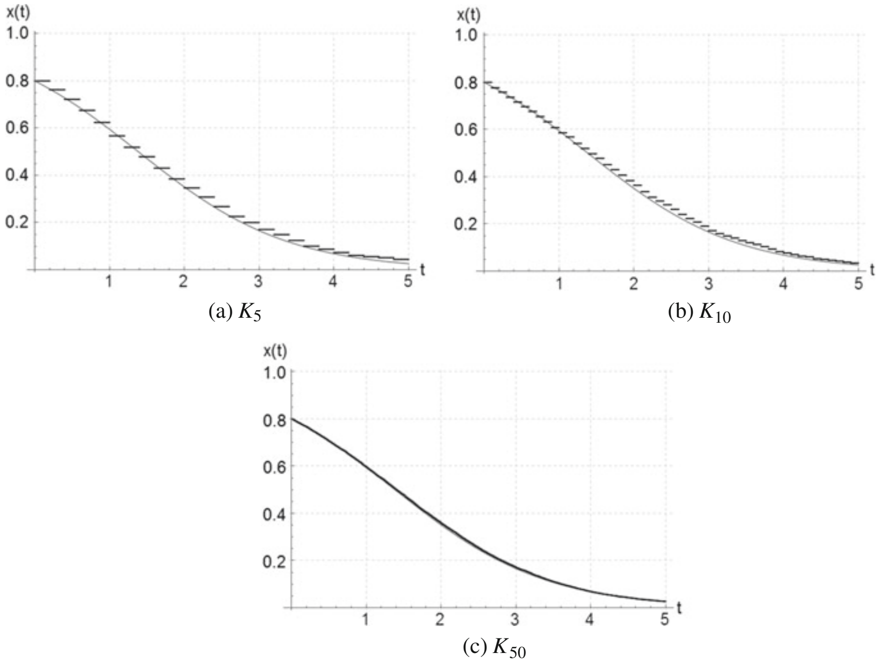


Fig. 4 Illustration to Theorem 7. Convergence of mean solutions $x_n^*(t)$ of evolutionary game $\mathscr{E}_{sc}$ on complete graphs (black step functions) to the solution $x(t)$ of replicator equation (6) (gray functions). The pictures depict the prisoner's dilemma scenario with $b = -1$ and $c = 2$, with initial condition $x(0) = 0.8$ and the mean solutions are computed from 500 simulations on complete graphs with $n = 5, 10, 50$ vertices

form limit, since the replicator equation (6) has at most one solution. This implies that the whole sequence $\{x_n^*(t)\}$ is uniformly convergent as well.[5]

The process of convergence is illustrated in Fig. 4. We emphasize the fact that our choice of evolutionary game on graphs, especially of the update order $\mathscr{T}_{sc}$ and the update rule φ_{sc}, is not unique. On the contrary, it is very specific and makes mostly sense only for complete graphs. There are other choices of update rules and update orders which lead to the convergence to trajectories of replicator equations. One could, for example, make update orders deterministic and consider non-constant step size where the length of the step size depends on the difference in utilities. We have chosen this particular example for the exact interpretations and a proof of convergence.

[5]The reason why Theorem 6 is not stated in this way directly is the fact that it could be applied also in cases when the problem (23) does not have a unique solution.

7 Summary and Final Remarks

The previous results indicate that the evolutionary games on graphs are not only a finite-dimensional counterpart of replicator equations but the exact solutions of replicator equations can be reconstructed as a limit of mean solutions of evolutionary games on complete graphs by intricate choices of update orders and update rules. The purpose of this paper has been to show this relationship in a clear way from a purely mathematical perspective. We note that, for example, the convergence of stability regions could be shown for any non-omitting update order (we discussed it only for synchronous and sequential update orders).

We emphasize that our main motivation was to show that the limit behaviour of evolutionary games on complete graphs correspond to the replicator equations. In contrast to the standard derivation of replicator equations [13, 25] we arrive to replicator equations from the finite population models. On the other hand, we admit that the updating probability p^* defined in (19) has weak justification in terms of biological interpretation for general graphs, since it contains the global property $\overline{u}$, which is the global average payoff in the population. In other words, this choice is an ad hoc choice for the case of complete graphs which we consider in this manuscript.

There are many open problems which are related to the mathematics of evolutionary games on graphs, even in the deterministic settings. We highlight especially the conditions ensuring the stability of fixed points, nonexistence of fixed points, existence of periodic solutions (is it as universal as the existence of coexistence equilibria [6]?) or the dynamics on special classes of graphs (even k-regular graphs are not fully described). The reader is invited to check the detailed commented lists of open problems in final sections of [5, 6].

Regarding this note, we highlight one additional open problem. Namely, the convergence of other classes of graphs. Note that we only considered the convergence of evolutionary games on complete graphs. But graph limits are much more general [16] and such limit processes could yield differential models generalizing traditional replicator equations.

Acknowledgements The author acknowledges the support by the Czech Science Foundation 15-00735S. Moreover, he is grateful to Hideaki Matsunaga for his patience, both referees for detailed comments and suggestions. Also he would like to thank Jeremias Epperlein, Stefan Siegmund and Vladimír Švígler for a beautiful cooperation on this project and the board of ISDE for their appreciation of our work, which means a lot for all of us.

References

1. Broom, M., Rychtář, J.: An analysis of the fixation probability of a mutant on special classes of non-directed graphs. Proc. Royal. Soc. Ser. A **464**, 2609–2627 (2008)
2. Broom, M., Hadjichrysanthou, C., Rychtář, J., Stadler, B.T.: Two results on evolutionary processes on general non-directed graphs. Proc. Royal. Soc. Ser. A **466**, 2795–2798 (2010)
3. Clutton-Brock, T.: Cooperation between non-kin in animal societies. Nature **462**, 51–57 (2009)

4. Dixit, A., Skeath, S., Reiley, D.: Games of strategy. W. W. Norton, New York (2009)
5. Epperlein, J., Siegmund, S., Stehlík, P.: Evolutionary games on graphs and discrete dynamical systems. J. Differ. Equ. Appl. **21**(2), 72–95 (2015)
6. Epperlein, J., Siegmund, S., Stehlík, P., Švígler, V.: Coexistence equilibria of evolutionary games on graphs under deterministic imitation dynamics. Discret. Contin. Dyn. Syst. Ser. B **21**(3), 803–813 (2016)
7. Fraňková, D.: Continuous dependence on a parameter of solutions of generalized differential equations. Časopis pro pěstování matematiky **114**(3), 230–261 (1989)
8. Gligor, M., Ausloos, M.: Clusters in weighted macroeconomic networks: the EU case introducing the overlapping index of GDP/capita fluctuation correlations. Eur. Phys. J. B **63**(4), 533–539 (2008)
9. Godsil, C., Royle, G.: Algebraic Graph Theory, 2nd edn. Springer, New York (2001)
10. Gross, T., Lima, C.J.D., Blasius, B.: Epidemic dynamics on an adaptive network. Phys. Rev. Lett. **96** (2006). article no. 208701
11. Hauert, C., Doebeli, M.: Spatial structure often inhibits the evolution of cooperation in the snowdrift game. Nature **428**, 643–646 (2004)
12. Hofbauer, J., Sigmund, K.: The Theory of Evolution and Dynamical Systems. Cambridge University Press, Cambridge (1988)
13. Hofbauer, J., Sigmund, K.: Evolutionary Game Dynamics. Bull. Am. Math. Soc. **40**, 479–519 (2003)
14. Libich, J., Stehlík, P.: Monetary policy facing fiscal indiscipline under generalized timing of actions. J. Institutional Theor. Econ. **168**, 393–431 (2012)
15. Libich, J., Nguyen, D.: Strategic Monetary-Fiscal Interactions in a downturn. Econ. Rec. **91**(293), 172–190 (2015)
16. Lovász, L.: Large Networks and Graph Limits. American Mathematical Society, Providence (2012)
17. Myerson, R.: Game Theory: Analysis of Conflict. Harvard University Press, Cambridge (1997)
18. Nash, J.F.: The bargaining problem. Econometrica **18**, 155–162 (1950)
19. Nowak, M.A.: Evolutionary Dynamics: Exploring the Equations of Life. Harvard University Press, Cambridge (2006)
20. Nowak, M.A.: Five rules for the evolution of cooperation. Science **314**, 1560–1563 (2006)
21. Nowak, M.A., May, R.M.: Evolutionary games and spatial chaos. Nature **359**, 826–829 (1992)
22. Ohtsuki, H., Nowak, M.A.: Evolutionary games on cycles. Proc. R. Soc. B **273**, 2249–2256 (2006)
23. Ohtsuki, H., Nowak, M.A.: Evolutionary stability on graphs. J. Theor. Biol. **251**, 698–707 (2008)
24. Perc, M., Szolnoki, A.: Coevolutionary games–a mini review. BioSystems **99**, 109–125 (2010)
25. Sigmund, K.: The Calculus of Selfishness. Princeton University Press, Princeton (2010)
26. Slavík, A.: Dynamic equations on time scales and generalized ordinary differential equations. J. Math. Anal. Appl. **385**, 534–550 (2012)
27. Smith, J.M.: The theory of games and the evolution of animal conflicts. J. Theor. Biol. **47**, 209–221 (1974)
28. Szabó, G., Fáth, G.: Evolutionary games on graphs. Phys. Rep. **446**, 97–216 (2007)
29. Traulsen, A., Claussen, J.C., Hauert, C.: Coevolutionary dynamics: from finite to infinite populations. Phys. Rev. Lett. **95** (2005). article no. 238701

Part II
Contributed Papers

Connection Between Continuous and Discrete Delay and Halanay type Inequalities

István Győri and László Horváth

Abstract The main results of the paper are complimenting, extending and improving several earlier results obtained for Halanay type discrete difference inequalities. The novel idea is that the discrete time results are derived from our recent related continuous time results by using suitable delay differential inequalities with piecewise constant arguments. The sharpness of the results are illustrated by examples.

Keywords Delay difference and differential inequalities · Piecewise constant arguments · Asymptotic behavior of the solutions

MSC2010 Classification 26D10 · 65Q10

1 Introduction

Halanay [12] proved an upper estimation for the set of the nonnegative solutions of an autonomous continuous time delay differential inequality with maxima. This, so called Halanay inequality became a powerful tool in the stability theory of delay differential equations, therefore several authors improved, generalized and applied it (see for instance Baker and Tang [4], Baker [3], Ivanov, Liz and Trofimchuk [12], Mohamad and Gopalsamy [16], and Wen, Yu and Wang [19]).

The set of nonnegative numbers and the set of nonnegative integers will be denoted by $\mathbb{R}_+$ and $\mathbb{N}$ respectively.

The authors of this work, in a recent comprehensive study [10] (see also Győri and Horváth [9]), complemented, extended and improved several results from the earlier literature for the differential inequality

I. Győri · L. Horváth (✉)
Department of Mathematics, University of Pannonia, Egyetem u. 10,
Veszprém 8200, Hungary
e-mail: lhorvath@almos.uni-pannon.hu

I. Győri
e-mail: gyori@almos.uni-pannon.hu

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_5

$$x'(t) \le -\alpha(t)\,x(t) + \beta(t)\,x(t-\tau(t)), \quad t \ge t_0, \tag{1}$$

and the Halanay type inequality

$$x'(t) \le -\alpha(t)\,x(t) + \beta(t) \sup_{t-\tau(t)\le s\le t} x(s), \quad t \ge t_0, \tag{2}$$

under the mild conditions

(A_1) $t_0 \in \mathbb{R}$ is fixed, the functions $\alpha : [t_0, \infty[\to \mathbb{R}$ and $\beta : [t_0, \infty[\to \mathbb{R}_+$ are locally integrable,
(A_2) $\tau : [t_0, \infty[\to \mathbb{R}_+$ is a measurable function and it obeys the inequality

$$t_0 - r \le t - \tau(t), \quad t \ge t_0$$

with a constant $r \ge 0$.

Based on the importance of the discrete time dynamical systems there are several works in the literature (see Agarwal, Young-Ho Kim and Sen [1] and [2], Liz and Ferreiro [13], Liz, Ivanov and Ferreiro [14], Liz, Tkachenko and Trofimchuk [15], Song, Shen and Yin [18], and Xu [20]) devoted to the reformulation of the continuous time results for the discrete time difference inequality

$$y(n+1) \le a(n)\,y(n) + b(n)\,y(n-k(n)), \quad n \ge n_0, \tag{3}$$

and the Halanay type difference inequality

$$y(n+1) \le a(n)\,y(n) + b(n) \max_{n-k(n)\le i\le n} y(i), \quad n \ge n_0. \tag{4}$$

where

$$0 < a(n) < 1, \quad b(n) \ge 0, \quad n \ge n_0, \tag{5}$$

and

$$k(n) \in \mathbb{N}, \quad n - k(n) \ge n_0 - l, \quad n \ge n_0 \tag{6}$$

with a fixed integer n_0 and a fixed $l \in \mathbb{N}$.

We say that a function $x : [t_0 - r, \infty[\to \mathbb{R}$ is a solution of the differential inequalities (1) or (2) if x is Borel measurable and bounded on $[t_0 - r, t_0]$, locally absolutely continuous on $[t_0, \infty[$, and x satisfies (1) or (2) almost everywhere on $[t_0, \infty[$, respectively.

We say that the real sequence $(y(n))_{n\ge n_0-l}$ is a solution of the difference inequalities (3) or (4) if it satisfies (3) or (4) for all $n \ge n_0$, respectively.

It is interesting to note that in all of the above papers the proof of the discrete version of a continuous case result is basically a repetition of the original proof, of course, with some suitable changes.

In this paper we use a novel idea. Namely, we show that the qualitative properties of the solutions of our discrete inequalities are equivalent to the qualitative properties of the solutions of some suitable delay differential inequalities with piecewise constant argument.

The theory of delay differential equations with piecewise constant argument (EPCA) was initiated and studied by Cooke and Wiener [6] and [7]. The idea of approximating the solutions of continuous time delay differential equations with the solutions of a suitable constructed EPCA has been suggested by Győri [8] who proved convergence of the method for linear and nonlinear delay equations on compact intervals. In Cooke and Győri [5], it was pointed out that the approximation may be extended to noncompact intervals. The interested readers may refer to the further paper Sepúlveda [17].

The integer part of a real number r will be denoted by $[r]$ ($[r]$ is the largest integer less than or equal to r).

The connection among the continuous time inequalities (1) and (2) and the discrete time inequalities (3) and (4), respectively is given by the relations

$$\alpha(t) := a_1([t]), \quad \beta(t) := b_1([t]), \quad t-\tau(t) := [t]-k([t]), \quad t \geq n_0, \quad (7)$$

where

$$a_1(n) := -\ln(a(n)), \quad b_1(n) := \frac{-\ln(a(n))}{1-a(n)} b(n), \quad n \geq n_0. \quad (8)$$

Clearly, $a_1(n) > 0$ and $b_1(n) \geq 0$ for all $n \geq n_0$.

We shall see that the delay differential inequality

$$x'(t) \leq -\alpha(t)x(t) + \beta(t)x(t-\tau(t)), \quad t \geq n_0, \quad (9)$$

and the Halanay type differential inequality

$$x'(t) \leq -\alpha(t)x(t) + \beta(t) \sup_{t-\tau(t)\leq s\leq t} x(s), \quad t \geq n_0, \quad (10)$$

where α, β and τ are defined in (7), are closely related to the difference inequalities (3) and (4).

Remark 1 (a) Here α and β are piecewise constant functions, while τ is piecewise continuous, and therefore they are Borel measurable. It follows from (5–8) that $\alpha(t) > 0$, $\beta(t) \geq 0$, $\tau(t) \geq 0$ and $t-\tau(t) \geq n_0 - l$ for all $t \geq n_0$.

These conditions guarantee that (A_1) and (A_2) are satisfied for (9) and (10).

(b) It is easy to check that $\beta(t) \leq \alpha(t)$ $(t \geq n_0)$ if and only if $a(n) + b(n) \leq 1$ $(n \geq n_0)$. It is also obvious that $\lim_{t\to\infty} (t - \tau(t)) = \infty$ if and only if $\lim_{n\to\infty} (n - k(n)) = \infty$.

The paper is organized as follows: Sect. 2 contains some important lemmas on the equivalent behavior of the nonnegative solutions of our discrete and the related continuous inequalities. These lemmas clearly show the important role of the delay differential inequalities with piecewise constant arguments. The main results which compliment, extend and improve several earlier results are presented in Sect. 3. Finally, in Sect. 4 we compliment, extend and improve the result of Mohamad and Gopalsamy [16] given for a discrete inequality. Here we also illustrate the sharpness of some conditions with some examples.

This work has been inspired by an open discussion during the 22nd International Conference on Difference Equations and Applications, Osaka, Japan July 24–29, 2016 about the relevance of the continuous and discrete time model equations and their connections, initiated by Professors S. Elaydi and J. M. Cushing.

2 Basic Lemmas on the Equivalence of the Asymptotic Behavior of the Solutions of Continuous and Discrete Inequalities

To study the convergence of the nonnegative solutions of either (3) or (4) we need some lemmas.

Lemma 2 *Suppose that (5) and (6) are satisfied. Let $(y(n))_{n\geq n_0-l}$ be a nonnegative solution of (3), and denote $(z(n))_{n\geq n_0-l}$ be the unique solution of the initial value problem*

$$\left.\begin{array}{c} z(n+1) = a(n)z(n) + b(n)z(n-k(n)), \quad n \geq n_0, \\ z(n_0 - l) := y(n_0 - l), \ldots, z(n_0) := y(n_0) \end{array}\right\}. \tag{11}$$

Then

$$y(n) \leq z(n), \quad n \geq n_0 - l.$$

Proof We can apply an easy induction argument, by using that $a(n) > 0$ and $b(n) \geq 0$ for all $n \geq n_0$. ■

Lemma 3 *Suppose that (5) and (6) are satisfied.*

(a) Let $(y(n))_{n\geq n_0-l}$ be a nonnegative solution of (3), and denote $(z(n))_{n\geq n_0-l}$ be the unique solution of the initial value problem (11). Define the function $x : [n_0 - l, \infty[\to \mathbb{R}$ by

$$x(t) := e^{-a_1(n)(t-n)} z(n) \tag{12}$$
$$+ \frac{b_1(n)}{a_1(n)} \left(1 - e^{-a_1(n)(t-n)}\right) z(n - k(n)), \quad n \le t < n+1, \quad n \ge n_0,$$

and

$$x(t) := y(n), \quad n \le t < n+1, \quad n_0 - l \le n < n_0. \tag{13}$$

Then x is a nonnegative solution of the differential inequality (9) such that

$$y(n) \le x(n), \quad n_0 - l \le n. \tag{14}$$

(b) The set of nonnegative solutions of the Halanay type difference inequality (4) is the same as the set of nonnegative solutions $(y(n))_{n \ge n_0 - l}$ *of all delay difference inequalities*

$$y(n+1) \le a(n) y(n) + b(n) y(n - p(n)), \quad n \ge n_0, \tag{15}$$

where $(p(n))_{n \ge n_0}$ *is an integer valued sequence satisfying*

$$0 \le p(n) \le k(n), \quad n \ge n_0. \tag{16}$$

Proof (a) Obviously, x is nonnegative, and

$$x(n) = z(n), \quad n \ge n_0 - l. \tag{17}$$

Since x is piecewise constant on $[n_0 - l, n_0]$, it is Borel measurable and bounded on $[n_0 - l, n_0]$. It is not hard to check that x is continuous on $[n_0, \infty[$, and

$$x'(t) = -a_1(n) e^{-a_1(n)(t-n)} z(n)$$
$$+ b_1(n) e^{-a_1(n)(t-n)} z(n - k(n)), \quad n \le t < n+1, \quad n \ge n_0, \tag{18}$$

where $x'(n)$ $(n \ge n_0)$ means right-hand derivative in (18). The left-hand derivative at n $(n \ge n_0 + 1)$ also exists, but the function x is not differentiable at n in general.

It follows that

$$\left|x'(t)\right| \le a_1(n) z(n) + b_1(n) z(n - k(n)), \quad n \le t < n+1, \quad n \ge n_0,$$

and hence the right-hand side derivative of x is bounded on $[n_0, t]$ for all $t > n_0$. We can see that x is locally absolutely continuous on $[n_0, \infty[$.

By (18), some easy calculation shows that

$$\begin{aligned} x'(t) &= -a_1(n)\,x(t) + b_1(n)\,x(n-k(n)) \\ &= -a_1(n)\,x(t) + b_1(n)\,x(t-\tau(t)), \quad n \le t < n+1, \quad n \ge n_0, \end{aligned}$$

thus x is a solution of the delay differential equation

$$x'(t) = -\alpha(t)\,x(t) + \beta(t)\,x(t-\tau(t)), \quad t \ge n_0,$$

and hence x is a solution of (9).

Finally, (14) follows from (17) and Lemma 2.

(b) Since the set $\{n-k(n), \ldots, n\}$ is finite for every $n \ge n_0$, for any fixed nonnegative solution $(y(n))_{n \ge n_0 - l}$ of (4) there exists an integer valued sequence $(p(n))_{n \ge n_0}$ (depending on this solution) which satisfies (16) and

$$\max_{n-k(n) \le i \le n} y(i) = y(n-p(n)), \quad n \ge n_0.$$

Thus $(y(n))$ obeys (15).

Conversely, if $(y(n))_{n \ge n_0 - l}$ is a nonnegative solution of (15) with a sequence $(p(n))_{n \ge n_0}$ satisfying (16), then $b(n) \ge 0$ $(n \ge n_0)$ yields that $(y(n))$ is a solution of (4) too.

The proof is complete. ■

Lemma 4 *Suppose that (5) and (6) are satisfied. Then*

(a) If every nonnegative solution of the differential inequality (9) tends to zero at infinity, then every nonnegative solution of the difference inequality (3) tends to zero at infinity.

(b) If

$$a(n) + b(n) \le 1, \quad (n \ge n_0) \tag{19}$$

and

$$\lim_{n\to\infty} (n-k(n)) = \infty, \tag{20}$$

then the reverse assertion in (a) also holds.

(c) If every nonnegative solution of the Halanay type differential inequality (10) tends to zero at infinity, then every nonnegative solution of the Halanay type difference inequality (4) tends to zero at infinity.

(d) If (19) and (20) are satisfied, then the reverse assertion in (c) also holds.

Proof (a) Let $(y(n))_{n \ge n_0 - l}$ be a nonnegative solution of (3). By Lemma 3 (a), the function x defined by (12) and (13) is a nonnegative solution of (9). Consequently,

$$\lim_{t\to\infty} x(t) = 0. \tag{21}$$

Since (14) is satisfied, (21) tells us that $\lim_{n-\infty} y(n) = 0$.

(b) Let $x : [n_0 - l, \infty[\to \mathbb{R}$ be a nonnegative solution of (9). Then for any fixed $n \geq n_0$ we have

$$x'(t) \leq -a_1(n)\, x(t) + b_1(n)\, x(n - k(n)), \quad \text{a.e. on} \quad [n, n+1[\,,$$

and hence the Gronwall–Bellman inequality yields

$$\begin{aligned} x(t) &\leq x(n)\, e^{-a_1(n)(t-n)} \\ &+ b_1(n)\, x(n-k(n)) \frac{1}{a_1(n)} \left(1 - e^{-a_1(n)(t-n)}\right), \quad n \leq t < n+1. \end{aligned} \tag{22}$$

Since x is continuous on $[n_0, \infty[$, (22) gives that

$$\begin{aligned} x(n+1) &\leq x(n)\, e^{-a_1(n)} + b_1(n)\, x(n-k(n)) \frac{1}{a_1(n)} \left(1 - e^{-a_1(n)}\right) \\ &= a(n)\, x(n) + b(n)\, x(n-k(n)), \quad n \geq n_0. \end{aligned}$$

It can be seen that $(x(n))_{n \geq n_0 - l}$ is a nonnegative solution of (3), and hence

$$\lim_{n\to\infty} x(n) = 0. \tag{23}$$

Thus (20) implies

$$\lim_{n\to\infty} x(n - k(n)) = 0. \tag{24}$$

By using $a_1(n) > 0$ $(n \geq n_0)$, we get that for every $n \geq n_0$

$$0 < e^{-a_1(n)(t-n)} \leq 1, \quad n \leq t < n+1. \tag{25}$$

From (8), (5) and (19) we can deduce that

$$0 < \frac{b_1(n)}{a_1(n)} = \frac{b(n)}{1 - a(n)} \leq 1, \quad n \geq n_0. \tag{26}$$

The inequality (22) together with (23–26) give that $\lim_{t\to\infty} x(t) = 0$.

(c) It follows from Lemma 3 (b) that every nonnegative solution of (4) tends to zero at infinity if and only if every nonnegative solution of (15) tends to zero at infinity for every integer valued sequence $(p(n))_{n\geq n_0}$ satisfying (16).

Now let $(p(n))_{n\geq n_0}$ be an integer valued sequence satisfying (16), and let $\eta : [n_0, \infty[\to \mathbb{R}_+$ be defined by

$$t-\eta(t):=[t]-p([t]), \quad t \geq n_0.$$

Then η is a nonnegative measurable function satisfying

$$n_0-l \leq t-\eta(t) \quad \text{and} \quad \eta(t) \leq \tau(t), \quad t \geq n_0.$$

Since every nonnegative solution of the delay differential inequality

$$x'(t) \leq -\alpha(t)x(t)+\beta(t)x(t-\eta(t)), \quad t \geq n_0, \tag{27}$$

is obviously a solution of (10), every nonnegative solution of (27) tends to zero at infinity.

By applying (a) to (27) in this case, it follows that every nonnegative solution of (15) tends to zero at infinity.

(d) Since every nonnegative solution of (3) is a solution of (4), every nonnegative solution of (3) tends to zero at infinity. It now follows from (b) that every nonnegative solution of (9) also tends to zero at infinity. Therefore by Remark 1 (b) and by (19), we have from Theorem 2.8 in [10] that every nonnegative solution of (10) tends to zero at infinity too.

The proof is complete. ■

We define the notion of exponential convergence.

Definition 5 (a) We say that every nonnegative solution $y : [n_0-l, \infty[\to \mathbb{R}$ of the difference inequalities (3) or (4) tends to zero exponentially at infinity if there are $\kappa > 0$ and $K \geq 0$ (independent of the solutions) such that

$$y(n) \leq K \max_{n_0-l \leq i \leq n_0} y(i)\, e^{-\kappa(n-n_0)}, \quad n \geq n_0,$$

where κ is called as rate of convergence.

(b) We say that every nonnegative solution $x : [n_0-l, \infty[\to \mathbb{R}$ of the differential inequalities (9) or (10) tends to zero exponentially at infinity if there are $\kappa > 0$ and $K \geq 0$ (independent of the solutions) such that

$$x(t) \leq K \sup_{n_0-l \leq s \leq n_0} x(s)\, e^{-\kappa(t-n_0)}, \quad t \geq n_0,$$

where κ is called as rate of convergence.

The next result is analogous to Lemma 4.

Lemma 6 *Suppose that (5) and (6) are satisfied. Then*

(a) If every nonnegative solution of the differential inequality (9) tends to zero exponentially at infinity with the convergence rate $\kappa > 0$, then every nonnegative solution of the difference inequality (3) tends to zero exponentially at infinity with the same convergence rate.

(b) If

$$a(n) + b(n) \leq 1, \quad (n \geq n_0), \tag{28}$$

and the sequence $(k(n))_{n \geq n_0}$ *is bounded, then the reverse assertion in (a) also holds.*

(c) If every nonnegative solution of the Halanay type differential inequality (10) tends to zero exponentially at infinity with the convergence rate $\kappa > 0$*, then every nonnegative solution of the Halanay type difference inequality (4) tends to zero exponentially at infinity with the same convergence rate.*

(d) If (28) is satisfied, and the sequence $(k(n))_{n \geq n_0}$ *is bounded, then the reverse assertion in (c) also holds.*

Proof (a) We can follow the contexture of the proof of Lemma 4 (a).

By Lemma 3 (a), the function $x : [n_0 - l, \infty[\to \mathbb{R}$ defined by (12) and (13) is a nonnegative solution of (9), and hence there are $\kappa > 0$ and $K \geq 0$ such that

$$x(t) \leq K \sup_{n_0 - l \leq s \leq n_0} x(s) e^{-\kappa(t - n_0)}, \quad t \geq n_0.$$

We obtain from this and (14) that

$$y(n) \leq x(n) \leq K \max_{n_0 - l \leq i \leq n_0} y(i) e^{-\kappa(n - n_0)}, \quad t \geq n_0,$$

which shows the result.

(b) Exactly as in the proof of Lemma 4 (b), we have that $(x(n))_{n \geq n_0 - l}$ is a nonnegative solution of (3), and hence there are $\kappa > 0$ and $K \geq 0$ such that

$$x(n) \leq K \max_{n_0 - l \leq i \leq n_0} x(i) e^{-\kappa(n - n_0)}, \quad n \geq n_0.$$

Now, inequality (22) gives that

$$\begin{aligned} x(t) \leq K \max_{n_0 - l \leq i \leq n_0} x(i) e^{-\kappa(n - n_0)} e^{-a_1(n)(t-n)} + \frac{b_1(n)}{a_1(n)} \left(1 - e^{-a_1(n)(t-n)}\right) \\ \times K \max_{n_0 - l \leq i \leq n_0} x(i) e^{-\kappa(n - k(n) - n_0)}, \quad n \leq t < n + 1. \end{aligned}$$

This implies by using (25) and (26), that

$$x(t) \leq K \max_{n_0 - l \leq i \leq n_0} x(i) e^{-\kappa(n - n_0)} \left(1 + e^{\kappa k(n)}\right), \quad n \leq t < n + 1.$$

Since $(k(n))$ is bounded $(0 \leq k(n) \leq \tau, \; n \geq n_0)$, we have

$$x(t) \leq K(1 + e^{\kappa\tau}) \max_{n_0 - l \leq i \leq n_0} x(i) e^{-\kappa(n - n_0)}$$

$$\leq K\left(1+e^{\kappa\tau}\right)e^{\kappa}\sup_{n_0-l\leq s\leq n_0}x\left(s\right)e^{-\kappa\left(t-n_0\right)}\quad t\geq n_0,$$

and thus the desired conclusion is obtained.

(c) Lemma 3 (b) shows that every nonnegative solution of (4) tends to zero exponentially at infinity with the rate of convergence $\kappa > 0$ if and only if every nonnegative solution of (15) tends to zero exponentially at infinity with the same rate of convergence κ for every integer valued sequence $(p\,(n))_{n\geq n_0}$ satisfying (16).

We can follow the proof as in Lemma 4 (c): every nonnegative solution of (27) tends to zero exponentially at infinity, and the rate of convergence $\kappa > 0$ is the same as for (10). By applying (a) to (27) in this case, it follows that every nonnegative solution of (15) tends to zero exponentially at infinity with the rate of convergence κ.

(d) We can prove exactly as in Lemma 4 (d) by using Theorem 2.11 in [10] instead of Theorem 2.8 in [10].

The proof is complete. ■

We close this section with the next result.

Lemma 7 *Suppose that (5) and $a\,(n)+b\,(n)\leq 1$ $(n\geq n_0)$ hold, and consider the functions α and β defined in (7). Then*

(a)

$$\int_{n_0}^{\infty}\left(\alpha\left(s\right)-\beta\left(s\right)\right)ds=\sum_{n=n_0}^{\infty}\frac{1-a\left(n\right)-b\left(n\right)}{1-a\left(n\right)}\ln\left(\frac{1}{a\left(n\right)}\right).\tag{29}$$

(b)

$$\liminf_{t\to\infty}\frac{1}{t-n_0}\int_{n_0}^{t}\left(\alpha\left(s\right)-\beta\left(s\right)\right)ds$$
$$=\liminf_{n\to\infty}\frac{1}{n-n_0}\sum_{i=n_0}^{n}\frac{1-a\left(i\right)-b\left(i\right)}{1-a\left(i\right)}\ln\left(\frac{1}{a\left(i\right)}\right).\tag{30}$$

Proof (a) Since $\beta\,(t)\leq\alpha\,(t)$ $(t\geq n_0)$, the integral exists. Now, (29) follows from the definitions of α and β.

(b) Clearly,

$$\frac{1}{t-n_0}\int_{n_0}^{t}\left(\alpha\left(s\right)-\beta\left(s\right)\right)ds$$
$$=\frac{[t]-1-n_0}{t-n_0}\left(\frac{1}{[t]-1-n_0}\sum_{i=n_0}^{[t]-1}\frac{1-a\left(i\right)-b\left(i\right)}{1-a\left(i\right)}\ln\left(\frac{1}{a\left(i\right)}\right)\right)$$

$$+\frac{t-[t]}{t-n_0}\cdot\frac{1-a([t])-b([t])}{1-a([t])}\ln\left(\frac{1}{a([t])}\right),\quad t\ge n_0+1,$$

and this implies (30).

The proof is complete. ■

3 Main Results

Theorem 8 *Suppose that (5), (6) and $a(n)+b(n)\le 1$ $(n\ge n_0)$ are satisfied.*

(a) Assume further that $\lim_{n\to\infty}(n-k(n))=\infty$. If every nonnegative solution of either the difference inequality (3) or the Halanay type difference inequality (4) tends to zero at infinity, then

$$\sum_{n=n_0}^{\infty}\frac{1-a(n)-b(n)}{1-a(n)}\ln\left(\frac{1}{a(n)}\right)=\infty. \tag{31}$$

(b) Assume further that the sequence $(k(n))_{n\ge n_0}$ is bounded. If every nonnegative solution of either the difference inequality (3) or the Halanay type difference inequality (4) tends to zero exponentially at infinity, then

$$\liminf_{n\to\infty}\frac{1}{n-n_0}\sum_{i=n_0}^{n}\frac{1-a(i)-b(i)}{1-a(i)}\ln\left(\frac{1}{a(i)}\right)>0.$$

Proof (a) By Remark 1, Theorem 3.1 (a) in [10] guarantees that if every nonnegative solution of either (9) or (10) tends to zero at infinity, then

$$\int_{n_0}^{\infty}(\alpha(s)-\beta(s))\,ds=\infty.$$

In light of Lemma 7 (a), either Lemma 4 (b) or Lemma 4 (d) can be applied.

(b) Similar reasoning, starting with Theorem 3.1 (b) in [10], gives that if every nonnegative solution of either (9) or (10) tends to zero exponentially at infinity, then

$$\liminf_{t\to\infty}\frac{1}{t-n_0}\int_{n_0}^{t}(\alpha(s)-\beta(s))\,ds>0.$$

The result follows from this by using Lemma 7 (b) and either Lemma 6 (b) or Lemma 6 (d).

The proof is complete. ■

Theorem 9 *Suppose that (5), (6), $a(n)+b(n)\le 1$ $(n\ge n_0)$ and $\lim\limits_{n\to\infty}(n-k(n))=\infty$ are satisfied. Assume further that there exists a constant $0<q<1$ such that*

$$\limsup_{n\to\infty}\sum_{i=n_0}^{n}\frac{q-qa(i)-b(i)}{1-a(i)}\ln\left(\frac{1}{a(i)}\right)=\infty.$$

Then every nonnegative solution of either the difference inequality (3) or the Halanay type difference inequality (4) tends to zero at infinity.

Proof Since

$$\limsup_{t\to\infty}\int_{n_0}^{t}(q\alpha(s)-\beta(s))\,ds$$
$$\ge\limsup_{n\to\infty}\sum_{i=n_0}^{n}\frac{q-qa(i)-b(i)}{1-a(i)}\ln\left(\frac{1}{a(i)}\right),$$

it follows that

$$\limsup_{t\to\infty}\int_{n_0}^{t}(q\alpha(s)-\beta(s))\,ds=\infty. \tag{32}$$

Then Remark 1, $\lim\limits_{t\to\infty}(t-\tau(t))=\infty$, and (32) show that Theorem 3.3 in [10] can be applied to the differential inequalities (9) and (10). According to this theorem, every nonnegative solution of either (9) or (10) tends to zero at infinity.

Lemma 4 (a) and Lemma 4 (c) imply the assertions.

The proof is complete. ■

In the next two results we use the condition: there exists a constant $0<q<1$ such that

$$b(n)\le q(1-a(n)),\quad n\ge n_0. \tag{33}$$

Clearly, (33) is equivalent to

$$\beta(t)\le q\alpha(t),\quad t\ge n_0. \tag{34}$$

Theorem 10 *Suppose that (5), (6), (33) and $\lim\limits_{n\to\infty}(n-k(n))=\infty$ are satisfied.*

Then every nonnegative solution of either the difference inequality (3) or the Halanay type difference inequality (4) tends to zero at infinity if and only if

$$\sum_{n=n_0}^{\infty} \ln\left(\frac{1}{a(n)}\right) = \infty. \tag{35}$$

Proof According to (33)

$$(1-q)\sum_{i=n_0}^{n} \ln\left(\frac{1}{a(i)}\right) \le \sum_{i=n_0}^{n} \frac{1-a(i)-b(i)}{1-a(i)} \ln\left(\frac{1}{a(i)}\right) \le \sum_{i=n_0}^{n} \ln\left(\frac{1}{a(i)}\right), \tag{36}$$

which implies that (35) holds if and only if

$$\sum_{n=n_0}^{\infty} \frac{1-a(n)-b(n)}{1-a(n)} \ln\left(\frac{1}{a(n)}\right) = \infty. \tag{37}$$

Remark 1 and (34) show that the conditions of Theorem 3.5 in [10] hold. By applying this theorem we have that every nonnegative solution of either (9) or (10) tends to zero at infinity if and only if

$$\int_{n_0}^{\infty} (\alpha(s) - \beta(s))\, ds = \infty,$$

but Lemma 7 (a) shows that this condition is equivalent to (37).

The result comes from Lemma 4 (a) and (b), and Lemma 4 (c) and (d), respectively. The proof is complete. ■

Remark 11 Let $n_0 := 1$. If $a(n) := e^{-\frac{1}{n}}$ $(n \ge 1)$, then (35) holds, while if $a(n) := e^{-\frac{1}{n^2}}$ $(n \ge 1)$, then (35) does not hold.

Theorem 12 *Suppose that (5), (6) and (33) are satisfied. Assume further that*

$$M := \sup_{n \ge n_0} \sum_{i=\max(n-k(n),n_0)}^{n} \frac{1-a(i)-b(i)}{1-a(i)} \ln\left(\frac{1}{a(i)}\right) < \infty. \tag{38}$$

(a_1) For every nonnegative solution $(y(n))_{n \ge n_0 - l}$ of the difference inequality (3) we have

$$y(n) \le \max_{n_0 - l \le i \le n_0} y(i) \exp\left(-\mu \sum_{i=n_0}^{n} \frac{1-a(i)-b(i)}{1-a(i)} \ln\left(\frac{1}{a(i)}\right)\right), \quad n \ge n_0, \tag{39}$$

where $\mu \in]0, 1]$ *is the unique root of the equation*

$$qe^{\mu M} + \mu (1 - q) - 1 = 0. \tag{40}$$

(a_2) For every nonnegative solution $(y(n))_{n \geq n_0 - l}$ *of the Halanay type difference inequality (4) satisfies (39) too.*

(b) Every nonnegative solution of either the difference inequality (3) or the Halanay type difference inequality (4) tends to zero exponentially at infinity if

$$\liminf_{n \to \infty} \frac{1}{n - n_0} \sum_{i=n_0}^{n} \ln \left(\frac{1}{a(i)} \right) > 0. \tag{41}$$

This condition is also necessary if the sequence $(k(n))_{n \geq n_0}$ *is bounded.*

Proof (a_1) Define the function $x : [n_0 - l, \infty[\to \mathbb{R}$ by (12) and (13). As we have seen in Lemma 3 (a), x is a nonnegative solution of (9).

By (38) and (7), it is clear that

$$M = \sup_{t \geq n_0} \int_{\max(t - \tau(t), n_0)}^{t} (\alpha(s) - \beta(s))\, ds < \infty,$$

and hence, by Remark 1, Theorem 3.6 (a) in [10] can be applied, which gives that

$$x(t) \leq \sup_{n_0 - l \leq s \leq n_0} x(s) \exp \left(-\mu \int_{n_0}^{t} (\alpha(s) - \beta(s))\, ds \right), \quad t \geq n_0. \tag{42}$$

Since $y(n) \leq x(n)$ for every $n \geq n_0$ (see (14)), and

$$\sup_{n_0 - l \leq s \leq n_0} x(s) = \max_{n_0 - l \leq i \leq n_0} y(i),$$

the result comes from (42).

(a_2) The condition (38) holds with the same constant M for every integer valued sequence $(p(n))_{n \geq n_0}$ satisfying (16), and hence Lemma 3 (b) and (a_1) can be applied.

(b) From (36) we have that (41) holds if and only if

$$\liminf_{n \to \infty} \frac{1}{n - n_0} \sum_{i=n_0}^{n} \frac{1 - a(i) - b(i)}{1 - a(i)} \ln \left(\frac{1}{a(i)} \right) > 0. \tag{43}$$

By Lemma 7 (b), the condition (43) is equivalent to

$$\liminf_{t\to\infty} \frac{1}{t-n_0} \int_{n_0}^{t} (\alpha(s) - \beta(s))\, ds > 0,$$

and therefore Theorem 3.6 (c) in [10] yields that every nonnegative solution of the differential inequalities (9) and (10) tends to zero exponentially at infinity. Thus Lemmas 6 (a) and (c) can be applied, respectively.

Under the additional assumption, the condition is also necessary by Theorem 8 (b).

The proof is complete. ■

Remark 13 Assume that the conditions of the previous theorem are satisfied. If (41) is also satisfied, then Remark 14 in [10] shows that $\lim_{n\to\infty} (n - k(n)) = \infty$.

Theorem 14 *Suppose that (5), (6) and $a(n) + b(n) \le 1$ $(n \ge n_0)$ are satisfied.*

(a_1) There exists a constant $\kappa \ge 0$ such that for every nonnegative solution $(y(n))_{n\ge n_0-l}$ of the difference equation (3)

$$y(n) \le \max_{n_0-l\le i\le n_0} y(i)\, e^{-\kappa(n-n_0)}, \quad n \ge n_0. \tag{44}$$

(a_2) There exists a constant $\kappa \ge 0$ such that for every nonnegative solution $(y(n))_{n\ge n_0-l}$ of the Halanay type difference inequality (4) satisfies (44).

(b) If the sequence $(k(n))_{n\ge n_0}$ is bounded and

$$\inf_{n\ge n_0} \frac{\ln\left(\frac{1}{a(n)}\right)(1 - a(n) - b(n))}{1 - a(n) + \ln\left(\frac{1}{a(n)}\right) b(n)} > 0, \tag{45}$$

then there exists $\kappa > 0$ such that (44) holds, and thus every nonnegative solution of either the difference inequality (3) or the Halanay type difference inequality (4) tends to zero exponentially at infinity.

Proof $\kappa = 0$ is obviously satisfies the inequality

$$\kappa + \frac{\ln\left(\frac{1}{a(n)}\right)}{1 - a(n)} b(n)\, e^{\kappa(t-n+k(n))} \le \ln\left(\frac{1}{a(n)}\right), \quad n \le t < n+1, \quad n \ge n_0, \tag{46}$$

and hence the definitions of the functions α, β and τ in (7) give that

$$\kappa + \beta(t)\, e^{\kappa\tau(t)} \le \alpha(t), \quad t \ge n_0. \tag{47}$$

By Remark 1, Theorem 3.9 (b_1) in [10] can be applied, which insures that for every nonnegative solution $x : [n_0 - l, \infty[\to \mathbb{R}$ of the differential inequalities (9) and (10)

$$x(t) \le \sup_{n_0 - l \le s \le n_0} x(s) e^{-\kappa(t-n_0)}, \quad n \ge n_0.$$

(a_1) By using the solution of (9) defined in (12) and (13), we have (44).

(a_2) Let $(p(n))_{n \ge n_0}$ be an integer valued sequence satisfying (16), and consider the delay differential inequality (27). Since

$$\kappa + \beta(t) e^{\kappa \eta(t)} \le \alpha(t), \quad t \ge n_0$$

also holds, we can apply Lemma 3 (b) and (a_1).

(b) We can see from the proofs of (a_1) and (a_2) that it is enough to show that (47) holds with a $\kappa > 0$.

Assume $0 \le k(n) \le K$ for all $n \ge n_0$. Since

$$\frac{\ln\left(\frac{1}{a(n)}\right)(1 - a(n) - b(n))}{1 - a(n) + \ln\left(\frac{1}{a(n)}\right) b(n)(t - n + k(n))}$$

$$\ge \frac{\ln\left(\frac{1}{a(n)}\right)(1 - a(n) - b(n))}{1 - a(n) + \ln\left(\frac{1}{a(n)}\right) b(n)(1 + k(n))}$$

$$\ge \frac{1}{K+1} \cdot \frac{\ln\left(\frac{1}{a(n)}\right)(1 - a(n) - b(n))}{1 - a(n) + \ln\left(\frac{1}{a(n)}\right) b(n)}, \quad n \le t < n+1, \quad n \ge n_0,$$

the definitions of the functions α, β and τ in (7) and (45) insure that

$$\inf_{n \ge n_0} \frac{\alpha(t) - \beta(t)}{1 + \beta(t)\tau(t)} > 0.$$

Now by Remark 1, Corollary 1 (c) in [10] shows that there exists $\kappa > 0$ such that (47) holds.

The proof is complete. ■

4 Applications

The following result can be found in the paper [16].

Theorem A *Let $h > 0$ and let $(y(n))$ be a nonnegative sequence satisfying*

$$y(n+1) \le \frac{1}{1+a_0(n)h} y(n) + \frac{b_0(n)h}{1+a_0(n)h} \left(\max_{n-\kappa(n) \le i \le n} y(i) \right), \quad n \ge n_0, \tag{48}$$

$$y(n) = \varphi(n) \text{ for } n \in \left[n_0 - \kappa^*, n_0\right], \tag{49}$$

where $\kappa(n)$ denotes an integer valued, nonnegative and bounded sequence defined for $n \in \mathbb{Z}$ and $\kappa^* = \max_{n \in \mathbb{Z}} \kappa(n)$ is a positive integer; $\varphi(n)$ is a real valued sequence defined for $n \in [n_0 - \kappa^*, n_0]$; the parameters $a_0(n)$ and $b_0(n)$ defined for $n \in \mathbb{Z}$ denote real valued, nonnegative and bounded sequences. Suppose

$$a_0(n) - b_0(n) \ge \sigma > 0, \quad n \in \mathbb{Z}.$$

Then there exists a real number $\tilde{\lambda} > 1$ such that

$$y(n) \le \left(\max_{n_0-\kappa^* \le i \le n_0} y(i) \right) \left(\frac{1}{\tilde{\lambda}} \right)^{n-n_0}, \quad n \ge n_0.$$

From Theorems 12 and 14 we could find more general results which are better and complementary to the above theorem.

Theorem 15 *Let $h > 0$ and let $(y(n))_{n \ge n_0 - \kappa^*}$ be a nonnegative sequence satisfying (48) and (49), where $\kappa(n)$ denotes an integer valued, nonnegative sequence defined for $n \ge n_0$ such that $n - \kappa(n) \ge n_0 - \kappa^*$ with a positive integer κ^*; the parameters $a_0(n)$ and $b_0(n)$ defined for $n \in \mathbb{Z}$ denote positive sequences. Suppose that there exists a constant $0 < q < 1$ such that*

$$b_0(n) \le q a_0(n), \quad n \ge n_0 \tag{50}$$

and

$$\sup_{n \ge n_0} \sum_{i=\max(n-\kappa(n), n_0)}^{n} \frac{a_0(i) - b_0(i)}{a_0(i)} \ln(1 + a_0(i)h) < \infty. \tag{51}$$

If

$$\liminf_{n \to \infty} \frac{1}{n - n_0} \sum_{i=n_0}^{n} \ln(1 + a_0(i)h) > 0,$$

then there exists a real number $\tilde{\lambda} > 1$ such that

$$y(n) \le \left(\max_{n_0-\kappa^* \le i \le n_0} y(i) \right) \left(\frac{1}{\tilde{\lambda}} \right)^{n-n_0}, \quad n \ge n_0.$$

Proof Since (5), (6) and (33) are satisfied with

$$a(n) := \frac{1}{1 + a_0(n)h}, \quad b(n) := \frac{b_0(n)h}{1 + a_0(n)h}, \quad k(n) := \kappa(n), \quad n \geq n_0,$$

Theorem 12 (b) can be applied.

The proof is complete. ■

The next two remarks show that Theorem 15 is an essential improvement of Theorem A.

Remark 16 Suppose that the conditions of Theorem A are satisfied. In this case $0 \leq a_0(n), b_0(n) < c$ $(n \geq n_0)$, therefore

$$\frac{b_0(n)}{a_0(n)} \leq \frac{b_0(n)}{b_0(n) + \sigma} \leq \frac{c}{c + \sigma} < 1,$$

and thus (50) holds. Since

$$\sum_{i=\max(n-\kappa(n),n_0)}^{n} \frac{a_0(n) - b_0(n)}{a_0(n)} \ln(1 + a_0(n)h)$$
$$\leq \sum_{i=\max(n-\kappa(n),n_0)}^{n} \ln(1 + a_0(n)h) \leq \kappa^* \ln(1 + ch) < \infty,$$

(51) holds too.

Remark 17 Choose $h := 1$,

$$a_0(n) := \begin{cases} 1, & n = 2j, \quad j \in \mathbb{N} \\ \frac{1}{n}, & n = 2j + 1, \quad j \in \mathbb{N} \end{cases},$$

and

$$b_0(n) := \frac{1}{2n + 1}, \quad n \in \mathbb{N}.$$

Then

$$b_0(n) \leq \frac{1}{2} a_0(n), \quad n \in \mathbb{N}.$$

It is easy to check that (51) is also satisfied, if the sequence $(\kappa(n))$ is bounded. We have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n} \ln(1 + a_0(i)) = \frac{1}{2} \ln(2),$$

and hence Theorem 15 can be applied. Theorem A can not be used, since

$$a_0(2j+1) - b_0(2j+1) \to 0 \text{ as } j \to \infty.$$

The next theorem gives applicable explicit conditions for the exponential convergence of the nonnegative solutions of (48).

Theorem 18 *Let $h > 0$ and let $(y(n))_{n \ge n_0 - \kappa^*}$ be a nonnegative sequence satisfying (48) and (49), where $\kappa(n)$ denotes an integer valued, nonnegative and bounded sequence defined for $n \ge n_0$ and $\kappa^* = \max_{n \ge n_0} \kappa(n)$ is a positive integer; the parameters $a_0(n)$ and $b_0(n)$ defined for $n \in \mathbb{Z}$ denote positive sequences. Suppose that*

$$b_0(n) \le a_0(n), \quad n \ge n_0 \tag{52}$$

and

$$\inf_{n \ge n_0} \frac{\ln(1 + a_0(n)h)(a_0(n) - b_0(n))}{a_0(n) + \ln(1 + a_0(n)h)\, b_0(n)} > 0. \tag{53}$$

Then there exists a real number $\tilde{\lambda} > 1$ such that

$$y(n) \le \left(\max_{n_0 - \kappa^* \le i \le n_0} y(i) \right) \left(\frac{1}{\tilde{\lambda}} \right)^{n - n_0}, \quad n \ge n_0.$$

Proof As in the proof of the previous theorem (5) and (6) are satisfied, and

$$a(n) + b(n) = \frac{1 + b_0(n)h}{1 + a_0(n)h} \le 1, \quad n \ge n_0.$$

Now, Theorem 14 (b) can be applied.

The proof is complete. ■

Remark 19 Suppose that the conditions of Theorem A are satisfied. It is obvious that (52) holds. We show that (53) is also satisfied. If $0 \le a_0(n), b_0(n) < c \;\; (n \ge n_0)$, then

$$\begin{aligned}
\frac{\ln\left(\frac{1}{a(n)}\right)(1 - a(n) - b(n))}{1 - a(n) + \ln\left(\frac{1}{a(n)}\right) b(n)} &= \frac{\ln(1 + a_0(n)h)(a_0(n) - b_0(n))}{a_0(n) + \ln(1 + a_0(n)h)\, b_0(n)} \\
&\ge \frac{\sigma}{c} \cdot \frac{\ln(1 + a_0(n)h)}{1 + \ln(1 + a_0(n)h)} \ge \frac{\sigma}{c} \cdot \frac{\ln(1 + \sigma h + b_0(n)h)}{1 + \ln(1 + ch)} \\
&\ge \frac{\sigma}{c} \cdot \frac{\ln(1 + \sigma h)}{1 + \ln(1 + ch)} > 0, \quad n \ge n_0.
\end{aligned}$$

It can be seen that Theorem A is a special case of Theorem 18.

Remark 20 Choose

$$a_0(n) := e^{2n+1} - 1, \quad b_0(n) := e^{n+1} - 1, \quad n \geq 1.$$

Then easy to check that (52) and (53) hold, and therefore Theorem 18 can be applied. Since the sequences $(a_0(n))_{n\geq 1}$ and $(b_0(n))_{n\geq 1}$ are not bounded, Theorem A can not be applied.

Acknowledgement The authors gratefully thank to the referee for the valuable comments. The research of the authors has been supported by Hungarian National Foundations for Scientific Research Grant No. K120186.

References

1. Agarwal, R.P., Kim, Y.H., Sen, S.K.: New discrete Halanay inequalities: stability of difference equations. Commun. Appl. Anal. **12**, 83–90 (2008)
2. Agarwal, R.P., Kim, Y.H., Sen, S.K.: Advanced discrete Halanay-type inequalities: stability of difference equations. J. Inequal. Appl. 11 (2009), Art. ID 535849
3. Baker, C.T.H.: Development and application of Halanay-type theory: evolutionary differential and difference equations with time lag. J. Comput. Appl. Math. **234**, 2663–2682 (2010)
4. Baker, C.T.H., Tang, A.: Generalized Halanay inequalities for Volterra functional differential equations and discretized versions. Volterra Equations and Applications (Arlington, Tex, Stability Control Theory Methods Appl., vol. 10. Gordon and Breach, Amsterdam **2000**, 39–55 (1996)
5. Cooke, K.L., Győri, I.: Numerical approximation of the solutions of delay differential equations on an infinite interval using piecewise constant arguments. Comput. Math. Appl. **28**, 81–92 (1994)
6. Cooke, K.L., Wiener, J.: Retarded differential equations with piecewise constant delays. J. Math. Anal. Appl. **99**, 265–297 (1984)
7. Cooke, K.L., Wiener, J.: Stability regions for linear equations with piecewise continuous delay. Comput. Math. Appl. **12**, 695–701 (1986)
8. Győri, I.: On approximation of the solutions of delay differential equations by using piecewise constant arguments. Int. J. Math. Math. Sci. **14**, 111–126 (1991)
9. Győri, I., Horváth, L.: Sharp Gronwall–Bellman type integral inequalities with delay. Electron. J. Qual. Theory Differ. Equ. **111**, 1–25 (2016)
10. Győri, I., Horváth, L.: Sharp estimation for the solutions of delay differential and Halanay type inequalities. Discret. Contin. Dyn. Syst. **37**, 3211–3242 (2017)
11. Halanay, A.: Differential Equations: Stability, Oscillations, Time Lags. Academic Press, New York (1966)
12. Ivanov, A., Liz, E., Trofimchuk, S.: Halanay inequality, Yorke 3/2 stability criterion, and differential equations with maxima. Tokohu Math. **54**, 277–295 (2002)
13. Liz, E., Ferreiro, J.B.: A note on the global stability of generalized difference equations. Appl. Math. Lett. **15**, 655–659 (2002)
14. Liz, E., Ivanov, A., Ferreiro, J.B.: Discrete Halanay-type inequalities and applications. Nonlinear Anal. **55**, 669–678 (2003)
15. Liz, E., Tkachenko, V., Trofımchuk, S.: Global stability in discrete population models with delayed-density dependence. Math. Biosci. **199**, 26–37 (2006)
16. Mohamad, S., Gopalsamy, K.: Continuous and discrete Halanay-type inequalities. Bull. Aust. Math. Soc. **61**, 371–385 (2000)

17. Sepúlveda, D.: Approximation of linear differential equations with variable delay on an infinite interval. arXiv:1607.06960
18. Song, Y., Shen, Y., Yin, Q.: New discrete Halanay-type inequalities and applications. Appl. Math. Lett. **26**, 258–263 (2013)
19. Wen, L., Yu, Y., Wang, W.: Generalized Halanay inequalities for dissipativity of Volterra functional differential equations. J. Math. Anal. Appl. **347**, 169–178 (2008)
20. Xu, L.: Generalized discrete Halanay inequalities and the asymptotic behavior of nonlinear discrete systems. Bull. Korean Math. Soc. **50**, 1555–1565 (2013)

Convergence of Finite Difference Schemes Applied to the Cauchy Problems of Quasi-linear Partial Differential Equations of the Normal Form

Nobuyuki Higashimori, Hiroshi Fujiwara and Yuusuke Iso

Abstract We consider the Cauchy problems of nonlinear partial differential equations of the normal form in the class of the analytic functions. We apply semi-discrete finite difference approximation which discretizes the problems only with respect to the time variable, and we give a result about convergence. The main result shows convergence of consistent finite difference schemes even without stability, and therefore shows independence between stability and convergence for finite difference schemes. Our theoretical result can be realized numerically on multiple-precision arithmetic environments.

Keywords Finite difference method · Nonlinear PDE · Cauchy problem · Unstable scheme · Multiple-precision arithmetic

1 Introduction

We consider consistent finite difference schemes applied to the Cauchy problems of quasi-linear partial differential equations of the normal form:

$$\frac{\partial u}{\partial t} = \sum_{j=1}^{n} a_j(t, x, u)\frac{\partial u}{\partial x_j} + f(t, x, u), \qquad u(t, x)\Big|_{t=0} = 0, \tag{1}$$

N. Higashimori (✉)
Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan
e-mail: nobuyuki@acs.i.kyoto-u.ac.jp

H. Fujiwara · Y. Iso
Graduate School of Informatics,
Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan
e-mail: fujiwara@acs.i.kyoto-u.ac.jp

Y. Iso
e-mail: iso@acs.i.kyoto-u.ac.jp

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_6

where $a_j(t, x, u)$ and $f(t, x, u)$ are C^1 functions which are analytic with respect to $x = (x_1, \ldots, x_n)$ and u. Since $a_j(t, x, u)$ and $f(t, x, u)$ can be extended holomorphically with respect to x and u in some complex domain, we consider the problems for complex-valued functions $u = u(t, x)$ of a real variable t and complex variables x. If $a_j(t, x, u)$ and $f(t, x, u)$ are analytic with respect to all the variables, unique existence of an analytic solution is known as the Kowalevskaya theorem [12], and its classical proof is based on the method of power series. By the idea of reducing the Cauchy problems to those of an abstract ordinary differential equation on a scale of Banach spaces, Nirenberg [8] and Nishida [9] relaxed the analyticity assumption with respect to t to continuity with respect to t. In this case, there exists a unique solution in a class of C^1 functions of t with values in spaces of analytic functions of x. We follow their approach to show convergence of finite difference schemes applied to the Cauchy problems of nonlinear partial differential equations.

In the argument of convergence analysis, we often refer to stability analysis. We know, by the Lax equivalence theorem [7], that stability and convergence are equivalent to each other for consistent finite difference schemes of the *well-posed* Cauchy problems for *linear* partial differential equations. However, as is shown by Dahlquist [2] for the wave equation and by Hayakawa [4] for linear equations with constant coefficients, solutions to consistent but unstable finite difference schemes can be convergent in the class of the analytic functions. We remark that Dahlquist [2] relied on the Fourier analysis of the scheme, and that Hayakawa [4] relied on the method of power series, and we note that the latter is a discrete analogue of a classical proof of the Kowalevskaya theorem.

In 2011, Iso [6] developed Hayakawa's result and obtained convergence of semi-discrete finite difference schemes for the linear Cauchy problems. His method is discretization of the argument given in Yamanaka [13] and Ovsjannikov [10], in which they proved the Kowalevskaya theorem for linear equations by analyzing abstract linear Cauchy problems on a scale of Banach spaces earlier than Nirenberg [8] and Nishida [9] did for the nonlinear case. The present paper generalizes the main result of Iso [6] for nonlinear equations and is based on the results of Nirenberg-Nishida. As Dahlquist [2], Hayakawa [4], and Iso [6] did, our main result indicates that stability and convergence of finite difference schemes are independent of each other.

This paper is organized as follows. In Sect. 2, we quote the Nirenberg–Nishida theorem to confirm unique existence of the solution to (1). We give our main result in Sect. 3, an outline of its proof in Sect. 4, and a numerical example in Sect. 5. The difference scheme given in Sect. 5 is unstable, and its convergence cannot be observed by the standard double precision arithmetic. We use a multiple precision arithmetic *exflib* [3] to construct our numerical results.

2 An Abstract Cauchy-Kowalevskaya Theorem

Following [8–10, 13], we introduce a scale of Banach spaces. Let $\{B_\rho\}_{0<\rho\leq\rho_0}$ be a family of parametrized Banach spaces B_ρ with norm $\|\cdot\|_\rho$ satisfying

$$B_\rho \subset B_{\rho'} \text{ and } \|u\|_{\rho'} \leq \|u\|_\rho \text{ for } 0 < \rho' < \rho \leq \rho_0 \text{ and } u \in B_\rho. \tag{2}$$

We assume that all B_ρ are linear subspaces of a certain linear space B_0. Let $F(t, u)$ be a mapping defined on a subset of $\mathbb{R} \times B_0$ into B_0, and we consider the Cauchy problem of the form

$$du/dt = F(t, u(t)), \quad u(0) = 0. \tag{3}$$

Suppose that there exist positive numbers η, R, C, and K such that

$$\text{If } 0 < \rho' < \rho \leq \rho_0, \text{ the operator } F \text{ maps } \{t \in \mathbb{R};\ |t| < \eta\} \times \{u \in B_\rho;\ \|u\|_\rho < R\} \text{ to } B_{\rho'} \text{ continuously;} \tag{4}$$

$$\|F(t,u) - F(t,v)\|_{\rho'} \leq C\|u - v\|_\rho/(\rho - \rho') \text{ for } 0 < \rho' < \rho \leq \rho_0, \ |t| < \eta, \text{ and } u, v \in B_\rho \text{ with } \|u\|_\rho < R, \|v\|_\rho < R; \tag{5}$$

$$\|F(t,u)\|_\rho \leq K/(\rho_0 - \rho) \text{ for } 0 < \rho < \rho_0, |t| < \eta, \text{ and } u \in B_{\rho_0} \text{ with } \|u\|_{\rho_0} < R. \tag{6}$$

Our main result is to show convergence of semi-discrete difference schemes for the Cauchy problem (3), and we should remark that unique solvability of (3) is guaranteed by the following theorem due to Nirenberg [8] and Nishida [9].

Theorem 1 ([8, 9]) *Under the hypotheses* (2) *and* (4)–(6)*, there exists a positive number* $a < \eta/\rho_0$ *such that there exists a unique function* $u(t)$ *which is a solution to the Cauchy problem* (3) *in the sense that*

$$for\ every\ positive\ \rho < \rho_0,\ the\ function\ u(t)\ is\ C^1\ with\ values\ in\ B_\rho\ on\ the\ interval\ \{t \in \mathbb{R};\ |t| < a(\rho_0 - \rho)\}, \|u(t)\|_\rho < R,\ and\ satisfies\ (3)\ on\ the\ same\ interval.$$

Remark 1 The proof given in [9] shows that the solution u satisfies $\|u(t)\|_\rho \leq R/2$ if ρ and t satisfy $0 < \rho < \rho_0 - |t|/a$.

Remark 2 As is stated in Nirenberg [8], the abstract ordinary differential equation (3) contains the Cauchy problems for nonlinear partial differential equations of the normal form, and it immediately implies that our case (1) is contained. Precisely speaking, the Cauchy problem for the quasi-linear partial differential equation (1) has a unique analytic solution by Theorem 1 if there exist positive numbers η, ρ_0, and R such that $a_j(t, x, u)$ and $f(t, x, u)$ are

- C^1 functions of (t, x, u) which are bounded along with their first derivatives,
- analytic functions of (x, u) in the polydisk

$$\left(\prod_{j=1}^{n}\{x_j \in \mathbb{C};\ \left|x_j\right| < \rho_0\}\right) \times \{u \in \mathbb{C};\ |u| < R\}$$

for every $t \in \{|t| < \eta\}$.

To prove this, let $F(t, u)$ be defined by

$$F(t, u) := \sum_{j=1}^{n} a_j(t, x, u)\partial u/\partial x_j + f(t, x, u),$$

and let B_ρ be the space of all bounded holomorphic functions on the polydisk

$$D_\rho^n := \prod_{j=1}^{n}\{x_j \in \mathbb{C};\ \left|x_j\right| < \rho\}$$

with a norm $\|u\|_\rho := \sup\{|u(x)|\,;\ x \in D_\rho^n\}$. Here we remark that the family $\{B_\rho\}$ is a scale of Banach spaces and that the conditions (4)–(6) are met. Then we apply Theorem 1 to the Cauchy problem (1). Details are given in Sect. 3 of Nirenberg [8].

3 Main Result

We consider a consistent semi-discrete finite difference scheme for the Cauchy problem (3). For a number $T > 0$ and an integer $N \geq 1$, we set $\Delta t := T/N$ and $t_k := k\Delta t$ $(k = 0, \ldots, N)$, and we consider the explicit semi-discrete finite difference scheme

$$\begin{aligned} \frac{u^{k+1} - u^k}{\Delta t} &= F(t_k, u^k) \quad (k = 0, \ldots, N-1), \\ u^0 &= 0. \end{aligned}$$

This is equivalent to

$$u^k = \Delta t \sum_{j=0}^{k-1} F(t_j, u^j) \quad (k = 0, 1, \ldots, N), \tag{7}$$

where the summation equals zero for $k = 0$. We remark that a finite dimensional approximation of the operator $F(t, \cdot)$ is not considered here.

Our main result is that elements $u^0, \ldots, u^N \in B_0$ are well-defined by formula (7) and that u^k is close to $u(t_k)$ if Δt is sufficiently small. To show the result we additionally pose the following hypotheses:

For $0 < \rho' < \rho \le \rho_0$, the restriction of F to $\{t \in \mathbb{R};\ |t| < \eta\} \times \{u \in B_\rho;\ \|u\|_\rho < R\}$ is a C^1 function with values in $B_{\rho'}$; (8)

The partial Fréchet derivative of F with respect to u is a bounded linear operator from B_ρ to $B_{\rho'}$ and its norm satisfies $\|\partial_u F(t,u)\|_{\rho\to\rho'} \le C/(\rho-\rho')$ for $0 < \rho' < \rho \le \rho_0$, $|t| < \eta$, $u \in B_\rho$, $\|u\|_\rho < R$; (9)

There is a positive number L such that the partial Fréchet derivative of F with respect to t satisfies $\|\partial_t F(t,u)\|_{\rho'} \le L/(\rho-\rho')^2$ for $0 < \rho' < \rho \le \rho_0$, $|t| < \eta$, $u \in B_\rho$, $\|u\|_\rho < R$. (10)

The hypotheses above imply that the solution $u(t)$ to (3) is a C^2 function of t. Precisely, the next proposition holds.

Proposition 1 *Suppose* (2)*,* (6)*, and* (8)–(10)*. Let $u(t)$ be the unique solution to* (3) *as stated in Theorem 1. Then $u(t)$ is a C^2 function from $\{t \in \mathbb{R};\ |t| < a(\rho_0-\rho)\}$ to B_ρ for every $0 < \rho < \rho_0$. Moreover, there is a number $V > 0$ such that*

$$\|u''(t)\|_\rho \le V/(\rho_0 - \rho - |t|/a)^2 \tag{11}$$

for every ρ and t with $0 < \rho < \rho_0 - |t|/a$.

Proof (Outline) Suppose $0 < \rho < \rho_0 - |t|/a$, and let $\rho(t) := (\rho + \rho_0 - |t|/a)/2$. Then we have $0 < \rho < \rho(t) < \rho_0 - |t|/a$, and $\|u(t)\|_{\rho(t)} \le R/2$ by *Remark* 1. By using (9) and (6), we have

$$\|u'(t)\|_\rho \le \|F(t,u(t)) - F(t,0)\|_\rho + \|F(t,0)\|_\rho \le \frac{C\|u(t)\|_{\rho(t)}}{\rho(t)-\rho} + \frac{K}{\rho_0-\rho},$$

and thus, for some $V_1 > 0$,

$$\|u'(t)\|_\rho \le V_1/(\rho_0 - \rho - |t|/a), \quad 0 < \rho < \rho_0 - |t|/a. \tag{12}$$

By differentiating both sides of $u' = F(t,u)$ and using (9), (10), and (12), we can obtain (11) for some $V > 0$. □

Note that (8) implies (4) and (9) implies (5), and that by Theorem 1 the unique solution to (3) exists. Our main result is the following theorem.

Theorem 2 *Suppose* (2), (6)*, and* (8)–(10)*. Let* $u(t)$ *be the unique solution to* (3) *as in Theorem 1. Let* c*,* r_0*, and* T *be real numbers satisfying*

$$0 < c < \min\{a, 1/4C\}, \quad 0 < r_0 < \rho_0, \quad 0 < T < cr_0. \tag{13}$$

Then there exists a number $S > 0$ *such that if* $S\Delta t < R/2$*, Eq.* (7) *determines* $u^0, \ldots, u^N \in B_0$*, and they satisfy*

$$\sup_{0 \le t_k \le c(r_0 - \rho)} \|u^k - u(t_k)\|_\rho \le S\Delta t \tag{14}$$

for every positive $\rho < r_0$*.*

Proof (Outline) We introduce truncation errors w_k $(0 \le k \le N-1)$ by

$$\frac{u(t_{k+1}) - u(t_k)}{\Delta t} = F(t_k, u(t_k)) + w_k. \tag{15}$$

By Taylor's theorem, there are $v_0, \ldots, v_{N-1} \in B_0$ such that

$$u(t_{k+1}) - u(t_k) = F(t_k, u(t_k))\Delta t + v_k \Delta t^2, \quad 0 \le k \le N-1,$$

hence $w_k = v_k \Delta t$. For each k, $0 \le k \le N-1$, v_k satisfies

$$\|v_k\|_\rho \le V/(\rho_0 - \rho - t_{k+1}/a)^2, \quad 0 < \rho < \rho_0 - t_{k+1}/a, \tag{16}$$

by Proposition 1. The theorem immediately follows from the next lemma.

Lemma 1 *Under the hypotheses of Theorem 2, there exists a number* $S > 0$ *such that, if* $S\Delta t < R/2$*, the formula*

$$e_k = \Delta t \sum_{j=0}^{k-1} \left\{ F(t_j, u(t_j) + e_j) - F(t_j, u(t_j)) - w_j \right\}, \quad 0 \le k \le N, \tag{17}$$

determines elements $e_0, \ldots, e_N \in B_0$*, and for every* k*,* $0 \le k \le N$*, we have*

$$\|e_k\|_\rho \le S\Delta t, \quad 0 < \rho \le r_0 - t_k/c. \tag{18}$$

Assuming Lemma 1 for the moment we continue the proof of Theorem 2. Suppose that $S\Delta t < R/2$. Then (17) determines $e_0, e_1, \ldots, e_N$, and (18) holds. Put $u^k := u(t_k) + e_k$. Then (15) and (17) yield (7), and (18) yields (14). The proof will be completed by proving Lemma 1. □

4 Outline of Proof of Lemma 1

The detail will be found in [5]. Suppose that c, r_0, and T satisfy (13). We introduce a family $\{Y_\alpha\}_{\alpha \ge c}$ of linear subspaces of $(B_0)^{N+1}$ defined by

$$Y_\alpha := \{f = (f_0, \dots, f_N) \in (B_0)^{N+1};\ [f]_\alpha < +\infty\}, \text{ where}$$
$$[f]_\alpha := \inf \left\{ M \ge 0;\ \|f_k\|_\rho \le \frac{M t_k}{r_0 - \rho - t_k/\alpha} \text{ for } 0 \le k \le N, 0 < \rho < r_0 - t_k/\alpha \right\}.$$

The following propositions hold:

- $f_0 = 0$ for all $f = (f_0, \dots, f_N) \in Y_\alpha$;
- Y_α is a Banach space with respect to the norm $[\,\cdot\,]_\alpha$;
- If $c \le \beta \le \alpha$, then $[f]_\beta \le [f]_\alpha$ for all $f \in Y_\alpha$ and therefore $Y_\alpha \subset Y_\beta \subset Y_c$;
- If $c \le \beta < \alpha$, $0 \le k \le N$, and $0 < \rho \le r_0 - t_k/\beta$, then

$$f_k \in B_\rho \quad \text{and} \quad \|f_k\|_\rho \le \frac{[f]_\alpha}{1/\beta - 1/\alpha} \quad \text{for all } f \in Y_\alpha. \tag{19}$$

We want to define a sequence $e^{(m)} = \big(0, e_1^{(m)}, \dots, e_N^{(m)}\big)$, $m = 0, 1, 2, \dots$, in Y_c by the following rule. Let $e^{(0)} := (0, 0, \dots, 0)$. If $e^{(m)}$ is defined for some $m \ge 0$, the next term $e^{(m+1)}$ has the components

$$e_k^{(m+1)} := \Delta t \sum_{j=0}^{k-1} \left\{ F\big(t_j, u(t_j) + e_j^{(m)}\big) - F(t_j, u(t_j)) - w_j \right\}, \quad 0 \le k \le N. \tag{20}$$

The goal is to show that the above rule defines a convergent sequence in Y_c and that the limit is a solution to (17).

(Step 1) Fix k and ρ with $1 \le k \le N$ and $0 < \rho < r_0 - t_k/a$. By (16), we have

$$\|v_j\|_\rho \le \frac{V}{(\rho_0 - \rho - t_{j+1}/a)^2} \le \frac{V}{(\rho_0 - r_0)(r_0 - \rho - t_k/a)}, \quad 0 \le j \le k-1.$$

Since $w_j = v_j \Delta t$, we get

$$\|e_k^{(1)}\|_\rho \le \Delta t \sum_{j=0}^{k-1} \|w_j\|_\rho \le \frac{(V'\Delta t) t_k}{r_0 - \rho - t_k/a}$$

with $V' = V/(\rho_0 - r_0)$. This shows that $e^{(1)} \in B_a$ and $[e^{(1)}]_a \le V'\Delta t$.

(Step 2) Take b_0 and c_0 with $c < c_0 < b_0 < \min\{a, 1/4C\}$ and put $\delta := 1 - c_0/b_0$. Then $b_0 = c_0/(1-\delta)$ and $0 < \delta < 1$. Take θ with $0 < \theta < 1$ and $4Cb_0 < \theta^2$ and put $b_m := c_0/(1 - \delta\theta^m)$. Note that

$$c < c_0 < \cdots < b_2 < b_1 < b_0 < a,$$
$$[\,\cdot\,]_c \le [\,\cdot\,]_{c_0} \le \cdots \le [\,\cdot\,]_{b_2} \le [\,\cdot\,]_{b_1} \le [\,\cdot\,]_{b_0} \le [\,\cdot\,]_a,$$
$$Y_a \subset Y_{b_0} \subset Y_{b_1} \subset Y_{b_2} \subset \cdots \subset Y_{c_0} \subset Y_c.$$

It follows from Step 1 that $e^{(1)} \in Y_{b_0}$ and $[e^{(1)}]_{b_0} \le V'\Delta t$. By (19), if $0 \le k \le N$ and $0 < \rho \le r_0 - t_k/b_1$, we have

$$e_k^{(1)} \in B_\rho, \quad \|e_k^{(1)}\|_\rho \le \frac{b_0 c_0 V'}{(b_0 - c_0)(1-\theta)} \Delta t. \tag{21}$$

(Step 3) Let $S := b_0 c_0 V'/(b_0 - c_0)(1-\theta)^2$, and suppose that $S\Delta t < R/2$. The aim here is to show that the following assertions hold for all integers $m \ge 0$:

$(1)_m$ $\quad e^{(0)}, \ldots, e^{(m+1)}$ are defined by (20) and belong to Y_{b_m};

$(2)_m$ $\quad [e^{(m+1)} - e^{(m)}]_{b_m} \le \theta^{2m} V' \Delta t$;

$(3)_m$ $\quad \|e_k^{(m+1)}\|_\rho \le (1-\theta^{m+1}) S\Delta t \quad$ for $0 \le k \le N, 0 < \rho \le r_0 - t_k/b_{m+1}$.

They are valid for $m = 0$ by Step 2. Next, suppose $(1)_m$, $(2)_m$, $(3)_m$ for some $m \ge 0$. The inductive step consists of three parts (A)–(C) below.

(A) We put $\lambda_k := r_0 - t_k/b_{m+1}$ for $1 \le k \le N$. The purpose is to show that

$$e_k^{(m+2)} := \Delta t \sum_{j=0}^{k-1} \Big\{ F\big(t_j, u(t_j) + e_j^{(m+1)}\big) - F(t_j, u(t_j)) - w_j \Big\} \in B_{\lambda_k}$$

On the right side, the sum of the terms $-w_j \Delta t$ equals $e_k^{(1)}$, which belongs to B_{λ_k} by (21). If $j < k$, we obtain $\|u(t_j) + e_j^{(m+1)}\|_{\lambda_j} < R$ from $(3)_m$ and *Remark* 1. Since $\lambda_k < \lambda_j$, assumption (8) implies $F\big(t_j, u(t_j) + e_j^{(m+1)}\big) \in B_{\lambda_k}$ and $F(t_j, u(t_j)) \in B_{\lambda_k}$. Hence $e_k^{(m+2)} \in B_{\lambda_k}$.

(B) Let $e^{(m+2)} := \big(0, e_1^{(m+2)}, \ldots, e_N^{(m+2)}\big)$. Suppose that k and ρ satisfy $1 \le k \le N$ and $0 < \rho < r_0 - t_k/b_{m+1}$, and put $\rho_j' := (\rho + r_0 - t_j/b_{m+1})/2$ for $0 \le j \le k-1$. Then we can estimate the difference $e_k^{(m+2)} - e_k^{(m+1)}$ as follows:

$$\Big\| e_k^{(m+2)} - e_k^{(m+1)} \Big\|_\rho \le \Delta t \sum_{j=0}^{k-1} \Big\| F\big(t_j, u(t_j) + e_j^{(m+1)}\big) - F\big(t_j, u(t_j) + e_j^{(m)}\big) \Big\|_\rho$$
$$\le \Delta t \sum_{j=0}^{k-1} \frac{C}{\rho_j' - \rho} \Big\| e_j^{(m+1)} - e_j^{(m)} \Big\|_{\rho_j'} \le \Delta t \sum_{j=0}^{k-1} \frac{C}{\rho_j' - \rho} \, \frac{[e^{(m+1)} - e^{(m)}]_{b_{m+1}} t_j}{r_0 - \rho_j' - t_j/b_{m+1}}$$
$$\le \frac{4Cb_{m+1}[e^{(m+1)} - e^{(m)}]_{b_{m+1}} t_k}{r_0 - \rho - t_k/b_{m+1}} \le \frac{\theta^2 [e^{(m+1)} - e^{(m)}]_{b_m} t_k}{r_0 - \rho - t_k/b_{m+1}}.$$

Thus we find

$$e^{(m+2)} - e^{(m+1)} \in Y_{b_{m+1}}, \tag{22}$$
$$[e^{(m+2)} - e^{(m+1)}]_{b_{m+1}} \le \theta^2 [e^{(m+1)} - e^{(m)}]_{b_m}, \tag{23}$$

so $(1)_{m+1}$ follows from (22) and $(1)_m$, and $(2)_{m+1}$ follows from (23) and $(2)_m$.

(C) If $0 \le k \le N$ and $0 < \rho \le r_0 - t_k/b_{m+2}$, we use (19) and $(2)_{m+1}$ to get

$$\left\| e_k^{(m+2)} - e_k^{(m+1)} \right\|_\rho \le \theta^{m+1}(1-\theta) S \Delta t.$$

Combining this with $(3)_m$, we obtain $(3)_{m+1}$. Thus we finish the inductive step, and assertions $(1)_m$, $(2)_m$, $(3)_m$ hold for all $m \ge 0$.

(Step 4) By Step 3, the sequence $\{e^{(m)}\}_{m\ge 0}$ is Cauchy in the Banach space Y_{c_0}. Let $e = (e_0, \dots, e_N)$ denote the limit. Then, for k and ρ satisfying $0 \le k \le N$ and $0 < \rho \le r_0 - t_k/c$, we find that $\|e_k^{(m)} - e_k\|_\rho \to 0$ $(m \to \infty)$. Finally, we take the limit of (20) and $(3)_m$ to obtain (17) and (18).

5 Numerical Example

Finally, we illustrate Theorem 2 by numerical solution of the Cauchy problem

$$u_t(t,x) + \{u(t,x) + g(x)\}\{u_x(t,x) + g'(x)\} = 0, \quad t > 0,\ x \in \mathbb{R}, \tag{24}$$
$$u(0,x) = 0, \quad x \in \mathbb{R},$$

where $g(x) = \sin(\pi x)$. We remark that by putting $v(t,x) = u(t,x) + g(x)$ the problem above is equivalent to the Cauchy problem for the inviscid Burgers equation

$$v_t(t,x) + v(t,x) v_x(t,x) = 0, \quad t > 0,\ x \in \mathbb{R}, \tag{25}$$
$$v(0,x) = g(x), \quad x \in \mathbb{R}. \tag{26}$$

To apply Theorem 2, we consider Eq. (24) for $t > 0$, $x \in \mathbb{C}$. We take B_ρ as the space of bounded holomorphic functions in $D_\rho := \{x \in \mathbb{C};\ |x| < \rho\}$ with sup norm, and put $F(u) := -(u+g)(u'+g')$, which is a nonlinear differential operator acting on $u = u(x)$. Then the family $\{B_\rho\}_{0<\rho\le\rho_0}$ satisfies (2), and Eq. (24) can be written in the form (3). For each $\rho > 0$, the normed space B_ρ is a Banach space since the uniform limit of holomorphic functions is holomorphic (for a proof, see [1, p.176, Theorem 1]). To verify (9), we note the estimate for the first derivative:

$$\|u'\|_{\rho'} \le \|u\|_\rho / (\rho - \rho'), \quad u \in B_{\rho'}, \quad 0 < \rho' < \rho. \tag{27}$$

Indeed, if $x_0 \in D_{\rho'}$, then the open disk $\Delta := \{\zeta \in \mathbb{C}; |\zeta - x_0| < \rho - \rho'\}$ lies in D_ρ along with the boundary $\partial\Delta$. Cauchy's integral formula states that

$$u(x) = \frac{1}{2\pi i}\int_{\partial\Delta}\frac{u(\zeta)}{\zeta - x}\,d\zeta, \quad x \in \Delta,$$

and differentiation with respect to x gives

$$u'(x) = \frac{1}{2\pi i}\int_{\partial\Delta}\frac{u(\zeta)}{(\zeta - x)^2}\,d\zeta, \quad x \in \Delta.$$

By putting $x = x_0$ we obtain

$$\left|u'(x_0)\right| \le \frac{1}{2\pi}\int_{\partial\Delta}\frac{|u(\zeta)|}{|\zeta - x_0|^2}\,|d\zeta| \le \frac{\|u\|_\rho}{\rho - \rho'}.$$

Since x_0 is arbitrary in $D_{\rho'}$, the estimate (27) follows. Now suppose that $u \in B_\rho$, $\|u\|_\rho < R$, and $0 < \rho' < \rho \le \rho_0$. For every $h \in B_\rho$ we have

$$F(u + h) - F(u) = T_u h + hh',$$

where T_u is the linear operator given by $T_u h := -(u + g)h' - h(u' + g')$. Then it follows from (27) that

$$\|F(u + h) - F(u) - T_u h\|_{\rho'} = o(\|h\|_\rho),$$
$$\|T_u h\|_{\rho'} \le C\|h\|_\rho/(\rho - \rho'), \quad h \in B_\rho,$$

where C is a positive number depending on R and $\|g\|_{\rho_0}$. Thus we see that the mapping $F : B_\rho \to B_{\rho'}$ is Fréchet differentiable at u, and that the Fréchet derivative T_u satisfies $\|T_u\|_{\rho\to\rho'} \le C/(\rho - \rho')$, hence (9) holds. The remaining hypotheses of Theorem 2 can be verified similarly. Hence the convergence of the semi-discrete scheme

$$\frac{u^{k+1}(x) - u^k(x)}{\Delta t} + \{u^k(x) + g(x)\}\{(u^k)'(x) + g'(x)\} = 0, \quad k \ge 0,\ x \in \mathbb{R},$$
$$u^0(x) = 0, \quad x \in \mathbb{R}.$$

in the sup norm follows.

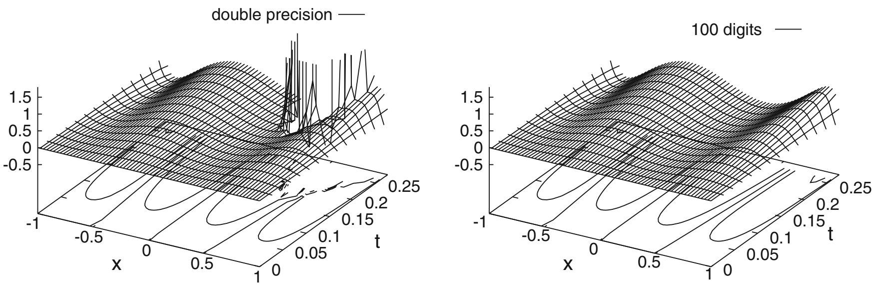
In our numerical computation, we discretize Eq. (24) by the forward difference in the t direction and also in the x direction. For positive numbers Δt and Δx, our finite difference scheme is

$$\frac{u^{k+1}(x) - u^k(x)}{\Delta t} + \{u^k(x) + g(x)\} \left(\frac{u^k(x + \Delta x) - u^k(x)}{\Delta x} + g'(x) \right) = 0, \quad (28)$$
$$k \geq 0,\ x \in \mathbb{R},$$
$$u^0(x) = 0, \quad x \in \mathbb{R}.$$

Since $g(x)$ takes both signs, there are points (t, x) whose domain of dependence for the scheme (28) does not contain the characteristic line for Eq. (24), and thus the convergence of the scheme cannot be proved in the framework of continuous functions on the real line.

Figures 1a, b show the results of numerical computation with double precision arithmetic and 100 decimal digits arithmetic on *exflib* [3], respectively, where the discretization parameters are $\Delta t = \Delta x = 0.005$. The result by double precision in Fig. 1a shows oscillation around $t = 0.2$ and $x = 0.5$, while the result by 100 decimal digits in Fig. 1b does not. This indicates that the oscillation comes from instability of the scheme and rounding errors.

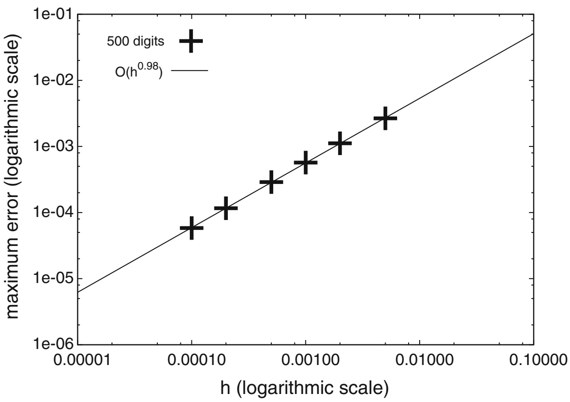
Figure 2 shows convergence of the scheme. The horizontal axis is $h = \Delta t = \Delta x$ and the vertical axis is the maximum error



(a) Profile computed with double precision (b) Profile computed with 100 decimal digits

Fig. 1 Profiles of numerical solution of the finite difference equation (28) with $\Delta t = \Delta x = 0.005$

Fig. 2 Rate of convergence by 500 decimal digits precision



$$\sup_{0<t_k<0.1,\ x\in\mathbb{R}} \left|u^k(x) - u(t_k, x)\right|.$$

The values $u(t_k, x)$ of the exact solution are given by $u(t_k, x) = v(t_k, x) - g(x)$, where v is the solution to the Cauchy problem (25)–(26). The solution v is constant on the characteristic lines for Eq. (25). Hence we can find the value $v(t_k, x) = v(0, x_0) = g(x_0)$ if (t_k, x) lies on the characteristic line $x = x_0 + g(x_0)t$ through $(0, x_0)$. We solved the equation $x = x_0 + g(x_0)t_k$ for x_0 by Newton's method. The line in Fig. 2 suggests that the maximum error is of order $O(h^{0.98})$ for h between 0.0001 and 0.01, which is almost consistent with the main result.

Acknowledgements This work is partially supported by JSPS Grant-in-Aid for scientific research (B) 25287028.

References

1. Ahlfors, L.V.: Complex Analysis, 3rd edn. McGraw-Hill, New York (1979)
2. Dahlquist, G.: Convergence and stability for a hyperbolic difference equation with analytic initial-values. Math. Scand. **2**, 91–102 (1954)
3. Fujiwara, H.: http://www-an.acs.i.kyoto-u.ac.jp/~fujiwara/exflib
4. Hayakawa, K.: Convergence of finite difference scheme and analytic data. Publ. Res. Inst. Math. Sci. **24**, 759–764 (1988)
5. Higashimori, N., Fujiwara, H.: Semi-discrete finite difference schemes for the nonlinear Cauchy problems of the normal form. (submitted to Proc. Japan Acad. Ser. A)
6. Iso, Y.: Convergence of a semi-discrete finite difference scheme applied to the abstract Cauchy problem on a scale of Banach spaces. Proc. Jpn. Acad. Ser. A **87**, 109–113 (2011)
7. Lax, P.D., Richtmyer, R.D.: Survey of the stability of linear finite difference equations. Commun. Pure Appl. Math. **9**, 267–293 (1956)
8. Nirenberg, L.: An abstract form of the nonlinear Cauchy-Kowalewski theorem. J. Differ. Geom. **6**, 561–576 (1972)
9. Nishida, T.: A note on a theorem of Nirenberg. J. Differ. Geom. **12**, 629–633 (1977)
10. Ovsjannikov, L.V.: Singular operator in the scale of Banach spaces. Dokl. Akad. Nauk SSSR **163-4**, 819–822 (1965) (Eng. Trans. Soviet Math. Dokl. **6**, 1025–1028 (1965))
11. Petrovsky, I.G.: Lectures on Partial Differential Equations. Dover, New York (1991) (Interscience, New York, (1954))
12. von Kowalevsky, S.: Zur Theorie der partiellen Differentialgleichungen. J.für die reine angew. Math. **80**, 1–32 (1875)
13. Yamanaka, T.: Note on Kowalevskaja's system of partial differential equations. Commentationes Math. Univ. St. Paul **9**, 7–10 (1961)

Operator Theoretic Phenomena of the Markov Operators which are Induced by Stochastic Difference Equations

Takashi Honda and Yukiko Iwata

Abstract We show the relation between the Jacobs-de Leeuw–Glicksberg decomposition of semigroups and the spectral decomposition of the Markov operators which are induced by stochastic difference equations by using our new results.

Keywords Stochastic difference equations · Spectral theory · Markov semigroups

1 Introduction

Density functions of a Markov process under some conditions are represented by a Markov operator. For example, a stochastic process $\{X_n\}_{n\geq 0}$ defined by

$$X_{n+1} = S(X_n) + Y_n$$

is a Markov process, where $S : \mathbb{R} \to \mathbb{R}$ is a dynamical system, $Y_0, Y_1, \cdots$ are independent random variables with values in $\mathbb{R}$ each having the same density g, and X_0 and $\{Y_n\}_{n\geq 0}$ are independent. Let f_n be the density function of X_n for each $n \geq 0$, and hence we have

$$f_{n+1}(x) = \int_{\mathbb{R}} f_n(y) g(x - S(y)) \mu(dy),$$

where μ is the Lebesgue measure on $\mathbb{R}$. This equation means that every density function f_n is represented by a Markov operator $T : L^1(\mathbb{R}) \to L^1(\mathbb{R})$ defined by

T. Honda (✉)
Faculty of Education, Department of Mathematics, Iwate University, Morioka 020-8550, Japan
e-mail: thonda7@iwate-u.ac.jp

Y. Iwata
Faculty of Liberal Arts, Department of Information Science, Tohoku Gakuin University, Sendai, Miyagi 981-3193, Japan
e-mail: iwata@mail.tohoku-gakuin.ac.jp

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_7

$$Tf(x) = \int_{\mathbb{R}} f(y)g(x - S(y))\mu(dy)$$

as $f_n = T^n f_0$.

In this paper, we are interested in asymptotic behavior of Markov operators on Banach spaces. Especially, we mainly interest some relation between the Jacobs-de Leeuw–Glicksberg decomposition of semigroups and the existence of a constrictor which attracts densities of the Markov operators on a Banach space, which are induced by stochastic difference equations. Jacobs [7] first obtained this splitting theorem under the reflexivity assumption. De Leeuw and Glicksberg [2] showed Theorem 1 in Sect. 2 and they also showed the similar splitting theorem for a non-abelian semigroup of linear contractions in a strictly convex Banach space with the strictly convex dual space.

The Jacobs-de Leeuw–Glicksberg decomposition holds for a cmplex Banach space. First, we shall show some relations the existence of a constrictor of a linear contractive operator between (continuous or discrete) semigroups of linear contractive operators in a cmplex Banach space by using recent results [5].

Second, we will consider a Markov operator T on a real $L^1(\Omega, \Sigma, \mu)$ space, wherer (Ω, Σ, μ) is a probability measure space. In general, Sine prove that if a Markov operator T on a real L^1 space is constrictive (see Definition 1 in Sect. 2), then the Jacobs-de Leeuw–Glicksberg decomposition and the spectral decomposition holds for the discrete semigroup $(T^n)_{n=1}^{\infty}$([12]). Moreover, the author gave a necessary and sufficient condition for a constrictive Markov operator T defined on a real L^1 space when a Markov operator T is an integral operator T with a stochastic kernel and satisfies some conditions (Iwata, [6]). One of our main results is Theorem 8 in which we gave another sufficient condition for a constrictive Markov operator T defined on a real L^p space $(1 \leq p < \infty)$.

2 Preliminares

We denote by $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{C}$ the sets of all positive integers, all real numbers and all complex numbers, respectively. For any scalar α, $\overline{\alpha}$ and $\mathrm{Re}\,(\alpha)$ are the complex conjugate and the real part of α. Let $(E, \|\cdot\|)$ be a real or complex Banach space with the dual space $(E^*, \|\cdot\|)$. We denote by $\langle x, x^*\rangle$ the dual pair of $x \in E$ and $x^* \in E^*$, i.e., $\langle x, x^*\rangle = x^*(x)$. For any scalar α, we have $\langle \alpha x, x^*\rangle = \langle x, \alpha x^*\rangle = \alpha\langle x, x^*\rangle$. For a subset $A \subset E$, w-cl$A$ is the closure of A in the weak topology.

A Banach space E is said to be *strictly convex* if $\|x + y\| < 2$ for $x, y \in E$ with $\|x\| = 1$, $\|y\| = 1$ and $x \neq y$.

With each $x \in E$, we associate the set

$$Jx = \{x^* \in E^* : \langle x, x^*\rangle = \|x\|^2 = \|x^*\|^2\}.$$

The multivalued operator $J : E \to E^*$ is called the *normalized duality mapping* of E. From the Hahn–Banach theorem, for every $x \in E \setminus \{0\}$, there is $x^* \in E^*$, $\|x^*\| = 1$ such that $\langle x, x^* \rangle = \|x\|$. Then, $Jx \neq \emptyset$ for each $x \in E$. If the dual space E^* is strictly convex, J is single-valued. Indeed, if $x_1^*, x_2^* \in J(x)$, then $\|x\| \|\frac{x_1^*+x_2^*}{2}\| \geq \langle x, \frac{x_1^*+x_2^*}{2} \rangle = \|x_1^*\|^2 = \|x_2^*\|^2 = \|x\|^2$. Since E^* is strictly convex, $x_1^* = x_2^*$. If E is reflexive, then J is a mapping of E onto E^*. Suppose E is reflexive, from the Hahn–Banach theorem, for every $x^* \in E^* \setminus \{0\}$, there is $x \in E$, $\|x\| = 1$ such that $\langle x, x^* \rangle = \|x^*\|$. Then, x^* must be an element of $J(\|x^*\|x)$. When E is a reflexive and strictly convex space with the strictly convex dual space, J is a single-valued, one-to-one and onto mapping. Then, we can define the single-valued mapping $J^{-1} : E^* \to E$ and we have $J^{-1} = J_*$, where J_* is the normalized duality mapping of E^*. When J is single-valued, we have $J(\alpha x) = \overline{\alpha} Jx$ for any scalar α. Indeed, $\langle \alpha x, \overline{\alpha} Jx \rangle = \alpha \overline{\alpha} \langle x, Jx \rangle = \|\alpha x\|^2 = \|\overline{\alpha} Jx\|^2$. See [1, 13] for more details.

Let A be a nonempty subset of a Banach space E and let A^* be a nonempty subset of the dual space E^*. We denote by spn$A$ and $\overline{\text{spn}}A$ the linear span and the closed linear span of A respectively. We define the *annihilator* $A_\perp^*$ of Y^* and the *annihilator* $A^\perp$ of Y as follows:

$$A_\perp^* = \{x \in E : \langle x, x^* \rangle = 0 \text{ for all } x^* \in A^*\}$$

and

$$A^\perp = \{x^* \in E^* : \langle x, x^* \rangle = 0 \text{ for all } x \in A\}.$$

Both subsets are closed linear subspaces of E and E^*, respectively. In a reflexive Banach space E, $A_\perp = A^\perp$ for $A \subset E = E^{**}$. If $A \subset B \subset E$ and $A^* \subset B^* \subset E^*$, then $B^\perp \subset A^\perp$ and $B_\perp^* \subset A_\perp^*$.

A mapping $T : E \to E$ in a Banach space E is called *nonexpansive* if it satisfies

$$\|Tx - Ty\| \leq \|x - y\|$$

for any $x, y \in E$. We call a linear nonexpansive mapping a *linear contaractive operator* and it is a bounded linear operator $T : E \to E$ such that $\|T\| \leq 1$. For a bounded linear operator $T : E \to E$, the *dual* operator of T is the operator $T^* : E^* \to E^*$ with $\langle Tx, x^* \rangle = \langle x, T^*x^* \rangle$ for $x \in E, x^* \in E^*$. We know that for any bounded linear operators S, T in E and scalar α, $\|T\| = \|T^*\|$, $(S + T)^* = S^* + T^*$ and $(\alpha T)^* = \alpha T^*$. For any mapping T in E, we denote the set of all fixed points of T by $F(T)$. A linear operator $P : E \to E$ is called *(linear) projection* if $P = P^2$. For a projection P of E onto $M \subset E$, we have $F(P) = M$. Indeed, if $x \in M$, there exists $y \in E$ such that $x = Py$. Then, $Px = P^2y = Py = x$ and $x \in F(P)$. Let I be the identity operator of E. If a linear operator $Q : E \to E$ satisfies that $Q^{-1}(0) = (I - Q)\,E$, then Q is a projection. Indeed, for any $x \in E$ we have $Qx = Q\,(Qx + (I - Q)x) = Q^2x$. See [11] for more details.

A nonempty set $\mathscr{S}$ of mappings in E is called a *semigroup* if

$$T, S \in \mathscr{S} \Rightarrow T \circ S \in \mathscr{S}$$

for all $T, S \in \mathscr{S}$. A semigroup $\mathscr{S}$ is called *abelian* if $T \circ S = S \circ T$ for all T, $S \in \mathscr{S}$. A sub-semigroup $\mathscr{J} \subset \mathscr{S}$ is called *ideal* if $\mathscr{J} \circ \mathscr{S} \subset \mathscr{J}$ and $\mathscr{S} \circ \mathscr{J} \subset \mathscr{J}$. The intersection of all ideals of $\mathscr{S}$ is called the *kernel* of $\mathscr{S}$ and denoted by $\mathscr{K}$; see [2]. In this case when $\mathscr{S}$ is a semigroup of bounded linear operators in E, we denote the set which consists of all of the dual operators T^* of $T \in \mathscr{S}$ by $\mathscr{S}^*$. The set $\mathscr{S}^*$ is a semigroup and we call it the *dual semigroup* of $\mathscr{S}$. Indeed, $T^* \circ S^* = (S \circ T)^*$ for any bounded linear operators T, S. A net $\{T_\alpha\}$ of bounded linear operators in E converges to T in the *weak operator topology* if and only if $T_\alpha x \rightharpoonup Tx$ weakly for all $x \in E$. A semigroup $\mathscr{S}$ of bounded linear operators in E is called *weakly almost periodic* if for any $x \in E$ the orbit $\mathscr{S}x = \{Tx \in E : T \in \mathscr{S}\}$ is conditionally weakly compact. If E is reflexive, any semigroup of linear contractive operator in E is weakly almost periodic. See [8] for more details.

Especially, we call operator semigroups indexed by non-negative integers or non-negative reals *one-parameter semigroup*. It is easy to see that any one-parameter semigroup is abelian. If a semigroup indexed by positive real numbers $\mathbb{R}^+$, we always assume that it is strongly continuous, that is

$$\lim_{t \to 0} \|T_{s+t}x - T_s x\| = 0 \qquad \forall s \geq 0, \quad x \in E.$$

We shall use the notation $(T_t)_{t \geq 0}$ for a one-parameter semigroup in the continuous parameter case, and $(T^n)_{n=1}^{\infty}$ for the discrete semigroup, generated by a single operator T. In this paper, we also use the notation $\mathscr{T} = (T_t)_{t \in J}$, where $J = \mathbb{R}^+$ or $J = \mathbb{N} \cup \{0\}$ for any one-parameter semigroup.

Let E be a real or complex Banach space and $\mathscr{T}$ be a weakly almost periodic semigroup of bounded operators on E. If E can be decomposed into the direct sum

$$E = E_{fl}(\mathscr{T}) \oplus E_{rev}(\mathscr{T})$$

with respect to $\mathscr{T}$, where

$$\begin{aligned} &E_{fl}(\mathscr{T}) := \{x \in E : \ 0 \in \text{w-cl}\{Tx\}_{T \in \mathscr{T}}\}, \quad \text{and} \\ &E_{rev}(\mathscr{T}) := \{x \in E : \ y \in \text{w-cl}\{Tx\}_{T \in \mathscr{T}} \ \Rightarrow \ x \in \text{w-cl}\{Ty\}_{T \in \mathscr{T}}\}, \end{aligned}$$

then we call this decomposition the Jacobs-deLeeuw–Glicksberg decomposition. This decomposition plays a very important role in our paper.

de Leeuw and Glicksberg proved the following decomposition theorem in 1961.

Theorem 1 *Let $\mathscr{T}$ be an abelian weakly almost periodic semigroup of bounded operators on a complex Banach space E and let Q be the unit in the kernel $\mathscr{K}$ of the closure of $\mathscr{T}$ in the weak operator topology. Then $E_{rev}(\mathscr{T}) = E_{uds}(\mathscr{T}) = QE$ and $E_{fl}(\mathscr{T}) = Q^{-1}(0) = (I - Q)E$. In particular, E is the direct sum of the closed invariant subspaces $E_{rev}(\mathscr{T})$ and $E_{fl}(\mathscr{T})$, i.e. $E = E_{fl}(\mathscr{T}) \oplus E_{rev}(\mathscr{T})$.*

See [3, 8] for more details.

If $\mathscr{T}$ is a one-parameter bounded semigroup of bounded operators on a complex Banach space E, Sine proved the following result (See Theorem 1.3.3, [3]).

Theorem 2 (Sine[12]) *Given a (continuous or descrete) one-parameter bounded semigroup $\mathscr{T}$ of bounded operators on a complex Banach space E, the following assertions are equivalent:*

1. *there is a compact subset $A \subset E$ such that*

$$\lim_{t\to\infty} \inf_{y\in A} \|T_t x - y\| = 0 \qquad \textit{for each } x \in B(E),$$

where $B(E)$ is the closed unit ball of E.

2. *there exists a $\mathscr{T}$-reducing decomposition $E = E_{fl}(\mathscr{T}) \oplus E_{rev}(\mathscr{T})$ with*
   - $E_{fl}(\mathscr{T}) = \{x \in E : \lim_{t\to\infty} \|T_t x\| = 0\}$
   - $E_{rev}(\mathscr{T}) = E_{uds}(\mathscr{T}), \quad \dim(E_{rev}(\mathscr{T})) < \infty$

Furthermore, Sine proved the following theorem for a linear contraction T on a real L^1 space (See Theorem 2, [12]).

Theorem 3 *Suppose T is a linear contraction on a real L^1 space. If there is a compact subset $A \subset E$ such that*

$$\lim_{n\to\infty} \inf_{y\in A} \|T^n x - y\| = 0 \qquad \textit{for each } x \in B(E),$$

then T is periodic on E_{rev} and is asymptotic periodic in the sense of that

$$\lim_{n\to\infty} \|T^n x - T^n \pi x\| = 0,$$

where π is the projection onto E_{rev}.

From Theorems 2 and 3, we define a constrictive Markov operator.

Definition 1 We call a linear contraction T on a Banach space $(E, \|\cdot\|)$ *constrictive* if there is a compact subset $A \subset E$ such that

$$\lim_{n\to\infty} \inf_{y\in A} \|T^n x - y\| = 0 \qquad \text{for each } x \in B(E),$$

where $B(E)$ is the closed unit ball of E. We call A a *constrictor* for T.

3 Main Results

An element $x \in E$ is called an eigenvector for $\mathscr{S}$, if there is a map $\lambda : \mathscr{S} \to \mathbb{C}$ with $Tx = \lambda(T)x, T \in \mathscr{S}$. Moreover, x is called an *eigenvector with unimodular eigenvalues* if $|\lambda(T)| = 1$ holds for all $T \in \mathscr{S}$. Let E_{uds} be the closure of the subspace of E spanned by all eigenvectors with unimodular eigenvalues. In [5], we defined the following subset of a complex Banach space.

Definition 2 Let E be a complex Banach space with the strictly convex dual space and let $\mathscr{T}$ be a nonexpansive semigroup in E. Then, we define that

$$U = \bigcap_{T \in \mathscr{T}} \bigcup_{0 \le \theta \le 2\pi} \{m \in E : \langle x - e^{i\theta} Tx, Jm \rangle = 0 \text{ for all } x \in E\}.$$

By using this set, we showed the following theorems.

Theorem 4 (Honda[5]) *Let E be a strictly convex and reflexive complex Banach space with the strictly convex dual space E^* and let $\mathscr{T}$ be an abelian semigroup of linear contractive operators in E. Then, we have*

$$E_{fl}(\mathscr{T}) = (JU)_{\perp} \quad \text{and}$$
$$E_{rev}(\mathscr{T}) = E_{uds}(\mathscr{T}) = \overline{\text{spn}} U$$

Theorem 5 (Honda[5]) *Let E be a strictly convex and reflexive complex Banach space with the strictly convex dual E^* and let $\mathscr{T}$ be an abelian semigroup of linear contractive operators in E. Then, $x \in E_{fl}(\mathscr{T})$ if and only if $\langle x, h^* \rangle = 0$ holds for all eigenvectors h^* of $\mathscr{T}^*$ having unimodular eigenvalues.*

From Theorem 2 of Sect. 2, we can obtain the following result of discrete and continuous semigoups, immediately.

Theorem 6 *Let E be a strictly convex and reflexive complex Banach space with the strictly convex dual space E^*. Given a one-parameter semigroup $\mathscr{T}$ of linear contractive operators on a complex Banach space E, the following assertions are equivalent:*

1. *there is a compact subset $A \subset E$ such that*

$$\lim_{t \to \infty} \inf_{y \in A} \|T_t x - y\| = 0 \qquad \text{for each } x \in B(E),$$

   where $B(E)$ is the closed unit ball of E.
2. *$x \in E_{fl}(\mathscr{T}) = \{x \in X : \lim_{t \to \infty} \|T_t x\| = 0\}$ if and only if $\langle x, h^* \rangle = 0$ holds for all eigenvectors h^* of $\mathscr{T}^*$ having unimodular eigenvalues.*

In descrete semigroups, we obtain more precise result.

Theorem 7 *Let E be a strictly convex and reflexive complex Banach space with the strictly convex dual space E^*. Given a linear contractive operator T on a complex Banach space E, the following assertions are equivalent:*

1. *T is constrictive.*
2. *$x \in E_{fl}(\mathscr{T}) = \{x \in X : \lim_{n\to\infty} \|T^n x\| = 0\}$ if and only if $\langle x, h^* \rangle = 0$ holds for all eigenvectors h^* of T^* having unimodular eigenvalues*

In the following, we will focus on an integral operator T with stochastic kernel and consider a situation in which T is constrictive Markov operator on a real L^1 space.

Let (Ω, Σ, μ) be a finite measure space on $\mathbb{R}$. We consider the real $L^1(\Omega)$ instead of a Banach space E of Sect. 2. We define an integral operator $T : L^1(\Omega) \to L^1(\Omega)$ by

$$Tx(\omega) = \int_\Omega K(\omega, \eta)x(\eta)\mu(d\eta) \qquad \text{for } x \in L^1(\Omega), \tag{1}$$

where $K : \Omega \times \Omega \to \mathbb{R}$ is a measurable function which satisfies

$$K(\omega, \eta) \geq 0 \quad \text{and} \quad \int_\Omega K(\omega, \eta)\mu(d\omega) = 1. \tag{2}$$

Any function K satisfying (2) is called a *stochastic kernel*. Clearly, the operator T satisfies the following properties:

P1 : $Tx \geq 0$ for all positive functions $x \in L^1(\Omega)$,
P2 : $\|Tx\| = \|x\|$ for all positive functions $x \in L^1(\Omega)$,
P3 : $T\mathbf{1}_\Omega(\omega) = \mathbf{1}_\Omega(\omega) \qquad$ for a.e. $\omega \in \Omega$.

where $\|\cdot\|$ denotes the L^1-norm and $\mathbf{1}_\Omega$ is the indicator function of Ω, defined by $\mathbf{1}_\Omega(\omega) = 1$ if $\omega \in \Omega$ and $\mathbf{1}_\Omega(\omega) = 0$ if $\omega \notin \Omega$.

Any linear operator $T : L^1 \to L^1$ satisfying P1 and P2 is called *Markov operator*. If an integral operator T on $L^1(\Omega)$ with stochastic kernel $K(\omega, \eta)$ satisfies P3, then the adjoint operator U of T is conservative, i.e.,

$$\text{if} \quad Uy \leq y \quad \text{for} \quad 0 \leq y \leq 1 \quad \text{then} \quad Uy = y.$$

(see Theorem 2.7 of Foguel [4]). Moreover, $T^k\mathbf{1}_\Omega = \mathbf{1}_\Omega$ for each $k \in \mathbb{N}$ and $T^k : L^1 \to L^1$ is also an integral Markov operator with a stochastic kernel $K_k(\omega, \eta)$, where $K_k(\omega, \eta)$is the compositions of $K(\omega, \eta)$ (see p.113 of Lasota and Macky [10]).

Lasota and Komorník proved that if a Markov operator is constrictive, then a spectral decomposition theorem holds on a "real" L^1 space, for which Sine prove in the general case(see [9], [3], [12]). This implies that constrictive Markov operators satisfy the Jacobs-de Leeuw–Glicksberg decomposition theorem. Moreover, Iwata give a necessary and sufficient condition for a constrictive Markov operator with a stochastic kernel defined on a real L^1 space([6]).

In the following theorem, we will give a sufficient condition of a constrictive Markov operator on a real $L^p(1 \le p < \infty)$ space.

Theorem 8 *Let* (Ω, Σ, μ) *be a finite measure space and* T *be an integral operator with a stochastic kernel* K *which satisfies that P1–P3 on real* $L^1(\Omega)$ *and* $\|Tx\|_p \le \|x\|_p$ *for some* $1 \le p < \infty$*, where* $\|\cdot\|_p$ *denotes the* L^p*-norm. If the sub* σ*-algebra*

$$\Sigma_0(T) = \{A \in \Sigma : \; T^n \mathbf{1}_A = \; \textit{characteristic function} \; \forall n \ge 0\}$$

has at most finitely many atoms and for each atom $W \in \Sigma_0(T)$*,*

$$\lim_{n\to\infty} \mu(A \setminus \mathrm{supp}(T^{dn}\mathbf{1}_B)) = 0 \qquad \forall A, B \subset W \quad \textit{of positive measure}, \tag{3}$$

where d *is the least common multiple of orders of atoms in* $\Sigma_0(T)$*, then* T *is constrictive on* $L^p(\Omega)$ *and*

$$\begin{aligned}
&L^p(\Omega) = E_{rev}(T) \oplus E_{fl}(T),\\
&E_{rev}(T) = \overline{\mathrm{spn}}\{\mathbf{1}_W : \; W \in \Sigma_0(T) \; \textit{is atom}\}, \textit{and}\\
&E_{fl}(T) = \{x \in L^p(\Omega) : \; \lim_{n\to\infty} \|T^n x\|_p = 0\}.
\end{aligned}$$

Proof The author proved that if T is an integral operator with stochastic kernel K which satisfies that P1–P3 on $L^1(\Omega)$, then sub σ-algebra $\Sigma_0(T)$ is atomic with respect to μ and all atoms in $\Sigma_0(T)$ are disjoint and cyclic (see Lemma 1, Iwata [6]). Thus we let $\{W_1, W_2, \cdots, W_m\}$ be the set of all atoms of $\Sigma_0(T)$. Since $\Sigma_0(T)$ has at most finitely many atoms, there exists the least common multiple d of orders of W_i, that is,

$$T^d \mathbf{1}_{W_i} = \mathbf{1}_{W_i} \qquad \text{for all } i = 1, \cdots, m.$$

Let R_{W_i} be the restriction of T^d to $L^1(W_i)$. We start by showing

$$\begin{aligned}
\Sigma_0(R_{W_i}) &:= \{A \in \Sigma : \; R^n_{W_i}\mathbf{1}_A = \text{characteristic function } \forall n \ge 0\}\\
&= \{\phi, W_i\}.
\end{aligned} \tag{4}$$

Suppose that there exists $A \in \Sigma_0(R_{W_i})$ such that $0 < \mu(A) < \mu(W_i)$, so that there exist $B_n \in \Sigma_0(R_{W_i})$ such that $R^n_{W_i}\mathbf{1}_A = \mathbf{1}_{B_n}$. Thus we have

$$R^n_{W_i}\mathbf{1}_{W_i \cap A^c} = \mathbf{1}_{W_i} - R^n_{W_i}\mathbf{1}_A = \mathbf{1}_{W_i \cap B_n^c}.$$

This implies that $\mathrm{supp}(R^n_{W_i}\mathbf{1}_{W_i\cap A^c}) = W_i \setminus (\mathrm{supp}(R^n_{W_i}\mathbf{1}_A))$. By (3), we have

$$0 = \lim_{n\to\infty} \mu(W_i \setminus \mathrm{supp}(R^n_{W_i}\mathbf{1}_A)) = \lim_{n\to\infty} \mu(\mathrm{supp}(R^n_{W_i}\mathbf{1}_{W_i\cap A^c})).$$

Similarly, we have $\mathrm{supp}(R_{W_i}^n \mathbf{1}_A) = W_i \setminus (\mathrm{supp}(R_{W_i}^n \mathbf{1}_{W_i \cap A^c}))$, then

$$0 = \lim_{n\to\infty} \mu(W_i \setminus \mathrm{supp}(R_{W_i}^n \mathbf{1}_{W_i \cap A^c})) = \lim_{n\to\infty} \mu(\mathrm{supp}(R_{W_i}^n \mathbf{1}_A)).$$

Therefore

$$0 = \mu(W_i) = \lim_{n\to\infty} \mu(\mathrm{supp}(R_{W_i}^n \mathbf{1}_A)) + \lim_{n\to\infty} \mu(\mathrm{supp}(R_{W_i}^n \mathbf{1}_{W_i \cap A^c})).$$

This contradicts $\mu(W_i) > 0$. Thus $\Sigma_0(R_{W_i}) = \{\phi, W_i\}$.

Define a new measure $\bar{\mu}$ on X by $d\bar{\mu} = \frac{\mathbf{1}_{W_i}}{\mu(W_i)} d\mu$. Since $\bar{\mu}(W_i) = 1$, $R_{W_i} : L^1(W_i, \bar{\mu}) \to L^1(W_i, \bar{\mu})$ is a Harris operator by 3 of Lemma 2 of Iwata [6]. Applying (v) of Proposition1 of Iwata [6] to R_{W_i}, we have

$$\lim_{n\to\infty} R_{W_i}^n x(\omega) = \frac{\int_{W_i} x(\omega)\bar{\mu}(d\omega)}{\bar{\mu}(W_i)} \mathbf{1}_{W_i}(\omega) \qquad \text{for } \bar{\mu}\text{-a.e.}\omega \in \Omega$$

for all $x \in L^p(W_i)$. Moreover, since $\|Tx\|_p \le \|x\|_p$, we have

$$\begin{aligned}
\left| R_{W_i}^n x - \frac{\int_{W_i} f(x)\bar{\mu}(dx)}{\bar{\mu}(W_i)} \mathbf{1}_{W_i}(x) \right| &= \left| R_{W_i}^n \left(x(\omega) - \frac{\int_{W_i} x(\omega)\bar{\mu}(d\omega)}{\bar{\mu}(W_i)} \mathbf{1}_{W_i}(\omega) \right) \right| \\
&\le R_{W_i}^n \left| x(\omega) - \frac{\int_{W_i} x(\omega)\bar{\mu}(d\omega)}{\bar{\mu}(W_i)} \mathbf{1}_{W_i}(\omega) \right| \\
&\le \left\| x - \frac{\int_{W_i} x(\omega)\bar{\mu}(d\omega)}{\bar{\mu}(W_i)} \mathbf{1}_{W_i} \right\|_{L^p(W_i, \bar{\mu})} \qquad \text{for } \bar{\mu}\text{-a.e.}\omega \in \Omega \quad \text{and} \quad \forall n \in \mathbb{N}.
\end{aligned}$$

Note that $\mathrm{supp}\left(R_{W_i}^n x - \frac{\int_{W_i} x(\omega)\mu(d\omega)}{\mu(W_i)} \mathbf{1}_{W_i}(\omega) \right) \subset W_i$. Thus, by the L^p Dominated Convergence Theorem, we obtain

$$\begin{aligned}
0 &= \lim_{n\to\infty} \left\| R_{W_i}^n x - \frac{\int_{W_i} x(\omega)\bar{\mu}(d\omega)}{\bar{\mu}(W_i)} \mathbf{1}_{W_i}(\omega) \right\|_{L^p(W_i, \bar{\mu})}^p \\
&= \lim_{n\to\infty} \int_\Omega \left| R_{W_i}^n x - \frac{\int_{W_i} x(\omega)\mu(d\omega)}{\mu(W_i)} \mathbf{1}_{W_i}(x) \right|^p \frac{\mathbf{1}_{W_i}}{\mu(W_i)} d\mu \\
&= \lim_{n\to\infty} \frac{1}{\mu(W_i)} \left\| R_{W_i}^n x - \frac{\int_{W_i} x(\omega)\mu(d\omega)}{\mu(W_i)} \mathbf{1}_{W_i}(\omega) \right\|_p^p
\end{aligned}$$

for all $x \in L^p(W_i)$. This implies that $\lim_{n\to\infty} \|R_{W_i}^n x - \frac{\langle x, \mathbf{1}_\Omega \rangle}{\mu(W_i)} \mathbf{1}_{W_i}\|_p = 0$ for all $x \in L^p(W_i)$. Then we have

$$L^p(W_i) = E_{rev}(R_{W_i}) \oplus L_0(R_{W_i}),$$

where

$$E_{rev}(R_{W_i}) = \{x \in L^p(W_i) : \ R_{W_i}x = x\} \quad \text{and,}$$
$$E_{fl}(R_{W_i}) = \{x \in L^p(W_i) : \ \lim_{n\to\infty} \|R_{W_i}^n x\|_p = 0\}.$$

Clearly, $E_{rev}(R_{W_i})$ is the closed linear span generated by $\mathbf{1}_{W_i}$. Besides, $x \in E_{fl}(R_{W_i})$ for any $x \in E_{fl}(T) = \{x \in L^p(\Omega) : \ \lim_{n\to\infty} \|T^n x\|_p = 0\}$ with $\mathrm{supp}(x) \subset W_i$. On the other hand, fix $x \in E_{fl}(R_{W_i})$ arbitrarily. Note that the sequence $\{\|T^n x\|_p\}_{n\geq 0}$ is decreasing and bounded because T is a Markov operator, and hence converges. Now $0 = \lim_{n\to\infty} \|R_{W_i}^n x\|_p = \lim_{n\to\infty} \|T^{dn} x\|_p$, then the sequence $\{\|T^n x\|_p\}_{n\geq 0}$ converges to 0. This implies that $x \in E_{fl}(T)$. Therefore we get

$$L^p(W_i) = E_{rev}(R_{W_i}) \oplus E_{fl}(T_{W_i}), \tag{5}$$

where $E_{fl}(T_{W_i}) = \{f \in L^1(W_i) : \ \lim_{n\to\infty} \|T^n f\|_p = 0\}$. Therefore we obtain

$$L^p(\Omega) = E_{rev}(T) \oplus E_{fl}(T),$$

where $E_{rev}(T) = \overline{\mathrm{spn}}\{\mathbf{1}_W : \ W \in \Sigma_0(T) \text{ is atom }\}$ and $E_{fl}(T) = \{x \in L^1(\Omega) : \lim_{n\to\infty} \|T^n x\|_p = 0\}$. This implies that T is constrictive on $L^p(\Omega)$ by Theorem 1.3.3 of Emel'yanov [3].

Acknowledgements This work was partially supported by JSPS Grant-in-Aid for Scientific Research (C) Grant Number 25400124.

References

1. Cioranescu, I.: Geometry of Banach spaces, duality mappings, and nonlinear problems. Kluwer Academic Publishers Group, Dordrecht (1990)
2. de Leeuw, K., Glicksberg, I.: Applications of almost periodic compactifications. Acta Math. **105**, 63–97 (1961)
3. Emel'yanov, Eduard Yu.: Non-spectral asymptotic analysis of one-parameter operator semigroups. Operator Theory: Advances and Applications, vol. 173, Birkhäuser Verlag, Basel (2007)
4. Foguel, Shaul, R.: The Ergodic Theory of Markov Processes. Van Nostrand Mathematical Studies. Van Nostrand Reinhold Co, New York (1969)
5. Honda, T.: A nonlinear analytic approach to the splitting theorem of Jacobs-de Leeuw-Glicksberg. Nonlinear Anal. **75**(16), 6002–6008 (2012)
6. Iwata, Y.: Constrictive Markov operators induced by Markov processes. Positivity. **20**(2), 355–367 (2016)
7. Jacobs, K.: Ergodentheorie und fastperiodische Funktionen auf Halbgruppen. Math. Z. **64**, 298–338 (1956)
8. Krengel, U.: Ergodic Theorems. Walter de Gruyter & Co., Berlin (1985)

9. Komorník, J., Lasota, A.: Asymptotic decomposition of Markov operators. Bull. Pol. Acad. Sci. Math. **35**(321–327), 5–6 (1987)
10. Lasota, Mackey, A., Michael, C.: Applied Mathematical Sciences. Chaos, fractals, and noise. Stochastic aspects of dynamics, 2nd edn. Springer, New York (1994)
11. Megginson, R.E.: An Introduction to Banach Space Theory. Springer, Berlin (1998)
12. Sine, R.: Constricted systems. Rocky Mt. J. Math. **21**(4), 1373–1383 (1991)
13. Takahashi, W.: Nonlinear Functional Analysis – Fixed Point Theory and Its Applications. Yokohama Publishers (2000)

On the Behavior of the Error in Numerical Iterative Method for PDE

Toshiyuki Kohno

Abstract The purpose of paper is to analyze the behavior of the error in the iterative method. Especially, we are interested in the classical iterative method such as SOR method and its preconditioning techniques to solve the linear system $Au = q$. In order to accelerate convergence, many researchers proposed several preconditioners [4–8]. There is also preconditioner available for both classical iterative and Krylov subspace methods. We focus on the behavior of error to find a good preconditioner. We treat difference equation derived from partial differential equation(PDE), because the coefficient matrix given by using difference approximation is easy to investigate. By examining the behavior of the error, we choose an effective preconditioner, and show the numerical results.

Keywords Preconditioner · Iterative method · PDE

1 Differential and Difference Equations

Many phenomena in sciences and engineering depend on more than one independent variable. The differential equation for the unknown function then involves partial derivatives of the function with respect to these independent variables. We consider the following partial differential equation(PDE),

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y), \qquad x, y \in \Omega \tag{1}$$

This equation is called the Poisson equation. Let Ω be a planar domain, and denote its boundary by $\partial\Omega$. To treat as the boundary value problem(BVP), we set

T. Kohno (✉)
Department of Information Science, Okayama University of Science, Ridai-cho 1-1,
Okayama 700-0005, Japan
e-mail: kohno@mis.ous.ac.jp

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_8

$$u = g(x, y), \quad x, y \in \partial\Omega.$$

We assume Ω is a square $\Omega = (0, 1) \times (0, 1)$. Then the boundary $\partial\Omega$ consists of four segments, witch are the four sides of the square. We divide the x interval [0, 1] into n equal parts and denote $h = \frac{1}{n}$ the x step size. Similarly, we divide the y. Then the grid points are (x_i, y_i), $1 \leq i, j \leq n + 1$, where $x_i = (i - 1)h$, $y_i = (j - 1)h$. For the differential equation at an interior grid point (x_i, y_i), we use the the three-points central difference to approximate the second derivative. Let $f_{ij} = f(x_i, y_i)$ and denote u_{ij} the finite difference approximation of $u(x_i, y_i)$. Then we have the following equation,

$$u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 2u_{i,j} = h^2 f_{ij}, \quad 2 \leq i, j \leq n. \tag{2}$$

It can be shown that the accuracy of the solution is of second order,

$$\max |u(x_i, y_i) - u_{ij}| = O(h^2) \tag{3}$$

In the analysis of numerical methods for solving boundary value problems, the truncation error is defined to the discrepancy between the difference equation and the differential equation. However, Eq. (2) represents a matrix equation with a simple five-diagonal matrix, is a manageable problem. In this paper, we consider about such a five-diagonal symmetric matrix.

2 Basic Iterative Method and Krylov Subspace Method

From the above Eq. (2), we have the following linear system,

$$Au = q. \tag{4}$$

We consider the splitting of A as following,

$$A = M - N, \tag{5}$$

where M is nonsingular. Hence we can construct a splitting-based iterative method as follows:

$$\boldsymbol{u}^{(k+1)} = M^{-1}N\boldsymbol{u}\ (k) + M^{-1}\boldsymbol{q}. \tag{6}$$

If the spectral radius of the iterative matrix $M^{-1}N$ is less than one, the sequence $\{u^{(k)}\}$ will converge to the solution of the linear system. We can express the matrix A as the matrix sum

$$A = D - E - F \tag{7}$$

where $D = \mathrm{diag}\{a_{11}, a_{22}, \ldots, a_{nn}\}$, and E and F are strictly lower and strictly upper triangular $n \times n$ matrices, respectively. When setting $M = D$, we have the point Jacobi iterative method. And if $M = D - E$, then the Gauss–Seidel iterative method. Moreover, we have the SOR iterative method by using $M = \frac{1}{\omega}(D - \omega E)$.

Next, we explain the Krylov subspace method. These techniques are based on projection processes onto Krylov subspaces which are subspaces spanned by vectors of the form $p(A)v$ where p is a polynomial. The general projection method for solving the linear system (4) is a method which seeks an approximate solution u_m from an affine subspace $\boldsymbol{u}_0 + \mathcal{K}_m$ of dimension m by imposing the Petrov–Galerkin condition

$$\boldsymbol{q} - A\boldsymbol{u}_m \perp \mathcal{L}_m, \tag{8}$$

where $\mathcal{L}_m$ is another subspace of dimension m. Here, u_0 represents an arbitrary initial guess to the solution. A Krylov subspace method is a method for which the subspace $\mathcal{K}_m$ is the Krylov subspace

$$\mathcal{K}_m(A, \boldsymbol{r}_0) = span\{\boldsymbol{r}_0, A\boldsymbol{r}_0, A^2\boldsymbol{r}_0, \cdots, A^{m-1}\boldsymbol{r}_0\}, \tag{9}$$

where $\boldsymbol{r}_0 = \boldsymbol{q} - A\boldsymbol{u}_0$. The different version of Krylov subspace methods arise from different choices of the subspace $\mathcal{L}_m$. There are methods such as the conjugate Gradient(CG), BiCG, conjugate gradient squared(CGS), and so on. The numerical comparison of these iterative methods is done only by graphically illustrating the norm of the residual vector $\boldsymbol{r}^{(k)} = \boldsymbol{q} - A\boldsymbol{u}^{(k)}$. We examine the behavior of the norm of residual vector of each iterative methods for the above BVP. Figure 1 shows the norm of

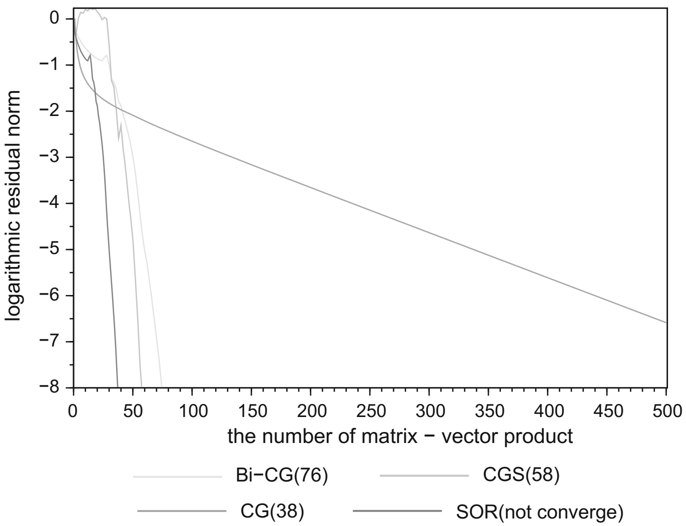


Fig. 1 the norm of residual vector of BiCG, CGS, CG, and SOR methods for the BVP

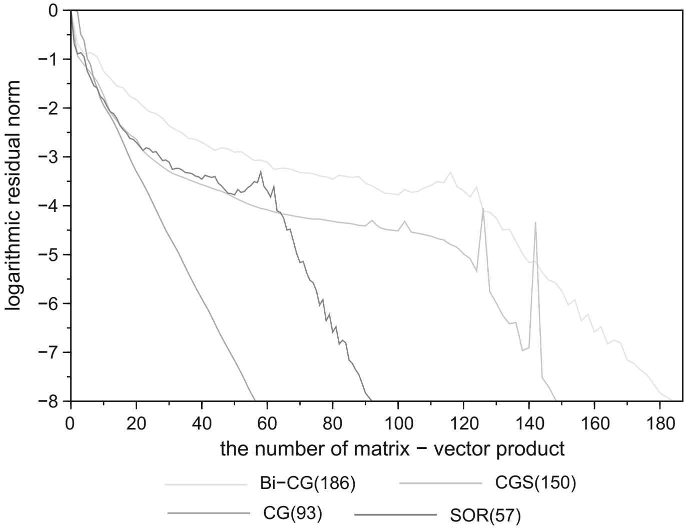


Fig. 2 the norm of residual vector for BiCG, CGS, CG, and SOR methods for a five-diagonal matrix

residual vector vs. the number of matrix-vector product for the Bi-CG, CGS, CG and SOR methods. The Horizontal axis shows the number of the product of matrix and vector, its number entered after the name of each method. The vertical axis is the logarithm of the norm of the residual vector. The SOR method does not converge, and other methods converge with less computation. The order of coefficient matrix is 400, this problem is symmetric, with small problem.

Next, we show the another five-diagonal symmetric problem derived from finite element method in Fig. 2. The order of a coefficient matrix is 176.

We find the twice big bouncing locations of CGS method in Fig. 2. And the behavior of the norm of the Bi-CG method is vibrating in the last steps. The SOR method ($\omega = 1$) indicates smooth convergence. This problem is convenient for the SOR method. We will try to examine the behavior of error more detail to find a good preconditioner.

3 Vector Visualisation

We will try to display the residual vectors in due to learn more about these results. We show the plots of residual vector for CGS method in Fig. 3, at first. Since the range of the logarithm value of the norm of the residual vector is 0 to -8, we showed the error with shading. In the graph, the shading is indicated by multiplying by -1. When the value is 0, the error is the largest and expressed in dark color. We put a color bar in Fig. 3.

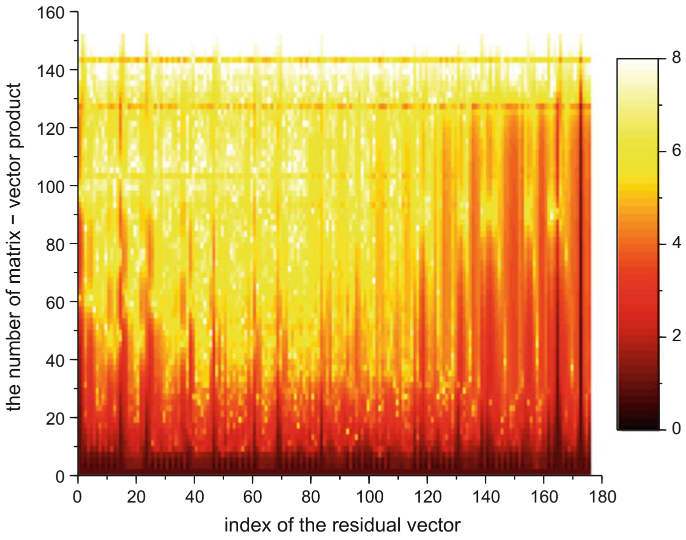


Fig. 3 the residual vector plot of CGS method

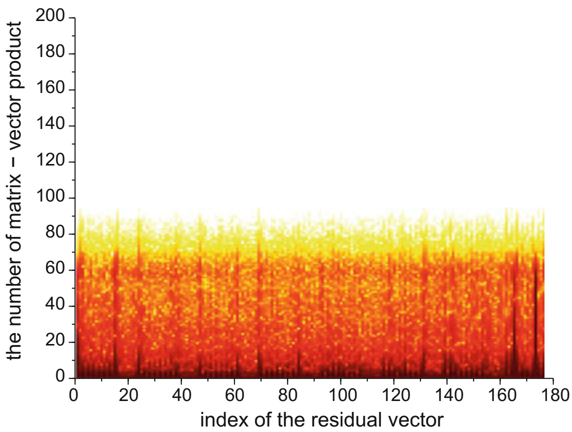


Fig. 4 CG method

Vertical axis is the number of matrix - vector product, horizontal axis is the index of residual vector, and we display the magnitude of the residual vector with gradation. At the beginning of the iterative process, the error is large, so it is expressed in dark color. After 30 iterations, it turns out that the proportion occupying the bright color increases. A phenomenon appears in which the clear error becomes large around 130 and 140 calculations. We found that the errors are large in each element of the residual vector. Also, we found that the error from the 120th to end of residual vector is large.

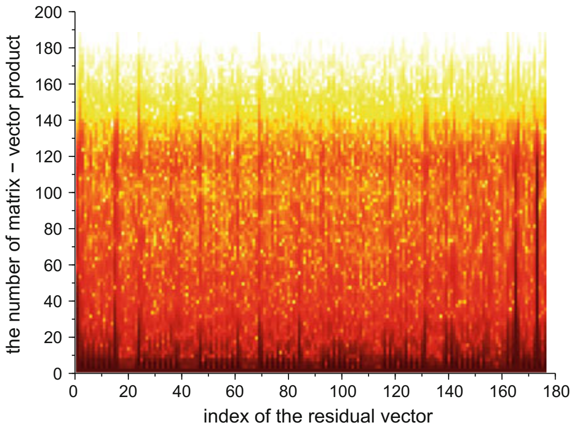


Fig. 5 Bi-CG method

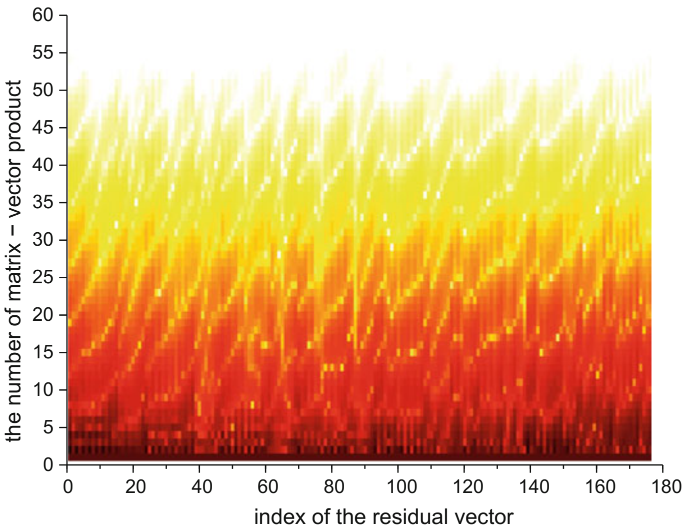


Fig. 6 SOR method

The results of the CG and BiCG method are shown in Figs. 4 and 5. Although the problem is a symmetric matrix, we see the error remains in the calculation process. In Fig. 6, we found that the SOR method has an unusual convergence situation. The number of calculations of the SOR method is a little, but the parts with the small error draw some oblique lines. We think that some oblique lines may indicate the features of the successive calculation of the SOR method. From this error behavior, we may be able to propose appropriate preconditioner.

4 Preconditioner and Comparison Theorem

Many preconditioners are proposed to accelerate its convergence for the basic iterative and Krylov subspace methods. For the classical iterative method, by using the some nonsingular matrix P, we have the preconditioned linear system

$$PA\boldsymbol{u} = P\boldsymbol{q}. \tag{10}$$

In 1994, Kohno et al. citekohno proposed the preconditioner $P = I + S(\alpha)\dot{\text{A}}$nd many researchers study the some preconditioners [4, 6–8]. We obtained the improved results in the convergence by some preconditioners, and we proved the comparison theorems.

We review some known results.

We write $A \leq B$ if $a_{ij} \leq b_{ij}$ holds for all elements of $A = (a_{ij})$ and $B = (b_{ij}) \in \boldsymbol{R}^{n\times n}$, calling A nonnegative if $A \geq O$. This definition carries immediately over to vectors by identifying them with $n \times 1$ matrices. In particular, we call the vector $\boldsymbol{v} \in \boldsymbol{R}^n$ positive (writing $\boldsymbol{v} > 0$) if all its elements are positive. Let $\boldsymbol{Z}^{n\times n}$ denote that set of all real $n \times n$ matrices which have non-positive off-diagonal elements. A nonsingular matrix $A \in \boldsymbol{Z}^{n\times n}$ is called an M-matrix if $A^{-1} \geq O$.

Definition 1 Let A be a real matrix. The representation $A = M - N$ is called a splitting of A if M is a nonsingular matrix. In addition, the splitting is

(i) convergent if $\rho(M^{-1}N) < 1$,
(ii) regular if $M^{-1} \geq O$ and $N \geq O$,
(iii) weak regular if $M^{-1} \geq O$ and $M^{-1}N \geq O$,
(iv) M-splitting if M is an M-matrix and $N \geq O$,

Definition 2 We call $A = M - N = (D - E) - F$ the Gauss–Seidel regular splitting of A if $(D - E)^{-1} \geq 0$ and $F \geq 0$, where D is the diagonal matrix and E and F are strictly lower and strictly upper triangular matrices of A, respectively.

Theorem 1 [3] *Let $A = M - N$ be a splitting.*

(i) If the splitting is regular or weak regular, then $\rho(M^{-1}N) < 1$ if and only if $A^{-1} \geq O$.
(ii) If the splitting is an M-splitting, then $\rho(M^{-1}N) < 1$ if and only if A is an M-matrix.

Theorem 2 [2] *Let $A \in \boldsymbol{Z}^{n\times n}$ be irreducible. Then each of the following conditions is equivalent to the statement: "A is a nonsingular M-matrix".*

(i) $A^{-1} > O$.
(ii) $A\boldsymbol{v} \geq 0$ for some $\boldsymbol{v} > 0$.

Lemma 1 [3] *Let $T \geq O$. If there exist $\boldsymbol{v} > 0$ and $\alpha > 0$ such that $T\boldsymbol{v} \leq \alpha\boldsymbol{v}$, then $\rho(T) \leq \alpha$. Moreover, if $T\boldsymbol{v} < \alpha\boldsymbol{v}$, then $\rho(T) < \alpha$.*

Theorem 3 [11] *Let $A = M - N$ be a regular splitting of the matrix A. Then, A is nonsingular with $A^{-1} \geq O$, if and only if $\rho(M^{-1}N) < 1$, where*

$$\rho(M^{-1}N) = \frac{\rho(A^{-1}N)}{1 + \rho(A^{-1}N)} < 1.$$

Theorem 4 [1] *Let $A = M_1 - N_1 = M_2 - N_2$ be two convergent splitting of A. Then the followings hold:*

(i) If $N_2 \geq N_1 \geq O$ and $N_i, i = 1, 2$ are monotone(i.e., the splittings are regular), then $M_1^{-1} \geq M_2^{-1}$.
(ii) If $M_1^{-1} \geq M_2^{-1}$ and $N_1\boldsymbol{x} \geq 0$, then $(M_1^{-1} - M_2^{-1})N_1\boldsymbol{x} \geq 0$, where $\boldsymbol{x}$ is the Perron vector of $G_1 = A^{-1}N$.
(iii) If $(M_1^{-1} - M_2^{-1})N_1\boldsymbol{x} \geq 0$, where $\boldsymbol{x}$ is the Perron vector of G_1, and if $A = M_i - N_i, i = 1, 2$ are weak regular splittings, then $\rho(M_1^{-1}N_1) \leq \rho(M_2^{-1}N_2)$.

By using above lemmas and Theorem 4, we derive comparison theorem for the general preconditioner P.

Theorem 5 *Let $A \in Z^{n \times n}$ be an irreducibly diagonally dominant matrix and $A = M - N$ be Gauss–Seidel regular splitting. The Gauss–Seidel iterative matrix $T = M^{-1}N$. Put the preconditioner P such that $PA = M_P - N_P$ is Gauss–Seidel regular splitting and*

$$M_p^{-1}P - M^{-1} \geq 0. \tag{11}$$

Then,

$$\rho(M_p^{-1}N_p) \leq \rho(T) < 1. \tag{12}$$

Proof Clearly, $A^{-1} \geq O$, thus from Theorem 3, $\rho(T) < 1$ holds. So by putting $A = P^{-1}(M_p - N_p)$, we have

$$A = M - N = P^{-1}(M_p - N_p).$$

As $A = M - N$ is Gauss–Seidel regular splitting, there exists a positive vector $\boldsymbol{v}$ satisfied the following equation

$$\rho(T)\boldsymbol{v} = T\boldsymbol{v}.$$

Then,

$$\boldsymbol{v} = \frac{1}{\rho(T)} M^{-1} N \boldsymbol{v} \geq 0.$$

Hence,

$$M\boldsymbol{v} = \frac{1}{\rho(T)} N \boldsymbol{v} \geq 0,$$

and,

$$\begin{aligned} A\boldsymbol{v} &= (M - N)\boldsymbol{v} = M(I - T)\boldsymbol{v} \\ &= \frac{1 - \rho(T)}{\rho(T)} N\boldsymbol{v} \geq O. \end{aligned}$$

From the condition, we have

$$\begin{aligned} \{(P^{-1}M_p)^{-1} - M^{-1}\}A\boldsymbol{v} &= (M_p^{-1}P - M^{-1})A\boldsymbol{v} \\ &= M_p^{-1}P\{P^{-1}(M_p - N_p)\}\boldsymbol{v} - (I - M^{-1}N)\boldsymbol{v} \\ &= (I - M_p^{-1}N_p)\boldsymbol{v} - (I - T)\boldsymbol{v} \\ &= T\boldsymbol{x} - M_p^{-1}N_p\boldsymbol{v} = \rho(T)\boldsymbol{v} - M_p^{-1}N_p\boldsymbol{v} \geq 0, \end{aligned} \tag{13}$$

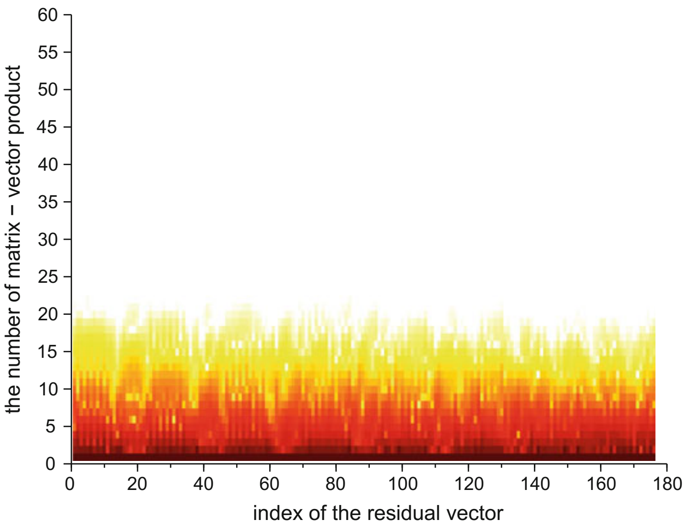


Fig. 7 Preconditioned SOR method with preconditioner $I + \beta U$

and by Lemma 1 implies

$$\rho(M_p^{-1}N_p) \leq \rho(T) < 1.$$

□

From the phenomenon that the parts with the small error draw some oblique lines, we consider that the convergence will be improved by using later elements of coefficient matrix A. The preconditioners using the behind element has been proposed [9, 10]. By using the preconditioner $P = (I + \beta U)$ where U is the upper codiagonal part of A, we obtained the good result in Fig. 7. We decided the parameter $\beta = 0.3$ from several numerical experiments.

5 Conclusion

We used the liner system obtained by discretization the PDE as a model problem. We think that it was impossible to select an appropriate preconditioner from the graph denoting the history of norm of the residual vector. By visualising the vector, we was able to choose a good preconditioner. For future work, we will experiment with another five-diagonal matrix.

Acknowledgements The author would like to thank the referees for pointing out some corrections. This study was supported by JSPS KAKENHI Grant Number JP 26400181.

References

1. Axelsson, O.: Iterative Solution Methods. Cambridge University Press, Cambridge (1994)
2. Berman, A., Plemmons, R.J.: Nonnegative Matrices in the Mathematical Sciences. Elsevier, Amsterdam (1994)
3. Frommer, A., Szyldm, D.B.: H-splitting and two-stage iterative methods. Num. Math. **63**, 345–356 (1992)
4. Gunawardena, A.D., Jain, S.K., Snyder, L.: Modified iterative methods for consistent linear systems. Linear Algebra Appl. **154–156**, 123–143 (1991)
5. Kohno, T., Kotakemori, H., Niki, H., Usui, M.: Improving modified Gauss–Seidel method for Z-matrices. Linear Algebra Appl. **267**, 113–123 (1997)
6. Kotakemori, H., Niki, H., Okamoto, N.: Accelerated iteration method for Z-matrices. J. Comput. Appl. Math. **75**, 87–97 (1996)
7. Kotakemori, H., Harada, K., Morimoto, M., Niki, H.: A comparison theorem for the iterative method with the preconditioner $(I + S_{\max})$. J. Comput. Appl. Math. **145**, 375–378 (2002)
8. Morimoto, M., Harada, K., Sakakihara, M., Sawami, H.: The Gauss–Seidel iterative method with the preconditioning matrix $(I + S + S_m)$. Jpn J. Ind. Appl. Math. **21**, 439–445 (2004)
9. Usui, M., Kohno, T., Niki, H.: Adaptive Gauss-Seidel method for linear systems. Int. J. Comput. Math. **51**, 119–125 (1994)
10. Usui, M., Niki, H., Kotakemori, H., Kohno, T.: The convergence theorem of $(I + \beta U)$-type preconditioned Gauss–Seidel iterative method. Jpn. J. Ind. Appl. Math. **6**(4), 307–316 (1996)
11. Varga, R.S.: Matrix Iterative Analysis. Springer, Berlin (2000)

Property B of the Four-Dimensional Neutral Difference System

Jana Krejčová

Abstract We deal with a four-dimensional nonlinear difference system with deviating arguments in the paper. The first equation of the system is of a neutral type. We study oscillatory and nonoscillatory solutions of neutral difference systems and their asymptotic properties. We establish sufficient conditions for the system to have strongly monotone solutions or Kneser solutions and then sufficient conditions for the system to have property B.

Keywords Property B · Strongly monotone solution · Kneser solution · Oscillatory solution · Nonoscillatory solution · Quickly oscillatory solution

1 Introduction

In this paper, we study asymptotic behavior of solutions of a four-dimensional system

$$\begin{aligned} \Delta\left(x_n + p_n x_{n-\sigma}\right) &= A_n\, f_1\left(y_n\right) \\ \Delta y_n &= B_n\, f_2\left(z_n\right) \\ \Delta z_n &= C_n\, f_3\left(w_n\right) \\ \Delta w_n &= D_n\, f_4\left(x_{\gamma_n}\right), \end{aligned} \tag{S}$$

where $n \in \mathbb{N}_0 = \{n_0, n_0+1, ...\}$, n_0 is a positive integer, σ is a nonnegative integer, $\{A_n\}, \{B_n\}, \{C_n\}, \{D_n\}$ are positive real sequences defined for $n \in \mathbb{N}_0$. Δ is the forward difference operator given by $\Delta x_n = x_{n+1} - x_n$.

The sequence $\gamma : \mathbb{N} \to \mathbb{N}$ satisfies

$$\lim_{n\to\infty} \gamma_n = \infty. \tag{H1}$$

The most common form of this sequence is $\gamma = n \pm \tau$, where $\tau \in \mathbb{N}$.

J. Krejčová (✉)
Faculty of Informatics and Statistics, University of Economics,
W. Churchill Sq. 4, 130 67 Prague, Czech Republic
e-mail: jana.krejcova@vse.cz

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_9

The sequence $\{p_n\}$ is a sequence of the real numbers and it satisfies

$$\lim_{n\to\infty} p_n = P, \text{ where } |P| < 1. \tag{H2}$$

Functions $f_i : \mathbb{R} \to \mathbb{R}$ for $i = 1, .., 4$ are invertible and satisfy

$$\frac{f_i(u)}{u} \geq 1, u \in \mathbb{R}\backslash 0. \tag{H3}$$

Nonlinear difference systems or difference equations are often studied when either

$$\sum_{n=n_0}^{\infty} A_n = \infty, \quad \sum_{n=n_0}^{\infty} C_n = \infty, \tag{H4}$$

or

$$\sum_{n=n_0}^{\infty} A_n = \infty, \quad \sum_{n=n_0}^{\infty} B_n = \infty, \quad \sum_{n=n_0}^{\infty} C_n = \infty. \tag{H5}$$

hold. If the condition (H5) holds, then we said that the system (S) is in the canonical form. In this article we study (S) with these conditions as well as without these conditions.

By a solution of the system (S) we mean a vector sequence (x, y, z, w) which satisfies the system (S) for $n \in \mathbb{N}_0$. We investigate oscillatory or nonoscillatory solutions. Therefore, the first important thing is to divide solutions into these groups.

Definition 1 The component x is said to be **nonoscillatory** if there exists $n_1 \geq n_0$ such that $x_n \geq 0$ (respectively $x_n \leq 0$) for all $n \geq n_1$. A solution of (S) is said to be nonoscillatory if all of its components x, y, z, w are nonoscillatory.

Definition 2 The component x is said to be **oscillatory** if for any $n_1 \geq n_0$ there exists $n \geq n_1$ such that $x_{n+1}x_n < 0$. If the component x satisfies $x_{n+1}x_n < 0$ for all $n \geq n_1$ then the component is said to be **quickly oscillatory**. A solution of (S) is said to be oscillatory (respectively quickly oscillatory) if all of its components x, y, z, w are oscillatory (respectively quickly oscillatory).

Definition 3 The system (S) has **weak property B** if every nonoscillatory solution of (S) satisfies

$$x_n z_n > 0 \quad \text{and } y_n w_n > 0 \quad \text{for large } n. \tag{1}$$

Definition 4 The system (S) has **property B** if any of its solutions either is oscillatory or satisfies either

$$\lim_{n\to\infty} |x_n| = \lim_{n\to\infty} |y_n| = \lim_{n\to\infty} |z_n| = \lim_{n\to\infty} |w_n| = \infty, \tag{2}$$

or

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n = \lim_{n\to\infty} z_n = \lim_{n\to\infty} w_n = 0. \tag{3}$$

Solutions satisfying (1) and $x_n y_n > 0$ are called *strongly monotone solutions*, while solutions satisfying (1) and $x_n y_n < 0$ are called *Kneser solutions*. Property B is defined in accordance with those for the higher-order differential equations or for the system of differential equations, see [8] and references therein. The system (S) is a prototype of even-order neutral systems and can be easily rewritten as a fourth-order nonlinear neutral difference equation. Equations with quasi-differences have been widely studied in the literature; see, for example, [2, 6, 9, 10]. In [6], oscillatory properties of solutions of the fourth-order difference equations are investigated. Their approach is based on studying the considered equation as a four-dimensional difference system, where $\{D_n\}$ is a negative real sequence. In [9], they studied asymptotic properties of neutral type difference equations. The problem of boundedness of solutions of the system (S) with $\gamma_n = n - \tau$ has been investigated in the recent paper [1].

The aim of this paper is to extend our results about asymptotic behavior of nonoscillatory solutions of (S). We are motivated by the paper [2], where asymptotic properties of (S) with $\{p_n\} = \{0\}$ have been investigated. We give sufficient conditions that (S) has weak property B and property B for the case when (S) is in the canonical form as well as without these assumptions. This completes the results from [6], where they study property A. We extend results from [1, 2, 7]. We continue in our previous research and we extend the results from our article [7], where we study system (S) in the canonical form with positive sequence $\{p_n\}$.

2 Oscillatory Solutions

Property B means that (S) has oscillatory or nonoscillatory solutions satisfying the asymptotic properties. We start with oscillatory solutions. Prototypes of oscillatory solutions of (S) are quickly oscillatory solutions, i.e. solutions of the form

$$x_n = (-1)^n q_n, \quad q_n > 0 \text{ for } n \in \mathbb{N}_0.$$

The following result can be seen as a necessary condition for their existence.

Theorem 1 *If*

$$\gamma_n \; mod\, 2 \neq n \; mod\, 2$$

and either

$$p_n \geq 0, \quad \sigma \text{ is even,} \tag{4}$$

or

$$p_n \le 0, \quad \sigma \text{ is odd,} \tag{5}$$

then the system (S) has no quickly oscillatory solutions.

Proof Let $x_n = (-1)^n q_n$ be a quickly oscillatory solution of (S). Then

$$\Delta\left(x_n + p_n x_{n-\sigma}\right) = (-1)^{n+1}\left(q_{n+1} + (-1)^{-\sigma} p_{n+1} q_{n+1-\sigma} + q_n + (-1)^{-\sigma} p_n q_{n-\sigma}\right).$$

Denote $S_n = q_{n+1} + (-1)^{-\sigma} p_{n+1} q_{n+1-\sigma} + q_n + (-1)^{-\sigma} p_n q_{n-\sigma}$. If (4) or (5) holds, then $S_n > 0$. From the first equation of (S) we have

$$y_n = f_1^{-1}\left(\frac{\Delta\left(x_n + p_n x_{n-\sigma}\right)}{A_n}\right) = (-1)^{n+1} f_1^{-1}\left(\frac{S_n}{A_n}\right).$$

Therefore

$$\Delta y_n = (-1)^n Y_n,$$

where $Y_n = f_1^{-1}\left(\frac{S_{n+1}}{A_{n+1}}\right) + f_1^{-1}\left(\frac{S_n}{A_n}\right) > 0$. From the second equation of (S) we obtain

$$z_n = f_2^{-1}\left(\frac{\Delta y_n}{B_n}\right), \quad \Delta z_n = (-1)^{n+1} Z_n,$$

where $Z_n = f_2^{-1}\left(\frac{Y_{n+1}}{B_{n+1}}\right) + f_2^{-1}\left(\frac{Y_n}{B_n}\right) > 0$. Repeating argument, we get from the third equation of (S)

$$w_n = f_3^{-1}\left(\frac{\Delta z_n}{C_n}\right), \quad \Delta w_n = (-1)^n W_n,$$

where $W_n = f_3^{-1}\left(\frac{Z_{n+1}}{C_{n+1}}\right) + f_3^{-1}\left(\frac{Z_n}{C_n}\right) > 0$. From here and from the fourth equation we have

$$(-1)^n W_n = D_n f_4\left((-1)^{\gamma_n} q_{\gamma_n}\right). \tag{6}$$

The signs of both sides of (6) are the same if and only if n and γ_n have the same remainder of division by two. □

By the method used in the proof of the previous theorem we can easily construct an example.

Example 1 Consider the equation

$$\begin{aligned} \Delta\left(x_n - \frac{1}{2}x_{n-1}\right) &= y_n \\ \Delta y_n &= z_n \\ \Delta z_n &= w_n \\ \Delta w_n &= 405\, x_{n-\tau}, \end{aligned} \tag{E1}$$

We have $p_n = -\frac{1}{2}$, $\sigma = 1$. Therefore, if τ is odd, (E1) has no quickly oscillatory solutions. If τ is even, the system can have a quickly oscillatory solution. Indeed, for $\tau = 2$ the system has the quickly oscillatory solution $x_n = (-1)^n 2^n$.

Example 2 Consider the equation

$$\begin{aligned} \Delta\left(x_n + \frac{1}{2}x_{n-2}\right) &= y_n \\ \Delta y_n &= z_n \\ \Delta z_n &= w_n \\ \Delta w_n &= \frac{729}{2}\, x_{n-\tau}, \end{aligned} \tag{E2}$$

We have $p_n = \frac{1}{2}$, $\sigma = 2$. Therefore, if τ is odd, (E2) has no quickly oscillatory solutions. If τ is even, the system can have a quickly oscillatory solution. Indeed, for $\tau = 2$ the system has the quickly oscillatory solution $x_n = (-1)^n 2^n$.

3 Nonoscillatory Solutions and Their Asymptotic Properties

If the system (S) has a solution (x, y, z, w), then it has the solution $(-x, -y, -z, -w)$ as well. Thus, throughout the paper, we can focus on solutions whose first component is eventually positive for large n.

We use the notation

$$s_n = x_n + p_n x_{n-\sigma}, \tag{7}$$

where $n \in \mathbb{N}_0$.

First, we point out some basic properties of (S) which we use to prove the main results of the paper. The first case of the following theorem was proved in [2, Lemma 1] and the proof of the second part can be prooved in the same way.

Lemma 1 *If* $\{p_n\} = \{0\}$*, then the solution* (x, y, z, w) *of (S) is nonoscillatory if and only if any of its components* x, y, z, w *is either positive or negative for large* n*. If* $\{p_n\} \neq \{0\}$ *and* x *is nonoscillatory, then components* y, z, w *and* s *are also nonoscillatory for large* n*.*

The following Lemma was proved in [7, Lemma 1] for the sequence $\{p_n\}$, where $0 \leq p_n < 1$. Now, we extend it for sequence $\{p_n\}$ satisfying (H2).

Lemma 2 *Let $\{x_n\}$ be eventually positive sequence and $\{p_n\}$ satisfies (H2), $n \in \mathbb{N}_0$. Let $\{s_n\}$ be the sequence defined by (7). Then $\{x_n\}$ is bounded if and only if $\{s_n\}$ is bounded. Moreover, if $\{s_n\}$ is positive and increasing for large n, then*

$$x_n \geq s_{n-\sigma}(1 - p_n) \quad \textit{for large } n. \tag{8}$$

Proof By (H2) and (7), the boundedness of x implies the boundedness of s. The opposite implication was proved in [6, Lemma 2] for $|P| < 1$. Therefore, we have to prove the assertion (8).

Assume $0 \leq P < 1$. We proved the estimation (8) in [7, Lemma 1].

Assume $-1 < P < 0$. If $\{s_n\}$ is positive and increasing, then $s_{n-\sigma} \leq s_n$. From the negativity of P we have $x_n \geq s_n$ and we get

$$x_n = s_n - p_n x_{n-\sigma} \geq s_n - p_n s_{n-\sigma} \geq s_{n-\sigma} - p_n s_{n-\sigma}.$$

□

The following lemma describes the possible types of nonoscillatory solutions.

Lemma 3 *Assume (H4). Then any nonoscillatory solution (x, y, z, w) of (S) with eventually positive x is one of the following types:*

type (a) $x_n > 0$ $y_n > 0$ $z_n > 0$ $w_n > 0$ for large n,
type (b) $x_n > 0$ $y_n > 0$ $z_n > 0$ $w_n < 0$ for large n,
type (c) $x_n > 0$ $y_n < 0$ $z_n > 0$ $w_n < 0$ for large n,
type (d) $x_n > 0$ $y_n < 0$ $z_n < 0$ $w_n < 0$ for large n,
type (e) $x_n > 0$ $y_n > 0$ $z_n < 0$ $w_n < 0$ for large n
type (f) $x_n > 0$ $y_n < 0$ $z_n > 0$ $w_n > 0$ for large n

Proof Let (x, y, z, w) be a nonoscillatory solution of (S) such that $x_n > 0$ for large n. There are eight possible types of these solutions. We prove that solutions of the following types do not exist.

type (i) $x_n > 0$ $y_n > 0$ $z_n < 0$ $w_n > 0$ for large n,
type (ii) $x_n > 0$ $y_n < 0$ $z_n < 0$ $w_n > 0$ for large n

Assume that there exist $n_1 \in \mathbb{N}_0$ and a solution such that $z_n < 0$, $w_n > 0$ for $n \geq n_1 \geq n_0$. From the fourth equation of (S) we have $\Delta w_n > 0$ and this implies that there exists $k > 0$ such that $w_n \geq k$ for large n. Using (H3) we have $f_3(w_n) \geq w_n \geq k$. By the summation of the third equation of (S) we have

$$z_n - z_{n_0} = \sum_{i=n_0}^{n-1} C_i f_3(w_i) \geq k \sum_{i=n_0}^{n-1} C_i.$$

Passing $n \to \infty$, we get a contradiction with the fact that $z_n < 0$. This excludes solutions of types (i) and (ii). □

Lemma 4 *Assume* (H5).

(i) *If* $0 \leq P < 1$, *then any nonoscillatory solution* (x, y, z, w) *of (S) with eventually positive* x *is of type (a), (b) or (c).*
(ii) *If* $-1 < P < 0$, *then any nonoscillatory solution* (x, y, z, w) *of (S) with eventually positive* x *is of type (a), (b), (c) or (d).*

Proof Assume that there exists a solution of type (e). Therefore, we have $z_n < 0$ and z is decreasing for all large n. This implies that there exists $l < 0$ such that $z_n \leq l$ for large n. From (H3) we get $f_2(z_n) \leq z_n \leq l$. By the summation of the second equation of (S) and passing $n \to \infty$ we get a contradiction with the positivity of y.

Assume that there exists a solution of type (f). Therefore, we have $z_n > 0$ and z is increasing for all large n. This implies that there exists $g > 0$ such that $z_n \geq g$ for large n. From (H3) we get $f_2(z_n) \geq z_n \geq g$. Using the summation of the second equation of (S) we get a contradiction with negativity of y. Thus, solutions of type (f) cannot exist.

The nonexistence of solutions of type (d) for $0 \leq P < 1$ was proved in our article [7]. □

By Definition 3, the system (S) has weak property B if there exist only nonoscillatory solutions of type (a) and (c). Solutions of type (a) are called strongly monotone and solutions of type (c) are called Kneser solutions. We have to determine some asymptotic properties of nonoscillatory solutions for the purpose of investigation property B. These properties are summarized in the following lemmas. Properties of strongly monotone solutions and Kneser solutions were proved in [7, Lemma 4, Lemma 5]. Therefore, there are presented without proofs.

Lemma 5 *Assume* (H4). *Then any solution of type (a) satisfies*

$$\lim_{n\to\infty} x_n = \infty, \quad \lim_{n\to\infty} z_n = \infty. \tag{9}$$

In addition, if (H5) *holds, then*

$$\lim_{n\to\infty} y_n = \infty. \tag{10}$$

Lemma 6 *Assume* (H4). *Then any solution of type (b) satisfies*

$$\lim_{n\to\infty} x_n = \infty, \quad \lim_{n\to\infty} w_n = 0. \tag{11}$$

Proof Let (x, y, z, w) be a solution of type (b). Because y is positive and increasing, there exists $k > 0$ such that $y_n \geq k$ for large n. By the summation of the first equation of (S) we get

$$s_n - s_{n_0} = \sum_{i=n_0}^{n-1} A_i f_1(y_i) \geq \sum_{i=n_0}^{n-1} A_i y_i \geq k \sum_{i=n_0}^{n-1} A_i.$$

Passing $n \to \infty$ we get $s_n \to \infty$. Lemma 2 implies that s is unbounded if and only if x is unbounded. Therefore $\lim_{n\to\infty} x_n = \infty$. Since w is negative and increasing, there exists $\lim_{n\to\infty} w_n = h, h \le 0$. Suppose $h < 0$, then from the summation of the third equation of (S) we get $z_n \to -\infty$ as $n \to \infty$, which is a contradiction with the boundedness of z. Therefore $\lim_{n\to\infty} w_n = 0$. □

Lemma 7 *Assume* (H4). *Then any solution of type (c) satisfies*

$$\lim_{n\to\infty} w_n = 0. \tag{12}$$

In addition, if (H5) *holds, then*

$$\lim_{n\to\infty} z_n = 0. \tag{13}$$

Lemma 8 *Assume* (H4). *Then any solution of type (d) satisfies*

$$\lim_{n\to\infty} x_n = \infty. \tag{14}$$

In addition, if (H5) *holds, then*

$$\lim_{n\to\infty} y_n = -\infty. \tag{15}$$

Proof Let (x, y, z, w) be a solution of type (d). Because y is negative and decreasing, there exists $k < 0$ such that $y_n \le k$ for large n. By the summation of the first equation of (S) we get

$$s_n - s_{n_0} = \sum_{i=n_0}^{n-1} A_i f_1(y_i) \le \sum_{i=n_0}^{n-1} A_i y_i \le k \sum_{i=n_0}^{n-1} A_i.$$

Passing $n \to \infty$ we get $s_n \to -\infty$. Lemma 2 implies that s is unbounded if and only if x is unbounded. Therefore $\lim_{n\to\infty} x_n = \infty$.

Since z is negative and decreasing, then using the same argument and the summation of the second equation of (S) we get $y_n \to -\infty$. □

We can continue to state sufficient conditions for the system (S) to have weak property B and property B.

4 Weak Property B and Property B

The first theorem gives the simple criterion that the system (S) has property B.

Theorem 2 *Assume* (H5). *If*

$$\sum_{n=n_0}^{\infty} D_n = \infty \tag{16}$$

holds, then the system (S) has property B.

Proof Assume that (x, y, z, w) is a nonoscillatory solution of the system (S) of type (b) or (d). Since x is positive and, by Lemmas 6 and 8, $\lim_{n\to\infty} x_n = \infty$, then there exists a real constant $k > 0$ such that $x_n \geq k$ for large n. By the summation of the fourth equation of (S) we get

$$w_n - w_{n_0} = \sum_{i=n_0}^{n-1} D_i f_4\left(x_{\gamma_i}\right) \geq \sum_{i=n_0}^{n-1} D_i x_{\gamma_i} \geq k \sum_{i=n_0}^{n-1} D_i. \tag{17}$$

Passing $n \to \infty$ we get the contradiction with the negativity of w. Thus, the system (S) does not have solutions of types (b) or (d).

If (x, y, z, w) is a solution of type (a), then using the same argument as in the previous and by (17) we get $w_n \to \infty$ for $n \to \infty$. From this fact and Lemma 5, we get that all solutions of type (a) satisfy (2).

If (x, y, z, w) is a solution of type (c), then there exists $\lim_{n\to\infty} x_n = h$, $h \geq 0$. Suppose $h > 0$, then by the summation of the fourth equation of (S) we get a contradiction with the negativity of w. Therefore, $\lim_{n\to\infty} x_n = 0$. Becuse y is negative and increasing, then there exists $\lim_{n\to\infty} y_n = l \leq 0$. Suppose $l < 0$. Using the summation of the first equation we get that $s_n \to -\infty$ for $n \to \infty$ which gives a contradiction with the boundedness of x. Therefore, $\lim_{n\to\infty} y_n = 0$. From that fact and Lemma 7 we get that all solutions of type (c) satisfy (3). Thus, system (S) has property B. □

By Theorem 2, the systems from Examples 1 and 2 have property B.

Remark 1 In view of Theorem 2, in the sequel, we assume $\sum_{n=n_0}^{\infty} D_n < \infty$.

We want to find conditions for (S) to have property B without satisfying (H4) and (H5). To ensure that we have to exclude solutions of types (b), (d), (e) and (f).

Theorem 3 *Let* (H1)–(H3) *hold. The system (S) has no solution of type (b) if any of the following conditions hold:*

(i)

$$\sum_{i=n_0}^{\infty} D_i \left(\sum_{j=n_0}^{\gamma_i - \sigma - 1} A_j \right) = \infty, \tag{18}$$

(ii)

$$\sum_{i=n_0}^{\infty} D_i \left(\sum_{j=n_0}^{i-1} C_j \right) = \infty. \tag{19}$$

Proof Assume that (x, y, z, w) is a type (b) solution.

(i) Since y is positive and increasing, there exists $k > 0$ such that $y_n \geq k$ for large n. By the summation of the first equation of (S) we get

$$s_n - s_{n_0} = \sum_{i=n_0}^{n-1} A_i f_1(y_i) \geq \sum_{i=n_0}^{n-1} A_i y_i \geq k \sum_{i=n_0}^{n-1} A_i .$$

Taking into account $\lim(1 - p_n) = 1 - P > 0$, there exists $p > 0$ such that $1 - p_n \geq p$, for large n. Using the summation of the fourth equation of (S) and (8) we have

$$w_n - w_{n_0} = \sum_{i=n_0}^{n-1} D_i f_4\left(x_{\gamma_i}\right) \geq \sum_{i=n_0}^{n-1} D_i x_{\gamma_i} \geq \sum_{i=n_0}^{n-1} D_i s_{\gamma_i - \sigma}(1 - p_{\gamma_i}) \geq \tag{20}$$

$$\geq p \sum_{i=n_0}^{n-1} D_i s_{\gamma_i - \sigma} \geq kp \sum_{i=n_0}^{n-1} D_i \left(\sum_{j=n_0}^{\gamma_i - \sigma - 1} A_j \right).$$

Passing $n \to \infty$ we get the contradiction with the negativity of w. Thus, solutions of type (b) do not exist.

(ii) Since $\lim_{n\to\infty} x_n = \infty$, there exists $k > 0$ such that $x_n \geq k$ for large n. By the summation of the fourth equation of (S) and using Lemma 6 we get

$$w_\infty - w_n = \sum_{i=n}^{\infty} D_i f_4\left(x_{\gamma_i}\right) \geq \sum_{i=n}^{\infty} D_i x_{\gamma_i} \geq k \sum_{i=n}^{\infty} D_i ,$$

Using the summation of the third equation of (S) we have

$$z_n - z_{n_0} = \sum_{i=n_0}^{n-1} C_i f_3(w_i) \leq \sum_{i=n_0}^{n-1} C_i w_i ,$$

$$-z_n + z_{n_0} \geq \sum_{i=n_0}^{n-1} C_i (-w_i) \geq k \sum_{i=n_0}^{n-1} C_i \left(\sum_{j=i}^{\infty} D_j \right).$$

Passing $n \to \infty$ and using the change of summation

$$\sum_{i=n_0}^{\infty} C_i \left(\sum_{j=i}^{\infty} D_j \right) = \sum_{i=n_0}^{\infty} D_i \left(\sum_{j=n_0}^{i-1} C_j \right) = \infty,$$

we get the contradiction with the boundedness of z. Thus, solutions of type (b) do not exist.

□

Theorem 4 *Let* (H1)–(H3) *hold. If*

$$\sum_{i=n_0}^{\infty} D_i \left(\sum_{j=n_0}^{\gamma_i+\sigma-1} A_j \right) = \infty \tag{21}$$

holds, then (S) has no solution of type (d).

Proof Assume that (x, y, z, w) is a type (d) solution. Since y is negative and decreasing, there exists $k < 0$ such that $y_n \leq k$ for large n. By the summation of the first equation of (S) we get

$$s_n - s_{n_0} = \sum_{i=n_0}^{n-1} A_i f_1(y_i) \leq \sum_{i=n_0}^{n-1} A_i y_i \leq k \sum_{i=n_0}^{n-1} A_i. \tag{22}$$

Using the summation of the fourth equation of (S) we have

$$w_n - w_{n_0} = \sum_{i=n_0}^{n-1} D_i f_4\left(x_{\gamma_i}\right) \geq \sum_{i=n_0}^{n-1} D_i x_{\gamma_i} \tag{23}$$

From (H2) and (7) we have $s_n \geq x_n - x_{n-\sigma} \geq -x_{n-\sigma}$ for large n. Thus

$$x_n \geq -s_{n+\sigma}. \tag{24}$$

Using this and (23) and (22) we get

$$w_n - w_{n_0} \geq \sum_{i=n_0}^{n-1} D_i \left(-s_{\gamma_i+\sigma}\right) \geq -k \sum_{i=n_0}^{n-1} D_i \left(\sum_{j=n_0}^{\gamma_i+\sigma-1} A_j \right).$$

Passing $n \to \infty$ we get the contradiction with the negativity of w. □

Theorem 5 *Let* (H1)–(H3) *hold. If*

$$\sum_{i=n_0}^{\infty} D_i \left(\sum_{j=n_0}^{\gamma_i-\sigma-1} A_j \left(\sum_{k=j}^{\infty} B_k \right) \right) = \infty \tag{25}$$

holds, then (S) has no solution of type (e).

Proof Assume that (x, y, z, w) is a type (e) solution. Since z is negative and decreasing, there exists $h < 0$ such that $z_n \le h$ for large n. By the summation of the second equation of (S) we get

$$y_\infty - y_n = \sum_{i=n}^{\infty} B_i f_2(z_i) \le \sum_{i=n}^{\infty} B_i z_i \le h \sum_{i=n}^{\infty} B_i.$$

Using the summation of the first equation of (S) we get

$$s_n - s_{n_0} = \sum_{i=n_0}^{n-1} A_i f_1(y_i) \ge \sum_{i=n_0}^{n-1} A_i y_i \ge -h \sum_{i=n_0}^{n-1} A_i \left(\sum_{j=i}^{\infty} B_j \right). \tag{26}$$

In case $p_n \le 0$, we get $s_n \le x_n$ from (7). Using this fact, the summation of the fourth equation of (S) and the estimation (26) we obtain

$$w_n - w_{n_0} = \sum_{i=n_0}^{n-1} D_i f_4\left(x_{\gamma_i}\right) \ge \sum_{i=n_0}^{n-1} D_i x_{\gamma_i} \ge \sum_{i=n_0}^{n-1} D_i s_{\gamma_i},$$

$$w_n - w_{n_0} \ge -h \sum_{i=n_0}^{n-1} D_i \left(\sum_{j=n_0}^{\gamma_i - 1} A_j \left(\sum_{k=j}^{\infty} B_k \right) \right). \tag{27}$$

In case $p_n > 0$ we use (8) and we get

$$w_n - w_{n_0} \ge \sum_{i=n_0}^{n-1} D_i x_{\gamma_i} \ge \sum_{i=n_0}^{n-1} D_i s_{\gamma_i - \sigma}(1 - p_{\gamma_i}),$$

$$w_n - w_{n_0} \ge -h(1 - P) \sum_{i=n_0}^{n-1} D_i \left(\sum_{j=n_0}^{\gamma_i - \sigma - 1} A_j \left(\sum_{k=j}^{\infty} B_k \right) \right). \tag{28}$$

Passing $n \to \infty$ we get the contradiction with the negativity of w in both cases (27), (28). □

Theorem 6 *Let* (H1)–(H3) *hold. If*

$$\sum_{i=n_0}^{\infty} B_i \left(\sum_{j=n_0}^{i-1} C_j \right) = \infty \tag{29}$$

holds, then (S) has no solution of type (f).

Proof Assume that (x, y, z, w) is a type (f) solution. Since w is positive and increasing, there exists $k > 0$ such that $w_n \geq k$ for large n. By the summation of the third equation of (S) we get

$$z_n - z_{n_0} = \sum_{i=n_0}^{n-1} C_i f_3(w_i) \geq \sum_{i=n_0}^{n-1} C_i w_i \geq k \sum_{i=n_0}^{n-1} C_i.$$

Using the summation of the second equation of (S) we have

$$y_n - y_{n_0} = \sum_{i=n_0}^{n-1} B_i f_2(z_i) \geq \sum_{i=n_0}^{n-1} B_i z_i \geq k \sum_{i=n_0}^{n-1} B_i \left(\sum_{j=n_0}^{i-1} C_j \right).$$

Passing $n \to \infty$ we get the contradiction with the negativity of y. □

If we observe conditions for the nonexistence of solutions of type (b), (d) and (e), then from the fact σ is a nonnegative integer and using the limit comparison criterion for series we get the following conclusion.

Remark 2 If the condition (18) holds, then conditions (21) and (25) hold too.

If we combine the conditions from previous theorems we get the sufficient conditions for system (S) to have weak property B.

Theorem 7 *Let* (H5) *hold. The system (S) has weak property B if one of these conditions hold*

(i) $0 \leq P < 1$ *and* (18),
(ii) $0 \leq P < 1$ *and* (19),
(iii) $-1 < P < 0$, (18) *and* (21),
(iv) $-1 < P < 0$, (19) *and* (21).

If we assume system (S) without conditions (H4) and (H5) we can use the following theorem.

Theorem 8 *Let* (18)*, and* (29) *hold. In addition, if*

$$\sum_{i=n_0}^{\infty} B_i \left(\sum_{j=i}^{\infty} C_j \right) = \infty \tag{30}$$

and

$$\sum_{i=n_0}^{\infty} C_i \left(\sum_{j=n_0}^{i-1} D_j \right) = \infty \tag{31}$$

hold, then the system (S) has weak property B.

Proof By Theorems 3, 4, 5 and 6 the system (S) does not have solutions of type (b), (d), (e) and (f). We prove that solutions of the following types do not exist.
type (i) $x_n > 0$ $y_n > 0$ $z_n < 0$ $w_n > 0$ for large n,
type (ii) $x_n > 0$ $y_n < 0$ $z_n < 0$ $w_n > 0$ for large n
Let (x, y, z, w) be a solution of type (i). Thus, w is positive increasing and there exists a constant $t > 0$ such that $w_n \geq t$ for large n. From the third equation of (S) we get

$$z_\infty - z_n \geq \sum_{i=n}^{\infty} C_i f_3(w_i) \geq \sum_{i=n}^{\infty} C_i w_i \geq t \sum_{i=n}^{\infty} C_i.$$

Substituting this into the second equation of (S) we obtain

$$y_n - y_{n_0} = \sum_{i=n_0}^{n-1} B_i f_2(z_i) \leq \sum_{i=n_0}^{n-1} B_i z_i \leq -t \sum_{i=n_0}^{n-1} B_i \left(\sum_{j=i}^{\infty} C_j \right).$$

Passing $n \to \infty$ we get the contradiction with the positivity of y.

Let (x, y, z, w) be a solution of type (ii). Thus, y is negative and decreasing and there exists a constant $k < 0$ such that $y_n \leq k$ for large n. From the summation of the first equation of (S) we get

$$s_n - s_{n_0} = \sum_{i=n_0}^{n-1} A_i f_1(y_i) \leq \sum_{i=n_0}^{n-1} A_i y_i \leq k \sum_{i=n_0}^{n-1} A_i.$$

Using the summation of the fourth equation of (S) and the estimation (24) we get

$$w_n - w_{n_0} \geq \sum_{i=n_0}^{n-1} D_i x_{\gamma_i} \geq \sum_{i=n_0}^{n-1} D_i (-s_{\gamma_i+\sigma}) \geq -k \sum_{i=n_0}^{n-1} D_i \left(\sum_{j=n_0}^{i-1} A_j \right).$$

Substituting this into the summation of the third equation of (S) we have

$$z_n - z_{n_0} = \sum_{i=n_0}^{n-1} C_i f_3(w_i) \geq \sum_{i=n_0}^{n-1} C_i w_i \geq -k \sum_{i=n_0}^{n-1} C_i \left(\sum_{j=n_0}^{i-1} D_j \left(\sum_{l=n_0}^{j-1} A_l \right) \right).$$

Passing $n \to \infty$ we get the contradiction with the negativity of z.
Therefore, the system (S) has only solutions of type (a) and (c). □

The following theorem gives us the conditions for system (S) to have property B when the system is not in the canonical form.

Theorem 9 *Let* (18)*,* (29)*,* (30) *and* (31) *hold. In addition, if*

$$\sum_{i=n_0}^{\infty} A_i \left(\sum_{j=n_0}^{i-1} B_j \right) = \infty \tag{32}$$

and

$$\sum_{i=n_0}^{\infty} A_i \left(\sum_{j=i}^{\infty} B_j \right) = \infty \tag{33}$$

hold, then the system (S) has property B.

Proof By Theorem 8, the system (S) has only solutions of type (a) and (c). First, assume that (x, y, z, w) is a solution of type (a). Thus, w is positive increasing and there exists a constant $k_1 > 0$ such that $w_n \geq k_1$ for large n. From the third equation of (S) we get

$$z_n - z_{n_0} = \sum_{i=n_0}^{n-1} C_i f_3(w_i) \geq \sum_{i=n_0}^{n-1} C_i w_i \geq k_1 \sum_{i=n_0}^{n-1} C_i.$$

Substituting this into the second equation of (S) we obtain

$$y_n - y_{n_0} = \sum_{i=n_0}^{n-1} B_i f_2(z_i) \geq \sum_{i=n_0}^{n-1} B_i z_i \geq k_1 \sum_{i=n_0}^{n-1} B_i \left(\sum_{j=n_0}^{i-1} C_j \right).$$

Passing $n \to \infty$ we have $y_n \to \infty$.

Using the same argument, there exists a constant $k_2 > 0$ such that $z_n \geq k_2$ for large n. From the summation of the first and the second equation of (S) we get

$$s_n - s_{n_0} = \sum_{i=n_0}^{n-1} A_i f_1(y_i) \geq k_2 \sum_{i=n_0}^{n-1} A_i \left(\sum_{j=n_0}^{i-1} B_j \right).$$

Passing $n \to \infty$ we have $s_n \to \infty$. By Lemma 2, $x_n \to \infty$ too.

There exists a constant $k_3 > 0$ such that $x_n \geq k_3$ for large n. From the summation of the third and the fourth equation of (S) we obtain

$$z_n - z_{n_0} = \sum_{i=n_0}^{n-1} C_i f_3(w_i) \geq k_3 \sum_{i=n_0}^{n-1} C_i \left(\sum_{j=n_0}^{i-1} D_j \right).$$

Passing $n \to \infty$ we have $z_n \to \infty$.

Since y is positive and increasing, there exists $k > 0$ such that $y_n \geq k$ for large n. From (20) we get $w_n \to \infty$ passing $n \to \infty$.

Now, assume that (x, y, z, w) is a solution of type (c). Because w is negative and increasing, there exists $\lim w_n = t_1 \leq 0$. First, assume that $t_1 < 0$. Using the summation of the second and third equation of (S) we get

$$y_n - y_{n_0} = \sum_{i=n_0}^{n-1} B_i f_2(z_i) \geq -t_1 \sum_{i=n_0}^{n-1} B_i \left(\sum_{j=i}^{\infty} C_j \right).$$

Passing $n \to \infty$ we get the contradiction with the boundedness of y, therefore $\lim_{n\to\infty} w_n = 0$.

Now, assume that $\lim x_n = t_2 \geq 0$. First, assume that $t_2 > 0$. Using the summation of the third and the fourth equation of (S) we have

$$z_n - z_{n_0} = \sum_{i=n_0}^{n-1} C_i f_3(w_i) \leq -t_2 \sum_{i=n_0}^{n-1} C_i \left(\sum_{j=i}^{\infty} D_j \right).$$

Passing $n \to \infty$ we get the contradiction with the boundedness of z, therefore $\lim_{n\to\infty} x_n = 0$.

Now, assume that $\lim y_n = t_3 \leq 0$. First, assume that $t_3 < 0$. Using the summation of the first and the fourth equation of (S) we obtain

$$w_n - w_{n_0} = \sum_{i=n_0}^{n-1} D_i f_4(x_{\gamma_i}) \geq \sum_{i=n_0}^{n-1} D_i x_{\gamma_i} \geq \sum_{i=n_0}^{n-1} D_i (-s_{\gamma_i+\sigma}) \geq -t_3 \sum_{i=n_0}^{n-1} D_i \left(\sum_{j=n_0}^{\gamma_i+\sigma-1} A_j \right).$$

Passing $n \to \infty$ we get the contradiction with the boundedness of w, therefore $\lim_{n\to\infty} y_n = 0$.

Finally, assume that $\lim z_n = t_4 \geq 0$. First, assume that $t_4 > 0$. Using the summation of the first and the second equation of (S) we get

$$s_n - s_{n_0} = \sum_{i=n_0}^{n-1} A_i f_1(y_i) \leq -t_4 \sum_{i=n_0}^{n-1} A_i \left(\sum_{j=i}^{\infty} B_j \right).$$

Because x is bounded, then s is bounded as well. Passing $n \to \infty$ we get the contradiction with the boundedness of s, therefore $\lim_{n\to\infty} z_n = 0$.

Now, we get the assertion by Definition 4. □

5 Concluding Remarks

We extend our results from [7] for the system (S) which is not in the canonical form and for the system with negative sequence p_n. Now, results of this paper may be extended with some simplification of conditions for (S) to have property B.

Acknowledgements This paper was processed with contribution of long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

References

1. Diblík, J. et. al: Bounded and unbounded non-oscillatory solutions of a four-dimensional nonlinear neutral difference system. Adv. Differ. Equ. **319**, 1–11 (2015)
2. Došlá, Z., Krejčová, J.: Nonoscillatory solutions of the four-dimensional difference system. Electron. J. Qual. Theory Differ. Equ. **4**, 1–11 (2011)
3. Došlá, Z., Krejčová, J.: Oscillation of a class of the fourth-order nonlinear difference equations. Adv. Differ. Equ. **99**, 1–14 (2012)
4. Došlá, Z., Krejčová, J.: Asymptotic and oscillatory properties of the fourth-order nonlinear difference equations. Appl. Math. Comput. **249**, 164–173 (2014)
5. Došlá, Z., Krejčová, J., Marini, M.: Minimal and maximal solutions of the fourth-order nonlinear difference equations. Differ. Equ. Appl. **20**(12), 1–14 (2014)
6. Jankowski, R., Schmeidel, E., Zonenberg, J.: Oscillatory properties of solutions of the fourth order difference equations with quasidifferences. Opusc. Math. **34**(4), (2014)
7. Krejčová, J.: Nonoscillatory solutions of the four-dimensional neutral difference system. Differ. Differ. Equ. Appl. **164**, 215–224 (2016)
8. Kiguradze, I.T., Chanturia, A.: Asymptot. Prop. Solut. Nonautonomous Ordinary Differ. Equ. Kluwer Acad. Publ. G, Dordrecht (1993)
9. Migda, M., Schmeidel, E.: Convergence of solutions of higher order neutral difference equations with quasi-differences. Tatr. Mt. Math. Publ. **63**, 205–213 (2015)
10. Thandapani, E., Vijaya, M.: Oscillatory and asymptotic behavior of fourth order quasilinear difference equations. Electron. J. Qual. Theory Differ. Equ. **64**, 1–15 (2009)

On the Structure of Polyhedral Positive Invariant Sets with Respect to Delay Difference Equations

Mohammed-Tahar Laraba, Sorin Olaru and Silviu-Iulian Niculescu

Abstract This chapter is dedicated to the study of the positive invariance of polyhedral sets with respect to dynamical systems described by discrete-time delay difference equations (DDEs). Set invariance in the original state-space, also referred to as $\mathscr{D}$-invariance, leads to conservative definitions due to its delay independent property. This limitation makes the $\mathscr{D}$-invariant sets only applicable to a limited class of systems. However, there exists a degree of freedom in the state-space transformations which can enable the positive invariant set-characterizations. In this work we revisit the set factorizations and extend their use in order to establish flexible set-theoretic analysis tools. With linear algebra structural results, it is shown that similarity transformations are a key element in the characterization of low complexity invariant sets within the class of convex polyhedral candidates. In short, it is shown that we can construct, in a low dimensional state-space, an invariant set for a dynamical system governed by a delay difference equation. The basic idea which enables the construction is a simple change of coordinates for the DDE. The obtained $\mathscr{D}$-invariant set exists in the new coordinates even if its existence necessary conditions are not fulfilled in the original state-space. This proves that the $\mathscr{D}$-invariance notion is dependent on the state-space representation of the dynamics. It is worth to recall as a term of comparison that the positive invariance for delay-free dynamics is independent of the state-space realization.

Keywords Linear delay difference equations · Positive invariance · Polyhedral sets

M.-T. Laraba (✉) · S. Olaru · S.-I. Niculescu
Laboratory of Signals and Systems of CentraleSupelec,
91192 Gif-sur-Yvette, France
e-mail: mohammed.laraba@l2s.centralesupelec.fr

S. Olaru
e-mail: sorin.olaru@l2s.centralesupelec.fr

S.-I. Niculescu
e-mail: silviu.niculescu@l2s.centralesupelec.fr

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_10

1 Introduction

Positive invariance is an essential concept with a wide range of applications in dynamical systems and control theory [1–3]. It serves as a basic tool in many control schemes such as model predictive control [4], fault tolerant control [5] and reference governor design [6].

Two popular constructions of positive invariant sets for delay difference equations exist. The first approach, referred to as Krasovskii approach, makes use of the fact that, in discrete-time framework, a finite dimensional extended state-space model can be constructed. The study is then simplified in the case of linear discrete time systems as long as the difficulties related to the infinite dimensionality of the state-space are avoided. A higher dimensional linear time invariant (LTI) system is obtained, its dimension is finite but is in direct relation with the delay value. The equivalent linear time invariant model provides an invariant set for the delay difference equation [7]. However, this approach suffers from an increased computational complexity with the delay's size and becomes impracticable when delays are relatively large. Hence, an alternative approach for the construction of invariant sets for DDEs referred to as Razumikhin approach and denoted as $\mathscr{D}$-invariance has been considered [8]. This approach has been formulated to obtain an invariant set for the DDE in the original state-space, which is independent from the delay value. Iterative procedures for the construction of $\mathscr{D}$-invariant sets as well as the relationship between time-varying DDE stability and $\mathscr{D}$-invariance were presented in [9–11]. However, the concept of $\mathscr{D}$-invariance is often conservative as long as the existence conditions are restrictive.

Recently, it has been recognized that $\mathscr{D}$-invariance can be seen from the geometrical point of view as a factorization of invariant set in the extended state-space [12]. It has been established that the invariance in the extended state-space corresponds to a minimal factorization while $\mathscr{D}$-invariance, under the constraints imposed by the dimension of the DDE, represents the maximal regular ordered factorization. This interesting result opens the way for factorizations which are in between the two representations by exploiting non-minimal state-space equations.

In this chapter, the link between the Razumikhin approach and the Krasovskii approach will be revisited using set factorization. The proposed framework yields a fitting trade-off between the conceptual generality of the extended state-space approach and the computational convenience of the $\mathscr{D}$-invariance approach. We show that $\mathscr{D}$-invariance, which can be seen as set factorization of an invariant set in the extended state-space, represents a particular realization of a broader family of invariant structures. The relationship between these families of invariant sets is established via set factorization and conjugacy.

After establishing the general result, a numerical example will be detailed for illustration. Therein, a dynamical system with a maximum delay equal to 2 and a state-space representation of dimension 2 will be studied. For this delay difference equation, the necessary conditions for the existence of $\mathscr{D}$-invariant sets proposed in [13] are not fulfilled. However, we propose a simple similarity transformation, which leads to a regular ordered factorization of the extended-state invariant set, and thus, allows the construction of a $\mathscr{D}$-invariant set in the novel basis.

The chapter is organized as follows. Section 2 presents some preliminary definitions and the existing results of [12] are recalled. In Sect. 3, the problem of existence and uniqueness of similarity transformations for the construction of $\mathscr{D}$-invariant sets is addressed. The shape of the similarity transformation which allows a regular ordered factorization is established in Sect. 3.1. In Sect. 4, a numerical example is given to illustrate the previous results and finally Sect. 5 draws concluding remarks.

2 Preliminary Definitions and Existing Results

Let us consider the following delay difference equation:

$$x(k) = \sum_{i=1}^{d} A_i x(k-i) \tag{1}$$

where $x(k) \in \mathbb{R}^n$ is the state vector at the time $k \in \mathbb{Z}_+$. Matrices $A_i \in \mathbb{R}^{n\times n}$ for $i = 1, \cdots, d$. For every interval Π of $\mathbb{R}_+$ we define $\mathbb{R}_\Pi := \mathbb{R} \cap \Pi$. The initial conditions are considered to be given by $x(-i) \in \mathbb{R}^n$, for $i \in \mathbb{Z}_{[1,d]}$. Given two sets $\mathscr{X}, \mathscr{Y} \subset \mathbb{R}^m$, $\mathscr{X} \oplus \mathscr{Y}$ and $\mathscr{X} \times \mathscr{Y}$ denote the Minkowski sum and the Cartesian product of these two sets, respectively:

$$\mathscr{X} \oplus \mathscr{Y} := \{z |\, \exists (x, y) \in (\mathscr{X}, \mathscr{Y}) \text{ such that } z = x + y\},$$

$$\mathscr{X} \times \mathscr{Y} := \{(x, y) |\, x \in \mathscr{X} \text{ and } y \in \mathscr{Y}\}.$$

An extended state-space representation can be constructed for any given (finite) delay realization. Using an augmented state vector

$$z(k) = \left[x(k)^T \cdots x(k-d+1)^T\right]^T$$

equation (1) can be rewritten as:

$$z(k) = A_e z(k-1) = \begin{bmatrix} A_1 & \dots & A_{d-1} & A_d \\ I & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix} z(k-1), \tag{2}$$

Definition 1 A set $\mathscr{P} \subseteq \mathbb{R}^n$ is called $\mathscr{D}$-invariant for the system (1) with initial conditions $x(-i) \in \mathscr{P}$ for all $i \in \mathbb{Z}_{[1,d]}$ if the state trajectory satisfies $x(k) \in \mathscr{P}, \forall k \in \mathbb{Z}_+$. ■

As already mentioned in the introduction, two main approaches exist in the literature dealing with positive invariant sets for discrete-time delay difference equations;

the invariant set in an extended state-space (2) on one side and the invariant set in the original state-space (1) (also called $\mathscr{D}$-invariant set) on the other side. The concept of cyclic invariance [14] proposes instead of a rigid set in $\mathbb{R}^{nd}$ or $\mathbb{R}^n$ a family of invariant sets and offers a certain degree of flexibility.

Definition 2 A family of (d tuples of) sets $\{\Omega_1, \dots, \Omega_d\}$ is called cyclic $\mathscr{D}$-invariant with respect to (1) if:

$$\begin{aligned} A_1\Omega_1 \oplus A_2\Omega_2 \oplus \cdots \oplus A_d\Omega_d \subseteq \Omega_d; \\ A_1\Omega_{d_m} \oplus A_2\Omega_1 \oplus \cdots \oplus A_d\Omega_{d-1} \subseteq \Omega_{d-1}; \\ \vdots \\ A_1\Omega_2 \oplus A_2\Omega_3 \oplus \cdots \oplus A_d\Omega_1 \subseteq \Omega_1. \end{aligned} \tag{3}$$

A generalization of the cyclic invariance notion to invariant family of sets was proposed by [15].

Definition 3 A family of (d tuples of) sets $\mathscr{F} \subset \mathbb{R}^{nd}$ is an invariant family with respect to (1) if for any tuple $\{\Omega_1, \Omega_2, \dots, \Omega_d\} \in \mathscr{F}$ there exists a set $\Omega_0 \subset \mathbb{R}^n$ such that $\{\Omega_0, \Omega_1 \dots, \Omega_{d-1}\} \in \mathscr{F}$ and

$$A_1\Omega_1 \oplus A_2\Omega_2 \oplus \cdots \oplus A_d\Omega_d \subseteq \Omega_0$$

The link between the two main representations for discrete-time delay difference equations and their invariant sets has received recently a unifying characterization via set factorization [12]. The reader interested in a more thorough introduction to set factorization may consult [16, 17]. Next, the basic notions in this respect are recalled.

Definition 4 A partition of a set of indices $P \subset \mathbb{Z}_{[1,m]}$ is the family of $'l'$ subsets P_k of P, which verify the following conditions:

- $\emptyset \notin \{P_k\}_{k=1}^{l}$,
- The subsets $\{P_k\}_{k=1}^{l}$ are said to cover P $i.e.$ $P = \bigcup_{i=1}^{l} P_i$,
- The elements of $\{P_k\}_{k=1}^{l}$ are pairwise disjoint ($P_i \cap P_j = \emptyset$ for $i \neq j$).

Given a subset $P_i \subset \mathbb{Z}_{[1,m]}$ and a set $\Omega \in \mathbb{R}^m$, $\Omega_{\downarrow P_i}$ denotes the projection of the set Ω on the subset of $\mathbb{R}^n$ with indices of Cartesian coordinates in P_i.

Definition 5 Let $\Omega \in \mathbb{R}^m$ and $\bigcup_{i=1}^{l} P_i$ be the partition of $\mathbb{Z}_{[1,m]}$

1. The set Ω *is factorized* according to the partition $\bigcup_{i=1}^{l} P_i = \mathbb{Z}_{[1,m]}$ if:

$$\Omega = \Omega_{\downarrow P_1} \times \cdots \times \Omega_{\downarrow P_l} \tag{4}$$

2. A *set factorization* (4) is said to be *balanced* if:

$$\text{card}\{P_1\} = \cdots = \text{card}\{P_l\}$$

3. A factorization is said to be *ordered* if it is defined by an *ordered* partition $P = \bigcup_{k=1}^{l} P_k$ satisfying:

$$\max\{P_i\} < \min\left\{P_j\right\},\ \forall i < j \in \mathbb{Z}_{[1,m]}, \tag{5}$$

4. A factorization is *regular* if is characterized by the equivalence of the factors

$$\Omega_{\downarrow P_1} = \cdots = \Omega_{\downarrow P_l} = S, \tag{6}$$

and

$$\Omega = \underbrace{S \times S \times \cdots \times S}_{l \text{ times}}. \tag{7}$$

Most of the factorization properties are related to the Cartesian product operation. It is clear that the set factorization is non-commutative. The exception is represented by the regular factorization which is commutative inside the given partition. Additionally, the regular factorizations are balanced but not necessarily ordered. From the structural point of view, the geometry of the factors is related to the geometry of the initial set. Convexity, for example, of a given set Ω implies the convexity of the factors. It is worth to be mentioned that the Cartesian product of several polyhedra is a polyhedron of higher dimension. It becomes clear that the polyhedral sets represent an interesting class of sets which can be used for the development in relationship with set factorization. In comparison, even if the projection of ellipsoidal sets is ellipsoidal, the Cartesian product of ellipsoids is not an ellipsoid rendering the ellipsoidal class of sets impracticable for set factorization despite the fact that they represent usual candidates for positive invariance with respect to linear time-invariant dynamics. The property of polyhedral factorization is recalled in the next proposition:

Proposition 1 *There exists a regular ordered factorization for a polyhedral set:*

$$\Omega = \left\{x \in \mathbb{R}^n : Fx \leq w\right\} \tag{8}$$

described by its minimal half space representation, if there exists a block diagonalization of the matrix F via a column permutation. ■

The relationship between $\mathcal{D}$-invariance and invariance in the extended state-space is formally stated in the next theorem without proof for brevity.

Theorem 1 *The system (1) admits a convex $\mathscr{D}$-invariant set if and only if there exists an invariant set for the system:*

$$z(k) = \begin{bmatrix} A_1 & \dots & A_{d-1} & A_d \\ I & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix} z(k-1), \tag{9}$$

which admits a regular ordered factorization. ■

Proposition 2 *Let*

$$\Omega = \left\{ x \in \mathbb{R}^{nd} | Fx \leq w \right\} \tag{10}$$

be an invariant set with respect to the system (2). A regular ordered factorization with dimension-n factors exists if there exists a transformation matrix $T \in \mathbb{R}^{(nd)\times(nd)}$ such that:

$$FT^{-1} = \begin{bmatrix} F_1 & 0 & \dots & 0 \\ 0 & F_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & F_d \end{bmatrix} \tag{11}$$

■

Corollary 1 *Let a delay difference equation be described by (1). There exists a $\mathscr{D}$-invariant set for this dynamical system in $\mathbb{R}^n$ if the following conditions are fulfiled:*

- *There exists a similarity transformation matrix T such that*

$$\begin{bmatrix} B_1 & \dots & B_{d-1} & B_d \\ I & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix} = T \begin{bmatrix} A_1 & \dots & A_{d-1} & A_d \\ I & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix} T^{-1} \tag{12}$$

- *There exists an invariant set with respect to the system*

$$\tilde{z}(k) = \begin{bmatrix} B_1 & \dots & B_{d-1} & B_d \\ I & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & I & 0 \end{bmatrix} \tilde{z}(k-1) \tag{13}$$

which admits a regular ordered factorization. ■

3 Existence and Uniqueness of Similarity Transformations in the Construction of $\mathscr{D}$-invariant Sets

We introduce in this section similarity transformations in the construction of $\mathscr{D}$-invariant sets all by preserving the dynamical model in the form of a delay difference equation in $\mathbb{R}^n$. Sylvester equations play a central role in many areas of applied mathematics and in particular in systems and control theory. Before introducing formally these equations, we need to introduce first Schur's lemma.

Lemma 1 *If A is a square $n \times n$ matrix, then A can be expressed as $A = QUQ^*$. Where Q^* is the trans-conjugate of the unitary matrix Q ($Q^{-1} = Q^*$); U is an upper triangular matrix (Schur form), containing the eigenvalues of A on its diagonal.* ■

Let us consider the equation $AX+XB = C$ where $A \in M_n$, $B \in M_m$ and $C \in M_{n\times m}$, where M_n denotes the set of square $n\times n$ matrices, and $M_{n\times m}$ denotes the set of $n\times m$ matrices. The Sylvester equation can be written in the form $(I_m \otimes A + B^T \otimes I_n) * Vect(X) = Vect(C)$, where $Vect(X)$ is the vertical concatenation of the columns of the matrix X, $I_{(.)}$ is the identity matrix. $\otimes$ denotes the Kronecker product of two matrices. The spectrum of a matrix $A \in M_n$ is the set of the eigenvalues of A, denoted by $\lambda(A)$, while the spectral radius is defined as $\rho(A) := \max\limits_{\xi\in\lambda(A)} (|\xi|)$. It is clear that $I_m \otimes A$ and $B^T \otimes I_n$ are two matrices belonging to M_{mn}.

Theorem 2 *If $\eta \in \lambda(A)$ and $v \in \mathbb{C}^n$ is the corresponding eigenvector of A, and if $\mu \in \lambda(B^T)$ and $w \in \mathbb{C}^m$ is the corresponding eigenvector of B^T, then $\eta + \mu$ is an eigenvalue of $I_m \otimes A + B^T \otimes I_n$ with $w \otimes v$ its corresponding eigenvector.*
Furthermore, if $\lambda(A) = \{\eta_1, \eta_2, ..., \eta_n\}$ and $\lambda(B^T) = \{\mu_1, \mu_2, ..., \mu_m\}$, then:

$$\lambda(I_m \otimes A + B^T \otimes I_n) = \left\{\eta_i + \mu_j; i \in \{1, \cdots, n\}, j \in \{1, \cdots, m\}\right\}.$$

Proof A and B^T are two square matrices of dimensions n and m respectively, their Schur decomposition is:

$$A = Q_A U_A Q_A^*,\ B^T = Q_{B^T} U_{B^T} Q_{B^T}^*, \tag{14}$$

where:

$$Q_A^* Q_A = I_n,\ Q_{B^T}^* Q_{B^T} = I_m. \tag{15}$$

It follows from (14) and (15) that:

$$\begin{cases} U_A = Q_A^* A Q_A \\ U_{B^T} = Q_{B^T}^* B^T Q_{B^T} \end{cases} \tag{16}$$

U_A and U_{B^T} in equation (16) are two upper triangular matrices. Let us now assume that W is the Kronecker product $W = Q_{B^T} \otimes Q_A \in M_{mn}$, then:

$$\begin{aligned} W^*W &= (Q_{B^T} \otimes Q_A)^*(Q_{B^T} \otimes Q_A) = (Q^*_{B^T} \otimes Q^*_A)(Q_{B^T} \otimes Q_A) \\ &= (Q^*_{B^T} Q_{B^T}) \otimes (Q^*_A Q_A) = I_{mn}. \end{aligned}$$

It holds also:

$$\begin{aligned} W^*(I_m \otimes A)W &= (Q^*_{B^T} \otimes Q^*_A)(I_m \otimes Q_A U_A Q^*_A)(Q_{B^T} \otimes Q_A) \\ &= (Q^*_{B^T} \otimes U_A Q^*_A)(Q_{B^T} \otimes Q_A) = (I_m \otimes U_A). \end{aligned} \tag{17}$$

and

$$\begin{aligned} W^*(B^T \otimes I_n)W &= (Q^*_{B^T} \otimes Q^*_A)(Q_{B^T} U_{B^T} Q^*_{B^T} \otimes I_n)(Q_{B^T} \otimes Q_A) \\ &= (U_{B^T} Q^*_{B^T} \otimes Q^*_A)(Q_{B^T} \otimes Q_A) = (U_{B^T} \otimes I_n). \end{aligned} \tag{18}$$

From the elements provided above, it becomes clear that:

$$\begin{aligned} W^*(I_m \otimes A + B^T \otimes I_n)W &= I_m \otimes U_A + U_{B^T} \otimes I_n \\ &= \begin{pmatrix} U_A & 0 & 0 & 0 \\ 0 & U_A & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & U_A \end{pmatrix} + \begin{pmatrix} \mu_1 I_n & * & * & * \\ 0 & \mu_2 I_n & * & * \\ 0 & 0 & \ddots & * \\ 0 & 0 & 0 & \mu_m I_n \end{pmatrix} \\ &= \begin{pmatrix} \mu_1 I_n + U_A & * & * & * \\ 0 & \mu_2 I_n + U_A & * & * \\ 0 & 0 & \ddots & * \\ 0 & 0 & 0 & \mu_m I_n + U_A \end{pmatrix} \end{aligned}$$

One can see that the diagonal elements of the resultant upper triangular matrix contain all sum pairs of eigenvalues of U_A, U_{B^T} which completes the proof. ■

Remark 1 If for some i and j, $\eta_i + \mu_j = 0$, then rank of $(I_m \otimes A + B^T \otimes I_n)$ is strictly less than nm, then the solution for the system $(I_m \otimes A + B^T \otimes I_n)Vect(X) = C$ is not unique.

Theorem 3 *The equation* $AX + XB = C$ *has a unique solution* $X \in M_{n\times m}$ *if and only if* $\lambda(A) \cap \lambda(-B) = \emptyset$

Proof Follows directly from Theorem 2. ■

In this work, we are interested in the similarity transformation as an auxiliary tool for the construction of $\mathscr{D}$-invariant sets while preserving the dynamical model in the form of a delay difference equation in $\mathbb{R}^n$, i.e. starting from an extended state-space model of a given dynamical system, we obtain another extended state-space model, which has the same dimension as the first one, via a simple change of coordinates. Such a similarity transformation represents a parametrization of the conditions for the existence of a regular ordered factorization.

In addition, one can see that solving the problem of determination of a matrix T in Eq. (12) is equivalent to the existence of an invertible matrix T which verifies a particular (homogeneous) Sylvester equation. In Theorem 2 it is shown that the Eq. (12)

can be rewritten as a linear (in our case homogeneous) equation of size $(nd)^2 \times (nd)^2$. This equation has non-trivial solutions if it is singular which is equivalent to matrices A and B having at least one common eigenvalue. It is clear that Eq. (12) represents a similarity transformation and that matrices A and B share the same set of eigenvalues. Subsequently, applying Theorem 3 guarantees that the transformation exists and more than that, it is not unique. In fact, the solution, in vectorial form, is the full null space of the matrix $(I_{nd} \otimes B - A^T \otimes I_{nd})$.

The similarity transformation corresponds to a transformation of the state $\tilde{z} = Tz$, where $\tilde{z}$ and z are the state vectors of the extended state-space realization. This results in several possible canonical forms. Different properties stand out more clearly in different realizations, and some forms may have advantages in some applications (recall for example the controllable and observable canonical forms in classical control theory). It is worth mentioning that most of dynamical properties of an LTI system, such as input-output properties and the impulse response and so on, are not changed by similarity transformations.

Remark 2 In general, algebraic equivalence[1] does not preserve stability properties of a dynamical system [19, 20], and for this a necessary and sufficient condition will be the *topological equivalence*, which is the algebraic condition plus the condition on the Euclidean norm of the matrix T [18].

In our case and since we are working in a time-invariant setting, it follows from [18] that two LTI systems are strictly equivalent whenever their phase vectors are related for all time t as $(t, \tilde{z}) = (t, Tz)$, where T is a nonsingular constant matrix, and obviously, strict equivalence implies topological equivalence.

When dealing with scalar systems (one state only), with simple linear algebra manipulations, it can be shown that the constraint imposed on the similarity transformation is very restrictive and allow only scaling type of change of coordinates on the original delay difference equation, without an impact on the regular ordered factorization.

Example 1 For illustration, let us consider dynamical matrix

$$A = \begin{bmatrix} a_1 & \dots & a_{d-1} & a_d \\ 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{bmatrix}; \tag{19}$$

there exists an infinite number of combinations of similarity transformations T, which satisfy equality (12). All matrices T are generated from the null space of the matrix $(I_d \otimes A - B^T \otimes I_d)$. This null space is the same as the one of $(I_d \otimes A - A^T \otimes I_d)$ since we consider here only scalar systems. For $d = 2$, the null space is generated by the two matrices:

[1]See [18] for a formal definition of *algebraic equivalence*.

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} a_0/a_1 & 1 \\ 1/a_1 & 0 \end{bmatrix} \right\}, \tag{20}$$

while for $d = 3$, the null space is generated by the three matrices:

$$\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} a_0/a_2 & a_1/a_2 & 1 \\ 1/a_2 & 0 & 0 \\ 0 & 1/a_2 & 0 \end{bmatrix}, \begin{bmatrix} a_1/a_2 & 1 & 0 \\ 0 & a_1/a_2 & 0 \\ 1/a_2 & -a_0/a_2 & 0 \end{bmatrix} \right\}. \tag{21}$$

Next we make a step forward towards the study of the structure of all matrices T which allow transformations by maintaining the dynamical model in the class of a DDE in $\mathbb{R}^n$. Specifically, we will be interested in the present work in the case of systems with two states $n = 2$ and a maximum delay $d = 2$ in (1).

Let us consider the extended dynamical system (2) in this case:

$$z(k) = \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} z(k-1) \Leftrightarrow x(k) = A_1 x(k-1) + A_2 x(k-2) \tag{22}$$

After the change of coordinates, in the novel basis:

$$\tilde{z}(k) = \begin{bmatrix} B_1 & B_2 \\ I & 0 \end{bmatrix} \tilde{z}_k = T \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} T^{-1} \tilde{z}(k-1) \tag{23}$$

The relationship between the augmented state in the two basis of coordinates is:

$$\tilde{z}(k) = T z(k) \tag{24}$$

$$\begin{bmatrix} \tilde{x}(k) \\ \tilde{x}(k-1) \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} x_(k) \\ x(k-1) \end{bmatrix} \tag{25}$$

$$\tilde{z}(k) = \begin{bmatrix} B_1 & B_2 \\ I & 0 \end{bmatrix} \tilde{z}(k-1) \Leftrightarrow \tilde{x}(k) = B_1 \tilde{x}(k-1) + B_2 \tilde{x}(k-2) \tag{26}$$

It follows from (25) that:

$$T_{21} x(k) + T_{22} x(k-1) = T_{11} x(k-1) + T_{12} x(k-2) \tag{27}$$

or equivalently:

$$\begin{aligned} T_{21} x(k) &= (T_{11} - T_{22}) x(k-1) + T_{12} x(k-2) \\ x(k) &= A_1 x(k-1) + A_2 x(k-2) \end{aligned} \tag{28}$$

From Eq. (28), we can easily derive the matrix T:

- If $T_{21} = \mathbf{O} \rightarrow T_{11} = T_{22} = T_*$ and $T_{12} = \mathbf{O}$

$$T = \begin{bmatrix} T_* & \mathbf{O} \\ \mathbf{O} & T_* \end{bmatrix} \tag{29}$$

- If $T_{21} = I_{2\times 2} \rightarrow T_{12} = A_2$ and $T_{11} = A_1 + T_{22}$

$$T = \begin{bmatrix} A_1 + T_{22} & A_2 \\ I_{2\times 2} & T_{22} \end{bmatrix} \tag{30}$$

- If $T_{21} \neq I_{2\times 2}$ and is invertible

$$\begin{cases} A_1 = T_{21}^{-1}(T_{11} - T_{22}) \\ A_2 = T_{21}^{-1} T_{12} \end{cases} \Leftrightarrow \begin{cases} T_{11} = T_{22} + T_{21} A_1 \\ T_{12} = T_{21} A_2 \end{cases}$$

$$T = \begin{bmatrix} T_{22} + T_{21} A_1 & T_{21} A_2 \\ T_{21} & T_{22} \end{bmatrix}$$

Remark 3 Note that the first case when $T_{21} = \mathbf{O}$, and the second one $T_{21} = I_{2\times 2}$ represent particular structures of a broader family of matrices T presented in the third case by relaxing the invertibility assumption for T_{21}. Just by setting $T_{21} = \mathbf{O}$ then $T_{21} = I_{2\times 2}$, Eq. (3) takes the form of Eq. (29) then (30) respectively.

All transition matrices for (28) are thus generated by the two matrices:

$$\left\{ \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}, \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} \right\}, \tag{31}$$

Based on the above particular forms, matrix T can be written, using the Kronecker product, in a compact form.

$$T = (I_{2\times 2} \otimes T_{22}) + (I_{2\times 2} \otimes T_{21}) A \tag{32}$$

It is worth noting that square matrices T_{21} and T_{22} can be chosen arbitrarily as long as they lead to an invertible matrix T.

3.1 Transformation Allowing Regular Ordered Factorization

Let us consider the change of coordinates $\tilde{z} = Tz$ applied to the dynamical system (22). For sake of simplicity, we examine in this section transformation of the form: $z = T^{-1}\tilde{z} = \Gamma \tilde{z}$, then Γ can be written as:

$$\Gamma = \begin{bmatrix} \gamma_{22} + \gamma_{21} B_1 & \gamma_{21} B_2 \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \tag{33}$$

which is equivalent to :

$$\Gamma = \left[(I_{2\times 2} \otimes \gamma_{22}) + (I_{2\times 2} \otimes \gamma_{21})B \right] \tag{34}$$

Let $\Omega = \left\{ z \in \mathbb{R}^4 | Fz \leq w \right\}$ be an invariant set in the extended state-space with respect to the dynamical system (22), then $\tilde{\Omega} = \left\{ \tilde{x} \in \mathbb{R}^4 | F\Gamma \tilde{z} \leq w \right\}$ will be the corresponding invariant set for the extended state realization (26).

$$\begin{aligned} F\Gamma &= F\left[(I_{2\times 2} \otimes \gamma_{22}) + (I_{2\times 2} \otimes \gamma_{21})B \right] \\ &= F(I_{2\times 2} \otimes \gamma_{22}) + F(I_{2\times 2} \otimes \gamma_{21})B \end{aligned} \tag{35}$$

$$\tilde{\Omega} = \left\{ \tilde{x} \in \mathbb{R}^4 | \left[F(I_{2\times 2} \otimes \gamma_{22}) + F(I_{2\times 2} \otimes \gamma_{21})B \right] \tilde{x} \leq w \right\} \tag{36}$$

$$F\Gamma = \left[F \begin{bmatrix} \gamma_{22} & 0_{2\times 2} \\ 0_{2\times 2} & \gamma_{22} \end{bmatrix} + F \begin{bmatrix} \gamma_{21} & 0_{2\times 2} \\ 0_{2\times 2} & \gamma_{21} \end{bmatrix} B \right] \tag{37}$$

Let F be:

$$F = \begin{bmatrix} F_1 & F_2 \\ F_3 & F_4 \end{bmatrix} \text{ then } F\Gamma = \begin{bmatrix} F_1(\gamma_{22} + \gamma_{21}B_1) + F_2\gamma_{21} & F_2\gamma_{22} + F_1\gamma_{21}B_2 \\ F_3(\gamma_{22} + \gamma_{21}B_1) + F_4\gamma_{21} & F_4\gamma_{22} + F_3\gamma_{21}B_2 \end{bmatrix}. \tag{38}$$

The necessary and sufficient condition for the existence of a factorization for the invariant set is that the matrix $F\Gamma$ has a lower triangular structure:

$$F\Gamma = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix} \tag{39}$$

which means that there exists γ_{22} and γ_{21} verifying the equality:

$$F_2\gamma_{22} + F_1\gamma_{21}B_2 = 0 \tag{40}$$

Remark 4 It is worth mentioning that the existence of a $\mathscr{D}$-invariant set in the new coordinates provides a region in the state-space which is invariant in the sens that all the state trajectories starting in this set remain inside it in the future. Furthermore, the existence of a $\mathscr{D}$-invariant set in general only implies that the system is stable independently from the delay value. If such a region does not exist for some linear delay difference equation, one can't conclude about the stability of the system (the existence of a $\mathscr{D}$-invariant set is a necessary (not sufficient) condition for the stability of DDEs).

3.2 Implications of the Results in Terms of Control System Concepts

It is well known that the existence of a Lyapunov Function is equivalent to the existence of a λ-contractive set. On the other hand, it was established that the existence of a Lyapunov-Razumikhin Function LRF is equivalent to the existence of a particular type of contractive sets, known as λ-$\mathscr{D}$-contractive sets, for the non-extended model (sub-level sets of a LRF being λ-$\mathscr{D}$-contractive sets). $\mathscr{D}$-invariance is a limit case of λ-$\mathscr{D}$-contractiveness (it would amount to choosing $\lambda = 1$).

Proposition 3 ([7]) *The existence of a $\mathscr{D}$-invariant set is equivalent to the existence of a Lyapunov-Razumikhin Function.*

The implication of the result in this work is stated in the following corollary:

Corollary 2 *The existence of a Lyapunov-Razumikhin Function depends on the DDE's state space representation.*

Proof If there exists a DDE for which the necessary condition for the existence of $\mathscr{D}$-invariant sets is not satisfied, then a LRF does not exist for the same DDE in virtue of Proposition 3. However, considering now a state space transformation which leads to the existence of a $\mathscr{D}$-invariant set, then the existence of a Lyapunov-Razumikhin Function is guaranteed in the novel representation of the DDE, and thus the dependence on the DDE's state realization is proved. ■

3.3 Polytopic State Constraints

We show that the proposed transformation is able to handle polytopic state limitations and guarantee constraints satisfaction. Suppose that the system states are subject to polytopic constraints $x(k) \in \mathscr{X}, \forall k \in \mathbb{Z}_+$ where $\mathscr{X}$ is a compact and convex set which contains the origin as an interior point:

$$\mathscr{X} = \left\{x \in \mathbb{R}^n |\ Fx \le f\right\} \tag{41}$$

It follows that the extended state space vector in (2) is subject to the constraint: $z(k) \in \mathscr{X}_{cst}, \forall k \in \mathbb{Z}_+$ where:

$$\begin{aligned} \mathscr{X}_{cst} &= \mathscr{X} \times \cdots \times \mathscr{X} \\ \mathscr{X}_{cst} &= \left\{ z \in \mathbb{R}^{nd} |\ \begin{bmatrix} F & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & F \end{bmatrix} z \le \begin{bmatrix} f \\ \vdots \\ f \end{bmatrix} \right\} \end{aligned} \tag{42}$$

Suppose now that there exists a similarity transformation of the form (12) allowing regular ordered factorization. Since $\tilde{z} = Tz$, the image of the set $\mathscr{X}_{cst}$ is obtained as:

$$\tilde{\mathscr{X}}_{cst} = \left\{ \tilde{z} \in \mathbb{R}^{nd} | \begin{bmatrix} F & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & F \end{bmatrix} T^{-1}\tilde{z} \leq \begin{bmatrix} f \\ \vdots \\ f \end{bmatrix} \right\} \tag{43}$$

However, there exits an invariant set Ω_{Be} wrt to the DDE in the new coordinates which admits a regular ordered factorization $\Omega_{Be} = \Omega \times \cdots \times \Omega$. Exploiting the scaling property of invariant sets, the maximal scaling factor ensuring constraints satisfaction in the original state space representation is obtained by solving the following optimization problem:

$$\max_{\lambda} \lambda \tag{44a}$$

subject to

$$\lambda > 0 \tag{44b}$$

$$\lambda \Omega_{Be} \subseteq \tilde{\mathscr{X}}_{cst} \tag{44c}$$

A numerical example is given in the following to show the effectiveness of the above results.

4 Illustrative Example

Let us consider the following dynamical system:

$$x(k) = A_1 x(k-1) + A_2 x(k-2), \tag{45}$$

$$A_1 = \begin{bmatrix} -0.5026 & 1.3088 \\ 0.5201 & 0.9026 \end{bmatrix} \tag{46}$$

$$A_2 = \begin{bmatrix} -0.059 & 0.4517 \\ -0.0935 & -0.7510 \end{bmatrix} \tag{47}$$

The necessary condition for the existence of a $\mathscr{D}$-contractive set proposed in [13] is not fulfilled. One can verify that the spectral radius of A_1 is not subunitary, $\rho(A_1) = 1.2837 > 1$. More than that, the necessary condition proposed in [21] is not verified. We can easily verify that the spectral radius of the sum $\rho(A_1 + A_2) = 1.1422 > 1$, and the set of generalized eigenvalues possesses four elements on the unit circle:

$$\gamma(U, V) = 0.4611 \pm 0.8873i, 0.6392 \pm 0.7690i,$$
$$0.144, -0.096, -10.410, 6.943.$$

The delay difference equation (45) does not admit a $\mathscr{D}$-invariant set. Note that the extended state-space representation has a strictly stable transition matrix, which allows the construction of invariant set $\Omega_{A_e} \subset \mathbb{R}^4$.

$$z(k) = \begin{bmatrix} -0.5026 & 1.3088 & -0.059 & 0.4517 \\ 0.5201 & 0.9026 & -0.0935 & -0.7510 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} z(k-1). \tag{48}$$

Let $\Omega_{A_e} = \left\{ z \in \mathbb{R}^4 | Fz \leq w \right\}$ be the extended invariant set with respect to (48). With linear algebra manipulations, we can find a similarity transformation T such that ($\tilde{z} = Tz$), which allows formulation the system $z(k) = Az(k-1)$ in the equivalent form $\tilde{z}(k) = B\tilde{z}(k-1)$, and there exists an invariant set with respect to this last dynamical system which admits a regular ordered factorization in $\mathbb{R}^2$. Let us take the transition matrix $T = (I_{2\times 2} \otimes T_{22}) + (I_{2\times 2} \otimes T_{21})A$,

$$T_{21} = \begin{bmatrix} -28.729 & 3.932 \\ 30.667 & -3.549 \end{bmatrix} \tag{49}$$

$$T_{22} = \begin{bmatrix} -3.161 & 34.342 \\ 4.831 & -37.413 \end{bmatrix}. \tag{50}$$

and

$$T = \begin{bmatrix} 13.322 & 0.289 & 1.326 & -15.930 \\ -12.427 & -0.479 & -1.476 & 16.517 \\ -28.729 & 3.932 & -3.161 & 34.342 \\ 30.667 & -3.549 & 4.831 & -37.413 \end{bmatrix} \tag{51}$$

The dynamical system in the new basis:

$$\tilde{z}(k) = B\tilde{z}(k-1) = TAT^{-1}\tilde{z}(k-1), \tag{52}$$

with

$$B = \begin{bmatrix} 0.20 & -0.34 & 0.24 & -0.17 \\ 0.34 & 0.20 & 0.17 & 0.24 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{53}$$

has a strictly stable transition matrix, which has the same set of eigenvalues as the system (48).

$$\lambda(A) = \lambda(B) = 0.6108 \pm 0.3697i, -0.4108 \pm 0.0297i.$$

This allows the construction of an invariant set $\Omega_{B_e} \subset \mathbb{R}^4$ which is factorizable, then the delay difference equation $\tilde{x}(k) = B_1\tilde{x}(k-1) + B_2\tilde{x}(k-2)$ admits a $\mathscr{D}$-invariant set $\Omega \subset \mathbb{R}^2$ in this novel basis. It can be shown that this particular choice of T_{21} and T_{22} verifies (40).

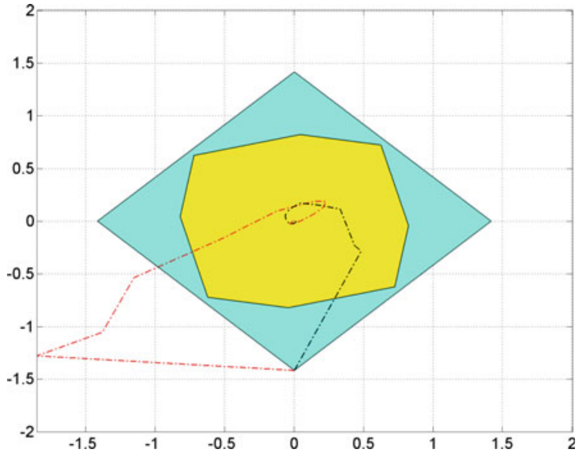


Fig. 1 Set Ω in cyan and the set iteration $B_1\Omega \oplus B_2\Omega$ in yellow

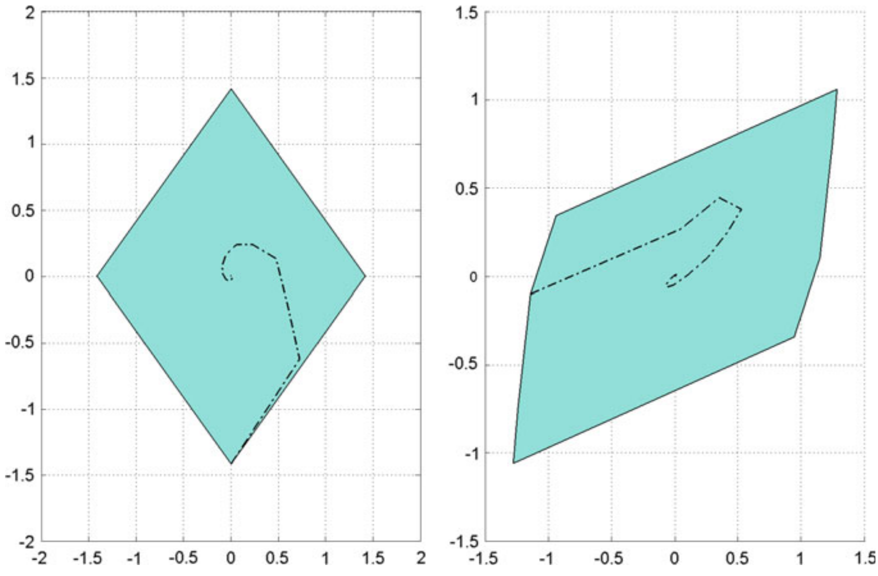


Fig. 2 State trajectories starting from the same initial state and the corresponding $\mathscr{D}$-invariant set (left) and the projection of Ω_{A_e} (right)

$$\Omega = \left\{ \tilde{x} \in \mathbb{R}^2 | \begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & \sqrt{2} \\ \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & -\sqrt{2} \end{bmatrix} \tilde{x} \leq \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} \right\} \tag{54}$$

$$\Omega_{B_e} = \left\{ \tilde{z} \in \mathbb{R}^4 | \begin{bmatrix} \sqrt{2} & -\sqrt{2} & 0 & 0 \\ -\sqrt{2} & \sqrt{2} & 0 & 0 \\ \sqrt{2} & \sqrt{2} & 0 & 0 \\ -\sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \\ 0 & 0 & -\sqrt{2} & \sqrt{2} \\ 0 & 0 & \sqrt{2} & \sqrt{2} \\ 0 & 0 & -\sqrt{2} & -\sqrt{2} \end{bmatrix} \tilde{z} \leq \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} \right\} \tag{55}$$

Figure 1 presents the $\mathscr{D}$-invariant set obtained. Dashed black line represents the state trajectory starting from the initial states $x(-1) = [0, -\sqrt{2}]'$ and $x(-2) = [0, 0]'$ with respect to the DDE $\tilde{x}(k) = B_1\tilde{x}(k-1) + B_2\tilde{x}(k-2)$. However, dashed red line represents the state trajectory starting from the same initial states with respect to the DDE (45).

We can see that the state trajectory in the original basis does not remain inside the blue set. It follows that this set is not $\mathscr{D}$-invariant with respect to the dynamics (45). However, in the new basis, the trajectory is converging to the origin and remains inside the blue set for all $k \in \mathbb{Z}_+$ which is a $\mathscr{D}$-invariant with respect to the dynamic $\tilde{x}(k) = B_1\tilde{x}(k-1)+B_2\tilde{x}(k-2)$. Dashed lines in Fig. 2 represent the state trajectories starting from the same initial state in different basis.

5 Conclusion

A unifying characterization of the link between invariance in the extended state-space and $\mathscr{D}$-invariance, via set factorization, was studied for discrete-time DDEs. Low complexity invariant sets were recalled and it was shown that set factorization combined with similarity transformations allow a flexible description of invariant sets in state-spaces of same dimension. Thus, a relaxation of the conservativeness of the existing $\mathscr{D}$-invariance constructions was delivered for a more flexible $\mathscr{D}$-invariance characterization. Since set invariance concept is at the basis of many control schemes related to constrained control analysis/design, it is shown how the transformation can be adapted in order to handle polytopic state constraints.

References

1. Bitsoris, G.: Positively invariant polyhedral sets of discrete-time linear systems. Int. J. Control **47**(6), 1713–1726 (1988)
2. Gilbert, E., Tan, K.: Linear systems with state and control constraints, the theory and application of maximal output admissible sets. IEEE Trans. Autom. Control **36**, 1008–1020 (1991)
3. Blanchini, F., Miani, S.: Set-theoretic Methods in Control. Birkhuser, Boston (2008)
4. Mayne, D., Rawlings, J., Rao, C., Scokaert, P.: Constrained model predictive control: stability and optimality. Automatica **36**, 789–814 (2000)
5. Olaru, S., De Dona, J.A., Seron, M.M., Stoican, F.: Positive invariant sets for fault tolerant multisensor control schemes. Int. J. Control **83**(12), 2622–2640 (2010)

6. Stoican, F., Olaru, S.: Set-theoretic Fault-tolerant Control in Multisensor Systems. Wiley, New York (2013)
7. Gielen, R.H., Lazar, M., Kolmanovsky, I.V.: Lyapunov methods for time-invariant delay difference inclusions. SIAM J. Control Optim. **50**(1), 110–132 (2012)
8. Laraba, M.T., Olaru, S., Niculescu, S.-I., Blanchini, F., Miani, S., Casagrande, D., Giordano, G.: Guide on set invariance for delay difference equations, Annual Reviews in Control, vol. 41
9. Lombardi, W., Olaru, S., Lazar, M., Niculescu, S.-I.: On positive invariance for delay difference equations. In: Proceedings of the IEEE American Control Conference, pp. 3674–3679 (2011)
10. Lombardi, W., Luca, A., Olaru, S., Niculescu, S.-I.: On the polyhedral set invariance conditions for time-delay systems. In: Proceedings of the 18th IFAC World Congress, Milano, Italy, pp. 308–313 (2011)
11. Laraba, M.T., Olaru, S., Niculescu, S.-I., Blanchini, F., Miani, S., Casagrande, D., Giordano, G.: Set invariance for delay difference equations. In: 12th IFAC Workshop on Time Delay Systems, Ann Arbor (MI), USA (2015)
12. Olaru, S., Stanković, N., Bitsoris, G., Niculescu, S.-I.: Low complexity invariant sets for time-delay systems: a set factorization approach. In: Low-Complexity Controllers for Time-Delay Systems, pp. 127–139. Springer, Berlin (2014)
13. Lombardi, W.: Constrained control for time-delay systems. Ph.D. dissertation, Supélec (2011)
14. Lombardi, W., Olaru, S., Bitsoris, G., Niculescu, S.-I.: Cyclic invariance for discrete time-delay systems. Automatica **48**(10), 2730–2733 (2012)
15. Raković, S.V., Gielen, R.H., Lazar, M.: Construction of invariant families of sets for linear systems with delay. In: Proceedings of the IEEE American Control Conference, pp. 6246–6251 (2012)
16. Halbwachs, N., Merchat, D., Parent-Vigouroux, C.: Cartesian factoring of polyhedra in linear relation analysis. In: Cousot R. (ed.) Static Analysis Springer Series Lecture Notes in Computer Science, no. 2694, pp. 355–365 (2003)
17. Halbwachs, N., Merchat, D., Gonnord, L.: Some ways to reduce the space dimension in polyhedra computations. Formal Methods Syst. Des **29**(1), 79–95 (2006)
18. Kalman, R.E.: Mathematical description of linear dynamical systems. J.S.I.A.M Control **1**(2), 152–192 (1963)
19. Kalman, R.E., Bertram, J.E.: Control system analysis and design via the second method of Lyapunov. J. Basic Engr. Trans. A.S.M.E. **82**, 371–393 (1960)
20. Hahn, W.: Theorie und anwendung der direkten methode yon Ljapunov. Springer, Berlin (1959)
21. Stankovic, N., Olaru, S., Niculescu, S.I.: Further remarks on asymptotic stability and set invariance for linear delay difference equations. Automatica **50**(8), 2191–2195 (2014)

On the Exponential Stability of Two-Dimensional Nonautonomous Difference Systems Which Have a Weighted Homogeneity of the Solution

Masakazu Onitsuka

Abstract The present paper is considered a two-dimensional difference system:

$$\Delta x(n) = a(n)x(n) + b(n)\phi_{p^*}(y(n)), \quad \Delta y(n) = c(n)\phi_p(x(n)) + d(n)y(n),$$

where all coefficients are real-valued sequences; p and p^* are positive numbers satisfying $1/p + 1/p^* = 1$; and $\phi_p(x) = |x|^{p-2}x$ for $x \neq 0$, and $\phi_p(0) = 0$. The aim of this paper is to clarify that uniform asymptotic stability and exponential stability are equivalent for the above system. To illustrate the obtained results, an example is given. In addition, a figure of a solution orbit which is drawn by a computer is also attached for a deeper understanding.

Keywords Exponential stability · Uniform asymptotic stability · Difference system · Weighted homogeneity

1 Introduction

Let $\mathbb{N}$ be the set of all natural numbers, and let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. We consider the two-dimensional nonlinear nonautonomous difference system of the form

$$\begin{aligned} \Delta x(n) &= a(n)x(n) + b(n)\phi_{p^*}(y(n)), \\ \Delta y(n) &= c(n)\phi_p(x(n)) + d(n)y(n) \end{aligned} \tag{1}$$

for $n \in \mathbb{N}_0$, where Δ is the forward difference operator $\Delta x(n) = x(n+1) - x(n)$; and $a(n)$, $b(n)$, $c(n)$ and $d(n)$ are real-valued sequences for $n \in \mathbb{N}_0$; the positive numbers p and p^* satisfy $1/p + 1/p^* = 1$; the real-valued function $\phi_q(x)$ is defined by

M. Onitsuka (✉)
Department of Applied Mathematics, Okayama University of Science, 700-0005 Okayama, Japan
e-mail: onitsuka@xmath.ous.ac.jp

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_11

$$\phi_q(x) = \begin{cases} |x|^{q-2}x & \text{if } x \neq 0, \\ 0 & \text{if } x = 0, \end{cases} \quad x \in \mathbb{R}$$

for $q > 1$. Note that ϕ_{p^*} is the inverse function of ϕ_p, and the numbers p and p^* are naturally greater than 1. Since $\phi_p(0) = 0 = \phi_{p^*}(0)$, system (1) has the zero solution $(x(n), y(n)) \equiv (0, 0)$.

When $a(n) \equiv 0$ and $b(n) \equiv 1$, (1) is reduced to the equation

$$\Delta\left(\phi_p(\Delta x(n))\right) - d(n)\phi_p(\Delta x(n)) - c(n)\phi_p(x(n)) = 0.$$

This equation is often called "*half-linear equation*" because the solution space is homogeneous, but not additive; that is, if $x_1(n)$ and $x_2(n)$ are solutions of half-linear equation, then $cx_1(n)$ is also a solution for any $c \in \mathbb{R}$, but $x_1(n) + x_2(n)$ is not always a solution when $p \neq 2$. Half-linear equation is originated from the study of ordinary differential equations. For example, the reader is referred to [1, 3, 4]. For difference equations, we can find in [2, 7–10, 16, 20, 21]. It is known that half-linear differential equations is a special case of the nonlinear differential system

$$x' = a(t)x + b(t)\phi_{p^*}(y),$$
$$y' = c(t)\phi_p(x) + d(t)y,$$

where $a(t)$, $b(t)$, $c(t)$ and $d(t)$ are continuous functions. For example, see [6, 12–15, 17–19]. Certainly, system (1) is a natural generalization of the half-linear difference equation. Note here that system (1) has a weighted homogeneous (homogeneous-like) property on the solution space, which will be proved in Sect. 2. In this paper, we will deal with the stability of the all solutions of (1) in the neighborhood of the zero solution.

Let Ω be an open connected set and that $\mathbf{0} \in \Omega$. We now consider the nonlinear nonautonomous difference system

$$\Delta\mathbf{x}(n) = \mathbf{f}(n, \mathbf{x}(n)), \quad \mathbf{f}(n, \mathbf{0}) = \mathbf{0} \tag{2}$$

for $n \in \mathbb{N}_0$, where $\mathbf{x}(n) \in \Omega \subset \mathbb{R}^k$ and $k \in \mathbb{N}$; $\mathbf{f} : \mathbb{N}_0 \times \Omega \to \Omega$ is continuous on Ω. If an initial condition $\mathbf{x}(n_0) = \mathbf{x}_0$ is given, then for $n \in \mathbb{N}_0$ there is a unique solution $\mathbf{x}(n) \equiv \mathbf{x}(n; n_0, \mathbf{x}_0)$ of (2) such that $\mathbf{x}(n_0; n_0, \mathbf{x}_0) = \mathbf{x}_0$. It is clear that (2) has the zero solution $\mathbf{x}(n) \equiv \mathbf{0}$. Now we give some definitions of the stability of the zero solution of (2). Let $\|\mathbf{x}\|$ be the Euclidean norm of $\mathbf{x}$. The zero solution is said to be *uniformly attractive* if there exists a $\delta_0 > 0$ and, for any $\varepsilon > 0$, there exists an $N(\varepsilon) \in \mathbb{N}$ such that $n_0 \in \mathbb{N}_0$ and $\|\mathbf{x}_0\| < \delta_0$ imply $\|\mathbf{x}(n; n_0, \mathbf{x}_0)\| < \varepsilon$ for all $n \geq n_0 + N(\varepsilon)$ and $n \in \mathbb{N}_0$. The zero solution is said to be *uniformly stable* if for any $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that $n_0 \in \mathbb{N}_0$ and $\|\mathbf{x}_0\| < \delta(\varepsilon)$ imply $\|\mathbf{x}(n; n_0, \mathbf{x}_0)\| < \varepsilon$ for all $n \geq n_0$ and $n \in \mathbb{N}_0$. The zero solution is *uniformly asymptotically stable* if it is uniformly attractive and is uniformly stable. The zero solution is said to be *exponentially stable (or exponentially asymptotically stable);*

if there exists a $\lambda > 0$ and, given any $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that $n_0 \in \mathbb{N}_0$ and $\|\mathbf{x}_0\| < \delta(\varepsilon)$ imply $\|\mathbf{x}(n; n_0, \mathbf{x}_0)\| \le \varepsilon e^{-\lambda(n-n_0)}$ for all $n \ge n_0$ and $n \in \mathbb{N}_0$. For example, we can refer to the books [5, 11]. For ordinary differential equations, see [13, 14, 22, 23]. From the definitions, exponential stability implies uniform asymptotic stability. However, uniform asymptotic stability does not imply exponential stability. For instance, the zero solution of the nonlinear scalar equation

$$\Delta x(n) = \left(\frac{1}{\sqrt{1 + x^2(n)}} - 1 \right) x(n) \tag{3}$$

is uniformly asymptotically stable, but it is not exponentially stable. It is clear that (3) has the zero solution $x(n; n_0, 0) \equiv 0$. The unique solution of (3) is given by

$$x(n; n_0, x_0) = \frac{x_0}{\sqrt{1 + x_0^2(n - n_0)}}.$$

Now we will show that the zero solution of (3) is uniformly asymptotically stable. Let $\delta_0 = 1$. For any $0 < \varepsilon < 1$, we set $N(\varepsilon) = \min\{n \in \mathbb{N} \,|\, \varepsilon^{-2} - 1 \le n\}$. We consider the solution of (3) with $|x_0| < \delta_0 = 1$. We may assume without loss of generality that $|x_0| \ne 0$. Then we have

$$|x(n; n_0, x_0)| = \frac{1}{\sqrt{|x_0|^{-2} + n - n_0}} < \frac{1}{\sqrt{1 + N}} \le \varepsilon$$

for $n \ge n_0 + N$. That is, the zero solution of (3) is uniformly attractive. For any $\varepsilon > 0$, we choose $\delta(\varepsilon) = \varepsilon$. We consider the solution of (3) with $|x_0| < \delta$. Then we have

$$|x(n; n_0, x_0)| < \frac{\delta}{\sqrt{1 + x_0^2(n - n_0)}} \le \delta = \varepsilon$$

for $n \ge n_0$. Namely, the zero solution of (3) is uniformly stable, and therefore, it is uniformly asymptotically stable. On the other hand, the zero solution of (3) is not exponentially stable since

$$|x(n; n_0, x_0)| e^{\Lambda(n-n_0)} = \frac{x_0 e^{\Lambda(n-n_0)}}{\sqrt{1 + x_0^2(n - n_0)}} \to \infty \quad \text{as} \quad n \to \infty$$

for any $\Lambda > 0$ and $|x_0| \ne 0$. Hence, uniform asymptotic stability does not always imply exponential stability.

In the special case that $p = 2$, (1) is reduced to the linear difference system

$$\begin{aligned} \Delta x(n) &= a(n)x(n) + b(n)y(n), \\ \Delta x(n) &= c(n)x(n) + d(n)y(n). \end{aligned} \tag{4}$$

It is well-known that the linear system has some good properties as follows. The solution space of (4) is homogeneous and additive, and (4) has a fundamental matrix $\Phi(n)$; that is, each column of $\Phi(n)$ satisfies (4) such that $\det \Phi(n) \neq 0$. Furthermore, if an initial data $(n_0, (x_0, y_0)) \in \mathbb{N}_0 \times \mathbb{R}^2$ is given, then we can find the formula of the solution

$$\begin{pmatrix} x(n) \\ y(n) \end{pmatrix} = \Phi(n)\Phi^{-1}(n_0) \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

Using this information, we can find a relationship of uniform asymptotic stability and exponential stability (see [5, p.186] and [11, p.287]).

Theorem A *If the zero solution of* (4) *is uniformly asymptotically stable, then it is exponentially stable.*

This theorem means that uniform asymptotic stability and exponential stability are equivalent for linear system (4). Now, the natural question arises. Will uniform asymptotic stability guarantee exponential stability, even if system (1) is nonlinear? The purpose of this paper is to answer the question. Note here that under the assumption $p \neq 2$, system (1) is nonlinear, and the right-hand side of (1) is not continuously differentiable at the origin since the function ϕ_q satisfies

$$\lim_{x \to 0} \frac{d}{dx}\phi_q(x) = \lim_{x \to 0}(q-1)|x|^{q-2} = \infty,$$

if $1 < q < 2$; that is, we cannot linearize (1) around the origin. Despite these difficulties, we can obtain the answer to the above question as follows.

Theorem 1 *If the zero solution of* (1) *is uniformly asymptotically stable, then it is exponentially stable.*

The obtained result means that uniform asymptotic stability and exponential stability are equivalent for system (1).

In the next section, we prepare some lemmas and a proposition which is the core of the proof of Theorem 1. In Sect. 3, we present the proof of Theorem 1. To illustrate the obtained results, we give an example in Sect. 4.

2 Uniform Attractivity, Uniform Stability and Exponential Stability

In this section, first we will present some conditions which are equivalent to uniformly attractive, uniformly stable and exponentially stable, respectively. Let $\|(x, y)\|_p = \sqrt[p]{|x|^p + |y|^p}$ for $(x, y) \in \mathbb{R}^2$.

Lemma 1 *The zero solution of* (1) *is uniformly attractive if and only if there exists a* $\gamma_0 > 0$ *and, for any* $\rho > 0$, *there exists an* $M(\rho) \in \mathbb{N}$ *such that* $n_0 \in \mathbb{N}_0$ *and* $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma_0$ *imply*

$$\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p < \rho$$

for all $n \geq n_0 + M(\rho)$ *and* $n \in \mathbb{N}_0$.

Proof If the zero solution of (1) is uniformly attractive, then there exists a $\delta_0 > 0$ and, for any $\varepsilon > 0$, there exists an $N(\varepsilon) \in \mathbb{N}$ such that $n_0 \in \mathbb{N}_0$ and $\|(x_0, y_0)\| < \delta_0$ imply $\|(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))\| < \varepsilon$ for all $n \geq n_0 + N$ and $n \in \mathbb{N}_0$. We define the following constants

$$\overline{p} = \max\{p, p^*\} \quad \text{and} \quad \gamma_0 = \min\left\{1, \left(\frac{\delta_0}{\sqrt{2}}\right)^{\frac{\overline{p}}{p}}\right\}.$$

For every $0 < \rho < 1$, we choose $M(\rho) = N(\rho^p/2)$.

Now we consider the solution $(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))$ of (1) with $n_0 \in \mathbb{N}_0$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma_0$. Since

$$\sqrt[p]{|x_0|^p + |y_0|^{p^*}} = \|(x_0, \phi_{p^*}(y_0))\|_p < \gamma_0$$

holds, we have

$$|x_0| < \gamma_0 \quad \text{and} \quad |y_0| < \gamma_0^{\frac{p}{p^*}}.$$

Using $0 < \gamma_0 \leq 1 \leq \overline{p}/p, \overline{p}/p^* \geq 1$ and the above inequalities, we obtain

$$\begin{aligned}
\|(x_0, y_0)\| &= \sqrt{x_0{}^2 + y_0{}^2} < \sqrt{\gamma_0{}^2 + \gamma_0{}^{\frac{2p}{p^*}}} \\
&= \sqrt{\min\left\{1, \left(\frac{\delta_0}{\sqrt{2}}\right)^{\frac{2\overline{p}}{p}}\right\} + \min\left\{1, \left(\frac{\delta_0}{\sqrt{2}}\right)^{\frac{2\overline{p}}{p^*}}\right\}} \\
&\leq \sqrt{2\min\left\{1, \left(\frac{\delta_0}{\sqrt{2}}\right)^2\right\}} \leq \delta_0, \qquad (5)
\end{aligned}$$

which implies

$$\|(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))\| < \frac{\rho^p}{2}$$

for $n \geq n_0 + M$ and $n \in \mathbb{N}_0$. Note here that we used $\varepsilon = \rho^p/2$. Therefore, we see that

$$|x(n; n_0, x_0, y_0)| < \frac{\rho^p}{2} \quad \text{and} \quad |y(n; n_0, x_0, y_0)| < \frac{\rho^p}{2}$$

for $n \geq n_0 + M$ and $n \in \mathbb{N}_0$. Since $0 < \rho^p/2 < 1 < p$ and $p^* > 1$ hold, we have

$$\left(\frac{\rho^p}{2}\right)^p < \frac{\rho^p}{2} \quad \text{and} \quad \left(\frac{\rho^p}{2}\right)^{p^*} < \frac{\rho^p}{2}.$$

Consequently, we conclude that

$$\|(x(n;n_0,x_0,y_0),\phi_{p^*}(y(n;n_0,x_0,y_0)))\|_p < \sqrt[p]{\frac{\rho^p}{2}+\frac{\rho^p}{2}} = \rho$$

for $n \geq n_0 + M$ and $n \in \mathbb{N}_0$. Thus, the necessity is true.

Conversely, we assume that there exists a $\gamma_0 > 0$ and, for any $\rho > 0$, there exists an $M(\rho) \in \mathbb{N}$ such that $n_0 \in \mathbb{N}_0$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma_0$ imply

$$\|(x(n;n_0,x_0,y_0),\phi_{p^*}(y(n;n_0,x_0,y_0)))\|_p < \rho$$

for all $n \geq n_0 + M$ and $n \in \mathbb{N}_0$. Define

$$\delta_0 = \min\left\{1, \frac{{\gamma_0}^p}{2}\right\}.$$

For every $0 < \varepsilon < 1$, we choose $N(\varepsilon) = M\left(\left(\varepsilon/\sqrt{2}\right)^{\overline{p}/p}\right)$.

We consider the solution $(x(n;n_0,x_0,y_0), y(n;n_0,x_0,y_0))$ of (1) with $n_0 \in \mathbb{N}_0$ and $\|(x_0,y_0)\| < \delta_0$. Since $|x_0| < \delta_0$, $|y_0| < \delta_0$, $0 < \delta_0 \leq 1 < p$ and $p^* > 1$, we have

$$\begin{aligned}\|(x_0,\phi_{p^*}(y_0))\|_p < \sqrt[p]{{\delta_0}^p + {\delta_0}^{p^*}} &= \sqrt[p]{\min\left\{1, \left(\frac{{\gamma_0}^p}{2}\right)^p\right\} + \min\left\{1, \left(\frac{{\gamma_0}^p}{2}\right)^{p^*}\right\}} \\ &\leq \sqrt[p]{2\min\left\{1, \frac{{\gamma_0}^p}{2}\right\}} \leq \gamma_0, \end{aligned} \tag{6}$$

and therefore,

$$\|(x(n;n_0,x_0,y_0),\phi_{p^*}(y(n;n_0,x_0,y_0)))\|_p < \left(\frac{\varepsilon}{\sqrt{2}}\right)^{\frac{\overline{p}}{p}}$$

for $n \geq n_0 + N$ and $n \in \mathbb{N}_0$. Note here that we used $\rho = \left(\varepsilon/\sqrt{2}\right)^{\overline{p}/p}$. From this inequality and $0 < \varepsilon^2/2 < 1 \leq \overline{p}/p$ and $\overline{p}/p^* \geq 1$, it follows that

$$x^2(n;n_0,x_0,y_0) < \left(\frac{\varepsilon^2}{2}\right)^{\frac{\overline{p}}{p}} \leq \frac{\varepsilon^2}{2} \quad \text{and} \quad y^2(n;n_0,x_0,y_0) < \left(\frac{\varepsilon^2}{2}\right)^{\frac{\overline{p}}{p^*}} \leq \frac{\varepsilon^2}{2}$$

for $n \geq n_0 + N$ and $n \in \mathbb{N}_0$. Consequently, we conclude that

$$\|(x(n;n_0,x_0,y_0), y(n;n_0,x_0,y_0))\| < \varepsilon$$

for $n \geq n_0 + N$ and $n \in \mathbb{N}_0$. This completes the proof of Lemma 1. □

Lemma 2 *The zero solution of* (1) *is uniformly stable if and only if for any* $\rho > 0$, *there exists a* $\gamma(\rho) > 0$ *such that* $n_0 \in \mathbb{N}_0$ *and* $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma(\rho)$ *imply*

$$\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p < \rho$$

for all $n \geq n_0$ *and* $n \in \mathbb{N}_0$.

Proof If the zero solution of (1) is uniformly stable, then for any $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that $n_0 \in \mathbb{N}_0$ and $\|(x_0, y_0)\| < \delta(\varepsilon)$ imply

$$\|(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))\| < \varepsilon$$

for all $n \geq n_0$ and $n \in \mathbb{N}_0$. Let $\overline{p} = \max\{p, p^*\}$. For every $0 < \rho < 1$, we choose

$$\gamma(\rho) = \min\left\{1, \left(\frac{1}{\sqrt{2}}\delta\left(\frac{\rho^p}{2}\right)\right)^{\frac{\overline{p}}{p}}\right\}.$$

Now we consider the solution $(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))$ of (1) with $n_0 \in \mathbb{N}_0$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma$. By the same estimate as in (5), we get $\|(x_0, y_0)\| < \delta$. Therefore, we see that

$$\|(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))\| < \frac{\rho^p}{2}$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$. Using the same argument as in the proof of Lemma 1, we conclude that $\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p < \rho$ for $n \geq n_0$ and $n \in \mathbb{N}_0$.

Conversely, we assume that for any $\rho > 0$, there exists a $\gamma(\rho) > 0$ such that $n_0 \in \mathbb{N}_0$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma(\rho)$ imply

$$\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p < \rho$$

for all $n \geq n_0$ and $n \in \mathbb{N}_0$. For every $0 < \varepsilon < 1$, we choose

$$\delta(\varepsilon) = \min\left\{1, \frac{1}{2}\gamma^p\left(\left(\frac{\varepsilon}{\sqrt{2}}\right)^{\frac{\overline{p}}{p}}\right)\right\}.$$

We consider the solution $(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))$ of (1) with $n_0 \in \mathbb{N}_0$ and $\|(x_0, y_0)\| < \delta$. By the same estimate as in (6), we obtain $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma$, and therefore,

$$\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p < \left(\frac{\varepsilon}{\sqrt{2}}\right)^{\frac{\overline{p}}{p}}$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$. Hence, we conclude that $\|(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))\| < \varepsilon$ for $n \geq n_0$ and $n \in \mathbb{N}_0$ by using the same argument as in the proof of Lemma 1. This completes the proof of Lemma 2. □

Lemma 3 *The zero solution of* (1) *is exponentially stable if and only if there exists a* $\mu > 0$ *and, given any* $\rho > 0$*, there exists a* $\gamma(\rho) > 0$ *such that* $n_0 \in \mathbb{N}_0$ *and* $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma(\rho)$ *imply*

$$\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p \leq \rho e^{-\mu(n-n_0)}$$

for all $n \geq n_0$ *and* $n \in \mathbb{N}_0$.

Proof If the zero solution of (1) is exponentially stable, then there exists a $\lambda > 0$ and, given any $\varepsilon > 0$, there exists a $\delta(\varepsilon) > 0$ such that $n_0 \in \mathbb{N}_0$ and $\|(x_0, y_0)\| < \delta(\varepsilon)$ imply $\|(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))\| \leq \varepsilon e^{-\lambda(n-n_0)}$ for all $n \geq n_0$ and $n \in \mathbb{N}_0$. Let $\mu = \lambda/p$ and $\overline{p} = \max\{p, p^*\}$. For every $0 < \rho < 1$, we determine

$$\gamma(\rho) = \min\left\{1, \left(\frac{1}{\sqrt{2}}\delta\left(\frac{\rho^p}{2}\right)\right)^{\frac{\overline{p}}{p}}\right\}.$$

We consider the solution $(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))$ of (1) with $n_0 \in \mathbb{N}_0$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma$. By the same estimate as in (5), we get $\|(x_0, y_0)\| < \delta$. Thus, we obtain

$$\|(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))\| \leq \frac{\rho^p}{2} e^{-p\mu(n-n_0)}$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$. Using

$$0 < \frac{\rho^p}{2} e^{-p\mu(n-n_0)} \leq \frac{\rho^p}{2} < 1 < p, \quad 1 < p^*$$

and the above inequality, we have

$$|x(n; n_0, x_0, y_0)|^p \leq \left(\frac{\rho^p}{2} e^{-p\mu(n-n_0)}\right)^p < \frac{\rho^p}{2} e^{-p\mu(n-n_0)}$$

and

$$|y(n; n_0, x_0, y_0)|^{p^*} \leq \left(\frac{\rho^p}{2} e^{-p\mu(n-n_0)}\right)^{p^*} < \frac{\rho^p}{2} e^{-p\mu(n-n_0)}$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$. Consequently, we conclude that

$$\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p < \rho e^{-\mu(n-n_0)}$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$.

Conversely, we suppose that there exists a $\mu > 0$ and, given any $\rho > 0$, there exists a $\gamma(\rho) > 0$ such that $n_0 \in \mathbb{N}_0$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma(\rho)$ imply

$$\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p \leq \rho e^{-\mu(n-n_0)}$$

for all $n \geq n_0$ and $n \in \mathbb{N}_0$. Let $\lambda = \mu p/\overline{p}$. For every $0 < \varepsilon < 1$, we choose

$$\delta(\varepsilon) = \min\left\{1, \frac{1}{2}\gamma^p\left(\left(\frac{\varepsilon}{\sqrt{2}}\right)^{\frac{\overline{p}}{p}}\right)\right\}.$$

We consider the solution $(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))$ of (1) with $n_0 \in \mathbb{N}_0$ and $\|(x_0, y_0)\| < \delta$. By the same estimate as in (6), we obtain $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma$, and therefore,

$$\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p \leq \left(\frac{\varepsilon}{\sqrt{2}}\right)^{\frac{\overline{p}}{p}} e^{-\frac{\overline{p}\lambda}{p}(n-n_0)}$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$. Using

$$0 < \frac{\varepsilon}{\sqrt{2}} e^{-\lambda(n-n_0)} \leq \frac{\varepsilon}{\sqrt{2}} < 1 \leq \frac{\overline{p}}{p} \quad \text{and} \quad 1 \leq \frac{\overline{p}}{p^*}$$

and the above inequality, we have

$$|x(n; n_0, x_0, y_0)| \leq \left(\frac{\varepsilon}{\sqrt{2}} e^{-\lambda(n-n_0)}\right)^{\frac{\overline{p}}{p}} \leq \frac{\varepsilon}{\sqrt{2}} e^{-\lambda(n-n_0)}$$

and

$$|y(n; n_0, x_0, y_0)| \leq \left(\frac{\varepsilon}{\sqrt{2}} e^{-\lambda(n-n_0)}\right)^{\frac{\overline{p}}{p^*}} \leq \frac{\varepsilon}{\sqrt{2}} e^{-\lambda(n-n_0)}$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$. Consequently, we see that $\|(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))\| \leq \varepsilon e^{-\lambda(n-n_0)}$ for $n \geq n_0$ and $n \in \mathbb{N}_0$. This completes the proof of Lemma 3. □

The following lemma means that system (1) has a weighted homogeneous property on the solution space.

Lemma 4 *If* $(x(n), y(n))$ *is a solution of* (1) *passing through a point* $(x_0, y_0) \in \mathbb{R}^2$ *at* $n = n_0 \in \mathbb{N}_0$*, then* $(\alpha x(n), \phi_p(\alpha) y(n))$ *is also a solution of* (1) *passing through a point* $(\alpha x_0, \phi_p(\alpha) y_0) \in \mathbb{R}^2$ *at* $n = n_0$ *for any* $\alpha \in \mathbb{R}$.

Proof Let $(x(n), y(n))$ be a solution of (1) passing through a point (x_0, y_0) at $n = n_0$. Define $\tilde{x}(n) = \alpha x(n)$ and $\tilde{y}(n) = \phi_p(\alpha) y(n)$ with $\alpha \in \mathbb{R}$. Then $(\tilde{x}(n_0), \tilde{y}(n_0)) = (\alpha x_0, \phi_p(\alpha) y_0)$ holds. Since ϕ_{p^*} is the inverse function of ϕ_p, we obtain

$$\Delta\tilde{x}(n) = \alpha\Delta x(n) = a(n)\alpha x(n)+b(n)\phi_{p^*}(\phi_p(\alpha)y(n)) = a(n)\tilde{x}(n)+b(n)\phi_{p^*}(\tilde{y}(n))$$

and

$$\Delta\tilde{y}(n) = \phi_p(\alpha)\Delta y(n) = c(n)\phi_p(\alpha x(n)) + d(n)\phi_p(\alpha)y(n) = c(n)\phi_p(\tilde{x}(n)) + d(n)\tilde{y}(n).$$

Thus, we see that $(\alpha x(n), \phi_p(\alpha)y(n))$ is a solution of (1) passing through a point $(\alpha x_0, \phi_p(\alpha)y_0)$ at $n = n_0$. □

The following result is the most important property for the proof of Theorem 1.

Proposition 5 *If the zero solution of* (1) *is uniformly attractive, then there exists a* $\gamma_0 > 0$ *and, for every* $\nu > 1$*, there exists an* $N(\nu) \in \mathbb{N}$ *such that* $n_0 \in \mathbb{N}_0$ *and* $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma_0\nu^{-(k-1)}$ *imply*

$$\|(x(n; n_0 + (k-1)N(\nu), x_0, y_0), \phi_{p^*}(y(n; n_0 + (k-1)N(\nu), x_0, y_0)))\|_p < \gamma_0\nu^{-k}$$

for all $n \geq n_0 + kN(\nu)$*,* $n \in \mathbb{N}_0$ *and* $k \in \mathbb{N}$*.*

Proof Lemma 1 implies that there exists a $\gamma_0 > 0$ and, for every $\nu > 1$, there exists an $M(\gamma_0/\nu) \in \mathbb{N}$ such that $\tau \in \mathbb{N}_0$ and $\|(\xi, \phi_{p^*}(\eta))\|_p < \gamma_0$ imply

$$\|(x(n; \tau, \xi, \eta), \phi_{p^*}(y(n; \tau, \xi, \eta)))\|_p < \frac{\gamma_0}{\nu}$$

for all $n \geq \tau + M$ and $n \in \mathbb{N}_0$. Let $N(\nu) = M(\gamma_0/\nu)$.

Let k be a natural number. Now we consider the solution

$$(x(n; n_0 + (k-1)N, x_0, y_0), y(n; n_0 + (k-1)N, x_0, y_0))$$

of (1) with $n_0 \in \mathbb{N}_0$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma_0\nu^{-(k-1)}$ for $n \geq n_0 + (k-1)N$ and $n \in \mathbb{N}_0$. Lemma 4 implies that

$$\left(\nu^{k-1}x(n; n_0 + (k-1)N, x_0, y_0), \phi_p\left(\nu^{k-1}\right)y(n; n_0 + (k-1)N, x_0, y_0)\right)$$

is also a solution of (1) passing through a point $\left(\nu^{k-1}x_0, \phi_p\left(\nu^{k-1}\right)y_0\right)$ at $n = n_0 + (k-1)N$. Since

$$\left\|\left(\nu^{k-1}x_0, \phi_{p^*}\left(\phi_p\left(\nu^{k-1}\right)y_0\right)\right)\right\|_p = \nu^{k-1}\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma_0$$

holds, we obtain

$$\begin{aligned}
\frac{\gamma_0}{\nu} &> \left\|\left(\nu^{k-1}x(n; n_0+(k-1)N, x_0, y_0), \phi_{p^*}\left(\phi_p\left(\nu^{k-1}\right)y(n; n_0+(k-1)N, x_0, y_0)\right)\right)\right\|_p \\
&= \|(\nu^{k-1}x(n; n_0+(k-1)N, x_0, y_0), \nu^{k-1}\phi_{p^*}(y(n; n_0+(k-1)N, x_0, y_0)))\|_p \\
&= \nu^{k-1}\|(x(n; n_0+(k-1)N, x_0, y_0), \phi_{p^*}(y(n; n_0+(k-1)N, x_0, y_0)))\|_p
\end{aligned}$$

for all $n \geq n_0 + (k-1)N + N = n_0 + kN$. Consequently, we see that

$$\|(x(n; n_0+(k-1)N, x_0, y_0), \phi_{p^*}(y(n; n_0+(k-1)N, x_0, y_0)))\|_p < \gamma_0 \nu^{-k}$$

for all $n \geq n_0 + kN$, $n \in \mathbb{N}_0$ and $k \in \mathbb{N}$. □

3 Proof of the Main Theorem

Now we present a proof of the main theorem.

Proof (Proof of Theorem 1) From uniform attractivity of (1) and Proposition 5, there exist a $\gamma_0 > 0$ and an $N(e) \in \mathbb{N}$ such that $n_0 \in \mathbb{N}_0$ and $\|(\xi, \phi_{p^*}(\eta))\|_p < \gamma_0 e^{-(k-1)}$ imply

$$\|(x(n; n_0 + (k-1)N, \xi, \eta), \phi_{p^*}(y(n; n_0 + (k-1)N, \xi, \eta)))\|_p < \gamma_0 e^{-k} \tag{7}$$

for all $n \geq n_0 + kN$, $n \in \mathbb{N}_0$ and $k \in \mathbb{N}$.

From uniform stability of (1) and Lemma 2, there exists a $\gamma(\gamma_0) > 0$ such that $n_0 \in \mathbb{N}_0$ and $\|(\xi, \phi_{p^*}(\eta))\|_p < \gamma$ imply

$$\|(x(n; n_0, \xi, \eta), \phi_{p^*}(y(n; n_0, \xi, \eta)))\|_p < \gamma_0 \tag{8}$$

for all $n \geq n_0$ and $n \in \mathbb{N}_0$. We set $\lambda = 1/N$. For every $\varepsilon > 0$, we choose

$$\delta(\varepsilon) = \frac{\gamma\varepsilon}{\gamma_0 e} > 0.$$

Now we consider the solution $(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))$ of (1) with $n_0 \in \mathbb{N}_0$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \delta$. For the sake of convenience, let

$$(x(n), y(n)) = (x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0)).$$

Lemma 4 implies that

$$\left(\frac{\gamma_0 e}{\varepsilon} x(n), \phi_p\left(\frac{\gamma_0 e}{\varepsilon}\right) y(n)\right)$$

is a solution of (1) passing through a point $((\gamma_0 e/\varepsilon)x_0, \phi_p(\gamma_0 e/\varepsilon) y_0)$ at $n = n_0$. Using $\delta = \gamma\varepsilon/(\gamma_0 e)$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \delta$, we have

$$\begin{aligned}\left\|\left(\frac{\gamma_0 e}{\varepsilon} x_0, \phi_{p^*}\left(\phi_p\left(\frac{\gamma_0 e}{\varepsilon}\right) y_0\right)\right)\right\|_p &= \left\|\left(\frac{\gamma_0 e}{\varepsilon} x_0, \frac{\gamma_0 e}{\varepsilon} \phi_{p^*}(y_0)\right)\right\|_p \\ &= \frac{\gamma_0 e}{\varepsilon} \|(x_0, \phi_{p^*}(y_0))\|_p < \gamma\end{aligned}$$

at $n = n_0$. From this inequality and (8) with

$$(\xi, \eta) = \left(\frac{\gamma_0 e}{\varepsilon} x_0, \phi_p\left(\frac{\gamma_0 e}{\varepsilon}\right) y_0\right),$$

we obtain

$$\frac{\gamma_0 e}{\varepsilon} \|(x(n), \phi_{p^*}(y(n)))\|_p = \left\| \left(\frac{\gamma_0 e}{\varepsilon} x(n), \phi_{p^*}\left(\phi_p\left(\frac{\gamma_0 e}{\varepsilon}\right) y(n)\right)\right) \right\|_p < \gamma_0 \tag{9}$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$. Consequently, we see that

$$\|(x(n), \phi_{p^*}(y(n)))\|_p < \frac{\varepsilon}{e}$$

for $n_0 \leq n \leq n_0 + N$ and $n \in \mathbb{N}_0$.

We note that

$$\left\| \left(\frac{\gamma_0 e}{\varepsilon} x_0, \phi_{p^*}\left(\phi_p\left(\frac{\gamma_0 e}{\varepsilon}\right) y_0\right)\right) \right\|_p < \gamma_0$$

holds at $n = n_0$ from (9). Then, from this inequality and (7) with

$$(\xi, \eta) = \left(\frac{\gamma_0 e}{\varepsilon} x_0, \phi_p\left(\frac{\gamma_0 e}{\varepsilon}\right) y_0\right), \quad k = 1,$$

we see that

$$\frac{\gamma_0 e}{\varepsilon} \|(x(n), \phi_{p^*}(y(n)))\|_p = \left\| \left(\frac{\gamma_0 e}{\varepsilon} x(n), \phi_{p^*}\left(\phi_p\left(\frac{\gamma_0 e}{\varepsilon}\right) y(n)\right)\right) \right\|_p < \frac{\gamma_0}{e} \tag{10}$$

for $n \geq n_0 + N$ and $n \in \mathbb{N}_0$. Consequently, we get

$$\|(x(n), \phi_{p^*}(y(n)))\|_p < \frac{\varepsilon}{e^2}$$

for $n_0 + N \leq n \leq n_0 + 2N$ and $n \in \mathbb{N}_0$.

Moreover, we note that

$$\left\| \left(\frac{\gamma_0 e}{\varepsilon} x(n_0 + N), \phi_{p^*}\left(\phi_p\left(\frac{\gamma_0 e}{\varepsilon}\right) y(n_0 + N)\right)\right) \right\|_p < \frac{\gamma_0}{e}$$

holds at $n = n_0 + N$ from (10). Using this inequality and (7) with

$$(\xi, \eta) = \left(\frac{\gamma_0 e}{\varepsilon} x(n_0 + N), \phi_p\left(\frac{\gamma_0 e}{\varepsilon}\right) y(n_0 + N)\right), \quad k = 2,$$

we see that

$$\frac{\gamma_0 e}{\varepsilon} \|(x(n), \phi_{p^*}(y(n)))\|_p = \left\| \left(\frac{\gamma_0 e}{\varepsilon} x(n), \phi_{p^*}\left(\phi_p\left(\frac{\gamma_0 e}{\varepsilon}\right) y(n)\right)\right) \right\|_p < \frac{\gamma_0}{e^2}$$

for $n \geq n_0 + 2N$ and $n \in \mathbb{N}_0$. Consequently, we have

$$\|(x(n), \phi_{p^*}(y(n)))\|_p < \frac{\varepsilon}{e^3}$$

for $n_0 + 2N \leq n \leq n_0 + 3N$ and $n \in \mathbb{N}_0$.

By the same process as in the above mentioned estimates, we conclude that

$$\|(x(n), \phi_{p^*}(y(n)))\|_p < \varepsilon e^{-k}$$

for $n_0 + (k-1)N \leq n \leq n_0 + kN$, $n \in \mathbb{N}_0$ and $k \in \mathbb{N}$. Since

$$-k \leq -\frac{1}{N}(n - n_0) = -\lambda(n - n_0)$$

holds, we have $\|(x(n), \phi_{p^*}(y(n)))\|_p < \varepsilon e^{-k} \leq \varepsilon e^{-\lambda(n-n_0)}$ for $n_0 + (k-1)N \leq n \leq n_0 + kN$, $n \in \mathbb{N}_0$ and $k \in \mathbb{N}$. Since we can divide the interval $[n_0, n_0 + kN]$ as

$$[n_0, n_0 + kN] = \bigcup_{i=1}^{k} [n_0 + (i-1)N, n_0 + iN]$$

for $k \in \mathbb{N}$, we conclude that

$$\|(x(n), \phi_{p^*}(y(n)))\|_p \leq \varepsilon e^{-\lambda(n-n_0)}$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$. Therefore, Lemma 3 implies that the zero solution of (1) is exponentially stable. This completes the proof of Theorem 1. □

4 Example and Simulation

In this section, we give an example. We consider the system of difference equations

$$\begin{aligned} \Delta x(n) &= -x(n) + f(n)\phi_{p^*}(y(n)), \\ \Delta y(n) &= -\phi_p(f(n))\phi_p(x(n)) - y(n), \end{aligned} \tag{11}$$

where $f(n) = (n+1)^{-\{1+(-1)^n\}/2}$ for $n \in \mathbb{N}_0$. Then the zero solution of (11) is uniformly asymptotically stable. Moreover, by Theorem 1, it is exponentially stable.

We will check this fact. Define the nonnegative function

$$V(n) = |x(n)|^p + |y(n)|^{p^*},$$

where $(x(n), y(n))$ is a solution of (11) for $n \geq n_0$ and $n \in \mathbb{N}_0$. Then we have

$$\begin{aligned}\Delta V(n) &= |x(n+1)|^p + |y(n+1)|^{p^*} - V(n) \\ &= |f(n)\phi_{p^*}(y(n))|^p + |\phi_p(f(n))\phi_p(x(n))|^{p^*} - V(n) \\ &= |f(n)|^p|y(n)|^{p^*} + |f(n)|^p|x(n)|^p - V(n) \\ &= (|f(n)|^p - 1)V(n) \end{aligned} \tag{12}$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$. From this and $f(n) = (n+1)^{-\{1+(-1)^n\}/2}$, we obtain

$$\Delta V(n) = \begin{cases} -\left(1 - \dfrac{1}{(n+1)^p}\right)V(n) & \text{if } n = 2k, \\ 0 & \text{if } n = 2k-1, \end{cases} \qquad k \in \mathbb{N}_0$$

for $n \geq n_0$ and $n \in \mathbb{N}_0$. If we regard V as a Lyapunov function, we can conclude that the derivative of V along any solution of (11) is nonpositive. Note here that if n is odd then $\Delta V(n) = 0$. This means that we cannot use well-known Lyapunov theorems for the (uniform) asymptotic stability, because these require that $\Delta V(n) < 0$ for all $n \in \mathbb{N}_0$ (see [11]). For this reason, we will prove directly that the zero solution is uniformly asymptotically stable by using function $V(n)$.

From (12) and

$$|f(n)| = (n+1)^{-\{1+(-1)^n\}/2} \leq 1,$$

we can easily check that

$$\begin{aligned} V(n) &= |f(n_0)|^p|f(n_0+1)|^p|f(n_0+2)|^p \cdots |f(n-2)|^p|f(n-1)|^pV(n_0) \\ &\leq |f(n-2)|^p|f(n-1)|^pV(n_0) \leq \frac{V(n_0)}{(n-1)^p} \end{aligned} \tag{13}$$

for $n \geq n_0 + 2$, and

$$V(n) \leq V(n_0) \tag{14}$$

for $n \geq n_0$.

First, we prove uniform attractivity by using Lemma 1. Let $\gamma_0 = 1$. For every $0 < \rho \leq 1$, we choose

$$M(\rho) = \min\left\{n \in \mathbb{N} \,\middle|\, \frac{1}{\rho} + 1 \leq n\right\} \geq 2.$$

We consider the solution $(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))$ of (11) with $n_0 \in \mathbb{N}_0$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma_0 = 1$. From (13) and $\|(x_0, \phi_{p^*}(y_0))\|_p = V^{1/p}(n_0)$, we have

$$\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p < \frac{1}{n-1} \leq \frac{1}{n_0 + M(\rho) - 1} \leq \rho$$

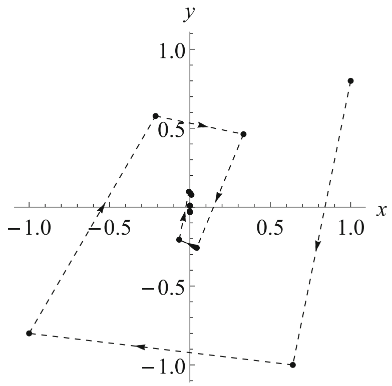


Fig. 1 A solution orbit of system (11) with $p = 3/2$ and an initial data $(n_0, (x_0, y_0)) = (0, (1, 0.8))$.

for all $n \geq n_0 + M(\rho)$ and $n \in \mathbb{N}_0$. Thus, Lemma 1 implies that the zero solution of (11) is uniformly attractive.

Next, we prove uniform stability by using Lemma 2. For every $\rho > 0$, we choose $\gamma(\rho) = \rho$. We consider the solution $(x(n; n_0, x_0, y_0), y(n; n_0, x_0, y_0))$ of (11) with $n_0 \in \mathbb{N}_0$ and $\|(x_0, \phi_{p^*}(y_0))\|_p < \gamma(\rho) = \rho$. From (14) and $\|(x_0, \phi_{p^*}(y_0))\|_p = V^{1/p}(n_0)$, we have

$$\|(x(n; n_0, x_0, y_0), \phi_{p^*}(y(n; n_0, x_0, y_0)))\|_p < \rho$$

for all $n \geq n_0$ and $n \in \mathbb{N}_0$. Thus, Lemma 2 implies that the zero solution of (11) is uniformly stable, and therefore, the zero solution of (11) is uniformly asymptotically stable. Using Theorem 1, we can conclude that the zero solution of (11) is exponentially stable.

Finally, to illustrate our example, we present a simulation. In Fig. 1, we draw a solution orbit of (11) with $p = 3/2$ starting from the point $(0, (1, 0.8)) \in \mathbb{N}_0 \times \mathbb{R}^2$. This solution tends to the zero solution exponentially.

References

1. Agarwal, R.P., Grace, S.R., O'Regan, D.: Oscillation theory for second order linear, half-linear, superlinear and sublinear dynamic equations. Kluwer Academic Publishers, Dordrecht (2002)
2. Cecchi, M., Došlá, Z., Marini, M., Vrkoč, I.: Asymptotic properties for half-linear difference equations. Math. Bohem. **131**(4), 347–363 (2006)
3. Došlý, O.: Half-linear differential equations, Handbook of differential equations, 161–357. Elsevier, Amsterdam (2004)
4. Došlý, O., Řehák, P.: Half-linear differential equations, North-Holland Mathematics Studies 202. Elsevier, Amsterdam (2005)
5. Elaydi, S.: An introduction to difference equations, 3rd edn. Undergraduate Texts in Mathematics. Springer, New York (2005)
6. Elbert, Á.: Asymptotic behaviour of autonomous half-linear differential systems on the plane. Studia Sci. Math. Hungar **19**(2–4), 447–464 (1984)

7. Fišnarová, S.: Oscillatory properties of half-linear difference equations: two-term perturbations. Adv. Differ. Equ. **2012**, 2012:101, 16 pp. (2012)
8. Hasil, P., Veselý, M.: Oscillation constants for half-linear difference equations with coefficients having mean values. Adv. Differ. Equ. **2015**, 2015:210, 18 pp. (2015)
9. Jiang, J., Tang, X.: Oscillation of second order half-linear difference equations (I). Appl. Math. Sci. (Ruse) **8**(37–40), 1957–1968 (2014)
10. Matucci, S., Řehák, P.: Rapidly varying decreasing solutions of half-linear difference equations. Math. Comput. Modelling **49**(7–8), 1692–1699 (2009)
11. Michel, A.N., Hou, L., Liu, D.: Stability of dynamical systems: Continuous, discontinuous, and discrete systems, Systems and Control: Foundations and Applications. Birkhäuser Boston Inc, Boston, MA (2008)
12. Mirzov, J.D.: Asymptotic properties of solutions of systems of nonlinear nonautonomous ordinary differential equations. Folia Facultatis Scientiarium Naturalium Universitatis Masarykianae Brunensis. Mathematica, 14. Masaryk University, Brno (2004)
13. Onitsuka M., Soeda, T.: Uniform asymptotic stability implies exponential stability for nonautonomous half-linear differential systems. Adv. Differ. Equ. **2015**, 2015:158, 24 pp. (2015)
14. Onitsuka, M., Sugie, J.: Uniform global asymptotic stability for half-linear differential systems with time-varying coefficients. Proc. Roy. Soc. Edinb. Sect. A **141**(5), 1083–1101 (2011)
15. Onitsuka, M., Tanaka, S.: Characteristic equation for autonomous planar half-linear differential systems. Acta. Math. Hungar. **152**(2), 336–364 (2017)
16. Řehák, P.: Oscillation criteria for second order half-linear difference equations. J. Differ. Equ. Appl. **7**(4), 483–505 (2001)
17. Sugie, J., Hata, S., Onitsuka, M.: Global attractivity for half-linear differential systems with periodic coefficients. J. Math. Anal. Appl. **371**(1), 95–112 (2010)
18. Sugie, J., Onitsuka, M.: Global asymptotic stability for half-linear differential systems with coefficients of indefinite sign. Arch. Math. (Brno) **44**(4), 317–334 (2008)
19. Sugie, J., Onitsuka, M., Yamaguchi, A.: Asymptotic behavior of solutions of nonautonomous half-linear differential systems. Studia Sci. Math. Hungar. **44**, 159–189 (2007)
20. Veselý, M., Hasil, P.: Oscillation and nonoscillation of asymptotically almost periodic half-linear difference equations. Abstr. Appl. Anal. **2013**, Art. ID 432936, 12 pp. (2013)
21. Wong, P.J.Y., Agarwal, R.P.: Oscillations and nonoscillations of half-linear difference equations generated by deviating arguments. Advances in difference equations, II. Comput. Math. Appl. **36**(10–12), 11–26 (1998)
22. Yoshizawa, T.: Stability theory by Liapunov's second method. The Mathematical Society of Japan, Tokyo (1966)
23. Yoshizawa, T.: Stability theory and the existence of periodic solutions and almost periodic solutions, Applied Mathematical Sciences 14. Springer-Verlag, New York, Heidelberg (1975)

A Corollary of a Theorem on Positive Solutions of Poincaré Difference Equations

Mihály Pituk

Abstract It is known that the exponential growth rate of every positive solution of a Poincaré difference equation is a nonnegative eigenvalue of the limiting equation with a positive eigenvector. In this note we show how this discrete result implies its continuous counterpart.

Keywords Poincaré difference equation · Growth rate · Cone positivity · Ordinary differential equation · Lyapunov exponent

1 Introduction and Main Result

Let $\mathbb{Z}$, $\mathbb{Z}^+$, $\mathbb{R}$ and $\mathbb{R}^+$ denote the set of integers, the set of nonnegative integers, the set of real numbers and the set of nonnegative real numbers, respectively. Given a positive integer k, $\mathbb{R}^k$ denotes the k-dimensional space of real column vectors with any norm $\|\cdot\|$. As usual, the symbol $\mathbb{R}^{k\times k}$ denotes the space of $k \times k$ matrices with real entries. The *induced norm* of a matrix $A \in \mathbb{R}^{k\times k}$ is defined by

$$\|A\| = \sup_{0\neq x\in\mathbb{R}^k} \frac{\|Ax\|}{\|x\|}.$$

A set K is said to be a *cone* in $\mathbb{R}^k$ if all three conditions below hold.

(i) K is a nonempty, convex and closed subset of $\mathbb{R}^k$,
(ii) $tK \subset K$ for all $t \geq 0$, where $tK = \{tx \mid x \in K\}$,
(iii) $K \cap (-K) = \{0\}$, where $-K = \{-x \mid x \in K\}$.

M. Pituk (✉)
Department of Mathematics, University of Pannonia, Egyetem út 10, 8200 Veszprém, Hungary
e-mail: pitukm@almos.uni-pannon.hu

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_12

Each cone K induces a partial ordering $\leq_K$ in $\mathbb{R}^k$ by $x \leq_K y$ if and only if $y - x \in K$. A vector $x \in \mathbb{R}^k$ is called *K-nonnegative* if $0 \leq_K x$. We say that $x \in \mathbb{R}^k$ is *K-positive* if $0 \leq_K x$ and $x \neq 0$. Thus, $x \in \mathbb{R}^k$ is K-positive if and only if $x \in K \setminus \{0\}$.

Consider the Poincaré difference equation

$$x(n+1) = (A + B(n))x(n), \qquad n \in \mathbb{Z}^+, \tag{1}$$

where $A \in \mathbb{R}^{k\times k}$ and $B : \mathbb{Z}^+ \to \mathbb{R}^{k\times k}$ satisfies

$$B(n) \to 0, \qquad n \to \infty. \tag{2}$$

A solution $x : \mathbb{Z}^+ \to \mathbb{R}^k$ of (1) is called *nonvanishing* if $x(n) \neq 0$ for all $n \in \mathbb{Z}^+$. According to a Perron type theorem [9], if (2) holds and x is a nonvanishing solution of (1), then the limit

$$\rho(x) = \lim_{n\to\infty} \sqrt[n]{\|x(n)\|} \tag{3}$$

exists and is equal to one of the modulus of eigenvalues of A. The quantity $\rho(x)$ is called the *exponential growth rate* of the solution x. Its logarithm is the Lyapunov exponent. For further related results, see [4, 7].

Let K be a cone in $\mathbb{R}^k$. A solution $x : \mathbb{Z}^+ \to \mathbb{R}^k$ of (1) is called K-positive if $x(n)$ is K-positive for all $n \in \mathbb{Z}^+$. In [8] we have shown the following improvement of the Perron type theorem for the K-positive solutions of (1).

Theorem 1 *Suppose* (2) *holds and let K be a cone in $\mathbb{R}^k$. If x is a K-positive solution of* (1)*, then its exponential growth rate $\rho(x)$ is a nonnegative eigenvalue of A with a K-positive eigenvector.*

The aim of the present note is to show how Theorem 1 can be used to prove its counterpart for the ordinary differential equation

$$y' = (C + D(t))y, \qquad t \in \mathbb{R}^+, \tag{4}$$

where $C \in \mathbb{R}^{k\times k}$ and $D : \mathbb{R}^+ \to \mathbb{R}^{k\times k}$ is a continuous matrix function satisfying

$$D(t) \to 0, \qquad t \to \infty, \tag{5}$$

or, more generally,

$$\int_t^{t+1} \|D(s)\|\, ds \to 0, \qquad t \to \infty. \tag{6}$$

It is known that if (6) holds and y is a nontrivial solution of (4), then the limit

$$\lambda(y) = \lim_{t\to\infty} \frac{\log \|y(t)\|}{t} \tag{7}$$

exists and is equal to the real part of one of the eigenvalues of C (see Theorem 5 in Chapter IV of [2]). The quantity $\lambda(y)$ is called the *strict Lyapunov exponent* of y. For further related results, see [6, 10].

Let K be a cone in $\mathbb{R}^k$. A solution $y : \mathbb{R}^+ \to \mathbb{R}^k$ of (4) is called K-positive if $y(t)$ is K-positive for all $t \in \mathbb{R}^+$. Our main result is the following analogue of Theorem 1 for the K-positive solutions of (4).

Theorem 2 *Suppose* (6) *holds and let K be a cone in $\mathbb{R}^k$. If y is a K-positive solution of* (4)*, then its strict Lyapunov exponent $\lambda(y)$ is a real eigenvalue of C with a K-positive eigenvector.*

In the special case when K is the nonnegative orthant in $\mathbb{R}^k$, the conclusion of Theorem 2 was proved in [11]. For general results on positive linear systems, see [1, 5].

2 Proof

The proof of Theorem 2 will be based on the following corollary of Theorem 1 for the nonhomogeneous difference equation

$$x(n+1) = Ax(n) + f(n), \qquad n \in \mathbb{Z}^+, \tag{8}$$

where $A \in \mathbb{R}^{k\times k}$ and $f : \mathbb{Z}^+ \to \mathbb{R}^k$.

Proposition 1 *Let K be a cone in $\mathbb{R}^k$. If x is a K-positive solution of* (8) *such that*

$$\frac{f(n)}{\|x(n)\|} \to 0, \qquad n \to \infty, \tag{9}$$

then its exponential growth rate $\rho(x)$ is a nonnegative eigenvalue of A with a K-positive eigenvector.

Proof Since conditions (3) and (9) are independent of the norm used we may (and do) consider the Euclidean norm. It is easily seen that under the hypotheses of the proposition x is a K-positive solution of (1) with

$$B(n) = \frac{f(n)[x(n)]^T}{\|x(n)\|^2}, \qquad n \in \mathbb{Z}^+,$$

where T denotes the transpose. It is easily shown that

$$\|B(n)\| = \frac{\|f(n)\|}{\|x(n)\|}, \qquad n \in \mathbb{Z}^+.$$

Therefore (9) implies (2) and the conclusion follows from Theorem 1. □

We will also need the following corollary of the Dunford functional calculus for bounded linear operators [3]. As usual, if $M \in \mathbb{R}^{k\times k}$, then $\sigma(M)$ denotes the *spectrum*, the set of eigenvalues, of M.

Proposition 2 *For every $M \in \mathbb{R}^{k\times k}$ and $t \in \mathbb{R}$, we have the spectral mapping formula*

$$\sigma(e^{tM}) = e^{t\sigma(M)} = \{e^{t\lambda} \mid \lambda \in \sigma(M)\}. \tag{10}$$

Now we are in a position to give a proof of Theorem 2.

Proof Let $h \in (0, 1)$ be fixed. By the variation of constants formula, we obtain for $n \in \mathbb{Z}^+$ and $s \geq nh$,

$$y(s) = e^{C(s-nh)}y(nh) + \int_{nh}^{s} e^{C(s-u)}D(u)y(u)\,du \tag{11}$$

and hence

$$\|y(s)\| \leq e^{\|C\|(s-nh)}\|y(nh)\| + \int_{nh}^{s} e^{\|C\|(s-u)}\|D(u)\|\|y(u)\|\,du.$$

From this, we find for $n \in \mathbb{Z}^+$ and $s \geq nh$,

$$e^{-\|C\|s}\|y(s)\| \leq e^{-\|C\|nh}\|y(nh)\| + \int_{nh}^{s} \|D(u)\|e^{-\|C\|u}\|y(u)\|\,du.$$

By Gronwall's lemma, we obtain for $n \in \mathbb{Z}^+$ and $s \geq nh$,

$$e^{-\|C\|s}\|y(s)\| \leq e^{-\|C\|nh}\|y(nh)\| \exp\left(\int_{nh}^{s} \|D(u)\|\,du\right)$$

and hence

$$\|y(s)\| \leq e^{\|C\|(s-nh)}\|y(nh)\| \exp\left(\int_{nh}^{s} \|D(u)\|\,du\right).$$

From this and (6), we obtain

$$\|y(s)\| \leq L\|y(nh)\| \qquad \text{whenever } n \in \mathbb{Z}^+ \text{ and } s \in [nh, (n+1)h], \tag{12}$$

where

$$L = e^{\|C\|h} \exp\left(\sup_{t\geq 0} \int_{t}^{t+1} \|D(u)\|\,du\right) < \infty.$$

Writing $s = (n+1)h$ in (11), we obtain for $n \in \mathbb{Z}^+$,

$$y((n+1)h) = e^{Ch}y(nh) + g(nh), \tag{13}$$

where

$$g(t) = \int_t^{t+h} e^{C(t+h-u)} D(u)y(u)\,du, \qquad t \in \mathbb{R}^+.$$

From (12), we find for $n \in \mathbb{Z}^+$,

$$\begin{aligned}\|g(nh)\| &\le \int_{nh}^{(n+1)h} e^{\|C\|((n+1)h-u)} \|D(u)\|\,\|y(u)\|\,du \\ &\le e^{\|C\|h} L \int_{nh}^{(n+1)h} \|D(u)\|\,\|y(nh)\|\,du.\end{aligned}$$

This, together with (6), implies

$$\frac{\|g(nh)\|}{\|y(nh)\|} \le e^{\|C\|h} L \int_{nh}^{(n+1)h} \|D(u)\|\,du \to 0, \qquad n \to \infty. \tag{14}$$

If we let

$$x(n) = y(nh), \qquad n \in \mathbb{Z}^+,$$

and

$$f(n) = g(nh), \qquad n \in \mathbb{Z}^+,$$

then (13) implies that x is a K-positive solution of (8) with $A = e^{Ch}$. Since (14) implies (9), we can apply Proposition 1. Therefore the limit

$$\rho(h) = \lim_{n\to\infty} \sqrt[n]{\|y(nh)\|} \tag{15}$$

is an eigenvalue of $A = e^{Ch}$ with a K-positive eigenvector $v(h)$. In view of the cone property (ii), we may (and do) assume that $\|v(h)\| = 1$. By Proposition 2, we have

$$\rho(h) = e^{h\lambda(h)} \qquad \text{for some } \lambda(h) \in \sigma(C). \tag{16}$$

Hence

$$e^{Ch} v(h) = e^{h\lambda(h)} v(h). \tag{17}$$

Take a sequence $h_j \in (0, 1)$ such that $h_j \to 0$ as $j \to \infty$. Since the sequences $\{\lambda(h_j)\}$ and $\{v(h_j)\}$ are bounded, there exists a subsequence $\{h_{j_i}\}$ of $\{h_j\}$ such that the limits

$$\lambda = \lim_{i\to\infty} \lambda(h_{j_i}) \tag{18}$$

and

$$v = \lim_{i\to\infty} v(h_{j_i}) \tag{19}$$

exist and are finite. Since K is a closed set, the limit vector v belongs to K. From this and the fact that $\|v\| = 1$, it follows that v is K-positive. Since $\sigma(C)$ is a finite set, (18) implies that $\lambda(h_{j_i}) = \lambda$ for all large i. This, together with (16) and (17), implies for all large i,

$$\rho(h_{j_i}) = e^{\lambda h_{j_i}} \tag{20}$$

and

$$e^{Ch_{j_i}} v(h_{j_i}) = e^{\lambda h_{j_i}} v(h_{j_i}) \tag{21}$$

and hence

$$\frac{e^{Ch_{j_i}} - I}{h_{j_i}} v(h_{j_i}) = \frac{e^{\lambda h_{j_i}} - 1}{h_{j_i}} v(h_{j_i}).$$

From this, letting $i \to \infty$ and taking into account that $h_{j_i} \to 0$ as $i \to \infty$, we find that

$$Cv = \lambda v.$$

Thus, λ is a real eigenvalue of C. Choose h_{j_i} such that (20) holds. From (15) and (20), we find that

$$\lambda(y) = \lim_{t\to\infty} \frac{\log \|y(t)\|}{t} = \lim_{n\to\infty} \frac{\log \|y(nh_{j_i})\|}{nh_{j_i}} = \frac{1}{h_{j_i}} \lim_{n\to\infty} \log \sqrt[n]{\|y(nh_{j_i})\|}$$
$$= \frac{1}{h_{j_i}} \log \lim_{n\to\infty} \sqrt[n]{\|y(nh_{j_i})\|} = \frac{1}{h_{j_i}} \log \rho(h_{j_i}) = \frac{1}{h_{j_i}} \log e^{\lambda h_{j_i}} = \lambda.$$

Thus, $\lambda(y)$ is a real eigenvalue of C and v is a corresponding K-positive eigenvector. □

Acknowledgements This work was supported in part by the Hungarian National Foundation for Scientific Research (OTKA) Grant No. K120186.

References

1. Berman, A., Plemmons, R.: Nonnegative Matrices in the Mathematical Sciences. Academic Press, New York (1979)
2. Coppel, W.A.: Stability and Asymptotic Behavior of Differential Equations. Heath, Boston (1965)
3. Dunford, N., Schwartz, J.T.: Linear Operators, Part I: General Theory. Interscience, New York (1988)
4. Elaydi, S.: An Introduction to Difference Equations. Springer, New York (2005)
5. Krause, U.: Positive Dynamical Systems in Discrete Time. De Gruyter, Berlin (2015)
6. Matsui, K., Matsunaga, H., Murakami, S.: Perron type theorems for functional differential equations with infinite delay in a Banach space. Nonlinear Anal. **69**, 3821–3837 (2008)
7. Matsunaga, H., Murakami, S.: Asymptotic behavior of solutions of functional difference equations. J. Math. Anal. Appl. **305**, 391–410 (2005)

8. Obaya, R., Pituk, M.: A variant of the Krein–Rutman theorem for Poincaré difference equations. J. Differ. Equ. Appl. **18**, 1751–1762 (2012)
9. Pituk, M.: More on Poincaré's and Perron's theorems for difference equations. J. Differ. Equ. Appl. **8**, 201–216 (2002)
10. Pituk, M.: A Perron theorem for functional differential equations. J. Math. Anal. Appl. **316**, 24–41 (2006)
11. Pituk, M.: A note on nonnegative solutions of a perturbed system of ordinary differential equations. Annales Univ. Sci. Budapest. **53**, 91–96 (2010)

The Case for Large Contraction in Functional Difference Equations

Youssef N. Raffoul

Abstract In this note we review some of the latest research on the qualitative analysis of solutions of difference equations using fixed point theory and Lyapunov functionals. It turns out that the use of fixed point theory alleviates some of the difficulties that arise from the use of Lyapunov functionals. Using fixed point theory requires us to find a mapping from suitable spaces that is a solution of the given difference equation. Once the suitable mapping is constructed there will be many fixed pint theorems to use, depending on the given equation, that yield a fixed point of that mapping and satisfies our initial value problem. In some cases a regular contraction argument will not be suitable and hence we replace it with what we call Large Contraction.

Keywords Large Contraction · Fixed point theory · Lyapunov functionals

1 Introduction

Most of real life applications are modeled by nonlinear systems for which implicit solutions can not be explicitly stated. This necessitates the qualitative analysis of such systems and in particular the study of how solutions behave with time. Biologists are interested in solutions remaining bounded and the exhibition of periodic behavior of solutions. For example, in [13] it is shown that there is a direct connection between boundedness of solutions and for solutions to exhibit a periodic behavior. In the paper [5] the authors considered a dynamical system and proved ultimate boundedness implied periodicity provided given functions are periodic. In addition, in the papers [10]–[12], the author used the notions of Lyapunov functionals and fixed point theory and obtained necessary conditions for the boundedness and ultimate boundedness and the existence of periodic solutions of functional difference equations of the form

$$x(n+1) = G(n, x(s);\ 0 \le s \le n) \stackrel{\text{def}}{=} G(n, x(\cdot)) \tag{1}$$

Y.N. Raffoul (✉)
Department of Mathematics, University of Dayton, Dayton, OH 45469-2316, US
e-mail: yraffoul1@udayton.edu

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_13

where $G : \mathbb{Z}^+ \times \mathbb{R}^k \to \mathbb{R}^k$ is continuous in x. During our analysis of (1), we encountered endless difficulties due to the pair of inequalities

$$W_1(|x(n)|) \leq V(n, x(\cdot)) \leq W_2(|x(n)|) \tag{2}$$

and

$$\Delta V(n, x(\cdot)) \leq -\rho W_3(|x(n)|) + K \tag{3}$$

for some constants ρ and $K \geq 0$.

In the past hundred and fifty years, Lyapunov functions/functionals have been exclusively and successfully used in the study of stability and existence of periodic and bounded solutions. This author has extensively used Lyapunov functions/functionals for the purpose of analyzing solutions of functional equations and each time the suitable Lyapunov functional presented us with unique difficulties that could only overcome by the imposition of severe conditions on the given coefficients. In practice, Lyspunov direct method requires pointwise conditions, while as so many real-life problems call for averages. Moreover, it is rare that we encounter a problem for which a suitable Lyapunov functional can be easily constructed. It is common knowledge among researchers that stability and boundedness results go hand in hand with the type of the Lyapunov functional that was used. To illustrate our concern, we consider the delay difference equation

$$x(t+1) = b(t)x(t) + a(t)x(t-\tau) + p(t), \;\; t \in \mathbb{Z}, \tag{4}$$

where $a, b, p : \mathbb{Z} \to \mathbb{R}$, τ is a positive integer, and assume the strong condition

$$|b(t)| < 1, \;\; \text{for all} \;\; t \in \mathbb{Z}. \tag{5}$$

For more on recent results regarding stability in difference equations we refer the reader to [2–4, 6, 8, 9, 14–16]. For the sake of completeness, we assume that there is a $\delta > 0$ such that

$$|b(t)| + \delta < 1, \tag{6}$$

and

$$|a(t)| \leq \delta, \;\; \text{and} \; |p(t)| \leq K, \;\; \text{for some positive constant} \;\; K. \tag{7}$$

Then all solutions of (4) are bounded. To see this we consider the Lyapunov function

$$V(t, x(\cdot)) = |x(t)| + \delta \sum_{s=t-\tau}^{t-1} |x(s)|.$$

Then along solutions of (4) we have

$$\begin{aligned}\Delta V &= |x(t+1)| - |x(t)| + \delta \sum_{s=t+1-\tau}^{t} |x(s)| - \delta \sum_{s=t-\tau}^{t-1} |x(s)| \\ &\leq |b(t)||x(t)| - |x(t)| + |a(t)||x(t-\tau)| + \delta \sum_{s=t+1-\tau}^{t} |x(s)| - \delta \sum_{s=t-\tau}^{t-1} |x(s)| + |p(t)| \\ &= \big(|b(t)| + \delta - 1\big)|x(t)| + \big(|a(t)| - \delta\big)|x(t-\tau)| + |p(t)| \\ &\leq \big(|b(t)| + \delta - 1\big)|x(t)| + |p(t)| \\ &\leq -\gamma|x(t)| + |p(t)|, \text{ for some positive constant } \gamma.\end{aligned}$$

It follows from the above pairs of inequalities in (2), (3) and of [12] that all solutions of (4) are bounded. It is evident conditions (5) and (6) are somewhat strong.

In this paper we use fixed point theory that will requires us to find a mapping from suitable spaces that is a solution of the given difference equation. Once the suitable mapping is constructed there will be many fixed pint theorems to use (depending on conditions) that yield a fixed point of that mapping and satisfies our initial value problem. As we shall see later, in some cases a regular contraction argument will not be suitable and hence we replace it with what we call Large Contraction. However, a general care must be taken when formulating the required mapping. For example, we consider the initial value problem

$$\Delta x(t) = g(t, x(t)), \;\; x(t_0) = x_0, \tag{8}$$

where $g : \mathbb{Z} \times \mathbb{R}^k \to \mathbb{R}^k$ is continuous in x. The question is, how can we show solutions of (8) are bounded. One would pick the set S as follows: for a given $(t_0, x_0) \in \mathbb{Z} \times \mathbb{R}^k$, let S be the set of functions $\phi : \mathbb{Z} \to \mathbb{R}^k$, which are bounded and satisfy $\phi(t_0) = x_0$. Next, we must select an appropriate mapping. If we try to write

$$\big(P\phi\big)(t) = x_0 + \sum_{s=t_0}^{t-1} g(s, \phi(s)),$$

then we instantly have difficulties. Suppose $\phi \in S$, then there is no way of arriving at $\big(P\phi\big)$ is bounded. That mapping will map a given bounded function ϕ right out of the set and there is no way of proving that there is a fixed point.

Let $g(t, x) = Ax + f(t, x)$, where A is a $k \times k$ real constant matrix with all its eigenvalues residing inside the unit circle and $f : \mathbb{Z} \times \mathbb{R}^k \to \mathbb{R}^k$ is continuous in x and bounded for bounded x. Now we assume the function f satisfies a Lipschitz condition. That is, there exists a positive constant L such that

$$|f(t, x) - f(t, y)| \leq L|x - y| \tag{9}$$

for $t \in \mathbb{Z}$ and $x, y \in \mathbb{R}^k$. For a given $(t_0, x_0) \in \mathbb{Z} \times \mathbb{R}^k$, by the variation of parameters, we have that for $t \geq t_0$

$$x(t) = A^{t-t_0}x_0 + \sum_{s=t_0}^{t-1} A^{t-s-1} f(s, x(s)). \tag{10}$$

Then $x(t)$ given by (10) is a solution of

$$x(t+1) = Ax(t) + f(t, x(t)),\ \ x(t_0) = x_0,$$

see [3] or [15].

Let $(S, \|\cdot\|)$ be a complete metric space of bounded sequences $\phi : \mathbb{Z} \to \mathbb{R}^k$ with the maximum metric and satisfying $\phi(t_0) = x_0$. By the assumption on A we can find positive constants l and $\eta \in (0, 1)$ such that $|A^t| \leq l\eta^t$, for $t \geq 0$. Define the mapping P using (10). As $f(t, x)$ is bounded for bounded x, it is easy to show that $P : S \to S$. Moreover, for $\phi_1, \phi_2 \in S$ we have that

$$|(P\phi_1)(t) - (P\phi_2)(t)| \leq \sum_{s=t_0}^{t-1} Ll\eta^{t-s-1}|\phi_1(s) - \phi_2(s)| \leq \frac{lL}{1-\eta}\|\phi_1 - \phi_2\|,$$

a contraction provided $lL/(1-\eta) < 1$. Hence, we have a unique fixed point ϕ, bounded and solution of our problem.

The reader should have been suspicious about the fact that we may easily create a linear term in $g(t, x)$ as the next example shows that this maybe a naive approach and may not work for all functions $g(t, x)$. Asking that $g(t, x) = Ax + f(t, x)$ is not much of an assumption since we can write

$$x(t+1) = Ax(t) + g(t, x(t)) - Ax(t).$$

As we have seen, we had to ask for $f(t, x)$ to satisfy (at least a local) contraction condition. Now, suppose our equation is scalar and that

$$x(t+1) = -x(t)^3 + h(t, x(t)),$$

where $h(t, x)$ satisfies a bound condition for bounded x. We put our equation in the form

$$x(t+1) = -x(t) + (x(t) - x(t)^3) + h(t, x(t)),$$

and by the variations of parameters formula we arrive at

$$x(t) = (-1)^t x_0 + \sum_{s=t_0}^{t-1} (-1)^{t-s-1}\{x(s) - x(s)^3 + h(s, x(s))\}. \tag{11}$$

In showing (11) define a contraction mapping, we encounter that for $x^2 + y^2 \le 1/2$,

$$|x - x^3 - y + y^3| \le |x - y| \left(1 - \frac{x^2 + y^2}{2}\right)$$

and the contraction constant tends to one as $x^2 + y^2 \to 0$.

As a consequence, the regular contraction mapping principle failed to produce any results. In the next section, we define a new concept on contraction, called Large Contraction and prove a parallel result to the Contraction Mapping Principle. Then based on the notion of Large Contraction, we introduce two theorems, in which Large Contraction is used in place of regular contraction.

2 Large Contraction; Boundedness

We begin this section by introducing the concept of Large Contraction.

Definition 1 *Let $(\mathscr{M}, d)$ be a metric space and $B : \mathscr{M} \to \mathscr{M}$. The map B is said to be large contraction if $\phi, \varphi \in \mathscr{M}$, with $\phi \ne \varphi$ then $d(B\phi, B\varphi) \le d(\phi, \varphi)$ and if for all $\varepsilon > 0$, there exists a $\delta \in (0, 1)$ such that*

$$[\phi, \varphi \in \mathscr{M}, d(\phi, \varphi) \ge \varepsilon] \Rightarrow d(B\phi, B\varphi) \le \delta d(\phi, \varphi).$$

The next theorems are alternative to the regular Contraction Mapping Principle, and, Krasnoselskii fixed point theorem in which we substitute Large Contraction for regular contraction. The proofs of the two theorems and the statement of Definition 1 can be found in [1].

Theorem 1 *Let $(\mathscr{M}, \rho)$ be a complete metric space and B be a large contraction. Suppose there are an $x \in \mathscr{M}$ and an $L > 0$ such that $\rho(x, B^n x) \le L$ for all $n \ge 1$. Then B has a unique fixed point in $\mathscr{M}$.*

Theorem 2 *Let $\mathscr{M}$ be a bounded convex nonempty subset of a Banach space $(\mathbb{B}, \|\cdot\|)$. Suppose that A and B map $\mathscr{M}$ into $\mathbb{B}$ such that*

i. $x, y \in \mathscr{M}$ implies $Ax + By \in \mathscr{M}$;
ii. A is compact and continuous;
iii. B is a large contraction mapping.

Then there exists $z \in \mathscr{M}$ with $z = Az + Bz$.

Next, we consider the completely nonlinear difference equation

$$x(t + 1) = a(t)x(t)^5 + p(t), \tag{12}$$

where $a, p : \mathbb{Z} \to \mathbb{R}$. To invert our equation, we create a linear term by letting

$$H(x) = -x + x^5. \tag{13}$$

It would become clearer later on that Hx is not a contraction and as a consequence the Contraction Mapping Principle can not be used. Instead, we will show that H is a Large Contraction and hence our mapping, to be constructed, will define a Large Contraction. Then we use Theorem 1 and show that solutions of (12) are bounded. This allows us to rewrite (12) in the form

$$x(t+1) - a(t)x(t) = a(t)H(x(t)) + p(t). \tag{14}$$

Let $x(0) = x_0$, then by the variation of parameters formula, one can easily show that for $t \geq 0$, $x(t)$ is a solution of (14) if and only if

$$x(t) = x_0 \prod_{s=0}^{t-1} a(s) + \sum_{s=0}^{t-1} \Big(a(s)H(x(s)) \prod_{u=s+1}^{t-1} a(u) \Big) + \sum_{s=0}^{t-1} \Big(p(s) \prod_{u=s+1}^{t-1} a(u) \Big). \tag{15}$$

We begin with the following lemma.

Lemma 1 *Let $\|\cdot\|$ denote the maximum norm. If*

$$\mathbb{M} = \left\{ \phi : \mathbb{Z} \to \mathbb{R} \mid \phi(0) = \phi_0, \quad \text{and } \ \|\phi\| \leq 5^{-1/4} \right\},$$

then the mapping H defined by (13) is a large contraction on the set $\mathbb{M}$.

Proof For any reals a and b we have the following inequalities

$$0 \leq (a+b)^4 = a^4 + b^4 + ab(4a^2 + 6ab + 4b^2),$$

and

$$-ab(a^2 + ab + b^2) \leq \frac{a^4 + b^4}{4} + \frac{a^2b^2}{2} \leq \frac{a^4 + b^4}{2}.$$

If $x, y \in \mathbb{M}$ with $x \neq y$, then $x(t)^4 + y(t)^4 < 1$. Hence, we arrive at

$$\begin{aligned} |H(u) - H(v)| &\leq |u - v| \left| 1 - \left(\frac{u^5 - v^5}{u - v} \right) \right| \\ &= |u - v| \left\{ 1 - u^4 - v^4 - uv(u^2 + uv + v^2) \right\} \\ &\leq |u - v| \left\{ 1 - \frac{(u^4 + v^4)}{2} \right\} \leq |u - v|, \end{aligned} \tag{16}$$

where we use the notations $u = x(t)$ and $v = y(t)$ for brevity. Now, we are ready to show that H is a large contraction on $\mathbb{M}$. For a given $\varepsilon \in (0, 1)$, suppose $x, y \in \mathbb{M}$ with $\|x - y\| \geq \varepsilon$. There are two cases:

a.
$$\frac{\varepsilon}{2} \leq |x(t) - y(t)| \ \text{ for some } \ t \in \mathbb{Z},$$

or

b.
$$|x(t) - y(t)| \leq \frac{\varepsilon}{2} \ \text{ for some } \ t \in \mathbb{Z}.$$

If $\varepsilon/2 \leq |x(t) - y(t)|$ for some $t \in \mathbb{Z}$, then

$$(\varepsilon/2)^4 \leq |x(t) - y(t)|^4 \leq 8(x(t)^4 + y(t)^4),$$

or

$$x(t)^4 + y(t)^4 \geq \frac{\varepsilon^4}{2^7}.$$

For all such t, we get by (16) that

$$|H(x(t)) - H(y(t))| \leq |x(t) - y(t)| \left(1 - \frac{\varepsilon^4}{2^7}\right).$$

On the other hand, if $|x(t) - y(t)| \leq \varepsilon/2$ for some $t \in \mathbb{Z}$, then along with (16) we find

$$|H(x(t)) - H(y(t))| \leq |x(t) - y(t)| \leq \frac{1}{2}\|x - y\|.$$

Hence, in both cases we have

$$|H(x(t)) - H(y(t))| \leq \min\left\{1 - \frac{\varepsilon^4}{2^7}, \frac{1}{2}\right\} \|x - y\|.$$

Thus, H is a large contraction on the set $\mathbb{M}$ with $\delta = \min\left\{1 - \varepsilon^4/2^7, 1/2\right\}$. The proof is complete. □

Remark 1 It is clear from inequality (16) that $(u^4 + v^4)/2 \to 0$, the contraction constant approaches one. Hence, Hx does not define a contraction mapping as we have claimed before.

For $\psi \in \mathbb{M}$, we define the map $B : \mathbb{M} \to \mathbb{M}$ by

$$(B\psi)(t) = \psi_0 \prod_{s=0}^{t-1} a(s) + \sum_{s=0}^{t-1} \Big(a(s)H(\psi(s)) \prod_{u=s+1}^{t-1} a(u)\Big) + \sum_{s=0}^{t-1} \Big(p(s) \prod_{u=s+1}^{t-1} a(u)\Big). \tag{17}$$

Lemma 2 *Assume for all $t \in \mathbb{Z}$*

$$|\psi_0|\Big|\prod_{s=0}^{t-1} a(s)\Big| + 4(5^{-5/4})\sum_{s=0}^{t-1}\Big|\prod_{u=s}^{t-1} a(u)\Big| + \sum_{s=0}^{t-1}\Big(\Big|p(s)\prod_{u=s+1}^{t-1} a(u)\Big|\Big) \le 5^{-1/4}. \tag{18}$$

If H is a large contraction on $\mathbb{M}$, then so is the mapping B.

Proof It is easy to see that

$$|H(x(t))| = |x(t) - x(t)^5| \le 4(5^{-5/4}) \text{ for all } x \in \mathbb{M}.$$

By Lemma 1 H is a large contraction on $\mathbb{M}$. Hence, for $x, y \in \mathbb{M}$ with $x \ne y$, we have $\|Hx - Hy\| \le \|x - y\|$. Hence,

$$\begin{aligned} |Bx(t) - By(t)| &\le \sum_{s=0}^{t-1} |H(x(s)) - H(y(s))|\Big|\prod_{u=s}^{t-1} a(u)\Big| \\ &\le 4(5^{-5/4})\sum_{s=0}^{t-1}\Big|\prod_{u=s}^{t-1} a(u)\Big|\|x - y\| \\ &= \|x - y\|. \end{aligned}$$

Taking maximum norm over the set $[0, \infty)$, we get that $\|Bx - By\| \le \|x - y\|$. Now, from the proof of Lemma 1, for a given $\varepsilon \in (0, 1)$, suppose $x, y \in \mathbb{M}$ with $\|x - y\| \ge \varepsilon$. Then $\delta = \min\{1 - \varepsilon^4/2^7, 1/2\}$, which implies that $0 < \delta < 1$. Hence, for all such $\varepsilon > 0$ we know that

$$[x, y \in \mathbb{M}, \|x - y\| \ge \varepsilon] \Rightarrow \|Hx - Hy\| \le \delta\|x - y\|.$$

Therefore, using (18), one easily verify that

$$\|Bx - By\| \le \delta\|x - y\|.$$

The proof is complete. □

We arrive at the following theorem in which we prove boundedness.

Theorem 3 *Assume* (18). *Then* (14) *has a unique solution in $\mathbb{M}$ which is bounded.*

Proof $(\mathbb{M}, \|\cdot\|)$ is a complete metric space of bounded sequences. For $\psi \in \mathbb{M}$ we must show that $(B\psi)(t) \in \mathbb{M}$. From (17) and the fact that

$$|H(x(t))| = |x(t) - x(t)^5| \le 4(5^{-5/4}) \quad \text{for all } x \in \mathbb{M},$$

we have

$$|(B\psi)(t)| \leq |\psi_0|\Big|\prod_{s=0}^{t-1} a(s)\Big| + 4(5^{-5/4})\sum_{s=0}^{t-1}\Big|\prod_{u=s}^{t-1} a(u)\Big| + \sum_{s=0}^{t-1}\Big(\Big|p(s)\prod_{u=s+1}^{t-1} a(u)\Big|\Big)$$
$$\leq 5^{-1/4}.$$

This shows that $(B\psi)(t) \in \mathbb{M}$. Lemma 2 implies the map B is a large contraction and hence by Theorem 1, the map B has a unique fixed point in $\mathbb{M}$ which is a solution of (14). This completes the proof. □

3 Large Contraction; Periodicity

In this section, we use Theorem 2 and prove the existence of a periodic solution of the nonlinear delay difference equation

$$x(t+1) = a(t)x(t)^5 + G(t, x(t-r)) + p(t), \ \ t \in \mathbb{Z}, \tag{19}$$

where r is a positive integer and

$$a(t+T) = a(t), \ \ p(t+T) = p(t), \ \ \text{and } G(t+T, \cdot) = G(t, \cdot) \tag{20}$$

and T is the least positive integer for which these hold. As before, for the sake of inversion, we rewrite (19) as

$$x(t+1) - a(t)x(t) = a(t)H(x(t)) + G(t, x(t-r)) + p(t), \tag{21}$$

where

$$H(x) = -x + x^5. \tag{22}$$

For more on periodic solutions in difference equations, we refer the reader to [10], [11], and [14] and the references therein. We begin with the following Lemma which we omit its proof.

Lemma 3 *Suppose that* $1 - \prod_{s=t-T}^{t-1} a(s) \neq 0$ *for all* $t \in \mathbb{Z}$. *Then* $x(t)$ *is a solution of* (21) *if and only if*

$$x(t) = \Big(1 - \prod_{s=t-T}^{t-1} a(s)\Big)^{-1} \sum_{u=t-T}^{t-1} \big(a(u)H(x(u)) + G(t, x(u-r)) + p(u)\big) \prod_{s=u+1}^{t-1} a(s).$$

Let P_T be the set of all sequences $x(t)$, periodic in t of period T. Then $(P_T, \|\cdot\|)$ is a Banach space when it is endowed with the maximum norm

$$\|x\| = \max_{t\in\mathbb{Z}} |x(t)| = \max_{t\in[0,T-1]} |x(t)|.$$

Set

$$\mathbb{M} = \{\varphi \in P_T : \|\varphi\| \le 5^{-1/4}\}. \tag{23}$$

Obviously, $\mathbb{M}$ is bounded and convex subset of the Banach space P_T. Let the map $A : \mathbb{M} \to P_T$ be defined by

$$(A\varphi)(t) = \Big(1 - \prod_{s=t-T}^{t-1} a(s)\Big)^{-1} \sum_{u=t-T}^{t-1} (G(t, \varphi(u-r)) + p(u)) \prod_{s=u+1}^{t-1} a(s),\ t \in \mathbb{Z}. \tag{24}$$

In a similar way, we set the map $B : \mathbb{M} \to P_T$ by

$$(B\psi)(t) = \Big(1 - \prod_{s=t-T}^{t-1} a(s)\Big)^{-1} \sum_{u=t-T}^{t-1} \big(a(u)H(\psi(u))\big) \prod_{s=u+1}^{t-1} a(s),\ t \in \mathbb{Z}. \tag{25}$$

It is clear from (24) and (25) that $A\varphi$ and $B\psi$ are T-periodic in t.

For simplicity we let

$$\eta := \Big|\Big(1 - \prod_{s=t-T}^{t-1} a(s)\Big)^{-1}\Big|.$$

Let

$$G(u, \psi(u-r)) = b(u)\psi(u-r)^5. \tag{26}$$

For $x \in \mathbb{M}$, we have

$$|x(t)|^5 \le 5^{-5/4},$$

and therefore,

$$\begin{aligned} G(u, x(u-r)) + p(u) &= b(u)x(u-r)^5 + p(u) \\ &\le 5^{-5/4}|b(u)| + |p(u)| \end{aligned} \tag{27}$$

and

$$|H(x(t))| = |x(t) - x(t)^5| \le 4(5^{-5/4}) \text{ for all } x \in \mathbb{M}.$$

We have the following theorem.

Theorem 4 *Suppose $G(u, \psi(u-r))$ is given by (26). Assume for all $t \in \mathbb{Z}$*

$$\eta \sum_{u=t-T}^{t-1} \Big(5^{-5/4}|b(u)| + |p(u)| + 4(5^{-5/4})|a(u)|\Big)\Big| \prod_{u=s+1}^{t-1} a(u)\Big| \leq 5^{-1/4}. \qquad (28)$$

Then (19) has a periodic solution.

Proof Using condition (28) and by a similar argument as in Lemma 2, one can easily show that B is a large contraction since H is a large contraction. Also, the map A is continuous and maps bounded sets into compact sets and hence it is compact. Moreover, for $\varphi, \psi \in \mathbb{M}$, we have by (28) that

$$A\varphi + B\psi : \mathbb{M} \to \mathbb{M}.$$

Hence an application of Theorem 2 implies the existence of a periodic solution in $\mathbb{M}$. This completes the proof. □

References

1. Burton, T.: Integral equations, implicit functions, and fixed points. Proc. Amer. Math. Soc. **124**, 2383–2390 (1996)
2. Cermák, Jan, *Difference equations in the qualitative theory of delay differential equations*, Proceedings of the Sixth International Conference on Difference Equations, 391–398, CRC, Boca Raton, FL, 2004
3. Elaydi, S.: An Introduction to Difference Equations. Springer, New York (1999)
4. Islam, Muhammad; Yankson, Ernest, *Boundedness and stability for nonlinear delay difference equations employing fixed point theory*, Electron. J. Qual. Theory Differ. Equ. 2005, No. 26, 18 pp
5. Kaufmann, E., Kosmatov, N., Raffoul, Y.: Nonlinear Dyn. Syst. Theory. The connection between boundedness and periodicity in nonlinear functional neutral dynamic equations on a time scale **9**(1), 89–98 (2009)
6. Kelley, W., Peterson, A.: Difference Equations. An Introduction with Applications. Academic Press, San Diego (2001)
7. Maroun, M., Raffoul, Y.: Periodic solutions in nonlinear neutral difference equations with functional delay. J. Korean Math. Soc. **42**(2), 255–268 (2005)
8. Migda, J.: Asymptotic behavior of solutions of nonlinear difference equations. Math. Bohem. **129**(4), 349–359 (2004)
9. Qian, C., Sun, Y.: On global attractivity of nonlinear delay difference equations with a forcing term. J. Differ. Equ. Appl. **11**(3), 227–243 (2005)
10. Raffoul, Y.: Stability and periodicity in discrete delayed equations. J. Math. Anal. Appl. **324**(2), 1356–1362 (2006)
11. Raffoul, Y.: Periodicity in general delay non-linear difference equations using fixed point theory. J. Differ. Equ. Appl. **10**(13–15), 1229–1242 (2004)
12. Raffoul, Y.: General theorems for stability and boundedness for nonlinear functional discrete systems. J. Math. Anal. Appl. **279**(2), 639–650 (2003)
13. Zhang, G.: Periodicity and attractivity of a nonlinear higher order difference equation. Appl. Math. Comput. **161**(2), 395–401 (2005)

14. Wan-Tong, L., Hai-Feng, H.: Positive periodic solutions of delay difference equations and applications in population dynamics. J. Comput. Appl. Math. **176**(2), 357–369 (2005)
15. Zhu, H., Huang, L.: Asymptotic behavior of solutions for a class of delay difference equation. Ann. Differ. Equ. **21**(1), 99–105 (2005)
16. Zhang, D.C., Shi, B.: Global behavior of solutions of a nonlinear difference equation. Appl. Math. Comput. **159**(1), 29–35 (2004)

Reaching Consensus via Polynomial Stochastic Operators: A General Study

Mansoor Saburov and Khikmat Saburov

Abstract In this paper, we consider a nonlinear protocol for a structured time-varying synchronous multi-agent system in which an opinion sharing dynamics is presented by non-autonomous polynomial stochastic operators associated with high-order stochastic hyper-matrices. We show that the proposed nonlinear protocol generates the Krause mean process. We provide a criterion to establish a consensus in the multi-agent system under the proposed nonlinear protocol.

Keywords Krause mean process · Markov chain with memory · Stochastic hyper-matrices · Polynomial stochastic operators · Consensus

1 Introduction

In the classical case, an opinion sharing dynamics of a structured time-varying synchronous multi-agent system is presented by the backward product of square stochastic matrices meanwhile a non-homogeneous Markov chain is presented by the forward product of square stochastic matrices. Therefore, the consensus in the multi-agent system and the ergodicity of the Markov chain are dual problems to each other. A more general model of the opinion sharing dynamics is *the Krause mean process* whereas *the Markov chain with memory* (or *the nonlinear Markov chain*) is a general model of the Markov chain. In this paper, we study a correlation between the Markov chains with memory (the nonlinear Markov chains) and the Krause mean processes. The reader may refer to the monographs [11, 12] for the great exposition of the Krause mean processes and the nonlinear Markov chains. A polynomial sto-

M. Saburov (✉)
Faculty of Science, International Islamic University Malaysia,
25200 Kuantan, Pahang, Malaysia
e-mail: msaburov@gmail.com

K. Saburov
Mathematical Modeling Lab, MIMOS BERHAD, Technology Park Malaysia,
57000 Kuala Lumpur, Malaysia
e-mail: khikmat.saburov@mimos.my

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_14

chastic operator is the simplest *nonlinear Markov operator*. Unlike linear stochastic operators, the structure of a set of all fixed points (stationary distributions) of polynomial stochastic operators (the Markov chains with memory) might be as complex as possible (see [16, 21]). In general, the analogy of the classical Perron–Frobenius theorem does not hold for polynomial stochastic operators associated with positive high-order stochastic hyper-matrix (see [16, 21]). However, under some extra conditions, the ergodicity of nonlinear Markov operators (polynomial stochastic operators associated with stochastic hyper-matrices) acting on the finite dimensional space has been studied in the paper [16]. In this paper, by exploring the same techniques, we are aiming to establish a consensus in the multi-agent system in which an opinion sharing dynamics is presented by non-autonomous polynomial stochastic operators associated with high-order stochastic hyper-matrices. We also show that the proposed nonlinear protocol generates the Krause mean process. It is worth mentioning that, in general, the Krause mean process eventually reaches to a consensus if and only if it eventually shrinks at some point (see [8–11]). In this paper, we improve Krause's result (see [8, 9]) in the special case where the mean process generates by non-autonomous polynomial stochastic operators associated with triply stochastic hyper-matrices. This is the novelty of the paper.

It is also worth mentioning that there are a lot of very recent researches on this topic done in time scale calculus, fractional calculus (see [14]).

2 A General Model of Opinion Dynamics

Opinion dynamics is the formation of opinions in a group of interacting individuals (decision units) so-called *agents*. Opinions could be assessments made by the agents of certain magnitudes as, for example, prices of goods or probabilities of events, in which they can be represented by nonnegative real numbers. In more complex cases opinions might be better modeled by vectors or more general mathematical objects (see [11]). The main problem is to find some conditions in which the opinions of all the agents converge to a common value. This is called *a consensus* among the agents. The quest for consensus depends very much on the structure of interaction among the agents. We first review a general model of opinion sharing dynamics of the multi-agent system presented in [5] which encompasses all classical models of opinion sharing dynamics [2–4].

Consider a group of m individuals $[m] = \{1, 2, \cdots, m\}$ acting together as a team or committee, each of whom can specify his/her own subjective distribution for some given task. It is assumed that if the individual i is informed of the distributions of each of the other members of the group then he/she might wish to revise his/her subjective distribution to accommodate the information.

Let $\mathbf{x}(t) = (x_1(t), \cdots, x_m(t))^T$ be the subjective distributions of the multi-agent system at the time t. Let $p_{ij}(t, \mathbf{x}(t))$ denote the weight that the individual i assigns to $x_j(t)$ when he/she makes the revision at the time $t + 1$. It was assumed that

$p_{ij}(t, \mathbf{x}(t)) \geq 0$ and $\sum_{j=1}^{m} p_{ij}(t, \mathbf{x}(t)) = 1$. After being informed of the subjective distributions of the other members of the group, the individual i revises his/her own subjective distribution from $x_i(t)$ to $x_i(t+1) = \sum_{j=1}^{m} p_{ij}(t, \mathbf{x}(t))x_j(t)$.

Let $\mathbb{P}(t, \mathbf{x}(t))$ denote an $m \times m$ row-stochastic matrix whose (ij) element is $p_{ij}(t, \mathbf{x}(t))$. *A general model of the structured time-varying synchronous system* is defined as follows

$$\mathbf{x}(t+1) = \mathbb{P}(t, \mathbf{x}(t))\,\mathbf{x}(t). \tag{1}$$

We may then obtain all classical models [2–5] by choosing suitable row-stochastic matrices $\mathbb{P}(t, \mathbf{x}(t))$.

We say that *a consensus* is reached in the structured time-varying synchronous multi-agent system (1) if $\mathbf{x}(t)$ converges to $\mathbf{c} = (c, \cdots, c)^T$ as $t \to \infty$. It is worth mentioning that the consensus $\mathbf{c} = \mathbf{c}(\mathbf{x}(0))$ might depend on an initial opinion $\mathbf{x}(0)$.

A more general model of an opinion sharing dynamics in which opinions are presented by vectors is called *a mean process*. The reader may refer to an excellent monograph written by Krause [11] for a detailed exposition of mean processes.

Let S be a non-empty convex subset of $\mathbb{R}^d$ and S^m be the $m-$fold Cartesian product of S. A sequence $\{\mathbf{x}(t)\}_{t=0}^{\infty} \subset S^m,\ \mathbf{x}(t) = (x_1(t), \cdots, x_m(t))^T$ is called *the Krause mean process* on S^m if $x_i(t+1) \in \mathbf{conv}\{x_1(t), \cdots, x_m(t)\}$ for all $1 \leq i \leq m$ and for all $t = 0, 1, \cdots$. In other words, a sequence $\{\mathbf{x}(t)\}_{t=0}^{\infty} \subset S^m$ is *the Krause mean process* if $\mathbf{conv}\{x_1(t+1), \cdots, x_m(t+1)\} \subset \mathbf{conv}\{x_1(t), \cdots, x_m(t)\}$ for all $t = 0, 1, \cdots$ where $\mathbf{conv}\{A\}$ is a convex hull of a set A. A mapping $T : S^m \to S^m$ is called *the Krause mean operator* if its trajectory $\{\mathbf{x}(t)\}_{t=0}^{\infty},\ \mathbf{x}(t) = T^t(\mathbf{x}(0))$ starting from any initial point $\mathbf{x}(0) \in S^m$ generates the Krause mean process on S^m.

It is worth mentioning that the nonlinear model of opinion sharing dynamics given by (1) is the Krause mean process due to the fact that the action of a stochastic matrix $\mathbb{P} = (p_{ij})_{i,j=1}^m$ on a vector $\mathbf{x} = (x_1, \cdots, x_m)^T$ can be viewed as formation of arithmetic means $(\mathbb{P}\mathbf{x})_i = \sum_{j=1}^{m} p_{ij}x_j$ with weights p_{ij}. The various kinds of nonlinear models of mean processes have been studied in the series of papers [5–10].

3 A Markov Chain with Memory

We know that the mean process and the Markov chain are dual processes to each other. We now recall some definitions in the theory of Markov chains [22].

Recall that a discrete-time *Markov chain* (or *a Markov chain with memory 1*) is a stochastic process with a sequence of random variables $\{X_t,\ t = 0, 1, 2\ldots\}$, which takes on values in a discrete finite state space $[m] = \{1, \ldots, m\}$ for a positive integer m such that

$$\begin{aligned} p_{i_1 j} &= \Pr\left(X_{t+1} = j | X_t = i_1, X_{t-1} = i_2, \ldots, X_1 = i_t, X_0 = i_{t+1}\right) \\ &= \Pr\left(X_{t+1} = j | X_t = i_1\right) \end{aligned}$$

where $i_1, \cdots, i_k, \cdots, i_{t+1}, j \in [m]$ and $\sum_{j=1}^{m} p_{i_1 j} = 1,\ p_{i_1 j} \geq 0,\ 1 \leq i_1, j \leq m.$

In other words, the probability of moving to the next state depends only on the present state and not on the previous states. A stochastic matrix $\mathbb{P} = \left(p_{i_1 j}\right)_{i_1, j=1}^{m}$ is called *one-step transition matrix* of the Markov chain.

Let $x_j(t) = \Pr\{X_t = j\}$ be the distribution of the state j at time t. The distribution of the Markov chain at time t is *a stochastic vector* $\mathbf{x}(t) = (x_1(t), \cdots, x_m(t))^T$, i.e., $\sum_{j=1}^{m} x_j(t) = 1$ and $x_j(t) \geq 0$ for any $1 \leq j \leq m$. The transition of the distributions from $\mathbf{x}(t)$ to $\mathbf{x}(t+1)$ is then governed by the rule

$$x_j(t+1) = \sum_{i_1=1}^{m} p_{i_1 j} x_{i_1}(t), \quad 1 \leq j \leq m.$$

Let $\Delta^{m-1} = \left\{x \in \mathbb{R}^m : \sum_{k=1}^{m} x_k = 1,\ x_k \geq 0,\ 1 \leq k \leq m\right\}$ be an $(m-1)-$dimensional simplex. A linear operator $\mathfrak{L} : \Delta^{m-1} \to \Delta^{m-1}$ associated with the stochastic matrix $\mathbb{P} = \left(p_{i_1 j}\right)_{i_1, j=1}^{m}$ as $\mathfrak{L}(\mathbf{x}) = \mathbf{x}^T \mathbb{P}$, i.e.,

$$(\mathfrak{L}(\mathbf{x}))_j = \sum_{i_1=1}^{m} p_{i_1 j} x_{i_1}, \quad 1 \leq j \leq m \tag{2}$$

is called *a linear Markov operator*.

A discrete-time *Markov chain with memory k* (*a k−order Markov chain*, see [1, 15]) is a stochastic process with a sequence of random variables $\{X_t,\ t = 0, 1, 2 \ldots\}$, which takes on values in a discrete finite state space $[m] = \{1, \ldots, m\}$ for a positive integer m such that

$$\begin{aligned} p_{i_1 \cdots i_k j} &= \Pr\left(X_{t+1} = j | X_t = i_1, X_{t-1} = i_2, \ldots, X_1 = i_t, X_0 = i_{t+1}\right) \\ &= \Pr\left(X_{t+1} = j | X_t = i_1, \cdots, X_{t-k+1} = i_k\right) \end{aligned}$$

where $i_1, \cdots, i_k, \cdots, i_{t+1}, j \in [m]$ and

$$\sum_{j=1}^{m} p_{i_1 \cdots i_k j} = 1, \qquad p_{i_1 \cdots i_k j} \geq 0, \quad 1 \leq i_1, \cdots, i_k, j \leq m.$$

In other words, the probability of moving to the next state depends *only* on the past k states (see [1, 15]). If $k = 1$ then we obtain the classical Markov chain. *An*

$(k+1)-$*order* $m-$*dimensional stochastic hyper-matrix* $\mathscr{P} = (p_{i_1\cdots i_k j})_{i_1,\cdots,i_k,j=1}^{m,\cdots,m,m}$ is called *the one-step transition hyper-matrix* of the Markov chain with memory k.

Let $x_j(t) = \Pr\{X_t = j\}$ be the distribution of the state j at time t. The distribution of the Markov chain with memory k at time t is a stochastic vector $\mathbf{x}(t) = (x_1(t), \cdots, x_m(t))^T$, i.e., $\sum_{j=1}^{m} x_j(t) = 1$ and $x_j(t) \geq 0$ for any $1 \leq j \leq m$. The transition of the distributions from $\mathbf{x}(t)$ to $\mathbf{x}(t+1)$ is then governed by the rule

$$x_j(t+1) = \sum_{1\leq i_1 i_2\cdots i_k \leq m} p_{i_1 i_2\cdots i_k j} x_{i_1}(t) x_{i_2}(t-1)\cdots x_{i_k}(t-k+1), \quad 1 \leq j \leq m.$$

A polynomial stochastic operator $\mathfrak{P} : \Delta^{m-1} \to \Delta^{m-1}$ associated with the $(k+1)-$order $m-$dimensional stochastic hyper-matrix $\mathscr{P} = (p_{i_1\cdots i_k j})_{i_1,\cdots,i_k,j=1}^{m,\cdots,m,m}$

$$(\mathfrak{P}(\mathbf{x}))_j = \sum_{1\leq i_1\cdots i_k\leq m} p_{i_1\cdots i_k j} x_{i_1}\cdots x_{i_k}, \quad 1 \leq j \leq m \tag{3}$$

is called *a nonlinear Markov operator* (see [12]). In the case $k = 1$, we obtain a linear Markov operator meanwhile in the case $k = 2$, we obtain *a quadratic stochastic operator* which has an incredible application in population genetics (see [13]).

4 Mean Processes Vs Markov Chains with Memory

In this section, we establish some correlation with the Krause mean processes and Markov chains with memory k. We first introduce some notions and notations.

Definition 1 A $(k+1)-$order $m-$dimensional hyper-matrix $\mathscr{P} = (p_{i_1\cdots i_k j})_{i_1,\cdots,i_k,j=1}^{m,\cdots,m,m}$ is called *stochastic* if one has that

$$\sum_{j=1}^{m} p_{i_1\cdots i_k j} = 1, \quad p_{i_1\cdots i_k j} \geq 0, \quad 1 \leq i_1, \cdots, i_k, j \leq m.$$

Definition 2 A $(k+1)-$order $m-$dimensional hyper-matrix $\mathscr{P} = (p_{i_1\cdots i_k j})_{i_1,\cdots,i_k,j=1}^{m,\cdots,m,m}$ is called *doubly stochastic* if one has that

$$\sum_{i_k=1}^{m} p_{i_1\cdots i_k j} = \sum_{j=1}^{m} p_{i_1\cdots i_k j} = 1, \quad p_{i_1\cdots i_k j} \geq 0, \quad 1 \leq i_1, \cdots, i_k, j \leq m.$$

Let $\mathscr{P} = (p_{i_1\cdots i_k j})_{i_1,\cdots,i_k j=1}^{m,\cdots,m,m}$ be the $(k+1)-$order $m-$dimensional doubly stochastic hyper-matrix and $\mathscr{P}_l = (p_{i_1\cdots i_k l})_{i_1,\cdots,i_k=1}^{m,\cdots,m}$ be its $k-$order $m-$dimensional l^{th} subhyper-matrix for fixed l. It is clear that $\mathscr{P}_l = (p_{i_1\cdots i_k l})_{i_1,\cdots,i_k=1}^{m,\cdots,m}$ is also stochastic hyper-matrix. In the sequel, we write $\mathscr{P} = (\mathscr{P}_1|\mathscr{P}_2|\cdots|\mathscr{P}_m)$ for the $(k+1)-$order $m-$dimensional doubly stochastic hyper-matrix.

We define a polynomial stochastic operator $\mathfrak{P} : \Delta^{m-1} \to \Delta^{m-1}$ associated with $(k+1)-$order $m-$dimensional doubly stochastic hyper-matrix $\mathscr{P} = (\mathscr{P}_1|\cdots|\mathscr{P}_m)$ as follows

$$(\mathfrak{P}(\mathbf{x}))_l = \sum_{i_1=1}^{m} \cdots \sum_{i_k=1}^{m} p_{i_1\cdots i_k l}\, x_{i_1}\cdots x_{i_k}, \qquad 1 \le l \le m. \tag{4}$$

We also define a polynomial stochastic operator $\mathfrak{P}_l : \Delta^{m-1} \to \Delta^{m-1}$ associated with the $k-$order $m-$dimensional stochastic hyper-matrix $\mathscr{P}_l = (p_{i_1\cdots i_k l})_{i_1,\cdots,i_k=1}^{m,\cdots,m}$ as

$$(\mathfrak{P}_l(\mathbf{x}))_j = \sum_{i_1=1}^{m} \cdots \sum_{i_{k-1}=1}^{m} p_{i_1\cdots i_{k-1} j l}\, x_{i_1}\cdots x_{i_{k-1}}, \qquad 1 \le j \le m \tag{5}$$

for all $l \in [m]$. It follows from (4) and (5) that

$$(\mathfrak{P}(\mathbf{x}))_l = \sum_{j=1}^{m} (\mathfrak{P}_l(\mathbf{x}))_j\, x_j = \big(\mathfrak{P}_l(\mathbf{x}), \mathbf{x}\big), \quad 1 \le l \le m$$

where $(\cdot,\cdot)$ stands for the standard inner product of two vectors.

Therefore, the polynomial stochastic operator $\mathfrak{P} : \Delta^{m-1} \to \Delta^{m-1}$ given by (4) can be written as follows

$$\mathfrak{P}(\mathbf{x}) = \Big(\big(\mathfrak{P}_1(\mathbf{x}), \mathbf{x}\big), \cdots, \big(\mathfrak{P}_m(\mathbf{x}), \mathbf{x}\big)\Big)^T \tag{6}$$

where $\mathfrak{P}_l : \Delta^{m-1} \to \Delta^{m-1}$ is defined by (5) for all $l \in [m]$.

We now define an $m \times m$ matrix as follows

$$\mathbb{P}(\mathbf{x}) = \begin{pmatrix} \big(\mathfrak{P}_1(\mathbf{x})\big)_1 & \big(\mathfrak{P}_1(\mathbf{x})\big)_2 & \cdots & \big(\mathfrak{P}_1(\mathbf{x})\big)_m \\ \big(\mathfrak{P}_2(\mathbf{x})\big)_1 & \big(\mathfrak{P}_2(\mathbf{x})\big)_2 & \cdots & \big(\mathfrak{P}_2(\mathbf{x})\big)_m \\ \vdots & \vdots & \ddots & \vdots \\ \big(\mathfrak{P}_m(\mathbf{x})\big)_1 & \big(\mathfrak{P}_m(\mathbf{x})\big)_2 & \cdots & \big(\mathfrak{P}_m(\mathbf{x})\big)_m \end{pmatrix}. \tag{7}$$

We show that $\mathbb{P}(\mathbf{x})$ is doubly stochastic matrix for every $\mathbf{x} \in \Delta^{m-1}$. In fact we know that $\mathbb{P}(\mathbf{x}) = \big(p_{ij}(\mathbf{x})\big)_{i,j=1}^{m}$ where

$$p_{ij}(\mathbf{x}) = \big(\mathfrak{P}_i(\mathbf{x})\big)_j = \sum_{i_1=1}^{m} \cdots \sum_{i_{k-1}=1}^{m} p_{i_1\cdots i_{k-1} j i}\, x_{i_1}\cdots x_{i_{k-1}}. \tag{8}$$

Therefore, we have that

$$\sum_{i=1}^{m} p_{ij}(\mathbf{x}) = \sum_{i_1=1}^{m} \cdots \sum_{i_{k-1}=1}^{m} \left(\sum_{i=1}^{m} p_{i_1 \cdots i_{k-1} j i} \right) x_{i_1} \cdots x_{i_{k-1}}$$
$$= \sum_{i_1=1}^{m} \cdots \sum_{i_{k-1}=1}^{m} x_{i_1} \cdots x_{i_{k-1}} = (x_1 + \cdots + x_m)^{k-1} = 1,$$
$$\sum_{j=1}^{m} p_{ij}(\mathbf{x}) = \sum_{i_1=1}^{m} \cdots \sum_{i_{k-1}=1}^{m} \left(\sum_{j=1}^{m} p_{i_1 \cdots i_{k-1} j i} \right) x_{i_1} \cdots x_{i_{k-1}}$$
$$= \sum_{i_1=1}^{m} \cdots \sum_{i_{k-1}=1}^{m} x_{i_1} \cdots x_{i_{k-1}} = (x_1 + \cdots + x_m)^{k-1} = 1.$$

Hence, it follows from (6) and (7) that

$$\mathfrak{P}(\mathbf{x}) = \mathbb{P}(\mathbf{x})\mathbf{x} \tag{9}$$

and we call it *a matrix form* of the polynomial stochastic operator (4) associated with the $(k+1)-$order $m-$dimensional doubly stochastic hyper-matrix.

Consequently, we prove the following result.

Proposition 1 *A polynomial stochastic operator* $\mathfrak{P} : \Delta^{m-1} \to \Delta^{m-1}$ *associated with* $(k+1)-$*order* $m-$*dimensional doubly stochastic hyper-matrix* $\mathscr{P} = (\mathscr{P}_1 | \cdots | \mathscr{P}_m)$ *generates the Krause mean process.*

5 Nonlinear Consensus via Polynomial Stochastic Operators

In this section, we provide a nonlinear protocol of multi-agent systems.

Definition 3 A $(k+1)-$order $m-$dimensional hyper-matrix $\mathscr{P} = (p_{i_1 \cdots i_k j})_{i_1, \cdots, i_k, j=1}^{m, \cdots, m, m}$ is called *triply stochastic* if one has that

$$\sum_{i_{k-1}=1}^{m} p_{i_1 \cdots i_{k-1} i_k j} = \sum_{i_k=1}^{m} p_{i_1 \cdots i_{k-1} i_k j} = \sum_{j=1}^{m} p_{i_1 \cdots i_{k-1} i_k j} = 1, \ p_{i_1 \cdots i_k j} \geq 0, \ i_1, \cdots, i_k, j \in [m].$$

Protocol A. Let $\{k(n)\}_{n=1}^{\infty}$ be a sequence of natural numbers such that $k(n) \geq 2$ for all $n \in \mathbb{N}$ and $\{\mathscr{P}_n\}_{n=1}^{\infty}$, $\mathscr{P}_n = \left(p^{(n)}_{i_1 \cdots i_{k(n)} j}\right)_{i_1, \cdots, i_{k(n)}, j=1}^{m, \cdots, m, m}$ be a sequence of $(k(n)+1)-$order $m-$dimensional triply stochastic hyper-matrices. Let $\{\mathfrak{P}_n\}_{n=1}^{\infty}$, $\mathfrak{P}_n : \Delta^{m-1} \to \Delta^{m-1}$ be a sequence of polynomial stochastic operators associated with $(k(n)+1)-$order $m-$dimensional triply stochastic hyper-matrices $\{\mathscr{P}_n\}_{n=1}^{\infty}$.

Suppose that an opinion sharing dynamics of the multi-agent system is generated by non-autonomous polynomial stochastic operators as follows

$$\mathbf{x}^{(n+1)} = \mathfrak{P}_{n+1}\left(\mathbf{x}^{(n)}\right), \quad \mathbf{x}^{(0)} \in \Delta^{m-1} \tag{10}$$

where $\mathbf{x}^{(n)} = \left(x_1^{(n)}, \cdots, x_m^{(n)}\right)^T$ is the subjective distribution after n revisions.

Remark 1 Some special cases of PROTOCOL A have been considered in the previous studies: the case $k(n) = 2$ for all $n \in \mathbb{N}$ and $\{\mathscr{P}_n\}_{n=1}^{\infty} = \{\mathscr{P}\}$ in [17, 18]; the case $k(n) = 2$ for all $n \in \mathbb{N}$ with any sequence $\{\mathscr{P}_n\}_{n=1}^{\infty}$ in [20]; the case $k(n) = k \geq 2$ for all $n \in \mathbb{N}$ and $\{\mathscr{P}_n\}_{n=1}^{\infty} = \{\mathscr{P}\}$ in [19]. In this paper, we unify, extend, and generalize all previous results presented in the papers [17–20].

Definition 4 We say that the multi-agent system presented by PROTOCOL A eventually reaches to a consensus if $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty}$ converges to the center $\mathbf{c} = (\frac{1}{m}, \cdots, \frac{1}{m})^T$ of the simplex Δ^{m-1} for any $\mathbf{x}^{(0)} \in \Delta^{m-1}$.

We now introduce some notations.

We say that $\mathbf{x} \geq 0$ (resp. $\mathbf{x} > 0$) if $x_i \geq 0$ (resp. $x_i > 0$) for all $i \in [m]$. Let $\mathrm{int}\Delta^{m-1} := \{\mathbf{x} \in \Delta^{m-1} : \mathbf{x} > 0\}$ be an interior of the simplex Δ^{m-1}. Let $M(\mathbf{x}) = \max\limits_{i\in[m]} x_i$, $m(\mathbf{x}) = \min\limits_{i\in[m]} x_i$ and $d(\mathbf{x}) = M(\mathbf{x}) - m(\mathbf{x})$ for any $\mathbf{x} \in \Delta^{m-1}$. It is clear that the functions $M(\cdot), m(\cdot), d(\cdot) : \Delta^{m-1} \to \mathbb{R}$ are continuous and $d(\mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{c} = (\frac{1}{m}, \cdots, \frac{1}{m})^T$.

We need the following simple but crucial lemma.

Lemma 1 ([18–20]) *A sequence* $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty} \subset \Delta^{m-1}$ *converges to the center* $\mathbf{c} = (\frac{1}{m}, \cdots, \frac{1}{m})^T$ *of the simplex* Δ^{m-1} *if and only if* $\lim\limits_{n\to\infty} d\left(\mathbf{x}^{(n)}\right) = 0$.

Let $\mathbf{e}_1, \cdots, \mathbf{e}_m$ be vertices of the simplex Δ^{m-1} and $\mathbf{e}_i^{(n+1)} = \mathfrak{P}_{n+1}\left(\mathbf{e}_i^{(n)}\right)$ with $\mathbf{e}_i^{(0)} := \mathbf{e}_i$ for all $i \in [m]$.

Theorem 1 *Suppose that an opinion sharing dynamics of the multi-agent system is described by* PROTOCOL A. *The multi-agent system eventually reaches to a consensus if and only if for every* $i \in [m]$ *there exists* $n(i) \in \mathbb{N}$ *such that* $\mathbf{e}_i^{(n(i))} > 0$.

Proof ONLY IF PART: Suppose that the multi-agent system eventually reaches to a consensus. It particularly means that for every $i \in [m]$ the sequence $\{\mathbf{e}_i^{(n)}\}_{n=0}^{\infty}$ converges to the center $\mathbf{c} = (\frac{1}{m}, \cdots, \frac{1}{m})^T$ of the simplex Δ^{m-1}. Since $\mathbf{c} \in \mathrm{int}\Delta^{m-1}$, there exists $n(i) \in \mathbb{N}$ such that $\mathbf{e}_i^{(n(i))} > 0$ for every $i \in [m]$.

IF PART: Suppose that for every $i \in [m]$ there exists $n(i) \in \mathbb{N}$ such that $\mathbf{e}_i^{(n(i))} > 0$. We want to show that the sequence $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty}$ defined by (10) starting from any initial point $\mathbf{x}^{(0)} \in \Delta^{m-1}$ converges to the center $\mathbf{c} = (\frac{1}{m}, \cdots, \frac{1}{m})^T$ of the simplex Δ^{m-1}. In order to prove it, due to Lemma 1, it is enough to show $\lim\limits_{n\to\infty} d\left(\mathbf{x}^{(n)}\right) = 0$.

We shall accomplish it in a few steps.

Step 1: If $\mathbf{x}^{(n_0)} > 0$ for some $n_0 \in \mathbb{N}$ then $\mathbf{x}^{(n)} > 0$ for any $n > n_0$.
Indeed, due to Proposition 1, the sequence $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty}$ generates a mean process. It means that $m(\mathbf{x}^{(n_0)}) \leq m(\mathbf{x}^{(n)}) \leq M(\mathbf{x}^{(n)}) \leq M(\mathbf{x}^{(n_0)})$ for any $n > n_0$. Since $m(\mathbf{x}^{(n_0)}) > 0$, we then obtain that $m(\mathbf{x}^{(n)}) > 0$ for any $n > n_0$. Therefore, we have that $\mathbf{x}^{(n)} > 0$ for any $n > n_0$.

Step 2: One has that $\mathbf{e}_i^{(n)} > 0$ for any $n > n_0$ and $i \in [m]$ where $n_0 = \max\limits_{i\in[m]} n(i)$.
Indeed, due to **Step 1** and $\mathbf{e}_i^{(n(i))} > 0$ for every $i \in [m]$, we have that $\mathbf{e}_i^{(n)} > 0$ for any $n > n_0$ and for any $i \in [m]$ where $n_0 = \max\limits_{i\in[m]} n(i)$.

Step 3: One has that $\mathbf{x}^{(n)} > 0$ for any $n > n_0$ and $\mathbf{x}^{(0)} \in \Delta^{m-1}$ where $n_0 = \max\limits_{i\in[m]} n(i)$.
In order to show it we first prove the following inequality for any $\mathbf{x}^{(0)} \in \Delta^{m-1}$

$$\mathbf{x}^{(n)} \geq x_1^{\mathrm{K(n)}}\mathbf{e}_1^{(n)} + x_2^{\mathrm{K(n)}}\mathbf{e}_2^{(n)} + \cdots + x_m^{\mathrm{K(n)}}\mathbf{e}_m^{(n)}, \quad n \in \mathbb{N} \tag{11}$$

where $\mathrm{K(n)} = k(1) \cdot k(2) \cdot \ldots \cdot k(n)$ for $n \in \mathbb{N}$. Let us first introduce some necessary notations.
Let $\mathscr{M}_{\mathscr{P}_n} : (\mathbb{R}^m)^{\times k(n)} \to \mathbb{R}^m$ be a multi-linear operator associated with $(k(n)+1)-$order $m-$dimensional stochastic hyper-matrix $\mathscr{P}_n = \left(p^{(n)}_{i_1\cdots i_{k(n)} j}\right)_{i_1,\cdots,i_{k(n)},j=1}^{m,\cdots,m,m}$ as follows

$$\mathscr{M}_{\mathscr{P}_n}\left(\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(k(n))}\right) = \sum_{i_1=1}^{m} \cdots \sum_{i_{k(n)}=1}^{m} y^{(1)}_{i_1} y^{(2)}_{i_2} \cdots y^{(k(n))}_{i_{k(n)}} \mathbf{p}_{i_1\cdots i_{k(n)}\bullet}$$

where $\mathbf{p}_{i_1\cdots i_{k(n)}\bullet} = (p_{i_1\cdots i_{k(n)}1}, \cdots, p_{i_1\cdots i_{k(n)}m}) \in \Delta^{m-1}$ for any $i_1, \cdots, i_{k(n)} \in [m]$. It is clear that $\mathscr{P}_n(\mathbf{x}) = \mathscr{M}_{\mathscr{P}_n}(\mathbf{x}, \mathbf{x}, \cdots, \mathbf{x})$ for any $\mathbf{x} \in \Delta^{m-1}$. Moreover, if $\mathbf{x} = \lambda_1\mathbf{v}_1 + \cdots + \lambda_q\mathbf{v}_q \in \Delta^{m-1}$ with $\mathbf{v}_1, \cdots, \mathbf{v}_q \in \Delta^{m-1}$, $\lambda_1 + \cdots + \lambda_q = 1$, and $\lambda_1, \cdots, \lambda_q \geq 0$ then

$$\begin{aligned}
\mathscr{P}_n(\mathbf{x}) &= \sum_{i_1=1}^{q} \cdots \sum_{i_{k(n)}=1}^{q} \lambda_{i_1} \cdots \lambda_{i_{k(n)}} \mathscr{M}_{\mathscr{P}_n}\left(\mathbf{v}_{i_1}, \cdots, \mathbf{v}_{i_{k(n)}}\right) \\
&= \lambda_1^{k(n)} \mathscr{P}_n(\mathbf{v}_1) + \cdots + \lambda_q^{k(n)} \mathscr{P}_n(\mathbf{v}_q) + \\
&\qquad + \sum_{\substack{\text{at least for two} \\ i_\mu, i_\nu:\ i_\mu \neq i_\nu}} \lambda_{i_1} \cdots \lambda_{i_{k(n)}} \mathscr{M}_{\mathscr{P}_n}\left(\mathbf{v}_{i_1}, \cdots, \mathbf{v}_{i_{k(n)}}\right)
\end{aligned} \tag{12}$$

Hence, it follows form (12) that

$$\begin{aligned}
\mathbf{x}^{(1)} &= \mathscr{P}_1(\mathbf{x}^{(0)}) = x_1^{\mathrm{K}(1)}\mathbf{e}_1^{(1)} + x_2^{\mathrm{K}(1)}\mathbf{e}_2^{(1)} + \cdots + x_m^{\mathrm{K}(1)}\mathbf{e}_m^{(1)} + \text{remaining parts},\\
\mathbf{x}^{(2)} &= \mathscr{P}_2(\mathbf{x}^{(1)}) = x_1^{\mathrm{K}(2)}\mathbf{e}_1^{(2)} + x_2^{\mathrm{K}(2)}\mathbf{e}_2^{(2)} + \cdots + x_m^{\mathrm{K}(2)}\mathbf{e}_m^{(2)} + \text{remaining parts},\\
&\vdots\\
\mathbf{x}^{(n)} &= \mathscr{P}_n(\mathbf{x}^{(n-1)}) = x_1^{\mathrm{K(n)}}\mathbf{e}_1^{(n)} + x_2^{\mathrm{K(n)}}\mathbf{e}_2^{(n)} + \cdots + x_m^{\mathrm{K(n)}}\mathbf{e}_m^{(n)} + \text{remaining parts}.
\end{aligned}$$

Consequently, the last equality yields the inequality (11).
Moreover, it follows from the inequality (11) and $\mathbf{e}_i^{(n)} > 0$ for any $n > n_0,\ i \in [m]$ (see **Step 2**) that $\mathbf{x}^{(n)} > 0$ for any $n > n_0$ and for any $\mathbf{x}^{(0)} \in \Delta^{m-1}$.

Step 4: One has that $\lim\limits_{n\to\infty} \mathbf{x}^{(n)} = \mathbf{c}$ for any $\mathbf{x}^{(0)} \in \Delta^{m-1}$.
As we already showed that (see **Step 3**)

$$0 < m(\mathbf{x}^{(n_0)}) \leq m(\mathbf{x}^{(n)}) \leq M(\mathbf{x}^{(n)}) \leq M(\mathbf{x}^{(n_0)}) < 1 \tag{13}$$

for any $n > n_0$ and for any $\mathbf{x}^{(0)} \in \Delta^{m-1}$.
We know that

$$\mathbf{x}^{(n+1)} = \mathbb{P}_{n+1}(\mathbf{x}^{(n)})\mathbf{x}^{(n)}$$

where $\mathbb{P}_{n+1}(\mathbf{x}^{(n)}) = \left(p_{ij}^{(n+1)}(\mathbf{x}^{(n)})\right)_{i,j=1}^{m}$ with

$$p_{ij}^{(n+1)}(\mathbf{x}^{(n)}) = \sum_{i_1=1}^{m}\cdots\sum_{i_{k(n+1)-1}=1}^{m} p^{(n+1)}_{i_1\cdots i_{k(n+1)-1}ji}\, x_{i_1}^{(n)}\cdots x_{i_{k(n+1)-1}}^{(n)}.$$

Since $\sum\limits_{i_{k(n+1)-1}=1}^{m} p^{(n+1)}_{i_1\cdots i_{k(n+1)-1}ji} = 1$ for any $n \in \mathbb{N}$, it follows from (13) that

$$0 < m\left(\mathbf{x}^{(n_0)}\right) \leq p_{ij}^{(n+1)}\left(\mathbf{x}^{(n)}\right) \leq M\left(\mathbf{x}^{(n_0)}\right) < 1$$

for any $i, j \in [m]$ and for any $n > n_0$.
We then obtain from the last inequality that

$$\begin{aligned}
x_i^{(n+1)} &= \sum_{j=1}^{m} p_{ij}^{(n+1)}\left(\mathbf{x}^{(n)}\right)\left(x_j^{(n)} - M\left(\mathbf{x}^{(n)}\right)\right) + M\left(\mathbf{x}^{(n)}\right)\\
&\leq m\left(\mathbf{x}^{(n_0)}\right)\left(m\left(\mathbf{x}^{(n)}\right) - M\left(\mathbf{x}^{(n)}\right)\right) + M\left(\mathbf{x}^{(n)}\right)\\
&= \left(1 - m\left(\mathbf{x}^{(n_0)}\right)\right) M\left(\mathbf{x}^{(n)}\right) + m\left(\mathbf{x}^{(n_0)}\right) m\left(\mathbf{x}^{(n)}\right),
\end{aligned} \tag{14}$$

$$
\begin{aligned}
x_i^{(n+1)} &= \sum_{j=1}^{m} p_{ij}^{(n+1)}\left(\mathbf{x}^{(n)}\right)\left(x_j^{(n)} - m\left(\mathbf{x}^{(n)}\right)\right) + m\left(\mathbf{x}^{(n)}\right) \\
&\geq m\left(\mathbf{x}^{(n_0)}\right)\left(M\left(\mathbf{x}^{(n)}\right) - m\left(\mathbf{x}^{(n)}\right)\right) + m\left(\mathbf{x}^{(n)}\right) \\
&= m\left(\mathbf{x}^{(n_0)}\right) M\left(\mathbf{x}^{(n)}\right) + \left(1 - m\left(\mathbf{x}^{(n_0)}\right)\right) m\left(\mathbf{x}^{(n)}\right) \qquad (15)
\end{aligned}
$$

for any $i \in [m]$. Hence, we obtain from (14) and (15) that

$$
\begin{aligned}
d\left(\mathbf{x}^{(n+1)}\right) &= M\left(\mathbf{x}^{(n+1)}\right) - m\left(\mathbf{x}^{(n+1)}\right) \\
&\leq \left(1 - 2m\left(\mathbf{x}^{(n_0)}\right)\right)\left(M\left(\mathbf{x}^{(n)}\right) - m\left(\mathbf{x}^{(n)}\right)\right) \\
&= \left(1 - 2m\left(\mathbf{x}^{(n_0)}\right)\right) d\left(\mathbf{x}^{(n)}\right) \qquad (16)
\end{aligned}
$$

for any $n > n_0$. Then, it follows from (16) that

$$
d\left(\mathbf{x}^{(n)}\right) \leq \left(1 - 2m\left(\mathbf{x}^{(n_0)}\right)\right)^{n-n_0} d\left(\mathbf{x}^{(n_0)}\right).
$$

Since $m\left(\mathbf{x}^{(n_0)}\right) > 0$ and $1 - 2m\left(\mathbf{x}^{(n_0)}\right) < 1$, we get that $\lim_{n\to\infty} d\left(\mathbf{x}^{(n)}\right) = 0$. This completes the proof.

Remark 2 Let $\mathfrak{D} : \Delta^{m-1} \to \Delta^{m-1}$, $\mathfrak{D}(\mathbf{x}) = \mathbb{D}\mathbf{x}$ be a linear doubly stochastic operator associated with a doubly stochastic matrix $\mathbb{D}$. It is well known in the ergodic theory of Markov chains that a trajectory $\{\mathbf{x}^{(n)}\}_{n=0}^{\infty}$, $\mathbf{x}^{(n+1)} = \mathfrak{D}(\mathbf{x}^{(n)})$ of the linear doubly stochastic operator starting from any initial point $\mathbf{x}^{(0)} \in \Delta^{m-1}$ converges to the center $\mathbf{c} = (\frac{1}{m}, \cdots, \frac{1}{m})^T$ of the simplex Δ^{m-1} if and only if for each $i \in [m]$ there exists $n(i)$ such that $\mathbf{e}_i^{(n(i))} > 0$ where $\mathbf{e}_i^{(n+1)} = \mathfrak{D}(\mathbf{e}_i^{(n)})$ with $\mathbf{e}_i^{(0)} := \mathbf{e}_i$. In the similar spirit, Theorem 1 is an analogy of this result for the nonlinear doubly stochastic operator associated with the triply stochastic hyper-matrix.

Corollary 1 *Suppose that an opinion sharing dynamics of the multi-agent system is described by* PROTOCOL A. *If* $\mathscr{P}_{n_0} > 0$ *for some* n_0, *i.e.,* $p_{i_1\cdots i_{k(n_0)}j}^{(n_0)} > 0$ *for any* $i_1, \cdots, i_{k(n_0)}, j \in [m]$ *then the multi-agent system eventually reaches to a consensus.*

We provide an example to support our theoretical result.

Example 1 Let $m \geq 3$, $\mathbf{e} = (1, \cdots, 1)$, and $\{\mathbf{a}_n\}_{n=1}^{\infty} \subset \Delta^{m-1}$ be a sequence of stochastic vectors. We define a sequence of operators $\mathfrak{P}_n : \Delta^{m-1} \to \Delta^{m-1}$ as

$$
\mathfrak{P}_n(\mathbf{x}) = \mathbf{a}_n \sum_{i=1}^{m} x_i^3 + 3\frac{\mathbf{e} - \mathbf{a}_n}{m-1}\sum_{i<j}(x_i^2 x_j + x_i x_j^2) + 6\frac{(m-3)\mathbf{e} + 2\mathbf{a}_n}{(m-1)(m-2)}\sum_{i<j<k} x_i x_j x_k.
$$

Obviously, we have that $\mathfrak{P}_n(\mathbf{e}_i) = \mathbf{a}_n$ for any $1 \leq i \leq m$ and $\mathfrak{P}_n(\mathbf{x}) \in \mathrm{int}\Delta^{m-1}$ for any $\mathbf{x} \notin \{\mathbf{e}_1, \cdots, \mathbf{e}_m\}$. Due to Theorem 1, the consensus is established in the system if and only if there exists $n_0 \in \mathbb{N}$ such that $\mathbf{a}_{n_0} \notin \{\mathbf{e}_1, \cdots, \mathbf{e}_m\}$. Consequently, if (and only if) at least one element of the sequence $\{\mathbf{a}_n\}_{n=1}^{\infty}$ is not the vertex of the simplex Δ^{m-1} then the consensus is established in the system.

Acknowledgements This work has been done under the MOHE grant FRGS14-141-0382. The first Author (M.S.) is grateful to the Junior Associate scheme of the Abdus Salam International Centre for Theoretical Physics, Trieste, Italy. He is also indebted to Professor Hideaki Matsunaga for his kind hospitality during the conference ICDEA2016, Osaka, Japan.

References

1. Berchtold, A., Raftery, A.: The mixture transition distribution model for high-order Markov chains and non-Gaussian time series. Stat. Sci. **7**, 328–356 (2002)
2. Berger, R.L.: A necessary and sufficient condition for reaching a consensus using DeGroot's method. J. Amer. Stat. Assoc. **76**, 415–418 (1981)
3. Chatterjee, S., Seneta, E.: Towards consensus: some convergence theorems on repeated averaging. J. Appl. Prob. **14**, 89–97 (1977)
4. De Groot, M.H.: Reaching a consensus. J. Amer. Stat. Assoc. **69**, 118–121 (1974)
5. Hegselmann, R., Krause, U.: Opinion dynamics and bounded confidence: models, analysis and simulation. J. Art. Soc. Soc. Sim. **5**(3), 1–33 (2002)
6. Hegselmann, R., Krause, U.: Opinion dynamics driven by various ways of averaging. Comp. Econ. **25**, 381–405 (2005)
7. Krause, U.: A discrete nonlinear and non-autonomous model of consensus formation. In: Elaydi, S., et al. (eds.) Communications in Difference Equations, pp. 227–236. Gordon and Breach, Amsterdam (2000)
8. Krause, U.: Compromise, consensus, and the iteration of means. Elem. Math. **64**, 1–8 (2009)
9. Krause, U.: Markov chains, Gauss soups, and compromise dynamics. J. Cont. Math. Anal. **44**(2), 111–116 (2009)
10. Krause, U.: Opinion dynamics – local and global. In: Liz, E., Manosa, V. (eds.) Proceedings of the Workshop Future Directions in Difference Equations, pp. 113–119. Universidade de Vigo, Vigo (2011)
11. Krause, U.: Positive Dynamical Systems in Discrete Time: Theory, Models, and Applications. Walter de Gruyter (2015)
12. Kolokoltsov, V.: Nonlinear Markov Processes and Kinetic Equations. Cambridge University Press (2010)
13. Lyubich, Y.I.: Mathematical Structures in Population Genetics. Springer (1992)
14. Lu, J., Yu, X., Chen, G., Yu, W.: Complex Systems and Networks: Dynamics. Springer, Controls and Applications (2016)
15. Raftery, A.: A model of high-order Markov chains. J. Roy. Stat. Soc. **47**, 528–539 (1985)
16. Saburov, M.: Ergodicity of nonlinear Markov operators on the finite dimensional space. Non. Anal. Theo. Met. Appl. **143**, 105–119 (2016)
17. Saburov, M., Saburov, Kh: Reaching a consensus in multi-agent systems: A time invariant nonlinear rule. J. Educ. Vocat. Res. **4**(5), 130–133 (2013)
18. Saburov, M., Saburov, Kh: Mathematical models of nonlinear uniform consensus. Sci. Asia **40**(4), 306–312 (2014)
19. Saburov, M., Saburov, Kh: Reaching a nonlinear consensus: polynomial stochastic operators. Inter. J. Cont. Auto. Sys. **12**(6), 1276–1282 (2014)
20. Saburov, M., Saburov, Kh: Reaching a nonlinear consensus: a discrete nonlinear time-varying case. Inter. J. Sys. Sci. **47**(10), 2449–2457 (2016)
21. Saburov, M., Yusof, N.A.: Counterexamples to the conjecture on stationary probability vectors of the second-order Markov chains. Linear Algebra Appl. **507**, 153–157 (2016)
22. Seneta, E.: Nonnegative Matrices and Markov Chains. Springer-Verlag (1981)

On the Stability of an SIR Epidemic Discrete Model

Kaori Saito

Abstract A mathematical epidemic discrete equation, which appears as a model for the spread of disease-causing, is treated. In this paper, we consider the asymptotic stability of a discrete SIR epidemic model by using the classical linearization method and some Liapunov functions.

Keywords SIR epidemic discrete model · Positive equilibrium points · Asymptotic stability

1 Introduction and Motivation

Over the last decade a great deal of articles have been devoted to the study of the dynamics of discrete epidemic models; see, e.g., [1, 2, 4, 7, 10, 11] and references therein. In this paper, imitating a discrete SIS epidemic model proposed in Jang and Elaydi's parer [7] by use of the nonstandard discretization technique of Mickens [8], we shall consider the following discrete SIR epidemic model

$$\begin{aligned} S_{n+1} - S_n &= b - \beta S_{n+1} I_n - \mu_1 S_{n+1}, \\ I_{n+1} - I_n &= \beta S_n I_n - (\mu_2 + \lambda) I_{n+1}, \\ R_{n+1} - R_n &= \lambda I_n - \mu_3 R_{n+1}, \quad n \geq 0. \end{aligned} \tag{1}$$

The initial condition of (1) is given by

$$S_0 \geq 0, \quad I_0 \geq 0 \quad \text{and} \quad R_0 \geq 0. \tag{2}$$

For equation (1), S_n denotes the number of the population susceptible to the disease, I_n denotes the number of infectious individual and R_n denotes the number who has been removed from the possibility of infection through full immunity. It is assumed

K. Saito (✉)
Iwate Prefectural University, Miyako College, 1-5-1 Kanan, Miyako, Iwate 027-0039, Japan
e-mail: saito_k@iwate-pu.ac.jp

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_15

that all newborns are susceptible. μ_1, μ_2 and μ_3 are positive constants. μ_1 is the death rate of the susceptible, μ_2 is the death rate of the infective and μ_3 is the death rate of the recovered. It is biologically natural to assume that

$$\mu_1 \leq \min\{\mu_2, \mu_3\}. \tag{3}$$

In addition, the positive constants b and λ represent the birth and death rates of the population and recovery rate of infectives, respectively. The positive constant β is the average number of contacts per infective per day. We can show the existence of a unique positive solution (S_n, I_n, R_n) of equation (1) with the initial condition (2).

In 1979, for ordinary differential equations, Anderson and May [2] have studied the asymptotic stability of the following epidemic differential equation

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta S(s)I(t) - \mu S(t) + \mu, \\ \frac{dI(t)}{dt} &= \beta S(t)I(t) - \mu I(t) - \lambda I(t), \\ \frac{dR(t)}{dt} &= \lambda I(t) - \mu R(t), \quad t \geq 0, \end{aligned} \tag{4}$$

where β, μ and λ are positive constants, and $S(t) + I(t) + R(t) = N(t)$ denotes the total number of a population at the time t. In [1], it is assumed that $N(t)$ is a constant, that is $N(t) = 1$ for all $t \geq 0$, and that the birth and death rates of population are the same values. Recently, Hamaya and Saito [5] have studied the property of permanence of the solution $(S(t, x), I(t, x), R(t, x))$ of partial differential equations with diffusion.

The purpose of this paper is to investigate the property of permanence and global asymptotic stability of solutions of equation (1). Notice that for our equation (1), the total number $N_n(= S_n + I_n + R_n)$ of population at discrete time n is not a solution of linear equation (inequality) compare with the equation treated in [4], which is a solution of linear equation (inequality). This point is a motivation of interest in the area.

2 Preliminary and Local Stability of Equilibrium Points

If I_n is known, then R_n can be obtained by equation (1). Therefore, we can rewrite to replace equation (1) with the following equation

$$\begin{aligned} S_{n+1} - S_n &= b - \beta S_{n+1} I_n - \mu_1 S_{n+1}, \\ I_{n+1} - I_n &= \beta S_n I_n - (\mu_2 + \lambda) I_{n+1}, \quad n \geq 0. \end{aligned} \tag{5}$$

For any parameters β, b, λ and μ_i $(i = 1, 2, 3)$, it is easy to check that the equilibrium solution of (1) with the initial condition (2) exists as follows:

(i) If $b > 0$, then equation (1) always has a disease free equilibrium $E_{S_0^*} = (S_0^*, 0, 0)$, where

$$S_0^* = \frac{b}{\mu_1}.$$

(ii) Furthermore, if

$$S_0^* > S^* \equiv \frac{\mu_2 + \lambda}{\beta}, \tag{6}$$

then equation (1) also has a unique positive endemic equilibrium $E^+ = (S^*, I^*, R^*)$, where

$$S^* = \frac{\mu_2 + \lambda}{\beta}, \quad I^* = \frac{b\beta - \mu_1(\mu_2 + \lambda)}{\beta(\mu_2 + \lambda)}, \quad R^* = \frac{\lambda}{\mu_3} I^*.$$

We discuss the behavior of solutions of equation (5).

Definition 1 Equation (1) is said to be permanent if there are positive constants ν_i and M_i $(i = 1, 2, 3)$ such that

$$\begin{aligned}
\nu_1 \le \liminf_{n\to\infty} S_n \le \limsup_{n\to\infty} S_n \le M_1, \\
\nu_2 \le \liminf_{n\to\infty} I_n \le \limsup_{n\to\infty} I_n \le M_2, \\
\nu_3 \le \liminf_{n\to\infty} R_n \le \limsup_{n\to\infty} R_n \le M_3
\end{aligned}$$

hold for any solution of (1) with the initial condition (2). Here ν_i and M_i $(i = 1, 2, 3)$ are independent of (2).

Now, we have the following theorems.

Theorem 1 *If $S_0^* < S^*$, then the disease free equilibrium $E_{S_0^*}$ of (1) is locally asymptotically stable. And if $S_0^* > S^*$, then $E_{S_0^*}$ is unstable.*

Proof It is sufficient to show the statement of Theorem 1 for equation (5). From (5), we obtain

$$S_{n+1} = \frac{S_n + b}{1 + \beta I_n + \mu_1}, \quad I_{n+1} = \frac{(1 + \beta S_n) I_n}{1 + \mu_2 + \lambda}.$$

For variables S and I, we can calculate the Jacobian matrix of

$$J(S, I) = \begin{pmatrix} \frac{1}{1+\beta I+\mu_1} & -\frac{(b+S)\beta}{(1+\beta I+\mu_1)^2} \\ \frac{\beta I}{1+\mu_2+\lambda} & \frac{\beta S+1}{1+\mu_2+\lambda} \end{pmatrix}.$$

In the case of the disease free equilibrium $E_{S_0^*} = (S_0^*, 0)$ of (5), Jacobian matrix is given by

$$J(E_0^*) = \begin{pmatrix} \frac{1}{1+\mu_1} & -\frac{(b+S_0^*)\beta}{(1+\mu_1)^2} \\ 0 & \frac{\beta S_0^*+1}{1+\mu_2+\lambda} \end{pmatrix}. \tag{7}$$

Since $J(E_0^*)$ is an upper triangular matrix, eigenvalues are diagonal elements itself of

$$\xi_1 = \frac{1}{1+\mu_1} \quad \text{and} \quad \xi_2 = \frac{\beta S_0^* + 1}{1+\mu_2+\lambda}.$$

If $S_0^* < S^*$, we have $0 < \xi_1 < 1$ and $0 < \xi_2 < 1$, and thus E_0^* is locally asymptotically stable. And if $S_0^* > S^*$, we have $\xi_2 > 1$, and hence $E_{S_0^*}$ is unstable. For equation (1), we obtain

$$R_{n+1} = \frac{R_n + \lambda I_n}{1+\mu_3}.$$

Jacobian matrix is similar given by

$$J(E_0^* = (S_0^*, 0, 0)) = \begin{pmatrix} \frac{1}{1+\mu_1} & -\frac{(b+S_0^*)\beta}{(1+\mu_1)^2} & 0 \\ 0 & \frac{\beta S_0^*+1}{1+\mu_2+\lambda} & 0 \\ 0 & \frac{\lambda}{1+\mu_3} & \frac{1}{1+\mu_3} \end{pmatrix}.$$

Thus, the third eigenvalue is $\xi_3 = \frac{1}{1+\mu_3}$ with $0 < \xi_3 < 1$. Therefore, $E_0^* = (S_0^*, 0, 0)$ of (1) is locally asymptotically stable if $S_0^* < S^*$. This proof is completed. □

Theorem 2 *If $S_0^* > S^*$, then the endemic equilibrium E^+ of 1 is locally asymptotically stable.*

Proof As the same reason in the proof of Theorem 1, we consider the endemic equilibrium $E^+ = (S^*, I^*)$ of (5), where

$$S^* = \frac{\mu_2+\lambda}{\beta} \quad \text{and} \quad I^* = \frac{\beta b - \mu_1(\mu_2+\lambda)}{\beta(\mu_2+\lambda)}. \tag{8}$$

From (7) and (8), Jacobian matrix is

$$J(E^+) = \begin{pmatrix} \frac{1}{1+\beta I^*+\mu_1} & -\frac{(b+S^*)\beta}{(1+\beta I^*+\mu_1)^2} \\ \frac{\beta I^*}{1+\mu_2+\lambda} & \frac{\beta S^*+1}{1+\mu_2+\lambda} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\mu_2+\lambda}{\beta b+\mu_2+\lambda} & -\frac{(\mu_2+\lambda)^2}{(\beta b+\mu_2+\lambda)} \\ \frac{\beta b-\mu_1(\mu_2+\lambda)}{(\mu_2+\lambda)(1+\mu_2+\lambda)} & 1 \end{pmatrix}.$$

The equilibrium point E^+ is local asymptotically stable if the condition

$$|\,\mathrm{tr}\, J(E^+)| < 1 + \det J(E^+) < 2 \tag{9}$$

yields (cf. p.188 in [3]). To do this, we first show that $|\,\mathrm{tr}\, J(E^+)| < 1 + \det J(E^+)$. By $S_0^* > S^*$ and $\beta b - \mu_1(\mu_2 + \lambda) > 0$, we have

$$\begin{aligned} |\,\mathrm{tr}\, J(E^+)| &= \frac{\mu_2+\lambda}{\beta b+\mu_2+\lambda} + 1 \\ &< 1 + \frac{\mu_2+\lambda}{\beta b+\mu_2+\lambda} + \frac{\beta b-\mu_1(\mu_2+\lambda)}{(\mu_2+\lambda)(1+\mu_2+\lambda)} \times \frac{(\mu_2+\lambda)^2}{\beta b+\mu_2+\lambda} \\ &= 1 + \det J(E^+). \end{aligned}$$

Next, we can show that $\det J(E^+) < 1$. We set

$$\mathscr{R}_0 = \frac{\beta b}{\mu_1(\mu_2+\lambda)}. \tag{10}$$

Then $\mathscr{R}_0 > 1$ by the assumption $S_0^* > S^*$. It is clear that

$$1 - \frac{\mu_1(\mu_2+\lambda)}{\beta b} < 1 + \frac{1}{\mu_2+\lambda}$$

and we obtain

$$0 < (\mu_2+\lambda)(\beta b - \mu_1(\mu_2+\lambda)) < \beta b(1+\mu_2+\lambda).$$

Thus, we have

$$\begin{aligned} \det J(E^+) &= \frac{(\mu_2+\lambda)(1+\mu_2+\lambda) + (\mu_2+\lambda)(\beta b - \mu_1(\mu_2+\lambda))}{(\beta b+\mu_2+\lambda)(1+\mu_2+\lambda)} \\ &< \frac{(\mu_2+\lambda)(1+\mu_2+\lambda) + \beta b(1+\mu_2+\lambda)}{(\beta b+\mu_2+\lambda)(1+\mu_2+\lambda)} = 1. \end{aligned}$$

Therefore, we have (9). For equation (1), we obtain the characteristic equation of

$$|J(E^+ = (S^*, I^*, R^*)) - tE_{3\times 3}| = \begin{vmatrix} \frac{\mu_2+\lambda}{\beta b+\mu_2+\lambda} - t & -\frac{(\mu_2+\lambda)^2}{(\beta b+\mu_2+\lambda)} & 0 \\ \frac{\beta b-\mu_1(\mu_2+\lambda)}{(\mu_2+\lambda)(1+\mu_2+\lambda)} & 1-t & 0 \\ 0 & \frac{\lambda}{1+\mu_3} & \frac{1}{1+\mu_3} - t \end{vmatrix}$$

$$= |J(E^+ = (S^*, I^*)) - tE_{2\times 2}| \times \left(\frac{1}{1+\mu_3} - t\right) = 0,$$

where $E_{j\times j}$ is the $j \times j (j = 2, 3)$ unit matrix. Thus, we have solutions t_1 and t_2 of $|J(E^+ = (S^*, I^*)) - tI_{2\times 2}| = 0$, and $t_3 = \frac{1}{1+\mu_3}$. Then, by (9) and Schur-Cohn criterion (cf. [3]), we have $|t_1| < 1$, $|t_2| < 1$ and $|t_3| < 1$. Therefore, the endemic equilibrium $E^+ = (S^*, I^*, R^*)$ of (1) is locally asymptotically stable if $S_0^* > S^*$. This completes the proof of Theorem 2. □

Remark 1 By using [7, Theorem 2.4], we can also prove that if $S_0^* > S^*$, then the endemic equilibrium $E^+ = (S^*, I^*)$ of (5) is locally asymptotically stable.

Remark 2 It is known that $\mathscr{R}_0$ given by (10) is the basic production number of equation (1). Notice that $\mathscr{R}_0 < 1$ is equivalent to $S_0^* < S^*$ in Theorem 1 and $\mathscr{R}_0 > 1$ is equivalent to $S_0^* > S^*$ in Theorem 2; see [6, 9].

3 Global Attractor and Permanence

In this section, we discuss the global attractivity of equilibrium points of (1) by using some Liapunov functions. However, in this time, we will omit the detail of proofs of the following Theorems 3 and 4 (cf. [5, 11]). Especially, Theorem 3 is proved by modifying the proof of Theorem 2.3 in [7]. In the proof of Theorem 4, we define the following Liapunov function of equation (5);

$$\begin{aligned} V_n &= V_n(S_n, I_n) \\ &= S_n - S^* + S^* \log \frac{S_n}{S^*} + I_n - I^* + I^* \log \frac{I_n}{I^*}. \end{aligned} \tag{11}$$

Then, it is clear that $V_n > 0$ for $(S_n, I_n) \neq (S^*, I^*)$ and $V_n = 0$ for $(S_n, I_n) = (S^*, I^*)$. Moreover, we have $\Delta V_n (= V_{n+1} - V_n) \leq 0$ for sufficient large $n > 0$. Thus, using LaSalle's invariance principle, we obtain (S_n, I_n) tends to (S^*, I^*) as $n \to \infty$.

Theorem 3 *If $S_0^* < S^*$, then every solution (S_n, I_n, R_n) of equation (1) with (2) satisfies*

$$\lim_{n\to\infty} I_n = 0, \quad \lim_{n\to\infty} R_n = 0 \quad \textit{and} \quad \lim_{n\to\infty} S_n = \frac{b}{\mu_1}.$$

Theorem 4 *If $S_0^* > S^*$ and $I_0 \neq 0$, then every solution (S_n, I_n, R_n) of equation (1) with (2) satisfies*

$$\lim_{n\to\infty} S_n = S^*, \quad \lim_{n\to\infty} I_n = I^* \quad \textit{and} \quad \lim_{n\to\infty} R_n = R^*.$$

Remark 3 Theorems 1 and 3 show that if $S_0^* < S^*$, then the disease free equilibrium $E_{S_0^*}$ of (1) is globally asymptotically stable. Also, Theorems 2 and 4 show that if $S_0^* > S^*$ and $I_0 \neq 0$, then the unique endemic equilibrium E^+ of (1) is globally asymptotically stable.

In [11], we have proved that equation (1) has the property of permanence Fig 1.

Theorem 5 *If $S_0^* > S^*$, then equation (1) is permanent.*

Finally, we give an example of Theorems 2 and 4.

Example 1 For simplicity, we demonstrate equation (5), where $\beta = 0.02$, $\mu_1 = 0.15$, $b = 5.0$, and $\mu_2 + \lambda = 0.3$. Then equation (5) becomes

$$\begin{aligned} S_{n+1} - S_n &= 5.0 - 0.02 S_{n+1} I_n - 0.1 S_{n+1}, \\ I_{n+1} - I_n &= 0.02 S_n I_n - 0.3 I_{n+1}, \quad n \geq 0, \end{aligned} \tag{12}$$

where

$$S_0^* = \frac{b}{\mu_1} = \frac{5.0}{0.15} = 33.\dot{3}, \quad S^* = \frac{\mu_2 + \lambda}{\beta} = \frac{0.3}{0.02} = 15.0, \quad \text{then} \quad S_0^* > S^*,$$

$$E^+ = (S^*, I^*) = (15.0, 9.1\dot{6}), \quad \text{and} \quad E_{S_0^*} = (S_0^*, 0) = (33.\dot{3}, 0),$$

$$I^* = \frac{b\beta - \mu_1(\mu_2 + \lambda)}{\beta(\mu_2 + \lambda)} = \frac{5.0 \times 0.02 - 0.15 \times 0.3}{0.02 \times 0.3} = 9.1\dot{6} > 0.$$

The initial condition is

$$S_0 = 150.0 > 0, \;\; I_0 = 3.0 > 0.$$

The following figures illustrate our theorem and suggest that the endemic equilibrium E^+ of equation (1) is globally asymptotically stable if the assumptions in Theorem 4 hold. In the figures, the horizontal axis shows n of a discrete time, and the

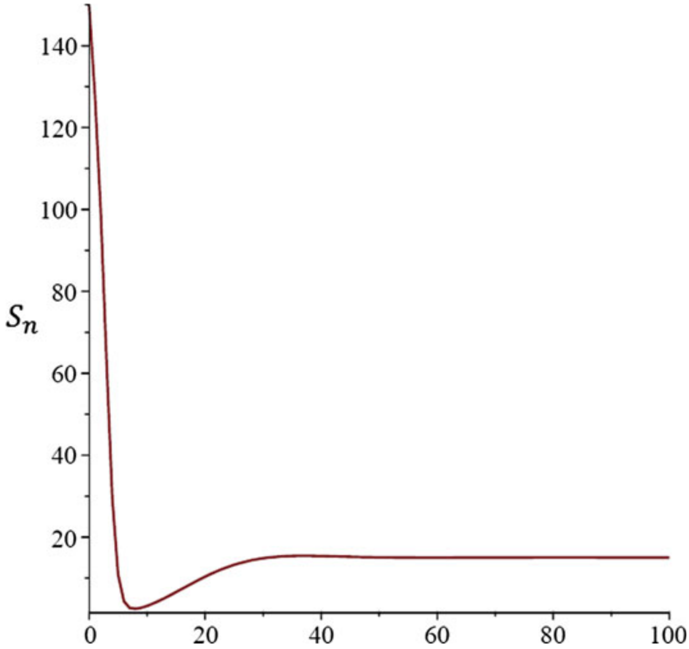


Fig. 1 The behavior of S_n with $S_0 = 150.0$

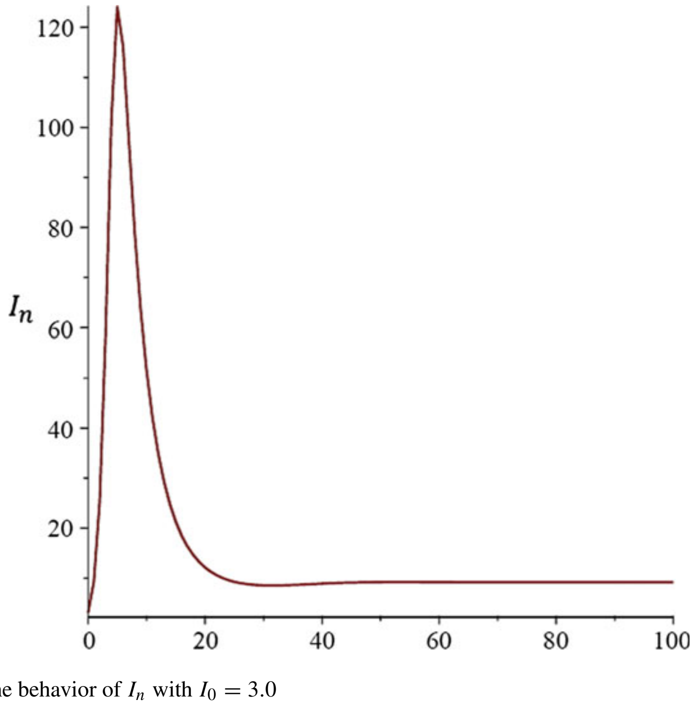


Fig. 2 The behavior of I_n with $I_0 = 3.0$

vertical axis shows the behavior of S_n or I_n, in the graph of trajectory of equation (12). That is, if $S_0^* > S^*$, each of S_n and I_n is closer to S^* and I^* respectively as $n \to \infty$ Fig. 2.

Acknowledgements I would like to thank Professor Saber Elaydi for useful comments in this paper. Moreover, I appreciate Professor Jim M. Cushing for useful advices. I am also grateful to Professor Toshiyuki Kohno for assistance with the numerical simulations. Finally, I am grateful to the referees for useful suggestions. The research of this paper is partially supported by the JSPS KAKENHI Grant Number 26400181.

References

1. Allen, L.J.: Some discrete-time SI, SIR, and SIS epidemic models. Math. Biosci. **124**, 83–105 (1994)
2. Anderson, R.M., May, R.M.: Population biology of infectious diseases. Part1. Nature **280**, 361–367 (1979)
3. Elaydi, S.: An Introduction to Difference Equations, Third edn. Springer, Berlin (2005)
4. Enatsu, Y., Nakata, Y., Muroya, Y.: Global stability for a class of discrete SIR epidemic models. Math. Biosci. Eng. **7**, 347–361 (2010)
5. Hamaya, Y., Saito, K.: Global asymptotic stability of a delayed SIR epidemic model with diffusion. Libertas Math. **36**(1), 53–72 (2016)
6. Inaba, H.: Mathematical Models for Demography and Epidemics, University of Tokyo Press (2002)
7. Jang, S., Elaydi, S.: Difference equations from discretization of a continuous epidemic model with immigration of infectives. Canad. Appl. Math. Quart. **11**(1), 93–105 (2003)
8. Mickens, R.: Nonstandard Finite Difference Methods of Differential Equations. World Scientific, Singapore (1994)
9. Murray, J.D.: Mathematical Biology, Third edn. Springer (2002)
10. Roeger, L.-I.W.: Dynamically consistent discrete-time SI and SIS epidemic models. Discret. Contin. Dyn. Syst. Suppl. 653–662 (2013)
11. Saito, K.: On the stability of SIR epidemic discrete models. To be submitted

Nonoscillation of Second-Order Linear Equations Involving a Generalized Difference Operator

Jitsuro Sugie and Masahiko Tanaka

Abstract Our purpose here is to establish nonoscillation criteria for the second-order linear difference equations of the form

$$\Delta_a(r_{n-1}\Delta_a x_{n-1}) + p_n x_n = 0.$$

Here, $\{r_n\}$ and $\{p_n\}$ are sequences of real numbers and Δ_a is the weighted difference operator defined by $\Delta_a x_n = x_{n+1} - a x_n$ with any positive constant a. A certain sequence determined from the constant a and two sequences $\{r_n\}$ and $\{p_n\}$ plays an important role in the results obtained. To be a little more precise, what should be paid attention to is a weighted sum of two adjacent terms of the sequence. The main tools for the proof of our results are Sturm's separation theorem and the Riccati transformation method. Our results are compared with several previous works by using some specific examples.

Keywords Linear difference equations · Nonoscillation · Riccati transformation · Sturm's separation theorem

1 Introduction

We consider the self-adjoint difference equation

$$\Delta_a(r_{n-1}\Delta_a x_{n-1}) + p_n x_n = 0, \qquad n = 1, 2, \ldots, \tag{1.1}$$

J. Sugie (✉) · M. Tanaka
Department of Mathematics, Shimane University, Matsue 690-8504, Japan
e-mail: jsugie@riko.shimane-u.ac.jp

M. Tanaka
e-mail: qut4527@yahoo.co.jp

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_16

where $\{r_n\}$ and $\{p_n\}$ are sequences satisfying

$$r_n > 0 \quad \text{for} \;\; n \in \mathbb{N} \cup \{0\} \tag{1.2}$$

and

$$p_n < a(r_n + r_{n-1}) \quad \text{for} \;\; n \in \mathbb{N}, \tag{1.3}$$

and Δ_a is the forward weighted difference operator defined by

$$\Delta_a x_n = x_{n+1} - a x_n$$

with $a > 0$. It is not necessary to assume that p_n is positive for $n \in \mathbb{N}$.

The null sequence $\{0\}$ is a solution of (1.1). This solution is called a trivial solution. Nontrivial solutions of (1.1) are classified into two groups by asymptotic behavior. Those belonging to one group are called *oscillatory* solutions and those belonging to the other group are called *nonoscillatory* solutions. An oscillatory solution $\{x_n\}$ has a subsequence $\{x_{n_i}\}$ such that $n_i \in \mathbb{N}$ tends to ∞ as $i \to \infty$ and $x_{n_i} x_{n_{i+1}} \leq 0$ for all $i \in \mathbb{N}$. A nonoscillatory solution $\{x_n\}$ has an $N \in \mathbb{N}$ such that $x_n > 0$ for $n \geq N$ or $x_n < 0$ for $n \geq N$. If $\{x_n\}$ is a solution of (1.1), then $\{-x_n\}$ is also a solution of (1.1). Hence, we may assume without loss of generality that a nonoscillatory solution of (1.1) are eventually positive.

The purpose of this paper is to give sufficient conditions which guarantee that all nontrivial solutions of (1.1) are nonoscillatory. Our conditions will be expressed with the relation between the positive constant a and two sequences $\{r_n\}$ and $\{p_n\}$.

Suppose that there is a subsequence $\{n_k\}$ of $\mathbb{N}$ tending to ∞ as $k \to \infty$ such that

$$p_{n_k} \geq a(r_{n_k} + r_{n_k - 1}).$$

If all nontrivial solutions of (1.1) are nonoscillatory, then by (1.2), we have

$$\begin{aligned} &\Delta_a(r_{n_k-1} \Delta_a x_{n_k-1}) + p_{n_k} x_{n_k} \\ &\quad = r_{n_k} x_{n_k+1} + \bigl\{p_{n_k} - a(r_{n_k} + r_{n_k-1})\bigr\} x_{n_k} + a^2 r_{n_k-1} x_{n_k-1} > 0 \end{aligned}$$

for all sufficiently large k. This is a contradiction. Hence, the inequality (1.3) is a necessary condition for all nontrivial solutions of (1.1) to be nonoscillatory, and therefore, it is natural to assume the inequality (1.3).

Several articles have reported oscillation and nonoscillation of solutions of difference equations which are expressed by using the forward weighted difference operator Δ_a and its generalized forms (for example, see [10, 15, 16, 18–20, 22]). This operator is a simple generalization of the usually forward difference operator Δ. Many studies have been made on oscillation problem of difference equations using the operator Δ. In those researches, we often notice a similarity between results of

$$\Delta(r_{n-1} \Delta x_{n-1}) + p_n x_n = 0 \tag{1.4}$$

and its continuous counter part

$$\big(r(t)x'\big)' + p(t)x = 0, \tag{1.5}$$

where $r, p\colon (0,\infty) \to \mathbb{R}$ are continuous functions, $r(t) > 0$ for $t > 0$. For example, Hooker [7] considered Eq. (1.4) and gave a discrete analogue of the well-known Hille-Wintner comparison theorem for second-order linear differential equations. About the classical Hille-Wintner comparison theorem, see [5, 21, 23].

A critical value that divides oscillation and nonoscillation of solutions of ordinary differential equations such as (1.5) is called an *oscillation constant* (refer to [5, 12, 17, 21]). The oscillation constant often becomes 1/4 for linear differential equations. For example, as known well, all nontrivial solutions of the Euler differential equation

$$x'' + \frac{\gamma}{t^2}x = 0$$

are nonoscillatory if and only if $\gamma \le 1/4$. Analogues of this result for Eq. (1.4) were shown in a series of papers of Hooker et al. [7, 9, 13] (see also the books [2, Chap. 6], [4, Chap. 7], [11, Chap. 6]). Their results can be easily extended to those that are applicable to Eq. (1.1). To present those results, we define

$$q_n = \frac{r_n^2}{\big\{a(r_n + r_{n-1}) - p_n\big\}\big\{a(r_{n+1} + r_n) - p_{n+1}\big\}} \tag{1.6}$$

for $n \in \mathbb{N}$. From assumptions (1.2) and (1.3), we see that $\{q_n\}$ is a positive sequence.

Theorem A *If $a^2 q_n \ge 1/(4-\varepsilon)$ for some $\varepsilon > 0$ and for all sufficiently large n, then all nontrivial solutions of* (1.1) *are oscillatory.*

Theorem B *If $a^2 q_n \le 1/4$ for all sufficiently large n, then all nontrivial solutions of* (1.1) *are nonoscillatory.*

Theorem C *If $a^2 q_{n_k} \ge 1$ for a sequence $\{n_k\}$ tending to ∞, then all nontrivial solutions of* (1.1) *are oscillatory.*

Theorems A and B are called "oscillation theorem" and "nonoscillation theorem", respectively. From these results, we see that the oscillation constant is $1/4$. In such a sense, Theorems A and B have a good balance. However, to apply Theorem B (respectively, Theorem A), the amount $a^2 q_n$ must be less than or equals to (respectively, greater than) $1/4$ for all sufficiently large n. These restrictions seem to be too strong.

In this paper, we pay our attention to a weighted sum of two adjacent terms of the sequence $\{q_n\}$. If the weighted sum is not greater than 1, then our result can be applied even if there is a subsequence $\{q_{n_k}\} \subset \{q_n\}$ such that n_k tends to ∞ as $k \to \infty$ and $a^2 q_{n_k}$ is greater than $1/4$ for $k \in \mathbb{N}$.

Theorem 1 *Assume* (1.2) *and* (1.3). *Suppose that there exists a sequence* $\{\alpha_k\}$ *with* $\alpha_k > 1$ *and either*

$$\frac{a^2\alpha_k}{\alpha_k - 1} q_{2k-1} + a^2\alpha_{k+1} q_{2k} \le 1 \tag{1.7}$$

or

$$\frac{a^2\alpha_k}{\alpha_k - 1} q_{2k} + a^2\alpha_{k+1} q_{2k+1} \le 1 \tag{1.8}$$

for all sufficiently large $k \in \mathbb{N}$. *Then all nontrivial solutions of* (1.1) *are nonoscillatory.*

2 Transformation to the Riccati Difference Equation

Sturm's separation theorem and the Riccati transformation method are very famous and useful in oscillation theory. For Sturm's separation theorem on difference equations, see [4, pp. 321–322] for example. Sturm's separation theorem ensures that oscillatory solutions and nonoscillatory solutions do not coexist in the same linear difference equation. Hence, if we find a nonoscillatory solution of (1.1), then all nontrivial solutions of (1.1) are nonoscillatory.

Let $\{x_n\}$ be a nonoscillatory solution of (1.1). Then, by (1.2), we can define

$$z_n = \frac{a(r_{n+1} + r_n) - p_{n+1}}{r_n} \frac{x_{n+1}}{x_n}$$

for all sufficiently large $n \in \mathbb{N}$. This conversion from $\{x_n\}$ to $\{z_n\}$ is called a *Riccati-type transformation*. The sequence $\{z_n\}$ satisfies the first-order non-linear difference equation

$$q_n z_n + \frac{a^2}{z_{n-1}} = 1 \tag{2.1}$$

for all sufficiently large n, where $\{q_n\}$ is the sequence defined in (1.6). Conversely, if Eq. (2.1) has a solution that is eventually positive, then there is a nonoscillatory solution of (1.1). Hence, from Sturm's separation theorem, we see that all non-trivial solutions of (1.1) are nonoscillatory. We therefore need only to find a positive solution of (2.1) in order to prove Theorem 1.

Proof of Theorem 1 Consider only the case that (1.7) holds, because the proof of the case that (1.8) holds is the same as that of the case that (1.7) holds.

We can find an $N \in \mathbb{N}$ so that $\alpha_k > 1$ and

$$a^2\alpha_{k+1} \le \frac{1}{q_{2k}}\left(1 - \frac{a^2\alpha_k}{\alpha_k - 1} q_{2k-1}\right)$$

for all $k \geq N$. Let $\{z_n\}$ be a solution of (2.1) satisfying $z_{2N-2} \geq a^2\alpha_N > a^2$. Then we obtain

$$z_{2N-1} = \frac{1}{q_{2N-1}}\left(1 - \frac{a^2}{z_{2N-2}}\right) \geq \frac{1}{q_{2N-1}}\left(1 - \frac{a^2}{a^2\alpha_N}\right) = \frac{\alpha_N - 1}{\alpha_N q_{2N-1}} > 0.$$

Hence, by (1.7) we have

$$z_{2N} = \frac{1}{q_{2N}}\left(1 - \frac{a^2}{z_{2N-1}}\right) \geq \frac{1}{q_{2N}}\left(1 - \frac{a^2\alpha_N}{\alpha_N - 1}\,q_{2N-1}\right) \geq a^2\alpha_{N+1} > a^2.$$

Similarly, we can estimate that

$$z_n \geq \begin{cases} \dfrac{\alpha_k - 1}{\alpha_k q_{2k-1}} & \text{if } \ n = 2k-1, \\ a^2\alpha_{k+1} & \text{if } \ n = 2k \end{cases}$$

for all $k \geq N$. Hence, the sequence $\{z_n\}$ is an eventually positive solution of (2.1). We therefore conclude that all nontrivial solutions of (1.1) are nonoscillatory. □

We can choose a constant sequence that is greater than 1 as the sequence $\{\alpha_k\}$ in Theorem 1. If $\alpha_k \equiv 2$, then $\alpha_k/(\alpha_k - 1) \equiv 2$. Hence, we have the following corollary of Theorem 1.

Corollary 1 *Assume* (1.2) *and* (1.3). *Suppose that either*

$$a^2(q_{2k-1} + q_{2k}) \leq \frac{1}{2} \tag{2.2}$$

or

$$a^2(q_{2k} + q_{2k+1}) \leq \frac{1}{2} \tag{2.3}$$

holds for all sufficiently large $k \in \mathbb{N}$. *Then all nontrivial solutions of* (1.1) *are nonoscillatory.*

Remark 1 If $q_n \leq 1/4a^2$ for all sufficiently large n, both inequalities (2.2) and (2.3) are naturally satisfied. Hence, Corollary 1 fully includes Theorem B.

Remark 2 Since the arithmetic mean of two positive numbers is not less than their geometric mean, condition (2.2) implies that

$$a^2\sqrt{q_{2k-1}q_{2k}} \leq \frac{1}{4} \tag{2.4}$$

is satisfied for all sufficiently large $k \in \mathbb{N}$. Note that we cannot weaken (2.2) to (2.4) in Corollary 1. For example, consider Eq. (1.1) with

$$ar_n = \begin{cases} 1/4 & \text{if } n = 2k-1, \\ 1 & \text{if } n = 2k \end{cases}$$

for $k \in \mathbb{N}$ and $p_n \equiv 1/4$. Then we have

$$a^2 q_n = \frac{a^2 r_n^2}{\{a(r_n + r_{n-1}) - p_n\}\{a(r_{n+1} + r_n) - p_{n+1}\}} = \begin{cases} 1/16 & \text{if } n = 2k-1, \\ 1 & \text{if } n = 2k. \end{cases}$$

Hence, the inequality (2.4) is satisfied. However, since $\{q_n\}$ has a subsequence that is identically $1/a^2$, Theorem C concludes that all nontrivial solutions are oscillatory.

In Theorem 1, the sequence $\{\alpha_k\}$ or its subsequence does not necessarily have to be constant. If $q_{2k-1} < 1/a^2$ for all sufficiently large $k \in \mathbb{N}$, then we can choose

$$\alpha_k = \frac{1}{1 - a\sqrt{q_{2k-1}}} > 1.$$

Since $\alpha_k/(\alpha_k - 1) = 1/(a\sqrt{q_{2k-1}})$ and $\alpha_{k+1} = 1/(1 - a\sqrt{q_{2k+1}})$, condition (1.7) becomes

$$a\sqrt{q_{2k-1}} + \frac{a^2 q_{2k}}{1 - a\sqrt{q_{2k+1}}} \le 1.$$

Similarly, we can check that condition (1.8) coincides with

$$a\sqrt{q_{2k}} + \frac{a^2 q_{2k+1}}{1 - a\sqrt{q_{2k+2}}} \le 1$$

by setting $1/(1 - a\sqrt{q_{2k}})$ on α_k if $q_{2k} < 1/a^2$ for all sufficiently large $k \in \mathbb{N}$. We therefore get the following corollary of Theorem 1.

Corollary 2 *Assume* (1.2) *and* (1.3). *Suppose that either*

$$a^2 q_{2k-1} < 1 \quad \text{and} \quad a^2 q_{2k} \le (1 - a\sqrt{q_{2k-1}})(1 - a\sqrt{q_{2k+1}}) \tag{2.5}$$

or

$$a^2 q_{2k} < 1 \quad \text{and} \quad a^2 q_{2k+1} \le (1 - a\sqrt{q_{2k}})(1 - a\sqrt{q_{2k+2}}) \tag{2.6}$$

holds for all sufficiently large $k \in \mathbb{N}$. *Then all nontrivial solutions of* (1.1) *are nonoscillatory.*

3 Comparison with Previous Studies

To compare our results given in Sects. 1 and 2 with previous works, we give some examples. First, we introduce several previous works that are related.

From various viewpoints, Hinton and Lewis [6] discussed the difference equation

$$c_n x_{n+1} + c_{n-1} x_{n-1} = b_n x_n, \tag{3.1}$$

which is equivalent to Eq. (1.4). They presented the following result on nonoscillation.

Theorem D *Suppose that all* $c_n = 1$,

$$\sum_{n=1}^{\infty} |b_n - 2| < \infty \quad \text{and} \quad \limsup_{k\to\infty} k \sum_{n=k}^{\infty} |b_n - 2| < \frac{1}{4}.$$

Then all nontrivial solutions of (3.1) *are nonoscillatory.*

Since $p_n = c_n + c_{n-1} - b_n$, we can rewrite conditions in Theorem D as

$$\sum_{n=1}^{\infty} |p_n| < \infty \quad \text{and} \quad \limsup_{k\to\infty} k \sum_{n=k}^{\infty} |p_n| < \frac{1}{4}. \tag{3.2}$$

Theorem D is a discrete analogue of Hille's nonoscillation result in [5]. Unfortunately, we can use this result to no periodic difference equations except for Eq. (1.4) with $r_n \equiv 1$ and $p_n \equiv 0$ (or Eq. (3.1) with $b_n \equiv 2$ and $c_n \equiv 0$).

Chen and Erbe [3] obtained oscillation and nonoscillation criteria for Eq. (1.4) using Riccati techniques. The Riccati difference equation that they used is different from that of this paper. Their main assumptions were

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \sum_{j=1}^{k} p_j > -\infty \tag{3.3}$$

and others (see Theorem 2.3 in [3]). Consider the case that $\{p_n\}$ is a periodic sequence with period m. Let

$$\ell = n - \left[\frac{n}{m}\right] m,$$

where $[d]$ means the greatest integer that is less than or equal to a real number d. Then, ℓ is an integer satisfying $0 \le \ell \le m - 1$. We can estimate that

$$\sum_{k=1}^{n}\sum_{j=1}^{k} p_j = \frac{m}{2}\left[\frac{n}{m}\right]\left(1+\left[\frac{n}{m}\right]\right)\sum_{j=1}^{m} p_j - \left[\frac{n}{m}\right]\sum_{j=1}^{m}(j-1)p_j$$
$$+\left[\frac{n}{m}\right]\sum_{j=1}^{m} p_j + \sum_{j=1}^{\ell} p_j$$
$$= \frac{m}{2}\left[\frac{n}{m}\right]\left(1+\left[\frac{n}{m}\right]\right)\sum_{j=1}^{m} p_j + \left[\frac{n}{m}\right]\sum_{j=1}^{m}(2-j)p_j + \sum_{j=1}^{\ell} p_j.$$

Here, we regard $\sum_{j=1}^{0} p_j$ as 0. Since $n/m - 1 < [n/m] \le n/m$, we see that

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{k} p_j = \frac{1}{2}\frac{m}{n}\left[\frac{n}{m}\right]\left(1+\left[\frac{n}{m}\right]\right)\sum_{j=1}^{m} p_j + \frac{1}{n}\left[\frac{n}{m}\right]\sum_{j=1}^{m}(2-j)p_j + \frac{1}{n}\sum_{j=1}^{\ell} p_j$$
$$< \frac{1}{2}\frac{m}{n}\left[\frac{n}{m}\right]\left(1+\left[\frac{n}{m}\right]\right)\sum_{j=1}^{m} p_j + \frac{1}{m}\left|\sum_{j=1}^{m}(2-j)p_j\right| + \frac{1}{n}\left|\sum_{j=1}^{\ell} p_j\right|.$$

If $\sum_{j=1}^{m} p_j < 0$, then

$$\frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{k} p_j < \frac{1}{2}\left(1-\frac{m}{n}\right)\left(1+\left[\frac{n}{m}\right]\right)\sum_{j=1}^{m} p_j + \frac{1}{m}\left|\sum_{j=1}^{m}(2-j)p_j\right| + \frac{1}{n}\left|\sum_{j=1}^{\ell} p_j\right|.$$

Since $\sum_{j=1}^{m}(2-j)p_j$ and $\sum_{j=1}^{\ell} p_j$ are finite, we see that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{k} p_j = -\infty.$$

This contradicts (3.3). Thus, the above result of Chen and Erbe cannot use when $\sum_{j=1}^{m} p_j < 0$.

Using Eq. (2.1) with $a = 1$ and a comparison theorem proved by Kwong [13, Theorem 2], Abu-Risha [1, Theorem 2.1] gave a necessary and sufficient condition for nonoscillation of solutions of (1.4). Although this result is simple, it does not give any concrete condition about the sequence $\{q_n\}$. To this end, Abu-Risha also presented an explicit condition between q_{n-1}, q_n and q_{n+1} for $n \in \mathbb{N}$. We can extend his result as follows.

Theorem E *All nontrivial solutions of* (1.1) *are nonoscillatory if there is an* $N \in \mathbb{N}$ *such that*

$$a^2\left(\sqrt{q_{n+1}} + \sqrt{q_n}\right)\left(\sqrt{q_n} + \sqrt{q_{n-1}}\right) \le 1 \tag{3.4}$$

holds for $n \ge N$.

Parhi [18, Theorem 2.1] showed that if $p_n \le 0$ for all sufficiently large $n \in \mathbb{N}$, then all nontrivial solutions of (1.1) are nonoscillatory. Ma [14, Lemma 2.3] already pointed out that the same result holds for Eq. (1.4).

Now, we give an example of Corollary 1.

Example 1 For any $a > 0$, let $r_0 = 3.4$,

$$r_n = \begin{cases} 4.4 & \text{if } n = 4k-3, \\ 3.5 & \text{if } n = 4k-2, \\ 5.8 & \text{if } n = 4k-1, \\ 3.4 & \text{if } n = 4k \end{cases} \quad \text{and} \quad p_n = \begin{cases} 3.8a & \text{if } n = 4k-3, \\ -8.1a & \text{if } n = 4k-2, \\ 5.3a & \text{if } n = 4k-1, \\ -15.8a & \text{if } n = 4k \end{cases}$$

with $k \in \mathbb{N}$. Then all nontrivial solutions of (1.1) are nonoscillatory.

It is clear that conditions (1.2) and (1.3) hold. Since

$$\begin{aligned} a^2 q_n &= \frac{a^2 r_n^2}{\{a(r_n + r_{n-1}) - p_n\}\{a(r_{n+1} + r_n) - p_{n+1}\}} \\ &= \begin{cases} 0.3025 & \text{if } n = 4k-3, \\ 0.19140625 & \text{if } n = 4k-2, \\ 0.3364 & \text{if } n = 4k-1, \\ 0.1156 & \text{if } n = 4k, \end{cases} \end{aligned} \tag{3.5}$$

we obtain

$$a^2(q_{4k-3} + q_{4k-2}) = 0.49390625 < 0.5$$

and

$$a^2(q_{4k-1} + q_{4k}) = 0.452 < 0.5$$

with $k \in \mathbb{N}$. Hence, the inequality (2.2) holds. Thus, by Corollary 1, all nontrivial solutions of (1.1) are nonoscillatory.

Here, we simulate a solution of (1.1) with the sequences $\{r_n\}$ and $\{p_n\}$ that were given in Example 1 (see Fig. 1). To make the behavior of the solution more apparent, we connect the dots x_{n-1} and x_n with a line segment and draw a line graph.

From Fig. 1, we see that the polygonal line tends to rise while repeating a vertical motion. Since the polygonal line does not cross the horizontal n-axis, this solution $\{x_n\}$ is nonoscillatory. Sturm's separation theorem guarantees that all nontrivial solutions are nonoscillatory.

We also simulate a solution $\{z_n\}$ of (2.1) (see Fig. 2). This solution corresponds to the solution of (1.1) drawn in Fig. 1.

Remark 3 The inequality (2.3) does not hold in Example 1. In fact,

$$a^2(q_{4k-2} + q_{4k-1}) = 0.52780625 > 0.5$$

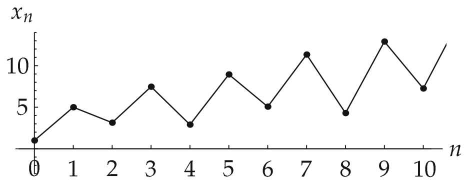


Fig. 1 This polygonal line displays the motion of a solution $\{x_n\}$ of (1.1) given in Example 1 with $a = 4/5$. The initial condition of the solution is $(x_0, x_1) = (1, 5)$.

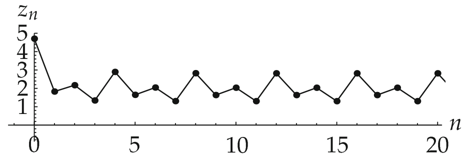


Fig. 2 Riccati's equation (2.1) has a positive solution $\{z_n\}$ when the sequence $\{q_n\}$ satisfies (3.5) with $a = 4/5$. The initial condition of the solution is $z_0 = 80/17$.

for any $k \in \mathbb{N}$.

Condition (3.2) is not satisfied, because $\sum_{n=1}^{\infty} |p_n|$ is infinity. Hence, Theorem D is useless to verify Example 1. Since

$$p_1 + p_2 + p_3 + p_4 = -14.8a < 0$$

and p_n is not always negative, the results of Chen and Erbe [3] and Parhi [18] mentioned above are also useless for Example 1. From (3.5) it is clear that Theorem B cannot be applied to Example 1. Using (3.5), we can compute

$$a(\sqrt{q_{n+1}} + \sqrt{q_n}) = \begin{cases} 0.9875 & \text{if } n = 4k-3, \\ 1.0175 & \text{if } n = 4k-2, \\ 0.92 & \text{if } n = 4k-1, \\ 0.89 & \text{if } n = 4k \end{cases}$$

with $k \in \mathbb{N}$. Hence, we have

$$a^2(\sqrt{q_{n+1}} + \sqrt{q_n})(\sqrt{q_n} + \sqrt{q_{n-1}}) = \begin{cases} 1.00478125 & \text{if } n = 4k-2, \\ 0.9361 & \text{if } n = 4k-1, \\ 0.8188 & \text{if } n = 4k, \\ 0.878875 & \text{if } n = 4k+1 \end{cases}$$

with $k \in \mathbb{N}$. There is no $N \in \mathbb{N}$ where the inequality (3.4) is satisfied for all $n \geq N$. Hence, Theorem E is also not available in Example 1.

Next, we give an example of Corollary 2.

Example 2 For any $a > 0$, let $r_0 = 4$,

$$r_n = \begin{cases} 10 & \text{if } n = 4k-3, \\ 2 & \text{if } n = 4k-2, \\ 1.5 & \text{if } n = 4k-1, \\ 4 & \text{if } n = 4k \end{cases} \quad \text{and} \quad p_n = \begin{cases} -6a & \text{if } n = 4k-3, \\ -13a & \text{if } n = 4k-2, \\ 2.5a & \text{if } n = 4k-1, \\ 0.5a & \text{if } n = 4k \end{cases}$$

with $k \in \mathbb{N}$. Then all nontrivial solutions of (1.1) are nonoscillatory.

Conditions (1.2) and (1.3) are satisfied obviously. It is easy to check that

$$a^2 q_n = \frac{a^2 r_n^2}{\{a(r_n + r_{n-1}) - p_n\}\{a(r_{n+1} + r_n) - p_{n+1}\}} = \begin{cases} 0.2 & \text{if } n = 4k-3, \\ 0.16 & \text{if } n = 4k-2, \\ 0.45 & \text{if } n = 4k-1, \\ 0.16 & \text{if } n = 4k. \end{cases} \tag{3.6}$$

Hence, the inequality (2.5) holds. In fact, $a^2 q_{4k-3} = 0.2 < 1$, $a^2 q_{4k-1} = 0.45 < 1$,

$$a^2 q_{4k-2} = 0.16 < (1 - \sqrt{0.2})(1 - \sqrt{0.45}) = (1 - a\sqrt{q_{4k-3}})(1 - a\sqrt{q_{4k-1}}),$$
$$a^2 q_{4k} = 0.16 < (1 - \sqrt{0.45})(1 - \sqrt{0.2}) = (1 - a\sqrt{q_{4k-1}})(1 - a\sqrt{q_{4k+1}})$$

for all $k \in \mathbb{N}$. Thus, by Corollary 2, all nontrivial solutions of (1.1) are nonoscillatory.

To illustrate Example 2, we give two figures. One is the polygonal line described a solution $\{x_n\}$ of (1.1) (see Fig. 3). This polygonal line seems like a series of mountains. The polygonal line does not intersect the horizontal n-axis. Hence, this solution is nonoscillatory. From Sturm's separation theorem, we see that all nontrivial solutions are nonoscillatory. The other is the polygonal line displayed the behavior of a solution of $\{z_n\}$ of (2.1) (see Fig. 4). This solution corresponds to the solution of (1.1) drawn in Fig. 3.

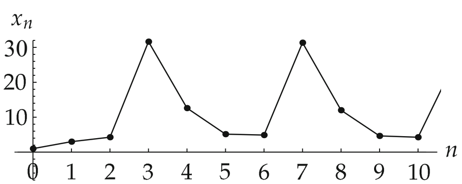


Fig. 3 This polygonal line displays the motion of a solution $\{x_n\}$ of (1.1) given in Example 2 with $a = 3/4$. The initial condition of the solution is $(x_0, x_1) = (1, 3)$.

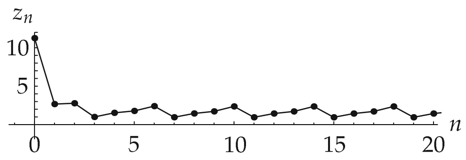


Fig. 4 Riccati's equation (2.1) has a positive solution $\{z_n\}$ when the sequence $\{q_n\}$ satisfies (3.6) with $a = 3/4$. The initial condition of the solution is $z_0 = 45/4$.

Remark 4 The inequality (2.6) does not hold in Example 2, because

$$a^2 q_{4k-1} = 0.45 > 0.36 = (1 - \sqrt{0.16})^2 = (1 - a\sqrt{q_{4k-2}})(1 - a\sqrt{q_{4k}})$$

with $k \in \mathbb{N}$.

Remark 5 We cannot apply Corollary 1 to Example 2, because both inequalities (2.2) and (2.3) are not satisfied. In fact, from (3.6) it follows that

$$a^2(q_{4k-1} + q_{4k}) = 0.61 > 0.5,$$
$$a^2(q_{4k-2} + q_{4k-1}) = 0.61 > 0.5$$

for all $k \in \mathbb{N}$.

From (3.6), we see that $a^2 q_{4k-1} > 1/4$ for $k \in \mathbb{N}$. Hence, Theorem B is useless for Example 2. Since $r_n \not\equiv 1$ and $\sum_{n=1}^{\infty} |p_n|$ is infinity, Theorem D is inapplicable to Example 2. The results of Chen and Erbe [3] and Parhi [18] mentioned above are also useless for Example 2, because

$$p_1 + p_2 + p_3 + p_4 = -16a < 0$$

and p_n is not always negative. It turns out from (3.6) that

$$a(\sqrt{q_{n+1}} + \sqrt{q_n}) = \begin{cases} (2+\sqrt{5})/5 & \text{if } n = 4k-3, \\ (3\sqrt{5}+4)/10 & \text{if } n = 4k-2, \\ (4+3\sqrt{5})/10 & \text{if } n = 4k-1, \\ (\sqrt{5}+2)/5 & \text{if } n = 4k \end{cases}$$

with $k \in \mathbb{N}$. Hence, we have

$$a^2\left(\sqrt{q_{n+1}} + \sqrt{q_n}\right)\left(\sqrt{q_n} + \sqrt{q_{n-1}}\right) = \begin{cases} 0.9072135954999581\cdots & \text{if } n = 4k-2, \\ 1.14665631459995\cdots & \text{if } n = 4k-1, \\ 0.9072135954999581\cdots & \text{if } n = 4k, \\ 0.7177708763999665\cdots & \text{if } n = 4k+1 \end{cases}$$

with $k \in \mathbb{N}$. There is no $N \in \mathbb{N}$ where the inequality (3.4) is satisfied for all $n \geq N$. Hence, Theorem E is also inapplicable to Example 2.

4 Further Nonoscillation Criteria

In Sect. 3, we have focused on a weighted sum of two adjacent terms q_{2k-1} and q_{2k} or two adjacent terms q_{2k} and q_{2k+1}. The most simple case of the weighted sum of two adjacent terms is the arithmetic mean between two terms q_{2k-1} and q_{2k} or two terms q_{2k} and q_{2k+1} (see Corollary 1). The weight does not have to be a constant and it is allowed to change depending on the value of q_n (see Corollary 2). In this section, by taking account of several sets of two adjacent terms, we extend Theorem 1 as follows.

Theorem 2 *Assume* (1.2) *and* (1.3). *Suppose that there exists an* $N \in \mathbb{N}$ *such that for any* $k \geq N$ *there are two sequences* $\{\alpha_k\}$ *and* $\{\beta_k\}$ *with* $\alpha_k > 1$ *and* $\beta_k > 1$. *If*

$$\frac{a^2\alpha_k}{\alpha_k - 1}\, q_{4k-3} + a^2\beta_k q_{4k-2} \leq 1 \tag{4.1}$$

and

$$\frac{a^2\beta_k}{\beta_k - 1}\, q_{4k-1} + a^2\alpha_{k+1} q_{4k} \leq 1, \tag{4.2}$$

then all nontrivial solutions of (1.1) *are nonoscillatory.*

Proof From (4.1) and (4.2) it follows that

$$a^2\beta_k \leq \frac{1}{q_{4k-2}}\left(1 - \frac{a^2\alpha_k}{\alpha_k - 1}\, q_{4k-3}\right)$$

and

$$a^2\alpha_{k+1} \leq \frac{1}{q_{4k}}\left(1 - \frac{a^2\beta_k}{\beta_k - 1}\, q_{4k-1}\right)$$

for all $k \geq N$. Consider a solution $\{z_n\}$ of (2.1) satisfying $z_{4N-4} \geq a^2\alpha_N > a^2$. Then we can check that

$$z_{4N-3} = \frac{1}{q_{4N-3}}\left(1 - \frac{a^2}{z_{4N-4}}\right) \geq \frac{1}{q_{4N-3}}\left(1 - \frac{a^2}{a^2\alpha_N}\right) = \frac{\alpha_N - 1}{\alpha_N q_{4N-3}} > 0,$$
$$z_{4N-2} = \frac{1}{q_{4N-2}}\left(1 - \frac{a^2}{z_{4N-3}}\right) \geq \frac{1}{q_{4N-2}}\left(1 - \frac{a^2\alpha_N}{\alpha_N - 1}\, q_{4N-3}\right) \geq a^2\beta_N > a^2,$$

$$z_{4N-1} = \frac{1}{q_{4N-1}}\left(1 - \frac{a^2}{z_{4N-2}}\right) \geq \frac{1}{q_{4N-1}}\left(1 - \frac{a^2}{a^2\beta_N}\right) = \frac{\beta_N - 1}{\beta_N q_{4N-1}} > 0,$$
$$z_{4N} = \frac{1}{q_{4N}}\left(1 - \frac{a^2}{z_{4N-1}}\right) \geq \frac{1}{q_{4N}}\left(1 - \frac{a^2\beta_N}{\beta_N - 1}q_{4N-1}\right) \geq a^2\alpha_{N+1} > a^2.$$

We inductively obtain

$$z_n \geq \begin{cases} \dfrac{\alpha_k - 1}{\alpha_k q_{4k-3}} & \text{if } n = 4k-3, \\ a^2\beta_k & \text{if } n = 4k-2, \\ \dfrac{\beta_k - 1}{\beta_k q_{4k-1}} & \text{if } n = 4k-1, \\ a^2\alpha_{k+1} & \text{if } n = 4k \end{cases}$$

with $k \geq N$. Hence, the sequence $\{z_n\}$ is a positive solution of (2.1). We therefore conclude that all nontrivial solutions of (1.1) are nonoscillatory. □

By the same way, we have the following result (we omit the proof).

Theorem 3 *Assume* (1.2) *and* (1.3). *Suppose that there exists an* $N \in \mathbb{N}$ *such that for any* $k \geq N$ *there are two sequences* $\{\alpha_k\}$ *and* $\{\beta_k\}$ *with* $\alpha_k > 1$ *and* $\beta_k > 1$. *If*

$$\frac{a^2\alpha_k}{\alpha_k - 1}q_{4k-2} + a^2\beta_k q_{4k-1} \leq 1 \tag{4.3}$$

and

$$\frac{a^2\beta_k}{\beta_k - 1}q_{4k} + a^2\alpha_{k+1}q_{4k+1} \leq 1, \tag{4.4}$$

then all nontrivial solutions of (1.1) *are nonoscillatory.*

Remark 6 If the inequalities (4.1) and (4.2) are satisfied for $k \in \mathbb{N}$ sufficiently large, then the inequality (1.7) also holds. In fact, let

$$\gamma_k = \begin{cases} \alpha_\ell & \text{if } k = 2\ell - 1, \\ \beta_\ell & \text{if } k = 2\ell \end{cases}$$

with $l \in \mathbb{N}$. Then, by (4.1) and (4.2) we obtain

$$\frac{\gamma_k}{\gamma_k - 1}a^2 q_{2k-1} + \gamma_{k+1}a^2 q_{2k} \leq 1;$$

namely, the inequality (1.7). Similarly, if the inequalities (4.3) and (4.4) are satisfied for $k \in \mathbb{N}$ sufficiently large, then the inequality (1.8) also holds. Hence, Theorems 2 and 3 also extend Theorem 1.

Let p be a real number that is larger than 1 and let p^* be the conjugate number of p; namely,

$$\frac{1}{p} + \frac{1}{p^*} = 1.$$

Then p^* is also greater than 1. We choose constants $\alpha > 1$ and $\beta > 1$ as the two sequences $\{\alpha_k\}$ and $\{\beta_k\}$ in Theorems 2 and 3, respectively. Then the inequalities (4.1)–(4.4) become

$$a^2(\alpha^* q_{4k-3} + \beta q_{4k-2}) \le 1, \tag{4.5}$$

$$a^2(\beta^* q_{4k-1} + \alpha q_{4k}) \le 1, \tag{4.6}$$

$$a^2(\alpha^* q_{4k-2} + \beta q_{4k-1}) \le 1, \tag{4.7}$$

$$a^2(\beta^* q_{4k} + \alpha q_{4k+1}) \le 1, \tag{4.8}$$

respectively. Hence, we have the following corollaries of Theorems 2 and 3.

Corollary 3 *Assume* (1.2) *and* (1.3). *Suppose that there exists an* $N \in \mathbb{N}$ *such that both* (4.5) *and* (4.6) *hold for* $k \ge N$. *Then all nontrivial solutions of* (1.1) *are nonoscillatory.*

Corollary 4 *Assume* (1.2) *and* (1.3). *Suppose that there exists an* $N \in \mathbb{N}$ *such that both* (4.7) *and* (4.8) *hold for* $k \ge N$. *Then all nontrivial solutions of* (1.1) *are nonoscillatory.*

We here give an example of Corollary 3.

Example 3 For any $a > 0$, let $r_0 = 2.4$,

$$r_n = \begin{cases} 3.1 & \text{if } n = 4k-3, \\ 1 & \text{if } n = 4k-2, \\ 2 & \text{if } n = 4k-1, \\ 2.4 & \text{if } n = 4k \end{cases} \quad \text{and} \quad p_n = \begin{cases} -3.5a & \text{if } n = 4k-3, \\ 1.1a & \text{if } n = 4k-2, \\ -6a & \text{if } n = 4k-1, \\ 2.4a & \text{if } n = 4k \end{cases}$$

with $k \in \mathbb{N}$. Then all nontrivial solutions of (1.1) are nonoscillatory.

Clearly, conditions (1.2) and (1.3) hold. The sequence $\{q_n\}$ satisfies

$$\begin{aligned} a^2 q_n &= \frac{a^2 r_n^2}{\{a(r_n + r_{n-1}) - p_n\}\{a(r_{n+1} + r_n) - p_{n+1}\}} \\ &= \begin{cases} 961/2700 & \text{if } n = 4k-3, \\ 1/27 & \text{if } n = 4k-2, \\ 2/9 & \text{if } n = 4k-1, \\ 8/25 & \text{if } n = 4k. \end{cases} \end{aligned} \tag{4.9}$$

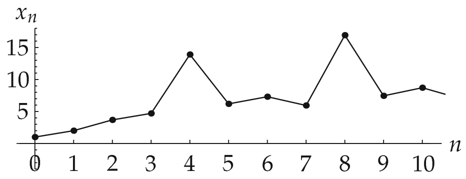


Fig. 5 This polygonal line displays the motion of a solution $\{x_n\}$ of (1.1) given in Example 3 with $a = 7/10$. The initial condition of the solution is $(x_0, x_1) = (1, 2)$.

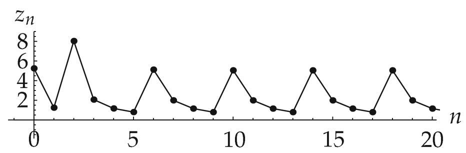


Fig. 6 Riccati's equation (2.1) has a positive solution $\{z_n\}$ when the sequence $\{q_n\}$ satisfies (4.9) with $a = 7/10$. The initial condition of the solution is $z_0 = 21/4$.

Let $\alpha = 2$ and $\beta = 3$. Then we obtain

$$\alpha^* a^2 q_{4k-3} + \beta a^2 q_{4k-2} = 2 \times \frac{961}{2700} + 3 \times \frac{1}{27} = \frac{1111}{1350} < 1$$

and

$$\beta^* a^2 q_{4k-1} + \alpha a^2 q_{4k} = \frac{3}{2} \times \frac{2}{9} + 2 \times \frac{8}{25} = \frac{73}{75} < 1;$$

namely, the inequalities (4.5) and (4.6) are satisfied for all $k \in \mathbb{N}$. Hence, by Corollary 3, all nontrivial solutions of (1.1) are nonoscillatory.

We confirm Example 3 by using two simulations. Figure 5 shows the behavoir of a solution $\{x_n\}$ of (1.1) given in Example 3. As in Figs. 1 and 3, the behavior of this solution is represented by a polygonal line. This polygonal line tends to rise slowly while moving up and down. Since the polygonal line does not meet the horizontal n-axis, this solution is nonoscillatory. Hence, by Sturm's separation theorem, all nontrivial solutions are nonoscillatory. Recall that each nonoscillatory solution of (1.1) corresponds to a positive solution of (2.1). Figure 6 displays the motion of the solution of (2.1) corresponding to the solution of (1.1) drawn in Fig. 5.

Remark 7 We cannot apply Corollary 1 to Example 3, because both inequalities (2.2) and (2.3) are not satisfied. In fact, from (4.9), we see that

$$a^2(q_{4k-1} + q_{4k}) = \frac{2}{9} + \frac{8}{25} = \frac{122}{225} > \frac{1}{2};$$
$$a^2(q_{4k} + q_{4k+1}) = \frac{8}{25} + \frac{961}{2700} = \frac{73}{108} > \frac{1}{2}$$

for all $k \in \mathbb{N}$.

Remark 8 For any $k \in \mathbb{N}$,

$$a^2 q_{4k} = \frac{8}{25} = 0.32 > 0.213237912254794\cdots$$
$$= \left(1 - \sqrt{\frac{2}{9}}\right)\left(1 - \sqrt{\frac{961}{2700}}\right) = (1 - a\sqrt{q_{4k-1}})(1 - a\sqrt{q_{4k+1}})$$

and

$$a^2 q_{4k+1} = \frac{961}{2700} = 0.3559259259259259\cdots > 0.3507306961112502\cdots$$
$$= \left(1 - \sqrt{\frac{8}{25}}\right)\left(1 - \sqrt{\frac{1}{27}}\right) = (1 - a\sqrt{q_{4k}})(1 - a\sqrt{q_{4k+2}}).$$

Hence, both inequalities (2.5) and (2.6) are not satisfied, and therefore, Corollary 2 cannot be applied to Example 3.

We can easily check that Theorems B and D are not applied to Example 2. It is also clear that the results of Chen and Erbe [3] and Parhi [18] mentioned in Sect. 3 cannot be applied to Example 2. From (4.9) it follows that

$$a(\sqrt{q_{n+1}} + \sqrt{q_n}) = \begin{cases} 41\sqrt{3}/90 & \text{if } n = 4k-3, \\ (3\sqrt{2} + \sqrt{3})/9 & \text{if } n = 4k-2, \\ 11\sqrt{2}/15 & \text{if } n = 4k-1, \\ (31\sqrt{3} + 36\sqrt{2})/90 & \text{if } n = 4k \end{cases}$$

with $k \in \mathbb{N}$. Hence, we have

$$a^2\left(\sqrt{q_{n+1}} + \sqrt{q_n}\right)\left(\sqrt{q_n} + \sqrt{q_{n-1}}\right) = \begin{cases} 0.5238114053855937\cdots & \text{if } n = 4k-2, \\ 0.6884769420045553\cdots & \text{if } n = 4k-1, \\ 1.205389631325233\cdots & \text{if } n = 4k, \\ 0.9170922049812309\cdots & \text{if } n = 4k+1 \end{cases}$$

with $k \in \mathbb{N}$. There is no $N \in \mathbb{N}$ where the inequality (3.4) is satisfied for all $n \geq N$. Hence, Theorem E is also inapplicable to Example 3.

Acknowledgements The first author's work was supported in part by Grant-in-Aid for Scientific Research No. 25400165 and 17K05327 from the Japan Society for the Promotion of Science.

References

1. Abu-Risha, M.H.: Oscillation of second-order linear difference equations. Appl. Math. Lett. **13**, 129–135 (2000)
2. Agarwal, R.P., Grace, S.R., O'Regan, D.: Oscillation Theory for Second Order Linear, Half-linear, Superlinear and Sublinear Dynamic Equations. Kluwer Academic Publishers, Dordrecht (2002)
3. Chen, S., Erbe, L.H.: Riccati techniques and discrete oscillations. J. Math. Anal. Appl. **142**, 468–487 (1989)
4. Elaydi, S.: An Introduction to Difference Equations, 3rd edn. Springer-Verlag, New York, Undergraduate Texts in Mathematics (2005)
5. Hille, E.: Non-oscillation theorems. Trans. Amer. Math. Soc. **64**, 234–252 (1948)
6. Hinton, D.B., Lewis, R.T.: Spectral analysis of second order difference equations. J. Math. Anal. Appl. **63**, 421–438 (1978)
7. Hooker, J.W.: A Hille-Wintner type comparison theorem for second order difference equations. Internat. J. Math. Math. Sci. **6**, 387–393 (1983)
8. Hooker, J.W., Patula, W.T.: Riccati type transformations for second-order linear difference equations. J. Math. Anal. Appl. **82**, 451–462 (1981)
9. Hooker, J.W., Kwong, M.K., Patula, W.T.: Oscillatory second order linear difference equations and Riccati equations. SIAM J. Math. Anal. **18**, 54–63 (1987)
10. Hosseinzadeh, H., Afrouzi, G.A.: Forward (r, s)-difference operator r, s and solving difference equations. Appl. Math. Sci. (Ruse) **2**, 3005–3013 (2008)
11. Kelley, W.G., Peterson, A.C.: Difference Equations: An Introduction with Applications. Academic Press, Boston (1991)
12. Kneser, A.: Untersuchungen über die reellen Nullstellen der Integrale linearer Differentialgleichungen. Math. Ann. **42**, 409–435 (1893)
13. Kwong, M.K., Hooker, J.W., Patula, W.T.: Riccati type transformations for second-order linear difference equations. II. J. Math. Anal. Appl. **107**, 182–196 (1985)
14. Ma, M.: Dominant and recessive solutions for second order self-adjoint linear difference equations. Appl. Math. Lett. **18**, 179–185 (2005)
15. Manuel, M.M.S., Kılıçman, A., Srinivasan, K., Babu, G.D.: Oscillation of solutions of some generalized nonlinear α-difference equations. Adv. Differ. Equ. 2014, 8 pp. (2014:109)
16. Manuel, M.M.S., Srinivasan, K., Dilip, D.S., Babu, G.D.: Asymptotic and oscillatory behaviour of certain nonlinear generalized α-difference equations. Int. J. Pure Appl. Math. **92**, 549–563 (2014)
17. Nehari, Z.: Oscillation criteria for second-order linear differential equations. Trans. Amer. Math. Soc. **85**, 428–445 (1957)
18. Parhi, N.: Oscillation and non-oscillation of solutions of second order difference equations involving generalized difference. Appl. Math. Comput. **218**, 458–468 (2011)
19. Parhi, N., Panda, A.: Nonoscillation and oscillation of solutions of a class of third order difference equations. J. Math. Anal. Appl. **336**, 213–223 (2007)
20. Popenda, J.: Oscillation and nonoscillation theorems for second-order difference equations. J. Math. Anal. Appl. **123**, 34–38 (1987)
21. Swanson, C.A.: Comparison and Oscillation Theory of Linear Differential Equations. Mathematics in Science and Engineering, vol. 48, Academic Press, New York (1968)
22. Tan, M.-C., Yang, E.-H.: Oscillation and nonoscillation theorems for second order nonlinear difference equations. J. Math. Anal. Appl. **276**, 239–247 (2002)
23. Wintner, A.: On the comparison theorem of Kneser-Hille. Math. Scand. **5**, 255–260 (1957)

An Evolutionary Game Model of Families' Voluntary Provision of Public Goods

Aiko Tanaka and Jun-ichi Itaya

Abstract We consider a two-stage voluntary provision model where individuals in a family contribute to a pure public good and/or a household public good, and an altruistic parent makes a non-negative income transfer to his or her child. The subgame perfect equilibrium derived in the model is analyzed using two evolutionary dynamics games (i.e., replicator dynamics and best response dynamics). As a result, the equilibria with ex-post transfers and pre-committed transfers coexist, and are unstable in the settings of replicator dynamics as well as best response dynamics, whereas the monomorphic states (i.e., all families undertake either ex-post or pre-committed transfers) are stable. An income redistribution policy does not alter the real allocations in the settings of both evolutionary dynamics games, because the resulting real allocations depend on only the total income of society and not on the distribution of individual income.

Keywords Voluntary provision · Subgame perfect equilibrium · Evolutionary game

1 Introduction

In this chapter, we study the private provision of public goods using a framework of evolutionary game theory. When an income transfer is made from a parent to a child, we can consider two cases: first, the case that the child acts after observing that the parent makes an income transfer to the child, and second, when the parent makes an income transfer after observing the child's action. Focusing our analysis on the timing when the parent makes an income transfer to the child, we call the

A. Tanaka (✉) · J.-i. Itaya
Graduate School of Economics and Business Administration,
Hokkaido University, Kita9 Nishi7, Kita-ku, Sapporo 060-0809, Japan
e-mail: atanaka@econ.hokudai.ac.jp

J.-i. Itaya
e-mail: itaya@econ.hokudai.ac.jp

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_17

former "ex-post" and the latter "pre-committed" in this chapter. For simplification, these income transfers are termed "bequests" and "gifts," respectively. We assume that a gift tax and inheritance tax do not exist.

In our model, we consider a family consisting of one selfish child who is not concerned with his parent's utility and one altruistic parent who is concerned with her child's utility. We assume that the parent and the child make voluntary contributions to public goods and that the parent makes a non-negative income transfer to the child. There exist a finite number of families, and they take an ex-post or pre-committed strategy about parent's transfer to the child. Further, it is assumed that two families are randomly paired from such a population and they play a two-stage game.

We study two types of evolutionary game dynamics.[1] The first is replicator dynamics, where the share of a strategy changes through time depending on the difference between a strategy's expected payoff and the group's average expected payoff. The child imitates the parent's strategy. The next generation's strategy share increases if the expected payoff using this strategy is larger than the average group payoff, and it is weeded out if the expected payoff by the strategy is smaller than the group average. The second is best-response dynamics, where the share of a strategy changes through time depending on the difference between the expected payoff of the strategy and the other. Families are able to monitor the strategy distribution precisely at a probability in a period, and after monitoring, they choose a strategy to maximize their own payoff.

We derive the subgame perfect equilibrium for the drawn two families, and check whether subgame perfect equilibrium is stable or not in replicator dynamics and best-response dynamics. Cornes, Itaya, and Tanaka [1] constructed a one-shot, two-stage voluntary provision model and we basically conform to it. In order to construct replicator dynamics, we need to know the parent's expected utility for ex-post and pre-committed, and group's average expected utility. We think that it is difficult to rank a parent's expected utility or group average utility for a subgame perfect equilibrium or to conclude which is the best strategy for the family. To overcome the difficulty of ranking the parent's expected utility for ex-post and pre-committed, and group's average expected utility for preference parameters, we will find a sufficient condition which endure that the Nash equilibrium is stable.

The rest of the chapter is organized as follows. We present the details of the model in Sect. 2. Section 3 shows the two-stage game subgame perfect equilibrium when two families are matched. We analyze the replicator dynamics in Sect. 4, the best-response dynamics in Sect. 5. We summarize with a conclusion in Sect. 5.

2 Model

We consider a population with a size normalized to 1 (when considering best-response dynamics, we assume a finite number of families, say, N). Two families are randomly

[1] See Samuelson [3] and Fudenberg and Levine [2] for details of the evolutionary dynamics.

paired from the population, and they play a random matching game. Each family consists of two agents, an altruistic parent and an egoistic child, and they act sequentially. For one of the families in the pair, call family i, whose members are identified by superscript i, the utility function of the parent who is altruistic towards her child and that of the child who is egoistic are given using Cobb–Douglas utilities as follows:

$$U_p^i = u_p^i + \alpha U_k^i = \ln(c_p^i)^r (G)^{1-r} + \alpha \ln(c_k^i)^s (G)^{1-s}, \quad i = 1, 2, \tag{1}$$

$$U_k^i = \ln(c_k^i)^s (G)^{1-s}, \quad i = 1, 2. \tag{2}$$

where c_p^i and c_k^i are the parent's and child's consumptions of the private good, respectively, G is an interfamily public good, and α is the parameter that measures the strength of the parent's altruism towards her child and is assumed to be common among families for simplification. Note also that under the Cobb-Douglas utility functions specified by (1) and (2), c_p^i (and c_k^i) and G are normal goods. We assume that $\alpha \in [0, 1]$, which implies that the parent neither cares about her child more than she cares about herself nor hates her child.

Public good G is an interfamily public good whose benefits spill over to members of the other family. Moreover, the public good is entirely supplied by voluntary contributions made by the parent and child, g_p^i and g_k^i, respectively, of family $i = 1, 2$. The public good is thus produced according to the following summation technology:

$$G = \sum_{i=1}^{2} \left(g_p^i + g_k^i\right). \tag{3}$$

Because a parent is altruistic to her child, she makes an non-negative income transfer to him. The familiy's strategy for income transfer are ex-post or pre-committed. For the ex-post strategy, at Stage 1, the child of family i chooses his own consumption c_k^i and contributes to the public good g_k^i. At Stage 2, after having observed the contributions made by the children of both families (g_k^1, g_k^2), the parent of family i chooses c_p^i and g_p^i (or equivalently, π^i and g_p^i) to maximize her utility function (1) subject to

$$c_p^i + \pi^i + g_p^i = y_p^i, \quad i = 1, 2, \tag{4}$$

$$c_k^i + g_k^i = y_k^i + \pi^i, \quad i = 1, 2, \tag{5}$$

where $\pi^i \geq 0$ represents the transfer from the parent of family i to her child, and y_p^i and y_k^i are the fixed incomes of the parent and child of family i, respectively.

On the other hand, a pre-committed act is as follows: the parent pre-commits to a fixed transfer before the child chooses his public good contribution. Given the pre-committed transfer π^i, the child chooses his contribution at Stage 2 to maximize the utility function.

There exist ex-post-type families at the rate $0 \leq x(t) \leq 1$, while there exist pre-committed-type families at the rate $1 - x(t)$. Subscript t denote the time, $t = 0, 1, \ldots, T$ and the population state at time t is denoted by $(x(t), 1 - x(t))$.

The transfer timing strategy of the family is decided by the parent of the family, so we assume that the family utility is equal to the parent's utility, $U_p^i(c_p^i, G; c_k^i)$ $i = 1, 2$. If both families are of the ex-post type, both families gain utility a. If one family is ex-post and the other is pre-committed, the ex-post family gains utility b, and the pre-committed family gains utility c. If both families are pre-committed, both families gain utility d. The payoffs a, b, c, and d in the following payoff matrix are the outcome of the subgame perfect equilibrium we derive in the next section. The payoff matrix is summarized as follows:

		family 2	
		ex-post	pre-committed
family 1	ex-post	$a,\ a$	$b,\ c$
	pre-committed	$c,\ b$	$d,\ d$

(6)

In this study, we analyze the conditions that improve the timing of the families' strategies. To see the change of population status dynamically, we analyze two dynamics: replicator dynamics and best-response dynamics, which are used in the analysis of the evolutionary game. For simplicity, the payoffs in (6) which is obtained from the subgame perfect equilibrium of two-stage game are used to analyze the evolutionary game dynamics, replicator dynamics and best-response dynamics.

3 Subgame Perfect Equilibrium

We obtain a subgame perfect equilibrium for three types of two-stage games which we describe in the previous section and calculate the values of the payoffs a, b, c, and d in the payoff matrix (6) in the following subsections.

3.1 Both Families are Ex-post

If two ex-post families are randomly drawn from the population, both gain utility a in the payoff matrix (6). We compute the value of a in this subsection. Using backward induction, parents maximize their utility (1) subject to (4) and (5) at the second stage. Assuming interior solutions, the first-order conditions are

$$\frac{\partial U_p^i}{\partial g_p^i} = \frac{-r}{y_p^i - \pi^i - g_p^i} + \frac{1 - r + \alpha(1-s)}{G} = 0, \quad i = 1, 2, \tag{7}$$

$$\frac{\partial U_p^i}{\partial \pi^i} = \frac{-r}{y_p^i - \pi^i - g_p^i} + \frac{\alpha s}{y_k^i + \pi^i - g_k^i} = 0, \quad i = 1, 2. \tag{8}$$

Solving (7) and (8) for g_p^i and π^i, using (3) yields

$$g_p^i = \frac{(1+\alpha)(y_p^i + y_k^i) - (r + \alpha s)(y_p^j + y_k^j)}{1 + \alpha(1+s) + r} - g_k^i, \quad i = 1, 2, \tag{9}$$

$$\pi^i = \frac{\alpha s(y_p^i + y_p^j + y_k^j) - (1 + \alpha + r)y_k^i}{1 + \alpha(1+s) + r} + g_k^i, \quad i = 1, 2. \tag{10}$$

At the first stage, the children maximize their utility (2) subject to (5) and taking into accounts parents' reactions (9) and (10). Assuming interior solutions, the first order conditions are [2]

$$\frac{\partial U_k^i}{\partial g_k^i} = \frac{s}{y_k^i + \pi^i(g_k^i) - g_k^i}\left(\frac{d\pi^i(g_k^i)}{dg_k^i} - 1\right) + \frac{1-s}{g_k^i + g_p^i(g_k^i) + g_p^j + g_k^j}\left(\frac{dg_p^i(g_k^i)}{dg_k^i} + 1\right) = 0, \quad i = 1, 2. \tag{11}$$

Substituting (9) into (3), we obtain

$$G = \frac{1 + \alpha - r - \alpha s}{1 + \alpha + r + \alpha s} Y,$$

where $Y = \sum_i = 1^2 \left(y_p^i + y_k^i\right) .c_p^i$ is obtained by substituting (9) and (10) into (4):

$$c_p^i = \frac{r}{1 + \alpha + r + \alpha s} Y, \quad i = 1, 2.$$

c_k^i is obtained by substituting (10) into (5):

$$c_k^i = \frac{\alpha s}{1 + \alpha + r + \alpha s} Y, \quad i = 1, 2.$$

Using (1), finally, a in the payoff matrix (6) is:

[2]Equation (11) are constantly $\partial U_k^i / \partial g_k^i = 0$ because $d\pi^i(g_k^i)/dg_k^i = 1$ and $dg_p^i(g_k^i)/dg_k^i = -1$ for $i = 1, 2,$ by (9) and (10). Then, (11) holds for any value $g_k^i > 0$. Then indeterminacy occurs for transfers and contributions to public goods.

$$a = U_p^i = \ln\left(\frac{r}{1+\alpha+r+\alpha s}Y\right)^r \left(\frac{1+\alpha-r-\alpha s}{1+\alpha+r+\alpha s}Y\right)^{1-r}$$
$$+ \alpha \ln\left(\frac{\alpha s}{1+\alpha+r+\alpha s}Y\right)^s \left(\frac{1+\alpha-r-\alpha s}{1+\alpha+r+\alpha s}Y\right)^{1-s}. \tag{12}$$

3.2 Both Families are Pre-committed

If two pre-committed type families are drawn from the population, they both have utility d in the payoff matrix (6). We compute the value of d as previous subsection:

$$d = \ln\left(\frac{r}{1+\alpha+r}Y\right)^r \left(\frac{(1+\alpha-r)(1-s)}{(1+\alpha+r)(1+s)}Y\right)^{1-r}$$
$$+ \alpha \ln\left(\frac{(1+\alpha-r)s}{(1+\alpha+r)(1+s)}Y\right)^s \left(\frac{(1+\alpha-r)(1-s)}{(1+\alpha+r)(1+s)}Y\right)^{1-s}. \tag{13}$$

3.3 One Ex-post Family and One Pre-committed Family

If one ex-post family and one pre-committed family are drawn from the population, the ex-post family gains utility b and the pre-committed type family gains utility c in the payoff matrix (6). We compute the values of b and c in this subsection. Without loss of generality, we consider that family 1 as the ex-post type family and family 2 as the pre-committed type family. Using backward induction, the parent of family 1 and the child of family 2 maximize their utility at the second stage. Assuming interior solutions, the first-order conditions are the same as (7) and (8) for $i = 1$ and (11) for $i = 2$. Solving these for g_p^1, π^1 and g_k^2, we obtain

$$g_p^1 = \frac{(1-s)(r+\alpha s)(g_p^2+\pi^2+y_k^2) - (1+\alpha-r-\alpha s)(y_p^1+y_k^1)}{-1-\alpha+rs+\alpha s^2} - g_k^1, \tag{14}$$
$$\pi^1 = \frac{-\alpha s(1-s)(g_p^2+\pi^2+y_p^1+y_k^2) + (1+\alpha-rs-\alpha s)y_k^1}{-1-\alpha+rs+\alpha s^2} + g_k^1, \tag{15}$$
$$g_k^2 = \frac{-(1+\alpha)(1-s)(\pi^2+y_k^2) + s(1+\alpha-r-\alpha s)(g_p^2+y_p^1+y_k^1)}{-1-\alpha+rs+\alpha s^2}. \tag{16}$$

At the first stage, given (14)–(16), the child of family 1 and the parent of family 2 maximize their utility. Assuming interior solutions, the first-order conditions are

$$\frac{\partial U_k^1}{\partial g_k^1} = \frac{s}{y_k^1 + \pi^1(g_k^1, g_p^2, \pi^2) - g_k^1}\left(\frac{\partial \pi^1(g_k^1, g_p^2, \pi^2)}{\partial g_k^1} - 1\right)$$
$$+ \frac{1-s}{g_k^1 + g_p^1(g_k^1, g_p^2, \pi^2) + g_p^2 + g_k^2(g_k^1, g_p^2, \pi^2)}$$
$$\times\left(\frac{\partial g_p^1(g_k^1, g_p^2, \pi^2)}{\partial g_k^1} + \frac{\partial g_k^2(g_k^1, g_p^2, \pi^2)}{\partial g_k^1} + 1\right) = 0, \tag{17}$$

$$\frac{\partial U_p^2}{\partial g_p^2} = \frac{-r}{y_p^2 - \pi^2 - g_p^2} + \frac{\alpha s}{y_k^2 + \pi^2 - g_k^2}\left(-\frac{\partial g_k^2(g_k^1, g_p^2, \pi^2)}{\partial g_p^2}\right)$$
$$+ \frac{\alpha(1-s) + 1 - r}{g_k^1 + g_p^1(g_k^1, g_p^2, \pi^2) + g_p^2 + g_k^2(g_k^1, g_p^2, \pi^2)}$$
$$\times\left(1 + \frac{\partial g_p^1(g_k^1, g_p^2, \pi^2)}{\partial g_p^2} + \frac{\partial g_k^2(g_k^1, g_p^2, \pi^2)}{\partial g_p^2}\right) = 0, \tag{18}$$

$$\frac{\partial U_p^2}{\partial \pi^2} = \frac{-r}{y_p^2 - \pi^2 - g_p^2} + \frac{\alpha s}{y_k^2 + \pi^2 - g_k^2}\left(1 - \frac{\partial g_k^2(g_k^1, g_p^2, \pi^2)}{\partial \pi^2}\right)$$
$$+ \frac{\alpha(1-s) + 1 - r}{g_k^1 + g_p^1(g_k^1, g_p^2, \pi^2) + g_p^2 + g_k^2(g_k^1, g_p^2, \pi^2)}$$
$$\times\left(\frac{\partial g_p^1(g_k^1, g_p^2, \pi^2)}{\partial \pi^2} + \frac{\partial g_k^2(g_k^1, g_p^2, \pi^2)}{\partial \pi^2}\right) = 0. \tag{19}$$

For (17), $\partial U_k^i / \partial g_k^i = 0$ holds. For (18) and (19),

$$-\frac{\partial g_k^2(g_k^1, g_p^2, \pi^2)}{\partial g_p^2} = 1 - \frac{\partial g_k^2(g_k^1, g_p^2, \pi^2)}{\partial \pi^2} = \frac{s(1+\alpha-r-\alpha s)}{1+\alpha-rs-\alpha s^2}$$

and

$$1 + \frac{\partial g_p^1(g_k^1, g_p^2, \pi^2)}{\partial g_p^2} + \frac{\partial g_k^2(g_k^1, g_p^2, \pi^2)}{\partial g_p^2}$$
$$= \frac{\partial g_p^1(g_k^1, g_p^2, \pi^2)}{\partial \pi^2} + \frac{\partial g_k^2(g_k^1, g_p^2, \pi^2)}{\partial \pi^2} = -\frac{1-s}{1+\alpha-r-\alpha s}$$

hold, such that, (18) and (19) can be combined to produce a single equation as follows:

$$\frac{-r}{y_p^2-\pi^2-g_p^2}+\frac{\alpha s^2(1+\alpha-r-\alpha s)}{(y_k^2+\pi^2-g_k^2)(1+\alpha-rs-\alpha s^2)}$$
$$-\frac{1+\alpha-rs-\alpha s^2}{(g_p^2+\pi^2+y_k^2+y_p^1+y_k^1)(1+\alpha-r-\alpha s)}=0. \tag{20}$$

Substituting (16) into (20) and solving for $\pi^2+g_p^2$,

$$g_p^2+\pi^2=\frac{-r(y_k^2+y_p^1+y_k^1)+(1+\alpha-r)y_p^2}{1+\alpha}, \tag{21}$$

Substituting (14), (16) and (21) into (3) provides,

$$G=\frac{(1+\alpha-r)(1+\alpha-r-\alpha s)(1-s)}{(1+\alpha)(1+\alpha-rs-\alpha s^2)}Y.$$

We can solve this as follows:

$$c_p^1=\frac{r(1+\alpha-r)(1-s)}{(1+\alpha)(1+\alpha-rs-\alpha s^2)}Y,$$
$$c_p^2=\frac{r}{1+\alpha}Y.$$

Therefore, the payoffs b and c in the payoff matrix (6) are respectively given by,

$$b=\ln\left(\frac{r(1+\alpha-r)(1-s)}{(1+\alpha)(1+\alpha-rs-\alpha s^2)}Y\right)^r$$
$$\times\left(\frac{(1+\alpha-r)(1+\alpha-r-\alpha s)(1-s)}{(1+\alpha)(1+\alpha-rs-\alpha s^2)}Y\right)^{1-r}$$
$$+\alpha\ln\left(\frac{\alpha s(1+\alpha-r)(1-s)}{(1+\alpha)(1+\alpha-rs-\alpha s^2)}Y\right)^s$$
$$\times\left(\frac{(1+\alpha-r)(1+\alpha-r-\alpha s)(1-s)}{(1+\alpha)(1+\alpha-rs-\alpha s^2)}Y\right)^{1-s}, \tag{22}$$

$$c=\ln\left(\frac{r}{1+\alpha}Y\right)^r\left(\frac{(1+\alpha-r)(1+\alpha-r-\alpha s)(1-s)}{(1+\alpha)(1+\alpha-rs-\alpha s^2)}Y\right)^{1-r}$$
$$+\alpha\ln\left(\frac{s(1+\alpha-r)(1+\alpha-r-\alpha s)}{(1+\alpha)(1+\alpha-rs-\alpha s^2)}Y\right)^s$$
$$\times\left(\frac{(1+\alpha-r)(1+\alpha-r-\alpha s)(1-s)}{(1+\alpha)(1+\alpha-rs-\alpha s^2)}Y\right)^{1-s}. \tag{23}$$

4 Replicator Dynamics

In this section, we analyze for replicator dynamics which is well known as the basic evolutionary dynamics literature. For replicator dynamics, the change in the population share of players who take a certain strategy is determined by the difference between the average payoff by the strategy and the average payoff of the population. If the average expected payoff of the strategy is larger (smaller) than that of the population, then the share of the strategy of the next generation will be increasing (decreasing). For time period $t = 1, 2, \ldots,$ replicator dynamics is

$$x(t+1) = x(t)\frac{ax(t) + b(1-x(t))}{x(t)\{ax(t) + b(1-x(t))\} + (1-x(t))\{cx(t) + d(1-x(t))\}}. \tag{24}$$

Therefore, the stationary points of this dynamics are in the case of $-a + b + c - d \neq 0$,

$$x^* = 0,\ 1,\ \frac{b-d}{-a+b+c-d}.$$

If $0 < (b-d)/(-a+b+c-d) < 1$, then this dynamics has an interior stationary point. This interior stationary point means that mixed state of both ex-post type families and pre-committed type families are coexisting. We show the conditions that the replicator dynamics (24) has an interior unstable stationary point. For stability, we define some stability notions according to Samuelson [3] as follows.

Stability *The state $x^* \in X$ is a* stationary state *of difference equation $x(t+\Delta t) - x(t) = f(x)$ defined on $X \subset \mathbb{R}^n$ if $f(x^*) = 0$. x^* is* stable *if it is a stationary point with the property that for every neighborhood V of x^*, there exists a neighborhood $U \subset V$ with the property that if an initial state $x_0 \in U$, then the solution of* (24) *$x(x_0, t) \in V$ for all $t > 0$. And x^* is* asymptotically stable *if it is stable and there exists a neighborhood W of x^* such that $x_0 \in W$ implies* $\lim_{t\to\infty} x(x_0, t) = x^*$.

First, we illustrate combinations (r, s) which satisfies $b - d < 0$ and $c - a < 0$ for $\alpha = 0.1, 0.2, \ldots, 0.9$ graphically. In the following nine figures, the horizontal axis indicates r and the vertical axis indicates s. The blue region is plotted for $b - d > 0$ and the red region is set for $c - a > 0$, so the white region indicates $b - d < 0$ and $c - a < 0$. (24) has an interior unstable stationary point $x^* = (b-d)/(-a+b+c-d)$ for (r, s) in the whiteregion. As shown, there is almost no area where $b - d > 0$ and $c - a > 0$ are overlapping. This implies that the interior stationary point of (24) is almost unstable. Although we can not provide a mathematically rigorous proof, we concentrate on the condition which the interior unstable stationary point exists, $b - d < 0$ and $c - a < 0$. Furthermore, by (24), $x^* = 0, 1$ are asymptotically stable.

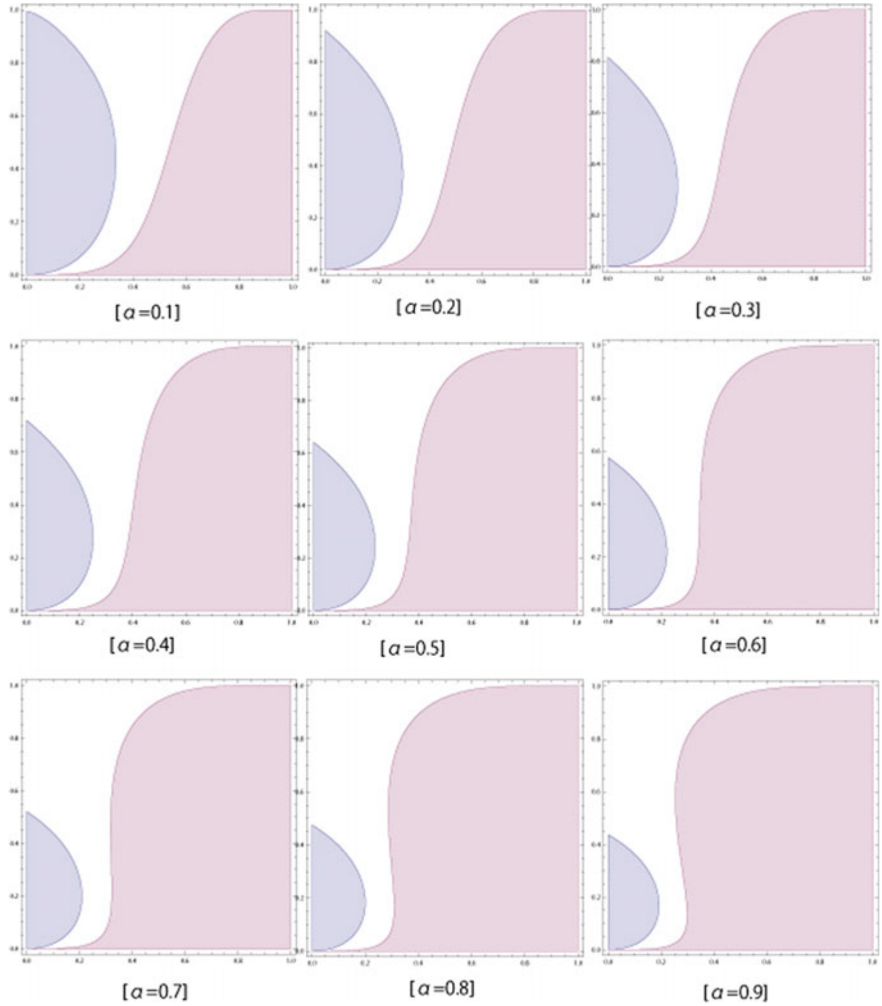


Fig. 1 The region (r, s) for replicator dynamics

These figures show a trend for this dynamics. The region (r, s) satisfying $b - d < 0$ expands, and the region (r, s) satisfying $c - a < 0$ shrinks as α increases.[3] The payoff b is the payoff of ex-post and the payoff d is the payoff of pre-committed if the opponent family takes pre-committed, as we see the payoff matrix (6), and $b - d$ is the payoff difference. $b - d < 0$ implies that if the opponent family takes pre-committed, then pre-committed is better than ex-post. Since the region (r, s) sat-

[3]We calculate the area for $b - d < 0$ and $c - a < 0$, the white region in Fig. 1. By simple calculation, we find the tendency that the blue region $b - d > 0$ shrinks and the red region $c - a > 0$ as α increases.

isfying $b - d < 0$ expands as α increases, this tendency implies that if the parent is more altruistic toward her own child, the region (r, s) satisfying that pre-committed tends to be better than ex-post expands if the opponent is pre-committed. The payoff $a$is the payoff of ex-post and the payoff c is the payoff of pre-committed if the opponent family takes ex-post, as we see the payoff matrix (6), and $c - a$ is the payoff difference. $c - a < 0$ implies that if the opponent family takes ex-post, then ex-post is better than pre-committed. Since the region (r, s) satisfying $c - a < 0$ shrinks as α increases, this tendency implies that if the parent is more altruistic toward her own child, the region (r, s) satisfying that pre-committed tends to be better than ex-post expands if the opponent is pre-committed. That is, if the parent is more altruistic toward her own child, the region (r, s)satisfying that pre-committed tends to be better than ex-post expands.

Second, we give the conditions for (r, s) that analytically satisfies $b - d < 0$ and $c - a < 0$, respectively.

Lemma 1 *If* $\frac{1-s}{2s} < r < 1$ *and* $s_0 < s < \frac{1}{\sqrt{3}}$ *where* $s_0 \in (0, 1)$ *is one real solution for* $s_0^3 + s_0^2 + s_0 - 1 = 0$*, then* $b - d < 0$.

Proof From (13) and (22), we can compute $b - d$ as follows:

$$\begin{aligned} b - d = {} & \alpha s \ln \alpha - (1 + \alpha) \ln(1 + \alpha) + r \ln(1 + \alpha - r) + (1 + \alpha) \ln(1 + \alpha + r) \\ & + (r + \alpha s) \ln(1 - s) + (1 + \alpha - r) \ln(1 + s) \\ & + (1 + \alpha - r - \alpha s) \ln(1 + \alpha - r - \alpha s) \\ & - (1 + \alpha) \ln(1 + \alpha - rs - \alpha s^2), \end{aligned} \tag{25}$$

If the parameter representing the strength of preference for private goods in the utility function of the child, $s = 0$, $b - d = (1 + \alpha)(-2 \ln(1 + \alpha) + \ln(1 + \alpha - r) + \ln(1 + \alpha + r)) < 0$. Differentiating (25) by s yields

$$\begin{aligned} \frac{\partial(b - d)}{\partial s} = & - \alpha + \frac{1 + \alpha - r}{1 + s} - \frac{r + \alpha s}{1 - s} + \frac{(1 + \alpha)(r + 2\alpha s)}{1 + \alpha - rs - \alpha s^2} \\ & + \alpha \ln \alpha + \alpha \ln(1 - s) - \alpha \ln(1 + \alpha - r - \alpha s) \\ = & \frac{1 + \alpha - r - \alpha r - s - \alpha s - rs + 2r^2 s - \alpha s^2 + 3\alpha r s^2 - \alpha s^3}{(1 - s)(1 + s)(1 + \alpha - rs - \alpha s^2)} \\ & + \alpha \ln \frac{\alpha(1 - s)}{1 + \alpha - r - \alpha s}. \end{aligned}$$

Because $\ln \frac{\alpha(1-s)}{1+\alpha-r-\alpha s} < 0$, the sufficient condition for $\partial(b - d)/\partial s$ to be negative is as follows:

$$1 + \alpha - r - \alpha r - s - \alpha s - rs + 2r^2 s - \alpha s^2 + 3\alpha r s^2 - \alpha s^3 < 0. \tag{26}$$

From (26), we obtain

$$\begin{aligned} &1+\alpha-r-\alpha r-s-\alpha s-rs+2r^2s-\alpha s^2+3\alpha rs^2-\alpha s^3 \\ &=\alpha(r(3s^2-1)-(s^3+s^2+s-1))+(1-r)(1-s-2rs). \end{aligned}$$

This is negative if

$$3s^2-1<0, s^3+s^2+s-1>0\text{and } 1-s(2r+1)<0.$$

Let

$$f(s)=s^3+s^2+s-1.$$

$f(s) = 0$ has one real solution $s_0 \in (0, 1)$ satisfying $f(s_0) = 0$:

$$s_0 = \frac{1}{3}\left(-1 - \frac{2}{(17 + 3\sqrt{33})^{\frac{1}{3}}} + (17 + 3\sqrt{33})^{\frac{1}{3}} \right) \approx 0.5437.$$

$f(s) > 0$ for $s_0 < s$. Furthermore, $3s^2 - 1 < 0$ for $0 < s < \frac{1}{\sqrt{3}}$ and $1 - s - 2rs < 0$ for $\frac{1-s}{2s} < r$. That is,

$$s_0 < s < \frac{1}{\sqrt{3}} \quad \text{and} \quad \frac{1-s}{2s} < r < 1. \tag{27}$$

is the sufficient condition for $b - d < 0$. □

Lemma 1 shows the sufficient condition for $b - d < 0$. If $b - d < 0$,the replicator dynamic (24) has no interior stable stationary point regardless of $c - a$. As we see the figures in Fig. 1, there is almost no area where $b - d > 0$ and $c - a > 0$ are overlapping, and the interior stationary point of (24) is almost unstable. Although we can not provide a mathematically rigorous proof, we provide the sufficient condition which the interior unstable stationary point exists, $b - d < 0$ and $c - a < 0$. Lemma 2 shows the sufficient condition for $c - a < 0$. Here, similarly, the replicator dynamic has no interior stable stationary point.

Lemma 2 *If $r < s$ and $r < r_0(s)$ where $r = r_0(s)$ as the value of r that satisfies $\lim_{\alpha \to 0}(c - a) = 0$, then $c - a < 0$ for sufficiently small $\alpha > 0$.*

Proof From (12) and (23), we can derive the following:

$$\begin{aligned} c - a = &-\alpha s \ln \alpha - (1 + \alpha) \ln(1 + \alpha) + (1 + \alpha - r) \ln(1 + \alpha - r) \\ &+ (1 + \alpha - r - \alpha s) \ln(1 - s) + \alpha s \ln(1 + \alpha - r - \alpha s) \\ &+ (1 + \alpha) \ln(1 + \alpha + r + \alpha s) \\ &- (1 + \alpha - r) \ln(1 + \alpha - rs - \alpha s^2). \end{aligned} \tag{28}$$

Differentiating (28) by r yields

$$\frac{\partial(c - a)}{\partial r} = \Omega^{-1} K + \ln \frac{(1 + \alpha - rs - \alpha s^2)}{(1 + \alpha - r)(1 - s)}.$$

where

$$\Omega = (1 + \alpha - r - \alpha s)(1 + \alpha + r + \alpha s)(1 + \alpha - rs - \alpha s^2),$$

and

$$\begin{aligned} K = &-r - 2\alpha r - \alpha^2 r + r^2 + \alpha r^2 + s + \alpha s - \alpha^2 s - \alpha^3 s \\ &- rs - \alpha rs + \alpha rs^2 + \alpha^2 rs^2 + \alpha^2 s^3 + \alpha^3 s^3. \end{aligned}$$

Consider the case where α is sufficiently close to 0. Then,

$$\lim_{\alpha\to 0}(c-a) = (1-r)\ln(1-r) + \ln(1+r) + (1-r)\ln(1-s) - (1-r)\ln(1-rs),$$

$$\lim_{r\to 0}(\lim_{\alpha\to 0}(c-a)) = \ln(1-s) < 0, \quad \lim_{r\to 1}(\lim_{\alpha\to 0}(c-a)) = \ln 2 > 0,$$

and

$$\begin{aligned}\lim_{\alpha\to 0}\frac{\partial(c-a)}{\partial r} &= -1 + \frac{1}{1+r} + \frac{(1-r)s}{1-rs} - \ln(1-r) - \ln(1-s) + \ln(1-rs)\\ &= \frac{s-r}{(1+r)(1-rs)} + \ln\frac{1-rs}{(1-r)(1-s)}.\end{aligned}$$

Here, $\lim_{\alpha\to 0}\frac{\partial(c-a)}{\partial r} > 0$ if $r < s$. Hence, if $r < s$ and α is sufficiently close to 0, then there exists $r = r_0(s)$ as the value of r that satisfies $\lim_{\alpha\to 0}(c-a) = 0$. If $r < r_0(s)$, $\lim_{\alpha\to 0}(c-a) < 0$. For r, the sufficient condition of $c-a<0$ with $\alpha = 0$ is

$$r < s \quad \text{and} \quad r < r_0(s). \tag{29}$$

Thus, there exists $\bar{\alpha} > 0$such that $c-a<0$ under (29) for $0 < \alpha < \bar{\alpha}$. □

Theorem 1 *Assume that* $\frac{1-s}{2s} < r < s$, $r < r_0(s)$ *and* $s_0 < s < \frac{1}{\sqrt{3}}$. *There exists sufficiently small* $\bar{\alpha} > 0$ *such that for* $0 < \alpha < \bar{\alpha}$, *the replicator dynamic (24) has an interior unstable stationary point.*

Proof Because $\frac{1-s}{2s} < s$ for s satisfying (27), by Lemmas 1 and 2, the sufficient condition satisfying (27) and (29) is

$$s_0 < s < \frac{1}{\sqrt{3}} \quad \text{and} \quad \frac{1-s}{2s} < r < \min\{s,\ r_0(s)\}.$$

If $0 < (b-d)/(-a+b+c-d) < 1$, this dynamic has an interior stationary point. If $-a+b+c-d < 0$, $b-d < 0$ and $c-a < 0$ are the conditions for this dynamic to have an interior unstable stationary point:

$$x^* = \frac{b-d}{-a+b+c-d},$$

where $b-d$ is (25) and $c-a$ is (28) for sufficiently small $\alpha > 0$. □

Remark 1 We present an example with parameter values that satisfy the hypothesis of Theorem 1. Let $s = 0.55$, $r = 0.5$, $\alpha = 0.01$. Then, $b-d \approx -0.1660$, $c-a \approx -0.1557$, $\frac{1-s}{2s} \approx 0.4091$, $r_0(s) \approx 0.6434$. For these values, the replicator dynamic (24) has the interior unstable stationary point $x^* \approx 0.5160$.

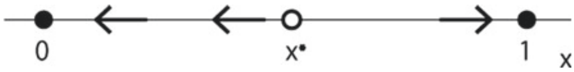


Fig. 2 The best-response dynamics

5 Best-Response Dynamics

In this section, we analyze the best-response dynamics. For best-response dynamics, the change in the population share of players who take a certain strategy is determined by the difference between the average payoff by the strategy and the average payoff of the other strategy's average payoff. If the average expected payoff of the strategy is larger (smaller) than that of the antagonistic strategy, then the share of the strategy of the next generation will be increasing (decreasing). For time period $t = 1, 2, ...,$ best-response dynamics is

$$x(t+1) = \frac{1}{t+1} BR(x(t), 1 - x(t)) + \frac{t}{t+1} x(t). \tag{30}$$

Therefore, the stationary points of this dynamics are in the case of $-a + b + c - d \neq 0$,

$$x^* = 0, 1, \frac{b - d}{-a + b + c - d}.$$

We apply the argument in the previous section to the existence condition of an interior unstable stationary point. As shown in the previous section, this dynamic has an interior unstable stationary point if $b - d < 0$ and $c - a < 0$. The following figure shows this dynamics. According to the figure, $x = 0, 1$ are asymptotically stable stationary points (Fig. 2).

The best-response dynamics in this section contains a instantaneous learning process. Fudenberg and Levine [2] denote "fictitious play" to suppose that players choose their actions to maximize the period's expected payoff given their prediction or assessment of the distribution of the opponent's actions in that period. In our model, the monitoring process corresponds to this notion. If a family obtains the correct information for the strategy distribution of the society, the family can change its strategy to the best-response one instantaneously. Although the replicator dynamics in the previous section does not have such a instantaneous learning process, it has inertia, i.e., partially weak rationality through the comparison to the group's average payoff.

Conclusion

This paper has shown two evolutionary dynamics of a subgame perfect solution for a two-stage game where two families privately contribute to public goods. Each family contains an altruistic parent and a selfish child, and the parent makes a non-negative income transfer to her own child; moreover, they act sequentially. In our model, replicator and best-response dynamics show that in the society, the coexistence of an ex-post transfer-type family and a pre-committed transfer-type family is not stable, whereas a monomorphic state (i.e., all families are ex-post or all families are pre-committed) is stable.[4] If altruism is sufficiently weak, income redistribution does not change the results of both evolutionary game approaches, because they depend on total income Y and not individual income distribution $(y_p^1, y_k^1, y_p^2, y_k^2)$. This means that any income redistribution policies may not affect the results. We believe it is important to analyze policies incorporating the gift tax and the inheritance tax into the model in future research. And it is an open question whether it can be applied in a more or slightly general utility function.

References

1. Cones, R.C., Itaya, J., Tanaka, A.: Private provision of public goods between families. J. Popul. Econ. **25**, 1451–1480 (2012)
2. Fudenberg, D., Levine, D.K.: The Theory of Learning in Games, 2nd edn. MIT Press, Cambridge (1999)
3. Samuelson, L.: Evolutionary Games and Equilibrium Selection, 1st edn. MIT Press, Cambridge (1998)

[4] From the OECD data, many countries have adopted both the gift tax, and the estate and inheritance tax. This implies that both ex-post transfer and pre-committed transfer coexist. OECD data suggest that our model should be modified or it has not reached a stationary point. It is a future challenge.

On the Periodic Behavior of a System of Piecewise Linear Difference Equations

W. Tikjha and E. Lapierre

Abstract In this article we consider the following system of piecewise linear difference equations: $x_{n+1} = |x_n| - y_n - 1$ and $y_{n+1} = x_n + |y_n| - 1$ where the initial condition (x_0, y_0) is an element of $\{(x, 0) : x > \frac{3}{2}\}$ and x_0 is not in a sequence of intervals $B_n = \{x : \frac{2^{2n+1}-1}{2^{2n}} < x \leq \frac{2^{2n+2}-1}{2^{2n+1}}\}$ for any integer n. We show that the solution to the system is eventually one of two particular prime period 4 solutions.

Keywords Difference equation · Periodic solution · Piecewise linear system

1 Introduction

For the convenience of the reader we are including the following definitions [3]. A *system of difference equations of the first order* is a system of the form

$$\begin{cases} x_{n+1} = f(x_n, y_n) \\ y_{n+1} = g(x_n, y_n) \end{cases}, \quad n = 0, 1, ... \tag{1}$$

where f and g are continuous functions which map $\mathbf{R}^2$ into $\mathbf{R}$.

W. Tikjha (✉)
Faculty of Science and Technology, Pibulsongkram Rajabhat University, Phitsanulok, Thailand
e-mail: wirottik@psru.ac.th

W. Tikjha
Centre of Excellence in Mathematics, PERDO, CHE, Bangkok, Thailand

E. Lapierre
Department of Mathematics, Johnson and Wales University, 8 Abbott Park Place, Providence, RI, USA
e-mail: Evelina.Lapierre@jwu.edu

S. Elaydi et al. (eds.), *Advances in Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 212,
DOI 10.1007/978-981-10-6409-8_18

A *solution* of the system of difference equations (1) is a sequence $\{(x_n, y_n)\}_{n=0}^{\infty}$ which satisfies the system for all $n \geq 0$. If we prescribe an *initial condition*

$$(x_0, y_0) \in \mathbf{R}^2$$

then

$$\begin{cases} x_1 = f(x_0, y_0) \\ y_1 = g(x_0, y_0) \end{cases}$$
$$\begin{cases} x_2 = f(x_1, y_1) \\ y_2 = g(x_1, y_1) \end{cases}$$
$$\vdots$$

and so the solution $\{(x_n, y_n)\}_{n=0}^{\infty}$ of the system of difference equations (1) exists for all $n \geq 0$ and is uniquely determined by the initial condition (x_0, y_0).

A solution of the system of difference equations (1) which is constant for all $n \geq 0$ is called an *equilibrium solution*. If

$$(x_n, y_n) = (\bar{x}, \bar{y}) \text{ for all } n \geq 0$$

is an equilibrium solution of the system of difference equations (1), then $(\bar{x}, \bar{y})$ is called an *equilibrium point*, or simply an *equilibrium* of the system of difference equations (1).

A solution $\{(x_n, y_n)\}_{n=0}^{\infty}$ of the system of difference equations (1) is called *periodic with period-p* (or a *period-p solution*) if there exists an integer $p \geq 1$ such that

$$(x_{n+p}, y_{n+p}) = (x_n, y_n) \text{ for all } n \geq 0. \tag{2}$$

We say that the solution is *periodic with prime period-p* if p is the smallest positive integer for which (2) holds. In this case, a p-tuple

$$\begin{pmatrix} x_{n+1}, y_{n+1} \\ x_{n+2}, y_{n+2} \\ x_{n+3}, y_{n+3} \\ \vdots \\ x_{n+p}, y_{n+p} \end{pmatrix}$$

of any p consecutive values of the solution is called a *p-cycle* of the system of difference equations (1).

A solution $\{(x_n, y_n)\}_{n=0}^{\infty}$ of the system of difference equations (1) is called *eventually periodic with period-p* if there exists an integer $N \geq 0$ such that $\{(x_n, y_n)\}_{n=N}^{\infty}$ is periodic with period-p; that is,

$$(x_{n+p}, y_{n+p}) = (x_n, y_n) \text{ for all } n \geq N. \tag{3}$$

Known methods to determine the local asymptotic stability and global stability are not easily applied to piecewise systems. This is why two of the most famous and enigmatic systems of difference equations are piecewise: the Lozi Map and the Gingerbread Man Map. See Ref. [1–3, 5, 6].

In 2008, Gerry Ladas and Ed Grove constructed the following family of systems of piecewise linear difference equations to gain a better understanding of such enigmatic systems

$$\begin{cases} x_{n+1} = |x_n| + ay_n + b \\ \\ y_{n+1} = x_n + c|y_n| + d \end{cases}, \quad n = 0, 1, \ldots \tag{4}$$

where the initial condition $(x_0, y_0) \in R^2$ and the parameters $a, b, c,$ and $d \in \{-1, 0, 1\}$. We believe these 81 systems are prototypes that will help us understand the global behavior of more complicated systems. See Ref. [3, 4, 7].

After discovering the global behavior of most of the systems, we noticed a few trends. Over half of the systems have exactly one equilibrium point, while some have two or three, and the remaining systems either have none or have infinitely many (which usually reside on a line). About a third have periodic solutions. The periodicities are 2, 3, 4, 5, 6, 9, and no other.

We were able to generalize a few systems. That is, we know the global behavior of some systems when one or more parameters are elements of $\mathbf{R}^+$, not just elements of $\{-1, 0, 1\}$ such as in the article [8].

In this paper, we consider a special case of the system above, specifically

$$\begin{cases} x_{n+1} = |x_n| - y_n - 1 \\ \\ y_{n+1} = x_n + |y_n| - 1 \end{cases}, \quad n = 0, 1, \ldots \tag{5}$$

where the initial condition (x_0, y_0) is any element of $\{(x, 0) : x > \frac{3}{2}\}$ and $x_0 \notin B_n = \{x : \frac{2^{2n+1}-1}{2^{2n}} < x \leq \frac{2^{2n+2}-1}{2^{2n+1}}\}$ for any integer n. We show that every solution of System (5) is eventually one of the prime period 4 solutions below:

$$P_{4.1} = \begin{pmatrix} -1, & -1 \\ 1, & -1 \\ 1, & 1 \\ -1, & 1 \end{pmatrix} \quad \text{or} \quad P_{4.2} = \begin{pmatrix} 1, & -3 \\ 3, & 3 \\ -1, & 5 \\ -5, & 3 \end{pmatrix}.$$

2 Main Results

Theorem 1 *Let* $\{(x_n, y_n)\}_{n=0}^{\infty}$ *be a solution of System* (5). *Let* $l = \{(x, 0) : x > \frac{3}{2}\}$ *and* $B_n = \{x : \frac{2^{2n+1}-1}{2^{2n}} < x \leq \frac{2^{2n+2}-1}{2^{2n+1}}\}$. *If* $(x_0, y_0) \in l$ *and* $x_0 \notin B_n$ *for any integer n, then* $\{(x_n, y_n)\}_{n=0}^{\infty}$ *is eventually the prime period 4 solution* $P_{4.1}$ *or* $P_{4.2}$.

The proof of this theorem requires the following two remarks.

Remark 1 Let $\{(x_n, y_n)\}_{n=0}^{\infty}$ be a solution of System (5). If there is a positive integer N such that $y_N = -x_N - 2 \geq 0$, then $\{(x_n, y_n)\}_{n=N+1}^{\infty}$ is the prime period 4 solution $P_{4.2}$.

Proof Suppose that (x_N, y_N) satisfies the hypothesis then

$$\begin{aligned} x_{N+1} &= |x_N| - y_N - 1 = -x_N + x_N + 2 - 1 = 1 \\ y_{N+1} &= x_N + |y_N| - 1 = x_N - x_N - 2 - 1 = -3, \end{aligned}$$

as required. □

Remark 2 Let $\{(x_n, y_n)\}_{n=0}^{\infty}$ be a solution of System (5). If there is a positive integer N such that $y_N = -x_N \geq 0$, then $\{(x_n, y_n)\}_{n=N+1}^{\infty}$ is the prime period 4 solution $P_{4.1}$.

Proof Suppose that (x_N, y_N) satisfies the hypothesis then

$$\begin{aligned} x_{N+1} &= |x_N| - y_N - 1 = -x_N + x_N - 1 = -1 \\ y_{N+1} &= x_N + |y_N| - 1 = x_N - x_N - 1 = -1, \end{aligned}$$

as required. □

We will now begin the proof of the theorem.

Proof Suppose that $(x_0, y_0) \in l$ and $x_0 \notin B_n$ for any integer n. Then

$$\begin{aligned} x_0 &> \tfrac{3}{2} \\ y_0 &= 0 \\[1em] x_1 &= |x_0| - y_0 - 1 = x_0 - 1 > 0 \\ y_1 &= x_0 + |y_0| - 1 = x_0 - 1 > 0 \\[1em] x_2 &= |x_1| - y_1 - 1 = -1 \\ y_2 &= x_1 + |y_1| - 1 = 2x_0 - 3 > 0 \\[1em] x_3 &= |x_2| - y_2 - 1 = -2x_0 + 3 < 0 \\ y_3 &= x_2 + |y_2| - 1 = 2x_0 - 5. \end{aligned}$$

If $x_0 \in \left[\frac{5}{2}, \infty\right)$ then $y_3 = 2x_0 - 5 \geq 0$. We now apply Remark 1 and we have

$$(x_4, y_4) = (1, -3) \in P_{4.2}.$$

If $x_0 \in \left(\frac{3}{2}, \frac{5}{2}\right)$ then we will show that the solution is eventually prime period 4 by mathematical induction.

For each integer $n \geq 1$, let

$$l_n = \frac{2^{2n} - 1}{2^{2n-1}}, \quad u_n = \frac{2^{2n} + 1}{2^{2n-1}}, \quad m_n = \frac{2^{2n+1} - 1}{2^{2n}}, \quad \delta_n = 2^{2n+1} - 1,$$

and let $P(n)$ be the following statement:
For $x_0 \in (l_n, u_n)$,

$$\begin{aligned} x_{4n} &= 1 \\ y_{4n} &= -2^{2n} x_0 + \delta_n. \end{aligned}$$

If $x_0 \in (l_n, m_n]$ then $y_{4n} \geq 0$ and so $(x_{4n+2}, y_{4n+2}) \in P_{4.1}$.
If $x_0 \in (m_n, u_n)$ then $y_{4n} < 0$ and so

$$\begin{aligned} x_{4n+1} &= 2^{2n} x_0 - \delta_n > 0 \\ y_{4n+1} &= 2^{2n} x_0 - \delta_n > 0 \end{aligned}$$

$$\begin{aligned} x_{4n+2} &= -1 \\ y_{4n+2} &= 2^{2n+1} x_0 - 2\delta_n - 1. \end{aligned}$$

If $x_0 \in (m_n, l_{n+1}]$ then $x_0 \in B_n$.
If $x_0 \in (l_{n+1}, u_n)$ then $y_{4n+2} > 0$ and so

$$\begin{aligned} x_{4n+3} &= -2^{2n+1} x_0 + 2\delta_n + 1 < 0 \\ y_{4n+3} &= 2^{2n+1} x_0 - 2\delta_n - 3. \end{aligned}$$

If $x_0 \in [u_{n+1}, u_n)$ then $y_{4n+3} \geq 0$ and so $(x_{4n+4}, y_{4n+4}) \in P_{4.2}$.
If $x_0 \in (l_{n+1}, u_{n+1})$ then $y_{4n+3} < 0$.

We shall first show that $P(1)$ is true.
For $x_0 \in (l_1, u_1) = \left(\frac{3}{2}, \frac{5}{2}\right)$ and $x_3 = -2x_0 + 3 < 0$, $y_3 = 2x_0 - 5 < 0$, we have

$$\begin{aligned} x_{4(1)} &= x_4 = |x_3| - y_3 - 1 = 1 \\ y_{4(1)} &= y_4 = x_3 + |y_3| - 1 = -4x_0 + 7 = -2^{2(1)} x_0 + \delta_1. \end{aligned}$$

If $x_0 \in (l_1, m_1] = \left(\frac{3}{2}, \frac{7}{4}\right]$ then $y_{4(1)} = -4x_0 + 7 \geq 0$ and

$$x_{4(1)+1} = x_5 = |x_4| - y_4 - 1 = 4x_0 - 7 \leq 0$$
$$y_{4(1)+1} = y_5 = x_4 + |y_4| - 1 = -4x_0 + 7 \geq 0.$$

We apply Remark 2, and see that $(x_6, y_6) = (-1, -1) \in P_{4.1}$.

If $x_0 \in (m_1, u_1) = \left(\frac{7}{4}, \frac{5}{2}\right)$ then $y_{4(1)} = -4x_0 + 7 < 0$ and so

$$x_{4(1)+1} = x_5 = |x_4| - y_4 - 1 = 4x_0 - 7 = 2^{2(1)}x_0 - \delta_1 > 0$$
$$y_{4(1)+1} = y_5 = x_4 + |y_4| - 1 = 4x_0 - 7 = 2^{2(1)}x_0 - \delta_1 > 0$$

$$x_{4(1)+2} = x_6 = |x_5| - y_5 - 1 = -1$$
$$y_{4(1)+2} = y_6 = x_5 + |y_5| - 1 = 8x_0 - 15 = 2^{2(1)+1}x_0 - 2\delta_1 - 1.$$

If $x_0 \in (m_1, l_{1+1}] = \left(\frac{7}{4}, \frac{15}{8}\right]$ then $x_0 \in B_1$.
If $x_0 \in (l_{1+1}, u_1) = \left(\frac{15}{8}, \frac{5}{2}\right)$ then $y_{4(1)+2} = 8x_0 - 15 > 0$ and so

$$x_{4(1)+3} = x_7 = |x_6| - y_6 - 1 = -8x_0 + 15 = -2^{2(1)+1}x_0 + 2\delta_1 + 1 < 0$$
$$y_{4(1)+3} = y_7 = x_6 + |y_6| - 1 = 8x_0 - 17 = 2^{2(1)+1}x_0 - 2\delta_1 - 3.$$

If $x_0 \in [u_{1+1}, u_1) = \left[\frac{17}{8}, \frac{5}{2}\right)$ then $y_{4(1)+3} = 8x_0 - 17 \geq 0$. We apply Remark 1 and see that $(x_{4(1)+4}, y_{4(1)+4}) = (1, -3) \in P_{4.2}$.
If $x_0 \in (l_{1+1}, u_{1+1}) = \left(\frac{15}{8}, \frac{17}{8}\right)$ then $y_{4(1)+3} = 8x_0 - 17 < 0$.
Therefore $P(1)$ is true, as required.

Suppose $P(k)$ is true for any positive integer k. If $x_0 \in (l_{k+1}, u_{k+1}) = \left(\frac{2^{2k+2}-1}{2^{2k+1}}, \frac{2^{2k+2}+1}{2^{2k+1}}\right)$ then

$$x_{4k+3} = -2^{2k+1}x_0 + 2\delta_k + 1 < 0$$
$$y_{4k+3} = 2^{2k+1}x_0 - 2\delta_k - 3 < 0.$$

So we have

$$x_{4(k+1)} = x_{4k+4} = |x_{4k+3}| - y_{4k+3} - 1 = 1$$
$$y_{4(k+1)} = y_{4k+4} = x_{4k+3} + |y_{4k+3}| - 1 = -2^{2k+2}x_0 + 4\delta_k + 3 = -2^{2(k+1)}x_0 + \delta_{k+1}.$$

If $x_0 \in (l_{k+1}, m_{k+1}] = \left(\frac{2^{2k+2}-1}{2^{2k+1}}, \frac{2^{2k+3}-1}{2^{2k+2}}\right]$ then $y_{4(k+1)} = -2^{2(k+1)}x_0 + \delta_{k+1} \geq 0$, and so

$$x_{4(k+1)+1} = x_{4k+5} = |x_{4k+4}| - y_{4k+4} - 1 = 2^{2(k+1)}x_0 - \delta_{k+1} \leq 0$$
$$y_{4(k+1)+1} = y_{4k+5} = x_{4k+4} + |y_{4k+4}| - 1 = -2^{2(k+1)}x_0 + \delta_{k+1} \geq 0.$$

We apply Remark 2, and see that $(x_{4k+6}, y_{4k+6}) = (-1, -1) \in P_{4.1}$.

If $x_0 \in (m_{k+1}, u_{k+1}) = \left(\frac{2^{2k+3}-1}{2^{2k+2}}, \frac{2^{2k+2}+1}{2^{2k+1}}\right)$ then $y_{4(k+1)} = -2^{2(k+1)}x_0 + \delta_{k+1} < 0$, and so

$$\begin{aligned} x_{4(k+1)+1} &= x_{4k+5} = |x_{4k+4}| - y_{4k+4} - 1 = 2^{2(k+1)}x_0 - \delta_{k+1} > 0 \\ y_{4(k+1)+1} &= y_{4k+5} = x_{4k+4} + |y_{4k+4}| - 1 = 2^{2(k+1)}x_0 - \delta_{k+1} > 0 \end{aligned}$$

$$\begin{aligned} x_{4(k+1)+2} &= x_{4k+6} = |x_{4k+5}| - y_{4k+5} - 1 = -1 \\ y_{4(k+1)+2} &= y_{4k+6} = x_{4k+5} + |y_{4k+5}| - 1 = 2^{2(k+1)+1}x_0 - 2\delta_{k+1} - 1. \end{aligned}$$

If $x_0 \in (m_{k+1}, l_{k+2}] = \left(\frac{2^{2k+3}-1}{2^{2k+2}}, \frac{2^{2k+4}-1}{2^{2k+3}}\right]$ then $x_0 \in B_{k+1}$.

If $x_0 \in (l_{k+2}, u_{k+1}) = \left(\frac{2^{2k+4}-1}{2^{2k+3}}, \frac{2^{2k+2}+1}{2^{2k+1}}\right)$ then
$y_{4(k+1)+2} = 2^{2(k+1)+1}x_0 - 2\delta_{k+1} - 1 > 0$, and so

$$\begin{aligned} x_{4(k+1)+3} &= x_{4k+7} = |x_{4k+6}| - y_{4k+6} - 1 = -2^{2(k+1)+1}x_0 + 2\delta_{k+1} + 1 < 0 \\ y_{4(k+1)+3} &= y_{4k+7} = x_{4k+6} + |y_{4k+6}| - 1 = 2^{2(k+1)+1}x_0 - 2\delta_{k+1} - 3. \end{aligned}$$

If $x_0 \in [u_{k+2}, u_{k+1}) = \left[\frac{2^{2k+4}+1}{2^{2k+3}}, \frac{2^{2k+2}+1}{2^{2k+1}}\right)$ then
$y_{4(k+1)+3} = 2^{2(k+1)+1}x_0 - 2\delta_{k+1} - 3 \geq 0$. We apply Remark 1, and see that $(x_{4(k+1)+4}, y_{4(k+1)+4}) = (1, -3) \in P_{4.2}$.

If $x_0 \in (l_{k+2}, u_{k+2}) = \left(\frac{2^{2k+4}-1}{2^{2k+3}}, \frac{2^{2k+4}+1}{2^{2k+3}}\right)$ then
$y_{4(k+1)+3} = 2^{2(k+1)+1}x_0 - 2\delta_{k+1} - 3 < 0$.

Therefore $P(k+1)$ is true. By mathematical induction $P(n)$ is true for any positive integer n.

We note that

$$\lim_{n\to\infty} l_n = \lim_{n\to\infty} u_n = \lim_{n\to\infty} m_n = 2.$$

When $(x_0, y_0) = (2, 0)$ we have $(x_1, y_1) = (1, 1) \in P_{4.1}$. Hence we can conclude that the solution to System (5) is eventually one of the two prime period 4 solutions $P_{4.1}$ or $P_{4.2}$ for every initial condition (x_0, y_0) such that $(x_0, y_0) \in \{(x, 0) : x > \frac{3}{2}\}$ and $x_0 \notin B_n$. □

3 Discussion and Conclusion

In this paper we showed that for any initial condition on a specific region of the positive x-axis the solution of System (2) will be one of two specific prime period 4 solutions. We would like to share our conjecture for the global behavior of this system.

Set

$$P_{3.1} = \begin{pmatrix} -\frac{1}{3}, & -1 \\ \frac{1}{3}, & -\frac{1}{3} \\ -\frac{1}{3}, & -\frac{1}{3} \end{pmatrix}, \quad P_{3.2} = \begin{pmatrix} \frac{3}{5}, & \frac{1}{5} \\ -\frac{3}{5}, & -\frac{1}{5} \\ -\frac{1}{5}, & -\frac{7}{5} \end{pmatrix}$$

$$P_{4.1} = \begin{pmatrix} -1, & -1 \\ 1, & -1 \\ 1, & 1 \\ -1, & 1 \end{pmatrix}, \quad \text{and} \quad P_{4.2} = \begin{pmatrix} 1, & -3 \\ 3, & 3 \\ -1, & 5 \\ -5, & 3 \end{pmatrix}.$$

Conjecture *Let* $\{(x_n, y_n)\}_{n=0}^{\infty}$ *be a solution of System (2) with* $(x_0, y_0) \in R^2$. *Then* $\{(x_n, y_n)\}_{n=0}^{\infty}$ *is the unique equilibrium* $\left(-\frac{1}{5}, -\frac{3}{5}\right)$, *or eventually the prime period 3 solution* $P_{3.1}$ *or* $P_{3.2}$, *or the prime period 4 solution* $P_{4.1}$ *or* $P_{4.2}$.

Acknowledgements This work was supported by the Thailand Research Fund [MRG5980053], National Research Council of Thailand and Pibulsongkram Rajabhat University. The first author is supported by the Centre of Excellence in Mathematics, CHE, Thailand.

References

1. Barnsley, M.F., Devaney, R.L., Mandelbrot, B.B., Peitgen, H.O., Saupe, D., Voss, R.F.: The Science of Fractal Images. Springer, New York (1991)
2. Devaney, R.L.: A piecewise linear model of the the zones of instability of an area-preserving map. Phys. D **10D**, 387–393 (1984)
3. Grove, E.A., Ladas, G.: Periodicities in Nonlinear Difference Equations. Chapman Hall/CRC Press, New York (2005)
4. Grove, E.A., Lapierre, E., Tikjha, W.: On the global behavior of $x_{n+1} = |x_n| - y_n - 1$ and $y_{n+1} = x_n + |y_n|$. Cubo. **14**, 125–166 (2012)
5. Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmos. Sci. **20**, 130–141 (1963)
6. Lozi, R.: Un attracteur etrange du type attracteur de Henon. J. Phys. **39**, 9–10 (1978)
7. Tikjha, W., Lapierre, E.G., Lenbury, Y.: On the global character of the system of piecewise linear difference equations $x_{n+1} = |x_n| - y_n - 1$ and $y_{n+1} = x_n - |y_n|$. Adv. Differ. Equ. (2010). doi:10.1155/2010/573281
8. Tikjha, W., Lapierre, E., Sitthiwirattham, T.: The stable equilibrium of a system of piecewise linear difference equations. Adv. Differ. Equ. **2017**, 67 (2017). doi:10.1186/s13662-017-1117-2