Springer Proceedings in Mathematics & Statistics

Ziyad AlSharawi
Jim M. Cushing
Saber Elaydi *Editors*

Theory and Applications of Difference Equations and Discrete Dynamical Systems

ICDEA, Sultan Qaboos University, Muscat, Oman, May 26–30, 2013

Springer

Springer Proceedings in Mathematics & Statistics

Volume 102

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 OR 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

Ziyad AlSharawi · Jim M. Cushing
Saber Elaydi
Editors

Theory and Applications of Difference Equations and Discrete Dynamical Systems

ICDEA, Sultan Qaboos University, Muscat, Oman, May 26–30, 2013

Springer

Editors
Ziyad AlSharawi
Department of Mathematics and Statistics
Sultan Qaboos University
Muscat
Oman

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

Jim M. Cushing
Department of Mathematics
University of Arizona
Tucson, AZ
USA

ISSN 2194-1009 ISSN 2194-1017 (electronic)
ISBN 978-3-662-44139-8 ISBN 978-3-662-44140-4 (eBook)
DOI 10.1007/978-3-662-44140-4

Library of Congress Control Number: 2014947681

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

Springer Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The International Conference on Difference Equations and Applications (ICDEA) is held annually under the auspices of the International Society of Difference Equations. At these conferences, leading mathematicians from around the world assemble to discuss and present their research in the theory, analysis, and applications of difference equations, discrete time dynamical systems, and related disciplines. The 19th annual conference (ICDEA 2013) was held at Sultan Qaboos University, Muscat, Oman, May 26–30, 2013.

At ICDEA 2013, ten invited speakers delivered plenary talks on various topics of discrete dynamical systems and difference equations and their applications to natural sciences. The opening plenary address was given by Professor Michal Misiurewicz who was awarded, at the conference, the second biannual Aulbach Prize by the International Society of Difference Equations in recognition of his significant contributions to difference equations and discrete dynamical systems. Included among the more than 60 talks presented at the conference were those in four special sessions on Applications of Dynamical Systems with Delays, Difference Equation Applications in the Biological Sciences, Continuous Dynamical Systems, and Topological Dynamics.

These proceedings contain articles written by participants at ICDEA 2013 and were selected by our panel of referees to ensure quality of scientific content. Four of the articles are survey papers prepared by the plenary speakers Azmy Ackleh, Arno Berger, Eduardo Liz, and Hinke Osinga.

Our gratitude and appreciation go to the organizers for their efforts that made possible the success of the conference; the members of the scientific committee who ensured the high standards of the conference's scientific activities; the administration of Sultan Qaboos University for providing its facilities and resources to conference participants; and last but not least, the sponsors for their generous financial contributions.

Ziyad AlSharawi
Jim M. Cushing
Saber Elaydi

ICDEA 2013 Sponsors

Conference Organizing and Scientific Committees

Organizing Committee

Asma Al-Ghassani, Sultan Qaboos University, Muscat, Oman
Nasser Al-Salti, Sultan Qaboos University, Muscat, Oman
Ziyad AlSharawi, Sultan Qaboos University, Muscat, Oman
Mohamed Rhouma, Qatar University, Doha, Qatar

Scientific Committee

Linda Allen, Texas Tech University, USA
Luis Alseda, Universitat Autonoma de Barcelona, Spain
Ziyad AlSharawi, Sultan Qaboos University, Oman
Saber Elaydi, Trinity University, USA
Yoshihiro Hamaya, Okayama University of Science, Japan
Allan Peterson, University of Nebraska, USA
Christian Pötzsche, Alpen-Adria-Universität Klagenfurt, Austria
George Sell, University of Minnesota, USA
Lubomir Snoha, Slovakia Matej Bel University, Slovakia

Contents

Speakers at ICDEA 2013

Syed Abbas
Indian Institute of Technology Mandi
School of Basic Sciences
Mandi, H.P.
175001, India
sabbas.iitk@iitmandi.ac.in

Duaa H. Abdelrahman
United Arab Emirates University
Department of Mathematical
Sciences
P.O. Box 15551
Al Ain, UAE
duaa-abdelrahman@uaeu.ac.ae

Azmy S. Ackleh
Department of Mathematics
University of Louisiana at Lafayette
Lafayette, Louisiana 70504-1010
USA
ackleh@louisiana.edu

Afaq Ahmad
Sultan Qaboos University
Electrical and Computer Engineering
P.O. Box 33, PC 123
Al Khod, Muscat, Oman
afaq@squ.edu.om

Mehiddin Al-Baali
Sultan Qaboos University
Department of Mathematics
and Statistics
P.O. Box 36, PC 123
Al Khod, Muscat, Oman
albaali@squ.edu.om

Muna A. Alhalawa
Technical University of Lisbon
Department of Mathematics
Av. Rovisco Pais, 1049-001 Lisboa
Portugal
muna.alhalawa@ist.utl.pt

Aija Anisimova
University of Latvia
Department of Mathematics
Latvia, Riga
Zellu 8, LV-1002
aija-anisimova@inbox.lv

M. Naim Anwar
United Arab Emirates University
Department of Mathematical Sciences
P.O. Box 15551
Al Ain, UAE
m.anwar@uaeu.ac.ae

Jaromír Baštinec
Brno University of Technology
Department of Mathematics
Technická 10, 616 00 Brno
Czech Republic
bastinec@feec.vutbr.cz

Arno Berger
University of Alberta
Mathematical and Statistical Sciences
Edmonton, Alberta T6G 2G1
Canada
berger@ualberta.ca

Fatema Berrabah
Djillali Liabès University
Department of Mathematics
P.O. Box 89
Sidi Bel Abbes, Algeria
berrabah_f@yahoo.fr

Henk Bruin
University of Vienna
Faculty of Mathematics
Nordbergstraße 15
1090 Vienna, Austria
henk.bruin@univie.ac.at

Jose S. Cánovas
Technical University of Cartagena
Applied Mathematics and Statistics
Cartagena, Murcia, Spain
jose.canovas@upct.es

Jim M. Cushing
University of Arizona
Department of Mathematics
617 N. Santa Rita
Tucson, Arizona 85721, USA
cushing@math.arizona.edu

Josef Diblík
Brno University of Technology
Department of Mathematics
Technická 10, 616 00 Brno
Czech Republic
diblik.j@fce.vutbr.cz

Yiming Ding
Chinese Academy
of Sciences
Wuhan Institute of Physics
and Mathematics
P.O. Box 71010
Wuhan 430071, China
ding@wipm.ac.cn

Matúš Dirbák
Matej Bel University
Department of Mathematics
Tajovského 40
Banská Bystrica, Slovakia
matus.dirbak@umb.sk

Zuzana Došlá
Masaryk University
Department of Mathematics
and Statistics
Kotlářská 2, CZ-61137
Brno, Czech Republic
dosla@math.muni.cz

Ondřej Došlý
Masaryk University Brno
Department of Mathematics
and Statistics
Kotlářská 2, CZ-611 37
Czech Republic
dosly@math.muni.cz

Saber Elaydi
Trinity University
Department of Mathematics
San Antonio, Texas 78212
USA
selaydi@trinity.edu

Julia V. Elyseeva
Moscow State University
of Technology
Department of Applied
Mathematics
Vadkovskii per. 3a, 101472
Moscow, Russia
elyseeva@mtu-net.ru

Selenne Garcia-Torres
University of Southern California
Department of Mathematics
Los Angeles, CA 90089
USA
garciato@usc.edu

Ahmad Ghaleb
Cairo University
Department of Mathematics
Giza, 12613
Egypt
afghaleb@sci.cu.edu.e.g

István Györi
University of Pannonia
Department of Mathematics
P.O. Box 158
8201 Veszprém, Hungary
gyori@almos.uni-pannon.hu

Rod Halburd
University College London
Department of Mathematics
Gower Street
London, WC1E 6BT, UK
R.Halburd@ucl.ac.uk

Yoshihiro Hamaya
Okayama University of Science
Department of Information Science
1-1 Ridai-chyo, Kita-ku
Okayama, Japan
P.O. Box 700-0005
hamaya@mis.ous.ac.jp

Petr Hasil
Mendel University in Brno
Department of Mathematics
Zemědělská 1, 613 00 Brno
Czech Republic
hasil@mendelu.cz

Roman Šimon Hilscher
Masaryk University
Department of Mathematics
and Statistics
Kotlářská 2, CZ-61137 Brno
Czech Republic
hilscher@math.muni.cz

Syeda D. Jabeen
Indian Statistical Institute
203 B. T. Road, PC 700108
Kolkata, India
syed-sdj@yahoo.com

Sebti Kerbal
Sultan Qaboos University
Department of Mathematics
and Statistics
P.O. Box 36, PC 123
Al Khod, Muscat, Oman
skerbal@squ.edu.om

Bernd Krauskopf
The University of Auckland
Department of Mathematics
Private Bag 92019
Auckland 1142, New Zealand
B.Krauskopf@auckland.ac.nz

Sanjeev Kumar
Institute of Basic Science
Department of Mathematics
Khandari, Agra-282002
India
sanjeevibs@yahoo.co.in

S. Lakshmanan
United Arab Emirates University
Department of Mathematical Sciences
P.O. Box 15551
Al Ain, UAE
frihan@uaeu.ac.ae

Eduardo Liz
University of Vigo
Marcosende Campus
Department of Applied Mathematics II
36310 Vigo, Spain
eliz@dma.uvigo.es

Shamil Makhmutov
Sultan Qaboos University
Department of Mathematics and Statistics
P.O. Box 36, PC 123
Al Khod, Muscat, Oman
makhm@squ.edu.om

Michał Misiurewicz
Purdue University Indianapolis
Department of Mathematical Sciences
Indianapolis, IN, USA
mmisiure@math.iupui.edu

Haniffa M. Nasir
University of Peradeniya
Department of Mathematics
Sri Lanka
nasirh@squ.edu.om

Piotr Oprocha
AGH University of Science and Technology
Faculty of Applied Mathematic
al. A. Mickiewicza 30, 30-059 Kraków
Poland
oprocha@agh.edu.pl

Hinke M. Osinga
The University of Auckland
Department of Mathematics
Private Bag 92019
Auckland 1142, New Zealand
h.m.osinga@auckland.ac.nz

Mihály Pituk
University of Pannonia
Department of Mathematics
P.O. Box 158
8201 Veszprém, Hungary
pitukm@almos.uni-pannon.hu

Michal Pospíšil
Electrical Engineering and Communication
Brno University of Technology
Technická 3058/10
616 00 Brno, Czech Republic
pospisilm@feec.vutbr.cz

Zdeněk Pospíšil
Masaryk University
Department of Mathematics and Statistics
611 37 Kotlářská 2
Brno, Czech Republic
pospisil@math.muni.cz

Anton Purnama
Sultan Qaboos University
Department of Mathematics and Statistics
P.O. Box 36, PC 123
Al Khod, Muscat, Oman
antonp@squ.edu.om

Martin Rasmussen
Imperial College London
Department of Mathematics
London SW7 2AZ
m.rasmussen@imperial.ac.uk

Andrejs Reinfelds
University of Latvia
Institute of Mathematics
Raiņa bulvāris 29
Rīga LV-1459, Latvia
reinf@latnet.lv

Szilárd Gy. Révész
Kuwait University
Department of Mathematics
P.O. Box 5969
Safat 13060 Kuwait
szilard@sci.kuniv.edu.kw

Mohamed B.H. Rhouma
Qatar University
Department of Mathematics
P.O. Box 2713
Doha, Qatar
rhouma@qu.edu.qa

Fathalla A. Rihan
United Arab Emirates University
Department of Mathematical
Sciences
P.O. Box 15551
Al Ain, UAE
frihan@uaeu.ac.ae

Alfonso Ruiz-Herrera
University of Granada
Department of Applied Mathematics
Granada, Spain
alfonsoruiz@ugr.es

Samir H. Saker
Mansoura University
Department of Mathematics
Mansoura 35516
Egypt
shsaker@mans.edu.e.g.

César M. Silva
Universidade da Beira Interior
Departamento de Mathematica
Rua Marquês d'Ávila e Bolama
6201-001 Covilhã, Portugal
csilva@ubi.pt

Jaroslav Smítal
Mathematics Institute
Silesian University
CZ-746 01 Opava, Czech Republic
jaroslav.smital@math.slu.cz

Nasser H. Sweilam
Cairo University
Department of Mathematics
Giza, 12613
Egypt
nsweilam@sci.cu.edu.e.g.

Yasuhiro Takeuchi
Aoyama Gakuin University
College of Science and Engineering
Fuchinobe, Chuo-Ku, Sagamihara-shi
Kanagawa 252-5258, Japan
takeuchi@gem.aoyama.ac.jp

Abdullahi Umar
Sultan Qaboos University
Department of Mathematics
and Statistics
P.O. Box 36, PC 123
Al Khod, Muscat, Oman
aumarh@squ.edu.om

Liudmila A. Uvarova
Moscow State University
of Technology
Department of Applied Mathematics
Vadkovskii per. 3a, 101472
Moscow, Russia
Uvarova_la@rambler.ru

Michal Veselý
Masaryk University
Department of Mathematics
and Statistics
Kotlářská 2, CZ-611 37
Brno, Czech Republic
michal.vesely@mail.muni.cz

Jiří Vítovec
University of Technology Brno
Institute of Mathematics
Technická 10, 616 00 Brno
Czech Republic
vitovec@feec.vutbr.cz

Petr Zemánek
Masaryk University, Faculty of Science
Department of Mathematics
and Statistics
Kotlářská 2, CZ-61137
Brno, Czech Republic
zemanekp@math.muni.cz

Part I
Papers by Invited Speakers

Competitive Exclusion Through Discrete Time Models

Azmy S. Ackleh and Paul L. Salceanu

Abstract In biology, the principle of competitive exclusion, largely attributed to the Russian biologist G. F. Gause, states that two species competing for common resources (food, territory etc.) cannot coexist, and that one of the species drives the other to extinction. We make a survey of discrete-time mathematical models that address this issue and point out the main mathematical methods used to prove the occurrence of competitive exclusion in these models. We also offer examples of models in which competitive exclusion fails to take place, or at least it is not the only outcome. Finally, we present an extension of the competitive exclusion results in [1, 5] to a more general model.

1 Introduction

An important tenet in mathematical biology is *the principle of competitive exclusion*. According to this tenet any two species that compete for common limited resources cannot coexist in the long term; one of the species will drive the other to extinction. This principle is attributed to the Russian biologist Gause [18] who, through lab experiments, studied many scenarios of interaction for competing species, among which some resulted in competitive exclusion. The principle was first illustrated by the Lotka-Volterra competition model and ever since it has been extensively studied in the literature.

In the field of mathematical biology (especially in ecology and epidemiology) numerous continuous time mathematical models, such as ordinary, partial and delay differential equations, have been dedicated to this idea. However, discrete time

A.S. Ackleh (✉) · P.L. Salceanu
University of Louisiana at Lafayette, Lafayette, LA 70504, USA
e-mail: ackleh@louisiana.edu

P.L. Salceanu
e-mail: salceanu@louisiana.edu

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_1

models have been used much less in the study competitive exclusion. Discrete time models allow for interesting and more general cases of coexistence (such as nontrivial interior attractors); thus contradicting the principle of competitive exclusion.

Cushing et al. [11] considers two discrete models for competition between two species. One model (a Leslie-Gower model), with Beverton-Holt-type nonlinearities (which we will, hereafter, also refer to as *weak nonlinearities*), has dynamics analogous to the Lotka-Volterra competition model. Namely, if the competition between the two species is strong, then every orbit converges either to the extinction equilibrium, or to one of the two boundary equilibria (thus competitive exclusion occurs in the latter case). On the other hand, if the competition is weak, then both species survive at the unique interior equilibrium. The other competition model considered assumes age structure in one of the two species and Ricker-type nonlinearities, and in this model coexistence can occur in the form of an asymptotically stable 2-cycle. *Strong nonlinearities* (involving exponential functions) in the species growth functions could even lead to "mixed-type" attractors, such as a stable coexistence 2-cycle and a stable exclusion equilibrium [12], or even to multiple coexistence attractors, some of which are chaotic, if structure is also added into the model [16, 17].

An interesting idea, exploited by Rael et al. [30], is to investigate if evolution that happens on a fast time scale (commensurate with that of population dynamics) can change the outcome of competition. Namely if, through evolution, it is possible to change the winner between two competing species, or to go from competitive exclusion to coexistence, thus explaining the "anomalies" observed in the lab experiments of Dawson [15], respectively Park et al. [24], involving two species of flour beetle insects (*Tribolium castaneum* and *Tribolium confusum*).

In [4], AlSharawi and Rhouma study an n-species Leslie-Gower model, but assume that competition impacts equally each species. They show that competitive exclusion is the only outcome, with one species surviving (at the equilibrium) and driving the rest of the species to extinction. In order to look for possibilities of coexistence among species, they modify the model by considering stocking and/or harvesting. Inspired by a paper of Ackleh et al. [1], Chow and Hsieh [5] show that competitive exclusion occurs in the n-species Leslie-Gower model from [4], even when the impact of competition is not the same for each species, but assuming that the mortality for species i is proportional to the size of the population. This result for two species ($n = 2$) can also be found in [1]. Ackleh et al. [2] further extend this model so that to allow for selection and mutation. This modification complicates the mathematics, but competitive exclusion still remains possible.

The work of Hsu et al. [21] addresses the issue of competitive exclusion in a more general framework, that of ordered Banach spaces. Under a set of assumptions made on the map defining the dynamics of the system, the authors show that the omega limit set of any interior point is either an interior equilibrium, if such an equilibrium exists, or a boundary equilibrium. One of the assumptions is that the map is strictly order-preserving. This assumption, in most cases, is difficult to verify in applications and general conditions sufficient to imply it are not known to the authors (see Smith [31] for planar systems). Ackleh and Zhang [3] use the results in [21] to analyze competition between two species of iris (a plant that produces sexually and asexually), each species being also structured by stage (seeds, juveniles and adults). The authors show that each species, in the absence of the other, can survive at an equilibrium with all coordinates positive, and that when they compete, the species the species having the largest equilibrium value drives the other species to extinction; provided the intraspecific competition efficiencies of both species are similar. When they consider different intraspecific competition efficiencies between the species they show that coexistence in the form of globally asymptotically stable equilibrium (in the interior of $\mathbb{R}^6_+$) is possible [34]. Cushing et al. [13] also consider a structured model (juveniles and adults) with two competing species, having similar boundary dynamics as in [3]. In their model they also obtain besides competitive exclusion, a globally asymptotically stable (in the interior of $\mathbb{R}^4_+$) interior equilibrium due to allowing for different interspecific competition efficiencies between the juvenile classes.

Smith and Zhao [33] investigate competitive exclusion among multiple microbial populations feeding on the same nutrient and prove that the microbial species with the break-even nutrient concentration wins the competition, driving all the other microbial population to extinction. A particular case of this model, containing only two microbial strains was previously analyzed in [32].

This survey is far from being a comprehensive one and it is mainly organized by considering two classes of competition models, with and without (age/stage) structure. This is because, especially with regard to the first class, there is no "general method" available to approach the issue of competitive exclusion in discrete time models. Another reason is that adding structure into a competition model has the potential of enriching model dynamics by violating the principle of competitive exclusion, thus allowing for (multiple) interior attractors [13]. Thus, our primary goals here are to observe the principle of competitive exclusion "at work" in different types of such models and to point out the main mathematical tools (specific to each model) used to investigate it.

2 Unstructured Models

In unstructured models each species is regarded as an aggregation of identical individuals, without taking into account different characteristics of individuals within the species. The majority of the unstructured competition models in discrete time, that predict competitive exclusion, use a Beverton-Holt-type population growth function, and this class of models is known as *Leslie-Gower*-type models [23]. Thus we begin with a survey of these models. In unstructured Leslie-Gower models with pure selection (individuals of one species produce individuals of the same species only), competitive exclusion is the most probable outcome. The following model, considered in [4], accounts for such a case:

$$x_i(t+1) = \frac{b_i x_i(t)}{1 + \sum_{j=1}^{n} c_j x_j(t)}, \quad i = 1, \ldots, n. \tag{1}$$

In this model, competition affects each species equally [notice that the right-hand side in every equation in (1) has the same denominator]. Then the largest of the "growth coefficients" $b_1, \dots, b_n$ dictates whether all species go extinct, or just one of them survives and all the other die out. More exactly, if $b_1 > b_2 > \cdots > b_n$, then $b_1 < 1$ implies $x_i(t) \to 0$ as $t \to \infty$, for all $i = 1, \dots, n$, and $b_1 > 1$ implies $x_i(t) \to 0$ as $t \to \infty$, for all $i = 2, \dots, n$, and $x_1(t) \to (b_1 - 1)/c_1$. Although, the strict inequalities among b_i's, $i = 2, \dots, n$ (assumed in [4]), could be replaced by weak inequalities, so that the assumption $b_1 > b_2 \geq \cdots \geq b_n$ would produce the same result (see also [2, 5]).

In [1], Ackleh et al. consider a modified version of (1), when the effect of competition is not the same for all species, but the mortality rate for each species is proportional to the sum total of the population sizes of all species (that is, the intraspecific competition is the same as any interspecific competition).

$$x_i(t+1) = \frac{a_i x_i(t)}{1 + b_i \sum_{j=1}^{n} x_j(t)}, \quad i = 1, \dots, n. \tag{2}$$

If

$$\frac{a_1 - 1}{b_1} > \frac{a_i - 1}{b_i} > 0, \quad \text{for all } i = 2, \dots, n, \tag{3}$$

then the boundary equilibrium $E_1 = ((a_1 - 1)/b_1, 0, \dots, 0)$ is asymptotically stable, while all the other nontrivial boundary equilibria $E_i = (0, \dots, (a_i - 1)/b_i, 0, \dots, 0)$ are unstable. This follows directly from the eigenvalues of the Jacobian matrix evaluated at these equilibria, which are easy to calculate. Further, it is shown that, when $n = 2$, E_1 also attracts all solutions with $x_1(0) > 0$. For this, the authors use Bendixson-Dulac criteria for difference equations, but also mention that the same result could be achieved by using the theory of competitive planar systems of difference equations. This result is extended to arbitrary n in [2, 5] (with different proofs). In [5] the proof is tedious and requires quite a few preliminary results (lemmas) in order to prepare it. Ackleh et al. offer a much shorter proof, based on the fact that the set $\{x \in \mathbb{R}_+^n \mid |x| \leq (a_1 - 1)/b_1\}$ is positively invariant and attracts all solutions. This follows by noting that the maximum of $|f(x)|$, where f denotes the map that gives the right hand side in (2), over the set $\mathscr{H}_c := \{x \in \mathbb{R}_+^n \mid |x| = c \leq (a_1 - 1)/b_1\}$ is attained at a point $x = (0, \dots 0, c, 0 \dots 0)$, hence it is equal to $ca_i/(1 + b_i c)$, for some $i \in \{1, \dots, n\}$. Further, based on (3), it follows that $\max_{x \in \mathscr{H}_c} |f(x)| \leq (a_1 - 1)/b_1$, which implies that the set $\{x \in \mathbb{R}_+^n \mid |x| \leq (a_1 - 1)/b_1\}$ is positively invariant. We present this idea in more detail in Sect. 4, when we prove a similar result for more general growth functions. Above, by abusing notation, $|x|$ denotes the L^1 norm of x. That is, $|x| = |x_1| + \cdots + |x_n|$, where $|x_i|$ is the absolute value of the real number x_i. The same notation will be used several times throughout this paper.

In case that some of the maximum carrying capacities corresponding to some of the species are equal, say when $(a_1 - 1)/b_1 = (a_2 - 1)/b_2 = \cdots = (a_k - 1)/b_k > (a_{k+1} - 1)/b_{k+1} \geq \cdots \geq (a_n - 1)/b_n$ (for some $k \in \{1, \dots, n\}$), it is shown (both in [2, 5]) that every solution of (2) with $x_1(0), \dots, x_l(0) > 0$ converges exponentially to a point on the hyperplane $\{(x_1,\ldots,x_k,0,\ldots,0) \mid x_i > 0,\ i=1,\ldots,k$, and $x_1+\cdots+x_k=(a_1-1)/b_1\}$.

When intraspecific competition is different from the interspecific competition, unstructured Leslie-Gower models still allow for competitive exclusion but also leave the door open for coexistence. As Cushing et al. show in [11] for a two-dimensional model of this type [see (4) below], the dynamics are similar to those predicted by the Lotka-Volterra model.

$$\begin{aligned} x_{t+1} &= b_1 \frac{x_t}{1+x_t+c_1 y_t} \\ y_{t+1} &= b_2 \frac{y_t}{1+c_2 x_t+y_t} \end{aligned} \tag{4}$$

Here the intraspecific coefficients [that is, the coefficients of x_t and y_t from the denominator in the first, respectively the second equation in (4)] are scaled to one. It is shown that when the growth coefficients b_1 and b_2 are not both greater than one, the model can have only boundary equilibria $E_0=(0,0)$, $E_1=(b_1-1,0)$ and $E_2=(0,b_2-1)$. In fact, if $b_1,\ b_2<1$ then E_0 is globally asymptotically stable in $\mathbb{R}^2_+$; if $b_1>1,\ b_2<1$ then E_1 is globally asymptotically stable in $\{(x,y) \mid x>0\}$; if $b_1<1,\ b_2>1$ then E_2 is globally asymptotically stable in $\{(x,y) \mid y>0\}$. So in the last two cases the model predicts competitive exclusion. However, when $b_1,\ b_2>1$, the two species can coexist at an interior equilibrium

$$E_3=\left(\frac{c_1(b_2-1)-(b_1-1)}{c_1c_2-1}, \frac{c_2(b_1-1)-(b_2-1)}{c_1c_2-1}\right).$$

The main results are presented below.

1. $c_1<(b_1-1)/(b_2-1)$ and $c_2>(b_2-1)/(b_1-1) \Rightarrow E_1$ is globally asymptotically stable in $\{(x,y) \mid x>0\}$;
2. $c_1>(b_1-1)/(b_2-1)$ and $c_2<(b_2-1)/(b_1-1) \Rightarrow E_2$ is globally asymptotically stable in $\{(x,y) \mid y>0\}$;
3. $c_1>(b_1-1)/(b_2-1)$ and $c_2>(b_2-1)/(b_1-1) \Rightarrow E_3$ is a saddle point, and all solutions starting in $\{(x,y) \mid x>0$ and $y>0\}$, but not on the (one-dimensional) global stable manifold of E_3, converge either to E_1 or to E_2;
4. $c_1<(b_1-1)/(b_2-1)$ and $c_2<(b_2-1)/(b_1-1) \Rightarrow E_3$ is globally asymptotically stable in $\{(x,y) \mid x>0$ and $y>0\}$.

Thus, cases 1, 2 and 3 above give competitive exclusion, but when interspecific competition is small, coexistence occurs (case 4).

As shown in [12], one of the "mechanisms" that can produce non-Lotka-Volterra type dynamics is to consider stronger nonlinearities to model species growth, for example by using Ricker-type functions.

$$\begin{aligned} x_{t+1} &= b_1 x_t e^{-c_{11}x_t-c_{12}y_t}+s_1x_t \\ y_{t+1} &= b_2 y_t e^{-c_{21}x_t-c_{22}y_t}+s_2y_t \end{aligned} \tag{5}$$

The parameters s_1 and s_2 in $[0, 1)$ account for the survival rates of species x and y, respectively. This model can have a coexistence equilibrium $E = ((\ln n_1 - c_1 \ln n_2)/(1-c_1c_2), (\ln n_2 - c_2 \ln n_1)/(1-c_1c_2))$ (whenever $n_i = b_i/(1-s_i) > 1$, $i = 1, 2$), but it is unstable. The single species dynamics (say for species y) are given by

$$y_{t+1} = b_2 y_t e^{-c_{22} y_t} + s_2 y_t. \tag{6}$$

At the critical value $b_2^{cr} = (1 - s_2)e^{2/(1-s_2)}$ of the parameter b_2, a period-doubling bifurcation occurs, resulting in an asymptotically stable 2-cycle (y_0^*, y_1^*) for (6), which results in an extinction 2-cycle $((0, y_0^*), (0, y_1^*))$ for (5). Denoting $r = c_2/c_1$ and $c = c_1$, model (5) becomes

$$\begin{aligned} x_{t+1} &= n_1(1 - s_1)x_t e^{-x_t - c y_t} + s_1 x_t \\ y_{t+1} &= n_2(1 - s_2)y_t e^{-r c x_t - y_t} + s_2 y_t. \end{aligned} \tag{7}$$

This is done in order to investigate the stability of the extinction 2-cycle as a function of a single inter-specific competition c.

It is shown in [12] that the Jacobian matrix of (7) evaluated at $((0, y_0^*), (0, y_1^*))$ has eigenvalues

$$\begin{aligned} \lambda_1 &= \lambda_1(c) = [n_1(1 - s_1)e^{-c y_1^*} + s_1][n_1(1 - s_1)e^{-c y_2^*} + s_1] \\ \lambda_2 &= [n_2(1 - s_2)(1 - y_1^*)e^{-y_1^*} + s_2][n_2(1 - s_2)(1 - y_2^*)e^{-y_2^*} + s_2] \end{aligned} \tag{8}$$

But $\lambda_2 < 1$ (based on b_2 being greater than, but near b_2^{cr}, which represents the existence condition for the extinction 2-cycle). Hence $\lambda_1 = \lambda_1(c)$ determines the stability of the extinction 2-cycle. Thus, since λ_1 is decreasing in c and $\lambda_1(0) = (n_1(1 - s_1) + s_1)^2 > 1$ and $\lim_{c\to\infty} \lambda_1(c) = s_1 s_2 < 1$, there exists a unique value $c = c^*$ such that $((0, y_0^*), (0, y_1^*))$ is stable for $c > c^*$, and unstable for $c < c^*$. The loss of stability of the extinction 2-cycle at the critical value of the parameter $c = c^*$ suggests existence of coexistence 2-cycles for $c < c^*$. The authors show that indeed this is the case, by Lyapunov-Schmidt expansions of $((0, y_0^*), (0, y_1^*))$, which they use, in turn, to estimate the bifurcation value c^* for the 2-cycles generated by the solution branch $(x, y, c) = (x, y(x), c(x))$ that gives the fixed points of the composite map for the right-hand side of (7). Also using Lyapunov-Schmidt expansions, the authors further give conditions for the stability of the coexistence 2-cycle (that is, conditions for $c'(0)$ to be negative), as well as for the existence of a *mixed attractor* (stable coexistence 2-cycle and stable exclusion equilibrium) in terms of the survival ratios s_1 and s_2. To complement their theoretical results, the authors offer a series of numerical simulations in order to illustrate other scenarios of mixed-type attractors, that include higher period cycles, quasi-periodic and even chaotic attractors.

Even though model (5) has richer dynamics, as compared to the previous models presented in this section, the performed mathematical analysis is a local one, thus conditions that differentiate in between *global* coexistence and competitive exclusion are not determined.

A particular form of (7) (with $s_1 = s_2 = 0$) was analyzed in [25, 31], even though the mixed attractor scenario (as mentioned above) was not observed (we give below the system used in [25]).

$$x_{n+1} = x_n e^{K - x_n - a y_n}$$
$$y_{n+1} = y_n e^{L - b x_n - y_n} \tag{9}$$

Using results concerning the dynamics of planar competitive systems (also developed in [31]), Smith [31] proved that when the carrying capacities of both species are smaller than one, either one of the two nonzero boundary equilibria, or the interior equilibrium, attracts all solutions starting in the interior of $\mathbb{R}^2_+$. With respect to model (9), these results say the following. Assume that $K,\ L \le 1$.

(a) If $aL/K < 1 < bK/L$ then the equilibrium $E_1 = (K, 0)$ attracts all points not on the y axis.
(b) If $bK/L < 1 < aL/K$ then the equilibrium $E_2 = (0, L)$ attracts all points not on the x axis.
(c) If $aL/K,\ bK/L < 1$ then the (unique) interior equilibrium E^* attracts all solutions starting in the interior of $\mathbb{R}^2_+$.

Further, for $bK/L,\ aL/K > 1$ (which corresponds to the interior equilibrium being a saddle point), it is shown that there exists a smooth curve Γ through the origin and E^* (which is, in fact, the global stable manifold of E^*), that "separates" $\mathbb{R}^2_+$ into two (disjoint) "relatively open" sets B_1 and B_2, that represent the basin of attraction of E_1 and E_2, respectively. Hence, for $K,\ L \le 1$, (9) predicts Lotka-Volterra-type dynamics.

In [25] the authors are mostly concerned with the local stability analysis for this model by determining the regions in the parameter space that dictate the local stability of each of the four equilibrium points mentioned above. Of particular interest are the *non-hyperbolic* cases, when the authors determine the stability of the nontrivial boundary equilibria E_1 and E_2, as well as of the interior equilibrium E^*, when one eigenvalue (but not both) of the Jacobian matrix evaluated at each of these equilibria has modulus equal to one. The methods involve the approximation of the (one-dimensional) center manifold near the equilibrium point using Taylor series, and then studying the dynamics on the center manifold (the Schwarzian derivative is also used for this). Further, using bifurcation analysis, the authors show that the coexistence equilibrium undergoes a *period-doubling bifurcation* leading to asymptotically stable coexistence periodic orbits of arbitrarily large periods, and eventually to chaos. Thus, as a consequence of strong nonlinearities, model (9) has the potential of violating the principle of competitive exclusion.

We conclude this section with a look at the *competitive exclusion—coexistence* dichotomy from the perspective of evolutionary game theory (EGT). Namely, we look at how evolution that occurs on a time scale comparable to that of the dynamics of the population, can influence (change) the outcome of competition between species, as discussed by Rael et al. in [30].

The mathematical model is a Leslie-Gower-type model of two competing species x_1 and x_2, in which the competition coefficients are not constant, but depend on a vector u, (with two components), that gives the phenotypic traits (or strategies) of the two species. In turn, the vector u depends on how the two species evolve over time, thus the full model consists of four difference equations:

$$\begin{aligned}
x_1(t+1) &= x_1(t)\frac{e^r}{1+c(u_1(t),u_1(t))x_1(t)+c(u_1(t),u_2(t))x_2(t)}\\
x_2(t+1) &= x_2(t)\frac{e^r}{1+c(u_2(t),u_1(t))x_1(t)+c(u_2(t),u_2(t))x_2(t)}\\
u_1(t+1) &= u_1(t)-\sigma_1^2\frac{\left[\frac{\partial}{\partial v}c(v,u_1)x_1(t)+\frac{\partial}{\partial v}c(v,u_2)x_2(t)\right]_{(u_1(t),u_2(t))}}{1+c(u_1(t),u_1(t))x_1(t)+c(u_1(t),u_2(t))x_2(t)}\\
u_2(t+1) &= 2_1(t)-\sigma_2^2\frac{\left[\frac{\partial}{\partial v}c(v,u_1)x_1(t)+\frac{\partial}{\partial v}c(v,u_2)x_2(t)\right]_{(u_1(t),u_2(t))}}{1+c(u_2(t),u_1(t))x_1(t)+c(u_2(t),u_2(t))x_2(t)}
\end{aligned}\tag{10}$$

This model considers two scenarios: with and without a *boxer effect*, meaning that the maximal competitive intensity does not (respectively does) occur between species having identical, or nearly identical, strategies. The main objective is to find out if it is possible to start with a strategy that would lead to competitive exclusion (if $u(t)$ remained constant for all $t \geq 0$) but that, in fact, evolves towards a strategy that results in the coexistence of the two species. Mathematically, this translates into solutions of (10) starting at $(x_1^0, x_2^0, u_1^0, u_2^0)$ (at $t = 0$), such that $(x_1(t), x_2(t), u_1(t), u_2(t)) \to (x_1^*, x_2^*, u_1^*, u_2^*)$ as $t \to \infty$, where $x_1^*, x_2^* > 0$, but solutions of the subsystem formed with the first two equations in (10) and keeping $u_1(t) = u_1^0$ and $u_1(t) = u_2^0$ constant for all t, would approach an extinction equilibrium (thus, implying evolution would change the outcome of the competition from competitive exclusion to coexistence). Through numerical simulations it is shown that this situation can, indeed, take place, but in the model without boxer effect it requires the initial trait to correspond to a globally attracting extinction equilibrium (with one of the two species absent) and not to a saddle interior equilibrium. The latter case is of special interest because it corresponds to the Park's "anomalous" experiment that showed coexistence of the two species of *Tribolium*. Thus, it is concluded that a boxer effect can produce new "evolutionary paths" and, in addition, it can give rise to coexistence equilibria that are ESS (evolutionary stable strategy).

3 Structured Models

From a pure mathematical perspective, competitive exclusion in structured discrete-time competition models poses a much higher degree of difficulty, and a general approach for this has not yet been established. However, Hsu et al. [21] offer such an approach for *monotone* systems. The setup is as follows. Let X_1 and X_2 be ordered Banach spaces with non-empty interiors, $X = X_1 \times X_2$, and $X^+ = X_1^+ \times X_2^+$. Let K be the cone with nonempty interior $int(K) = Int(X_1^+) \times (-Int(X_2^+))$, that generates the following partial order relations. Thus, for any $x = (x_1, x_2)$ and $\bar{x} = (\bar{x}_1, \bar{x}_2)$ in X^+, we have:

$$\begin{aligned} x \leq_K \bar{x} \ &\Leftrightarrow x_1 \leq \bar{x_1} \text{ and } x_2 \geq \bar{x_2}; \\ x <_K \bar{x} \ &\Leftrightarrow x_1 \leq \bar{x}_1,\ x_2 \geq \bar{x_2} \text{ and } x \neq \bar{x}; \\ x \ll_K \bar{x} \ &\Leftrightarrow x_1 < \bar{x_1} \text{ and } x_2 > \bar{x_2}. \end{aligned} \tag{11}$$

Let $T : X^+ \mapsto X^+$ be an order compact, continuous map, having the following properties:

(H1) $x <_K \bar{x} \Rightarrow T(x) <_K T(\bar{x})$;
(H2) $T(0) = 0$ and $\exists\ U$ a neighborhood of 0 such that $\forall\ x \in U\backslash\{0\}$, $\exists\ n = n(x)$ such that $T^n(x) \notin U$.
(H3) $T(X_1^+ \times \{0\}) \subset X_1^+ \times \{0\}$ and $\exists\ \hat{x}_1 \gg 0$ such that $T(\hat{x}_1, 0) = (\hat{x}_1, 0)$ and $T^n(x_1, 0) \rightarrow (\hat{x}_1, 0)$, $\forall\ x_1 > 0$. Assume a symmetric condition for T on $\{0\} \times X_2^+$, with fixed point $(0, \tilde{x}_2)$.
(H4) If $x, y \in X^+$ satisfy $x <_K y$ and either x or y belong to $Int(X^+)$ then $T(x) \ll_K T(y)$. If $x = (x_1, x_2) \in X^+$ satisfies $x_i \neq 0$, $i = 1, 2$, then $T(x) \gg 0$.

Under these assumptions, if the map T does not have a fixed point in $Int(X^+)$, then either $E_1 = (\hat{x}_1, 0)$, or $E_2 = (0, \tilde{x}_2)$ attracts all solutions of $x_{n+1} = T(x_n)$ starting in $Int(X^+) \cap (I := [0, \hat{x}_1] \times [0, \tilde{x}_2])$. For solutions starting in $Int(X^+) \setminus I$, solutions converge either to E_1, or to E_2, depending on the initial condition.

Thus, in a competition modeled by such a map T, if the two species (x_1 and x_2) cannot coexist at an interior equilibrium, the outcome predicted by the model is competitive exclusion. However, for solutions with both species present and starting outside I, we do not know which species persists and which dies out.

The model of Ackleh and Zhang [3], used to investigate the competition between the Louisiana blue flag iris (*I. hexagona*), which is a wild species of iris, and the yellow flag iris (*I. pseudacorus*), which is a cultivated species, fits in the above framework:

$$\begin{aligned} x_{t+1}^A &= b_1^A z_t^A \\ y_{t+1}^A &= s_1^A x_t^A + b_2^A z_t^A \\ z_{t+1}^A &= s_2^A(\phi_t) y_t^A + s_3^A(\phi_t) z_t^A \\ x_{t+1}^B &= b_1^B z_t^B \\ y_{t+1}^B &= s_1^B x_t^B + b_2^B z_t^B \\ z_{t+1}^B &= s_2^B(\phi_t) y_t^B + s_3^B(\phi_t) z_t^B \end{aligned} \tag{12}$$

The two species are denoted by A and B, while x, y and z stand for seeds, juveniles and adults, respectively. ϕ denotes the total number of plants (juvenile and adult) for both species together. The survivorship functions s_i^k, $k = A, B$, are assumed to satisfy $s_1^k \in (0, 1)$, $0 < s_2^k(0) \le s_3^k(0) < 1$ and $(s_2^k)'(x) \le (s_3^k)'(x) < 0$ for all $x \ge 0$. In addition, for $i = 2, 3$, $s_1^k \in C^1[0, \infty)$, $(s_2^k(x)x)' > 0$ for all $x \ge 0$, $\lim_{x\to\infty} s_i^k(x) = 0$, $\lim_{x\to\infty} s_i^k(x)x < \infty$.

In order to verify assumption (H3) above, Ackleh and Zhang use the fact that the single species dynamics are given by a monotone system. The nontrivial boundary equilibria are denoted by $E_1 = (\hat{x}^A, \hat{y}^A, \hat{z}^A, 0, 0, 0)$, respectively $E_2 = (0, 0, 0, \hat{x}^B, \hat{y}^B, \hat{z}^B)$. The partial order relation $\le_K$ is chosen such that $x = (x_1, x_2)$ $y = (y_1, y_2) \in \mathbb{R}_+^3 \times \mathbb{R}_+^3$ satisfy $x \le_K y$ if and only if $x_1 \le y_1$ and $x_2 \ge y_2$. Assumption (H4) is replaced by the slightly weaker

(H4)' If $x = (x_1, x_2)$ or $y = (y_1, y_2)$ are in $Int(\mathbb{R}_+^3 \times \mathbb{R}_+^3)$ and $x <_K y$ then $T^l(x) \ll T^l(y)$ for some $l \ge 0$. If $x = (x_1, x_2) \in \mathbb{R}_+^3 \times \mathbb{R}_+^3$ and $x_i \ne (0, 0, 0)$, $i = 1, 2$, then $T^m \gg 0$ for some $m \ge 0$.

It is shown in [3] that, whenever locally asymptotically stable, E_1 also attracts all solutions having non-zero components corresponding to species A if and only if $\hat{y}^A + \hat{z}^A > \hat{y}^B + \hat{z}^B$ (with an analogous statement for species B), and that coexistence occurs only in the extreme case $\hat{y}^A + \hat{z}^A = \hat{y}^B + \hat{z}^B$.

The authors then extend their model to allow for different intraspecific competition efficiencies between the two species [34], where they consider the following discrete two-species model:

$$\begin{aligned}
x_{t+1}^A &= b_1^A z_t^A \\
y_{t+1}^A &= s_1^A x_t^A + b_2^A z_t^A \\
z_{t+1}^A &= s_2^A(\phi_1) y_t^A + s_3^A(\phi_1) z_t^A \\
x_{t+1}^B &= b_1^B z_t^B \\
y_{t+1}^B &= s_1^B x_t^B + b_2^B z_t^B \\
z_{t+1}^B &= s_2^B(\phi_2) y_t^B + s_3^B(\phi_2) z_t^B .
\end{aligned} \tag{13}$$

Here it is assumed, due to competition, that the survivorship of both juveniles and adults for the two species depend on a weighted total number of plants, $\phi_1 =: \phi_t^A + c_1\phi_t^B$ and $\phi_2 =: c_2\phi_t^A + \phi_t^B$, respectively (where $\phi_t^A = y_t^A + z_t^A$ is the total number of plants (juveniles and adults) for species A, and $\phi_t^B = y_t^B + z_t^B$ is the total number of plants for species B, at time t). The competition coefficients $c_1 > 0$, $c_2 > 0$ represent a measure of the strength of interspecific competition between the two species.

Note that this model reduces to a model of type (12) in the case $c_1 = c_2 = 1$. The main goal in [34] is to extend the analysis to the case $c_1, c_2 > 0$. In addition to the global analysis of model (13) with weak nonlinear survivorship functions (Beverton-Holt type), the authors also establish the local stability with Ricker type survivorship functions and perform numerical bifurcation analysis on these strong nonlinearities. They define the net reproductive number of species A and B at density level ϕ by

$$R^A(\phi) = \frac{b_1^A s_1^A s_2^A(\phi) + s_2^A(\phi) b_2^A}{1 - s_3^A(\phi)} \quad \text{and} \quad R^B(\phi) = \frac{b_1^B s_1^B s_2^B(\phi) + s_2^B(\phi) b_2^B}{1 - s_3^B(\phi)}.$$

The *inherent* net reproductive number of species A is $R^A(0)$ and of species B is $R^B(0)$. Notice that $R^A(\phi)$ and $R^B(\phi)$ are decreasing functions with $\lim_{\phi\to\infty} R^A(\phi) = \lim_{\phi\to\infty} R^B(\phi) = 0$. Thus, if $R^A(0) > 1$ and $R^B(0) > 1$ then there exist unique positive real numbers α_1 and α_2 such that $R^A(\alpha_1) = 1$ and $R^B(\alpha_2) = 1$.

Clearly, model (13) has the trivial equilibrium denoted by $E_0 = (0, 0, 0, 0, 0, 0)$. Furthermore, if $R^A(0) > 1$ and $R^B(0) > 1$, then each species (living alone) has a positive globally asymptotically stable interior fixed point. Denote this equilibrium for species A by $E^A = (\hat{x}^A, \hat{y}^A, \hat{z}^A)$ and for species B by $E^B = (\hat{x}^B, \hat{y}^B, \hat{z}^B)$. Thus, it follows that model (13) has two nontrivial boundary equilibria given by $E_1 = (E^A, 0) \in \mathbb{R}^6_+$ and $E_2 = (0, E^B) \in \mathbb{R}^6_+$. Therefore, $\alpha_1 = \hat{y}^A + \hat{z}^A$ and $\alpha_2 = \hat{y}^B + \hat{z}^B$. Furthermore, if either

$$\alpha_2 < c_2\alpha_1 \quad \text{and} \quad \alpha_1 < c_1\alpha_2$$

or

$$\alpha_2 > c_2\alpha_1 \quad \text{and} \quad \alpha_1 > c_1\alpha_2 \tag{14}$$

is satisfied, then system (13) has a unique interior equilibrium

$$E_3 = (b_1^A \bar{z}^A, (s_1^A b_1^A + b_2^A)\bar{z}^A, \bar{z}^A, b_1^B \bar{z}^B, (s_1^B b_1^B + b_2^B)\bar{z}^B, \bar{z}^B),$$

where

$$\bar{z}^A = \frac{\phi^A}{1 + s_1^A b_1^A + b_2^A}, \quad \bar{z}^B = \frac{\phi^B}{1 + s_1^B b_1^B + b_2^B}.$$

Relying on the monotonicity of the system when the nonlinearities are weak, and using the theory of [21], Zhang and Ackleh establish the following:

(a) If $c_2 > \frac{\alpha_2}{\alpha_1}$ and $c_1 < \frac{\alpha_1}{\alpha_2}$, then the boundary equilibrium E_1 is globally asymptotically stable;
(b) If $c_2 < \frac{\alpha_2}{\alpha_1}$ and $c_1 > \frac{\alpha_1}{\alpha_2}$, then the boundary equilibrium E_2 is globally asymptotically stable;
(c) If $c_2 > \frac{\alpha_2}{\alpha_1}$ and $c_1 > \frac{\alpha_1}{\alpha_2}$, then both boundary equilibria are locally asymptotically stable;
(d) If $c_2 < \frac{\alpha_2}{\alpha_1}$ and $c_1 < \frac{\alpha_1}{\alpha_2}$, then the unique interior equilibrium E_3 is globally attractive.

Then the authors study the local stability in the case of strong nonlinearities and investigate the model dynamics by using numerical simulations and bifurcation diagrams. The numerical results indicate that the criteria given in (a–d) above can be used to predict the competition outcome even for strong nonlinearities. While under such strong nonlinearities competitive exclusion, coexistence and bistability can be obtained, the surviving species may have richer dynamics including periodic or chaotic attractors.

A slightly different competition model between two species, containing only two age groups (juveniles and adults) for each species, has been considered by Cushing et al. [13].

$$
\begin{aligned}
J_{t+1} &= b_1 \frac{1}{1 + d_1 A_t} A_t \\
A_{t+1} &= s_1 \frac{1}{1 + J_t + c_1 j_t} J_t \\
j_{t+1} &= b_2 \frac{1}{1 + d_2 a_t} a_t \\
a_{t+1} &= s_2 \frac{1}{1 + c_2 J_t + j_t} j_t
\end{aligned} \tag{15}
$$

A particular feature of this model, that facilitates its analysis, is that its variables that are two time-steps apart form two decoupled, unstructured systems, one for each species.

$$
\begin{aligned}
J_{t+2} &= b_1 s_1 \frac{1}{1 + (1 + d_1 s_1) J_t + c_1 j_t} J_t \\
A_{t+2} &= b_1 s_1 \frac{1}{1 + (d_1 + b_1) A_t + c_1 b_2 a_t (1 + d_1 A_t)/(1 + d_2 a_t)} A_t \\
j_{t+2} &= b_2 s_2 \frac{1}{1 + c_2 J_t + (1 + d_2 s_2) j_t} j_t \\
a_{t+2} &= b_2 s_2 \frac{1}{1 + c_2 b_1 A_t (1 + d_2 a_t)/(1 + d_1 A_t) A_t + (d_2 + b_2) a_t} a_t
\end{aligned} \tag{16}
$$

The two submodels corresponding to juveniles, respectively adults, in (16) are of the form (4). Thus, they produce similar Lotka-Volterra-type dynamics for (15), in the cases when the inter-specific competition coefficients c_1 and c_2 are not both large. More exactly, assuming that $b_i s_i > 1$, $i = 1, 2$, we have the following:

(a) if $c_1 < (1+d_2 s_2)(b_1 s_1 - 1)/(b_2 s_2 - 1)$ and $c_2 < (1+d_1 s_1)(b_2 s_2 - 1)/(b_1 s_1 - 1)$, then there exists a coexistence equilibrium that is globally asymptotically stable in $Int(\mathbb{R}^4_+)$;
(b) if $c_1 < (1+d_2 s_2)(b_1 s_1 - 1)/(b_2 s_2 - 1)$ and $c_2 > (1+d_1 s_1)(b_2 s_2 - 1)/(b_1 s_1 - 1)$, then the exclusion equilibrium $(J_e, A_e, 0, 0) = ((b_1 s_1 - 1)/(1 + d_1 s_1), (b_1 s_1 - 1)/(b_1 + d_1), 0, 0)$ is asymptotically stable and attracts all solutions of (15) starting in $Int(\mathbb{R}^4_+)$;

(c) if $c_1 > (1+d_2 s_2)(b_1 s_1 - 1)/(b_2 s_2 - 1)$ and $c_2 < (1+d_1 s_1)(b_2 s_2 - 1)/(b_1 s_1 - 1)$, then the exclusion equilibrium $(0, 0, j_e, a_e) = (0, 0, (b_2 s_2 - 1)/(1+d_2 s_2), (b_2 s_2 - 1)/(b_2 + d_2))$ is asymptotically stable and attracts all solutions of (15) starting in $Int(\mathbb{R}^4_+)$.

However, when both c_1 and c_2 are large, which corresponds to the case when the interior equilibrium is unstable (saddle), there exists a nontrivial coexistence (local) attractor. Thus,

(d) if $c_1 > (1+d_2 s_2)(b_1 s_1 - 1)/(b_2 s_2 - 1)$ and $c_2 > (1+d_1 s_1)(b_2 s_2 - 1)/(b_1 s_1 - 1)$, then the coexistence equilibrium (for which explicit formula is given in [13]) is a saddle and the 2-cycle $\{(J_e, 0, 0, a_e), (0, A_e, j_e, 0)\}$ is locally asymptotically (orbitally) stable.

Ackleh et al. [2] re-visited model (2), but this time allowing for mutation. That is, individuals from species j can give birth, with probability γ_{ij} to individuals belonging to other species, i. Thus, model (2) becomes

$$x_i(t+1) = \frac{(\gamma_{ii} a_i + 1) x_i(t) + \sum_{j \neq i} \gamma_{ij} a_j x_j(t)}{1 + b_i \sum_j x_j(t)}, \quad a_i > 0, b_i > 0. \tag{17}$$

When the *mutation matrix* Γ is block-diagonal

$$\Gamma = \begin{pmatrix} \Gamma_1 & 0 & 0 & 0 & 0 \\ 0 & \Gamma_2 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \dots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \Gamma_l \end{pmatrix}, \tag{18}$$

(17) can be regarded as a structured model of the form

$$x^{(i)}(t+1) = A_i(x(t)) x^{(i)}(t), \; i = 1, \dots, l, \tag{19}$$

where $x = (x^{(1)}, \dots, x^{(l)})$, $x^{(i)} \in \mathbb{R}^{n_i}$. Hence $n_1 + \dots + n_l = n$, but we assume that $n_1 \geq 2$ for some i, in order to have a model different from (2).

Let $b_1^M = \min_{j=1}^{n_1} b_j$ and $b_i^m = \max_{j=1}^{n_i}$, $i = 2, \dots, l$. Then from (19) we have

$$\begin{aligned} x^{(1)}(t+1) &\geq \frac{1}{1 + b_1^M |x(t)|} \tilde{A}_1 x^{(1)}(t) \\ x^{(i)}(t+1) &\leq \frac{1}{1 + b_i^m |x(t)|} \tilde{A}_i x^{(i)}(t), \quad i = 2, \dots, l, \end{aligned} \tag{20}$$

where $\tilde{A}_1 = (1 + b_1^M |x|) A_i$ and $\tilde{A}_i = (1 + b_i^m |x|) A_i$, $i = 2, \dots, l$. It is shown in [2] that, when each Γ_i, $i = 1, \dots, n$, is irreducible and $(r_j / r_1) \max\{1, b_1^M / b_j^m\} < 1$ for

$j \geq 2$, where r_i denotes the spectral radius of $\tilde{A}_i$, $i = 1, \ldots, n$, then $x^{(j)}(t) \to 0$, as $t \to \infty$. This condition, which leads to competitive exclusion, is weaker than (3). This is primarily due to the fact that the methods of proof for the unstructured model (2) could not be extended to (17).

Modeling the growth functions using stronger nonlinearities in structured models also complicate the global dynamics (as compared to the classical Lotka-Volterra models). To illustrate this fact, the authors in [11] consider the following Ricker competition model, when one of the two species is structured by age (juveniles and adults).

$$\begin{aligned} J_{t+1} &= b_1 A_t \exp(-c_{11} A_t - c_{12} y_t) \\ A_{t+1} &= (1-\mu) J_t \\ y_{t+1} &= b_2 y_t \exp(-c_{21} J_t - c_{22} y_t). \end{aligned} \tag{21}$$

It is shown that this model does not produce Lotka-Volterra-type dynamics, in the sense that local stability of the non-zero boundary equilibria does not necessarily imply competitive exclusion (as in cases 1–3 on page XX). The authors prove the existence of coexistence 2-cycles by showing that the composite map [that is $F \circ F$, where F is the right-hand-side of (21)] has non negative fixed points. For this, they set $J = 0$, and then the fixed points of F^2 satisfy

$$\begin{aligned} b_1(1-\mu)\exp(-c_{11}A - c_{12}y) &= 1 \\ b_2^2 \exp(-c_{22}y - c_{21}A\frac{1}{1-\mu} - b_2 c_{22} y e^{-c_{22}y}) &= 1. \end{aligned} \tag{22}$$

From (22), a scalar equation for y is obtained:

$$b_2 c_{22} y e^{-c_{22}y} = \left(\frac{c_{12}c_{21}}{(1-\mu)c_{11}}\right) y + \left(2\ln b_2 - \frac{c_{21}}{(1-\mu)c_{11}} \ln b_1(1-\mu)\right). \tag{23}$$

The graphs of the two functions given by the left and right-had-sides in (23) can have at most two intersection points. Such an intersection point y^* gives a fixed point (A^*, y^*) for (22) which, in turn, yields a 2-cycle $\{(0, A^*, y^*), (\bar{J}, 0, \bar{y})\}$ for (21). For a particular set of parameters, a 2-cycle is computed numerically and then verified to be asymptotically stable, while the corresponding boundary equilibria are also asymptotically stable, hence model (21) exhibits a "multiple attractor", making coexistence possible.

Along the same lines, there is the age structured LPA (larvae, pupae, adult) two-species competition model (see [8–10, 16, 17]).

$$\begin{aligned} l_{t+1} &= b a_t e^{-c_{el} l_t - c_{ea} a_t - c_{eL} L_t - c_{eA} A_t} \\ p_{t+1} &= l_t(1-\mu_l) \\ a_{t+1} &= p_t e^{-c_{pa} a_t - c_{pA} A_t} + a_t(1-\mu_a) \\ L_{t+1} &= B A_t e^{-c_{El} l_t - c_{Ea} a_t - c_{EL} L_t - c_{EA} A_t} \\ P_{t+1} &= L_t(1-\mu_L) \\ A_{t+1} &= P_t e^{-c_{Pa} a_t - c_{PA} A_t} + A_t(1-\mu_A) \end{aligned} \tag{24}$$

This model is a natural extension of the well-known single species LPA model (see [6, 7, 14, 19, 20, 22]) and mainly developed to try to explain some of the "anomalies" (multiple coexistence attractors) observed in the lab experiments of Park and his collaborators (see [24, 26–29]) involving two closely related species, *T. confusum* and *T. castaneum*.

Through a series of numerical simulations, Edmunds [16] and Edmunds et al.[17] show that various scenarios can take place in regard to the global dynamics generated by (24):

1. a globally attracting 2-cycle, with one of the species extinct;
2. two attractors: one 2-cycle with one of the species extinct and a coexistence attractor (which could be periodic, or even chaotic);
3. three attractors: one 2-cycle with one of the species extinct and two coexistence attractor (one of which being chaotic);
4. two extinction attractors (with one of the species "missing"): one 2-cycle and one chaotic, and a coexistence 2-cycle;
5. two extinction attractors (with one of the species "missing"): one 2-cycle and one chaotic.

However, the basins of attraction of these attractors have complicated structure, some of them being fractals. Thus, predictions as to which species survives when the initial population is near these boundaries is practically impossible, in the presence of stochasticity.

Extending the work in [32], Smith and Zhao [33] consider a chemostat model where m species compete for a common food source.

$$
\begin{aligned}
x^i_{n+1} &= A_i(S_n)x^i_n, \quad i = 1, \ldots, m, \\
S_{n+1} &= (1-E)\left(S_n - \sum_{j=1}^{m} f_j(S_n)U^j_n\right) + ES^0.
\end{aligned}
\tag{25}
$$

Each species (at time step n) is structured by size and denoted in the model by the vector $x^i_n \in \mathbb{R}^{r_i}$, and $U^i_n = |x^i_n|$. S_n denotes the nutrient concentration (at time step n). The nutrient uptake rate f is of Michaelis-Menten form $f(S) = MS/(a + S)$, with M being the maximum uptake rate, and a being the half-saturation (or Michaelis-Menten) constant (in fact, the paper allows more general forms of f, the Michaelis-Menten form being just a particular case). The projection matrix for the ith population is

$$
\begin{pmatrix}
1-P_i & 0 & \ldots & & & M_iP_i \\
M_iP_i & 1-P_i & 0 & \ldots & & 0 \\
0 & M_iP_i & 1-P_i & 0 & \ldots & 0 \\
\vdots & & \ddots & & & \vdots \\
0 & & \cdots & 0 & M_iP_i & 1-P_i
\end{pmatrix},
$$

where $M_i = 2^{1/r_i}$ and $P_i = f_i(S)/(M_i - 1)$. It follows then that the total biomass for the ith species satisfies the equation

$$U^i_{n+1} = (1-E)(1+f_i(S_n))U^i_n, \quad i = 1, \dots, m. \tag{26}$$

Denoting $S_n + \sum_{i=1}^n U^i_n$ by Σ_n, Eqs. (25) and (26) imply

$$\Sigma^i_{n+1} = (1-E)\Sigma_n + ES^0. \tag{27}$$

In turn, from (27) it follows that $\Sigma_n \to S^0$ as $n \to \infty$. Thus, replacing S_n by $S^0 - \sum_{i=1}^n U^i_n$, (26) becomes an unstructured, *limiting* system. The assumptions made on f guarantee that the equation

$$(1-E)(1+f_i(S)) = 1$$

has at most one solution and, whenever such a solution exists, it is denoted by λ_i (and referred to as the break-even nutrient concentration for the ith population). Otherwise, λ_i is set to be ∞. It is shown then that the total biomass of the species having the lowest λ_i, with $\lambda_i < S^0$, settles at a positive equilibrium equal to $S^0 - \lambda_1$, while all the other species go extinct. This is done using the Lyapunov function $W : \mathbb{R}^m \to \mathbb{R}$, $W(x) = |x|$, and the LaSalle invariance principle. Further, using properties on *internally chain transitive sets* (omega limit sets of bounded solutions being such sets) the above results regarding the limiting system are "lifted" back to the system formed by (26) together with the second equation in (25), from where it follows that every solution of (25) with $x^1_0 \neq 0$ converges to $(\bar{x}^1, 0, \dots 0, \lambda_1)$, where $\bar{x}^1 = ((S^0 - \lambda_1)/r_1, \dots, (S^0 - \lambda_1)/r_1)$.

4 Extending the Competitive Exclusion Result in [5]

In this section we revisit the model (2), but consider the growth functions to be of a more general form. Specifically, we consider the model

$$x_i(t+1) = g_i(|x(t)|)x_i(t), \quad i = 1, \dots, n. \tag{28}$$

Each $g_i : \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be non increasing and to have the property that there exists a unique value $\alpha_i > 0$ such that $g_i(\alpha_i) = 1$. We also assume that $y \mapsto yg_i(y)$ is non decreasing for all $i = 1, \dots, n$ and $y \in [0, \max \alpha_i]$. The proof of the following result is in the spirit of the approach developed in [2].

Theorem 1 *Assume that* $\alpha_1 = \alpha_2 = \cdots = \alpha_k > \alpha_{k+1} \geq \cdots \geq \alpha_n$*, for some* $k \in \{1, \dots, n\}$*. Then* $x_i(t) \to 0$ *as* $t \to \infty$*, for all* $i = k+1, \dots, n$*, and* $x(t) \to \bar{x}$ *as* $t \to \infty$*, where* $\bar{x} = \bar{x}(x(0))$ *is an equilibrium with* $|\bar{x}| = \alpha_1$.

Proof Let $f = (f_1, \dots, f_n)$, where $f_i(x) := g_i(|x|)x_i$, $i = 1, \dots, n$. Let $\mathscr{H}_c = \{x \in \mathbb{R}^n_+ \mid |x| = c > 0\}$ and assume $c \leq \alpha_1$. Then, for $x \in \mathscr{H}_c$, we have

$$|f(x)| = \sum_{i=1}^{n} g_i(c)x_i. \tag{29}$$

Note that the right hand side in (29) is a linear function of x, hence the maximum of $|f(x)|$, for $x \in \mathscr{H}_c$, is attained at a point $x = (0, \dots 0, c, 0 \dots 0)$.

$$\max_{x \in \mathscr{H}_c} |f(x)| = g_i(c)c, \quad \text{for some } i \in \{1, \dots, n\}. \tag{30}$$

But

$$g_i(c)c \leq g_i(\alpha_1)\alpha_1 \leq g_i(\alpha_i)\alpha_1 = \alpha_1. \tag{31}$$

This implies that $\mathscr{S} := \{x \in \mathbb{R}^n_+ \mid |x| \leq \alpha_1\}$ is positively invariant for (28).

Now let $\tilde{x} \in \mathbb{R}^n_+$ such that $\tilde{x}_1 + \cdots + \tilde{x}_k > 0$ and consider the solution $x(t)$ of (28) with $x(0) = \tilde{x}$. Without loss of generality we consider $\tilde{x}_i > 0$ for all $i = 1, \dots, k$. Then we have one of the following two cases:

(a) $\tilde{x} \in \mathscr{S}$. Then $x(t) \in \mathscr{S}$ for all $t \geq 0$, which implies that $x_i(t+1) \geq x_i(t)$, $i = 1, \dots, k$, for all $t \geq 0$. Hence $x_i(t)$ is convergent to an $\bar{x}_i > 0$, $i = 1, \dots, k$, as $t \to \infty$. Then, from the equation for x_1 in (28), it follows that $|x(t)| \to \alpha_1$. This implies that $\lim_{t\to\infty} x_i(t+1)/x_i(t) = g_i(\alpha_1) < 1$, for all $i = k+1, \dots, n$, hence $x_i(t) \to 0$, $i = k+1, \dots, n$. Thus, $x(t) \to \bar{x} = (\bar{x}_1, \dots, \bar{x}_k, 0, \dots, 0)$, where $\bar{x}$ must be an equilibrium with $|\bar{x}| = \alpha_1$.

(b) $\tilde{x} \notin \mathscr{S}$. If $x(t) \notin \mathscr{S}$ for all $t \geq 0$ then

$$x_i(t+1) \leq g_i(\alpha_1)x_i(t), \quad \forall\, t \geq 0, \quad i = 1, \dots, n. \tag{32}$$

If $i \notin \{1, \dots, k\}$, then (32) implies that $x_i(t) \to 0$ as $t \to \infty$, because $\alpha_1 > \alpha_i$. If $i \in \{1, \dots, k\}$, then (32) implies that $x_i(t+1) \leq x_i(t)$ for all $t \geq 0$, hence $x_i(t)$ converges to some $\bar{x}_i$, as $t \to \infty$. Then, from the equation for x_1 in (28), it follows that $|x(t)| \to \alpha_1$. So again, $x(t) \to \bar{x} = (\bar{x}_1, \dots, \bar{x}_k, 0, \dots, 0)$, where $\bar{x}$ must be an equilibrium with $|\bar{x}| = \alpha_1$.

If $x(t) \in \mathscr{S}$ for some $t > 0$, then, without loss of generality we can consider that we are in case *(a)* above.

□

Remark 1 Suppose that the growth function $g_i(y) = a_i e^{-b_i y}$, i.e., is of the Ricker type. Then, if $\ln(a_1)/b_1 \leq 1/b_i$, $i = 1, \dots, n$, the function $g_i(y)y$ is increasing on the interval $y \in [0, \alpha_1]$ and hence Theorem 1 applies.

Acknowledgments The work of A.S. Ackleh is partially supported by the National Science Foundation under grant # DMS-1312963. The work of P.L. Salceanu is partially supported by the Louisiana Board of Regents under grant # LEQSF(2012-15)-RD-A-29.

References

1. Ackleh, A.S., Dib, Y., Jang, S.: A discrete-time Beverton-Holt competition model. In: Allen, L.J.S., Aulback, B., Elaydi, S., Sacker, R. (eds.) Difference Equations and Discrete Dynamical Systems, pp. 1–10. World Scientific, Singapore (2005)
2. Ackleh, A.S., Sacker, R.J., Salceanu, P.L.: On a discrete selection-mutation model. Manuscript under review.
3. Ackleh, A.S., Zhang, P.: Competitive exclusion in a discrete stage-structured two species model. Math. Model. Nat. Phenom. **6**, 156–175 (2009)
4. AlSharawi, Z., Rhouma, M.: Coexistence and extinction in a competitive exclusion Leslie/Gower model with harvesting and stocking. J. Diff. Equ. Appl. **15**, 1031–1053 (2009)
5. Chow, Y., Hsieh, J.: On multi-dimensional discrete-time Beverton-Holt competition models. J. Diff. Equ. Appl. **19**, 491–506 (2013)
6. Costantino, R.F., Desharnais, R.A., Cushing, J.M., Dennis, B.: Chaotic dynamics in an insect population. Science **275**, 389–391 (1997)
7. Costantino, R.F., Cushing, J.M., Dennis, B., Desharnais, R.A., Henson, S.M.: Resonant population cycles in temporally fluctuating habitats. Bull. Math. Biol. **60**, 247–273 (1998)
8. Cushing, J.M.: An Introduction to Structured Population Dynamics. SIAM, Philadelphia (1998)
9. Cushing, J.M.: The LPA model. Fields Inst. Commun. **43**, 29–55 (2004)
10. Cushing, J.M., Costantino, R.F., Dennis, B., Desharnais, R.A., Henson, S.M.: Chaos in Ecology. Academic Press/Elsevier, San Diego (2003)
11. Cushing, J.M., Levarge, S., Chitnis, N., Henson, S.M.: Some discrete competition models and the competitive exclusion principle. J. Diff. Equ. Appl. **10**, 1139–1151 (2004)
12. Cushing, J.M., Henson, S.M., Blackburn, C.C.: Multiple mixed-type attractors in a competition model. J. Biol. Dyn. **4**, 347–362 (2007)
13. Cushing, J.M., Henson, S.M., Roeger, L.-I.: Coexistence of competing juvenile-adult structured populations. J. Biol. Dyn. **1**, 201–231 (2007)
14. Desharnais, R.A., Costantino, R.F., Cushing, J.M., Henson, S.M., Dennis, B.: Chaos and population control of insect outbreaks. Ecol. Lett. **4**, 229–235 (2001)
15. Dowson, P.S.: Developmental rate and competitive ability in Tribolium. III. Competition in unfavorable environments. J. Stored Prod. Res. **3**, 193–198 (1967)
16. Edmunds, J.L.: Multiple attractors in a discrete competition model. Theor. Popul. Biol. **72**, 379–388 (2007)
17. Edmunds, J.L., Cushing, J.M., Costantino, R.F., Henson, S.M., Dennis, B., Desharnais, R.A.: Park's Tribolium competition experiments: a non-equilibrium species coexistence hypothesis. J. Anim. Ecol. **72**, 703–712 (2003)
18. Gause, G.F.: The Struggle for Existence. The Williams & Wilkins Company, Baltimore (1934)
19. Henson, S.M., Costantino, R.F., Cushing, J.M., Desharnais, R.A., Dennis, B., King, A.A.: Lattice effects observed in chaotic dynamics of experimental populations. Science **294**, 602–605 (2001)
20. Henson, S.M., Cushing, J.M., Costantino, R.F., Dennis, B., Desharnais, R.A.: Phase switching in population cycles. Proc. Roy. Soc. Lond. B **265**, 2229–2234 (1998)
21. Hsu, S.B., Smith, H.L., Waltman, P.: Competitive exclusion and coexistence for competitive systems on ordered Banach spaces. Trans. Am. Math. Soc. **348**, 4083–4094 (1996)
22. King, A.A., Desharnais, R.A., Henson, S.M., Costantino, R.F., Cushing, J.M., Dennis, B.: Random perturbations and lattice effects in chaotic population dynamics. Science **297**, 2163a (2002)

23. Leslie, P.H., Gower, J.C.: The properties of a stochastic model for two competing species. Biometrika **45**, 316–330 (1958)
24. Leslie, P.H., Park, T., Mertz, D.B.: The effect of varying the initial numbers on the outcome of competition between two Tribolium species. J. Anim. Ecol. **37**, 9–23 (1968)
25. Luis, R., Elaydi, S., Oliveira, H.: Stability of a Ricker-type competition model and the competitive exclusion principle. J. Biol. Dyn. **5**, 636–660 (2011)
26. Park, T.: Experimental studies of interspecies competition. I. Competition between populations of the flour beetles Tribolium confusum Duval and Tribolium castaneum Herbst. Ecol. Monogr. **18**, 265–308 (1948)
27. Park, T.: Experimental studies of interspecies competition. II. Temperature, humidity and competition in two species of Tribolium. Physiol. Zool. **27**, 177–238 (1954)
28. Park, T.: Experimental studies of interspecies competition. III. Relation of initial species proportion to the competitive outcome in populations of Tribolium. Physiol. Zool. **30**, 22–40 (1957)
29. Park, T., Leslie, P.H., Mertz, D.B.: Genetic strains and competition in populations of Tribolium. Physiol. Zool. **37**, 97–162 (1964)
30. Rael, R.C., Vincent, T.L., Cushing, J.M.: Competitive outcomes changed by evolution. J. Biol. Dyn. **5**, 227–252 (2011)
31. Smith, H.L.: Planar competitive and cooperative difference equations. J. Diff. Equ. Appl. **3**, 335–357 (1998)
32. Smith, H.L.: A discrete, size-structured model of microbial growth and competition in the chemostat. J. Math. Biol. **34**, 734–754 (1996)
33. Smith, H.L., Zhao, X.-Q.: Competitive exclusion in a discrete-time, size-structured chemostat model. Discrete Contin. Dyn. Syst. Ser. B **1**, 183–191 (2001)
34. Zhang, P., Ackleh, A.S.: A discrete stage-structured two species competition model with sexual and clonal reproduction. J. Biol. Dyn. **6**, 2–16 (2012)

Benford Solutions of Linear Difference Equations

Arno Berger and Gideon Eshun

Abstract Benford's Law (BL), a notorious gem of mathematics folklore, asserts that leading digits of numerical data are usually not equidistributed, as might be expected, but rather follow one particular logarithmic distribution. Since first recorded by Newcomb in 1881, this apparently counter-intuitive phenomenon has attracted much interest from scientists and mathematicians alike. This article presents a comprehensive overview of the theory of BL for autonomous linear difference equations. Necessary and sufficient conditions are given for solutions of such equations to conform to BL in its strongest form. The results extend and unify previous results in the literature. Their scope and limitations are illustrated by numerous instructive examples.

1 Introduction

The study of digits generated by dynamical processes is a classical and rather wide subject that continues to attract interest from disciplines as diverse as ergodic and number theory [1, 14, 15, 27, 30], statistics [18, 21, 32], political science [16, 31, 40], and accounting [12, 13, 33, 37, 38]. Across these disciplines, one recurring theme is the surprising ubiquity of a logarithmic distribution of digits often referred to as *Benford's Law* (BL). The most well-known special case of BL is the so-called *first-digit law* which asserts that

$$\mathbb{P}\,(\text{leading digit } = d) = \log\left(1 + d^{-1}\right), \quad \forall d = 1, 2, \ldots, 9, \qquad (1)$$

A. Berger (✉) · G. Eshun
University of Alberta, Edmonton, AB, Canada
e-mail: berger@ualberta.ca

G. Eshun
e-mail: geshun@ualberta.ca

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_2

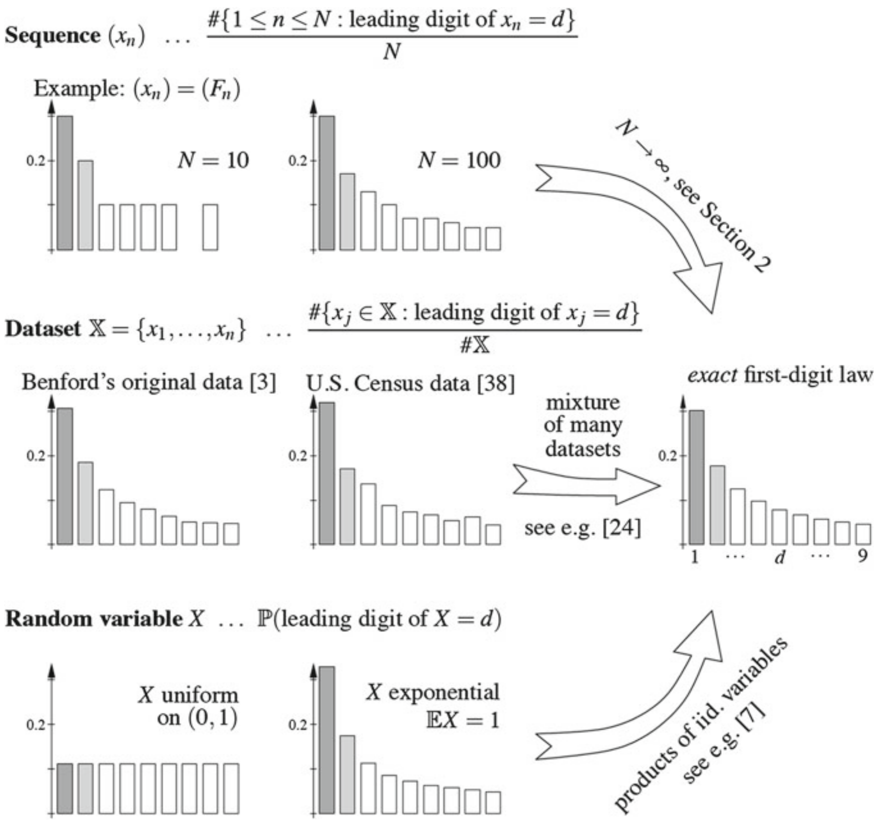

Fig. 1 Different interpretations of (1) for sequences, datasets, and random variables, respectively, and scenarios that may lead to conformance to the first-digit law

where *leading digit* refers to the first significant (decimal) digit (see Sect. 2 for rigorous definitions) and log is the base-10 logarithm; for example, the leading digit of $e = 2.718$ is 2, whereas the leading digit of $-e^e = -15.15$ is 1. Note that (1) is heavily skewed towards the smaller digits: For instance, the leading digit is almost seven times as likely to equal 1 (probability $\log 2 = 30.10\,\%$) as it is to equal 9 (probability $\log \frac{10}{9} = 4.57\,\%$).

Ever since first recorded by Newcomb [36] in 1881 and re-discovered by Benford [3] in 1938, examples of data and systems conforming to (1) in one form or another have been discussed extensively, for instance in real-life data (e.g. [19, 41]), stochastic processes (e.g. [44]), and in deterministic sequences (e.g. (n!) and the prime numbers [17]). There now exists a large body of literature devoted to the mechanisms whereby mathematical objects, such as e.g. sequences or random variables, do or do not satisfy (1) or variants thereof, see also Fig. 1. Beyond mathematics, BL has found diverse applications throughout the sciences. Given that the ubiquity of BL in these fields is still somewhat of a mystery [8], some BL-based tools (e.g. for fraud detection in

1	121393	20365011074	3416454622906707
1	196418	32951280099	5527939700884757
2	317811	53316291173	8944394323791464
3	514229	86267571272	14472334024676221
5	832040	139583862445	23416728348467685
8	1346269	225851433717	37889062373143906
13	2178309	365435296162	61305790721611591
21	3524578	591286729879	99194853094755497
34	5702887	956722026041	160500643816367088
55	9227465	1548008755920	259695496911122585
89	14930352	2504730781961	420196140727489673
144	24157817	4052739537881	679891637638612258
233	39088169	6557470319842	1100087778366101931
377	63245986	10610209857723	1779979416004714189
610	102334155	17167680177565	2880067194370816120
987	165580141	27777890035288	4660046610375530309
1597	267914296	44945570212853	7540113804746346429
2584	433494437	72723460248141	12200160415121876738
4181	701408733	117669030460994	19740274219868223167
6765	1134903170	190392490709135	31940434634990099905
10946	1836311903	308061521170129	51680708854858323072
17711	2971215073	498454011879264	83621143489848422977
28657	4807526976	806515533049393	135301852344706746049
46368	7778742049	1304969544928657	218922995834555169026
75025	12586269025	2111485077978050	354224848179261915075

	#	*exact* BL
1	30	30.10
2	18	17.60
3	13	12.49
4	9	9.69
5	8	7.91
6	6	6.69
7	5	5.79
8	7	5.11
9	4	4.57

Fig. 2 Already the first one-hundred Fibonacci numbers conform to BL quite well

tax, census, election or image processing data) have proved remarkably successful in practice. This in turn has triggered further research on the many unique features of BL [22, 23, 45]. It still rings true that, as Raimi [39] observed almost 40 years ago,

> [t]his particular logarithmic distribution of the first digits, while not universal, is so common and yet so surprising at first glance that it has given rise to a varied literature, among the authors of which are mathematicians, statisticians, economists, engineers, physicists and amateurs.

As of this writing, an online database [4] devoted exclusively to BL lists more than 750 references.

Due to their important role as elementary models throughout science, linear difference and differential equations have, from very early on, been studied for their conformance to (1). A simple early example [11, 20, 28, 48] is the sequence $(x_n) = (F_n) = (1, 1, 2, 3, 5, \ldots)$ of Fibonacci numbers, i.e. $F_1 = F_2 = 1$ and $F_n = F_{n-1} + F_{n-2}$ for all $n \geq 3$, which satisfies (1) in the sense that

$$\lim_{N\to\infty} \frac{\#\{1 \leq n \leq N : \text{leading digit of } x_n = d\}}{N} = \log(1 + d^{-1}), \quad \forall d = 1, 2, \ldots, 9, \tag{2}$$

see also Fig. 2. Another simple case in point is $(x_n) = (2^n)$ for which (2) also holds [2, §24.4]. On the other hand, the sequence of primes $(x_n) = (2, 3, 5, 7, 11, \ldots)$ does not satisfy (2), as was in essence observed already by [47], yet may conform to BL in some weaker sense [14, 42].

Both positive examples mentioned above, i.e. the sequences (F_n) and (2^n), are obviously solutions of (very simple) *autonomous linear difference equations*. Building on earlier work, notably [5, 26, 35, 43], it is the purpose of this article to provide a comprehensive overview of the theory of BL for such equations. Thus the central question throughout is as follows: Given $d \in \mathbb{N}$ and real numbers $a_1, a_2, \ldots, a_{d-1}, a_d$ with $a_d \neq 0$, consider the (autonomous, dth order) linear difference equation

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_{d-1} x_{n-d+1} + a_d x_{n-d}, \quad n \geq d+1. \qquad (3)$$

Under which conditions on $a_1, a_2, \ldots, a_{d-1}, a_d$, and presumably also on the initial values $x_1, x_2, \ldots, x_d$, does the solution (x_n) of (3) satisfy (2)? Early work in this regard seems to have led merely to *sufficient* conditions that are either restrictive or difficult to state. By contrast, two of the main results presented here (Corollary 3.7 and Theorem 4.11) provide easy-to-state, *necessary and sufficient* conditions for every non-trivial solution of (3) to conform to (1) in a sense much stronger than (2). The classical results in the literature are then but simple special cases.

The organisation of this article is as follows. Section 2 introduces the formal definitions and analytic tools required for the analysis. In Sect. 3, difference equations (3), as well as the matrices associated with them are studied under the additional assumption of *positivity*. Though restrictive, this assumption holds for some important applications, and it yields a particularly simple answer to the central question raised earlier. Dropping the positivity assumption, Sect. 4 studies the case of general equations and matrices. The emergence of *resonances*, the key problem in the general case, is dealt with by means of a tailor-made definition (Definition 4.2). While the main results (Theorems 4.5 and 4.11) are stated in full generality, proofs are given here only under an additional non-degeneracy condition (and the interested reader is referred to the authors' forthcoming work [6] for complete proofs). Finally, Sect. 5 demonstrates how the presented results can be used to explain the "cancellation of resonance" phenomenon first observed in the context of finite-state Markov chains [10].

2 Basic Definitions and Tools

Throughout, the following, mostly standard notation is adhered to. The sets of natural, non-negative integer, integer, rational, positive real, real, and complex numbers are symbolised by $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}^+$, $\mathbb{R}$, and $\mathbb{C}$, respectively. The cardinality of any finite set $Z \subset \mathbb{C}$ is $\#Z$. The real part, imaginary part, complex conjugate, and absolute value (modulus) of $z \in \mathbb{C}$ is denoted by $\Re z$, $\Im z$, $\overline{z}$, and $|z|$, respectively. Let $\mathbb{S} := \{z \in \mathbb{C} : |z| = 1\}$. The argument $\arg z$ of $z \neq 0$ is understood to be the unique number in $(-\pi, \pi]$ for which $z = |z| e^{\imath \arg z}$; for convenience, let $\arg 0 := 0$. For any set $Z \subset \mathbb{C}$ and number $w \in \mathbb{C}$, define $wZ := \{wz : z \in Z\}$. Thus for instance $w\mathbb{S} = \{z \in \mathbb{C} : |z| = |w|\}$ for every $w \in \mathbb{C}$. Given $Z \subset \mathbb{C}$, denote by $\operatorname{span}_{\mathbb{Q}} Z$ the smallest subspace of $\mathbb{C}$ (over $\mathbb{Q}$) containing Z; equivalently, if $Z \neq \varnothing$ then $\mathrm{span}_{\mathbb{Q}} Z$ is the set of all *finite* rational linear combinations of elements of Z, i.e.

$$\mathrm{span}_{\mathbb{Q}} Z = \left\{ \rho_1 z_1 + \rho_2 z_2 + \cdots + \rho_n z_n : n \in \mathbb{N}, \rho_1, \rho_2, \ldots, \rho_n \in \mathbb{Q}, z_1, z_2, \ldots, z_n \in Z \right\};$$

note that $\mathrm{span}_{\mathbb{Q}} \varnothing = \{0\}$. For every $x \in \mathbb{R}^+$, $\log x$ and $\ln x$ are, respectively, the base-10 and the natural (base-e) logarithm of x; for convenience, set $\log 0 := \ln 0 := 0$. For every $x \in \mathbb{R}$, denote by $\lfloor x \rfloor$ the largest integer not larger than x, hence $\langle x \rangle := x - \lfloor x \rfloor$ is the non-integer (or fractional) part of x.

Given $x \in \mathbb{R} \backslash \{0\}$, there exists a unique $S(x) \in [1, 10)$ such that $|x| = S(x) 10^k$ for some (necessarily unique) integer k. The number $S(x)$ is the (*decimal*) *significand* of x. Note that

$$S(x) = 10^{\langle \log |x| \rangle}, \quad \forall x \in \mathbb{R} \backslash \{0\};$$

for convenience let $S(0) := 0$. For $x \neq 0$, the integer $\lfloor S(x) \rfloor \in \{1, 2, \ldots, 9\}$ is the *first significant* (*decimal*) *digit* of x. More generally, for every $m \in \mathbb{N}$, the integer $\lfloor 10^{m-1} S(x) \rfloor - 10 \lfloor 10^{m-2} S(x) \rfloor \in \{0, 1, \ldots, 9\}$ is the mth *significant* (*decimal*) *digit* of x, see e.g. [7, Prop. 2.5].

Throughout this article, conformance to (1) for solutions of difference equations is studied using the following terminology.

Definition 2.1 A sequence (x_n) of real numbers is a *Benford sequence*, or simply *Benford*, if

$$\lim_{N \to \infty} \frac{\#\{1 \leq n \leq N : S(x_n) \leq t\}}{N} = \log t, \quad \forall t \in [1, 10). \tag{4}$$

Note that every Benford sequence (x_n) satisfies (2). For the purpose of this work, the following well-known characterization of the Benford property is indispensable.

Proposition 2.2 [17, Thm. 1] *A sequence* (x_n) *is Benford if and only if the sequence* $(\log |x_n|)$ *is uniformly distributed modulo one.*

The term *uniformly distributed modulo one* is henceforth abbreviated *u.d.* mod 1. In view of Proposition 2.2, a few basic facts regarding uniform distribution of sequences are used throughout; for an authoritative overall account on the subject, the reader is referred to [29].

Proposition 2.3 [29, Sect. I.2] *The following statements are equivalent for any sequence* (y_n) *in* $\mathbb{R}$*:*

(i) (y_n) *is u.d.* mod 1*;*
(ii) *For every* $\varepsilon > 0$ *there exists a sequence* (z_n) *that is u.d.* mod 1*, and*

$$\limsup_{N \to \infty} \frac{\#\{1 \leq n \leq N : |y_n - z_n| > \varepsilon\}}{N} < \varepsilon;$$

(iii) *Whenever* (z_n) *is convergent then* $(y_n + z_n)$ *is u.d.* mod 1*;*

(iv) (py_n) *is u.d.* mod 1 *for every non-zero integer p;*
(v) $(y_n + \alpha \log n)$ *is u.d.* mod 1 *for every* $\alpha \in \mathbb{R}$.

One of the simplest yet also most fundamental examples of a sequence u.d. mod 1 is $(n\vartheta)$ with $\vartheta \in \mathbb{R}\backslash\mathbb{Q}$. The following, therefore, is an immediate consequence of Propositions 2.2 and 2.3.

Proposition 2.4 *Let* (x_n) *be a sequence in* $\mathbb{R}$, *and* $\alpha \in \mathbb{R}\backslash\{0\}$. *If* $\lim_{n\to\infty} x_n/\alpha^n$ *exists (in* $\mathbb{R}$*) and is non-zero, then* (x_n) *is Benford if and only if* $\log|\alpha|$ *is irrational.*

Example 2.5 Since $\log 2$ is irrational (even transcendental), (2^n) is Benford, and so is the sequence (F_n) of Fibonacci numbers because, with $\varphi = \frac{1}{2}(1+\sqrt{5})$, $\lim_{n\to\infty} F_n/\varphi^n = 1/\sqrt{5} \neq 0$, and $\log\varphi$ is irrational as well. $\triangleleft$

Remark The Benford property can be studied w.r.t. any integer base $b \geq 2$, simply by replacing the decimal significand $S(x)$ in (4) with the base-b significand $S_b(x) = b^{\langle \log_b |x| \rangle}$, where $\log_b$ denotes the base-b logarithm. With the obvious modifications, the results in this work carry over to arbitrary base $b \in \mathbb{N}\backslash\{1\}$, cf. [5–7]. For the sake of clarity, however, only the familiar case $b = 10$ is considered from now on.

When studying solutions of linear difference equations, sequences of a particular form are often encountered, and the following lemma clarifies their properties.

Lemma 2.6 *Let* $\alpha \in \mathbb{R}$, $z \in \mathbb{C}\backslash\{0\}$, *and* (z_n) *a sequence in* $\mathbb{C}$ *with* $\lim_{n\to\infty} z_n = 0$. *If* $\vartheta_1, \vartheta_2 \in \mathbb{R}$ *are irrational then the following statements are equivalent:*

(i) $\vartheta_1 \notin \operatorname{span}_{\mathbb{Q}}\{1, \vartheta_2\}$;
(ii) *The sequence* (y_n) *with*

$$y_n = n\vartheta_1 + \alpha \log n + \log\left|\Re\left(ze^{\imath\pi n\vartheta_2} + z_n\right)\right|, \quad n \in \mathbb{N},$$

is u.d. mod 1.

Proof If $\vartheta_1 \notin \operatorname{span}_{\mathbb{Q}}\{1, \vartheta_2\}$ then $1, \vartheta_1, \vartheta_2$ are rationally independent, and [5, Lem. 2.9] shows that (y_n) is u.d. mod 1. On the other hand, if $\vartheta_1 \in \operatorname{span}_{\mathbb{Q}}\{1, \vartheta_2\}$ then $k_1\vartheta_1 = k_0 + k_2\vartheta_2$, where k_0, k_1, k_2 are appropriate integers with $k_1k_2 \neq 0$; assume w.l.o.g. that $k_1 > 0$. Consider now the sequence (η_n) with

$$\eta_n = n\vartheta_1 + \log\left|\Re(ze^{\imath\pi n\vartheta_2})\right| + \tfrac{k_2}{k_1}\left(\tfrac{1}{2} + \tfrac{\arg z}{\pi}\right) - \log|z|, \quad n \in \mathbb{N}.$$

If (y_n) was u.d. mod 1, then so would be (η_n), and hence also $(k_1\eta_n)$, by Proposition 2.3. Moreover, for every $n \in \mathbb{N}$,

$$\begin{aligned}\langle k_1\eta_n \rangle &= \left\langle k_2\left(n\vartheta_2 + \tfrac{1}{2} + \tfrac{\arg z}{\pi}\right) + k_1 \log\left|\sin\left(\pi\left(n\vartheta_2 + \tfrac{1}{2} + \tfrac{\arg z}{\pi}\right)\right)\right|\right\rangle \\ &= \left\langle f\left(n\vartheta_2 + \tfrac{1}{2} + \tfrac{\arg z}{\pi}\right)\right\rangle,\end{aligned}$$

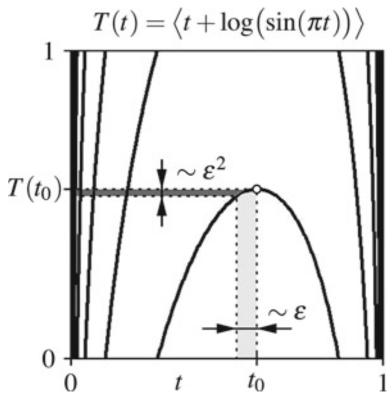

Fig. 3 The map T does not preserve $\lambda_{0,1}$, see the proof of Lemma 2.6

with the measurable function $f : \mathbb{R} \to \mathbb{R}$ given by $f(t) = k_2 t + k_1 \log|\sin(\pi t)|$. Note that $f(t+1) - f(t) \in \mathbb{Z}$ for all $t \in \mathbb{R}$, and so f induces the measurable map $T := \langle f \rangle$ on [0, 1). Recall now that the sequence $(n\vartheta_2 + \frac{1}{2} + \frac{\arg z}{\pi})$ is u.d. mod 1 because ϑ_2 is irrational. For every continuous, 1-periodic function $g : \mathbb{R} \to \mathbb{R}$, therefore,

$$\frac{1}{N}\sum_{n=1}^{N} g\left(\langle k_1\eta_n\rangle\right) = \frac{1}{N}\sum_{n=1}^{N} g \circ f\left(n\vartheta_2 + \tfrac{1}{2} + \tfrac{\arg z}{\pi}\right) \overset{N\to\infty}{\longrightarrow} \int_0^1 g \circ f(t)\,\mathrm{d}t$$
$$= \int_{[0,1]} g \circ T\,\mathrm{d}\lambda_{0,1} = \int_{[0,1]} g\,\mathrm{d}\left(\lambda_{0,1} \circ T^{-1}\right),$$

because $g \circ f$ is Riemann integrable on [0, 1]. On the other hand, if $(k_1\eta_n)$ was u.d. mod 1 then $\lim_{N\to\infty} N^{-1}\sum_{n=1}^{N} g(\langle k_1\eta_n\rangle) = \int_{[0,1]} g\,\mathrm{d}\lambda_{0,1}$ for every g, and hence $\lambda_{0,1} \circ T^{-1} = \lambda_{0,1}$. However, it is intuitively clear that the latter equality of measures does not hold. To see this formally, note that f is smooth on (0, 1) and has a (unique) non-degenerate maximum at some $0 < t_0 < 1$. Thus if $\lambda_{0,1} \circ T^{-1} = \lambda_{0,1}$ then, for all $\varepsilon > 0$ sufficiently small,

$$\frac{f(t_0-\varepsilon) - f(t_0-2\varepsilon)}{\varepsilon} = \frac{\lambda_{0,1}\left([T(t_0-2\varepsilon), T(t_0-\varepsilon)]\right)}{\varepsilon}$$
$$= \frac{\lambda_{0,1} \circ T^{-1}\left([T(t_0-2\varepsilon), T(t_0-\varepsilon)]\right)}{\varepsilon}$$
$$\geq \frac{\lambda_{0,1}([t_0-2\varepsilon, t_0-\varepsilon])}{\varepsilon} = 1,$$

which is impossible since $f'(t_0) = 0$, see also Fig. 3 which depicts the special case $k_1 = k_2 = 1$. Hence $(k_1\eta_n)$ is not u.d. mod 1, and neither are (η_n) and (y_n). □

Although it would be possible to study the Benford property of solutions of (3) directly, the analysis in subsequent sections becomes more transparent by means of a standard matrix-vector approach. To this end, associate with (3) the matrix

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_{d-1} & a_d \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \in \mathbb{R}^{d \times d}, \tag{5}$$

which is invertible as $a_d \neq 0$, and recall that, given initial values $x_1, x_2, \ldots, x_d \in \mathbb{R}$, the solution of (3) can be expressed neatly in the form

$$x_n = e_d^\top A^n y, \quad \text{where } y = A^{-1} \begin{bmatrix} x_d \\ \vdots \\ x_2 \\ x_1 \end{bmatrix} \in \mathbb{R}^d ; \tag{6}$$

here $e_1, e_2, \ldots, e_d$ represent the standard basis of $\mathbb{R}^d$, and $x^\top$ denotes the transpose of $x \in \mathbb{R}^d$, with $x^\top y$ being understood simply as the real number $\sum_{j=1}^d x_j y_j$. In what follows, therefore, the following, more general question suggested by (6) will be addressed: Under which conditions is $\left(x^\top A^n y\right)$ Benford, where A is any fixed real $d \times d$-matrix and $x, y \in \mathbb{R}^d$ are given vectors? Note that specifically choosing $x = e_j$ and $y = e_k$, with $j, k \in \{1, 2, \ldots, d\}$, simply yields $e_j^\top A^n e_k = [A^n]_{jk}$, i.e. the entry of A^n at the position (j, k). Also, if $A \in \mathbb{R}^{d \times d}$ is given by (5) then *every* sequence $\left(x^\top A^n y\right)$ solves (3), and (6) establishes a one-to-one correspondence between all sequences of the form $\left(e_d^\top A^n y\right)$ and *all* solutions of (3).

In the analysis of powers of matrices in the subsequent sections, d always is a fixed but usually unspecified positive integer. For every $x \in \mathbb{R}^d$, the number $|x| \geq 0$ is the *Euclidean norm* of x, i.e. $|x| = \sqrt{x^\top x} = \sqrt{\sum_{j=1}^d x_j^2}$. A vector $x \in \mathbb{R}^d$ is a *unit* vector if $|x| = 1$. The $d \times d$-identity matrix is I_d. For every matrix $A \in \mathbb{R}^{d \times d}$, its spectrum, i.e. the set of its eigenvalues, is denoted by $\sigma(A)$. Thus $\sigma(A) \subset \mathbb{C}$ is non-empty, contains at most d numbers and is symmetric w.r.t. the real axis, i.e. all non-real elements of $\sigma(A)$ come in complex-conjugate pairs. The number $r_\sigma(A) := \max\{|\lambda| : \lambda \in \sigma(A)\} \geq 0$ is the *spectral radius* of A. Note that $r_\sigma(A) > 0$ unless A is *nilpotent*, i.e. unless $A^N = 0$ for some $N \in \mathbb{N}$. For every $A \in \mathbb{R}^{d \times d}$, the number $|A|$ is the (*spectral*) *norm* of A, as induced by $|\cdot|$, i.e. $|A| = \max\{|Ax| : |x| = 1\}$. It is well-known that $|A| = \sqrt{r_\sigma(A^\top A)} \geq r_\sigma(A)$.

3 A Simple Special Case: Positive Matrices

The analysis of sequences $(x^\top A^n y)$ is especially simple if the matrix A or one of its powers happens to be positive. Recall that $A \in \mathbb{R}^{d \times d}$ is *positive*, in symbols $A > 0$, if $[A]_{jk} > 0$ for every $j, k \in \{1, 2, \ldots, d\}$. The following classical result, due to O

Perron, lists some of the remarkable properties of positive matrices, as they pertain to the present section. For a concise formulation, call $x \in \mathbb{R}^d$ *positive* (*non-negative*), in symbols $x > 0$ ($x \geq 0$), if $x_j > 0$ ($x_j \geq 0$) for every $j \in \{1, 2, \ldots, d\}$.

Proposition 3.1 [25, Sect. 8.2] *Assume that $A \in \mathbb{R}^{d\times d}$ is positive. Then:*

(i) *The number $r_\sigma(A) > 0$ is an (algebraically) simple eigenvalue of A, i.e. a simple root of the characteristic polynomial of A;*
(ii) $|\lambda| < r_\sigma(A)$ *for every eigenvalue $\lambda \neq r_\sigma(A)$ of A;*
(iii) *There exists a positive eigenvector q, unique up to multiplication by a positive number, corresponding to the eigenvalue $r_\sigma(A)$;*
(iv) *The limit $Q := \lim_{n\to\infty} A^n/r_\sigma(A)^n$ exists, and $Q > 0$ satisfies $Q^2 = Q$ as well as $AQ = QA = r_\sigma(A)Q$. (In fact, Q is a rank-one projection with $Qq = q$.)*

Recall that (α^n) with $\alpha > 0$ is Benford if and only if $\log\alpha$ is irrational. The following is a generalization of this simple fact to arbitrary dimension. Informally put, it asserts that as far as the Benford property is concerned, matrices with some positive power behave just like the one-dimensional sequence $\big(r_\sigma(A)^n\big)$.

Theorem 3.2 *Let A be a real $d \times d$-matrix, and assume that $A^N > 0$ for some $N \in \mathbb{N}$. Then the following four statements are equivalent:*

(i) *The number $\log r_\sigma(A)$ is irrational;*
(ii) *The sequence $(x^\top A^n y)$ is Benford for every $x, y \neq 0$ with $x \geq 0$ and $y \geq 0$;*
(iii) *The sequence $(|A^n x|)$ is Benford for every $x \neq 0$ with $x \geq 0$;*
(iv) *The sequence $(|A^n|)$ is Benford.*

Proof Since $A^N > 0$, the number $r_\sigma(A^N) = r_\sigma(A)^N > 0$ is an algebraically simple eigenvalue of A^N, by Proposition 3.1. It follows that exactly one of the two numbers $r_\sigma(A) > 0$ and $-r_\sigma(A) < 0$ is an algebraically simple eigenvalue of A. Denote this eigenvalue by λ_0, and let P be the spectral projection associated with it, that is, $P = bc^\top/b^\top c$ where b, c are eigenvectors of, respectively, A and $A^\top$ corresponding to the eigenvalue λ_0, i.e. $Ab = \lambda_0 b$ and $A^\top c = \lambda_0 c$. Thus $P^2 = P$ and $AP = PA = \lambda_0 P$. Moreover, the matrix $R := A - \lambda_0 P$ clearly satisfies $AR = RA$ and $PR = RP = 0$, and hence

$$A^n = \lambda_0^n P + R^n, \quad \forall n \in \mathbb{N}. \tag{7}$$

Since $|\lambda| < |\lambda_0| = r_\sigma(A)$ for every eigenvalue λ of R, $\lim_{n\to\infty} R^n/r_\sigma(A)^n = 0$, and an evaluation of (7) along even multiples of N yields

$$\lim_{n\to\infty} \frac{(A^N)^{2n}}{r_\sigma(A^N)^{2n}} = \lim_{n\to\infty} \left(\frac{\lambda_0^{2nN}}{r_\sigma(A)^{2nN}} P + \frac{R^{2nN}}{r_\sigma(A)^{2nN}} \right) = P\,.$$

This shows that $P = Q > 0$, with Q according to Proposition 3.1(iv) applied to A^N.

With these preparations, the asserted equivalences are now easily established. Indeed, given any $x, y \neq 0$ with $x \geq 0$ and $y \geq 0$, the vector Qy is positive, and

$$\frac{|x^\top A^n y|}{r_\sigma(A)^n} = \frac{|\lambda_0^n x^\top Qy + x^\top R^n y|}{r_\sigma(A)^n} = \left| x^\top Qy + \frac{x^\top R^n y}{\lambda_0^n} \right| \stackrel{n\to\infty}{\longrightarrow} x^\top Qy > 0, \quad (8)$$

together with Proposition 2.4, shows that $(x^\top A^n y)$ is Benford if and only if $\log r_\sigma(A)$ is irrational. A similar argument applies to $(|A^n x|)$, as

$$\frac{|A^n x|}{r_\sigma(A)^n} = \frac{|\lambda_0^n Qx + R^n x|}{r_\sigma(A)^n} = \left| Qx + \frac{R^n x}{\lambda_0^n} \right| \stackrel{n\to\infty}{\longrightarrow} |Qx| > 0,$$

whenever $x \geq 0$, $x \neq 0$, and also to $(|A^n|)$, as

$$\frac{|A^n|}{r_\sigma(A)^n} = \frac{|\lambda_0^n Q + R^n|}{r_\sigma(A)^n} = \left| Q + \frac{R^n}{\lambda_0^n} \right| \stackrel{n\to\infty}{\longrightarrow} |Q| > 0. \qquad \square$$

Remark The proof of Theorem 3.2 shows that if $\log r_\sigma(A)$ is rational then $(x^\top A^n y)$ and $(|A^n x|)$ are not Benford for *any* $x, y \geq 0$, and neither is $(|A^n|)$ Benford. Also, in (iii) and (iv), the Euclidean norm $|\cdot|$ can be replaced by any norm on, respectively, $\mathbb{R}^d$ and $\mathbb{R}^{d\times d}$.

Corollary 3.3 *Let $A \in \mathbb{R}^{d\times d}$, and assume that $A^N > 0$ for some $N \in \mathbb{N}$. Then, for every $j, k \in \{1, 2, \ldots, d\}$, the sequence $([A^n]_{jk})$ is Benford if and only if $\log r_\sigma(A)$ is irrational.*

Example 3.4 The matrix associated with the Fibonacci recursion

$$x_n = x_{n-1} + x_{n-2}, \quad n \geq 3, \qquad (9)$$

is $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, with $r_\sigma(A) = \varphi = \frac{1}{2}(1 + \sqrt{5})$. While A is non-negative, i.e. $[A]_{jk} \geq 0$ for all j, k, but fails to be positive, the matrix A^2 is positive, and so is A^n for every $n \geq 2$. Since $\log r_\sigma(A)$ is irrational (even transcendental), every entry of (A^n) is Benford. This is consistent with the fact that

$$A^n = \begin{bmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{bmatrix}, \quad n \geq 2,$$

and the sequence (F_n) is Benford.

Consider now the sequence (x_n) with $x_n = e_1^\top A^n (3e_2 - e_1)$. Recall that (x_n) thus defined also solves (9). However, since $3e_2 - e_1$ is not non-negative, Theorem 3.2 does not allow to decide whether $(x_n) = (2, 1, 3, 4, 7, \ldots)$, traditionally referred to as the sequence of *Lucas numbers* and denoted (L_n), is Benford. Corollary 3.7 below shows very easily that this is indeed the case. ◁

Example 3.5 Consider the (symmetric) matrix

$$B = \begin{bmatrix} -3 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 6 \end{bmatrix},$$

the characteristic polynomial of which is

$$p_B(\lambda) = \det(B - \lambda I_3) = -\lambda^3 + 3\lambda^2 + 20\lambda - 3\,.$$

It is readily confirmed that p_B has three different real roots. If $\lambda = \pm 10^{m/n}$ was a root of p_B with any relatively prime $m \in \mathbb{Z}$ and $n \in \mathbb{N}$, then $n \le 3$, and 10^m would divide $|\det B^n| = |\det B|^n = 3^n$, hence $m = 0$, that is, $\lambda = \pm 1$. But $p_B(\pm 1) = \pm 19 \neq 0$. It follows that $r_\sigma(B)$, albeit algebraic, is not a rational power of 10, and so $\log r_\sigma(B)$ is irrational (even transcendental). Moreover, B^n contains both positive and negative entries for $n = 1, 2, \ldots, 7$, yet

$$B^8 = \begin{bmatrix} 13841 & 1929 & 37034 \\ 1929 & 56662 & 335235 \\ 37034 & 335235 & 2031038 \end{bmatrix} > 0,$$

hence Theorem 3.2 and Corollary 3.3 apply. In particular, every entry of (B^n) is Benford. Note that the actual value of $r_\sigma(B)$,

$$r_\sigma(B) = 1 + \tfrac{2\sqrt{69}}{3} \cos\left(\tfrac{1}{3} \arccos \tfrac{57\sqrt{69}}{1058}\right) = 6.165,$$

is not needed at all to draw this conclusion. ◁

Example 3.6 When $A > 0$ and $\log r_\sigma(A)$ is irrational, the sequence $(x^\top A^n y)$ may nevertheless *not* be Benford for some non-zero $x, y \in \mathbb{R}^d$. By Theorem 3.2, such x, y cannot both be non-negative. For instance, the matrix $A = \begin{bmatrix} 5 & 15 \\ 15 & 5 \end{bmatrix}$ is positive, and $\log r_\sigma(A) = 1 + \log 2$ is irrational, yet $\bigl(e_1^\top A^n(e_1 - e_2)\bigr) = \bigl((-10)^n\bigr)$ is not Benford. On the other hand, even if $r_\sigma(B)$ is rational, $(x^\top B^n y)$ may be Benford for *some* $x, y \in \mathbb{R}^d$. Again, x, y cannot both be non-negative, by virtue of Theorem 3.2. Concretely, $B = \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix} > 0$ has $\log r_\sigma(B) = 1$ rational, yet $\bigl(e_1^\top B^n(e_1 - e_2)\bigr) = (2^n)$ is Benford. ◁

Corollary 3.7 *Let (x_n) be a solution of the linear difference equation*

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_{d-1} x_{n-d+1} + a_d x_{n-d}, \quad n \ge d + 1,$$

with $a_1, a_2, \ldots, a_{d-1}, a_d > 0$. Assume that the numbers $x_1, x_2, \ldots, x_d$ are non-negative, and at least one is positive. Then (x_n) is Benford if and only if $\log \zeta$ is *irrational, where* $z = \zeta$ *is the right-most root of* $z^d = a_1 z^{d-1} + a_2 z^{d-2} + \cdots + a_{d-1} z + a_d$.

Proof The associated matrix A according to (5) is non-negative, and $A^n > 0$ for $n \ge d$. Moreover, A has the characteristic polynomial

$$p_A(\lambda) = (-1)^d (\lambda^d - a_1 \lambda^{d-1} - a_2 \lambda^{d-2} - \cdots - a_{d-1}\lambda - a_d)\,.$$

Since $x_n = x^\top A^{n-1} y$ with $x = e_d \ge 0$ and $y = \sum_{j=1}^d x_{d+1-j} e_j \ge 0$, the claim follows directly from Theorem 3.2. □

Example 3.8 Every solution of (9) with $x_1 x_2 > 0$ is Benford. (For the case $x_1 < 0$ simply note that $(-x_n)$ is a solution of (9) as well.) Evidently, this includes the Fibonacci sequence, where $x_1 = x_2 = 1$, but also the Lucas numbers, where $x_1 = 2$, $x_2 = 1$. As they stand, however, Theorem 3.2 and Corollary 3.7 do not allow to decide whether the solution of (9) with, say, $x_1 = 2$, $x_2 = -3$ is Benford.

More generally, every solution (x_n) with $x_1 x_2 > 0$ of

$$x_n = a_1 x_{n-1} + a_2 x_{n-2}, \quad n \ge 3, \tag{10}$$

where a_1, a_2 are positive *integers*, is Benford if and only if $10^{2m} - a_2 \ne a_1 \cdot 10^m$ for every $m = 0, 1, \ldots, \lfloor \log(a_1 + a_2) \rfloor$. Again, this leaves open the question regarding the Benford property of solutions of (10) with $x_1 x_2 < 0$. The results of the next section allow to settle this question without any further calculation: Except for the trivial case $x_1 = x_2 = 0$, *every* solution of (10) is Benford if and only if

$$|10^{2m} - a_2| \ne a_1 \cdot 10^m, \quad \forall m = 0, 1, \ldots, \lfloor \log(a_1 + a_2) \rfloor\,. \tag{11}$$

For the Fibonacci recursion (9), for instance, (11) reduces to $|1 - 1| \ne 1$, which is obviously true. Thus, apart from $x_n \equiv 0$, *every* solution of (9) is Benford. ◁

The following examples aim at illustrating the scope and limitations of Theorem 3.2. Although the latter is easy to state and prove, and quite useful in a variety of situations, its overall applicability is somewhat limited because

- it does not apply in general if the matrix in question fails to have a positive power, see Example 3.9;
- even if it applies, the Benford property of individual solutions of a linear difference equation (3), or equivalently of sequences $(x^\top A^n y)$ with A according to (5) and *arbitrary* $x, y \in \mathbb{R}^d$, is generally unrelated to the Benford property of $(x^\top A^n y)$ with *non-negative* x, y, see Example 3.10;
- it does not apply to various sequences that are closely related to (A^n) and often of interest in their own right, for instance $(A^{n+1} - r_\sigma(A) A^n)$, see Example 3.11.

In view of these limitations, in the next section the Benford property is studied more generally for sequences $(x^\top A^n y)$ with *arbitrary* $A \in \mathbb{R}^{d\times d}$ and $x, y \in \mathbb{R}^d$.

Example 3.9 Theorem 3.2 may fail if $A \in \mathbb{R}^{d\times d}$ does not have a positive power. Simply consider the (non-negative) matrix $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, for which $\log r_\sigma(A) = \log 2$ is irrational, yet $([A^n]_{jk})$ is constant and hence not Benford except for $j = k = 1$. Neither is $(|A^n e_2|)$ Benford. Thus the implications (i)$\Rightarrow$(ii) and (i)$\Rightarrow$(iii) in Theorem 3.2 do not even hold for non-negative matrices. As will be seen in the next section, however, (ii)$\Rightarrow$(i) and (iii)$\Rightarrow$(i) remain true for arbitrary (non-nilpotent) matrices in that if $(x^\top A^n y)$ or $(|A^n x|)$ is, for every $x, y \in \mathbb{R}^d$, either Benford or vanishes for all $n \geq d$ then $\log r_\sigma(A)$ is irrational. Similarly, if A does not have a positive power then $(|A^n|)$ may not be Benford even when $\log r_\sigma(A)$ is irrational, see Example 4.10.

Note also that even if B does not have any positive power, all entries of (B^n), or in fact all non-trivial sequences $(x^\top B^n y)$, may nevertheless be Benford, as the example $B = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ shows, for which

$$B^n = 2^{n-1} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad n \in \mathbb{N}. \qquad \lhd$$

Example 3.10 Consider the difference equation

$$x_n = \tfrac{1}{2}(x_{n-1} + x_{n-2}), \quad n \geq 3, \tag{12}$$

and the associated matrix

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{bmatrix}. \tag{13}$$

Similarly to Example 3.4, $A \geq 0$ and $A^2 > 0$. In addition, A evidently has the property that the entries in each of its rows add up to 1. Thus A is a (*row-*) *stochastic* matrix. It is well known (and easy to see) that $r_\sigma(A) = 1$ for every (row- or column-) stochastic matrix. According to Theorem 3.2, none of the sequences $(x^\top A^n y)$ with $x, y \geq 0$ is Benford. In fact, a short calculation yields

$$A^n = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} + \frac{(-\frac{1}{2})^n}{3}\begin{bmatrix} 1 & -1 \\ -2 & 2 \end{bmatrix}, \quad n \in \mathbb{N}_0, \tag{14}$$

showing that each sequence $(x^\top A^n y)$ converges to a finite limit (which is positive unless $x = 0$ or $y = 0$) and hence cannot be Benford. Recall that each such sequence is a solution of (12). On the other hand, the solution of (12) with $x_1 = -2$, $x_2 = 1$ is $(x_n) = \big((-\frac{1}{2})^{n-2}\big)$ and clearly Benford. Thus a solution of a linear difference equation may be Benford even if the associated matrix A has a positive power but does not satisfy (i)–(iv) in Theorem 3.2.

To see that the reverse situation—some solution of a difference equation is not Benford despite the associated matrix having a positive power and satisfying (i)–(iv) in Theorem 3.2—can also occur, consider

$$x_n = 19x_{n-1} + 20x_{n-2}, \quad n \geq 3. \tag{15}$$

The solution of (15) with $x_1 = -1, x_2 = 1$ is $\big((-1)^n\big)$ and hence not Benford. On the other hand, the associated matrix $B = \begin{bmatrix} 19 & 20 \\ 1 & 0 \end{bmatrix}$ has a positive power as $B^2 > 0$, and $\log r_\sigma(B) = 1 + \log 2$ is irrational. ◁

Example 3.11 If $A^N > 0$ for some $N \in \mathbb{N}$ then exactly one of the two numbers $\lambda_0 = r_\sigma(A) > 0$ or $\lambda_0 = -r_\sigma(A) < 0$ is an eigenvalue of A, and $Q := \lim_{n\to\infty} A^n/\lambda_0^n$ exists and is a positive matrix. This fact, which has been instrumental in the proof of Theorem 3.2, is of particular interest in the case of A being a stochastic matrix, i.e. for $A \geq 0$ and each row (or column) of A summing up to 1. In this case, $\lambda_0 = r_\sigma(A) = 1$, and hence $Q = \lim_{n\to\infty} A^n$. Often, one is interested in the (Benford) properties of $(A^n - Q)$ and $(A^{n+1} - A^n)$. Entries of these sequences may well be Benford, notwithstanding the fact that Theorem 3.2 does not apply and $\log r_\sigma(A) = 0$ is rational. For instance, with A from (13), it follows from (14) that

$$A^n - Q = \frac{(-\frac{1}{2})^n}{3} \begin{bmatrix} 1 & -1 \\ -2 & 2 \end{bmatrix}, \quad n \in \mathbb{N}_0,$$

but also

$$A^{n+1} - A^n = (-\tfrac{1}{2})^{n+1} \begin{bmatrix} 1 & -1 \\ -2 & 2 \end{bmatrix}, \quad n \in \mathbb{N}_0,$$

and hence every entry of both $(A^n - Q)$ and $(A^{n+1} - A^n)$ is Benford. In general, note that $AQ = QA = \lambda_0 Q$, and consequently the sequences

$$\big([A^n - \lambda_0^n Q]_{jk}\big) = \big(e_j^\top (A^n - \lambda_0^n Q) e_k\big) = \big(e_j^\top A^n (I_d - Q) e_k\big)$$

as well as

$$\big([A^{n+1} - \lambda_0 A^n]_{jk}\big) = \big(e_j^\top A^n (A - \lambda_0 I_d) e_k\big)$$

are all of the form $(x^\top A^n y)$ with $x = e_j$ and the appropriate $y \in \mathbb{R}^d$ where, however, $y \geq 0$ may not hold and consequently Theorem 3.2 may not apply. ◁

With a view towards Theorem 3.2, how does one decide in practice whether a given $d \times d$-matrix A has a positive power? Comprehensive answers to this question appear to be documented in the literature only for $A \geq 0$, that is, for non-negative matrices. In this case, Wielandt's Theorem [25, Cor. 8.5.9] asserts that $A^N > 0$ for some $N \in \mathbb{N}$ (if and) only if $A^{d^2-2d+2} > 0$. The number $d^2 - 2d + 2$ is smallest possible in general, but can be reduced in many special cases, see [25, Sect. 8.5]. An equivalent condition is that A be *irreducible* and *aperiodic*, i.e. for any two indices $j, k \in \{1, 2, \ldots, d\}$ there exists a positive integer $N(j, k)$ such that $[A^n]_{jk} > 0$ for every $n \geq N(j, k)$; see e.g. [25, Sect. 8.4]. Note that if $A \geq 0$ but the matrix A^n is not positive for *any* $n \in \mathbb{N}$ then there exists $j, k \in \{1, 2, \ldots, d\}$ such that eventually the sequence $([A^n]_{jk})$ vanishes periodically and hence cannot be Benford. Overall, by combining these known facts, Theorem 3.2 can be re-stated specifically for non-negative matrices.

Theorem 3.12 *Let $A \in \mathbb{R}^{d\times d}$ be non-negative. Then the following three statements are equivalent:*

(i) *A is irreducible and aperiodic, and $\log r_\sigma(A)$ is irrational;*
(ii) *$A^{d^2-2d+2} > 0$ and $\log r_\sigma(A)$ is irrational;*
(iii) *The sequence $(x^\top A^n y)$ is Benford for every $x, y \neq 0$ with $x \geq 0$ and $y \geq 0$.*

Moreover, if (i)–(iii) hold then, for every $x \neq 0$ with $x \geq 0$, the sequence $(|A^n x|)$ is Benford, and so is $(|A^n|)$.

Proof If A^{d^2-2d+2} is not positive then neither is A^n for any n, by Wielandt's Theorem, and hence A cannot be irreducible and aperiodic. Thus (i)$\Rightarrow$(ii). According to Theorem 3.2, (iii) follows from (ii). Assume in turn that (i) does not hold. Then either A is not irreducible and aperiodic, or $\log r_\sigma(A)$ is rational. In the former case, $([A^n]_{jk}) = (e_j^\top A^n e_k)$ vanishes periodically for some $j, k \in \{1, 2, \ldots, d\}$, hence (iii) fails with $x = e_j \geq 0$ and $y = e_k \geq 0$. In the latter case, assume w.l.o.g. that A is irreducible and aperiodic. Then (iii) fails again, by virtue of Theorem 3.2. Overall, (iii)$\Rightarrow$(i). Finally, the assertions regarding $(|A^n x|)$ and $(|A^n|)$ are obvious from Theorem 3.2. □

Remark The non-negative matrix $A = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$ has neither of the properties (i)–(iii) in Theorem 3.12, and yet $(|A^n x|)$ is Benford for every $x \neq 0$, and so is $(|A^n|)$.

In general, i.e. without the assumption that A be non-negative, the clear-cut situation of the non-negative case persists only for the special cases $d = 1$ (trivial) and $d = 2$ (a simple exercise), where $A^N > 0$ for some $N \in \mathbb{N}$ (if and) only if $A^2 > 0$. In stark contrast, if $d \geq 3$ then the minimal positive integer N with $A^N > 0$ can be arbitrarily large. For example, for every $\alpha \in \mathbb{R}$ the (symmetric) 3×3-matrix

$$A_\alpha := \begin{bmatrix} 10 - 10^{4\alpha} & 10^{\alpha+1}\sqrt{2}(10^{2\alpha-1}+1) & 9\cdot 10^{2\alpha} \\ 10^{\alpha+1}\sqrt{2}(10^{2\alpha-1}+1) & 18\cdot 10^{2\alpha} & 10^{\alpha}\sqrt{2}(10^{2\alpha+1}+1) \\ 9\cdot 10^{2\alpha} & 10^{\alpha}\sqrt{2}(10^{2\alpha+1}+1) & 10^{4\alpha+1}-1 \end{bmatrix}$$

is positive precisely if $|\alpha| < \frac{1}{4}$, and for $|\alpha| \geq \frac{1}{4}$ a short calculation shows that

$$\min\{n \in \mathbb{N} : A_\alpha^n > 0\} = 2\lfloor|\alpha|\rfloor + 2 > 2|\alpha|.$$

Note that, for every $\alpha \in \mathbb{R}$, the matrix A_α has at most *one* negative entry (and is, in the terminology of [25, Exc. 8.3.9], *essentially non-negative*).

The example of A_α demonstrates that unlike in the non-negative case, for $d \geq 3$ the minmal exponent N with $A^N > 0$ does not admit an upper bound independent of A. Still, the property that $A^N > 0$ for *some* $N \in \mathbb{N}$ can be characterized rather neatly.

Proposition 3.13 *The following properties are equivalent for every $A \in \mathbb{R}^{d\times d}$:*

(i) $A^N > 0$ *for some* $N \in \mathbb{N}$;
(ii) $A^{2n} > 0$ *for all sufficiently large* $n \in \mathbb{N}$;
(iii) *Either* $\lambda_0 = r_\sigma(A) > 0$ *or* $\lambda_0 = -r_\sigma(A) < 0$ *is an algebraically simple eigenvalue of* A *with* $|\lambda| < r_\sigma(A)$ *for every* $\lambda \in \sigma(A)\backslash\{\lambda_0\}$*, and the spectral projection* Q *associated with* λ_0 *is positive, i.e.*

$$Q = \frac{bc^\top}{b^\top c} > 0, \tag{16}$$

where b *and* c *are eigenvectors of, respectively,* A *and* $A^\top$ *corresponding to the eigenvalue* λ_0*, that is,* $Ab = \lambda_0 b$ *and* $A^\top c = \lambda_0 c$.

Applying this result for instance to the 3×3-matrix B of Example 3.5 yields, with $\lambda_0 = r_\sigma(B) = 6.165$,

$$Q = 10^{-4}\begin{bmatrix} 3.158 & 28.95 & 175.3 \\ 28.95 & 265.3 & 1606 \\ 175.3 & 1606 & 9731 \end{bmatrix} > 0,$$

and hence immediately shows that $B^N > 0$ for *some* $N \in \mathbb{N}$. (In Example 3.5, the minimal such N was seen to be $N = 8$.) On the other hand, for the matrix B considered in Example 3.9, the spectral projection associated with $\lambda_0 = r_\sigma(B) = 2$,

$$Q = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

is not positive, and neither is $B^n = 2^{n-1}B = 2^n Q$ positive for any $n \in \mathbb{N}$.

Example 3.14 Theorems 3.2 and 3.12 are especially easy to apply if A is an *integer* matrix, i.e. if $[A]_{jk} \in \mathbb{Z}$ for every j, k. In this case, an explicit calculation of $r_\sigma(A)$ is not required. In fact, if $A \in \mathbb{Z}^{d\times d}$ with $d \geq 2$ and $A^N > 0$ for some $N \in \mathbb{N}$ then $\log r_\sigma(A)$ is irrational (even transcendental) provided that

$$\begin{array}{l}\text{none of the numbers } \pm 10^m, \text{ with } m = 1, 2, \ldots, \lfloor d \, \log \|A\|_\infty \rfloor \\ \text{and } \|A\|_\infty := \max_j \sum_{k=1}^d \left|[A]_{jk}\right|, \text{ is an eigenvalue of any of} \\ \text{the } d \text{ matrices } A, A^2, \ldots, A^d. \end{array} \tag{17}$$

Even simpler to check is the condition that

$$\det A \text{ is not divisible by } 10, \tag{18}$$

which implies (17) and hence also guarantees the irrationality of $\log r_\sigma(A)$.

For example, the matrix A associated with (9) is an integer matrix with $A^2 > 0$, and $\det A = -1$ obviously satisfies (18). Hence, as already seen in Example 3.4, Theorem 3.2 applies, and $(x^\top A^n y)$ is Benford for all $x, y \geq 0$ with $x \neq 0$ and $y \neq 0$. Similarly, for the matrix B discussed in Example 3.5, $B^8 > 0$, and $\det B = -3$ is not divisible by 10, hence $\log r_\sigma(B)$ is irrational, and again Theorem 3.2 can be applied without determining the actual value of $r_\sigma(B)$.

For another example, consider the matrix

$$C = \begin{bmatrix} -3 & -1 & -1 \\ -2 & 1 & -3 \\ 1 & -3 & -1 \end{bmatrix},$$

for which $\|C\|_\infty = 6$. As before, $C^8 > 0$, hence Theorem 3.2 applies. Note that $\det C = 30$, and so (18) fails. However, (17) holds, as $\lfloor 3 \log \|C\|_\infty \rfloor = 2$ and none of the four integers ± 10, $\pm 10^2$ is an eigenvalue of any of the three matrices

$$C, \quad C^2 = \begin{bmatrix} 10 & 5 & 7 \\ 1 & 12 & 2 \\ 2 & -1 & 9 \end{bmatrix}, \quad C^3 = \begin{bmatrix} -33 & -26 & -32 \\ -25 & 5 & -39 \\ 5 & -30 & -8 \end{bmatrix},$$

as can easily be checked e.g. by means of row-reductions. Again, therefore, $\log r_\sigma(C)$ is irrational. ◁

4 The Case of Arbitrary Matrices

Given an arbitrary real $d \times d$-matrix A, this section presents a necessary and sufficient condition for the sequence $(x^\top A^n y)$ to be, for any vectors $x, y \in \mathbb{R}^d$, either Benford or identically zero for $n \geq d$. As explained earlier, the result also allows to characterize the Benford property for solutions of any linear difference equation. To provide the reader with some intuition as to which properties of such equations, or the matrices associated with them, may affect the Benford property, first a few simple examples are discussed.

Example 4.1 (i) Let the sequence (x_n) be defined recursively as

$$x_n = x_{n-1} - x_{n-2}, \quad n \geq 3, \tag{19}$$

with given $x_1, x_2 \in \mathbb{R}$. From the explicit representation for (x_n),

$$x_n = (x_1 - x_2) \cos\left(\tfrac{1}{3}\pi n\right) + \tfrac{1}{\sqrt{3}}(x_1 + x_2) \sin\left(\tfrac{1}{3}\pi n\right), \quad n \in \mathbb{N},$$

it is clear that $x_{n+6} = x_n$ for all n, i.e. (x_n) is 6-periodic. This oscillatory behaviour of (x_n) corresponds to the fact that the eigenvalues of (19), i.e. of the matrix associated

with it, are $\lambda = e^{\pm \imath \pi/3}$ and hence lie on the unit circle $\mathbb{S}$. For no choice of x_1, x_2, therefore, is (x_n) Benford.
(ii) Consider the linear 3-step recursion

$$x_n = 2x_{n-1} + 10x_{n-2} - 20x_{n-3}, \quad n \geq 4. \tag{20}$$

For any $x_1, x_2, x_3 \in \mathbb{R}$, the value of x_n is given explicitly by

$$x_n = \alpha_1 2^n + \alpha_2 10^{n/2} + \alpha_3 (-1)^n 10^{n/2},$$

with the constants $\alpha_1, \alpha_2, \alpha_3$ according to

$$\alpha_1 = \tfrac{1}{12}(10x_1 - x_3), \quad \alpha_{2,3} = \tfrac{1}{60}(x_3 + 3x_2 - 10x_1) \pm \tfrac{1}{12\sqrt{10}}(x_3 - 4x_1).$$

Clearly, $\limsup_{n\to\infty} |x_n| = +\infty$ unless $x_1 = x_2 = x_3 = 0$, so unlike in (i) the sequence (x_n) is not bounded. However, if $|\alpha_2| \neq |\alpha_3|$ then

$$\log|x_n| = \frac{n}{2} + \log\left|\alpha_1 10^{-n(\frac{1}{2} - \log 2)} + \alpha_2 + (-1)^n \alpha_3\right| \approx \frac{n}{2} + \log|\alpha_2 + (-1)^n \alpha_3|,$$

showing that $\big(S(x_n)\big)$ is asymptotically 2-periodic and hence (x_n) is not Benford. Similarly, if $|\alpha_2| = |\alpha_3| \neq 0$ then $\big(S(x_n)\big)$ is convergent along even (if $\alpha_2 = \alpha_3$) or odd (if $\alpha_2 = -\alpha_3$) indices n, and again (x_n) is not Benford. Only if $\alpha_2 = \alpha_3 = 0$ yet $\alpha_1 \neq 0$, or equivalently if $x_3 = 2x_2 = 4x_1 \neq 0$ is (x_n) Benford. Obviously, the oscillatory behaviour of $\big(S(x_n)\big)$ in this example is due to the characteristic equation $\lambda^3 = 2\lambda^2 + 10\lambda - 20$ associated with (20) having two roots with the same modulus but opposite signs, namely $\lambda = \pm\sqrt{10}$.
(iii) Let $\gamma = \cos(\pi \log 2) = 0.5851$ and define (x_n) recursively as

$$x_n = 4\gamma x_{n-1} - 4x_{n-2}, \quad n \geq 3, \tag{21}$$

with given $x_1, x_2 \in \mathbb{R}$. As before, an explicit formula for x_n is easily derived as

$$\begin{aligned} x_n &= 2^{n-2}(4\gamma x_1 - x_2)\cos(\pi n \log 2) + 2^{n-2}\frac{\gamma x_2 - 2x_1(2\gamma^2 - 1)}{\sqrt{1-\gamma^2}}\sin(\pi n \log 2) \\ &= 2^n \beta \cos(\pi n \log 2 + \xi), \end{aligned}$$

with the appropriate $\beta \geq 0$ and $\xi \in \mathbb{R}$. Although somewhat oscillatory, the sequence (x_n) is clearly unbounded. However, if $(x_1, x_2) \neq (0, 0)$ then $\beta > 0$, and

$$\log|x_n| = n \log 2 + \log \beta + \log|\cos(\pi n \log 2 + \xi)|, \quad n \in \mathbb{N},$$

together with Lemma 2.6, where $\vartheta_1 = \vartheta_2 = \log 2$, $\alpha = 0$, $z = e^{\imath \xi}$, and $z_n \equiv 0$, shows that (x_n) is not Benford. The reason for this can be seen in the fact that, while

$\log|\lambda| = \log 2$ is irrational for the roots $\lambda = 2e^{\pm \imath\pi \log 2}$ of the characteristic equation associated with (21), there clearly is a rational dependence between the two real numbers $\log|\lambda|$ and $\frac{1}{2\pi}\arg\lambda$, namely $\log|\lambda| - 2(\frac{1}{2\pi}\arg\lambda) = 0$. ◁

The above examples indicate that, under the perspective of BL, the main difficulty when dealing with multi-dimensional systems is their potential for more or less cyclic behaviour, either of the orbits themselves or of their significands. (In the case of *positive* matrices, as seen in the previous section, cyclicality does not occur or, more correctly, remains hidden.) To precisely denominate this difficulty, the following terminology will prove useful. Recall that, given any set $Z \subset \mathbb{C}$, $\operatorname{span}_{\mathbb{Q}} Z$ denotes the smallest linear subspace of $\mathbb{C}$ (over $\mathbb{Q}$) containing Z.

Definition 4.2 A non-empty set $Z \subset \mathbb{C}$ with $|z| = r$ for some $r > 0$ and all $z \in Z$, i.e. $Z \subset r\mathbb{S}$, is *non-resonant* if its associated set $\Delta_Z \subset \mathbb{R}$, defined as

$$\Delta_Z := \left\{1 + \frac{\arg z - \arg w}{2\pi} : z, w \in Z\right\}$$

satisfies the following two conditions:

(i) $\Delta_Z \cap \mathbb{Q} = \{1\}$;
(ii) $\log r \notin \operatorname{span}_{\mathbb{Q}} \Delta_Z$.

An arbitrary set $Z \subset \mathbb{C}$ is *non-resonant* if, for every $r > 0$, the set $Z \cap r\mathbb{S}$ is either non-resonant or empty; otherwise Z is *resonant*.

Note that by its very definition the set Δ_Z always satisfies $1 \in \Delta_Z \subset (0, 2)$ and is symmetric w.r.t. the point 1. The empty set $\varnothing$ and the singleton $\{0\}$ are non-resonant. On the other hand, $Z \subset \mathbb{C}$ is certainly resonant if either $\{-r, r\} \subset Z$ for some $r > 0$, in which case (i) is violated, or $Z \cap \mathbb{S} \neq \varnothing$, which causes (ii) to fail.

Example 4.3 The singleton $\{z\}$ with $z \in \mathbb{C}$ is non-resonant if and only if either $z = 0$ or $\log|z| \notin \mathbb{Q}$. Similarly, the set $\{z, \overline{z}\}$ with $z \in \mathbb{C}\backslash\mathbb{R}$ is non-resonant if and only if the three numbers 1, $\log|z|$ and $\frac{1}{2\pi}\arg z$ are rationally independent, i.e. linearly independent over $\mathbb{Q}$. ◁

Remark If $Z \subset r\mathbb{S}$ then, for every $z \in Z$,

$$\operatorname{span}_{\mathbb{Q}} \Delta_Z = \operatorname{span}_{\mathbb{Q}} \left(\{1\} \cup \left\{\frac{\arg z - \arg w}{2\pi} : w \in Z\right\}\right),$$

which shows that the dimension of $\operatorname{span}_{\mathbb{Q}} \Delta_Z$, as a linear space over $\mathbb{Q}$, is at most $\#Z$. Also, if $Z \subset r\mathbb{S}$ is symmetric w.r.t. the real axis, then the condition (ii) in Definition 4.2 is equivalent to $\log r \notin \operatorname{span}_{\mathbb{Q}}\big(\{1\} \cup \{\frac{1}{2\pi}\arg z : z \in Z\}\big)$, cf. [5, Def. 3.1].

Recall that the behaviour of (A^n) is completely determined by the eigenvalues of A, together with the corresponding (generalized) eigenvectors. As far as BL is concerned, the key question turns out to be whether or not $\sigma(A)$ is non-resonant. Clearly $\log r_\sigma(A)$ is irrational whenever $\sigma(A)$ is non-resonant (and A is not nilpotent), but the converse is not true in general.

Example 4.4 The spectrum of the matrix A associated with the Fibonacci recursion (9), $\sigma(A) = \{-\varphi^{-1}, \varphi\}$, is non-resonant. On the other hand, the matrices

$$B = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 10 & -20 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 4\gamma & -4 \\ 1 & 0 \end{bmatrix},$$

associated with the difference equations (19), (20), and (21), respectively, all have a resonant spectrum. Indeed, $\sigma(B) = \{e^{\pm \imath\pi/3}\}$, and hence $\Delta_{\sigma(B)} = \{\frac{2}{3}, 1, \frac{4}{3}\}$ contains rational numbers other than 1, which violates (i) in Definition 4.2. Also, $\log|e^{\pm\imath\pi/3}| = 0$, and so (ii) is violated, too. Similarly, $\sigma(C) = \{2, \pm\sqrt{10}\}$, and with $Z = \sigma(C) \cap \sqrt{10}\mathbb{S} = \{\pm\sqrt{10}\}$ again both (i) and (ii) in Definition 4.2 fail. Finally, $\sigma(D) = \{2e^{\pm\imath\pi \log 2}\}$, and so $\Delta_{\sigma(D)} = \{1, 1 \pm \log 2\}$ satisfies (i), yet (ii) is violated as $\log 2 \in \operatorname{span}_{\mathbb{Q}} \Delta_{\sigma(D)} = \operatorname{span}_{\mathbb{Q}}\{1, \log 2\}$. ◁

The following theorem is the main result of the present section. Like Theorems 3.2 and 3.12, but without any assumptions on A, it extends to arbitrary dimensions the simple fact that for the sequence $(x\alpha^n y)$ with $\alpha \in \mathbb{R}\backslash\{0\}$ to be either Benford (if $xy \neq 0$) or trivial (if $xy = 0$) it is necessary and sufficient that $\log|\alpha|$ be irrational. To concisely formulate the result, call $(x^\top A^n y)$ and $(|A^n x|)$ with $A \in \mathbb{R}^{d\times d}$ and $x, y \in \mathbb{R}^d$ *terminating* if, respectively, $x^\top A^n y = 0$ or $A^n x = 0$ for all $n \geq d$; similarly, $(|A^n|)$ is terminating if $A^n = 0$ for all $n \geq d$.

Theorem 4.5 *Let A be a real $d \times d$-matrix. Then the following statements are equivalent:*

(i) *The set $\sigma(A)$ is non-resonant;*
(ii) *For every $x, y \in \mathbb{R}^d$, the sequence $(x^\top A^n y)$ is Benford or terminating.*

Moreover, if (i) *and* (ii) *hold then, for every $x \in \mathbb{R}^d$, the sequence $(|A^n x|)$ is Benford or terminating, and so is $(|A^n|)$.*

For a full proof of Theorem 4.5, the reader is referred to [6]. A simplified variant that applies to *most* matrices is given at the end of this section. From the argument, it will transpire that "terminating" can be replaced by "zero" (meaning "identically zero") whenever A is invertible, i.e. whenever $0 \notin \sigma(A)$. Before that, however, a few examples, corollaries, and remarks are presented.

Example 4.6 (i) As seen in Example 4.4, the (invertible) matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ associated with (9) has non-resonant spectrum. For every $x, y \in \mathbb{R}^2$, therefore, $(x^\top A^n y)$ is either Benford or zero. The latter happens precisely if x and y are proportional to, respectively, the eigenvector $\varphi e_1 + e_2$, corresponding to the eigenvalue φ of A, and to the eigenvector $\varphi e_2 - e_1$, corresponding to $-\varphi^{-1}$, or vice versa. In particular, the sequences $(F_n) = (e_1^\top A^n e_1)$ and $(L_n) = \left(e_1^\top A^n (3e_2 - e_1)\right)$ are Benford, as has already been observed in Examples 3.4 and 3.8.

Note that (F_n^2), for instance, is also Benford. This follows from Proposition 2.3 but can be seen directly as well by noticing that $\left(F_n^2 + \frac{2}{5}(-1)^n\right)$ is a solution of

$$x_n = 3x_{n-1} - x_{n-2}, \quad n \geq 3,$$

and the associated matrix $\begin{bmatrix} 3 & -1 \\ 1 & 0 \end{bmatrix}$ has non-resonant spectrum $\{\varphi^2, \varphi^{-2}\}$.

(ii) The 3×3-matrix B considered in Example 3.5 has non-resonant spectrum, as it has three real eigenvalues of different absolute value, none of which is of the form $\pm 10^{m/n}$ with $m \in \mathbb{Z}$ and $n \in \mathbb{N}$. As in (i), every sequence $(x^\top B^n y)$ is either Benford or zero, with the latter being the case precisely if x and y are proportional to eigenvectors of B corresponding to two *different* eigenvalues. Note that even for this conclusion, which is stronger than the one reached in Example 3.5, it is not necessary to know $\sigma(B)$ explicitly. In fact, unlike in Example 3.5 it is not even necessary to know that $B^N > 0$ for some N. ◁

Example 4.7 For the matrix $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ one finds $\sigma(A) = \{\sqrt{2}e^{\pm \imath\pi/4}\}$ which is resonant, as $\Delta_{\sigma(A)} = \{\frac{3}{4}, 1, \frac{5}{4}\}$. By Theorem 4.5, there must be $x, y \in \mathbb{R}^2$ for which $(x^\top A^n y)$ is neither Benford nor zero. Indeed, observe for instance that

$$e_1^\top A^n e_1 = e_1^\top 2^{n/2} \begin{bmatrix} \cos(\frac{1}{4}\pi n) & -\sin(\frac{1}{4}\pi n) \\ \sin(\frac{1}{4}\pi n) & \cos(\frac{1}{4}\pi n) \end{bmatrix} e_1 = 2^{n/2}\cos(\tfrac{1}{4}\pi n), \quad n \in \mathbb{N}_0,$$

and hence $(e_1^\top A^n e_1)$ is neither Benford (because $e_1^\top A^{4n-2} e_1 = 0$ for all n) nor zero (because $e_1^\top A^{8n} e_1 = 2^{4n} \neq 0$ for all n). Note, however, that this of course does not rule out the possibility that *some* sequences derived from (A^n) may be Benford nevertheless. For instance, $(|A^n|) = (2^{n/2})$ is Benford. For another concrete example, fix any $x \neq 0$ and, for each $n \in \mathbb{N}$, denote by E_n the area of the triangle with vertices at $A^n x$, $A^{n-1}x$, and the origin. Then

$$E_n = \tfrac{1}{2}\left|\det(A^n x, A^{n-1}x)\right| = 2^{n-2}|x|^2, \quad n \in \mathbb{N},$$

so (E_n) is Benford, see Fig. 4. Note also that while $\sigma(A)$ is resonant, the set $\sigma(A^4) = \{-4\}$ is not. (The reverse implication is easily seen to hold for all $d \in \mathbb{N}$ and $A \in \mathbb{R}^{d\times d}$: If $\sigma(A)$ is non-resonant then so is $\sigma(A^n)$ for every $n \in \mathbb{N}$.) ◁

Example 4.8 For the matrix $B = \begin{bmatrix} 19 & 20 \\ 1 & 0 \end{bmatrix}$, first encountered in Example 3.10, $\sigma(B) = \{-1, 20\}$ is resonant. Consequently, there must be $x, y \in \mathbb{R}^2$ for which $(x^\top B^n y)$ is neither Benford nor zero. In essence, this has already been observed in Example 3.10, with $x = e_1$ and $y = e_1 - e_2$, for which $(x^\top B^n y) = \big((-1)^n\big)$. Note that failure of $(x^\top B^n y)$ to be Benford can occur only if $(20x_1 + x_2)(y_1 + y_2) = 0$. For *most* $x, y \in \mathbb{R}^2$, therefore, $(x^\top B^n y)$ is either Benford or zero. ◁

Example 4.9 This example briefly reviews matrices and difference equations from earlier examples in the light of Theorem 4.5

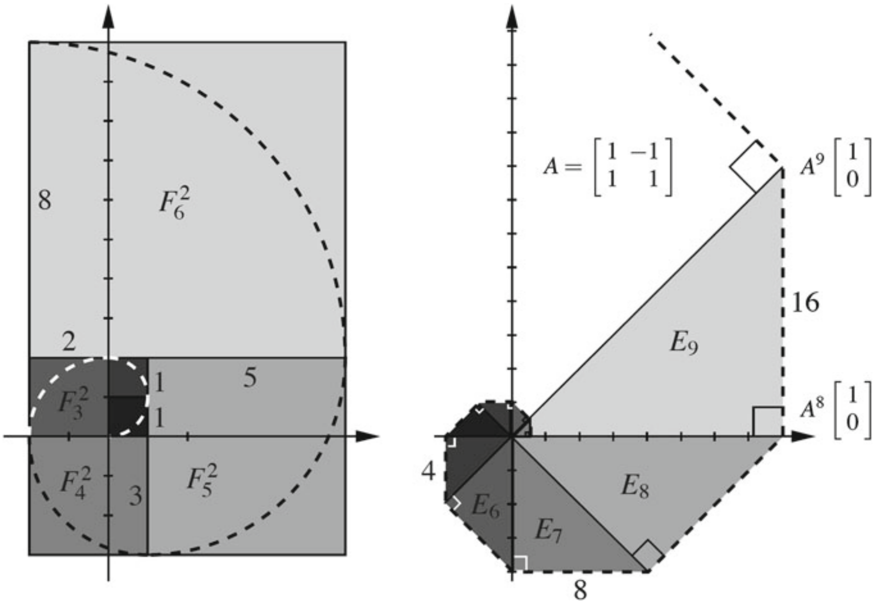

Fig. 4 Two Benford sequences, (F_n^2) and (E_n), derived from linear 2-dimensional systems, see Examples 4.6 and 4.7; note that $\sigma(A)$ is resonant for the matrix A associated with (E_n)

(i) The matrices $A = \begin{bmatrix} 5 & 15 \\ 15 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix}$ both have resonant spectrum, $\sigma(A) = \{-10, 20\}$ and $\sigma(B) = \{2, 10\}$, which corroborates the observation, made in Example 3.6, that for some $x, y \in \mathbb{R}^2$, $(x^\top A^n y)$ is neither Benford nor zero, and similarly for B. Note, however, that $(x^\top A^n y)$ is Benford whenever $x, y \in \mathbb{R}^2$ are not multiples of $e_1 - e_2$, and hence for *most* $x, y \in \mathbb{R}^2$, whereas $(x^\top B^n y)$ can be Benford *only* if x or y is a multiple of $e_1 - e_2$.

(ii) While Theorem 3.2 did not apply to $B = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ in Example 3.9, every sequence $(x^\top B^n y)$ was seen to be Benford or terminating. This observation is consistent with $\sigma(B) = \{0, 2\}$ being non-resonant.

(iii) As is the case for every stochastic matrix, the matrix $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{bmatrix}$ in Examples 3.10 and 3.11, has resonant spectrum $\sigma(A) = \{-\frac{1}{2}, 1\}$, and for most $x, y \in \mathbb{R}^2$, $(x^\top A^n y)$ is not Benford. The question, already raised in Example 3.11, whether, say, entries of $(A^{n+1} - A^n)$ can be Benford nevertheless is addressed in Sect. 5. $\triangleleft$

Example 4.10 Unlike in Theorem 3.2, within the wider scope of Theorem 4.5 the sequence $(|A^n x|)$ may, for every $x \in \mathbb{R}^d$, be Benford or terminating even if (i) and (ii) do not hold. Similarly, $(|A^n|)$ may be Benford. For an example, consider the 3×3-matrix

$$A = 10^{\varphi^2} \begin{bmatrix} \cos(2\pi\varphi) & -\sin(2\pi\varphi) & 0 \\ \sin(2\pi\varphi) & \cos(2\pi\varphi) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where $\varphi = \frac{1}{2}(1+\sqrt{5})$, as usual, and hence $\varphi^2 = \varphi + 1$. The spectrum $\sigma(A) = \{10^{\varphi^2}, 10^{\varphi^2} e^{\pm\imath 2\pi\varphi}\}$ is resonant because

$$\tfrac{1}{2}(3+\sqrt{5}) = \varphi^2 = \log 10^{\varphi^2} \in \operatorname{span}_{\mathbb{Q}} \Delta_{\sigma(A)} = \operatorname{span}_{\mathbb{Q}}\{1, \sqrt{5}\}.$$

Nevertheless, for every $x \in \mathbb{R}^3$,

$$|A^n x| = 10^{n\varphi^2} \left\| \begin{bmatrix} x_1\cos(2\pi n\varphi) - x_2 \sin(2\pi n\varphi) \\ x_1 \sin(2\pi n\varphi) + x_2\cos(2\pi n\varphi) \\ x_3 \end{bmatrix} \right\| = 10^{n\varphi^2}|x|,$$

and since φ^2 is irrational, $(|A^n x|)$ is Benford whenever $x \neq 0$. Similarly, note that $10^{-n\varphi^2}A$ is an isometry for every n, and $(|A^n|) = (10^{n\varphi^2})$ is Benford. However, by Theorem 4.5, not *every* sequence $(x^\top A^n y)$ can be Benford or zero. That $(e_2^\top A^n e_1) = \big(10^{n\varphi^2}\sin(2\pi n\varphi)\big)$, for instance, is neither can be seen easily using Lemma 2.6.

Consider now also the matrix

$$B = 10^{\varphi^2} \begin{bmatrix} \cos(2\pi\varphi) & -\sin(2\pi\varphi) & \sin(\pi\varphi)\cos(\pi\varphi) \\ \sin(2\pi\varphi) & \cos(2\pi\varphi) & \sin(\pi\varphi)^2 \\ 0 & 0 & 1 \end{bmatrix}.$$

Clearly, $\sigma(B) = \sigma(A)$, so the spectrum of B is resonant as well. A short calculation confirms that

$$B^n = 10^{n\varphi^2} \begin{bmatrix} \cos(2\pi n\varphi) & -\sin(2\pi n\varphi) & \sin(\pi n\varphi)\cos(\pi n\varphi) \\ \sin(2\pi n\varphi) & \cos(2\pi n\varphi) & \sin(\pi n\varphi)^2 \\ 0 & 0 & 1 \end{bmatrix}, \quad n \in \mathbb{N}_0,$$

from which it follows for instance that

$$|B^n\sqrt{2}e_3| = 10^{n\varphi^2}\sqrt{3-\cos(2\pi n\varphi)}, \quad n \in \mathbb{N}_0,$$

and consequently, using $\varphi^2 = \varphi + 1$,

$$\big\langle \log|B^n\sqrt{2}e_3|\big\rangle = \big\langle n\varphi + \tfrac{1}{2}\log\big(3-\cos(2\pi n\varphi)\big)\big\rangle = \langle f(n\varphi)\rangle,$$

with the smooth function $f:\mathbb{R}\to\mathbb{R}$ given by $f(s) = s + \frac{1}{2}\log\big(3-\cos(2\pi s)\big)$. Recall that $(n\varphi)$ is u.d. mod 1. As in the proof of Lemma 2.6, consider the piecewise smooth map $T = \langle f\rangle$ on $[0,1)$ induced by f. Since T is a bijection of $[0,1)$ with non-constant slope, $\lambda_{0,1}\circ T^{-1} \neq \lambda_{0,1}$. This in turn means that $(|B^n e_3|)$, and in fact

$(|B^n x|)$ for *most* $x \in \mathbb{R}^3$, is neither Benford nor zero. Similarly,

$$|B^n| = 10^{n\varphi^2}\sqrt{1 + \tfrac{1}{2}\sin(\pi n\varphi)^2 + \tfrac{1}{2}|\sin(\pi n\varphi)|\sqrt{4 + \sin(\pi n\varphi)^2}}, \quad n \in \mathbb{N}_0,$$

and a completely analogous argument shows that $(|B^n|)$ is not Benford either.

As evidenced by this example, the property of a matrix $A \in \mathbb{R}^{d\times d}$ that $(|A^n x|)$ is, for every $x \in \mathbb{R}^d$, either Benford or terminating is *not* a spectral property, i.e. it cannot be decided upon, at least for $d \geq 3$, by using $\sigma(A)$ alone. Similarly, $(|A^n|)$ being Benford is not a spectral property of A. $\triangleleft$

Remark According to Theorem 4.5, non-resonance of $\sigma(A)$ is, for any invertible $A \in \mathbb{R}^{d\times d}$, equivalent to the widespread generation of Benford sequences of the form $(x^\top A^n y)$. Most $d \times d$-matrices are invertible with non-resonant spectrum, under a topological as well as a measure-theoretic perspective. To put this more formally, let

$$\mathscr{G}_d := \{A \in \mathbb{R}^{d\times d} : A \text{ is invertible and } \sigma(A) \text{ is non-resonant}\}.$$

Thus for example $\mathscr{G}_1 = \{[\alpha] : \alpha \in \mathbb{R}\backslash\{0\}, |\alpha| \neq 10^\rho \text{ for every } \rho \in \mathbb{Q}\}$. While the complement of $\mathscr{G}_d$ is dense in $\mathbb{R}^{d\times d}$, it is a topologically small set: $\mathbb{R}^{d\times d}\backslash\mathscr{G}_d$ is *of first category*, i.e. a countable union of nowhere dense sets. A (topologically) typical ("generic") $d \times d$-matrix therefore belongs to $\mathscr{G}_d$. Similarly, if A is an $\mathbb{R}^{d\times d}$-valued random variable, that is, a random matrix, whose distribution is a.c. with respect to the d^2-dimensional Lebesgue measure on $\mathbb{R}^{d\times d}$, then $\mathbb{P}(A \in \mathscr{G}_d) = 1$, i.e. with probability one A is invertible and $\sigma(A)$ non-resonant.

The next result is a corollary of Theorem 4.5 for difference equations and analogous to Corollary 3.7 but without any positivity assumptions on coefficients or initial values.

Theorem 4.11 *The following statements are equivalent for the difference equation*

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_{d-1}x_{n-d+1} + a_d x_{n-d}, \quad n \geq d+1, \tag{22}$$

where $a_1, a_2, \ldots, a_{d-1}, a_d \in \mathbb{R}$ *with* $a_d \neq 0$*:*

(i) *The set* $\{z \in \mathbb{C} : z^d = a_1 z^{d-1} + a_2 z^{d-2} + \cdots + a_{d-1}z + a_d\}$ *is non-resonant;*
(ii) *Every solution* (x_n) *of* (22) *is Benford, unless* $x_n \equiv 0$.

While the reader is again referred to [6] for a full proof of Theorem 4.11, a simplified argument applicable to most $a_1, a_2, \ldots, a_{d-1}, a_d$ is given at the end of the present section, following the very similar proof of Theorem 4.5.

Example 4.12 Since $\{z \in \mathbb{C} : z^2 = z + 1\} = \{-\varphi^{-1}, \varphi\}$ is non-resonant, every solution of (9) except for $x_n \equiv 0$ is Benford, as was already seen in Example 4.6.

More generally, consider the second-order difference equation

$$x_n = a_1 x_{n-1} + a_2 x_{n-2}, \quad n \geq 3, \tag{23}$$

where a_1, a_2 are non-zero *integers*, and $a_2 > 0$. The set $\{z \in \mathbb{C} : z^2 = a_1 z + a_2\}$ consists of two real numbers with different absolute value, and is resonant if and only if one of them is of the form $\pm 10^N$ for some $N \in \mathbb{N}_0$. It follows that every solution (x_n) of (23), except for the trivial $x_n \equiv 0$, is Benford if and only if

$$|10^{2m} - a_2| \neq |a_1| 10^m, \quad \forall m = 0, 1, \ldots, \lfloor \log(|a_1| + a_2) \rfloor. \tag{24}$$

For example, for $a_1 = 2, a_2 = 5$, condition (24) reduces to $|1 - 5| \neq 2$. As the latter is obviously correct, every solution of

$$x_n = 2x_{n-1} + 5x_{n-2}, \quad n \geq 3,$$

is Benford unless $x_1 = x_2 = 0$. On the other hand, for

$$x_n = 19x_{n-1} + 20x_{n-2}, \quad n \geq 3,$$

(24) reads $|10^{2m} - 20| \neq 19 \cdot 10^m$ for $m = 0, 1$, which is violated for $m = 0$. This corroborates the observation, already made in Example 3.10, that $\big((-1)^n\big)$ is a solution that is neither Benford nor zero. ◁

Remark Earlier, weaker forms and variants of the implication (i)⇒(ii) in Theorems 4.5 and 4.11, or special cases thereof, can be traced back at least to [42] and may also be found in [5, 7, 9, 26, 35, 43]. The reverse implication (ii)⇒(i) seems to have been addressed only for $d < 4$, see [7, Thm. 5.37]. A case in point is the 4×4-matrix

$$A = 10^{\sqrt{2}} \begin{bmatrix} \cos(2\pi\sqrt{3}) & -\sin(2\pi\sqrt{3}) & 0 & 0 \\ \sin(2\pi\sqrt{3}) & \cos(2\pi\sqrt{3}) & 0 & 0 \\ 0 & 0 & \cos(4\pi\sqrt{3}) & -\sin(4\pi\sqrt{3}) \\ 0 & 0 & \sin(4\pi\sqrt{3}) & \cos(4\pi\sqrt{3}) \end{bmatrix}.$$

In [7, Ex. 5.36], it was observed that $(x^\top A^n y)$ is Benford or zero for every $x, y \in \mathbb{R}^4$—despite the fact that A fails to be *Benford regular*, a property introduced there that is more restrictive than the non-resonance of $\sigma(A)$. This mismatch is resolved by Theorem 4.5, simply by noticing that $\sigma(A) = \{10^{\sqrt{2}} e^{\pm \imath 2\pi\sqrt{3}}, 10^{\sqrt{2}} e^{\pm \imath 4\pi\sqrt{3}}\}$ is indeed non-resonant.

Example 4.13 While satisfying theoretically, Theorems 4.5 and 4.11 may be difficult to use in practice, even if A is an integer 2×2-matrix (in Theorem 4.5), or $d = 2$ and a_1, a_2 are integers (in Theorem 4.11). To illustrate the basic difficulty, consider the innocent-looking difference equation

$$x_n = 2x_{n-1} - 5x_{n-2}, \quad n \geq 3. \tag{25}$$

	1	2	3	4	5	6	7	8	9
$N = 10$	50.00	10.00	10.00	10.00	10.00	0.00	10.00	0.00	0.00
$N = 100$	20.00	11.00	13.00	7.00	10.00	12.00	12.00	9.00	6.00
$N = 1000$	29.80	17.20	13.80	8.90	7.80	6.80	6.50	5.20	4.00
$N = 10000$	29.99	17.23	12.78	9.51	7.92	6.61	6.01	5.19	4.76
exact BL	30.10	17.60	12.49	9.69	7.91	6.69	5.79	5.11	4.57

	0	1	2	3	4	5	6	7	8	9
$N = 10$	40.00	40.00	10.00	0.00	0.00	0.00	0.00	0.00	0.00	10.00
$N = 100$	15.00	11.00	8.00	10.00	15.00	7.00	10.00	9.00	7.00	8.00
$N = 1000$	10.80	10.70	11.10	9.50	11.60	10.10	10.70	9.30	8.40	7.80
$N = 10000$	11.44	11.48	11.18	9.98	10.22	10.03	9.27	9.04	8.85	8.51
exact BL	11.96	11.38	10.88	10.43	10.03	9.66	9.33	9.03	8.75	8.49

Fig. 5 Relative frequencies of the first (top) and second significant digits for the first N terms of the solution (x_n) of (25) with $x_1 = x_2 = 1$, see Example 4.13; the data suggests that (x_n) is Benford

For the set $Z = \{z \in \mathbb{C} : z^2 = 2z - 5\} = \{1 \pm 2\imath\} = \{\sqrt{5}e^{\pm\imath \arctan 2}\}$ it is not hard to see that $\Delta_Z = \{1, 1 \pm \frac{1}{\pi} \arctan 2\}$ satisfies (i) in Definition 4.2. Thus the non-resonance of Z is equivalent to $\log 5 \notin \text{span}_{\mathbb{Q}} \Delta_Z = \text{span}_{\mathbb{Q}}\{1, \frac{1}{\pi} \arctan 2\}$. While $\log 5$ and $\frac{1}{\pi} \arctan 2$ can both be shown to be transcendental, it seems to be unknown whether or not $1, \log 5, \frac{1}{\pi} \arctan 2$ are rationally independent [46]. In other words, it is not known whether the set Z is non-resonant. In the likely case that it is, every solution of (25), except for $x_n \equiv 0$, would be Benford; otherwise, none would. Experimental evidence strongly supports the former alternative, see Fig. 5. ◁

The practical difficulty alluded to in Example 4.13 can be avoided altogether only if all eigenvalues of A, or all roots of $z^d = a_1 z^{d-1} + a_2 z^{d-2} + \cdots + a_{d-1} z + a_d$ are *real*. In this situation, the following simple observation may be helpful.

Proposition 4.14 *A set $Z \subset \mathbb{R}$ is non-resonant if and only if every $z \in Z\backslash\{0\}$ satisfies*

$$\log |z| \notin \mathbb{Q} \quad \text{and} \quad |w| \neq |z| \quad \text{for every } w \in Z\backslash\{z\}.$$

The remainder of this section is devoted to presenting proofs of Theorems 4.5 and 4.11. Both proofs are given here only under the additional assumption that,

$$\text{for every } r > 0\text{, the set } \sigma(A) \cap r\mathbb{S} \text{ contains at most two elements,} \tag{26}$$

i.e. the matrix A has at most two eigenvalues of modulus r, which may take the form of the real pair $-r, r$, or a non-real pair $\lambda, \overline{\lambda}$ with $|\lambda| = r$. (For complete proofs without this assumption, the reader is referred to [6]. Note that the matrices *not* satisfying (26) form a nowhere dense nullset in $\mathbb{R}^{d\times d}$.) For convenience, let

$$\sigma^+(A) := \{\lambda \in \sigma(A) : \Im\lambda \geq 0\}\backslash\{0\}.$$

Proof of Theorem 4.5 If $\sigma^+(A) = \varnothing$, then A is nilpotent, $\sigma(A) = \{0\}$ is non-resonant, every sequence $(x^\top A^n y)$ is identically zero for $n \geq d$, and the claimed equivalence trivially holds. Thus, from now on assume that $\sigma^+(A)$ is not empty.

Recall that, given any $x, y \in \mathbb{R}^d$, the value of $x^\top A^n y$ can be written in the form

$$x^\top A^n y = \Re\left(\sum\nolimits_{\lambda\in\sigma^+(A)} p_\lambda(n)\lambda^n\right), \quad n \geq d, \tag{27}$$

where p_λ is, for every $\lambda \in \sigma^+(A)$, a (possibly non-real) polynomial of degree at most $d-1$; moreover, p_λ is real whenever $\lambda \in \mathbb{R}$. The representation (27) follows for instance from the Jordan Normal Form Theorem. Note that p_λ also depends on x, y, but for the sake of notational clarity this dependence is not displayed explicitly.

To establish the asserted equivalence, assume first that $\sigma(A)$ is non-resonant and, given any $x, y \in \mathbb{R}^d$, that $p_\lambda \neq 0$ for some $\lambda \in \sigma^+(A)$. (Otherwise $x^\top A^n y = 0$ for all $n \geq d$.) Let

$$r := \max\{|\lambda| : \lambda \in \sigma^+(A), p_\lambda \neq 0\} > 0.$$

Recall that $\sigma(A) \cap r\mathbb{S}$ contains at most two elements. Note also that r and $-r$ cannot both be eigenvalues of A, as otherwise $\sigma(A)$ would be resonant. Hence either exactly one of the two numbers r, $-r$ is an eigenvalue of A, and $\log r$ is irrational, or else $\sigma(A) \cap r\mathbb{S} = \{re^{\pm\imath\pi\vartheta}\}$ with the appropriate irrational $0 < \vartheta < 1$, and

$$\log r \notin \operatorname{span}_{\mathbb{Q}}\{1, 1 \pm \vartheta\} = \operatorname{span}_{\mathbb{Q}}\{1, \vartheta\}.$$

In the former case, assume w.l.o.g. that r is an eigenvalue. (The case of $-r$ being an eigenvalue is completely analogous.) Recall that $|\lambda| < r$ for every other eigenvalue λ of A with $p_\lambda \neq 0$. Denote by $k \in \{0, 1, \ldots, d-1\}$ the degree of the polynomial p_r, and let $\gamma := \lim_{n\to\infty} p_r(n)/n^k$. Note that γ is non-zero and real. From (27), it follows that

$$|x^\top A^n y| = \left|p_r(n)r^n + \sum\nolimits_{\lambda\in\sigma^+(A):|\lambda|<r} p_\lambda(n)\lambda^n\right| = r^n n^k|\gamma + z_n|, \quad n \geq d,$$

with the (real) sequence (z_n) given by

$$z_n = \frac{p_r(n)}{n^k} - \gamma + \frac{1}{r^n n^k}\sum\nolimits_{\lambda\in\sigma^+(A):|\lambda|<r} p_\lambda(n)\lambda^n, \quad n \geq d.$$

Clearly, $\lim_{n\to\infty} z_n = 0$. Since $\log r$ is irrational and

$$\log|x^\top A^n y| = n\log r + k\log n + \log|\gamma + z_n|, \quad n \geq d,$$

Proposition 2.3 implies that $(x^\top A^n y)$ is Benford.

In the other case, the matrix A has $\lambda_0 = re^{\imath\pi\vartheta}$ and its conjugate $\overline{\lambda_0} = re^{-\imath\pi\vartheta}$ as eigenvalues, and $|\lambda| < r$ for every other eigenvalue λ of A with $p_\lambda \neq 0$. With k denoting the degree of p_{λ_0}, let again $\gamma := \lim_{n\to\infty} p_{\lambda_0}(n)/n^k$, and note that γ may now be non-real, yet is non-zero as before. Deduce from (27) that

$$|x^\top A^n y| = \left|\Re\left(p_{\lambda_0}(n) r^n e^{\imath\pi n\vartheta} + \sum\nolimits_{\lambda\in\sigma^+(A):|\lambda|<r} p_\lambda(n)\lambda^n\right)\right| = r^n n^k |\Re(\gamma e^{\imath\pi n\vartheta} + z_n)|,$$

with the (possibly non-real) sequence (z_n), given by

$$z_n = \left(\frac{p_{\lambda_0}(n)}{n^k} - \gamma\right) e^{\imath\pi n\vartheta} + \frac{1}{r^n n^k} \sum\nolimits_{\lambda\in\sigma^+(A):|\lambda|<r} p_\lambda(n)\lambda^n,$$

again satisfying $\lim_{n\to\infty} z_n = 0$. Since ϑ is irrational due to the non-resonance of $\sigma(A)$, the set $I := \{n \in \mathbb{N} : \Re(\gamma e^{\imath\pi n\vartheta} + z_n) = 0\}$ has density zero, that is, $\lim_{n\to\infty} \frac{1}{n}\#(I \cap \{1, 2, \ldots, n\}) = 0$, and

$$\log|x^\top A^n y| = n\log r + k\log n + \log|\Re(\gamma e^{\imath\pi n\vartheta} + z_n)|, \quad \forall n \in \mathbb{N}\backslash I. \tag{28}$$

(The reader familiar with the Skolem–Mahler–Lech Theorem [34, Thm. A] will notice that I is actually *finite*, though this much stronger property is not needed here.) Lemma 2.6 with $\vartheta_1 = \log r$, $\vartheta_2 = \vartheta$, $\alpha = k$, $z = \gamma$ shows that the sequence on the right in (28) is u.d. mod 1, and so is $(\log|x^\top A^n y|)$, by Proposition 2.3. Thus $(x^\top A^n y)$ is Benford, and the proof of (i)$\Rightarrow$(ii) is complete.

To establish the reverse implication, assume that $\sigma(A)$ is resonant. Then, for some $r_0 > 0$ and with $Z := \sigma(A) \cap r_0\mathbb{S}$, the set Δ_Z contains rational numbers other than 1, or $\log r_0 \in \mathrm{span}_{\mathbb{Q}}\Delta_Z$, or both. Assume first that $1 + \rho \in \Delta_Z$ for some rational number $\rho > 0$. This implies that Z contains exactly two elements, either $r_0, -r_0$ or else $r_0 e^{\pm\imath\pi\rho}$. In the former case, let b, c be unit eigenvectors corresponding to, respectively, the eigenvalues r_0 and $-r_0$ of A, and let $x := y := b + c$. Then

$$x^\top A^n y = (b + c)^\top \big(r_0^n b + (-r_0)^n c\big) = (1 + b^\top c)\big(r_0^n + (-r_0)^n\big).$$

By the Cauchy–Schwarz inequality, $1 + b^\top c > 0$. Hence $x^\top A^n y = 0$ for all odd n but $x^\top A^n y > 0$ for all even n, and $(x^\top A^n y)$ is neither Benford nor terminating. In the case of non-real eigenvalues, there exist linearly independent unit vectors $b, c \in \mathbb{R}^d$ such that, for every $n \in \mathbb{N}_0$,

$$A^n b = r_0^n\cos(\pi n\rho)b - r_0^n\sin(\pi n\rho)c, \quad A^n c = r_0^n\sin(\pi n\rho)b + r_0^n\cos(\pi n\rho)c. \tag{29}$$

Hence with $x := y := b + c$,

$$x^\top A^n y = r_0^n (b+c)^\top \big((\cos(\pi n\rho) + \sin(\pi n\rho))b + (\cos(\pi n\rho) - \sin(\pi n\rho))c\big)$$
$$= 2(1+b^\top c) r_0^n \cos(\pi n \rho),$$

and again $x^\top A^n y = 0$ periodically but no identically. Thus $(x^\top A^n y)$ is neither Benford nor terminating.

It remains to consider the case where $\#(\Delta_Z \cap \mathbb{Q}) \le 1$ for every $Z = \sigma(A) \cap r\mathbb{S}$ and $r > 0$, yet $\log r_0 \in \mathrm{span}_{\mathbb{Q}} \Delta_Z$ for some $r_0 > 0$. Again it is helpful to distinguish two cases: either $\sigma(A) \cap r_0\mathbb{S} \subset \mathbb{R}$ or $\sigma(A) \cap r_0\mathbb{S} \subset \mathbb{C}\backslash\mathbb{R}$. In the former case, exactly one of the two numbers r_0 and $-r_0$ is an eigenvalue of A. The argument for $-r_0$ being analogous, assume w.l.o.g. that $\sigma(A) \cap r_0\mathbb{S} = \{r_0\}$. Then $\Delta_Z = \{1\}$ and hence $\log r_0$ is rational. Taking $x := y := b$, where b is any eigenvector of A corresponding to the eigenvalue r_0, yields $x^\top A^n y = r_0^n |b|^2$, and $(x^\top A^n y)$ is neither Benford nor terminating. In the other case, i.e. for $Z = \sigma(A) \cap r_0\mathbb{S} = \{r_0 e^{\pm \imath \pi \vartheta}\}$ with some irrational $0 < \vartheta < 1$, pick again linearly independent unit vectors $b, c \in \mathbb{R}^d$ such that (29) holds for all n, with ρ replaced by ϑ. With $x := y := b + c$, it follows that

$$\log |x^\top A^n y| = n \log r_0 + \log |2(1 + b^\top c)\cos(\pi n \vartheta)|.$$

Recall that $\log r_0 \in \mathrm{span}_{\mathbb{Q}}\{1, \vartheta\}$, by assumption. An application of Lemma 2.6 with $\vartheta_1 = \log r_0$, $\vartheta_2 = \vartheta$, $\alpha = 0$, $z = 2(1 + b^\top c) > 0$ and $z_n \equiv 0$ shows that the sequence $(x^\top A^n y)$ is not Benford. Clearly, it is not terminating either. Thus (ii) $\Rightarrow$ (i), as claimed.

To complete the proof, the assertions regarding $(|A^n x|)$ and $(|A^n|)$ have to be verified. The above argument establishing (i) $\Rightarrow$ (ii) can be used to verify the former assertion because, for every $x \in \mathbb{R}^d$, $|A^n x|^2 = \sum_{j=1}^d (e_j^\top A^n x)^2$, and so $(\log |A^n x|)$ is easily seen to be u.d. mod 1 using Lemma 2.6. Finally, if A is not nilpotent (otherwise $(|A^n|)$ obviously is terminating) assume first that $\sigma(A) \cap r_\sigma(A)\mathbb{S} = \{r_\sigma(A) e^{\pm \imath 2\pi\vartheta}\}$ for some irrational $0 < \vartheta < \frac{1}{2}$. Then, with the appropriate integer $k \ge 0$,

$$(A^n)^\top A^n = r_\sigma(A)^{2n} n^{2k} \big(B(n\vartheta)^\top B(n\vartheta) + C_n\big), \quad n \in \mathbb{N},$$

where the function $B : \mathbb{R} \to \mathbb{R}^{d\times d}$ is 1-periodic, real-analytic and does not vanish identically, and (C_n) is a sequence in $\mathbb{R}^{d\times d}$ with $|C_n| \to 0$. It follows that

$$\log |A^n| = n \log r_\sigma(A) + k \log n + \tfrac{1}{2} \log r_\sigma \big(B(n\vartheta)^\top B(n\vartheta) + C_n\big).$$

Note that $r_\sigma\big(B(t)^\top B(t)\big) > 0$ for all but finitely many $t \in [0, 1)$. By the assumed non-resonance of $\sigma(A)$, $\log r_\sigma(A) \notin \mathrm{span}_{\mathbb{Q}}\{1, \vartheta\}$, and hence [5, Lem. 2.9] shows that $(|A^n|)$ is Benford. As the argument for the case $\sigma(A) \cap r_\sigma(A)\mathbb{S} \subset \mathbb{R}$ is completely analogous, the proof is complete. $\square$

Proof of Theorem 4.11 Note first that the matrix A associated with (22) via (5) is invertible because $a_d \ne 0$. Hence (27) is valid for all $n \in \mathbb{N}$, and the sequence $(x^\top A^n y)$ vanishes identically unless $p_\lambda \ne 0$ for some $\lambda \in \sigma^+(A)$. Also note that

$\sigma(A)$ simply equals $\{z \in \mathbb{C} : z^d = a_1 z^{d-1} + a_2 z^{d-2} + \cdots + a_{d-1} z + a_d\}$. Thus, if the latter set is non-resonant then $(x_n) = (e_d^\top A^n y)$ with $y = A^{-1} \sum_{j=1}^d x_{d+1-j} e_j$ either is Benford or else vanishes identically. This shows that (i) ⇒ (ii).

To establish the reverse implication, assume that $\sigma(A)$ is resonant, and distinguish cases just as in the above proof of Theorem 4.5. If $\sigma(A)$ is resonant due to failure of (i) in Definition 4.2 then, for some $r_0 > 0$ and rational number $\rho \in (0, 1)$, either $\{-r_0, r_0\} \subset \sigma(A)$ or $\{r_0 e^{\pm \imath \pi \rho}\} \subset \sigma(A)$. In the former case, $(x_n) = \big(r_0^n + (-r_0)^n\big)$ solves (22) and is neither Benford nor zero. In the latter case, the same is true for $(x_n) = \big(r_0^n \cos(\pi n \rho)\big)$. If, on the other hand, $\sigma(A)$ is resonant due to failure of (ii) then, for some $r_0 > 0$ and irrational $\vartheta \in (0, 1)$, either $r_0 \in \sigma(A)$ and $\log r_0 \in \mathbb{Q}$, or else $\{r_0 e^{\pm \imath \pi \vartheta}\} \subset \sigma(A)$ and $\log r_0 \in \mathrm{span}_{\mathbb{Q}}\{1, \vartheta\}$. In the first case, $(x_n) = (r_0^n)$ solves (22) and is neither Benford nor zero. In the second case, $(x_n) = \big(r_0^n \cos(n\pi\vartheta)\big)$ is a non-zero solution of (22) that is not Benford since

$$\big(\log |r_0^n \cos(n\pi\vartheta)|\big) = \big(n \log r_0 + \log|\cos(n\pi\vartheta)|\big)$$

is not u.d. mod 1, by Lemma 2.6. Overall, (ii) ⇒ (i), and the proof is complete. □

5 An Application to Markov Chains

If A is a real $d \times d$-matrix and $\log r_\sigma(A)$ is rational then, as an immediate consequence of Theorem 4.5, the sequence $(x^\top A^n y)$ is, for most $x, y \in \mathbb{R}^d$, *not* Benford. If, in addition, A happens to have a positive power then, for instance, none of the entries $([A^n]_{jk})$ is Benford, according to Corollary 3.3. Even in this situation, however, it is quite possible that all entries of $(A^{n+1} - r_\sigma(A) A^n)$, and in fact most sequences $\big(x^\top (A^{n+1} - r_\sigma(A) A^n) y\big)$, are Benford. This phenomenon has already been observed in Examples 3.10 and 3.11 for the matrix $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{bmatrix}$, for which $\log r_\sigma(A) = \log 1 = 0$, and yet

$$A^{n+1} - A^n = \left(-\tfrac{1}{2}\right)^{n+1} \begin{bmatrix} 1 & -1 \\ -2 & 2 \end{bmatrix}, \quad n \in \mathbb{N}_0,$$

hence most sequences $(x^\top (A^{n+1} - A^n) y)$ are Benford. The purpose of the present section is to study this "cancellation of resonance" scenario and to demonstrate how it can be understood easily by utilizing the results from previous sections. The scenario is of particular interest in the case of *stochastic* matrices which often arise as transition probability matrices of finite-state Markov chains. (As observed in Example 4.9(iii), the spectrum of every stochastic matrix is resonant.) However, "cancellation of resonance" may occur whenever A has a dominant simple eigenvalue, and it is in this more general and transparent setting that the main result, Theorem 5.1 below, is formulated. The specific result for Markov chains is then a simple corollary (Corollary 5.4).

Assume, therefore, that the real $d \times d$-matrix A has a dominant eigenvalue λ_0 that is algebraically simple, i.e. $|\lambda| < |\lambda_0|$ for every $\lambda \in \sigma(A)\backslash\{\lambda_0\}$, and λ_0 is a simple root of the characteristic polynomial of A. Note that λ_0 is necessarily a real number, and $r_\sigma(A) = |\lambda_0|$. It is not hard to see that under these assumptions the limit

$$Q_A := \lim_{n\to\infty} \frac{A^n}{\lambda_0^n} \tag{30}$$

exists. Moreover, it is clear from (30) that $Q_A A = A Q_A = \lambda_0 Q_A$, but also $Q_A^2 = Q_A$. In fact, Q_A is nothing but the spectral projection associated with λ_0 and can also be represented in the form

$$Q_A = \frac{bc^\top}{b^\top c}, \tag{31}$$

where b, c are eigenvectors of, respectively, A and $A^\top$ corresponding to the eigenvalue λ_0. A dominant, algebraically simple eigenvalue is often observed in practice. For instance, it occurs whenever $A^N > 0$ for some $N \in \mathbb{N}$, see Proposition 3.1. (In this case even $Q_A > 0$.) But it also occurs for matrices such as e.g.

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix},$$

of which no power is positive.

Consider now the sequences $(A^{n+1} - \lambda_0 A^n)$ and $(A^n - \lambda_0^n Q_A)$, both of which in a sense measure the speed of convergence in (30) and therefore are often of interest in their own right. Using the results of Sect. 4, the Benford behaviour of these sequences is easily analysed.

Theorem 5.1 *Assume $A \in \mathbb{R}^{d\times d}$ has a dominant eigenvalue λ_0 that is algebraically simple, and let Q_A be the associated projection according to* (31). *Then the following three statements are equivalent:*

(i) *The set $\sigma(A)\backslash\{\lambda_0\}$ is non-resonant;*
(ii) *The sequence $(x^\top(A^{n+1} - \lambda_0 A^n)y)$ is Benford or terminating for every $x, y \in \mathbb{R}^d$;*
(iii) *The sequence $(x^\top(A^n - \lambda_0^n Q_A)y)$ is Benford or terminating for every $x, y \in \mathbb{R}^d$.*

Proof Since all assertions are trivially correct for $d = 1$, assume $d \geq 2$ from now on, and hence $\lambda_0 \neq 0$. As in the proof of Theorem 3.2, let $R := A - \lambda_0 Q_A$ and observe that $AR = RA$ as well as $Q_A R = 0 = R Q_A$, and hence

$$A^n = \lambda_0^n Q_A + R^n, \quad \forall n \geq 1. \tag{32}$$

Note that, for every $\lambda \in \sigma(A)\backslash\{\lambda_0\}$ and $x \in \mathbb{R}^d$ with $(A - \lambda I_d)^m x = 0$ (i.e. $x \neq 0$ is a generalized eigenvector of A corresponding to the eigenvalue $\lambda \neq \lambda_0$),

$$0 = c^\top (A - \lambda I_d)^m x = \left((A^\top - \lambda I_d)^m c\right)^\top x = (\lambda_0 - \lambda)^m c^\top x,$$

and so $c^\top x = 0$, which in turn implies $Q_A x = 0$, and hence $A^n x = R^n x$ for all n, by (32), and $\lambda \in \sigma(R)$. On the other hand, $Ab = \lambda_0 b = \lambda_0 Q_A b$ and therefore $Rb = 0$. Thus $0 \in \sigma(R)$. Also, if $Rx = \lambda_0 x$ for some $x \in \mathbb{R}^d$ then (32) yields

$$Q_A x = \lim_{n\to\infty} \frac{A^n x}{\lambda_0^n} = Q_A x + x,$$

hence $x = 0$. In other words, $\lambda_0 \notin \sigma(R)$, and overall $\sigma(R) = (\sigma(A)\backslash\{\lambda_0\}) \cup \{0\}$, showing that $\sigma(R)$ is non-resonant if and only if $\sigma(A)\backslash\{\lambda_0\}$ is non-resonant. Moreover, deduce from (30) and (32) that

$$A^{n+1} - \lambda_0 A^n = R^n (R - \lambda_0 I_d), \quad A^n - \lambda_0^n Q_A = R^n, \quad \forall n \geq 1.$$

Since $R - \lambda_0 I_d$ is invertible, the asserted equivalences are now obvious from Theorem 4.5. □

Example 5.2 (i) The (positive) matrix $B = \begin{bmatrix} 6 & 4 \\ 4 & 6 \end{bmatrix}$ first encountered in Example 3.6 has the dominant simple eigenvalue $\lambda_0 = 10$. Thus Theorem 5.1 applies, with

$$Q_B = \lim_{n\to\infty} \frac{B^n}{\lambda_0^n} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Since $\sigma(B)\backslash\{10\} = \{2\}$ is non-resonant, every sequence $(x^\top(B^{n+1} - 10B^n)y)$ and $(x^\top(B^n - 10^n Q_B)y)$ is Benford or terminating. This can also be seen directly from

$$B^{n+1} - 10B^n = -2^{n+2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad B^n - 10^n Q_B = 2^{n-1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad n \in \mathbb{N}_0.$$

(ii) For the (non-negative) matrix $B = \begin{bmatrix} 19 & 20 \\ 1 & 0 \end{bmatrix}$ from Example 3.10, $\lambda_0 = 20$ is a dominant simple eigenvalue, and

$$Q_B = \frac{1}{21}\begin{bmatrix} 20 & 20 \\ 1 & 1 \end{bmatrix}.$$

However, $\sigma(B)\backslash\{20\} = \{-1\}$ is resonant, and hence some (in fact, most) sequences $(x^\top(B^{n+1} - 20B^n)y)$ and $(x^\top(B^n - 20^n Q_B)y)$ are neither Benford nor terminating. Again, this can be confirmed by an explicit calculation as well.

(iii) The matrix $C = \begin{bmatrix} 6 & -8 \\ 4 & -6 \end{bmatrix}$ does not have a dominant eigenvalue, as $\sigma(C) = \{\pm 2\}$, and hence Theorem 5.1 does not apply. Correspondingly, the limit $\lim_{n\to\infty} C^n/2^n$ does not exist. Note, however, that every entry of $(C^{n+1} - 2C^n)$, for instance, is Benford, as

$$C^{n+1} - 2C^n = 2(-2)^{n+1} \begin{bmatrix} -1 & 2 \\ -1 & 2 \end{bmatrix}, \quad n \in \mathbb{N}_0. \qquad \triangleleft$$

Example 5.3 (i) The spectrum of

$$A = \begin{bmatrix} -10 & 15 & 15 \\ -24 & 29 & 27 \\ 24 & -24 & -22 \end{bmatrix}$$

equals $\sigma(A) = \{-10, 2, 5\}$, and hence is resonant, yet $\lambda_0 = -10$ is a dominant simple eigenvalue, and $\sigma(A)\backslash\{-10\} = \{2, 5\}$ is non-resonant. By Theorem 5.1, every entry of $(A^{n+1} + 10A^n)$ and $(A^n - (-10)^n Q_A)$, in particular, is Benford or terminating, where

$$Q_A = \begin{bmatrix} 1 & -1 & -1 \\ 2 & -2 & -2 \\ -2 & 2 & 2 \end{bmatrix}.$$

(ii) Consider the (non-negative) 3×3-matrix

$$B = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Clearly, $\lambda_0 = 3$ is a dominant eigenvalue, and $\sigma(B)\backslash\{3\} = \{2\}$ is non-resonant. However,

$$B^n = \begin{bmatrix} 3^n & n3^{n-1} & 0 \\ 0 & 3^n & 0 \\ 0 & 0 & 2^n \end{bmatrix}, \quad n \in \mathbb{N}_0,$$

and so $\lim_{n\to\infty} B^n/3^n$ does not exist. The reason for this is that the eigenvalue λ_0, although dominant, is not simple. Thus Theorem 5.1 does not apply. Nevertheless, $(x^\top(B^{n+1} - 3B^n)y)$ is Benford or terminating for every $x, y \in \mathbb{R}^3$. $\triangleleft$

Remark A close inspection of the proof of Theorem 5.1 shows that the assumption of algebraic simplicity for λ_0 can be relaxed somewhat. As a matter of fact, Theorem 5.1 remains unchanged if the dominant eigenvalue λ_0 is merely assumed to be *semi-simple*, meaning that its algebraic and geometric multiplicities coincide or, equivalently, that $A - \lambda_0 I_d$ and $(A - \lambda_0 I_d)^2$ have the same rank.

Arguably the most important application of Theorem 5.1 is to stochastic matrices. Recall that $A \in \mathbb{R}^{d\times d}$ is row-stochastic (column-stochastic) if $A \geq 0$ and the entries in each row (column) add up to 1; recall also that $r_\sigma(A) = 1 \in \sigma(A)$ for every (row- or column-) stochastic matrix. In probability textbooks, the letters P, Q etc. are traditionally used to denote stochastic matrices, a tradition adhered to for the remainder of this section. If $P \in \mathbb{R}^{d\times d}$ is a (row-) stochastic matrix, then it can naturally be interpreted as the matrix of 1-step transition probabilities of a time-homogeneous d-state Markov chain (X_n), i.e. (X_n) is a discrete-time Markov process on d symbols, $s_1, s_2, \ldots, s_d$, and, for every $n \in \mathbb{N}$,

$$\mathbb{P}(X_{n+1} = s_k | X_n = s_j) = [P]_{jk} \quad \forall j, k \in \{1, 2, \ldots, d\}. \tag{33}$$

As a consequence of (33), the N-step transition probabilities are simply given by the entries of P^N, that is, for every $n \in \mathbb{N}$,

$$\mathbb{P}(X_{n+N} = s_k | X_n = s_j) = [P^N]_{jk} \quad \forall j, k \in \{1, 2, \ldots, d\}.$$

Thus the long-term behaviour of the stochastic process (X_n) is governed by the sequence of (stochastic) matrices (P^n). Moreover, if $|\lambda| < 1$ for every eigenvalue $\lambda \neq 1$ of P then $Q_P := \lim_{n\to\infty} P^n$ exists and is itself a stochastic matrix. A common problem in many Markov chain models is to estimate Q_P through numerical simulation. In this context, the sequences $(P^{n+1} - P^n)$ and $(P^n - Q_P)$ are of special interest, as they both in a sense measure the speed of convergence of $P^n \to Q_P$. They are also rich sources of Benford sequences.

Corollary 5.4 [10, Thm. 12] *Assume that the stochastic matrix $P \in \mathbb{R}^{d\times d}$ is irreducible and aperiodic, and let $Q_P := \lim_{n\to\infty} P^n$. If $\sigma(P)\backslash\{1\}$ is non-resonant then, for every $j, k \in \{1, 2, \ldots, d\}$, the sequences $([P^{n+1} - P^n]_{jk})$ and $([P^n - Q_P]_{jk})$ are Benford or terminating.*

Proof Since P is irreducible and aperiodic, $P^N > 0$ for some $N \in \mathbb{N}$, and hence $\lambda_0 = 1$ is a dominant, algebraically simple eigenvalue of P. The claim then follows from Theorem 5.1. □

Example 5.5 For the stochastic matrix

$$P = \frac{1}{10}\begin{bmatrix} 9 & 0 & 1 \\ 6 & 3 & 1 \\ 1 & 0 & 9 \end{bmatrix},$$

$\sigma(P)\backslash\{1\} = \left\{\frac{3}{10}, \frac{4}{5}\right\}$ is non-resonant. Note that P fails to be irreducible, and hence Corollary 5.4 does not apply directly. However, $\lambda_0 = 1$ obviously is dominant and simple, and so Theorem 5.1 can be used to deduce that every entry of $(P^{n+1} - P^n)$ and $(P^n - Q_P)$ is Benford or terminating, with

$$Q_P = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

◁

Example 5.6 Consider the (irreducible and aperiodic) stochastic matrix

$$P = \frac{1}{30}\begin{bmatrix} 14 & 11 & 5 \\ 11 & 14 & 5 \\ 5 & 5 & 20 \end{bmatrix},$$

for which $\sigma(P)\backslash\{1\} = \{\frac{1}{10}, \frac{1}{2}\}$ is resonant. While Corollary 5.4 does not apply, Theorem 5.1 shows that there exist $x, y \in \mathbb{R}^3$ for which $(x^\top(P^{n+1} - P^n)y)$, for instance, is neither Benford nor terminating. For a concrete example that is neither, simply take $x = e_1$, $y = e_1 - e_2$, which yields $(x^\top(P^{n+1} - P^n)y) = (10^{-n})$. On the other hand, with

$$Q_P = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

it is straightforward to check that all entries of $(P^{n+1} - P^n)$ and $(P^n - Q_p)$ are Benford. Thus the non-resonance of $\sigma(P)\backslash\{1\}$ is not necessary for the latter property. In other words, the implication in Corollary 5.4 cannot in general be reversed. Moreover, the property asserted by Corollary 5.4, i.e. the property that all entries of $(P^{n+1} - P^n)$ and $(P^n - Q_P)$ are Benford or terminating, is not a spectral property of P. To see this, consider for example

$$\widetilde{P} = \frac{1}{10}\begin{bmatrix} 6 & 3 & 1 \\ 3 & 4 & 3 \\ 1 & 3 & 6 \end{bmatrix},$$

and note that $\sigma(\widetilde{P}) = \sigma(P)$ and $Q_{\widetilde{P}} = Q_P$. Again, it is readily confirmed that, for instance $([\widetilde{P}^{n+1} - \widetilde{P}^n]_{22}) = (-\frac{3}{5}10^{-n})$ and $([\widetilde{P}^n - Q_{\widetilde{P}}]_{22}) = (\frac{2}{3}10^{-n})$, and both sequences are neither Benford nor terminating. $\triangleleft$

The situation described in Corollary 5.4 is very common among stochastic matrices. To put this more formally, denote by $\mathscr{P}_d$ the family of all (row-) stochastic $d \times d$-matrices, that is

$$\mathscr{P}_d = \left\{P \in \mathbb{R}^{d\times d} : P \geq 0, \sum\nolimits_{k=1}^{d} [P]_{jk} = 1 \ \forall j = 1, 2, \ldots, d\right\}.$$

The set $\mathscr{P}_d$ is a compact and convex subset of $\mathbb{R}^{d\times d}$. For example,

$$\mathscr{P}_1 = \{[1]\} \quad \text{and} \quad \mathscr{P}_2 = \left\{\begin{bmatrix} s & 1-s \\ 1-t & t \end{bmatrix} : 0 \leq s, t \leq 1\right\}.$$

Note that $\mathscr{P}_d$ can be identified with a d-fold copy of the standard $(d-1)$-simplex, that is, $\mathscr{P}_d \simeq \{x \in \mathbb{R}^d : x \geq 0, \sum_{j=1}^{d} x_j = 1\}^d$, and hence carries the (normalized) $d(d-1)$-dimensional Lebesgue measure Leb. Consider now

$$\mathcal{H}_d := \Big\{ P \in \mathcal{P}_d : P \text{ is irreducible and aperiodic, and } \sigma(P)\backslash\{1\} \text{ is non-resonant} \Big\}.$$

Thus $\mathcal{H}_d$ is exactly the family of stochastic matrices covered by Corollary 5.4. For instance, $\mathcal{H}_1 = \{[1]\} = \mathcal{P}_1$,

$$\mathcal{H}_2 = \left\{ \begin{bmatrix} s & 1-s \\ 1-t & t \end{bmatrix} : 0 \le s, t < 1,\ s+t = 1 \text{ or } \log|s+t-1| \notin \mathbb{Q} \right\},$$

and in both cases $\mathcal{H}_d$ constitues *most* of $\mathcal{P}_d$. The latter can be shown to be true in general: For every $d \in \mathbb{N}$, the complement of $\mathcal{H}_d$ in $\mathcal{P}_d$ is a set of first category and has Leb-measure zero. Thus if P is a $\mathcal{P}_d$-valued random variable, i.e. a random stochastic matrix, whose distribution is absolutely continuous (w.r.t. Leb, which means that $\mathbb{P}(P \in C) = 0$ whenever $C \subset \mathcal{P}_d$ and $\text{Leb}(C) = 0$), then $\mathbb{P}(P \in \mathcal{H}_d) = 1$. Together with Corollary 5.4, this implies

Corollary 5.7 [10, Thm. 17] *If the random stochastic matrix P has an absolutely continuous distribution then with probability one, P is irreducible and aperiodic, and every sequence $([P^{n+1} - P^n]_{jk})$ and $([P^n - Q_P]_{jk})$ is Benford or terminating.*

Note that for example the random stochastic matrix P has an absolutely continuous distribution whenever its d rows are chosen independently according to the same density on the standard $(d-1)$-simplex.

While the above genericity properties are very similar to the corresponding results for arbitrary matrices (see the *Remark* on p. 23), they do not follow directly from the latter. In fact, they are somewhat harder to prove, as they assert (topological as well as measure-theoretic) prevalence of $\mathcal{H}_d$ within the space $\mathcal{P}_d$ which, as a subset of $\mathbb{R}^{d\times d}$, is itself a nowhere dense nullset. The interested reader may want to consult [10] for details.

Acknowledgments This work was supported by an NSERC Discovery Grant. The authors are grateful to T. Hill, A. Weiss, and R. Zweimüller for helpful conversations.

References

1. Anderson, T.C., Rolen, L., Stoehr, R.: Benford's law for coefficients of modular forms and partition functions. Proc. Am. Math. Soc. **139**, 1533–1541 (2011)
2. Arnold, V.I.: Ordinary Differential Equations. Springer, Berlin (1992)
3. Benford, F.: The law of anomalous numbers. Proc. Am. Philos. Soc. **78**, 551–572 (1938)
4. Benford Online Bibliography. http://www.benfordonline.net
5. Berger, A.: Multi-dimensional dynamical systems and Benford's law. Discrete Contin. Dyn. Syst. **13**, 219–237 (2005)
6. Berger, A., Eshun, G.: A characterization of Benford's law in discrete-time linear systems. (In preparation) (2014)
7. Berger, A., Hill, T.P.: A basic theory of Benford's law. Probab. Surv. **8**, 1–126 (2011)
8. Berger, A., Hill, T.P.: Benfords law strikes back: no simple explanation in sight for mathematical gem. Math. Intell. **33**, 8591 (2011)

9. Berger, A., Siegmund, S.: On the distribution of mantissae in nonautonomous difference equations. J. Difference Equ. Appl. **13**, 829–845 (2007)
10. Berger, A., Hill, T.P., Kaynar, B., Ridder, A.: Finite-state Markov chains obey Benford's law. SIAM J. Matrix Anal. Appl. **32**, 665–684 (2011)
11. Brown, J.L., Duncan, R.L.: Modulo one uniform distribution of the sequence of logarithms of certain recursive sequences. Fibonacci Quart. **8**, 482–486 (1970)
12. Carslaw, C.A.P.N.: Anomalies in income numbers: evidence of goal oriented behavior. Account. Rev. **63**, 321–327 (1988)
13. Cho, W.K.T., Gaines, B.J.: Breaking the (Benford) law: statistical fraud detection in campaign finance. Am. Statist. **61**, 218–223 (2007)
14. Cohen, D.I.A., Katz, T.M.: Prime numbers and the first digit phenomenon. J. Number Theory **18**, 261–268 (1984)
15. Dajani, K., Kraikamp, C.: Ergodic Theory of Numbers. Carus Mathematical Monographs 29. Mathematical Association of America, Washington DC (2002)
16. Deckert, J., Myagkov, M., Ordeshook, P.C.: Benford's law and the detection of election fraud. Polit. Anal. **19**, 245–268 (2011)
17. Diaconis, P.: The distribution of leading digits and uniform distribution mod 1. Ann. Probab. **5**, 72–81 (1977)
18. Diekmann, A.: Not the first digit! using Benford's law to detect fraudulent scientific data. J. Appl. Stat. **34**, 321–329 (2007)
19. Docampo, S., del Mar Trigo, M., Aira, M.J., Cabezudo, B., Flores-Moya, A.: Benfords law applied to aerobiological data and its potential as a quality control tool. Aerobiologia **25**, 275–283 (2009)
20. Duncan, R.L.: An application of uniform distributions to the Fibonacci numbers. Fibonacci Quart. **5**, 137–140 (1967)
21. Geyer, C.L., Williamson, P.P.: Detecting fraud in data sets using Benford's law. Commun. Stat. Simul. Comput. **33**, 229–246 (2004)
22. Hill, T.P.: Base-invariance implies Benford's law. Proc. Am. Math. Soc. **123**, 887–895 (1995)
23. Hill, T.P.: The significant-digit phenomenon. Am. Math. Monthly **102**, 322–327 (1995)
24. Hill, T.P.: A statistical derivation of the significant-digit law. Stat. Sci. **10**, 354–363 (1995)
25. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (1985)
26. Kanemitsu, S., Nagasaka, K., Rauzy, G., Shiue, J-S.: On Benford's law: the first digit problem. In: Probability Theory and Mathematical Statistics (Kyoto, 1986). Lecture Notes in Math. 1299, 158–169 (1988)
27. Kontorovich, A.V., Miller, S.J.: Benford's law, values of L-functions and the $3x + 1$ problem. Acta Arith. **120**, 269–297 (2005)
28. Kuipers, L.: Remark on a paper by R.L. Duncan concerning the uniform distribution mod 1 of the sequence of the logarithms of the fibonacci numbers. Fibonacci Quart. 7, 465–466, 473 (1969)
29. Kuipers, L., Niederreiter, H.: Uniform Distribution of Sequences. Wiley, New York (1974)
30. Lagarias, J.C., Soundararajan, K.: Benford's law for the $3x + 1$ function. J. Lond. Math. Soc. **74**, 289–303 (2006)
31. Mebane, W.R.: Fraud in the 2009 presidential election in Iran? Chance **23**, 6–15 (2010)
32. Miller, S.J., Nigrini, M.J.: Order statistics and Benford's law. Int. J. Math. Math. Sci. 382948 (2008)
33. Moore, G.B., Benjamin, C.O.: Using Benford's law for fraud detection. Intern. Auditing **19**, 4–9 (2004)
34. Myerson, G., van der Poorten, A.J.: Some problems concerning recurrence sequences. Am. Math. Monthly **102**, 698–705 (1995)
35. Nagasaka, K., Shiue, J.-S.: Benford's law for linear recurrence sequences. Tsukuba J. Math. **11**, 341–351 (1987)
36. Newcomb, S.: Note on the frequency of use of the different digits in natural numbers. Am. J. Math. **4**, 39–40 (1881)

37. Nigrini, M.J.: A taxpayer compliance application of Benford's law. J. Am. Tax. Association **18**, 72–91 (1996)
38. Nigrini, M.J.: Digital Analysis Using Benford's Law: Tests Statistics for Auditors. Global Audit, Vancouver (2000)
39. Raimi, R.A.: The first digit problem. Am. Math. Monthly **83**, 521–538 (1976)
40. Roukema, B.F.: A first-digit anomaly in the 2009 Iranian presidential election. Preprint arXiv:0906.2789 (2009)
41. Sambridge, M., Tkalčić, H., Jackson, A.: Benford's law in the natural sciences. Geophys. Res. Lett. **37**, L22301 (2010)
42. Schatte, P.: On H_∞-summability and the uniform distribution of sequences. Math. Nachr. **113**, 237–243 (1983)
43. Schatte, P.: On the uniform distribution of certain sequences and Benford's law. Math. Nachr. **136**, 271–273 (1988)
44. Schürger, K.: Extensions of black-scholes processes and Benford's law. Stochastic Process. Appl. **118**, 1219–1243 (2008)
45. Strzałka, D.: On some properties of Benford's law. J. Kore. Math. Soc. **47**, 1055–1075 (2010)
46. Waldschmidt, M.: Diophantine Approximation on Linear Algebraic Groups. Springer, Berlin (2000)
47. Wintner, A.: On the cyclical distribution of the logarithms of the prime numbers. Q. J. Math. **6**, 65–68 (1935)
48. Wlodarski, J.: Fibonacci and lucas numbers tend to obey Benfords law. Fibonacci Quart. **9**, 87–88 (1971)

Harvesting and Dynamics in Some One-Dimensional Population Models

Eduardo Liz and Frank M. Hilker

Abstract We review some dynamical effects induced by constant effort harvesting in single-species discrete-time population models. We choose three different forms for the density-dependent recruitment function, which include the overcompensatory Ricker map for semelparous species; a modified Ricker model allowing for adult survivorship; and a model with both strong Allee effect and overcompensation which results from incorporating mate limitation in the Ricker model. We show that these simple models exhibit some interesting (and sometimes unexpected) phenomena such as the *hydra effect*; bubbling; sudden collapses; and essential extinction. We underline the importance of two often underestimated issues that turn out to be crucial for management: census timing and intervention time.

1 Introduction

We consider discrete-time single-species models governed by a first-order difference equation

$$x_{n+1} = F(x_n), \quad n = 0, 1, 2, \dots, \tag{1}$$

E. Liz (✉)
Departamento de Matemática Aplicada II, Universidad de Vigo,
36310 Vigo, Spain
e-mail: eliz@dma.uvigo.es

F.M. Hilker
Centre for Mathematical Biology, Department of Mathematical Sciences, University of Bath,
Bath BA2 7AY, UK
e-mail: frank.hilker@uni-osnabrueck.de

F.M. Hilker
Institute of Environmental Systems Research, School of Mathematics/Computer Science,
Osnabrück University,
49069 Osnabrück, Germany

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_3

where x_n denotes the population size at the n–th generation, and F is the so-called stock–recruitment (production) function. These models are well-suited for semelparous populations [7, Chap. 4], but they also fit well to populations where a fraction of adults survive the reproduction season [3, Sect. 7.5]. We will focus our attention on the unimodal Ricker map [12]

$$F(x) = x\mathrm{e}^{r(1-x)}, \quad r > 0, \tag{2}$$

and two modifications of it. The first one assumes a survivorship rate α, $\alpha \in (0, 1)$, of adults and reads

$$F(x) = \alpha x + (1-\alpha)x\mathrm{e}^{r(1-x)}. \tag{3}$$

We refer to this function as the Ricker–Clark map; see, e.g., [3, 9, 19].

The second modification of the Ricker model that we will consider exhibits a strong Allee effect, that is, there is a critical population size (Allee threshold) below which the population cannot survive [4]. It has been used by Schreiber [14] to model mate limitation, and we will refer to it as the Ricker–Schreiber map. Its production function is

$$F(x) = \frac{\beta x}{1+\beta x} x\,\mathrm{e}^{r(1-x)}, \quad \beta > 0. \tag{4}$$

Our aim is to show how a strategy of constant effort harvesting changes the dynamics of the difference Eq. (1) when the production map F is given by (2), (3) or (4).

2 The Ricker Model and the Hydra Effect

In this section we consider the Ricker function (2), which is a prototype for overcompensatory production. See Fig. 1 for a graphic representation when $r = 3$.

A strategy of constant effort harvesting assumes that a percentage γx of the population is removed at every period. Thus, harvesting a population following the Ricker map after recruitment gives

$$x_{n+1} = (1-\gamma)x_n\,\mathrm{e}^{r(1-x_n)}, \quad n = 0, 1, 2, \ldots, \tag{5}$$

where $\gamma \in (0, 1)$. The bifurcation diagram of (5) for varying γ (Fig. 2) shows the well-known effects of increasing harvesting:

- Reducing complexity: if the unharvested population is unstable, a sufficiently large harvesting effort leads the system to a globally stable positive equilibrium through a series of period-halving bifurcations (see, e.g., [8]).
- Overharvesting leads to extinction after a transcritical bifurcation at $\gamma = 1 - \mathrm{e}^{-r}$.

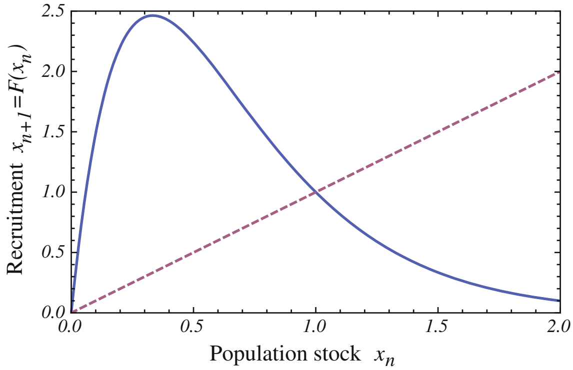

Fig. 1 Graphic representation of the Ricker stock–recruitment curve $F(x) = x\mathrm{e}^{3(1-x)}$. This *curve* is overcompensatory; this means that after a critical value of the population size, recruitment decreases with increasing population size. The intersection with the line $y = x$ (*dashed line*) is the positive equilibrium $x = 1$ (obtained from the carrying capacity after normalization)

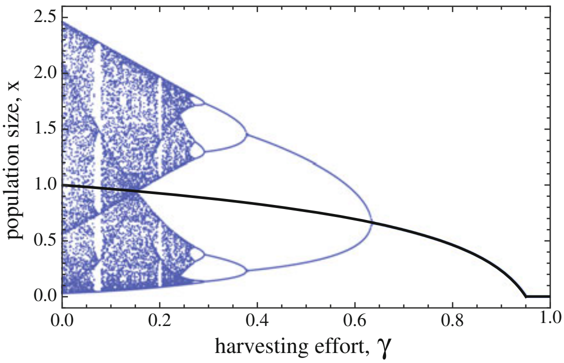

Fig. 2 Bifurcation diagram for Eq. (5) with $r = 3$ and $\gamma \in (0, 1)$. For each value of γ (with step 0.001), we produce 300 iterations of (5) with a random initial condition $x_0 \in [0, 2.5]$, and plot the last 20 iterates to let the transients die out. The *bold line* corresponds to the average population size

2.1 Census Timing and the Hydra Effect

The hydra effect is a term recently coined by P. A. Abrams and co-authors to define a seemingly paradoxical increase in the size of a population in response to an increase in its per-capita mortality [1]. One of the simplest models where this effect can be observed is a modified version of Eq. (5), namely,

$$x_{n+1} = (1-\gamma)x_n\, \mathrm{e}^{r(1-(1-\gamma)x_n)}, \quad n = 0, 1, 2, \ldots \tag{6}$$

This equation has been studied in [8, 11, 15]; see also [1] for other recruitment functions. In particular, Ref. [11] proves that the average population size for any

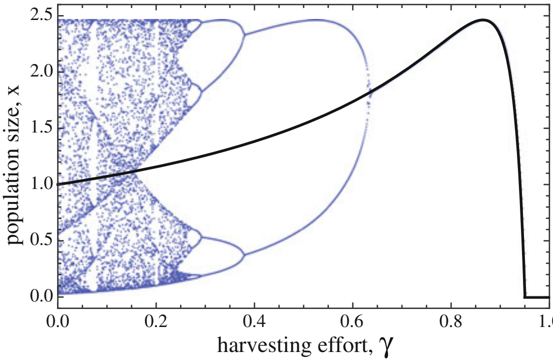

Fig. 3 Bifurcation diagram for Eq. (6) with $r = 3$ and $\gamma \in (0, 1)$. The *bold line* corresponds to the average population size

initial condition $x_0 > 0$ is an increasing function of the harvesting effort γ for all $\gamma \in \left(0, 1 - e^{1-r}\right)$. The average population size is defined by the formula

$$\phi(x_0, \gamma) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} x_i(\gamma).$$

The bifurcation diagram for $r = 3$ is shown in Fig. 3. The bold line corresponds to the average population size.

What is the relationship between models (5) and (6)? From an ecological point of view, both are models with only two processes: reproduction and harvesting. The only difference between them is the moment at which the population size is measured. Indeed, if we denote by $F(x) = xe^{r(1-x)}$ the recruitment function and by $h(x) = (1-\gamma)x$ the harvesting action, Eq. (5) corresponds to *census after harvesting*. That is, it can be written in the form

$$x_{n+1} = h(F(x_n)), \quad n = 0, 1, 2, \ldots \tag{7}$$

On the other hand, Eq. (6) corresponds to *census after reproduction*. That is, it can be written in the form

$$x_{n+1} = F(h(x_n)), \quad n = 0, 1, 2, \ldots \tag{8}$$

Following another analogy [3, see Sect. 7.1], Eq. (7) measures the dynamics of the parent stock, whereas (8) measures the dynamics of the recruits.

From a mathematical point of view, Eqs. (7) and (8) are dynamically equivalent [11]; this means that they share the same properties of stability, periodicity and chaos. However, what we observe in Figs. 2 and 3 does not appear to be the same. In particular, while (7) does not exhibit the hydra effect, (8) does do. This fact stresses the necessity of taking into account census timing when a mathematical model is used for management purposes; using Clark's terms [3, see Sect. 7.1], the same harvesting

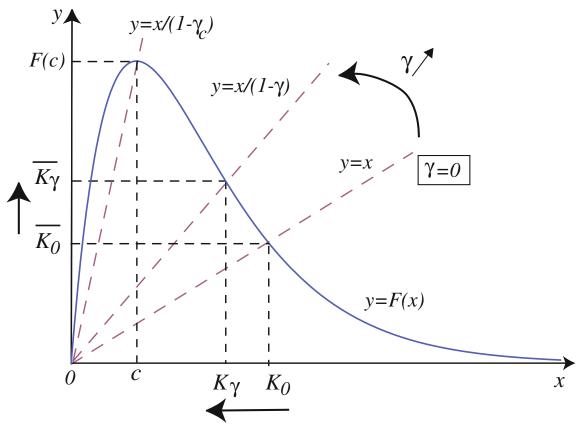

Fig. 4 Geometric representation of the positive equilibria of models (5) and (6). The equilibrium K_γ of the former decreases as γ is increased, while the equilibrium $\overline{K_\gamma}$ of the latter increases with γ until the critical value γ_c, for which the *line* $y = x/(1-\gamma)$ intersects the *curve* $y = F(x)$ at its maximum value $(c, F(c))$

model can exhibit a hydra effect when we census recruits, but it does not if we census the parent stock. Of course, the hydra effect is still present in the recruits but "hidden" since we do not measure it. For more discussion on this topic, see [5] and references therein.

We notice that the hydra effect in discrete single-species models can only occur if the density dependence is overcompensatory [1, 15]. Actually, it is easy to explain using the geometric interpretation of the positive equilibrium. Recall that for the Ricker model, the average population size matches the equilibrium even when the population oscillates. It is easy to check that the equilibrium K_γ of (5) is the projection on the x–axis of the intersection of the curve $y = F(x)$ with the line $y = x/(1-\gamma)$, and the positive equilibrium of (6) is $\overline{K_\gamma} = F(K_\gamma)$, which is the projection of the same intersection point on the y–axis. This simple observation explains why increasing harvesting produces a hydra effect in model (6) but not in (5). See Fig. 4.

2.2 Variable Harvest Timing and Its Impact on Population Dynamics

As emphasized in [6], the timing of harvesting may profoundly influence the impact on the population. The main reason is that if population growth is compensatory, then if individuals are removed at early stages in the season, the remaining individuals reproduce better. Seno [15] proposed one of the simplest models that considers harvesting at a specific point of time within the season. It assumes that individuals accumulate energy for reproduction in the course of the season and takes into account density-dependent effects in the population dynamics for the parts of the season before and after harvesting. For the Ricker map, the model is

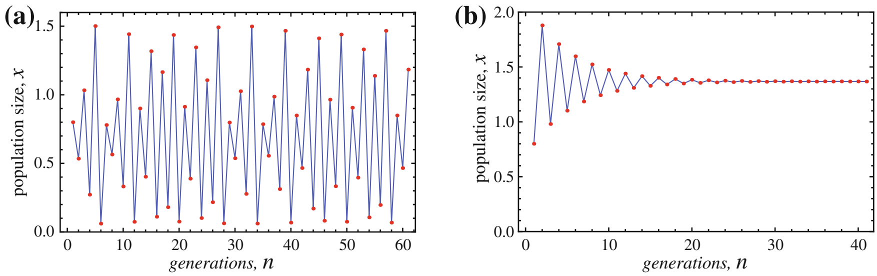

Fig. 5 Time series for Eq. (9) with $\gamma = 0.7$, $r = 4$ and different harvesting times: **a** chaotic solution for $\theta = 1$ (harvesting at the end of the reproductive season); **b** asymptotically stable positive equilibrium for $\theta = 0.7$

$$x_{n+1} = (1-\gamma)x_n \left(\theta e^{r(1-x_n)} + (1-\theta)e^{r(1-(1-\gamma)x_n)}\right), \quad n = 0, 1, 2, \dots \quad (9)$$

where $\theta \in [0, 1]$ is the moment of time in the season (assuming its length is 1) when harvest intervention takes place. The main conclusion of Seno's paper is that the hydra effect in (9) occurs for low values of θ, that is to say, the earlier we harvest, the more the average population size is increased.

It is easily seen that the right-hand side of (9) is a convex combination of the right-hand sides of Eqs. (5) and (6); cf. [2]. In other words, it can be written as

$$x_{n+1} = \theta h(F(x_n)) + (1-\theta)F(h(x_n)), \quad n = 0, 1, 2, \dots,$$

where h and F have the same meaning as in (7). Using this fact, it is easy to prove that the intervention time θ does not change the critical value of the harvesting effort γ driving the population to extinction, and that the average population size is a decreasing function of θ.

Since we have seen that the cases $\theta = 0$ and $\theta = 1$ are dynamically equivalent, an interesting problem is how intervention time affects the qualitative behaviour of the model with harvesting. This problem has been addressed in [2], and a major conclusion is that harvesting at intermediate values of the reproductive season may reduce complexity. To illustrate this fact, Fig. 5 shows the time series of a solution of Eq. (9) with $\theta = 1$ (which reduces to (5)) and $\theta = 0.7$. While the former is chaotic, the latter has a globally attracting positive equilibrium.

3 The Ricker–Clark Model and the Bubbling Effect

As we have seen in the previous section, one of the characteristics of the Ricker model with harvesting (5) is that an increasing harvesting effort stabilizes the positive equilibrium. Actually, the opposite effect is not posible: harvesting cannot destabilize

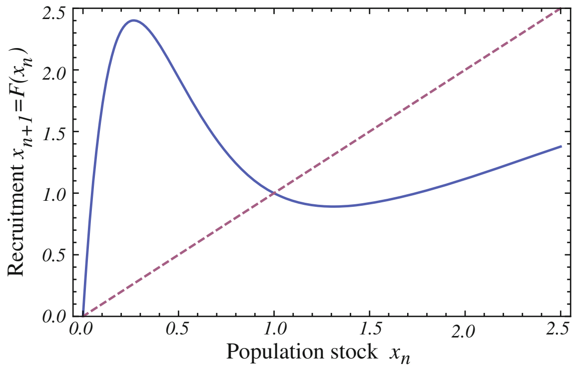

Fig. 6 Graphic representation of the Ricker–Clark stock–recruitment curve $F(x) = 0.55x + 0.45xe^{3(1-x)}$. In this case there are two critical points (a local maximum and a local minimum). The only positive equilibrium is the carrying capacity (normalized to 1), as in the Ricker map

a stable positive equilibrium [11]. However, it has recently become apparent that harvesting/fishing can magnify fluctuations in exploited populations, and some hypotheses have been proposed (see, e.g., [16] and references therein).

One of the simplest mechanisms giving rise to destabilization with increasing harvesting effort in deterministic models of discrete-time single-species populations is to allow a certain percentage of the adult population to survive the season. This yields the Ricker–Clark production function (3). Contrary to the usual Ricker map, function (3) is usually bimodal, and this fact leads to richer dynamics. See Fig. 6 for a graphic representation when $r = 4$ and $\alpha = 0.55$.

The influence of harvesting in a population governed by Eq. (1) with the Ricker–Clark function (3) has been recently studied in [9] for constant quota harvesting and in [11] for constant effort harvesting. In both cases, it was shown that for certain parameter ranges (of the adult survivorship α and production rate r) increasing harvesting can destabilize the positive equilibrium and, more generally, harvesting can magnify fluctuations of population abundance, even inducing chaotic oscillations [11, Sect. 3.2].

Consider the Ricker–Clark map with constant effort harvesting

$$x_{n+1} = (1-\gamma)\left(\alpha x_n + (1-\alpha)x_n\, e^{r(1-x_n)}\right), \quad n = 0, 1, 2, \ldots \tag{10}$$

The destabilization that occurs for increased harvesting can be explained by a *bubbling* effect, which essentially consists of a period-doubling bifurcation followed by a period-halving bifurcation; these bifurcations produce a *bubble* in the bifurcation diagram. See Fig. 7.

In Fig. 8 we visualize the bubbling effect as well as population extinction due to overharvesting by presenting time series predicted by the model for selected harvesting efforts.

In Ref. [11, Theorem 2], the exact parameter ranges are given for which a bubble occurs in Eq. (10). What is necessary is a combination of high production rates ($r > 3$) and intermediate survivorship rates. Similar conclusions are obtained for

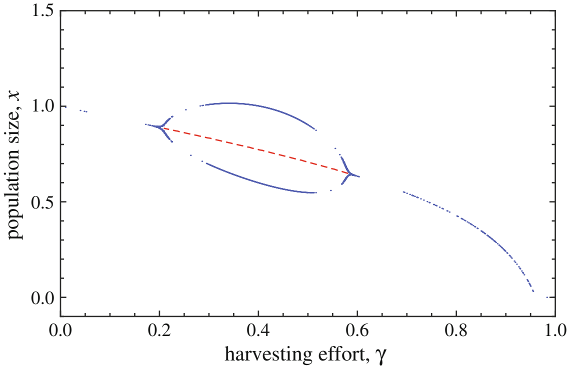

Fig. 7 Bifurcation diagram showing a bubbling effect for Eq. (10) with $\alpha = 0.55$, $r = 4$ and $\gamma \in (0, 1)$. Inside the *bubble*, the equilibrium is unstable (*dashed curve*), and there is an attracting periodic orbit of period two

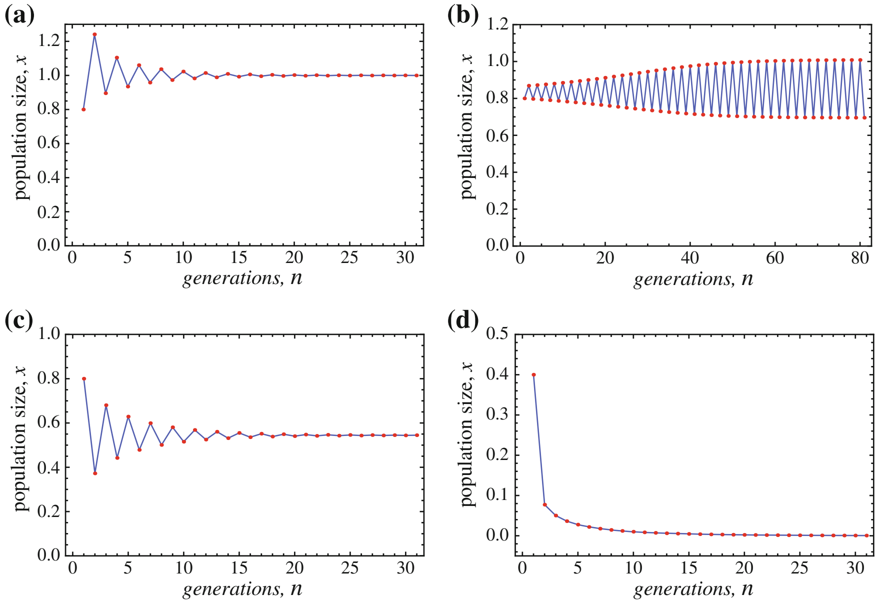

Fig. 8 Time series for Eq. (10) with $\alpha = 0.55$, $r = 4$ and different harvesting rates: **a** asymptotically stable equilibrium for the unharvested population ($\gamma = 0$); **b** sustained oscillations for a capture rate of 30 %; **c** the equilibrium is again stable when the harvesting rate is 70 %; **d** overharvesting ($\gamma = 0.96$) drives the population extinct

a stage-structured model with two age classes (juveniles and adults) if only adult harvesting is allowed [10, 20]. Bubbling can also occur if juveniles and adults are harvested with the same rate, but not if juveniles are the only harvesting target (for more details, see [10]). Note that forms of bubbling have also been observed in population models with constant feedback control (here, constant immigration), but only when varying the production rate rather than the harvesting parameter [18].

If we consider intervention time using Seno's model as we did in Sect. 2, we arrive at a similar conclusion: intermediate harvesting times can be stabilizing; actually, a suitable value of the timing parameter θ can avoid the bubbling effect (see [2] for more details).

4 The Ricker–Schreiber Model: Sudden Collapses and Essential Extinction

A common feature of the models studied in the previous sections is that overharvesting (leading to population extinction) takes place after a transcritical bifurcation, i.e. when the harvesting effort has passed a critical value γ^*. Actually, for values of γ slightly smaller than γ^*, the positive equilibrium is globally asymptotically stable and decays continuously to zero. In some sense, this means that extinction can be prevented if harvesting pressure is increased only gradually (although the decay to zero can be very fast, especially if we census after reproduction, see Fig. 3).

But there are populations for which the transition from a stable positive equilibrium to extinction is discontinuous, producing a so-called sudden collapse. This phenomenon is typical of a strategy of constant quota harvesting [9, 13], but it can also happen for constant effort harvesting if the population model exhibits a strong Allee effect.

The last model we consider in this paper is also based on the Ricker map, but it includes a factor for positive density dependence that induces a strong Allee effect. It is the Ricker–Schreiber model

$$x_{n+1} = \frac{\beta x_n}{1+\beta x_n} x_n \, e^{r(1-x_n)}, \quad n = 0, 1, 2, \ldots, \tag{11}$$

which was already introduced in Eq. (4). Parameter β represents the carrying capacity of the population in the absence of mate limitation multiplied by an individual's efficiency to find a mate [14, Sect. 2.1]. See Fig. 9 for a graphic representation when $r = 3.5$ and $\beta = 0.5$.

There are three generic possibilities for the dynamics of model (10): extinction; bistability between extinction and survival; and essential extinction [13]. The latter means that extinction occurs for a randomly chosen initial condition with probability one. For fixed values of β and r, different harvesting efforts result in the three generic possibilities. For example, we consider the model with constant effort harvesting

$$x_{n+1} = (1-\gamma)\frac{4x_n}{1+4x_n} x_n \, \mathrm{e}^{4(1-x_n)}, \quad n = 0, 1, 2, \ldots \tag{12}$$

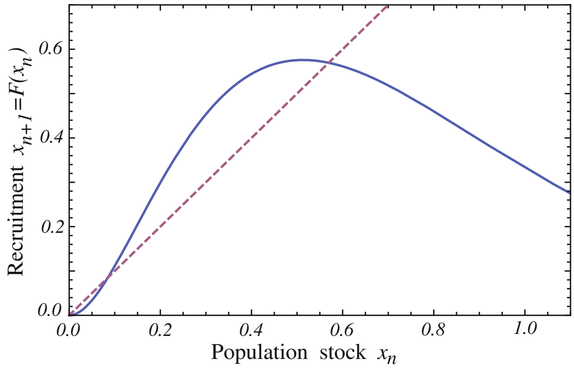

Fig. 9 Graphic representation of the Ricker–Schreiber stock–recruitment curve $F(x) = (0.5x^2/(1 + 0.5x))e^{3.5(1-x)}$. The Allee threshold is the smaller positive equilibrium; population sizes below it cannot survive

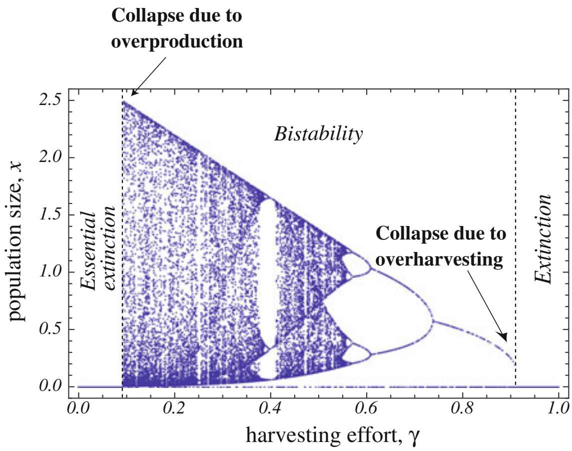

Fig. 10 Bifurcation diagram of the Ricker–Schreiber model (12), using the harvesting effort γ as the bifurcation parameter. The three generic possibilities are observed; in case of bistability, the nontrivial attractor is complex for low harvesting rates and becomes an attracting positive equilibrium after a series of period-halving bifurcations for larger harvesting effort

For $\gamma = 0$, there is essential extinction (c.f. [14, p. 205]). When constant effort harvesting is applied (see the bifurcation diagram in Fig. 10), a boundary collision switches the dynamics from essential extinction to bistability at a value of $\gamma_1 = 0.09384$. A tangent bifurcation leads to extinction at $\gamma_2 = 0.91104$, which corresponds to a sudden collapse due to overharvesting. Between γ_1 and γ_2, the dynamics of the nontrivial attractor ranges from chaos to asymptotic stability of the larger positive equilibrium.

We call the reader's attention to an unusual behaviour of extinction: populations can persist within a band of medium to high harvesting efforts, whereas extinction occurs for lower and very high harvesting efforts. This phenomenon is also typical of a strategy of constant quota harvesting, and has been uncovered by Sinha and Parthasarathy [17]. For constant effort harvesting in models with Allee effects, it was first demonstrated by Schreiber [14].

The influence of harvest timing in the model (10) has been considered in [2]. Here we just state the main conclusions for the model

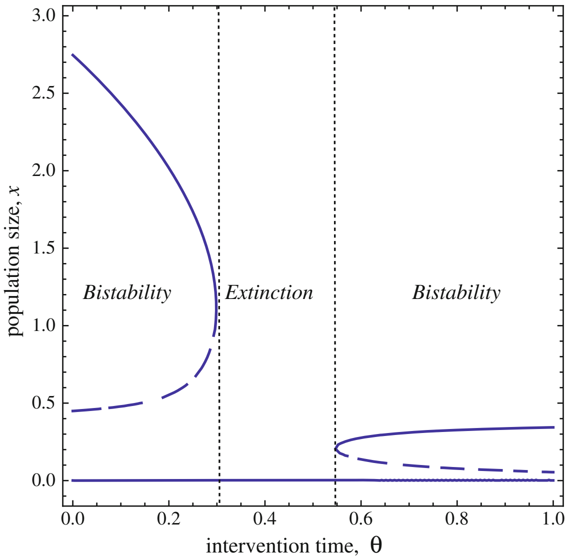

Fig. 11 Bifurcation diagram of model (13) with $F_{4,4}(x) = (4x/(1+4x))x\,e^{4(1-x)}$ and harvesting rate $\gamma = 0.875$. For early and late harvesting times, the population can survive at moderate sizes, but the same harvesting effort drives the population extinct if the harvest takes place at intermediate moments of the season

$$x_{n+1} = \theta(1-\gamma)F_{\beta,r}(x_n) + (1-\theta)F_{\beta,r}((1-\gamma)x_n), \quad n = 0, 1, 2, \ldots, \tag{13}$$

where $\theta \in [0, 1]$ and $F_{\beta,r}(x) = (\beta x/(1+\beta x))x\,e^{r(1-x)}$.

- For moderate harvesting efforts, intermediate values of the harvest timing θ can stabilize the larger positive equilibrium and hence facilitate stabilization—similarly to the models considered in Sects. 2 and 3.
- For large harvesting efforts (close to the regime of overharvesting), intermediate values of the harvest timing θ can render the population more vulnerable to extinction. In this scenario, the population can persist for early- or late-season harvesting, but goes extinct for mid-season harvesting; see Fig. 11. The underlying reason is that intervention time θ does change the overharvesting effort, i.e. the critical value of the harvesting effort at which the system switches from survival to extinction. This is in contrast to the models considered in Sects. 2 and 3.
- For low harvesting efforts (close to the regime of essential extinction), intermediate values of the harvest timing θ can prevent essential extinction, which would occur for early or late harvesting.

Hence, intermediate harvest times can be both beneficial (for small and moderate harvesting efforts) and detrimental (for large harvesting efforts). See [2] for more details.

5 Conclusions

In this contribution, we have reviewed the impact of harvesting effort and harvest timing on population dynamics. While a large part of the literature is mainly concerned with the yield obtained from harvesting (e.g., [3]), we have focused on (i) the abundance of the exploited population and (ii) the complexity of the dynamics induced by harvesting, in particular whether harvesting can be stabilizing or destabilizing. Both aspects are crucial for the yield as well as for the sustainability of the population. In the overview of this contribution, we have exclusively considered single-species discrete-time population models. However, they represent a fair amount of different ecological situations as they take into account overcompensation (scramble competition); adult survival (iteroparity) and critical depensation (strong Allee effect).

Regarding **population abundance,** the most interesting phenomenon is the *hydra effect* [1, 11, 15]. Average population abundance can increase in response to an increase in the per-capita mortality rate. This phenomenon underlines the importance of **census timing**, as the hydra effect in parts of the population may be "hidden" from observation and go unnoticed [5].

Regarding the **complexity** of the dynamics, increased harvesting typically stabilizes population dynamics, but in the presence of adult survivorship it can also be destabilizing. Typical mechanisms are period-halving bifurcations and bubbling [8, 11].

In compensatory models (i.e., without Allee effect), **harvest timing** does not affect the critical harvesting effort leading to overexploitation and population extinction. Harvesting at an intermediate moment of the season can reduce dynamic complexity, preventing chaos and sometimes stabilizing the positive equilibrium. In models with a strong Allee effect, intermediate harvest timing can enhance both persistence as well as extinction; the actual outcome depends on the magnitude of the harvesting effort [2].

In models with a *strong Allee effect,* population extinction due to overharvesting may occur in form of a **sudden collapse** rather than gradually. Intermediate harvesting rates, however, may help the population to survive, preventing essential extinction due to overproduction [14].

To conclude, we emphasize that a good knowledge of the population dynamics is crucially important for designing management programmes of exploited populations. For example, does population growth exhibit exact or undercompensation, overcompensation or depensation? Is the population semelparous or iteroparous? Once the underlying population dynamics is known, it can be equally important to address the aspects of census timing (how many times and at what moments in the seasons the population is measured) and harvest timing. Kokko [6, p. 143] highlighted already in 2001 that

> Timing of harvesting may profoundly influence the impact on the population.

In this overview, we have collected further theoretical mechanisms demonstrating the role of harvest timing and that it should not be neglected in comparison to harvesting effort.

Acknowledgments E. Liz was partially supported by the Spanish Government and FEDER, grant MTM2010–14837. He acknowledges the support and nice hospitality received from the Organizing Committee of the 19th International Conference on Difference Equations & Applications, particularly to Dr. Ziyad AlSharawi. F. M. Hilker acknowledges a Santander research travel grant that facilitated his visit to Universidad de Vigo to collaborate with the first author.

References

1. Abrams, P.A.: When does greater mortality increase population size? The long story and diverse mechanisms underlying the hydra effect. Ecol. Lett. **12**, 462–474 (2009)
2. Cid, B., Hilker, F.M., Liz, E.: Harvest timing and its population dynamic consequences in a discrete single-species model. Math. Biosci. **248**, 78–87 (2014)
3. Clark, C.W.: Mathematical Bioeconomics: The Optimal Management of Renewable Resources. Wiley, Hoboken (1990)
4. Courchamp, F., Berec, L., Gascoigne, J.: Allee Effects in Ecology and Conservation. Oxford University Press, New York (2008)
5. Hilker, F.M., Liz, E.: Harvesting, census timing and “hidden" hydra effects. Ecol. Complex. **14**, 95–107 (2013)
6. Kokko, H.: Optimal and suboptimal use of compensatory responses to harvesting: timing of hunting as an example. Wildl. Biol. **7**, 141–150 (2001)
7. Kot, M.: Elements of Mathematical Ecology. Cambridge University Press, Cambridge (2001)
8. Liz, E.: How to control chaotic behaviour and population size with proportional feedback. Phys. Lett. A **374**, 725–728 (2010)
9. Liz, E.: Complex dynamics of survival and extinction in simple population models with harvesting. Theor. Ecol. **3**, 209–221 (2010)
10. Liz, E., Pilarczyk, P.: Global dynamics in a stage-structured discrete-time population model with harvesting. J. Theor. Biol. **297**, 148–165 (2012)
11. Liz, E., Ruiz-Herrera, A.: The hydra effect, bubbles, and chaos in a simple discrete population model with constant effort harvesting. J. Math. Biol. **65**, 997–1016 (2012)
12. Ricker, W.E.: Stock and recruitment. J. Fish. Res. Bd. Can. **11**, 559–623 (1954)
13. Schreiber, S.J.: Chaos and population disappearances in simple ecological models. J. Math. Biol. **42**, 239–260 (2001)
14. Schreiber, S.J.: Allee effect, extinctions, and chaotic transients in simple population models. Theor. Popul. Biol. **64**, 201–209 (2003)
15. Seno, H.: A paradox in discrete single species population dynamics with harvesting/thinning. Math. Biosci. **214**, 63–69 (2008)
16. Shelton, A.O., Mangel, M.: Fluctuations of fish populations and the magnifying effects of fishing. Proc. Natl. Acad. Sci. USA **108**, 7075–7080 (2011)
17. Sinha, S., Parthasarathy, S.: Unusual dynamics of extinction in a simple ecological model. Proc. Natl. Acad. Sci. USA **93**, 1504–1508 (1996)
18. Stone, L.: Period-doubling reversals and chaos in simple ecological models. Nature **365**, 617–620 (1993)
19. Thieme, H.R.: Mathematics in Population Biology. Princeton University Press, Princeton (2003)
20. Zipkin, E.F., Kraft, C.E., Cooch, E.G., Sullivan, P.J.: When can efforts to control nuisance and invasive species backfire? Ecol. Appl. **19**, 1585–1595 (2009)

Chaos and Wild Chaos in Lorenz-Type Systems

Hinke M Osinga, Bernd Krauskopf and Stefanie Hittmeyer

Abstract This contribution provides a geometric perspective on the type of chaotic dynamics that one finds in the original Lorenz system and in a higher-dimensional Lorenz-type system. The latter provides an example of a system that features robustness of homoclinic tangencies; one also speaks of 'wild chaos' in contrast to the 'classical chaos' where homoclinic tangencies accumulate on one another, but do not occur robustly in open intervals in parameter space. Specifically, we discuss the manifestation of chaotic dynamics in the three-dimensional phase space of the Lorenz system, and illustrate the geometry behind the process that results in its description by a one-dimensional noninvertible map. For the higher-dimensional Lorenz-type system, the corresponding reduction process leads to a two-dimensional noninvertible map introduced in 2006 by Bamón, Kiwi, and Rivera-Letelier [arXiv 0508045] as a system displaying wild chaos. We present the geometric ingredients—in the form of different types of tangency bifurcations—that one encounters on the route to wild chaos.

1 Introduction

The Lorenz system was introduced and studied by meteorologist Edward Lorenz in the 1960s as an extremely simplified model for atmospheric convection dynamics [35]. Famously, Lorenz discovered sensitive dependence on the initial condition, and the Lorenz system has arguably become the best-known example of a chaotic system. It is given as the vector field

H.M. Osinga (✉) · B. Krauskopf · S. Hittmeyer
Department of Mathematics, The University of Auckland,
Private Bag 92019, Auckland 1142, New Zealand
e-mail: h.m.osinga@auckland.ac.nz

B. Krauskopf
e-mail: b.krauskopf@auckland.ac.nz

S. Hittmeyer
e-mail: stefanie.hittmeyer@auckland.ac.nz

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_4

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = \rho x - y - xz, \\ \dot{z} = -\beta z + xy. \end{cases} \tag{1}$$

The system is invariant under the symmetry of a rotation about the z-axis by π. The now classical choice of the parameters for which Lorenz found a chaotic attractor is

$$\sigma = 10, \quad \rho = 28, \quad \beta = 2\frac{2}{3}. \tag{2}$$

For these parameters, (1) has three equilibria: the origin $\mathbf{0}$ and a symmetrically related pair of secondary equilibria $p^{\pm}$, which are all saddles. The chaotic attractor is often called the Lorenz or butterfly attractor. It has two 'wings,' which are centred at p^- and p^+ (which are not part of the attractor). Importantly, the Lorenz attractor contains $\mathbf{0}$ and its one-dimensional unstable manifold $W^u(\mathbf{0})$, that is, the two trajectories that converge to $\mathbf{0}$ in backward time. The Lorenz attractor actually consists of infinitely many layers or sheets that are connected along $W^u(\mathbf{0})$, which forms the 'outer boundary' of the attractor. This is already sketched and studied in the original paper by Lorenz [35]; an illustration of the different layers of the Lorenz attractor can be found in the paper by Perelló [40] and it is reproduced in [14].

Figure 1 shows a computed version of the Lorenz attractor, which was rendered as a surface from computed orbit segments of several suitably chosen families; see [14] for details of this computation. Also shown in Fig. 1 is the one-dimensional unstable manifold $W^u(\mathbf{0})$, with its left and right branches rendered in different shades; observe how $W^u(\mathbf{0})$ forms the outer boundary of the Lorenz attractor. Our visualisation in Fig. 1 is quite different from most images of the Lorenz attractor that are obtained with numerical simulation. Starting from some initial condition, and letting transients die down, the Lorenz attractor is typically visualised by plotting (a long part of) the remaining trajectory. In this way, the part of the Lorenz attractor closest to the origin is generally missed, as it is not 'visited' very often by trajectories; hence, most published images show a considerably smaller part of the Lorenz attractor.

The Lorenz system (1) has been studied since the 1970s via the concept of the *geometric Lorenz attractor*, which is an abstract geometric model introduced by Guckenheimer [24], Guckenheimer and Williams [26], and Afrajmovich, Bykov and Shilnikov [1, 2]; see also [8, 44]. The key is that the geometric Lorenz attractor displays all the features observed in the Lorenz system, and that it can be reduced rigorously to a one-dimensional noninvertible map. This reduction is done in two steps. First of all, one considers the Poincaré return map to the horizontal section through the points $p^{\pm}$ (given by $z = \rho - 1$). Locally this map is a diffeomorphism that has a stable foliation (near the classic parameter values), that is, an invariant foliation that is uniformly contracted by the Poincaré return map. The map on the quotient space of this foliation is a one-dimensional noninvertible map, called the *Lorenz map*, and it describes the dynamics on the geometric Lorenz attractor exactly. It can be shown with standard methods that the Lorenz map has chaotic dynamics; see, for example, [25]. In 1999 Tucker [45] famously provided a computer-assisted proof that, for the classical parameter values (2), the Lorenz system (1) satisfies the

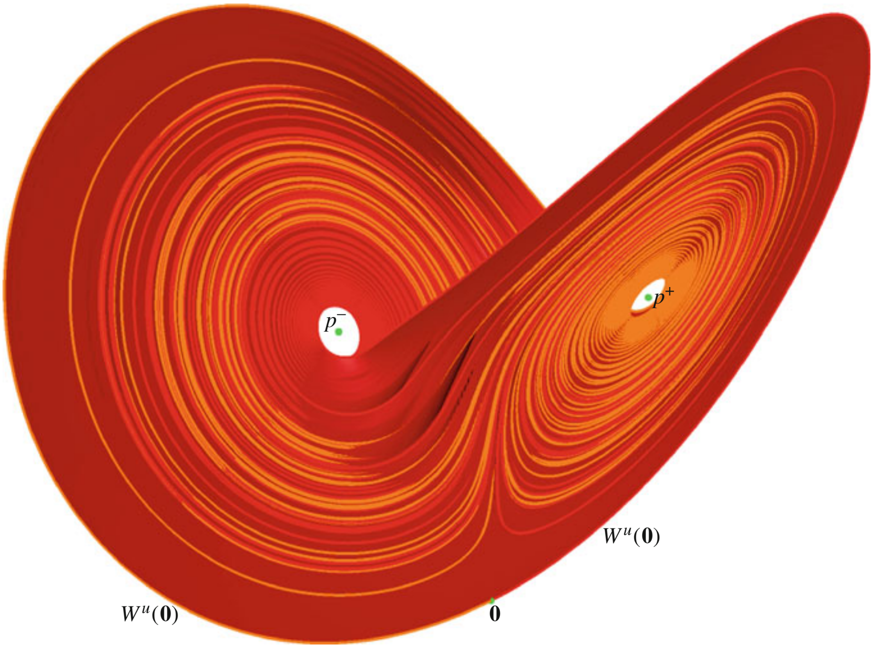

Fig. 1 The Lorenz attractor as computed and rendered as a surface, with the equilibria $\mathbf{0}$ and $p^{\pm}$ and the manifold $W^u(\mathbf{0})$

technical conditions of this geometric construction, thereby showing that the Lorenz attractor is indeed a chaotic attractor.

The question how chaos arises in the Lorenz system has also been considered, where ρ is chosen traditionally as the parameter that is varied [15, 43]. For small $\rho > 1$ all typical initial conditions simply end up at either p^- or p^+, which are the only attractors of (1). As ρ is increased, a first homoclinic bifurcation at $\rho \approx 13.9265$ is encountered; here both branches of the one-dimensional unstable manifold $W^u(\mathbf{0})$ of the origin return to $\mathbf{0}$ to form a pair of homoclinic connections. This global bifurcation creates not only a pair of (symmetrically related) saddle periodic orbits, but also a hyperbolic set of saddle type. The result is what has been called preturbulence [29], which is characterised by the existence of arbitrarily long chaotic transients before the system settles down to either p^- or p^+ (still the only attractors). At $\rho \approx 24.0579$ one encounters a pair of heteroclinic cycles between the origin and the pair of saddle periodic orbits, and this results in the creation of a chaotic attractor. The chaotic attractor, which is the closure of $W^u(\mathbf{0})$, coexists with the two stable equilibria until they become saddles in a Hopf bifurcation at $\rho = 470/19 \approx 24.7368$. After the Hopf bifurcation and up to $\rho = 28$, the chaotic attractor is the only attractor.

A crucial role in the organisation of the dynamics of the Lorenz system (1) is played by the stable manifold $W^s(\mathbf{0})$ of the origin $\mathbf{0}$, which we refer to as the Lorenz manifold. The origin $\mathbf{0}$ is a saddle equilibrium (for $\rho > 1$) with two stable directions and one unstable direction, and $W^s(\mathbf{0})$ is a smooth surface that consists of all points in $\mathbb{R}^3$ that end up at $\mathbf{0}$. Before the first homoclinic bifurcation, $W^s(\mathbf{0})$ forms the boundary between the two attractors p^- and p^+. In the preturbulent regime after the first homoclinic bifurcation $W^s(\mathbf{0})$ is still part of the basin boundary of $p^{\pm}$, but it is much more complicated topologically as it is involved in organising arbitrarily long transients.

More importantly for the purpose of this paper, the Lorenz manifold $W^s(\mathbf{0})$ organises the dynamics in the chaotic regime [14, 15]. Owing to the sensitive dependence on the initial condition, $W^s(\mathbf{0})$ is dense in phase space. Moreover, the interaction of the Lorenz manifold $W^s(\mathbf{0})$ with the unstable manifold $W^u(\mathbf{0})$ gives rise to infinitely many further homoclinic bifurcations when ρ is varied. Closely related is the fact that there are infinitely many homoclinic tangencies between the two-dimensional stable and unstable manifolds of the saddle periodic orbits that lie dense in the chaotic Lorenz attractor. More generally, such tangencies of a three-dimensional vector field correspond directly (by taking a Poincaré return map) to homoclinic tangencies of the one-dimensional stable and unstable manifolds of fixed or periodic points of a planar diffeomorphism such as the Hénon map [27], which is another well-known chaotic system. Near a homoclinic tangency one can construct Smale horseshoe dynamics, that is, a chaotic hyperbolic set of saddle type. Moreover, any homoclinic tangency of a one-parameter family of three-dimensional vector fields, or planar diffeomorphisms, is accumulated in parameter space by other homoclinic tangencies [39], leading to an infinite sequence of homoclinic tangency points accumulating on other homoclinic tangency points. This is one of the characterising properties of 'classical chaos' that arises in vector fields of dimension three and in diffeomorphisms of dimension two, for which the Lorenz system and the Hénon map are standard examples; see, for example, textbooks such as [4, 25, 42].

At a homoclinic tangency of a hyperbolic set (such as a periodic orbit) there is a nontransversal intersection of its stable and unstable manifolds. In particular, the point of homoclinic tangency is nonwandering and its tangent bundle cannot be decomposed into stable and unstable subspaces. As a result, the system is not uniformly hyperbolic, or simply, it is *nonhyperbolic* at a homoclinic tangency. In other words, in 'classical chaos' one finds infinitely many accumulating points of nonhyperbolicity. A property is said to be *robust* (in the C^1-topology) if there is an open neighbourhood in the space of vector fields or diffeomorphisms such that all these systems have said property. As it turns out, it has been argued that nonhyperbolicity and homoclinic tangencies do not occur *robustly* in three-dimensional vector fields or two-dimensional diffeomorphisms [36].

On the other hand, robust homoclinic tangencies and, hence, robust nonhyperbolicity can be found in vector fields of dimension at least four and in diffeomorphisms of dimension at least three [11]. Any system with this property is said to display *wild chaos* [37]. There are several constructions of diffeomorphisms that feature robust nonhyperbolicity [3, 6, 7, 21, 22]. Moreover, in [49] it is shown that a four-dimensional vector field model of calcium dynamics in a neuronal cell has a heterodimensional cycle between two saddle periodic orbits, which is directly associated with robust hyperbolicity [11, 29]. It is also possible to construct an n-dimensional vector field with robust homoclinic tangencies of a singular attractor and, hence, with wild chaos. Turaev and Shilnikov presented such an example for $n \geq 4$ in [46, 47]. We consider here the example for $n \geq 5$ due to Bamón, Kiwi, and Rivera-Letelier [9], which is constructed as a Lorenz-type system. It suffices to consider their construction for $n = 5$; the associated attractor is called *Lorenz-like* because it is effectively a higher-dimensional version of the geometric Lorenz attractor. The dynamics of the five-dimensional Lorenz-type vector field is described by a four-dimensional diffeomorphism given as the Poincaré return map to a suitable codimension-one section. On this section there is a two-dimensional stable foliation, and the resulting quotient map is now a noninvertible map of the plane. This map is given in [9] in explicit form; in fact, Bamón, Kiwi, and Rivera-Letelier construct their example by starting from the noninvertible map, lifting it to the four-dimensional Poincaré return map and then suspending this diffeomorphism to obtain an abstract five-dimensional Lorenz-type vector field. In a small neighbourhood of a specific point in parameter space, they then show that the planar noninvertible map is robustly nonhyperbolic.

The goal of this paper is to determine and illustrate the geometry behind chaos in the Lorenz system and wild chaos in the five-dimensional Lorenz-type system. This study is made possible by advanced numerical methods—based on solving families of boundary value problems—for the computation of two-dimensional global manifolds of vector fields [15, 30, 31, 33] and tangency bifurcations involving stable and unstable sets of noninvertible planar maps [10, 28]; their implementation is done in the packages AUTO [13] and Cl_MatContM [18, 23], respectively. Section 2 is concerned with the Lorenz system. Our starting point in Sect. 2.1 is the discussion of how the three-dimensional phase space is organised globally by the two-dimensional Lorenz manifold $W^s(\mathbf{0})$ of the origin in the presence of the classical Lorenz attractor. We then discuss in Sect. 2.2 the geometry behind the description of the dynamics on the Lorenz attractor by the one-dimensional Lorenz map. The two-dimensional Lorenz-like map is introduced in Sect. 3 and its basic properties are discussed. The transition from simple to wild chaos is the subject of Sect. 3.1, where we show how different types of tangency bifurcations are involved in creating increasingly complicated dynamics. In Sect. 3.2 we present a two-parameter bifurcation diagram with curves of the different tangency bifurcations, which allows us to identify a large region where we conjecture wild chaos to be found. Finally, Sect. 4 summarises the results and briefly discusses avenues for future research.

2 Chaos in the Lorenz System

In this section we consider the chaotic dynamics of the Lorenz system (1) for the classical parameter values given in (2). We first consider the organisation of the full phase space and then illustrate the geometry behind the reduction to the one-dimensional Lorenz map.

2.1 Global Organisation of the Phase Space

The Lorenz attractor is the only attractor of (1) for $\rho = 28$. Its basin is the entire phase space $\mathbb{R}^3$ with the exception of the symmetric pair of secondary equilibria $p^{\pm}$ and their one-dimensional stable manifolds $W^s(p^{\pm})$. Recall that the origin $\mathbf{0}$ and its one-dimensional unstable manifold $W^u(\mathbf{0})$ are part of the chaotic attractor. This also means that the two-dimensional stable manifold $W^s(\mathbf{0})$ lies in the basin of the Lorenz attractor. Moreover, locally near $\mathbf{0}$ the invariant surface $W^s(\mathbf{0})$ determines the dynamics in the following sense: initial conditions on one side of $W^s(\mathbf{0})$ flow away from the origin into the left wing of the attractor (towards negative values of x) and those on the other side flow away from the origin into the right wing of the attractor (towards positive values of x). The sensitive dependence on initial conditions of the dynamics on the Lorenz attractor has global consequences throughout the phase space. Any open sphere in phase space, no matter how small, must contain two points that eventually move over the Lorenz attractor differently: at some point in time one trajectory is, say, on the left wing, while the other is on the right wing. This means that, locally near the attractor, the two trajectories are on either side of $W^s(\mathbf{0})$. This implies that $W^s(\mathbf{0})$ must divide the open sphere into two open halves, each containing one of the two initial conditions. In turn this proves that $W^s(\mathbf{0})$ lies dense in the basin of the Lorenz attractor and, hence, also in $\mathbb{R}^3$.

According to the stable and unstable manifold theorem [38], locally near $\mathbf{0}$ the surface $W^s(\mathbf{0})$ is a small topological disk that is tangent to the two-dimensional stable linear eigenspace $E^s(\mathbf{0})$ spanned by the eigenvectors of the two negative real eigenvalues. This disk can be imagined to grow while its boundary maintains a fixed geodesic distance (arclength of the shortest path on $W^s(\mathbf{0})$) to the origin $\mathbf{0}$. At any stage of this growth process one is dealing with a smooth embedding of the standard unit disk into $\mathbb{R}^3$ yet, as it grows, this topological disk fills out $\mathbb{R}^3$ densely. Hence, loosely speaking, one can imagine the surface $W^s(\mathbf{0})$ as a growing, space-filling pancake.

We developed a numerical method for the computation of two-dimensional stable and unstable manifolds, called the geodesic level set (GLS) method [30]. This method is based on the idea of growing such a manifold by adding geodesic bands to it at each step. With the GLS method we are able to compute a first part of $W^s(\mathbf{0})$ as a surface up to a considerable geodesic distance. On the other hand, in order to examine the denseness of $W^s(\mathbf{0})$ in $\mathbb{R}^3$ a different approach is needed. Namely, we consider the intersection set $\widehat{W}^s(\mathbf{0}) := W^s(\mathbf{0}) \cap S_R$ with a suitably chosen sphere S_R, which is then computed directly by defining a boundary value problem such that its solutions are orbit segments with begin point on S_R and end point in $E^s(\mathbf{0})$ near $\mathbf{0}$; see [5, 15] for details. More specifically, we choose the centre of S_R as the point $(0, 0, \rho - 1)$ on the z-axis, which lies exactly in the middle of the line that connects the two equilibria $p^{\pm}$. The radius R of S_R is chosen such that the Lorenz attractor is well inside S_R, and the second intersection points in $\widehat{W}^s(p^{\pm}) := W^s(p^{\pm}) \cap S_R$ of the small-amplitude branches of $W^s(p^{\pm})$ lie on the 'equator' of S_R—for $\rho = 28$ as considered here, this gives $R = 70.7099$; see [15] for details.

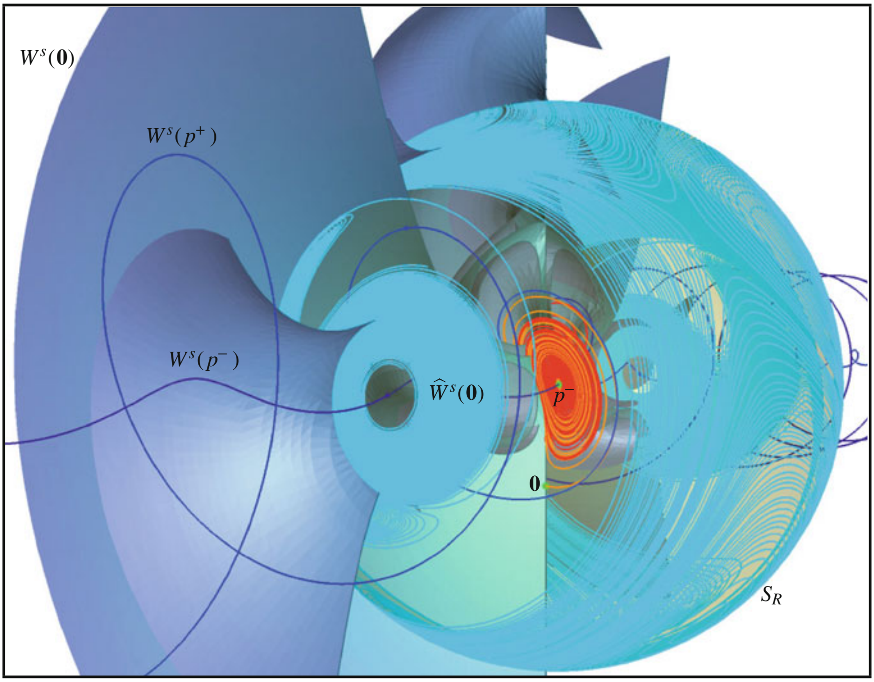

Fig. 2 The Lorenz manifold $W^s(\mathbf{0})$ for $\rho = 28$ intersecting the sphere S_R with $R = 70.7099$ in the set $\widehat{W}^s(\mathbf{0})$; also shown are the equilibria $\mathbf{0}$ and p^- and the one-dimensional manifolds $W^u(\mathbf{0})$ and $W^s(p^{\pm})$

Figure 2 illustrates the geometry of how $W^s(\mathbf{0})$ intersects the sphere S_R. The view is from a point with negative x- and y-coordinates, and only one half of the computed part of the surface $W^s(\mathbf{0})$ is shown, namely, the part with $y \geq 0$. The sphere S_R is rendered transparent. Inside S_R, we can clearly see the equilibria $\mathbf{0}$ and p^-, with p^+ obscured by $W^s(\mathbf{0})$. The one-dimensional unstable manifold $W^u(\mathbf{0})$, with its left and right branches rendered in different shades, gives an idea of the location of the Lorenz attractor. Also shown in Fig. 2 are the two one-dimensional stable manifolds $W^s(p^{\pm})$, each drawn in different shades. Note that the small-amplitude branch of $W^s(p^-)$ indeed intersects S_R along its equator, while the large-amplitude branch of $W^s(p^+)$ intersects S_R at a point higher up and closer to the z-axis. Recall that $p^{\pm} \cup W^s(p^{\pm})$ forms the complement of the basin of the Lorenz attractor. The surface $W^s(\mathbf{0})$ can be seen to wrap around the curves $W^s(p^{\pm})$, which it cannot intersect. The part of $W^s(\mathbf{0})$ that is shown, which was computed up to geodesic distance 162.5, generates the beginnings of what appear to be only three intersection curves in $\widehat{W}^s(\mathbf{0})$. It is clear that an impractically large piece of $W^s(\mathbf{0})$ would need to be computed to generate the many curves in $\widehat{W}^s(\mathbf{0})$ that are shown in Fig. 2; this is why $\widehat{W}^s(\mathbf{0})$ is computed directly.

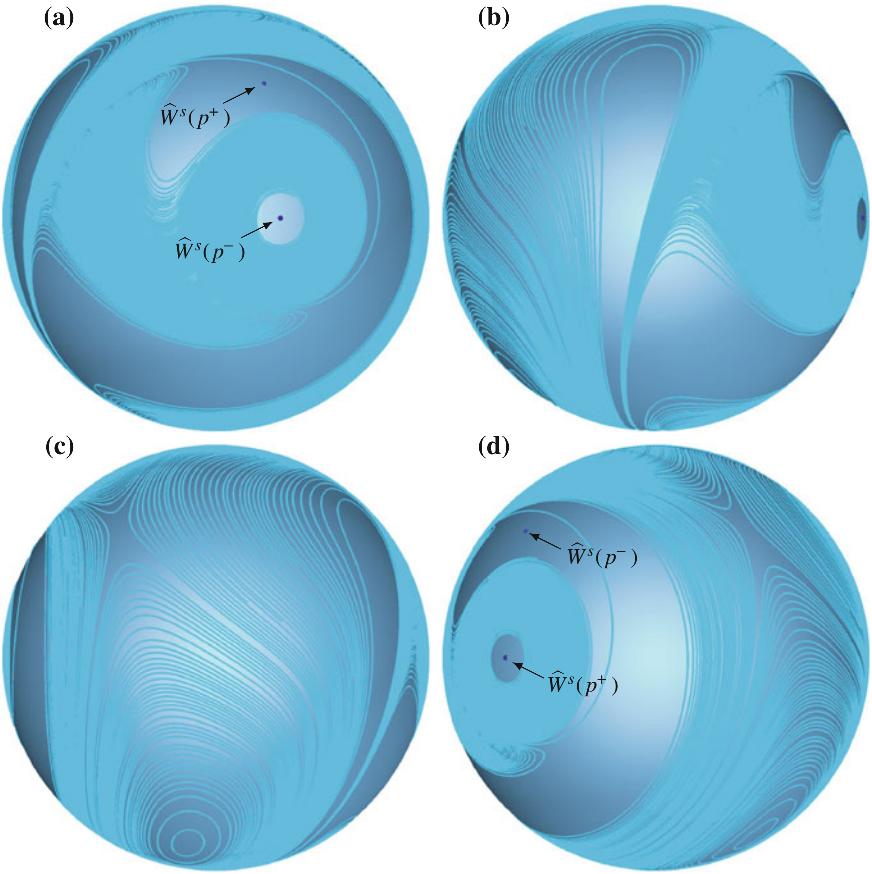

Fig. 3 The set $\widehat{W}^s(\mathbf{0})$ for $\rho = 28$ on the sphere S_R with $R = 70.7099$; also shown are $\widehat{W}^s(p^{\pm})$

Figure 3 shows four different views of the computed intersection curves in $\widehat{W}^s(\mathbf{0})$ on the sphere S_R; also shown on S_R are the points in $\widehat{W}^s(p^{\pm})$. In all views, the vertical axis is the z-axis of (1). In Fig. 3a, the horizontal axis is the direction defined by $(\cos\theta, -\sin\theta)$, with $\theta = 3\pi/10$ (in other words, the (x, y)-plane was rotated clockwise by $3\pi/10$ about the z-axis). The view points in panels (b)–(d) are consecutively rotated by a further $\pi/4$ radians about the z-axis; note that a further rotation over $\pi/4$ would show the symmetrical version of Fig. 3a with $\widehat{W}^s(p^-)$ and $\widehat{W}^s(p^+)$ interchanged. Figure 3 is designed to illustrate how $W^s(\mathbf{0})$ fills the phase space $\mathbb{R}^3$ by showing the intersection set $\widehat{W}^s(\mathbf{0})$ on the sphere S_R. Notice the intricate structure of how the curves in $\widehat{W}^s(\mathbf{0})$ fill up S_R; see also [15]. As one might expect, the computed curves in $\widehat{W}^s(\mathbf{0})$ are not distributed evenly on S_R, and there are several larger regions on S_R without computed curves in $\widehat{W}^s(\mathbf{0})$. This is due to the fact that a finite computation is performed to show an infinite process. More specifically, the curves in $\widehat{W}^s(\mathbf{0})$ that are shown in Fig. 3 have the property that the overall integration time of the associated computed orbit segments is no larger than 7.0; see [15] for details of the computational setup. This bound already leads to a considerable computation generating 350 MB of AUTO data and 377 individual curves in $\widehat{W}^s(\mathbf{0})$. As we have checked, these regions fill up with additional curves in $\widehat{W}^s(\mathbf{0})$ if one allows for a larger bound on the integration time of orbit segments; however, the number of curves thus obtained and, hence, the duration and data produced grow exponentially with the bound on the integration time. Figure 3 provides a good illustration of the space-filling nature of the surface $W^s(\mathbf{0})$ in phase space that, in turn, constitutes a global geometric interpretation of the sensitive dependence of the Lorenz system (1) on the initial conditions.

2.2 From Lorenz Attractor to Lorenz Map

The first step in the reduction process resulting in the description of the dynamics on the Lorenz attractor by the Lorenz map is to consider the Poincaré return map to the horizontal plane Σ_ρ through the secondary equilibria $p^\pm$, which is given by

$$\Sigma_\rho := \{(x, y, z) \in \mathbb{R}^3 \mid z = \rho - 1\}. \tag{3}$$

Geometrically, this means that one needs to consider the intersection sets with Σ_ρ of the relevant invariant objects of the vector field (1). Figure 4 illustrates the situation. The Lorenz attractor, represented by the unstable manifold $W^u(\mathbf{0})$ accumulating on it, can be found in the middle of the image. It is intersected by Σ_ρ, which is rendered transparent, at the height of the equilibria $p^\pm$. The stable manifold $W^s(\mathbf{0})$ is shown as computed up to geodesic distance 162.5; the parts of $W^s(\mathbf{0})$ below and above the plane Σ_ρ are rendered solid and transparent, respectively. The outer boundary of the computed part of $W^s(\mathbf{0})$ (the geodesic level set of geodesic distance 162.5) is highlighted to help illustrate the complicated geometry of this surface, which is topologically a disk. The surface $W^s(\mathbf{0})$ can be seen in Fig. 4 to intersect Σ_ρ in several curves of the set $\overline{W}^s(\mathbf{0}) := W^s(\mathbf{0}) \cap \Sigma_\rho$. One of them is the primary intersection curve $\overline{W}^s_0(\mathbf{0})$, which is invariant under the symmetry of a rotation by π about the z-axis and contains the point $(0, 0, \rho - 1)$. Also shown in Fig. 4 are the one-dimensional manifolds $W^s(p^\pm)$, which intersect Σ_ρ in discrete points.

It is important to realise, as can easily be checked from (1), that the flow is tangent to Σ_ρ along the *tangency locus*

$$C = \{(x, y, \rho - 1) \in \mathbb{R}^3 \mid x\, y = \beta(\rho - 1)\}. \tag{4}$$

The set C consists of two hyperbolas, which contain the equilibria $p^\pm \in \Sigma_\rho$, respecively. In between the two hyperbolas the vector field points downward (towards negative z), which is indicated by the symbol $\otimes$ in Fig. 4. In the regions to the other side of C the vector field points upward (towards positive z), which is indicated by

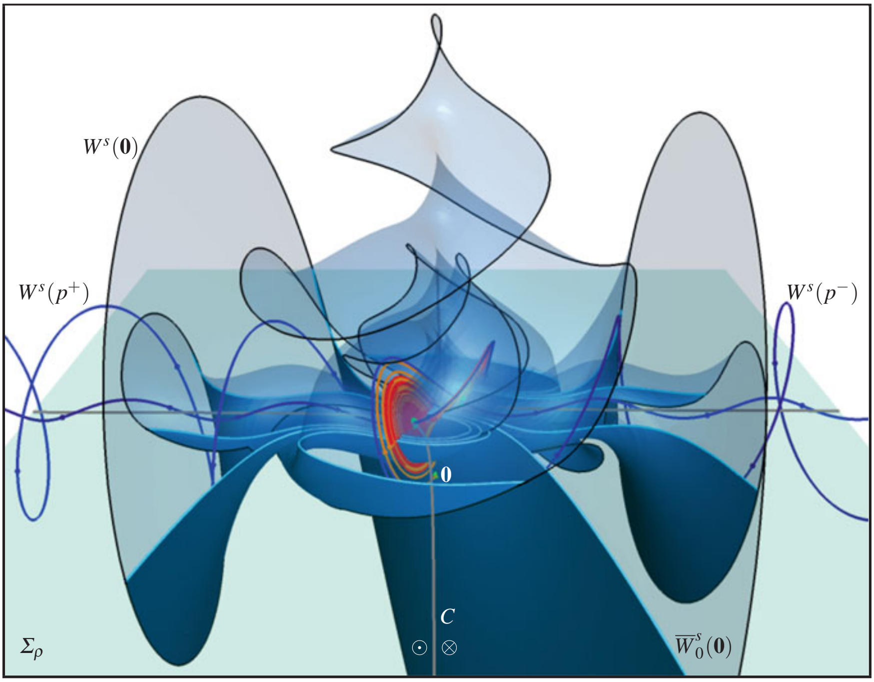

Fig. 4 The manifold $W^s(\mathbf{0})$ for $\rho = 28$ computed up to geodesic distance 162.5 and its intersection with the plane Σ_ρ; the section Σ_ρ and the part of $W^s(\mathbf{0})$ above it are rendered transparent. Also shown are the equilibria $\mathbf{0}$ and $p^\pm$, the one-dimensional manifolds $W^u(\mathbf{0})$ and $W^s(p^\pm)$, and the tangency locus C on Σ_ρ

the symbol $\odot$. As a result, the Poincaré return map, defined as the first return to the section, is not a diffeomorphism on the entire plane Σ_ρ. This is why one defines the local Poincaré return map only on the central region of Σ_ρ where the direction of the flow is downward [24], that is, in between the two hyperbolas of C; technically, this means that one considers the second return to Σ_ρ.

However, this local Poincaré map on the central region is still not a diffeomorphism. Namely, points along the primary intersection curve $\overline{W}_0^s(\mathbf{0})$ converge to $\mathbf{0} \notin \Sigma_\rho$ under the flow and, hence, do not return to the section Σ_ρ. This means that the Poincaré map is not defined on $\overline{W}_0^s(\mathbf{0})$. Trajectories through points to the left of $\overline{W}_0^s(\mathbf{0})$ spiral around p^- while those through points to the right of $\overline{W}_0^s(\mathbf{0})$ spiral around p^+ before intersecting the central region of Σ_ρ again. Hence, the local Poincaré map has a discontinuity across the curve $\overline{W}_0^s(\mathbf{0})$, and it maps each of the two complimentary regions either side of $\overline{W}_0^s(\mathbf{0})$ over the entire central region between the two hyperbolas in C.

Figure 5 shows the respective invariant objects in the plane Σ_ρ, which can be identified with the (x, y)-plane (with fixed $z = \rho - 1$). By construction, the equilibria $p^\pm$ lie in Σ_ρ and on the tangency locus C that bounds the central region indicated

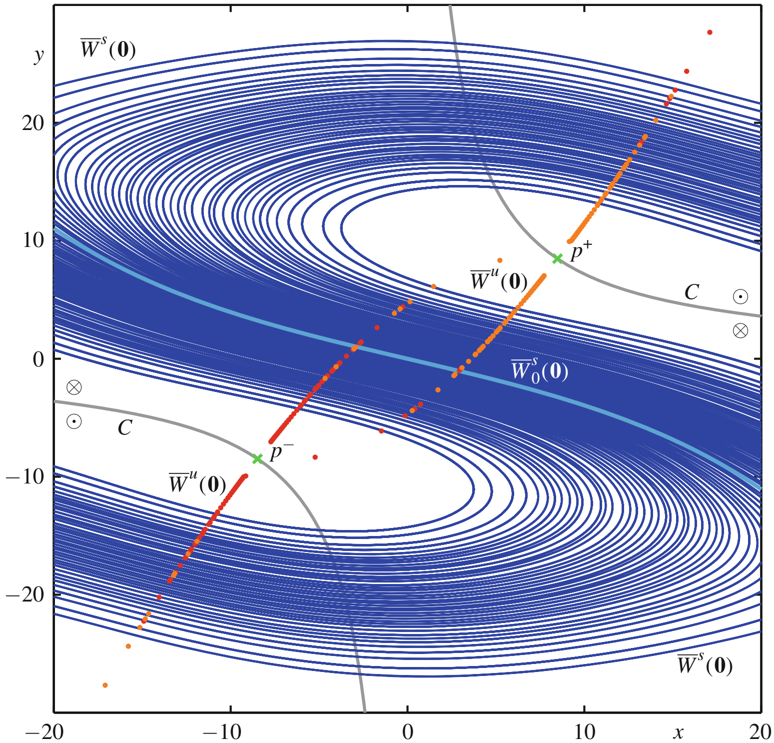

Fig. 5 The invariant objects of (1) for $\rho = 28$ in the plane Σ_ρ; compare with Fig. 4. Shown are the equilibria $p^\pm$, the tangency locus C, the intersection set of the Lorenz attractor as represented by $\overline{W}^u(\mathbf{0})$, and curves in $\overline{W}^s(\mathbf{0})$; the primary intersection curve $\overline{W}^s_0(\mathbf{0})$ is highlighted, and it divides the central region labelled $\otimes$ where the direction of the flow is downward

by the symbol $\otimes$. The Lorenz attractor is represented by the intersection points $\overline{W}^u(\mathbf{0}) := W^u(\mathbf{0}) \cap \Sigma_\rho$ of the unstable manifold $W^u(\mathbf{0})$. These intersection points appear to intersect Σ_ρ in four disjoint curves, two of which lie in the central region; note that the Lorenz attractor does not contain the points $p^\pm$ and compare with Fig. 1. Also shown in Fig. 5 are many curves of the intersection set $\overline{W}^s(\mathbf{0})$, and the primary curve $\overline{W}^s_0(\mathbf{0})$ is highligthed. Curves in the set $\overline{W}^s(\mathbf{0})$ were computed directly by imposing the boundary condition that the corresponding orbit segments have their begin point in Σ_ρ; by contrast, in Fig. 4 the shown curves in $\overline{W}^s(\mathbf{0})$ were obtained from the computed part of the two-dimensional manifold $W^s(\mathbf{0})$.

The reduction of the Poincaré map to the Lorenz map for $\rho = 28$ relies on the fact that the geometric Lorenz system—the abstract version of the Lorenz system—admits an invariant stable foliation in some neighbourhood of the chaotic attractor [1, 26, 41, 48]. This means that leaves of this foliation are mapped to leaves, and the dynamics on the leaves is a contraction. When restricted to said neighbourhood of the attractor, the curves in $\overline{W}^s(\mathbf{0})$ generate the stable foliation by means of taking their closure. Hence, Fig. 5 provides an illustration of the stable foliation by showing a large number of curves in $\overline{W^s}(\mathbf{0})$. The leaves of this foliation intersect the segment of the diagonal between p^- and p^+ in unique points. The one-dimensional Lorenz map is defined on this diagonal segment—or, rather, on the corresponding interval of the variable x—and it describes how leaves are mapped to leaves under the Poincaré map on the central region of Σ_ρ. The Lorenz map is topologically conjugate to the map

$$x \mapsto \begin{cases} 1 - \eta \ |x|^\alpha, & x \in [-1, 0), \\ -1 + \eta \ |x|^\alpha, & x \in (0, 1], \end{cases} \tag{5}$$

with $0 < \alpha < 1$, $\eta \in (1, 2)$ and $\alpha\,\eta > 1$; see [25]. Here, α is the ratio between the magnitudes of the weak stable and unstable eigenvalues of the equilibrium $\mathbf{0}$ of the Lorenz system (1). The Lorenz map is not invertible because it maps the subinterval $[-1, 0)$ to a much larger subinterval in $[-1, 1]$; due to symmetry, the same is true for the subinterval $(0, 1]$. Moreover, the Lorenz map has a discontinuity at 0, which is also referred to as the *critical point*; note that 0 corresponds to the point $(0, 0, \rho - 1) \in \overline{W^s}(\mathbf{0})$ that never returns to Σ_ρ. The critical point 0 has infinitely many preimages under the Lorenz map, because all points on $\overline{W^s}(\mathbf{0})$ eventually map to 0; compare with Fig. 5. One can also take the point of view that the critical point 0 of the Lorenz map represents the origin $\mathbf{0}$ of the Lorenz system; then the (symmetrically related) first intersection points of $\overline{W^u}(\mathbf{0})$ in the central region of Σ_ρ can be thought of as the forward (set-valued) image of the critical point 0. In particular, whenever these two points map to the critical point 0 under some iterate of the Lorenz map then this corresponds to a homoclinic orbit of $\mathbf{0}$ in the full Lorenz system.

The Lorenz map of the form (5) is a rigorous description of the dynamics of the Lorenz system (1) provided that there is an invariant stable foliation. There is every indication that this is indeed the case in this entire ρ-range of $0 < \rho \leq 30.1$ [43]. Indeed, the Lorenz map has been used to study the (emergence of) chaotic dynamics for increasing ρ up to $\rho = 28$ [24, 29, 43]. On the other hand, it is known that for larger values of ρ the Lorenz system has 'cusped horseshoes,' the dynamics of which is definitely not represented faithfully by the one-dimensional Lorenz map [24, 43]. By which mechanism the stable foliation is lost near $\rho \approx 30.1$ is the subject of ongoing research [12].

A closely-related concept is the so-called Lorenz template [19, 20, 24, 35]. Geometrically, the Lorenz template is obtained from the Lorenz attractor in Fig. 1 by the identification of points on the diagonal segment in between p^- and p^+ with points on the Lorenz attractor via the projection along leaves of the stable foliation in Σ_ρ. More specifically, consider the points corresponding to the stable projections of the first intersection points of the two sides of $W^u(\mathbf{0})$ with Σ_ρ in the central region where the direction of the flow is downward. The diagonal segment connecting these two points contains the point $(0, 0, \rho - 1)$. Initial conditions on the diagonal segment on either side of $(0, 0, \rho - 1)$ sweep out two surfaces as the flow takes them around p^- and p^+, respectively, until they return to the central region of Σ_ρ as two curves (that are very close to the intersection of the Lorenz attractor with Σ_ρ). Projection along stable leaves then identifies these two end curves with the initial diagonal segment. This segment can, hence, be thought of as the start and finish line on a branched two-manifold, that is, the topological object obtained by 'glueing' the two surfaces together along the diagonal in the central region of Σ_ρ; this branched two-manifold is the Lorenz template. In particular, the Lorenz template allows one to describe the symbolic dynamics of the knot-types in $\mathbb{R}^3$ of periodic orbits in the Lorenz system [19]. Notice that the dynamics from start to finish on the Lorenz template is exactly given by the Lorenz map.

3 Wild Chaos in a Lorenz-Type System of Dimension Five

The reduction process for the three-dimensional (geometric) Lorenz system can also be applied to systems with phase-space dimension $n \geq 4$. In direct analogy, one obtains an invariant foliation in a suitable $(n-1)$-dimensional cross-section with leaves of codimension one and dimension $n-2$; this would require that, near the Lorenz attractor, the additional directions are all stronger than those on the Lorenz attractor. Projection along stable leaves then results in a one-dimensional Lorenz map, meaning that the dynamics of such a vector field for $n \geq 4$ is just like that of the Lorenz system (1) itself.

To obtain a Lorenz-type vector field in higher dimensions with different dynamics from that of the Lorenz system (1), one needs to consider an example where the Poincaré map in a cross-section admits an invariant stable foliation of codimension at least two. In 2006, Bamón, Kiwi, and Rivera-Letelier [9] constructed such an abstract n-dimensional Lorenz-type vector field for $n \geq 5$ with a stable foliation of codimension two and dimension $n-3$ in the $(n-1)$-dimensional cross-section; the minimal case $n = 5$ contains all the geometric ingredients, and we restrict to it for simplicity in the discussion that follows. The central object in [9] is the corresponding two-dimensional noninvertible quotient map, which is given on the punctured complex plane as

$$\begin{aligned} f : \mathbb{C}\backslash\{0\} &\to \mathbb{C} \\ z &\mapsto (1-\lambda+\lambda\,|z|^a)\left(\frac{z}{|z|}\right)^2 + 1, \end{aligned} \tag{6}$$

with parameters $a, \lambda \in \mathbb{R}$ in the ranges $0 < a < 1$ and $0 < \lambda < 1$. Notice the term $|z|^a$, indicating a clear similarity with the form (5) of the one-dimensional Lorenz map.

A planar noninvertible map can have richer dynamics than a one-dimensional noninvertible map, where homoclinic tangencies can be at most dense in parameter space. Indeed, [9] provides a proof that there exists a small open region near the point $(a, \lambda) = (1, 1)$ in the (a, λ)-plane, such that the map (6) has a homoclinic tangency for every point from this parameter region; hence, homoclinic tangencies occur robustly, and the map, as well as the associated Lorenz-type vector field in $\mathbb{R}^5$, exhibit wild chaos.

As is the case for the one-dimensional Lorenz map, the origin in $\mathbb{C}$ does not have a well-defined image under (6). Hence, this point is a critical point, which we refer to as J_0. The critical point J_0 arises, as in Sect. 2.2, from the fact that it lies on the three-dimensional stable manifold of an equilibrium e of the five-dimensional Lorenz-type vector field, where e does not lie in the cross-section on which the four-dimensional Poincaré map is defined. The equilibrium e has a two-dimensional unstable manifold that corresponds in the planar map (6) to the critical circle

$$J_1 = \{z \in \mathbb{C} \mid \ |z - 1| = 1 - \lambda\}, \tag{7}$$

with radius $1 - \lambda$ around the point $z = 1$. The equilibrium e plays the role of the origin $\mathbf{0}$ of the Lorenz system (1) and, in complete analogy, the critical circle J_1 can be interpreted as the set-valued image of the critical point J_0. The map (6) maps the punctured complex plane $\mathbb{C}\backslash J_0$ in a two-to-one fashion—by angle doubling due to the term $(z/|z|)^2$—to the region outside the circle J_1; the centre of the angle-doubling is shifted by 1 with respect to $J_0 = 0$. Dynamics and bifurcations of this type of map are the subject of [28], where we consider a more general family with an additional complex parameter c for the shift; it is set to $c = 1$ in (6) for simplicity and in accordance with the formulation of the map in [9].

Our goal here is to present geometric mechanisms that are involved in the transition from simple dynamics to wild chaos in the map (6) as the point $(a, \lambda) = (1, 1)$ is approached. Key ingredients in this transition are different types of global bifurcations. The map (6) has fixed points and periodic points, which correspond to periodic orbits of the associated vector field. If they are saddles then these points have stable and unstable invariant sets, which are the generalisations of stable and unstable manifolds to the context of noninvertible maps; see, for example, [16, 17, 32] for more details. Points on the stable set $W^s(p)$ of a saddle periodic point p converge to p under iteration of f^k where k is the (minimal) period of p; note that $k = 1$ if p is a fixed point. Similarly, points on the unstable set $W^u(p)$ of p converge to p via a particular sequence of preimages of f^k. Note that $W^s(p)$ and $W^u(p)$ of the map (6) are one-dimensional objects, but they are typically not manifolds. The stable set $W^s(p)$ consists of a primary manifold $W^s_0(p)$ that contains p, and all preimages of $W^s_0(p)$, so that the stable set is typically a disjoint family of infinitely many one-dimensional manifolds. The unstable set may be an immersed one-dimensional manifold; however, the sequence of preimages of points in $W^u(p)$ may not be unique, in which case $W^u(p)$ has self-intersections. The stable and unstable sets of a saddle fixed or periodic point of the map (6) correspond to four-dimensional stable and two-dimensional unstable manifolds of the corresponding saddle periodic orbit in the five-dimensional Lorenz-type vector field.

Clearly, the stable and unstable sets of a fixed or periodic point p can become tangent, which is referred to as a *homoclinic tangency* and corresponds to a tangency between the respective manifolds of the associated periodic orbit in the five-dimensional Lorenz-type vector field. To characterise the additional global bifurcations that arise in the map (6) it is convenient to consider the *backward critical set*

$$\mathscr{J}^- := \cup_{k=0}^{\infty} f^{-k}(J_0),$$

of all preimages of the critical point J_0, and the *forward critical set*

$$\mathscr{J}^+ := \cup_{k=0}^{\infty} f^{k}(J_1),$$

of all images of the critical circle J_1. Note that $\mathscr{J}^-$ consists of potentially infinitely many discrete points, while $\mathscr{J}^+$ consists of infinitely many closed curves; we refer to $\mathscr{J} = \mathscr{J}^- \cup \mathscr{J}^+$ as the *critical set*. With this notation, we can define three further tangency bifurcations: the *forward critical tangency* where the stable set $W^s(p)$ becomes tangent to the circles in the forward critical set $\mathscr{J}^+$; the *backward critical tangency* where a sequence of points in the backward critical set $\mathscr{J}^-$ lies on the unstable set $W^u(p)$; and the *forward-backward critical tangency* where a sequence of points in the backward critical set $\mathscr{J}^-$ lies on the forward critical set $\mathscr{J}^+$. These three global bifurcation involving the critical set $\mathscr{J}$, as well as the homoclinic bifurcation, are encountered and discussed here as part of the transition to wild chaos. They are of codimension one, that is, they are encountered generically at isolated points when a single parameter is changed; their unfoldings are presented in detail in [28]. Note that a forward or backward critical tangency corresponds to a heteroclinic bifurcation between the corresponding periodic orbit and the equilibrium e of the Lorenz-type vector field. The forward-backward critical tangency, on the other hand, corresponds to the existence of an isolated homoclinic orbit of the saddle equilibrium e of the five-dimensional Lorenz-type vector field; it is the higher-dimensional analogue of how a homoclinic bifurcation in the Lorenz system (1) is described by the one-dimensional Lorenz map.

3.1 The Transition for Increasing $a = \lambda$

We now show a series of phase portraits as panels (a)–(l) of Fig. 6 that illustrate the bifurcations that are encountered in the transition to wild chaos and generate the robustness of homoclinic tangencies; more specifically, we increase a and λ along the diagonal $a = \lambda$ towards the point $(a, \lambda) = (1, 1)$, near which wild chaos was proven to exist [9]. To facilitate the visualisations, we project the complex plane $\mathbb{C}$ onto the Poincaré disk by stereographic projection, where the unit circle, that is, the boundary of the Poincaré disk represents the directions to infinity. In each phase portrait we show a suitable number of points in the backward critical set $\mathscr{J}^-$ (as dots) and the closed curves in the forward critical set $\mathscr{J}^+$. We remark that the circle J_1 with radius $1 - \lambda$ appears distorted in all phase portraits as a result of stereographic projection. For the values $a, \lambda \in \mathbb{R}$ that we consider, the map (6) has one fixed point p on the positive real line and a complex-conjugate pair of fixed points $q^{\pm}$. We plot these fixed points p and $q^{\pm}$, as well as the stable set $W^s(p)$ and unstable set $W^u(p)$ of the saddle point p; throughout, the points in $\mathscr{J}^-$ are branch points of the stable set $W^s(p)$. Notice that all phase portraits are symmetric with respect to complex conjugation, owing to the fact that $a, \lambda \in \mathbb{R}$. The phase portraits in Fig. 6 were obtained from computations of the transformed map on the Poincaré disk as follows: the fixed points p and $q^{\pm}$ can be found readily; $\mathscr{J}^-$ is represented by all backward images of J_0 under up to eleven backward iterations, that is, by $\cup_{k=0}^{11} f^{-k}(J_0)$; similarly, $\mathscr{J}^+$ is represented by J_1 and its next fourteen forward iterations; to obtain $W^s(p)$, we take advantage of the complex-conjugate symmetry and note that the primary manifold $W_0^s(p)$ is the real halfline $(0, \infty)$, which is the real interval $(0, 1]$ on the Poincaré disk; we computed eleven backward iterates of $W_0^s(p)$; finally, $W^u(p)$ was found by computing a first piece of arclength 5 and then plotting it and its next six iterates (in this way, we ensure that $W^u(p)$ maintains a suitable and comparable arclength as parameters are changed).

Figure 6a is for $a = \lambda = 0.7$, when the map (6) does not have chaotic dynamics, and all typical orbits converge to one of the two attracting fixed points $q^{\pm}$. The two branches of the unstable set $W^u(p)$ (which is an immersed manifold in this case) spiral towards q^+ and q^-, respectively. The preimages of $W_0^s(p)$ are organised in such a way that every point in the backward critical set $\mathscr{J}^-$ connects four branches of $W^s(p)$. Moreover, $\mathscr{J}^-$ accumulates on the boundary of the Poincaré disk. The forward critical set $\mathscr{J}^+$, on the other hand, accumulates on the unstable set $W^u(p)$. Figure 6b shows the phase portrait for $a = \lambda = 0.72$, just after a Neimark-Sacker bifurcation (or Hopf bifurcation for maps) [34]. The fixed points $q^{\pm}$ are now repellors and $W^u(p)$ and $\mathscr{J}^+$ accumulate on two invariant closed curves (not shown), which correspond to invariant tori in the associated Lorenz-type vector field. As a and λ change, these invariant closed curves undergo various bifurcations (associated with resonance phenomena) that we do not discuss here. Figure 6c shows the phase portrait for $a = \lambda = 0.73277$, approximately at the moment that $W^s(p)$ and $W^u(p)$ have a first homoclinic tangency. Since $W^u(p)$ accumulates on itself, this first homoclinic tangency is accumulated in parameter space, on the side of larger $a = \lambda$, by infinitely many homoclinic tangencies. As is shown in Fig. 6d for $a = \lambda = 0.745$, after the first homoclinic tangency there is a homoclinic tangle between $W^s(p)$ and $W^u(p)$. Therefore, the system is now chaotic in the classical sense, meaning that any homoclinic tangency between $W^s(p)$ and $W^u(p)$ is accumulated by further homoclinic tangencies with associated saddle hyperbolic sets and horseshoe dynamics; see, for example, [11, 39]. Notice also that $W^s(p)$ accumulates on itself and the two branches of the unstable set $W^u(p)$ now intersect. Moreover, the forward critical set $\mathscr{J}^+$ accumulates on $W^u(p)$, so that the first homoclinic tangency is also accumulated in parameter space, on the side of larger $a = \lambda$, by infinitely many forward critical tangencies; indeed, in Fig. 6d there is a tangle between $W^s(p)$ and $\mathscr{J}^+$ as a result. Furthermore, the forward critical tangencies have the effect that the points in $\mathscr{J}^-$ are branch points to infinitely many, instead of four branches of $W^s(p)$; see also [28]. In Fig. 6d, this can be seen at the origin, where an additional eight branches are shown to connect to 0; these are preimages of the two additional branches of $W^s(p)$ that intersect J_1.

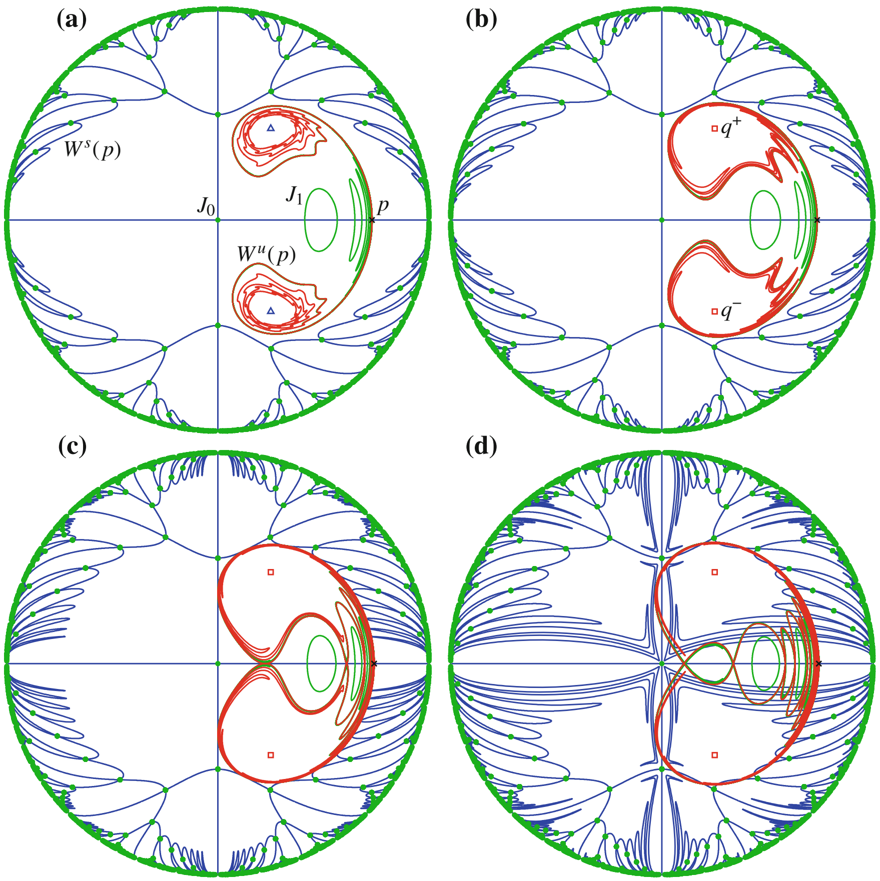

Fig. 6 The objects p (cross), $q^{\pm}$ (triangles when attracting, and squares when repelling), $W^s(p)$, $W^u(p)$, $\mathscr{J}^-$ and $\mathscr{J}^+$ on the Poincaré disk; from (**a**) to (**d**) $a = \lambda$ take the values 0.7, 0.72, 0.73277 and 0.745; from (**e**) to (**h**) $a = \lambda$ take the values 0.76302, 0.765, 0.77 and 0.8; and from (**i**) to (**l**) $a = \lambda$ take the values 0.85, 0.87, 0.9 and 0.95

In Fig. 6e for $a = \lambda = 0.76302$ one encounters the first backward critical tangency, where the unstable set $W^u(p)$ goes through the critical point $J_0 = 0$, which implies that $W^u(p)$ contains two sequences of preimages of J_0 (two because of symmetry). Since $W^u(p)$ accumulates on itself, this first backward critical tangency is accumulated in parameter space, on the side of larger $a = \lambda$, by infinitely many backward critical tangencies. Observe from Fig. 6f for $a = \lambda = 0.765$ how these interactions with J_0 induce effects near J_1 and its images. As a result of this first backward critical tangency, $W^u(p)$ has points of self-intersection on each of its two branches (in addition to the intersections between the two branches). Consider the region $\mathscr{A}$ enclosed by the first segments of the two branches of $W^u(p)$ up to when they meet on the real line. Before the backward critical tangency all points of $\mathscr{J}^-$

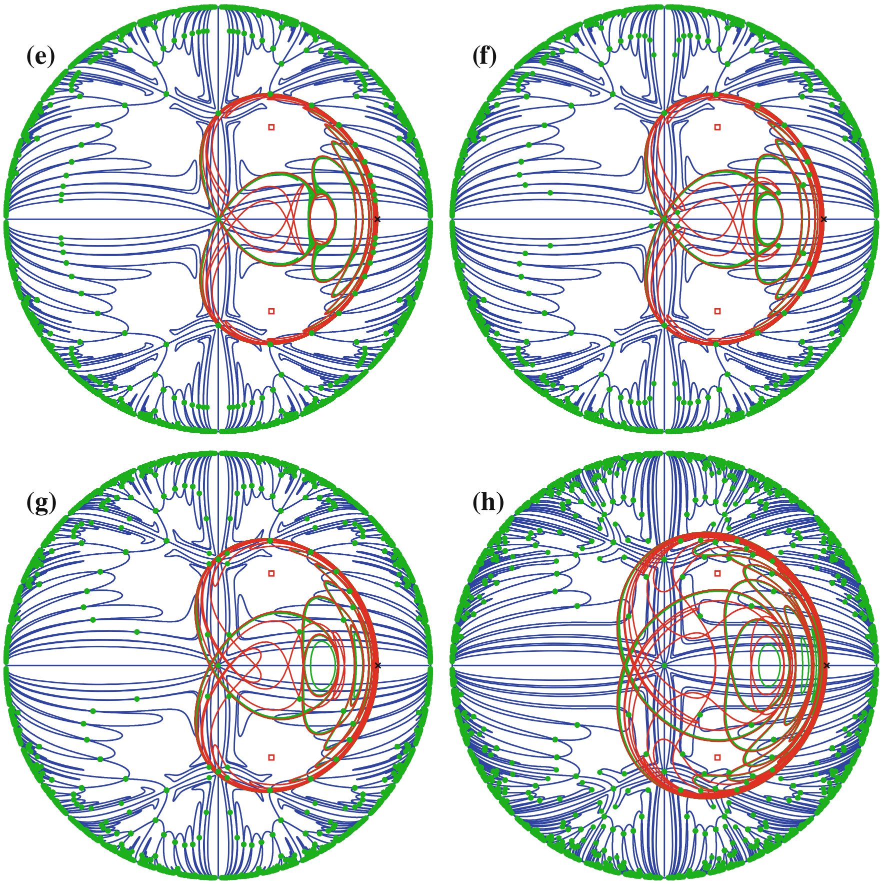

Fig. 6 (continued)

lie outside the region $\mathscr{A}$. In this and the accumulating further backward critical tangencies, more and more points of $\mathscr{J}^-$ move inside this region; see also Fig. 6g and h for $a = \lambda = 0.77$ and $a = \lambda = 0.8$, respectively. Moreover, the map (6) has a chaotic attractor in the region $\mathscr{A}$, which is the closure of the unstable set $W^u(p)$ and, hence, also contains p. Because the forward critical set $\mathscr{J}^+$ accumulates on $W^u(p)$, the first backward critical tangency is also accumulated in parameter space, on the side of larger $a = \lambda$, by infinitely many forward-backward critical tangencies between $\mathscr{J}^+$ and $\mathscr{J}^-$. The forward-backward critical tangencies lead to the disappearance of certain sequences of backward orbits of J_0 from the backward critical set $\mathscr{J}^-$; moreover, the closed curves in $\mathscr{J}^+$ develop self-intersections in the process. These effects of the forward-backward critical tangencies are difficult to discern in the phase portraits (f)–(l) of Fig. 6; see [28] for details and illustrations.

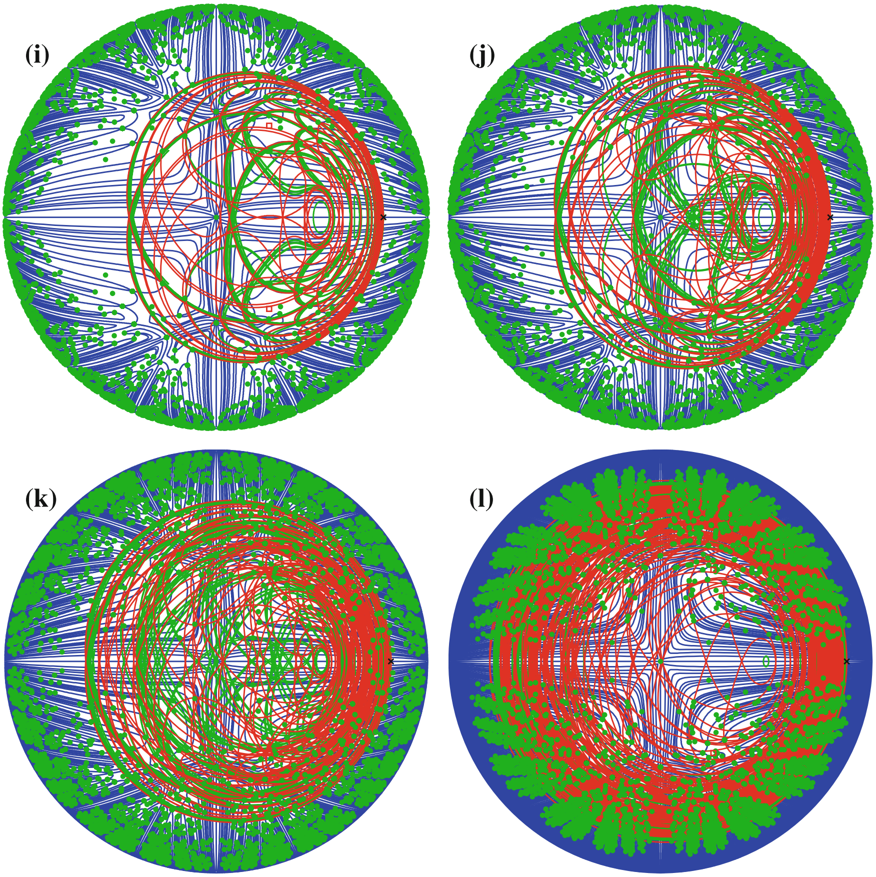

Fig. 6 (continued)

When $a = \lambda$ is increased further, $W^u(p)$ and, thus, the region $\mathscr{A}$ grows and incorporates more and more points of $\mathscr{J}^-$; see Fig. 6i–k for $a = \lambda = 0.85$, $a = \lambda = 0.87$ and $a = \lambda = 0.9$, respectively. At the same time, the sets $W^s(p)$, $W^u(p)$ and $\mathscr{J}$ seem to become denser in the Poincaré disk, leading to ever more associated tangency bifurcations when $a = \lambda$ is increased. As Bamón, Kiwi, and Rivera-Letelier showed in [9], near $a = \lambda = 1$ the tangency bifurcations between stable and unstable sets of the hyperbolic saddle of (6) occur robustly. This means that there exists $0 \ll w^* < 1$, such that one finds a homoclinic tangency of the hyperbolic saddle for every point $(a, \lambda) \in (w^*, 1) \times (w^*, 1)$. We believe that Fig. 6l for $a = \lambda = 0.95$ gives some impression of what wild chaos, that is, the robustness of homoclinic tangencies might look like. The saddle point p is only one of uncountably infinitely many nonwandering points; yet the sets $W^s(p)$ and $W^u(p)$ and the critical set $\mathscr{J}$ already fill out the Poincaré disk increasingly densely.

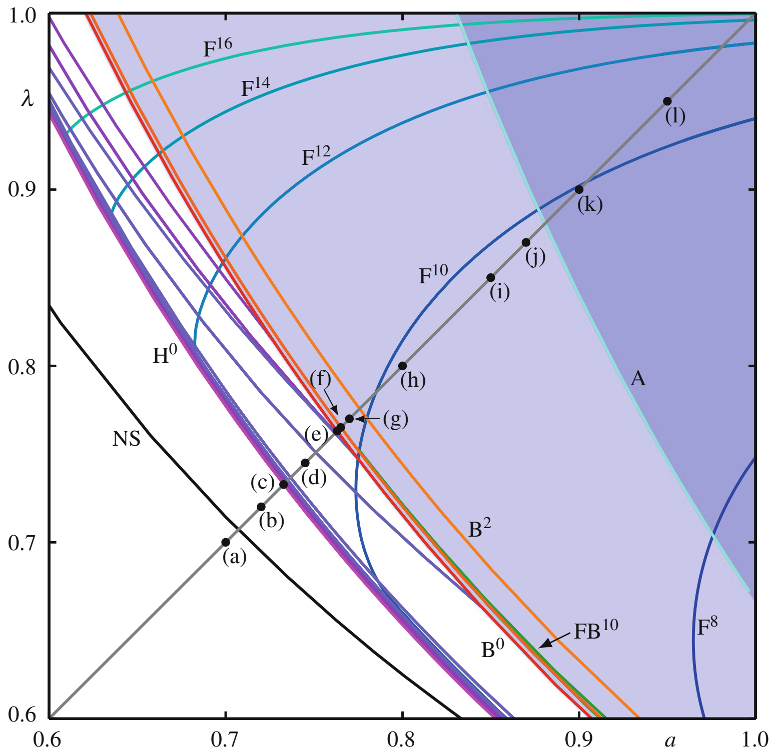

Fig. 7 Bifurcation diagram of (6) in the (a, λ)-plane for $a, \lambda \in \mathbb{R}$. Shown are the curve NS of Neimarck-Sacker bifurcation of $q^{\pm}$, the curve H^0 of first homoclinic tangency and two further nearby curves of homoclinic tangencies, the curves F^k for $k \in \{8, 10, 12, 14, 16\}$ of forward critical tangencies, the curve B^0 of first backward critical tangency and two further nearby curves of backward critical tangencies (one of which is labelled B^2), the curve FB^{10} of forward-backward critical tangency, and the curve A (cyan) along which $\det(Df(p)) = 1$. The labelled points along the diagonal $a = \lambda$ correspond to the panels of Fig. 6

3.2 The Bifurcation Diagram in the (a, λ)-plane

The bifurcations that are encountered as $a = \lambda$ is increased towards $a = \lambda = 1$ can be continued as curves when a and λ are allowed to vary independently. For tangency bifurcations this is done via the formulation of a suitable boundary value problem. These computations are based on the technique for continuing a locus of homoclinic tangency described in [10], which has been implemented in Cl_MatContM [18, 23]; details on how we adapted this method can be found in [28]. Figure 7 shows the resulting bifurcation diagram of (6) in the (a, λ)-plane; the points labelled (a)–(l) along the diagonal are the parameter points of the phase portraits of Fig. 6. Starting from the lower-left corner, one first encounters the Neimarck-Sacker bifurcation NS. The system then becomes chaotic when the curve H^0 of homoclinic tangency between $W^s(p)$ and $W^u(p)$ is crossed. As we already discussed, there are many more homoclinic tangencies that accumulate on H^0 and two of them are shown in Fig. 7. These curves of secondary homoclinic tangencies turn around and cross the diagonal at least twice, between the points (c) and (d) and between the points (d) and (e); they each end on the curve B^0 of first backward critical tangency where $W^u(p)$ interacts with $\mathscr{J}^-$. Also shown in Fig. 7 are five curves F^k of forward critical tangency between the primary manifold $W^s_0(p)$ and $f^{(k-1)}(J_1)$, namely, those for $k = 8$, 10, 12, 14 and 16. Observe how each curve F^k passes very close to H^0 before turning away towards the right boundary of Fig. 7 and note that F^k for $k = 12, 14$ and 16 cross the diagonal very close to the curve H^0. The curve B^0 is accumulated by curves of further backward critical tangencies, for example, the curve B^2. Figure 7 also shows the curve FB^{10} of forward-backward critical tangency between J_0 and $f^9(J_1)$, which lies very close to B^0.

While the proof in [9] is valid only very close to the point $a = \lambda = 1$, the bifurcation diagram in Fig. 7 suggests that one might expect to encounter wild chaos in a much larger region of the (a, λ)-plane. As soon as B^0 is crossed, infinitely many forward-backward critical tangencies have occured, which are codimension-one homoclinic bifurcations of the equilibrium e of the five-dimensional Lorenz-type vector field; as such, they play the role of the homoclinic bifurcation in the Lorenz system (1). Apart from this geometric ingredient, the proof in [9] also requires that the parameters are such that (6) is area-expanding in a neighbourhood of the chaotic attractor. In [28] we conjecture that homoclinic tangencies occur robustly to the right of the first backward critical tangency B^0; this region is shaded in Fig. 7. This is based on the suggestion that (6) is area-expanding in a neighbourhood of a subset of the attractor in this region. A sufficient (but not necessary) condition to ensure this area-expanding property is that the product of the eigenvalues of p exceeds 1. The curve A in Fig. 7 is the locus where $\det(Df(p)) = 1$, and (6) is area-expanding in a neighbourhood of the chaotic attractor in the darker shaded region to the right of A. Hence, in this darker region wild chaos should certainly be expected. In particular, this means that the phase portraits of Fig. 6k and l, and possibly also those of Fig. 6g–j, are already from the regime of wild chaos.

4 Conclusions

We presented a geometric perspective of the techniques used to prove the existence of chaos in the Lorenz system (1). The same approach can also be applied to the study of wild chaos in higher-dimensional Lorenz-type vector fields. We focussed here on the two-dimensional noninvertible map (6) by Bámon, Kiwi and Rivera-Letelier [9] and discussed how interactions between its invariant objects are directly related to homoclinic and heteroclinic bifurcations of the associated five-dimensional Lorenz-type vector field. In this way, we were able to describe geometric changes in (6) during the transition from non-chaotic, via chaotic to wild chaotic dynamics. Our numerical results provide guidance for further theoretical study. In particular, we proposed the conjecture that the wild chaotic regime for (6) starts as soon as the first backward critical tangency bifurcation has occurred. Due to the accumulative nature of the respective objects, the first backward critical tangency induces infinitely many forward-backward critical tangencies, which emerge as a main ingredient for wild chaos. It remains to show that, in this regime, the attractor has the necessary area-expanding properties. The numerical methods we employed can be used to investigate other two-dimensional noninvertible maps and associated vector fields. In particular, it is of interest to explore possible routes to wild chaos in these other examples.

Acknowledgments The work on the Lorenz system presented here has been performed in collaboration with Eusebius Doedel. We acknowlegde his contribution to the computation of the Lorenz attractor as shown in Fig. 1 and of the Lorenz manifold on the sphere in Figs. 2 and 3; moreover, the leaves of the stable foliation in Fig. 5 were computed with AUTO demo files that he developed recently. HMO and BK thank the organisers of ICDEA 2013 for their support, financial and otherwise.

References

1. Afrajmovich, V.S., Bykov, V.V., Sil′nikov, L.P.: The origin and structure of the Lorenz attractor, (Trans: Dokl. Akad. Nauk SSSR 234(2), 336–339 (1977)). Sov. Phys. Dokl. **22**, 253–255 (1977)
2. Afrajmovich, V.S., Bykov, V.V., Sil′nikov, L.P.: On structurally unstable attracting limit sets of Lorenz attractor type. Trans. Mosc. Math. Soc. **44**, 153–216 (1983)
3. Alligood, K.T., Sander, E., Yorke, J.A.: Crossing bifurcations and unstable dimension variability. Phys. Rev. Lett. **96**, 244103 (2006)
4. Alligood, K.T., Sauer, T.D., Yorke, J.A.: Chaos—An Introduction to Dynamical Systems. Springer, New York (1996)
5. Aguirre, P., Doedel, E.J., Krauskopf, B., Osinga, H.M.: Investigating the consequences of global bifurcations for two-dimensional invariant manifolds of vector fields. Discr. Contin. Dynam. Syst.—Ser. A **29**(4), 1309–1344 (2011)
6. Asaoka, M.: Hyperbolic sets exhibiting C^1-persistent homoclinic tangency for higher dimensions. Proc. Amer. Math. Soc. **136**, 677–686 (2008)
7. Asaoka, M.: Erratum to Hyperbolic sets exhibiting C^1-persistent homoclinic tangency for higher dimensions. Proc. Amer. Math. Soc. **138**, 1533 (2010)
8. Bunimovich, L.A., Sinai, J.G.: Stochasticity of the attractor in the Lorenz model. In: Gaponov-Grekhov, A.V. (ed.) Nonlinear Waves, Proceedings of Winter School, pp. 212–226. Nauka Press, Moscow (1979)
9. Bamón, R., Kiwi, J., Rivera-Letelier, J.: Wild Lorenz like attractors. arXiv 0508045 (2006)
10. Beyn, W.-J., Kleinkauf, J.-M.: The numerical computation of homoclinic orbits for maps. SIAM J. Numer. Anal. **34**, 1207–1236 (1997)
11. Bonatti, C., Díaz, L., Viana, M.: Dynamics Beyond Uniform Hyperbolicity. A global geometric and probabilistic perspective. Encycl. Math. Sci. **102**. (2005)
12. Creaser, J., Krauskopf, B., Osinga, H.M.: α-flips in the Lorenz System. The University of Auckland, Auckland (2014)
13. Doedel, E. J.: AUTO-07P: Continuation and bifurcation software for ordinary differential equations. with major contributions from Champneys, A. R., Fairgrieve, T. F., Kuznetsov, Yu. A., Oldeman, B. E., Paffenroth, R. C., Sandstede, B., Wang, X. J., Zhang, C. Available at http://cmvl.cs.concordia.ca/auto

14. Doedel, E.J., Krauskopf, B., Osinga, H.M.: Global bifurcations of the Lorenz manifold. Nonlinearity **19**(12), 2947–2972 (2006)
15. Doedel, E. J., Krauskopf, B., Osinga, H. M.: Global invariant manifolds in the transition to preturbulence in the Lorenz system. Indag. Math. (N.S.) **22**(3–4), 222–240 (2011)
16. England, J.P., Krauskopf, B., Osinga, H.M.: Computing one-dimensional stable manifolds and stable sets of planar maps without the inverse. SIAM J. Appl. Dynam. Syst. **3**(2), 161–190 (2004)
17. England, J.P., Krauskopf, B., Osinga, H.M.: Bifurcations of stable sets in noninvertible planar maps. Internat. J. Bifur. Chaos Appl. Sci. Engrg. **15**(3), 891–904 (2005)
18. Ghaziani, R.K., Govaerts, W., Kuznetsov, Y.A., Meijer, H.G.E.: Numerical continuation of connecting orbits of maps in Matlab. J. Differ. Equ. Appl. **15**, 849–875 (2009)
19. Ghrist, R., Holmes, P. J., Sullivan, M. C.: Knots and Links in Three-Dimensional Flows. Lecture Notes in Mathematics 1654. Springer, Berlin (1997)
20. Gilmore, R., Lefranc, M.: The Topology of Chaos: Alice in Stretch and Squeeze Land. Wiley-Interscience, New York (2004)
21. Gonchenko, S.V., Ovsyannikov, I.I., Simó, C., Turaev, D.: Three-dimensional Hénon-like maps and wild Lorenz-like attractors. Internat. J. Bifur. Chaos Appl. Sci. Engrg. **15**, 3493–3508 (2005)
22. Gonchenko, S.V., Shilnikov, L.P., Turaev, D.: On global bifurcations in three-dimensional diffeomorphisms leading to wild Lorenz-like attractors. Regul. Chaotic Dyn. **14**, 137–147 (2009)
23. Govaerts, W., Kuznetsov, Y.A., Ghaziani, R.K., Meijer, H.G.E.: Cl_MatContM: a toolbox for continuation and bifurcation of cycles of maps (2008). Available via http://sourceforge.net/projects/matcont
24. Guckenheimer, J.: A strange strange attractor. In: Marsden, J.E., McCracken, M. (eds.) The Hopf Bifurcation and its Applications, pp. 368–382. Springer, New York (1976)
25. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields, 2nd edn. Springer, New York (1986)
26. Guckenheimer, J., Williams, R.F.: Structural stability of Lorenz attractors. Publ. Math. IHES **50**, 59–72 (1979)
27. Hénon, M.: A two-dimensional mapping with a strange attractor. Commun. Math. Phys. **50**(1), 69–77 (1976)
28. Hittmeyer, S., Krauskopf, B., Osinga, H.M.: Interacting global invariant sets in a planar map model of wild chaos. SIAM J. Appl. Dynam. Syst. **12**(3), 1280–1329 (2013)
29. Kaplan, J.L., Yorke, J.A.: Preturbulence: a regime observed in a fluid flow model of Lorenz. Commun. Math. Phys. **67**, 93–108 (1979)
30. Krauskopf, B., Osinga, H.M.: Computing geodesic level sets on global (un)stable manifolds of vector fields. SIAM J. Appl. Dynam. Sys. **2**(4), 546–569 (2003)
31. Krauskopf, B., Osinga, H.M.: Computing invariant manifolds via the continuation of orbit segments. In: Krauskopf, B., Osinga, H.M., Galán-Vioque, J. (eds.) Numerical Continuation Methods for Dynamical Systems. Understanding Complex Systems, pp. 117–157. Springer, New York (2007)
32. Krauskopf, B., Osinga, H.M., Peckham, B.B.: Unfolding the cusp-cusp bifurcation of planar endomorphisms. SIAM J. Appl. Dynam. Syst. **6**(2), 403–440 (2007)
33. Krauskopf, B., Osinga, H.M., Doedel, E.J., Henderson, M.E., Guckenheimer, J., Vladimirsky, A., Dellnitz, M., Junge, O.: A survey of methods for computing (un)stable manifolds of vector fields. Internat. J. Bifur. Chaos Appl. Sci. Engrg. **15**(3), 763–791 (2005)
34. Kuznetsov, Yu, A.: Elements of Applied Bifurcation Theory, 2nd edn. Springer, New York (1998)
35. Lorenz, E.N.: Deterministic nonperiodic flows. J. Atmosph. Sci. **20**, 130–141 (1963)
36. Moreira, C.G.: There are no C^1-stable intersections of regular Cantor sets. Acta Math. **206**, 311–323 (2011)
37. Newhouse, S.E.: The abundance of wild hyperbolic sets and nonsmooth stable sets for diffeomorphisms. Publ. Math. IHES **50**(1), 101–151 (1979)

38. Palis, J., de Melo, W.: Geometric Theory of Dynamical Systems. Springer, New York (1982)
39. Palis, J., Takens, F.: Hyperbolicity and Sensitive Chaotic Dynamics at Homoclinic Bifurcations. Cambridge University Press, Cambridge (1993)
40. Perelló, C.: Intertwining invariant manifolds and Lorenz attractor. In: Global theory of dynamical systems. Proceedings of International Conference, Northwestern University, Evanston, Ill., 1979, Lecture Notes in Math, vol. 819, pp. 375–378. Springer, Berlin (1979)
41. Rand, D.: The topological classification of Lorenz attractors. Math. Proc. Cambridge Philos. Soc. **83**, 451–460 (1978)
42. Robinson, C.: Dynamical Systems: Stability, Symbolic Dynamics, and Chaos, 2nd edn. CRC Press LLC, Boca Raton (1999)
43. Sparrow, C.: The Lorenz Equations: Bifurcations, Chaos and Strange Attractors. Springer, New York (1982)
44. Sinai, J.G., Vul, E.B.: Hyperbolicity conditions for the Lorenz model. Physica D **2**, 3–7 (1981)
45. Tucker, W.: The Lorenz attractor exists. C. R. Acad. Sci. Paris Sér. I Math. **328**(12), 1197–1202 (1999)
46. Turaev, D.V., Shilnikov, L.P.: An example of a wild strange attractor. Mat. Sb. **189**, 291–314 (1998)
47. Turaev, D.V., Shilnikov, L.P.: Pseudo-hyperbolicity and the problem on periodic perturbations of Lorenz-like attractors. Russian Dokl. Math. **77**, 17–21 (2008)
48. Williams, R.F.: The structure of Lorenz attractors. Publ. Math. IHES **50**, 101–152 (1979)
49. Zhang, W., Krauskopf, B., Kirk, V.: How to find a codimension-one heteroclinic cycle between two periodic orbits. Discr. Contin. Dynam. Syst.–Ser. A **32**(8), 2825–2851 (2012)

Part II
Contributed Papers

Almost Automorphic Sequences and Their Application to a Model of a Cellular Neural Network

Syed Abbas

Abstract In this paper we show the almost automorphic sequence solution of a model of a cellular neural networks with piecewise constant argument. We convert the model into a corresponding difference equation model and then show the existence and global attractivity of solutions.

Keywords Almost automorphic sequence · Difference equations · Neural networks

1 Introduction

Since the introduction of almost periodic functions by Bohr [1], there have been many important generalizations of these functions. One of the important generalization of almost periodic functions is that of almost automorphic functions introduced by Bochner [2].

It is natural to inquire about a discrete counterpart of these functions. The theory of discrete almost automorphic sequence develops in parallel with that of almost automorphic functions. As we know, difference equations play an important role in many fields like numerical analysis, population dynamics, etc. Hence, many mathematicians have investigated the almost periodicity/almost automorphy of the solution to difference equations (see: [3–9]). Application of these functions in differential equations has been studied by several authors (for example [10–12] and references therein).

Neural networks are important in artificial networks because of their richness as theoretical models of collective dynamics. Many authors have recently studied the dynamics of neural-networks. For example, the network model proposed by Hopfield [13] is described by an ordinary differential equation of the form

S. Abbas (✉)
School of Basic Sciences, Indian Institute of Technology Mandi,
Mandi 175001, India
e-mail: sabbas.iitk@gmail.com

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_5

$$C_i \frac{dx_i(t)}{dt} = -\frac{1}{R_i} x_i(t) + \sum_{j=1}^{n} T_{ij} f_j(x_j(t)), \quad 1 \le i \le n, \ t \ge 0, \tag{1}$$

where the variable $x_i(t)$ denotes the voltage of the input of the ith neuron. Each neuron is characterized by an input capacitance C_i and a transfer function $f_i(x)$. The connection matrix element T_{ij} has a value $\frac{1}{R_{ij}}$ when the non-inverting output of the jth neuron is connected to the input of the ith neuron through a resistance R_{ij}, and a value $\frac{-1}{R_{ij}}$ when the inverting output of the jth neuron is connected to the input of the ith neuron through a resistance R_{ij}. The parallel resistance at the input of each neuron is defined by $R_i = \left(\sum_{j=1}^{n} |Tij| \right)^{-1}$. The nonlinear transfer function $g_i(u)$ is sigmoidal, saturating at 1 with maximum slope at $u = 0$. By defining

$$b_i = \frac{1}{C_i R_i}, \ a_{ij} = \frac{T_{ij}}{C_i}$$

we can re-write the differential equations as

$$\frac{dx_i(t)}{dt} = -b_i x_i(t) + \sum_{j=1}^{n} a_{ij} f_j(x_j(t)), 1 \le i \le n.$$

The theory of discrete almost automorphic sequences has been developed in a paper by Araya et al. [7]. Using this foundation, we establish the existence of a unique almost automorphic sequence solution for a general discrete model of cellular neural network. A cellular neural network is a nonlinear dynamic circuit consisting of many processing units called cells arranged in two or three dimensional array. This is useful in the areas of signal processing, image processing, pattern classification and associative memories. Hence, the application of cellular networks is of great interest to many researchers. For more details on neural networks, the interested reader may consult [3, 5, 6, 8, 14–16]. In [8, 14–16] the authors dealt with the global exponential stability and the existence of a periodic solution of a cellular neural network with delays using the general method of Lyapunov functional. The discrete analogue of continuous time cellular network models is important for theoretical analysis as well as for implementation. Thus, it is essential to formulate a discrete time analogue of continuous time network. A reasonable method is to discretize the continuous time network. For detailed analysis on the discretization method, the reader may consult Mohamad and Gopalsamy [16], Stewart [17].

The purpose of this paper is to study the problem of existence, uniqueness and exponential attractivity of almost automorphic solution of the following difference-differential equation model of a neural network,

$$\frac{dx_i(t)}{dt} = -\, a_i([t]) x_i(t) + \sum_{j=1}^{m} b_{ij}([t]) f_j(x_j(t))$$

$$+\sum_{j=1}^{m} c_{ij}([t])f_j(x_j(s-\tau_{ij})ds) + I_i([t]),$$
$$x_i(t) = \phi_i(t), \quad t \in [-\tau_{ij}, 0], \tag{2}$$

where $i = 1, 2, \ldots, m$ and $[\cdot]$ denote the greatest integer function. The function $x_i(t)$ is the potential of the cell i at time t and f_i is the nonlinear output function. b_{ij} and c_{ij} denote the strengths of connectivity between the cells i and j at the instants t and $t - \tau_{ij}$, respectively. τ_{ij} is the time delay required in processing and transmitting a signal from jth cell to the ith cell. We denote the ith component of an external input source from outside the network to the cell i by I_i. This is a generalization of the result of paper [5] for delayed model.

2 Preliminaries and Main Results

Assume X be a real or complex Banach space endowed with the norm $\|\cdot\|_X$.

Definition 1 A function $f : \mathbb{Z} \to X$ is said to be almost automorphic sequence if for every sequence of integer $\{k_l\}_{l\in\mathbb{N}}$ there exists a subsequence $\{k_n\}_{n\in\mathbb{N}}$ such that

$$f(k + k_n) \to g(k)$$

and

$$g(k - k_n) \to f(k)$$

for each $k \in \mathbb{Z}$. This is also equivalent to

$$\lim_{n\to\infty}\lim_{m\to\infty} f(k + k_n - k_m) = f(k)$$

for each $k \in \mathbb{Z}$.

Denote by *AAS(X)* the set of all almost automorphic sequences from $\mathbb{Z}$ to X. Then $(AAS(X), \|\cdot\|_{AAS(X)})$ is a Banach space with the supremum norm given by

$$\|u\|_{AAS(X)} = \sup_{k\in\mathbb{Z}} \|u(k)\|_X.$$

Definition 2 A function $f : \mathbb{Z} \times X \to X$ is said to be almost automorphic sequence in k for each $x \in X$ if for every sequence of integers $\{k_l\}_{l\in\mathbb{N}}$ there exists a subsequence $\{k_n\}_{n\in\mathbb{N}}$ such that

$$f(k + k_n, x) \to g(k, x)$$

and

$$g(k - k_n, x) \to f(k, x)$$

for each $k \in \mathbb{Z}$ and $x \in X$.

The set of all such functions is denoted by $AAS(\mathbb{Z} \times X, X)$.

For any almost automorphic functions $f(t)$ over $\mathbb{R}$, the sequence $\{x_n\}$ define by $x(n) = f(n)$ for $n \in \mathbb{Z}$ is almost automorphic.

Example Consider the function

$$f(k) = \text{signum}(\cos 2\pi k\theta).$$

This function is almost automorphic.

The discrete analogue of the model (2) is given by

$$\begin{aligned} x_i(n+1) = x_i(n)e^{-a_i(n)} + \frac{1 - e^{-a_i(n)}}{a_i(n)} \Big\{ &\sum_{j=1}^{m} b_{ij}(n) f_j(x_j(n)) \\ &+ \sum_{j=1}^{m} c_{ij}(n) f_j(x_j(n-\tau)) + I_i(n) \Big\}, \\ &x_i(n) = \phi_i(n), \quad \tau \le n \le 0, \end{aligned} \tag{3}$$

where $i = 1, 2, \ldots, m, \; n \in \mathbb{Z}$. We prove the existence and global attractivity of almost automorphic sequence solutions of Eq. (3) in this section. For more details of this kind of model without delay term and $c_{ij} = 0$, we refer to Huang et al. [9] in which the authors proved the existence of an almost periodic sequence solution.

The assumptions described below are necessary to show the existence of almost automorphic solutions of Eq. (3).

- (A_1) $a_i(n) > 0$ is almost automorphic sequence and $b_{ij}(n), \; c_{ij}(n), \; I_i(n)$ are almost automorphic sequence for $i, j = 1, 2, \ldots, m$.
- (A_2) There exist positive constants M_j and L_i such that $|f_j(x)| \le M_j$ and $|f_i(x) - f_i(y)| \le L_i|x - y|$ for each $x, y \in \mathbb{R}$ and $j = 1, 2, \ldots, m; \; i = 1, 2, \ldots, m$.

For discrete Eq. (3), let us introduce the following notations,

$$C_i(n) = e^{-a_i(n)}, \quad D_{ij}(n) = b_{ij}(n)\frac{1 - e^{-a_i(n)}}{a_i(n)},$$

$$E_{ij}(n) = c_{ij}(n)\frac{1 - e^{-a_i(n)}}{a_i(n)}, \quad F_i(n) = I_i(n)\frac{1 - e^{-a_i(n)}}{a_i(n)}.$$

Using the above notations, we can re-write Eq. (3) as,

$$x_i(n+1) = C_i(n)x_i(n) + \sum_{j=1}^{m} \Big(D_{ij}(n)f_j(x_j(n)) + E_{ij}(n)f_j(x_j(n-\tau)) \Big) + F_i(n), \quad (4)$$

for $i = 1, 2, \ldots, m$. The Eq. (4) is clearly a difference equation and in this work we focus on this equation only.

Denote:

$$C_i^* = \sup_{n\in\mathbb{Z}} |C_i(n)|, \quad I_i^* = \sup_{n\in\mathbb{Z}} |I_i(n)|,$$

$$D_{ij}^* = \sup_{n\in\mathbb{Z}} |D_{ij}(n)|, \quad E_{ij}^* = \sup_{n\in\mathbb{Z}} |E_{ij}(n)|, \quad F_i^* = \sup_{n\in\mathbb{Z}} |F_i(n)|,$$

$$b_{ij}^* = \sup_{n\in\mathbb{Z}} |b_{ij}(n)|, \quad a_i^* = \inf_{n\in\mathbb{Z}} a_i(n), \quad P_i = \sum_{j=1}^{m} (D_{ij}^* + E_{ij}^*)M_j + F_i^*.$$

Definition 3 A solution $x(\nu) = (x_1(\nu), \ldots, x_m(\nu))^T$ of (4) is said to be globally attractive if for any other solution $y(\nu) = (y_1(\nu), \ldots, y_m(\nu))^T$ of (4), we have

$$\lim_{\nu\to\infty} |x_i(\nu) - y_i(\nu)| = 0.$$

Lemma 1 *Suppose assumption* ($\mathbf{A_1}$) *holds, then* $C_i \in AAS$ *and* D_{ij}, E_{ij}, $F_i \in AAS$ *for* $i, j = 1, 2, \ldots, m$.

Proof From the assumption ($\mathbf{A_1}$) we know that $a_i(n)$ is almost automorphic sequence. Thus for any sequence k_l there exists a subsequence k_m such that

$$a_i(n+k_m) \to a_{i1}(n) \text{ and } a_{i1}(n-k_m) \to a_i(n).$$

Denoting $C_{i1}(n) = e^{-a_{i1}(n)}$, we have

$$\begin{aligned} &|C_i(n+k_m) - C_{i1}(n)| \\ &= |e^{-a_i(n+k_m)} - e^{-a_{i1}(n)}| \le |a_i(n+k_m) - a_{i1}(n)| \to 0, \end{aligned} \quad (5)$$

as $m \to \infty$. Also

$$\begin{aligned} &|C_{i1}(n-k_m) - C_i(n)| \\ &= |e^{-a_{i1}(n-k_m)} - e^{-a_i(n)}| \le |a_{i1}(n-k_m) - a_i(n)| \to 0, \end{aligned} \quad (6)$$

as $m \to \infty$. Thus one can conclude that $C_i(n)$ are almost automorphic. Now since b_{ij} and I_i are almost automorphic, we have

$$b_{ij}(n+k_m) \to \bar{b}_{ij}(n) \text{ and } \bar{b}_{ij}(n-k_m) \to b_{ij}(n)$$

and

$$I_i(n+k_m) \to \bar{I}_i(n) \text{ and } \bar{I}_i(n-k_m) \to I_i(n).$$

By assuming $D_{ij}(n) = b_{ij}(n)\frac{1-e^{-a_i(n)}}{a_i(n)}$ and $\bar{D}_{ij}(n) = \bar{b}_{ij}(n)\frac{1-e^{-a_i(n)}}{a_i(n)}$, we obtain

$$\begin{aligned}
&|D_{ij}(n+k_m) - \bar{D}_{ij}(n)| \\
&= \left| b_{ij}(n+k_m)\frac{1-e^{-a_i(n+k_m)}}{a_i(n+k_m)} - \bar{b}_{ij}(n)\frac{1-e^{-a_i(n)}}{a_i(n)} \right| \\
&\le |b_{ij}(n+k_m) - \bar{b}_{ij}(n)| \times \left| \frac{1-e^{-a_i(n+k_m)}}{a_i(n+k_m)} \right| \\
&\quad + |\bar{b}_{ij}(n)| \times \left| \frac{1-e^{-a_i(n+k_m)}}{a_i(n+k_m)} - \frac{1-e^{-a_i(n)}}{a_i(n)} \right| \\
&\quad \to \infty, \qquad \text{as } m \to \infty.
\end{aligned} \tag{7}$$

Similarly

$$\begin{aligned}
&|\bar{D}_{ij}(n-k_m) - D_{ij}(n)| \\
&= \left| \bar{b}_{ij}(n-k_m)\frac{1-e^{-a_i(n-k_m)}}{a_i(n-k_m)} - b_{ij}(n)\frac{1-e^{-a_i(n)}}{a_i(n)} \right| \\
&\le |\bar{b}_{ij}(n-k_m) - b_{ij}(n)| \times \left| \frac{1-e^{-a_i(n-k_m)}}{a_i(n-k_m)} \right| \\
&\quad + |b_{ij}(n)| \times \left| \frac{1-e^{-a_i(n-k_m)}}{a_i(n-k_m)} - \frac{1-e^{-a_i(n)}}{a_i(n)} \right| \\
&\quad \to \infty, \qquad \text{as } m \to \infty.
\end{aligned} \tag{8}$$

From the above analysis, we conclude that $D'_{ij}s$ are almost automorphic sequences. By similar analysis, it is not difficult to show that the sequences E_{ij} and F_i are also almost automorphic.

Lemma 2 *Under the assumptions* $(\mathbf{A_1})$, $(\mathbf{A_2})$, *every solution of (4) is bounded.*

Proof One can easily observe that the relation

$$C_i(n)x_i(n) - R_i \le x_i(n+1) \le C_i(n)x_i(n) + R_i,$$

where $R_i = \sum_{j=1}^{m} D^*_{ij} + F^*_i$ holds. Consider the following difference equations

$$\bar{x}_i(n+1) = C_i(n)\bar{x}_i(n) + R_i,$$

where $\bar{x}_i(0) = x_i(0)$. Using induction we have

$$\bar{x}_i(n) = \prod_{k=1}^{n} C_i(k)\bar{x}_i(0) + R_i\Big(\sum_{l=1}^{n-1}\prod_{k=1}^{l} C_i(k) + 1\Big)$$
$$\leq e^{-na_i^*}\bar{x}_i(0) + R_i(\sum_{l=1}^{n-1} e^{-la_i^*} + 1)$$
$$\leq |\bar{x}_i(0)| + \frac{R_i}{1 - e^{-a_i^*}}. \quad (9)$$

One can easily observe that $x_i(n) \leq \bar{x}_i(n)$. Now using the difference equation

$$\tilde{x}_i(n+1) = C_i(n)\tilde{x}_i(n) - R_i$$

and doing the similar calculation we get

$$\tilde{x}_i(n) \geq -|\bar{x}_i(0)| - \frac{R_i}{1 - e^{-a_i^*}}.$$

Combining the above two, we get the following estimate

$$-|x_i(0)| - \frac{R_i}{1 - e^{-a_i^*}} \leq x_i(n) \leq |x_i(0)| + \frac{R_i}{1 - e^{-a_i^*}}.$$

Thus x_i are bounded.

Now consider the following difference equations

$$x_i(n+1) = C_i(n)x_i(n) + F_i(n). \quad (10)$$

Lemma 3 *Under assumption* ($\mathbf{A_1}$), *there exists a almost automorphic sequence solution of (10).*

Proof Using the induction argument, one obtain

$$x_i(n+1) = \prod_{k=0}^{n} C_i(k)x_i(0) + \sum_{l=0}^{n}\prod_{k=n-l+1}^{n} C_i(k)F_i(n-l)$$
$$= e^{-\sum_{k=0}^{n} a_i(k)}x_i(0) + \sum_{l=0}^{n} I_i(n-l)\frac{1 - e^{-a_i(n-l)}}{a_i(n-l)}e^{-\sum_{k=n-l+1}^{n} a_i(k)}.$$

Consider the sequence

$$\hat{x}_i(n) = \sum_{l=0}^{\infty} I_i(n-l)\frac{1 - e^{-a_i(n-l)}}{a_i(n-l)}e^{-\sum_{k=n-l+1}^{n} a_i(k)}.$$

Since

$$|\hat{x}_i(n)| \leq \sum_{l=0}^{\infty} I_i^* \frac{1-e^{-a_i^*}}{a_i^*} e^{-(l-1)a_i^*} = \sum_{l=0}^{\infty} e^{a_i^*} \frac{1-e^{-a_i^*}}{a_i^*} e^{-la_i^*} \leq \frac{I_i^* e^{a_i^*}}{a_i^*}.$$

Thus the sequence $\hat{x}_i(n)$ is well defined. It is easy to verify that

$$\hat{x}_i(n+1) = C_i(n)\hat{x}_i(n) + F_i(n).$$

Hence the sequence $\hat{x}_i = \{\hat{x}_i(n)\}$ is bounded. Now define

$$\hat{x}_i(n) = \sum_{l=0}^{\infty} I_i(n-l) \frac{1-e^{-a_i(n-l)}}{a_i(n-l)} e^{-\sum_{k=n-l+1}^{n} a_i(k)}$$

and

$$\hat{y}_i(n) = \sum_{l=0}^{\infty} \bar{I}_i(n-l) \frac{1-e^{-a_i(n-l)}}{a_i(n-l)} e^{-\sum_{k=n-l+1}^{n} a_i(k)}.$$

For any sequence k_l there exists a sequence k_m such that

$$\begin{aligned}
&|\hat{x}_i(n+k_m) - \hat{y}_i(n)| \\
&= \Big| \sum_{l=0}^{\infty} I_i(n+k_m-l) \frac{1-e^{-a_i(n+k_m-l)}}{a_i(n+k_m-l)} e^{-\sum_{k=n-l+1}^{n} a_i(k+k_m)} \\
&\quad - \sum_{l=0}^{\infty} \bar{I}_i(n-l) \frac{1-e^{-a_i(n-l)}}{a_i(n-l)} e^{-\sum_{k=n-l+1}^{n} a_i(k)} \Big| \\
&\leq \sum_{l=0}^{\infty} \Big(|I_i(n+k_m-l) - \bar{I}_i(n-l)| \frac{1-e^{-a_i(n+k_m-l)}}{|a_i(n+k_m-l)|} e^{-\sum_{k=n-l+1}^{n} |a_i(k+k_m)|} \\
&\quad + \Big|\bar{I}_i(n-l)\Big| \times \Big| \frac{1-e^{-a_i(n+k_m-l)}}{a_i(n+k_m-l)} e^{-\sum_{k=n-l+1}^{n} a_i(k+k_m)} \\
&\quad - \frac{1-e^{-a_i(n-l)}}{a_i(n-l)} e^{-\sum_{k=n-l+1}^{n} a_i(k)} \Big| \Big) \\
&\leq \epsilon \frac{e^{a_i^*}}{a_i^*} + \epsilon \frac{I_i^* e^{a_i^*}}{a_i^{*2}} \leq \frac{e^{a_i^*}}{a_i^*}\epsilon + \frac{I_i^* e^{a_i^*}}{a_i^{*2}}\epsilon. \qquad (11)
\end{aligned}$$

From the above calculations, we obtain

$$\hat{x}_i(n+k_m) \to \hat{y}_i(n) \quad m \to \infty.$$

Similarly one can show that

$$\hat{y}_i(n - k_m) \to \hat{x}_i(n) \quad m \to \infty.$$

Hence $\hat{x}_i$ are almost automorphic sequences.

Theorem 1 *If assumptions* $(\mathbf{A_1})$, $(\mathbf{A_2})$ *hold, then there exists a unique almost automorphic sequence solution of (4) which is globally attractive, provided*

$$\max_{1\le i\le m}\left\{C_i^* + \sum_{j=1}^{m}(D_{ij}^* + E_{ij}^*)L_j\right\} < 1.$$

Proof Denote a metric $d : AAS \times AAS \to \mathbb{R}^+$, by

$$d(x, y) = \sup_{n\in\mathbb{Z}} \max_{1\le i\le m} |x_i(n) - y_i(n)|.$$

Now define a mapping $F : AAS \to AAS$ by $Fx = y$, where

$$Fx = (F_1x, F_2x, \dots, F_mx)^T$$

such that $F_ix = y_i$ and $y_i = \{y_i(n)\}$. Define

$$y_i(n+1) = C_i(n)x_i(n) + \sum_{j=1}^{m}\Big(D_{ij}(n)f_j(x_j(n)) + E_{ij}(n)f_j(x_j(n-\tau))\Big) + F_i(n),$$

where $\hat{x}_i$ are almost automorphic sequence solution of (10). Using Lemma 1 and assumption (A_2), we see that F maps almost automorphic sequences into almost automorphic sequences. Now denote

$$\max_{1\le i\le m}\{C_i^* + \sum_{j=1}^{m}(D_{ij}^* + E_{ij}^*)L_j\} = r < 1.$$

For $x, y \in AAS$, we have

$$\begin{aligned}\|Fx - Fy\| &= \sup_{n\in\mathbb{Z}} \max_{1\le i\le m} \sum_{j=1}^{m} \Big|[(D_{ij}(n)(f_j(x_j(n)) - f_j(y_j(n))) \\ &\quad + (E_{ij}(n)(f_j(x_j(n-\tau)) - f_j(y_j(n-\tau)))]\Big| \\ &\le \sup_{n\in\mathbb{Z}} \max_{1\le i\le m} \sum_{j=1}^{m} D_{ij}^* L_j |x_j(n) - y_j(n)|\end{aligned}$$

$$
\begin{aligned}
&+ \sup_{n\in\mathbb{Z}} \max_{1\le i\le m} \sum_{j=1}^{m} E_{ij}^{*} L_j |x_j(n-\tau) - y_j(n-\tau)| \\
&\le \max_{1\le i\le m} \{C_i^{*} + \sum_{j=1}^{m} (D_{ij}^{*} + E_{ij}^{*}) L_j\} \|x-y\| \\
&\le r\|x-y\|. \qquad (12)
\end{aligned}
$$

Hence F is a contraction. It follows that Eq. (4) has a unique almost automorphic sequence x.

Let y be any sequence satisfying Eq. (4). Consider $Q(n) = x(n) - y(n)$, then we get

$$
\begin{aligned}
Q_i(n+1) = C_i(n)Q_i(n) + \sum_{j=1}^{m} \Big(& D_{ij}(n)(f_j(x_j(n)) - f_j(y_j(n))) \\
& + E_{ij}(n)(f_j(x_j(n-\tau)) - f_j(y_j(n-\tau))) \Big). \qquad (13)
\end{aligned}
$$

Taking modulus of both sides one has

$$
|Q_i(n+1)| \le C_i^{*}|Q_i(n)| + \sum_{j=1}^{m} D_{ij}^{*} L_j |Q_j(n)| + \sum_{j=1}^{m} E_{ij}^{*} L_j |Q_j(n-\tau)|.
$$

Defining $Q(n) = \max_{1\le i\le m} |Q_i(n)|$, we have

$$
|Q(n+1)| \le C_i^{*}|Q(n)| + \sum_{j=1}^{m} (D_{ij}^{*} L_j Q(n) + E_{ij}^{*} L_j Q(n)) \le rQ(n). \qquad (14)
$$

By induction we have

$$
Q(n) \le r^n Q(0).
$$

Hence

$$
|x_i(n) - y_i(n)| \to 0 \quad as\ n \to \infty.
$$

Thus x is a unique globally attractive almost automorphic sequence solution of (4).

Acknowledgments This work is supported by the project “ITM/SG/SAB/001”.

References

1. Bohr, H.: Zur Theorie der fastperiodischen Funktionen I. Acta Math. **45**, 29–127 (1925)
2. Bochner, S.: Continuous mappings of almost automorphic and almost periodic functions. Rev. Mod. Phys. **52**, 907–910 (1964)
3. Abbas, S.: Pseudo almost periodic sequence solutions of discrete time cellular neural networks. Nonlinear Analysis Model. Control **14**(3), 283–301 (2009)
4. Abbas, S.: Weighted pseudo almost automorphic sequences and their applications. Elect. J. Differ. Equ. **2010**(121), 1–14 (2010)
5. Abbas, S., Xia, Y.: Existence and attractivity of k-almost automorphic sequence solution of a model of cellular neural networks with delay. Acta Math. Scientia **33**(1), 290–302 (2013)
6. Abbas, S.: Existence and attractivity of k-pseudo almost automorphic sequence solution of a model of bidirectional neural networks. Acta. Appl. Math. **119**(1), 57–74 (2012)
7. Araya, D., Castro, R., Lizama, C., Almost automorphic solutions of difference equations. Adv. Differ. Equ. 2009, Art. ID 591380, 15 pp.
8. Cao, J.: New results concerning exponential stability and periodic solutions of delayed cellular neural networks. Phys. Lett. A. **307**(2–3), 136–147 (2003)
9. Huang, Z., Wang, X., Gao, F.: The existence and global attractivity of almost periodic sequence solution of discrete time neural networks. Phys. Lett. A **350**(3–4), 182–191 (2006)
10. Abbas, S.: Pseudo almost automorphic solutions of fractional order neutral differential equation. Semigroup Forum **81**(3), 393–404 (2010)
11. Abbas, S.: Pseudo almost automorphic solutions of some nonlinear integro-differential equations. Comput. Math. Appl. **62**(5), 2259–2272 (2011)
12. NGuerekata, G.M.: Topics in Almost Automorphy. Springer, New York (2005)
13. Hopfield, J.J.: Neural networks and physical systems with emergent collect computational abilities. Proc. Natl. Acad. Sci. **79**, 2554–2558 (1982)
14. Cao, J., Zhou, D.: Stability analysis of delayed cellular neural networks. Neural Netw. **11**(9), 1601–1605 (1998)
15. Chen, T.: Global exponential stability of delayed hopfield neural networks. Neural Netw. **14**(8), 977–980 (2001)
16. Mohamad, S., Gopalsamy, K.: Exponential stability of continuous-time and discrete-time cellular neural networks with delays. Appl. Math. Comp. **135**(1), 17–38 (2003)
17. Stewart, I.: Warning-handle with care. Nature **355**, 16–17 (1992)

Advances in Periodic Difference Equations with Open Problems

Ziyad AlSharawi, Jose S. Cánovas and Antonio Linero

Abstract In this paper, we review some recent results on the dynamics of semi-dynamical systems generated by the iteration of a periodic sequence of continuous maps. In particular, we state several open problems focused on the structure of periodic orbits, forcing between periodic orbits, sharing periodic orbits, folding and unfolding periodic systems, and on applications of periodic systems.

1 Introduction

Let $C(I)$ denote the set of continuous maps $f : I \to I$ where I is a compact subinterval of the real line. We consider $f_0, \ldots, f_{p-1} \in C(I)$. These maps generate a semi-dynamical system [33], which we denote by $(I, [f_0, \ldots, f_{p-1}])$. For any $x \in I$, the orbit through x is denoted by $\mathrm{Orb}(x, [f_0, \ldots, f_{p-1}])$ and given by the solution of the non-autonomous difference equation

$$\begin{cases} x_0 = x, \\ x_{n+1} = f_{n \bmod \mathrm{p}}(x_n). \end{cases} \tag{1}$$

The number p is called the period of the system and it is always considered to be minimal.

J.S. Cánovas (✉)
Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena, Paseo de Alfonso XIII, 30203 Cartagena, Murcia, Spain
e-mail: jose.canovas@upct.es

Z. AlSharawi
Department of Mathematics and Statistics, Sultan Qaboos University, 36 PC 123 Al-Khod, Muscat, Sultanate of Oman
e-mail: alsha1zm@alsharawi.info

A. Linero
Departamento de Matemáticas, Universidad de Murcia,Campus de Espinardo, 30100 Murcia, Spain
e-mail: lineroba@um.es

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_6

Periodic difference equations have been studied by several authors recently (see for instance [6, 8, 9, 11, 15, 16, 23, 28–30, 33]. The interest for studying periodic discrete systems is motivated by applications in population dynamics, the (see e.g. [27] and [38]) and economic dynamics (see the so-called duopolies (see e.g. [44] and [47]). When one observes the orbit (x_n) of a point $x \in I$, it is easy to note that the subsequences $(x_{pn}), (x_{pn+1}), \dots, (x_{pn+p-1})$ are respectively the orbits of the initial points

$$x_0, \quad f_0(x_0), \quad \dots \quad , (f_{p-2} \circ \dots \circ f_0)(x_0)$$

under the iteration of the individual maps

$$F_0, \quad F_1, \quad F_2, \quad \dots \quad , F_{p-1},$$

where $F_j = f_{(p-1+j) \bmod p} \circ \dots \circ f_{j+1} \circ f_j$ for $j = 0, 1, \dots p-1$. Therefore, one might expect that the dynamics can be completely given by the dynamics of the above individual maps. Indeed, this is true for some characteristics of the dynamical system. For instance, the topological entropy of $[f_0, \dots, f_{p-1}]$ can be computed by means of the topological entropy of $f_{p-1} \circ \dots \circ f_0$ (see [40]). On the other hand, the ω-limit set $\omega(x, [f_0, \dots, f_{p-1}])$, which is the set of limit points of the orbit with initial condition x, can be obtained from the equality

$$\begin{aligned} \omega(x, [f_0, \dots, f_p]) &= \omega(x, f_{p-1} \circ \dots \circ f_0) \cup \omega(f_0(x), f_0 \circ f_{p-1} \circ \dots \circ f_1) \cup \dots \\ &\quad \dots \cup \omega((f_{p-2} \circ \dots \circ f_0)(x), f_{p-2} \circ \dots \circ f_0 \circ f_{p-1}) \\ &= \omega(x, F_0) \cup \omega(f_0(x), F_1) \cup \dots \cup \omega((f_{p-2} \circ \dots \circ f_0)(x), F_{p-1}) \end{aligned}$$

where each $\omega(z_j, F_j)$ is meant the set of limit points of the orbit of $z_j = (f_{j-1} \circ \dots \circ f_0)(x)$ under the interval map F_j, $j = 0, 1, \dots, p-1$, (here, $z_0 = x$).

The aim of this paper is to show that, even when many dynamical properties can be studied by the folded dynamical systems, there are several open problems that deserve investigation. The paper is organized in sections and each section covers a topic that includes some proposed open problems.

In Sect. 2, we deal with the set of periods $\mathrm{Per}[f_0, \dots, f_{p-1}]$ of periodic non-autonomous systems. In Sect. 3 we analyze how this set can be altered by the effect of folding some maps of the periodic non-autonomous system in order to obtain a new system, of possibly shorter period. After this, we present in Sect. 4 the question of studying the resulting period when we combine strings of two given periodic sequences. Another problem related to periodic orbits appears in Sect. 5: it is an open problem to determine whether or not the intersection of the sets of periodic points of two commuting interval maps is empty. Section 6 is devoted to the Parrondo's paradox. Finally, we present some interesting applications related with the dynamics of population models described by periodic non-autonomous systems.

2 Periodic Orbits in Periodic Non-autonomous Systems

When Eq. (1) is composed of one map, say f, an orbit $\mathrm{Orb}(x, f) = (x_n)$ is said to be periodic if there is $q \in \mathbb{N} := \{1, 2, \ldots\}$ such that $x_{n+q} = x_n$ for all $n \geq 0$. The smallest number q satisfying this condition is called the period or order, and denoted by $\mathrm{ord}_f(x)$. In the case of discrete dynamical systems on the interval I, the well–known Sharkovsky's theorem characterizes the set of periods of f, denoted $\mathrm{Per}(f)$. More precisely, we consider the following order in the set of natural numbers $\mathbb{N}$.

$$3 >_s 5 >_s 7 >_s \cdots >_s 2 \cdot 3 >_s 2 \cdot 5 >_s 2 \cdot 7 >_s \cdots$$
$$2^n \cdot 3 >_s 2^n \cdot 5 >_s 2^n \cdot 7 >_s \cdots >_s 2^{n+1} >_s 2^n >_s \cdots >_s 2 >_s 1.$$

For $n \in \mathbb{N} \cup \{2^\infty\}$, define $\mathscr{S}(n) = \{m \in \mathbb{N} : n >_s m\} \cup \{n\}$ and $\mathscr{S}(2^\infty) = \{2^n : n \in \mathbb{N} \cup \{0\}\}$. Sharkovsky's theorem states that if f has a periodic orbit (periodic sequence) of period n, then it has periodic points (periodic sequences) of period $m \in \mathscr{S}(n)$. Moreover, for any $n \in \mathbb{N} \cup \{2^\infty\}$ there is $f \in C(I)$ such that $\mathrm{Per}(f) = \mathscr{S}(n)$ (see [52] or [32] for a recent proof of Sharkovsky's theorem).

In the case of periodic non-autonomous systems, a generalization of Sharkovsky's theorem is given in [11]. For a fixed $p \in \mathbb{N}$, consider a p-periodic system defined by the maps $[f_0, \ldots, f_{p-1}]$. For $q \in \mathbb{N}$, define the clusters

$$\mathscr{A}_{p,q} = \{n : \mathrm{lcm}(n, p) = q \cdot p\} = \{n : q = \frac{n}{\gcd(n, p)}\}.$$

Notice that $p \cdot q \in \mathscr{A}_{p,q}$. Now, define the equivalence relation "$\sim_p$" on $\mathbb{N}$ by stating that $n \sim_p m$, $n, m \in \mathbb{N}$, if and only if n and m belong to the same set $\mathscr{A}_{p,q}$ for some $q \in \mathbb{N}$. If we denote any equivalence class $\mathscr{A}_{p,q}$ by $[q]$, we define the order on $\mathbb{N}/\sim_p$ by $[n] >_s [m]$ if and only if $n >_s m$. Now, $[q] \in \mathrm{Per}([f_0, \ldots, f_{p-1}])/\sim_p$ denotes that $\mathscr{A}_{p,q} \cap \mathrm{Per}([f_0, \ldots, f_{p-1}])$ is nonempty. The generalization in [11] shows if $[n] \in \mathrm{Per}([f_0, \ldots, f_{p-1}])/\sim_p$, then for any $[m] \in \mathbb{N}/\sim_p$ such that $[n] >_s [m]$, we have $[m] \in \mathrm{Per}([f_0, \ldots, f_{p-1}])/\sim_p$. The proof of this result is based on two facts: Sharkovsky's theorem and the fact that if $m \in \mathrm{Per}([f_0, \ldots, f_{p-1}])$ and $m \in [q]$, then $q \in \mathrm{Per}(f_{p-1} \circ \ldots \circ f_0)$.

When $p = 2$, another approach was used in [23] to characterize the structure of the set of periods $\mathrm{Per}([f_0, f_1])$. More precisely, set

$$\mathbb{N}^* := \mathbb{N} \setminus (\{2n - 1 : n \in \mathbb{N}\} \cup \{2\}).$$

The following result is given in [23].

Theorem 1 *Each of the following holds true for a 2-periodic system:*

(a) *If* $[f_0, f_1]$ *has a periodic orbit of period* $n \in \mathbb{N}^* \cup \{2^\infty\}$, *then* $\mathscr{S}(n) \setminus \{1, 2\} \subset \mathrm{Per}[f_0, f_1]$.
(b) *If* $2n + 1 \in \mathrm{Per}[f_0, f_1]$, $n \geq 1$, *then* $\mathscr{S}(2 \cdot 3) \setminus \{1\} \subset \mathrm{Per}[f_0, f_1]$.

(c) *There is a* 2*-periodic system* $[f_0, f_1]$ *such that* $\mathrm{Per}([f_0, f_1])$ *is* $\{1\}$*,* $\{2\}$ *or* $\{1, 2\}$.

(d) *For any* $n \in \mathbb{N}^* \cup \{2^\infty\}$*, there is a* 2*-periodic system* $[f_0, f_1]$ *such that one of the following is satisfied.*

d.1. $\mathrm{Per}([f_0, f_1]) = \mathscr{S}(n)$.
d.2. $\mathrm{Per}([f_0, f_1]) = \mathscr{S}(n)\setminus\{1\}$.
d.3. $\mathrm{Per}([f_0, f_1]) = \mathscr{S}(n)\setminus\{2\}$.

(e) *For any subset of odd numbers* $\mathrm{Imp} \subseteq \{2n+1 : n \in \mathbb{N}\}$ *there is a* 2*-periodic system* $[f_0, f_1]$ *such that one of the following is satisfied.*

e.1. $\mathrm{Per}([f_0, f_1]) = \mathrm{Imp} \cup (\mathscr{S}(2 \cdot 3)\setminus\{1\})$.
e.2. $\mathrm{Per}([f_0, f_1]) = \mathrm{Imp} \cup \mathscr{S}(2 \cdot 3)$.

Notice that the case $\mathrm{Per}([f_0, f_1]) = \mathrm{Imp} \cup (\mathscr{S}(2 \cdot 3)\setminus\{2\})$ is not allowed, that is, if $2n+1 \in \mathrm{Per}([f_0, f_1])$ for some $n \in \mathbb{N}$, then automatically $2 \in \mathrm{Per}([f_0, f_1])$. In addition, for $n \in \mathbb{N}^* \cup \{2^\infty\}$, $n \neq 2 \cdot 3$, there are no continuous maps $f_0, f_1 \in C(I)$ such that $\mathrm{Per}([f_0, f_1]) = \mathrm{Imp} \cup (\mathscr{S}(n)\setminus\{1\})$ or $\mathrm{Per}([f_0, f_1]) = \mathrm{Imp} \cup (\mathscr{S}(n)\setminus\{2\})$ or $\mathrm{Per}([f_0, f_1]) = \mathrm{Imp} \cup \mathscr{S}(n)$.

The following frame summarizes the forcing (where $n >_2 m$ is meant that the presence of a period n in the alternated system forces the existence of periodic sequences having order m):

$$\{2 \cdot n + 1 : n \in \mathbb{N}\} >_2 2 \cdot 3 >_2 2 \cdot 5 >_2 2 \cdot 7 >_2 \dots$$
$$2^n \cdot 3 >_2 2^n \cdot 5 >_2 2^n \cdot 7 >_2 \dots >_2 2^n >_2 \dots >_2 2^2 >_2 (1 \text{ or/and } 2).$$

The generalization of Sharkovsky's theorem given by AlSharawi et al. in [11] lacks the details about the forcing relationship within each equivalence class $[q] = \mathscr{A}_{p,q}$. On the other hand, the result of Cánovas and Linero in [23] gives the exact forcing between cycles when $p = 2$, but lacks the generality for $p > 2$. Since the results in [11] and [21], several attempts have been made to give the exact forcing within each equivalence class $[q]$ [6, 8, 15]. Although progress has been made in special cases, the general case is still open, which motivates our first open problem.

Open Problem 1 *Extend Theorem* 1 *to periodic sequences of maps of arbitrary period, i.e., characterize the set of periods* $\mathrm{Per}([f_0, \dots, f_{p-1}])$ *for any positive integer* p*.*

In each equivalence class or cluster $[q] = \mathscr{A}_{p,q}$, there is one period, namely pq, that does not depend on the intersection between the maps $f_0, f_1, \dots, f_{p-1}$, while the other periods need certain intersections between the maps [8, 15]. This observation leads to dividing the periods into the ones that are generic properties of the intersections, and the ones that are generic properties of the iterations. Indeed, establishing a connection between those two sets was one of the objectives in [6, 8, 12]. Examples were constructed [6] to show that the existence of $m \in \mathrm{Per}([f_0, \dots, f_{p-1}])$ for some $m \in \mathscr{A}_{p,q}$, where $q > 1$ is a power of 2, does not guarantee that $pq \in \text{Per}([f_0, \ldots, f_{p-1}])$. However, the following problem is still open.

Open Problem 2 *Suppose $m \in \text{Per}([f_0, \ldots, f_{p-1}])$ for some $m \in \mathscr{A}_{p,q}$, where $q > 1$ is an odd number. Prove that $pq \in \text{Per}([f_0, \ldots, f_{p-1}])$.*

It is well known that Sharkovsky's theorem works under certain modifications for other one-dimensional spaces, as the circle -with some modifications as a consequence of the degree of a circle map-, even for classes of n-dimensional continuous maps, for instance the so-called triangular maps $G(x_1, x_2, \ldots, x_n) = (g_1(x_1), g_2(x_1, x_2), \ldots, g_n(x_1, x_2, \ldots, x_n))$ (see [5] for more details). To this respect:

Open Problem 3 *Extend Theorem 1 to periodic sequences of two continuous circle maps, that is, characterize the set of periods $\text{Per}([f_0, f_1])$, where $f_0, f_1 : \mathbb{S}^1 \to \mathbb{S}^1$ are continuous.*

3 Folding in Periodic Systems

For a p-periodic system $[f_0, \ldots, f_{p-1}]$, it is possible to fold some of the maps to obtain a system of possibly shorter period. For instance, an obvious situation is the p-fold map $F_0 := f_{p-1} \circ f_{p-2} \circ \cdots \circ f_0$, which changes the p-periodic non-autonomous system into an autonomous system. In general, for any $1 \leq k \leq p-1$, we can fold the maps

$$f_{k-1} \circ \cdots \circ f_0 =: F_0, \quad f_{2k-1} \circ \cdots \circ f_k =: F_1, \quad f_{3k-1} \circ \cdots \circ f_{2k} =: F_2, \cdots \quad (2)$$

to form another periodic system $[F_0, F_1, \ldots, F_{\frac{p}{\gcd(p,k)}}]$. Some caution must be made here about the period of the new periodic system. It may not be $\frac{p}{\gcd(p,k)}$. Indeed, take the 4-periodic system $[f, f^{-1}, f^{-1}, f]$ and $k = 2$, then the folded system $[F_0, F_1]$ is 1-periodic rather than 2-periodic. The notion of folding was introduced in [6]. The scenario of having $[F_0, F_1, \ldots, F_{\frac{p}{\gcd(p,k)}}]$ with a period less than $\frac{p}{\gcd(p,k)}$ is called degenerate scenario and has been avoided. The case when k is a divisor of p is studied in [12] and connections between the cycles of the folded and unfolded systems have been established. However, for general k, the connection between the folded and unfolded systems still open for further investigation, which motives the next open problem.

Open Problem 4 *Consider the p-periodic system $[f_0, \ldots, f_{p-1}]$, and let $1 \leq k \leq p$. Consider the maps $F_0, F_1, \ldots, F_{\frac{p}{\gcd(p,k)}}$ as defined in (2). What is the relationship between $Per([f_0, \ldots, f_{p-1}])$ and $Per([F_0, F_1, \ldots, F_{\frac{p}{\gcd(p,k)}}])$?*

Periodic difference equations with delay of the form $y_{n+1} = g_{n \bmod p}(y_{n-(k-1)})$ have been studied by AlSharawi et al. in [10], and a characterization of the periodic structures was given. To visualize the orbits in this case, consider $p = 6$ and $k = 4$, then orbits of $y_{n+1} = g_{n \bmod 6}(y_{n-3})$ can be written in matrix form as

$$\begin{matrix} y_{-3} & y_{-2} & y_{-1} & y_0 \\ g_0 & g_1 & g_2 & g_3 \\ g_4 & g_5 & g_0 & g_1 \\ g_2 & g_3 & g_4 & g_5. \end{matrix}$$

From the columns of this matrix, observe that each column gives a periodic system of period $\frac{p}{\gcd(p,k)}$, which can be assumed to be the minimal period. This observation motivates the following problem.

Open Problem 5 *Suppose there is a p-periodic system* $[f_0, \ldots, f_{p-1}]$, *which is unknown to us, but we know one of its folded systems* $[F_0, F_1, \ldots, F_{\frac{p}{\gcd(p,k)}}]$. *What kind of similarity (if any) in periodic structure do we have between the unfolded p-periodic system* $[f_0, \ldots, f_{p-1}]$ *and the periodic system with delay* $x_{n+1} = F_n(x_{n-(k-1)})$?

4 Merging Periodic Sequences

In the unfolding process of a $\frac{p}{k}$−periodic system $[F_0, F_1, \ldots, F_{\frac{p}{k}}]$, we find ourselves dealing with sequences that are merged in a certain way [12]. Therefore, we find the notion of merging two periodic sequences to be related to the topic of the previous section, and therefore, we find it is worth addressing here. Suppose we have two periodic sequences $\{a_n\}$ and $\{b_n\}$ of periods q_1 and q_2, respectively. The two sequences can be thought of as two periodic signals or codes coming out of two machines. After each string of length k_1 produced by the first machine (a k_1 string of $\{a_n\}$), the second machine releases a k_2 string (a k_2 string of $\{b_n\}$). The obtained signal has the structure $[a_n, b_n] :=$

$$\overbrace{a_1, a_2, \ldots, a_{k_1}}^{k_1 \text{ string}}, \overbrace{b_1, b_2, \ldots, b_{k_2}}^{k_2 \text{ string}}, \overbrace{a_{k_1+1}, a_{k_1+2}, \ldots, a_{2k_1}}^{k_1 \text{ string}}, b_{k_2+1}, \ldots \tag{3}$$

Before we proceed, we clarify the notion by an illustrative example.

Example 1 For $n \in \mathbb{N}$, consider $a_n = n \bmod 4$ and $b_n = 4 + (n \bmod 6)$. Thus, $\{a_n\}$ is periodic of period $q_1 := 4$ and $\{b_n\}$ is periodic of period $q_2 := 6$.

(i) If $k_1 = 2$ and $k_2 = 3$, then

$$[a_n, b_n] = \{0, 1, 4, 5, 6, 2, 3, 7, 8, 9, 0, 1, 4, 5, 6, \cdots\},$$

and the period of the formed sequence is 10.

(ii) If $k_1 = 2$ and $k_2 = 4$, then

$$[a_n, b_n] = \{0, 1, 4, 5, 6, 7, 2, 3, 8, 9, 4, 5, 0, 1, 6, 7, 8, 9,\\ 2, 3, 4, 5, 6, 7, 0, 1, 8, 9, 4, 5, 2, 3, 6, 7, 8, 9, 0, 1 \cdots\}$$

and the period of the formed sequence is 36.

For a better understanding of the structure of the formed sequence $[a_n, b_n]$, we write its elements in matrix form as follows:

$$\begin{array}{cccccccc} a_0 & a_1 & \cdots & a_{k_1-1} & b_0 & b_1 & \cdots & b_{k_2-1} \\ a_{k_1} & a_{k_1+1} & \cdots & a_{2k_1-1} & b_{k_2} & b_{k_2+1} & \cdots & b_{2k_2-1} \\ a_{2k_1} & a_{2k_1+1} & \cdots & a_{3k_1-1} & b_{2k_2} & b_{2k_2+1} & \cdots & b_{3k_2-1} \\ a_{3k_1} & a_{3k_1+1} & \cdots & a_{4k_1-1} & b_{3k_2} & b_{3k_2+1} & \cdots & b_{4k_2-1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \ddots \end{array} \tag{4}$$

Now, reading this matrix row by row from left to right gives the formed sequence $[a_n, b_n]$. It is obvious that we obtain a periodic sequence in each column. In some cases, it is easy to deduce the new period of the sequence obtained from the merging process, for instance when the merged sequences are disjoint (we establish here the result without proof):

Lemma 1 *For $n \in \mathbb{N}$, let $\{a_n\}$ and $\{b_n\}$ be two disjoint periodic sequences of periods q_1 and q_2, respectively. Also, let $1 \le k_1 \le q_1$ and $1 \le k_2 \le q_2$. The period of the sequence $[a_n, b_n]$ formed in (3) is of minimal period kq, where $k = k_1 + k_2$ and $q = \operatorname{lcm}\left(\frac{q_1}{\gcd(k_1, q_1)}, \frac{q_2}{\gcd(k_2, q_2)}\right)$.*

The condition to have the two sequences disjoint is a luxury that one may not have, which leads us to state the following general problem:

Open Problem 6 *For $n \in \mathbb{N}$, let $\{a_n\}$ and $\{b_n\}$ be two periodic sequences of periods q_1 and q_2, respectively. Find the minimal period of the sequence $[a_n, b_n]$ formed in (3).*

5 Commuting Maps and the Problem of Sharing Periodic Orbits

Denote by $\mathrm{Fix}(f)$ and $\mathrm{P}(f)$ the set of fixed and periodic points of a map $f \in C(I)$, respectively. We consider two maps $f_0, f_1 \in C(I)$ such that they commute, that is, $f_0 \circ f_1 = f_1 \circ f_0$.

In the fifties of the 20-th century, some authors posed independently the problem of proving whether two commuting continuous interval maps share fixed points. The problem is answered in affirmative for polynomials, as J. F. Ritt pointed out in [49].

Other cases, under restrictive conditions, have also a positive answer (for instance, we mention [31] or [53]). The question on whether $\mathrm{Fix}(f_0) \cap \mathrm{Fix}(f_1)$ is nonempty, that is, f_0 and f_1 have a common fixed point, was open for a long time [37]. Finally, Boyce [20] and Huneke [36] found simultaneously counterexamples which show that in general the answer is negative, there exist two continuous commuting interval maps which do not share any fixed point. The counterexamples constructed in [20, 36] are given by two maps f_0 and f_1 that share periodic points.

Since then, the research on this topic was concentrated in several directions. For instance, to extend the problem to other compact metric spaces or to particular classes of continuous maps [34, 39, 41]. The problem has been also posed in terms of sharing periodic points which are not necessarily fixed points (see [3, 55]). Then, it can be expected to raise the following question:

Open Problem 7 *Is it true that* $\mathrm{P}(f_0) \cap \mathrm{P}(f_1) \neq \emptyset$ *for commuting continuous interval maps* $f_0, f_1 \in C(I)$*?*

Fixed and periodic points are the strongest type of recurrence in dynamical systems. There are weaker notions of recurrence that contain the sets of fixed and periodic points. Namely, a point $x \in X$ is called *recurrent* if for any open neighborhood U of x there is an increasing sequence $\{n_i\}_{i=1}^{\infty}$ such that $f^{n_i}(x) \in U$. If the sequence $\{n_i\}_{i=1}^{\infty}$ has bounded gaps, the point is called *uniformly recurrent*. If $n_i = ki$ for some $k \in \mathbb{N}$ the point is called *almost periodic*. Denote by $\mathrm{Rec}(f)$, $\mathrm{UR}(f)$ and $\mathrm{AP}(f)$ the sets of recurrent, uniformly recurrent and almost periodic points. It is clear from the definitions that

$$\mathrm{Fix}(f) \subseteq \mathrm{P}(f) \subseteq \mathrm{AP}(f) \subseteq \mathrm{UR}(f) \subseteq \mathrm{Rec}(f).$$

Following the Sharkovsky's order of natural numbers, let $\mathscr{T}_1 = \{f \in C(I) : \mathrm{P}(f)$ is closed$\}$, $\mathscr{T}_2 = \{f \in C(I) : f$ has periodic points of period 2^n, $n \geq 0\}$ and $\mathscr{T}_3 = \{f \in C(I) : f$ has a periodic point which is not a power of two$\}$. The next result was proved in [22].

Theorem 2 *Assume* $f_0, f_1 \in C(I)$ *commute. Then*

(a) If $f_0 \in \mathscr{T}_1$*, then* $\mathrm{Fix}(f_0) \cap \mathrm{P}(f_1) \neq \emptyset$.
(b) If $f_0 \in \mathscr{T}_2$*, then* $\mathrm{Fix}(f_0) \cap \mathrm{AP}(f_1) \neq \emptyset$.
(c) If $f_0 \in \mathscr{T}_3$*, then* $\mathrm{Fix}(f_0) \cap \mathrm{UR}(f_1) \neq \emptyset$.

The above result proves Open Problem 7 for maps which are simple from the point of view of dynamics. Note that maps of type $\mathscr{T}_3$ have positive topological entropy, and therefore, they are chaotic in the sense of Li and Yorke. Chaotic maps in the sense of Li and Yorke may also exist in the family $\mathscr{T}_2$, but they cannot be found in $\mathscr{T}_1$, which contains the set of continuous interval maps with finite set of periods (see [18] or [22]). So, if one has to look for counterexamples for Open Problem 7, he/she should construct maps having both infinitely many periodic points. In addition, they cannot have a finite number of monotonicity pieces (see [20, 36]).

We finish this section by a problem that links Theorem 1 and Open Problem .

Open Problem 8 *Find* $\mathrm{Per}([f_0, f_1])$ *for commuting maps* $f_0, f_1 \in C(I)$. *More precisely, can the examples of Theorem 1 be constructed such that* f_0 *and* f_1 *commute?*

6 The Parrondo's Paradox

Parrondo's paradox [35] has become an active area of research in many applied sciences like Physics [46], Economy [56] or Biomathematics [57]. As a first approach, we can say that it appears when we alternate different games in a stochastic or deterministic way. Parrondo's paradox exists when the behavior of individual systems and the combined one are completely different. For discrete dynamical systems, the paradox was formulated in [4] by showing that the phenomenon "chaos + chaos = order" and "order + order = chaos" are possible when considering periodic combinations of 1-dimensional quadratic maps. Similar results have been obtained by Boyarsky and collaborators in the random combination of piecewise smooth maps [19]. On the other hand, it was shown in [21] that in some particular cases the paradox is not possible.

The dynamic Parrondo's paradox was studied in detail in [24] as follows. For a map $f \in C(I)$, $I = [0, 1]$, denote by $\mathscr{D}(f)$ the set of dynamic properties of f (for instance to have positive topological entropy or exhibiting chaos in the sense of Li and Yorke), and define $\mathscr{D}([f_0, f_1])$ similarly. Let $J = [a, b] \subseteq I = [0, 1]$ and denote by $\varphi_J : J \to I$ a linear map such that $\varphi_J(a) = 0$ and $\varphi_J(b) = 1$. Define $f_J : [0, 1] \to [0, 1]$ by $f_J(x) = \varphi_J^{-1} \circ f \circ \varphi_J(x)$ if $x \in J$, $f_J(0) = 0$, $f_J(1) = 1$, and linear on any connected component of $[0, 1] \setminus J$. A dynamic property $P \in \mathscr{D}(f)$ is an L-property if for any continuous map f and any compact subinterval $J \subseteq [0, 1]$, it is held that $P \in \mathscr{D}(f) \cap \mathscr{D}(f_J)$. The fact that if $f_0, f_1 \in C(I)$, then the dynamics of the sequence $[f_0, f_1]$ is complicated (or simple) if and only if the dynamics of $f_0 \circ f_1$ is complicated (or simple), jointly with L-properties are the key for analyzing the Parrondo's paradox as the following result shows [24].

Theorem 3 *Let* P_i, $i = 1, 2, 3$, *be L-properties. Then there are* $f_0, f_1 \in C(I)$ *such that* $P_1 \in \mathscr{D}(f_0)$, $P_2 \in \mathscr{D}(f_1)$ *and* $P_3 \in \mathscr{D}(f_0 \circ f_1)$.

In particular, we can construct maps such that f_0 and f_1 have a complicated (simple) L-property and $f_0 \circ f_1$ has not this property. $[P \in \mathscr{D}(f_0) \cap \mathscr{D}(f_1)$ and $P \notin \mathscr{D}(f_0 \circ f_1)]$. For instance, we consider the topological entropy (see e.g. [2] or [5] for definition and basic properties of topological entropy), which is a useful tool to decide whether a map has a complicated dynamics. From Theorem 3, we can construct two continuous interval maps, f_0 and f_1, with zero topological entropy (and hence simple) such that $f_0 \circ f_1$ has positive topological entropy (and therefore a complicated dynamics), because the properties zero topological entropy and positive topological entropy are L-properties. However, we must emphasize that, although Theorem 3 shows the existence of paradox in a general way for a very large list of dynamical properties, the constructions made for proving it cannot be done when we consider fixed maps like for instance a member of the logistic family $f_a(x) = ax(1-x)$, $1 \leq a \leq 4$ and $x \in [0, 1]$.

Open Problem 9 *State an analogous result to Theorem 3 when the maps f_0 and f_1 commute.*

Consider the well–known logistic family $f_a(x) = ax(1-x)$, $a \in [1, 4]$. More precisely, we consider two maps f_a and f_b and wonder about the existence of paradox for parameters a and b. In [26], the paradox existence is shown for several parameters values. However, although $f_a \circ f_b$ may exhibit the paradox, it is observed that several combinations like $f_a \circ f_b \circ f_a$ do not exhibit the paradox.

Open Problem 10 *Characterize the set of parameters $a, b \in [1, 4]$ such that any combination of maps f_a and f_b exhibit the Parrondo's paradox. Is this set nonempty?*

Denote the topological entropy of a continuous map f by $h(f)$. In [25], numerical simulations show that Parrondo's paradox cannot be exhibited by maps with positive topological entropy, that is, if $\min\{h(f_a), h(f_b)\} > 0$, then numerical simulations show that $h(f_a \circ f_b) > 0$, and therefore the Parrondo's paradox cannot be exhibited (see Fig. 1).

Open Problem 11 *In the logistic family, prove or disprove that* $\min\{h(f_a), h(f_b)\} > 0$ *implies that* $h(f_a \circ f_b) > 0$.

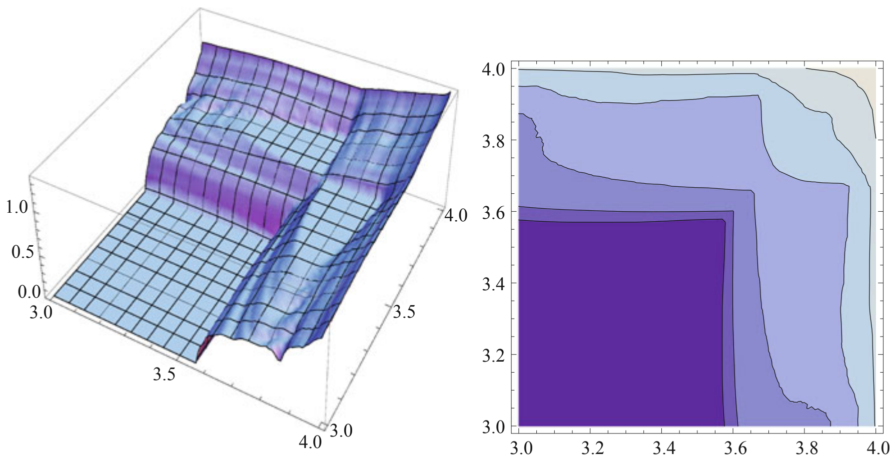

Fig. 1 On the *left* we show the topological entropy of F with accuracy 10^{-3}. On the *right*, we present the projection of topological entropy where stronger colors means lower topological entropy. The *darkest region* represents those parameter values for which the topological entropy is zero up to the prescribed accuracy

7 Applications

In population dynamics, one dimensional models usually take the form $x_{n+1} = x_n f(x_n)$, where x_n represents the density of a population at discrete time n and $f(t)$ is a function that reflects certain characteristics of the studied species. For instance, $f(t) = \frac{k\mu}{k+(\mu-1)t}$, $\mu > 1$ is used for the Beverton-Holt model [17], $f(t) = at(b-t)$ is used for the logistic model [45], and $f(t) = be^{-kt}$ is used for the Ricker model [48]. Forcing periodic harvesting or stocking in a deterministic environment leads to investigating the dynamics of models in the from

$$x_{n+1} = x_n f(x_n) \pm h_n, \tag{5}$$

where $\{h_n\}$ is a p-periodic sequence that represents harvesting or stocking quotas. See [7] and the references therein for more details and some open questions. We find Problem 3.1 in [7] to be suitable within the context of this paper.

Open Problem 12 *Consider Eq.* (5) *with stocking (i.e.* $+h_n$*) and assume this equation has a global attractor (like when* $f(t) = \frac{bt}{1+t}$ [14]*). Let* $\{\hat{h}_n\}$ *be a permutation of* $\{h_n\}$*. Define* x_{av} *and* $\hat{x}_{av}$ *to be the average of the global attractors associated with* $\{h_n\}$ *and* $\{\hat{h}_n\}$*, respectively. How does* x_{av} *relate to* $\hat{x}_{av}$*?*

Although it is tempting to believe that increasing constant yield harvesting in population models leads to a decline in the population, recent results show otherwise [50, 54] and the phenomenon is known as the hydra effect [1]. In fact, this notion led to a fertile area of research; see for instance [42, 43, 51] and the references therein. When we confine models in Eq. (5) to contest-competition models and force the harvesting to be constant yield harvesting (i.e., $h_n = h$ for all n), then the hydra effect is not known to take place. For instance, if we take the Beverton-Holt model with harvesting [13], then the global attractor is decreasing in h. However, the effect of periodic harvesting is not characterized yet, which motivates our next problem.

Open Problem 13 *Consider Eq.* (5) *with harvesting (i.e.* $-h_n$*). Investigate the effect of ordering the elements of the sequence* $\{h_n\}$ *on the population. On other words, assume that we have a periodic sequence of harvesting quotas but we have the freedom to permute its elements. Which permutation of* $\{h_n\}$ *plays on the advantage of the population in terms of the basin of attraction and in terms of the arithmetic average of the attractor?*

Acknowledgments J.S. Cánovas and A. Linero have been partially supported by the grants MTM2011-23221 from Ministerio de Ciencia e Innovación (Spain) and 08667/PI/08 from Programa de Generación de Conocimiento Científico de Excelencia de la Fundación Séneca, Agencia de Ciencia y Tecnología de la Comunidad Autónoma de la Región de Murcia (II PCTRM 2007–10).

References

1. Abrams, P.A.: When does greater mortality increase population size? The long story and diverse mechanisms underlying the hydra effect. Ecol. Lett. **12**, 462–474 (2009)
2. Adler, R.L., Konheim, A.G., McAndrew, M.H.: Topological entropy. Trans. Amer. Math. Soc. **114**, 309–319 (1965)
3. Alikhani-Koopaei, A.: On common fixed points and periodic points of commuting functions. Int. J. Math. Math. Sci. **21**, 269–276 (1998)
4. Almeida, J., Peralta-Salas, D., Romera, M.: Can two chaotic systems give rise to order? Phys. D **200**, 124–132 (2005)
5. Alsedá, L., Llibre, J., Misiurewicz, M.: Combinatorial dynamics and entropy in dimension one. World Scientific, Singapore (1993)
6. Al-Salman, A., AlSharawi, Z.: A new characterization of periodic oscillations in periodic difference equations. Chaos Solitons Fractals **44**, 921–928 (2011)
7. AlSharawi, Z.: Harvesting and stocking in contest competition models: open problems and conjectures, Pre-print
8. AlSharawi, Z.: Periodic orbits in periodic discrete dynamics. Comput. Math. Appl. **56**, 1966–1974 (2008)
9. AlSharawi, Z., Angelos, J.: On the periodic logistic equation. Appl. Math. Comput. **180**, 342–352 (2006)
10. AlSharawi, Z., Angelos, J., Elaydi, S.: Existence and stability of periodic orbits of periodic difference equations with delays. Int. J. Bifur. Chaos Appl. Sci. Engrg. **18**, 203–217 (2008)
11. AlSharawi, Z., Angelos, J., Elaydi, S., Rakesh, L.: An extension of Sharkovsky's theorem to periodic difference equations. J. Math. Anal. Appl. **316**, 128–141 (2006)
12. AlSharawi, Z., Cánovas, J. S., Linero, A.: Folding and unfolding in periodic difference equations, pre-print
13. AlSharawi, Z., Rhouma, M.: The Beverton-Holt model with periodic and conditional harvesting. J. Biol. Dyn. **3**, 463–478 (2009)
14. AlSharawi, Z., Rhouma, M.: The discrete Beverton-Holt model with periodic harvesting in a periodically fluctuating environment. Adv. Diff. Equ. (2010), Article ID 215875, doi:10.1155/2010/215875
15. Alves, J.F.: What we need to find out the periods of a periodic difference equation. J. Diff. Equ. Appl. **15**, 833–847 (2009)
16. Alves, J.F., Silva, L.: Periodic paths on nonautonomous graphs. Linear Algebra Appl. **437**, 1003–1015 (2012)
17. Beverton, R.J.H., Holt, S.J.: On the Dynamics of Exploited Fish Populations. The Blackburn Press, New Jersey (2004)
18. Block, L., Coppel, W.A.: Dynamics in One Dimension. Lecture Notes in Mathematics. Springer, Berlin (1992)
19. Boyarsky, A., Gora, P., Islam, M.S.: Randomly chosen chaotic maps can give rise to nearly ordered behavior. Phys. D **210**, 284–294 (2005)
20. Boyce, W.M.: Commuting functions with no common fixed point. Trans. Amer. Math. Soc. **137**, 77–92 (1969)
21. Cánovas, J.S.: Analyzing when the dynamic Parrondo's paradox is not possible. Int. J. Bifur. Chaos Appl. Sci. Engrg. **20**, 2975–2978 (2010)
22. Cánovas, J.S., Linero, A.: On the dynamics of composition of commuting interval maps. J. Math. Anal. Appl. **305**, 296–303 (2005)
23. Cánovas, J.S., Linero, A.: Periodic structure of alternating continuous interval maps. J. Difference Equ. Appl. **12**, 847–858 (2006)
24. Cánovas, J.S., Linero, A., Peralta-Salas, D.: Dynamic Parrondo's paradox. Phys. D **218**, 177–184 (2006)
25. Cánovas, J.S., Muñoz, M.: Computing topological entropy for periodic sequences of unimodal maps, pre-print

26. Cánovas, J.S., Muñoz, M.: Revisiting Parrondo's paradox for the logistic family. Fluct. Noise Lett. (2013). doi:10.1142/S0219477513500156
27. Cushing, J., Henson, S.: The effect of periodic habitat fluctuations on a nonlinear insect population model. J. Math. Biol. **36**, 201–226 (1997)
28. D'Aniello, E., Oliveira, H.: Pitchfork bifurcation for non-autonomous interval maps. J. Difference Equ. Appl. **15**, 291–302 (2009)
29. D'Aniello, E., Steele, T.H.: Stability in the family of ω-limit sets of alternating systems. J. Math. Anal. Appl. **389**, 1191–1203 (2012)
30. D'Aniello, E., Steele, T.H.: The ω-limit sets of alternating systems. J. Diff. Equ. Appl. **17**, 1793–1799 (2011)
31. DeMarr, R.: Common fixed points for commuting contraction mappings. Pacific J. Math. **13**, 1139–1141 (1963)
32. Du, B.S.: A simple proof of Sharkovsky's theorem. Amer. Math. Monthly **111**, 595–599 (2004)
33. Elaydi, S., Sacker, R.: Periodic difference equations, population biology and the Cushing-Henson conjectures. Math. Biosci. **201**, 195–207 (2006)
34. Grinč, M.: On common fixed points of commuting triangular maps. Bull. Polish Acad. Sci. Math. **47**, 61–67 (1999)
35. Harmer, G.P., Abbott, D.: Game theory: losing strategies can win by Parrondo's paradox. Nature **402**, 864 (1999)
36. Huneke, J.P.: On common fixed points of commuting functions on an interval. Trans. Amer. Math. Soc. **139**, 371–381 (1969)
37. Isbell, J.R.: Commuting mappings of trees. Bull. Amer. Math. Soc. **63**, 419 (1957)
38. Jillson, D.A.: Insect populations respond to fluctuating environments. Nature **288**, 699–700 (1980)
39. Jungck, G.: Common fixed points for commuting and compatible maps on compacta. Proc. Amer. Math. Soc. **103**, 977–983 (1988)
40. Kolyada, S., Snoha, L'.: Topological entropy of nonautononous dynamical systems. Random Comput. Dyn. **4**, 205–233 (1996)
41. Linero, A.: Common fixed points for commuting Cournot maps. Real Anal. Exchange **28**, 121–145 (2002)
42. Liz, E.: Complex dynamics of survival and extinction in simple population models with harvesting. Theor. Ecol. **3**, 209–221 (2010)
43. Liz, E., Ruiz-Herrera, A.: The hydra effect, bubbles, and chaos in a simple discrete population model with constant effort harvesting. J. Math. Biol. **65**, 997–1016 (2012)
44. Matsumoto, A., Nonaka, Y.: Statistical dynamics in a chaotic Cournot model with complementary goods. J. Econ. Behav. Organ. **61**, 769–783 (2006)
45. May, R.M.: Simple mathematical models with very complicated dynamics. Nature **261**, 459–467 (1976)
46. Parrondo, J.M.R., Harmer, G.P., Abbott, D.: New paradoxical games based on Brownian ratchets. Phys. Rev. Lett. **85**, 5226–5229 (2000)
47. Puu, T.: Chaos in duopoly pricing. Chaos Solitons Fractals **1**, 573–581 (1991)
48. Ricker, W.E.: Stock and recruitment. J. Fish. Res. Board Canada **11**, 559–623 (1954)
49. Ritt, J.F.: Permutable rational functions. Trans. Amer. Math. Soc. **25**, 399–448 (1923)
50. Schreiber, S.J.: Chaos and population disappearances in simple ecological models. J. Math. Biol. **42**, 239–260 (2001)
51. Seno, H.: A paradox in discrete single species population dynamics with harvesting/thinning. Math. Biosci. **214**, 63–69 (2008)
52. Sharkovsky, A.N.: Coexistence of cycles of a continuous transformation of a line into itself, Ukrain. Mat. Zh. 16, 61–71 (1964) (in Russian)
53. Shields, A.L.: On fixed points of commuting analytic functions. Proc. Amer. Math. Soc. **15**, 703–706 (1964)
54. Sinha, S., Parthasarathy, S.: Unusual dynamics of extinction in a simple ecological model. Proc. Natl. Acad. Sci. U S A **93**, 1504–1508 (1996)

55. Steele, T.H.: A note on periodic points and commuting functions, Real Anal. Exch. **24**, 781–790 (1998/9)
56. Spurgin, R., Tamarkin, M.: Switching investments can be a bad idea when Parrondo's paradox applies. J. Behav. Finance **6**, 15–18 (2005)
57. Wolf, D.M., Vazirani, V.V., Arkin, A.P.: Diversity in times of adversity: probabilistic strategies in microbial survival games. J. Theoret. Biol. **234**, 227–253 (2005)

An Evolutionary Beverton-Holt Model

J.M. Cushing

Abstract The classic Beverton-Holt (discrete logistic) difference equation, which arises in population dynamics, has a globally asymptotically stable equilibrium (for positive initial conditions) if its coefficients are constants. If the coefficients change in time, then the equation becomes nonautonomous and the asymptotic dynamics might not be as simple. One reason the coefficients can change in time is their evolution by natural selection. If the model coefficients are functions of a heritable phenotypic trait subject to natural selection then, by standard methods for modeling evolution, the model becomes a planar system of coupled difference equations, consisting of a Beverton-Holt type equation for the population dynamics and a difference equation for the dynamics of the mean phenotypic trait. We consider a case when the trait equation uncouples from the population dynamic equation and obtain criteria under which the evolutionary system has globally asymptotically stable equilibria or periodic solutions.

1 Introduction

The well-known difference equation

$$x_{t+1} = b\frac{1}{1+cx_t}x_t, \qquad b, c > 0 \tag{1}$$

arose historically in population dynamics as a discrete analog of logistic growth [6] (also see [7–10]). It has been used as a basic model in many studies in population, ecological and evolutionary dynamics in the same way that the logistic differential

J.M. Cushing (✉)
Department of Mathematics and The Interdisciplinary Program in Applied Mathematics, University of Arizona, 617 N. Santa Rita, Tucson, AZ 85721, USA
e-mail: cushing@math.arizona.edu

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_7

equation is used as a starting point for innumerable differential equation models in these fields. The equation has been used, for example, in the fisheries industry where it is known as the Beverton-Holt equation, a name that is now widely used for the equation.

The dynamics of (1) are well known. For $x_0 > 0$ the equilibrium $x^e = 0$ (which we will refer to as the *extinction equilibrium*) is globally asymptotically stable (GAS) if $b < 1$. If $b > 1$ then $x = 0$ is unstable and the *survival equilibrium* $x^e = (b-1)/c$ is GAS. Moreover, solution sequences are monotone (and hence the strong analogy with the logistic differential equation).

There are numerous biological reasons to consider the coefficients b and c in Eq. (1) not to be constants. For example, these parameters might change in time due to changing environmental conditions, to physiological cycles, etc. Their fluctuations might be stochastic or deterministically regular, even periodic (modeling seasonal, monthly, or daily environmental oscillations). These fluctuations in (1) give rise respectively to a stochastic, a nonautonomous, and a periodically forced difference equation. The mathematical literature on the latter case has, for both (1) and for more general scalar difference equations, grown considerably in the last decade.

Another biological reason for which parameters in population dynamic models might change is time is evolutionary adaptation. In this case, the coefficients b and/or c in (1) are assumed to be functions of a phenotypic trait or several phenotypic traits that are subject to Darwinian evolution through natural selection and hence that change in time. If a scalar v represents a quantified phenotypic trait (e.g. body size, age, color, metabolic rate, aggressiveness, etc.) whose value affects and determines the values of b and c experienced by an individual that inherits the trait v (or by a mutant or an invader with trait v), then we write $b = b(v)$ and $c = c(v)$. It is up to the modeler to assign specific properties to these two functions that reflect the biological situation of interest. It can be the case that these parameters are not only affected by the trait value v of the individual, but by the traits of other individuals in the population (for example, through competition for resources). One way to model this case (called frequency dependence) is to assume that b and/or c also depend on the mean trait u of the population and to write

$$b = b(v, u), \quad c = c(v, u).$$

A method for modeling the change in the mean trait u over time, as subject to evolution, is by means of the equations

$$x_{t+1} = b(v, u_t) \left. \frac{1}{1 + c(v, u_t) x_t} \right|_{v=u_t} x_t \tag{2}$$

$$u_{t+1} = u_t + \sigma^2 \left. \frac{\partial \ln r(x, v, u_t)}{\partial v} \right|_{v=u_t} \tag{3}$$

where

$$r(x, v, u) := b(v, u) \frac{1}{1 + c(v, u)x}$$

[1, 4, 5, 11]. The equation for u_t (the trait dynamics) states that change in the mean trait is proportional to the fitness gradient where here fitness is defined as $\ln r(x, v, u)$. The parameter σ^2 (the constant of proportionality in the assumed evolution law) has different biological interpretations that depend on the assumptions made in modeling evolution [1]. Generally, however, σ^2 is proportional to the variance of the trait in the population, which is assumed constant over time. In any case, σ^2 measures how fast evolution occurs, and we refer to it as the *speed of evolution*.

The model equations (2) constitute a planar system of difference equations in which the population dynamics of x_t and the evolutionary dynamics of u_t are in general coupled. A method known as adaptive dynamics uncouples the trait equation from the population dynamic equation by making the assumption that evolutionary and population dynamics occur on (infinitely) different time scales [1, 4]. In this paper, we consider a (fairly general) case in which the trait dynamics uncouple from the population dynamics without the necessity of this strong assumption about differing time scales.

We will assume that b depends only on v (i.e., only on the trait inherited by the individual and not on the traits of others in the population). We also assume that c, which is surrogate for intraspecific competition, is a function of the difference $v - u$; that is to say, the amount of competition felt by an individual depends on how different its trait is from that of others, as represented by the mean u. Letting R and R^+ denote the real numbers and positive real numbers respectively, we make the following assumption:

$$b = b(v) \text{ and } c = c(v - u) \text{ where } b, c \in C^2\left(R, R^+\right) \quad \text{and} \quad c'(0) = 0. \tag{4}$$

The ecological reason for the assumption $c'(0) = 0$ is that it is often assumed in evolutionary game theory models that an individual experiences maximum competition when its trait equals the population mean, i.e. the competition coefficient c is maximized when $v = u$. In this case, the evolutionary Beverton-Holt model (2) becomes

$$x_{t+1} = b(u_t) \frac{1}{1 + c_0 x_t} x_t \tag{5}$$

$$u_{t+1} = u_t + \sigma^2 \frac{b'(u_t)}{b(u_t)} \tag{6}$$

where $c_0 = c(0) > 0$. The prime denotes differentiation: $b'(u_t) = \partial b(v)/\partial v|_{v=u_t}$. Note that (6) is uncoupled from Eq. (5), but not vice versa.

The evolutionary models (2) and (3) are included in the general models studied in [3]. The theorems in [3] extend the fundamental bifurcation theorem for general population models that occurs as extinction states destabilize. These theorems apply to the evolutionary Beverton-Holt model (5)–(6) and provide criteria under which equilibrium states are locally stable or unstable. However, in this paper we will obtain a more general, independent analysis (including global dynamics) of (5)–(6) by taking advantage of the fact that (6) is uncoupled from (5) and of the fact that the global dynamics of (5) are well-known if evolution does not take place (i. e. $\sigma^2 = 0$ and u_t remains constant at u_0 for all time).

2 Asymptotic Dynamics of the Evolutionary Beverton-Holt Model

A *critical trait (mean) u* is one for which $b'(u) = 0$. Note that critical traits are equilibria of the scalar trait equation (5). Also note that (x^e, u^e) is an equilibrium of (5)–(6) if and only if u^e is a critical trait. If u^e is a critical trait, there exist two equilibria, one with $x^e = 0$ and another with $x^e = (b(u^e) - 1)/c_0$. We call the equilibrium $(0, u^e)$ an *extinction equilibrium*. We define a *survival equilibrium* (x^e, u^e) of the system (5)–(6) to be an equilibrium associated with trait u^e for which $x^e > 0$. The survival equilibria of (5)–(6) are

$$\left(x_+^e, u^e\right) = \left(\frac{b(u^e) - 1}{c_0}, u^e\right) \tag{7}$$

where u^e is any critical trait that satisfies

$$b\left(u^e\right) > 1.$$

Let $U \subseteq R$. We say that an equilibrium (x^e, u^e), $x^e \geq 0$, is *globally attracting on* $R^+ \times U$ if $(x_0, u_0) \in R^+ \times U$ implies $\lim_{t \to +\infty} (x_t, u_t) = (x^e, u^e)$. If, in addition, (x^e, u^e) is a locally asymptotically stable equilibrium of (5)–(6), then we say it is *globally asymptotically stable (GAS) on* $R^+ \times U$.

Note that because (6) is uncoupled from (5), the local asymptotic stability of an equilibrium (x^e, u^e) of (5)–(6) implies that u^e is a locally asymptotically stable equilibrium of (6).

Theorem 1 *Assume (4) and that* u^e *is a critical trait. We have the following facts about the extinction equilibrium* $(\hat{0}, u^e)$ *and the survival equilibrium (7) of the evolutionary system (5)–(6).*

(a) (Extinction equilibria). If $b(u^e) > 1$ *or if*

$$\left|1 + \sigma^2 \frac{b''(u^e)}{b(u^e)}\right| > 1 \tag{8}$$

then the extinction equilibrium $(\hat{0}, u^e)$ *is unstable.*

If $b(u^e) < 1$ *and*

$$\left|1 + \sigma^2 \frac{b''(u^e)}{b(u^e)}\right| < 1 \tag{9}$$

then there exists an open neighborhood U *of* u^e *such that the extinction equilibrium* $(\hat{0}, u^e)$ *is GAS on* $R^+ \times U$.

(b) *(Survival equilibria) If* $b(u^e) > 1$ *and (8) hold, then the survival equilibrium (7) is unstable.*

If $b(u^e) > 1$ *and (9) hold, then there exists an open neighborhood* U *of* u^e *such that the survival equilibrium (7) is GAS on* $R^+ \times U$.

Note that the instability inequality (8) holds if $b''(u^e) > 0$.

Proof The Jacobian $J(x, u)$ of (5)–(6), when evaluated at any equilibrium (x^e, u^e), namely

$$J\left(x^e, u^e\right) = \begin{pmatrix} b(u^e) \frac{1}{(1+c_0 x^e)^2} & 0 \\ 0 & 1 + \sigma^2 \frac{b''(u^e)}{b(u^e)} \end{pmatrix} \tag{10}$$

has eigenvalues

$$\lambda_1 = b\left(u^e\right) \frac{1}{(1 + c_0 x^e)^2}, \quad \lambda_2 = 1 + \sigma^2 \frac{b''(u^e)}{b(u^e)}.$$

(a) The instability assertions follow from the linearization principle when either λ_1 or λ_2 have absolute value greater than one. By the linearization principle, the extinction equilibrium $(0, u^e)$ is locally asymptotically stable when both λ_1 or λ_2 have absolute value less than one. What remains to show, in this case, is its global asymptotic stability when $b(u^e) < 1$. Since u^e is a locally asymptotically stable equilibrium of the trait equation (6), we can find a $\delta > 0$ be such that $|u_0 - u^e| < \delta$ implies $\lim_{t\to+\infty} u_t = u^e$. Let (x_t, u_t) be any solution of (5) with an initial condition (x_0, u_0) that satisfies $x_0 > 0$ and $|u_0 - u^e| < \delta$. We need to show that

$$\lim_{t\to+\infty} (x_t, u_t) = (0, u^e). \tag{11}$$

Since we already know that $\lim_{t\to+\infty} u_t = u^e$, we need only show $\lim_{t\to+\infty} x_t = 0$. Since $b(u^e) < 1$ we can choose a real number β satisfying $b(u^e) < \beta < 1$ and, since $\lim_{t\to+\infty} u_t = u^e$, there exists a $T > 0$ such that $t \geq T$ implies $b(u_t) \leq \beta$. From (5) we have

$$0 \leq x_{t+1} = b(u_t) \frac{1}{1 + c_0 x_t} x_t \leq \beta x_t$$

for all $t \geq T$. It follows that $\lim_{t\to+\infty} x(t) = 0$, i.e. (11) holds.

(b) For a survival equilibrium (7) we have

$$0 < \lambda_1 = \frac{1}{b(u^e)} < 1.$$

When (8) holds, instability follows because $|\lambda_2| > 1$.

On the other hand, when (9) holds, then $|\lambda_2| < 1$ and the survival equilibrium is locally asymptotic stability. What remains to prove, in this case, is its global stability. Since u^e is a locally asymptotically stable equilibrium of the trait equation (6), we can find a $\delta > 0$ be such that $|u_0 - u^e| < \delta$ implies $\lim_{t\to+\infty} u_t = u^e$. Let (x_t, u_t) be any solution of (5)–(6) with an initial condition (x_0, u_0) that satisfies $x_0 > 0$ and $|u_0 - u^e| < \delta$. We need to show

$$\lim_{t\to+\infty} (x_t, u_t) = (x^e, u^e). \tag{12}$$

Since we already know that $\lim_{t\to+\infty} u_t = u^e$, we need only show $\lim_{t\to+\infty} x_t = x_+^e$. The sequence x_t is positive for all t and satisfies the asymptotically autonomous equation (5). A straightforward calculation shows that $y_t := 1/x_t$ satisfies the linear, asymptotically autonomous equation

$$0 < y_{t+1} = \frac{1}{b(u_t)} y_t + \frac{c_0}{b(u_t)}.$$

Lemma 1 in the Appendix implies

$$\lim_{t\to+} y_t = \frac{c_0}{b(u^e) - 1}$$

from which we obtain

$$\lim_{t\to+} x_t = \frac{b(u^e) - 1}{c_0} = x^e$$

and hence (12) holds. This completes the proof.

With regard to evolutionary convergence (i.e. the stability of a survival equilibrium), Theorem 1b requires (9) hold. For this inequality to hold it is necessary that $b''(u^e) < 0$ and

$$\sigma^2 < -2\frac{b(u^e)}{b''(u^e)}, \tag{13}$$

that is to say, that the speed of evolution be not too fast. The open neighborhood U in Theorem 1(b) can be taken to be the basin (interval) of attraction of u^e as a stable equilibrium of the trait equation (6).

If the inequality (13) is reversed, that is to say, if the speed of evolution is too fast, then Theorem 1(a) implies the survival equilibrium is unstable and the expected exchange of stability between equilibrium branches as $b(u^e)$ increases through 1 does not occur. This destabilization of the survival equilibrium is due to a period doubling bifurcation in the trait equation (6) as σ^2 increases through the critical value $-2b(u^e)/b''(u^e)$ where the derivative of

$$u + \sigma^2 \frac{b'(u)}{b(u)}$$

equals -1 at $u = u^e$. In such a case, the uncoupled trait equation (5) can have locally asymptotically stable periodic cycles. In this case we have the following theorem for the dynamics of the evolutionary system (5). We say that a p-periodic solution (ξ_t, υ_t) of the system (5)–(6) is a *survival* p-periodic solution if $\xi_t > 0$ for all t.

Theorem 2 *Assume (4). Suppose υ_t is a hyperbolic, locally asymptotic p-periodic solution of the trait equation (6). If*

$$\prod_{t=0}^{p-1} b(\upsilon_t) > 1 \tag{14}$$

then there exists a survival p-periodic solution (ξ_t, υ_t) of the evolutionary system (5)–(6) which is globally asymptotically stable on $R^+ \times U$ for some open neighborhood U of υ_0.

Proof The x_t component of a solution pair (x_t, u_t) of (5)–(6) satisfies the nonautonomous equation

$$x_{t+1} = b(u_t) \frac{1}{1 + c_0 x_t} x_t, \quad x_0 > 0 \tag{15}$$

where u_t satisfies the uncoupled trait equation (6). Suppose u_t approaches the p-periodic solution υ_t as $t \to +\infty$. Then the nonautonomous equation (15) is asymptotically periodic with limiting equation

$$x_{t+1} = b(\upsilon_t) \frac{1}{1 + c_0 x_t} x_t, \quad x_0 > 0. \tag{16}$$

Defining $y_t = 1/x_t$ we obtain the linear, asymptotically periodic equation

$$y_{t+1} = \frac{1}{b(u_t)} y_t + \frac{c_0}{b(u_t)}, \quad y_0 > 0. \tag{17}$$

By Theorem 1 in [2] the periodically forced limiting equation

$$y_{t+1} = \frac{1}{b(\upsilon_t)} y_t + \frac{c_0}{b(\upsilon_t)} \tag{18}$$

has a unique, positive p-periodic solution θ_t.

To show the periodic solution θ_t is attracting, we consider the p-fold composite equation of (17), which is a linear (asymptotically autonomous) equation of the form (21) in Lemma 1 in the Appendix, namely,

$$y_{t+1} = \alpha_t y_t + \beta_t$$

with

$$\alpha_t = \frac{1}{\prod_{i=0}^{p-1} b(u_t)} \to \alpha^e = \frac{1}{\prod_{i=0}^{p-1} b(\upsilon_i)} < 1$$

and a sequence $\beta_t > 0$ (a formula for which we do not need) which approaches a limit β^e as $t \to \infty$. By Lemma 1, inequality (14) implies the solution of the p-fold composite with initial condition $y_0 > 0$ approaches as $t \to \infty$ the equilibrium $\beta^e/(1-\alpha^e)$ of the limiting equation of the composite. This limit is, in fact, the first point θ_0 in the p-periodic solution θ_t. Repeating this argument using y_i, $1 \le i \le p-1$, as a starting point, we find that the solution of the composite approaches the ith point θ_i in the p-periodic solution θ_t.

All of this is to say that the solution of (17) approaches the periodic solution θ_t of the limiting equation (18), which in turns implies the solution x_t of (15) approaches the periodic solution $\xi_t = 1/\theta_t$ of the limiting equation (16). All that remains to prove is that p-periodic solution (ξ_t, υ_t) of (5)–(6) is locally asymptotically stable. This is done by investigating the eigenvalues of the Jacobian of the p-fold composite map arising from (5)–(6). This Jacobian is equal to the product $\Pi_{i=0}^{p-1} J(\xi_t, \upsilon_t)$ where

$$J(x,u) = \begin{pmatrix} \lambda_1(x,u) & b'(u)\frac{1}{1+c_0 x}x \\ 0 & \lambda_2(u) \end{pmatrix}$$

is the Jacobian of (5)–(6). Here

$$\lambda_1(x,u) := b(u)\frac{1}{(1+c_0 x)^2}, \qquad \lambda_2(x,u) := 1 + \sigma^2 \frac{b(u)\,b''(u) - \left(b'(u)\right)^2}{b^2(u)}.$$

Therefore

$$\prod_{i=0}^{p-1} J(\xi_t, \upsilon_t) = \begin{pmatrix} \pi_1 & * \\ 0 & \pi_2 \end{pmatrix}, \quad \pi_1 := \prod_{i=0}^{p-1} \lambda_1(\xi_t, \upsilon_t), \qquad \pi_2 := \prod_{i=0}^{p-1} \lambda_2(\upsilon_t),$$

where the asterisk does not concern us, since the eigenvalues of this matrix lie along the diagonal. The assumption that υ_t is a hyperbolic, locally asymptotically stable periodic solution implies that $|\pi_2| < 1$ and hence that the stability of the periodic solution (ξ_t, υ_t) is determined by the π_1.

Note that

$$\prod_{i=0}^{p-1} \xi_{i+1} = \prod_{i=0}^{p-1} b(\upsilon_i) \frac{1}{1+c_0\xi_i}\xi_i = \left(\prod_{i=0}^{p-1} b(\upsilon_i) \frac{1}{1+c_0\xi_i}\right) \prod_{i=0}^{p-1} \xi_i.$$

Since ξ_t is p-periodic, it follows that $\prod_{i=0}^{p-1} \xi_{i+1} = \prod_{i=0}^{p-1} \xi_i$ and hence

$$1 = \prod_{i=0}^{p-1} b(\upsilon_i) \frac{1}{1+c_0\xi_i} = \prod_{i=0}^{p-1} b(\upsilon_i) \prod_{i=0}^{p-1} \frac{1}{1+c_0\xi_i}.$$

Then

$$0 < \pi_1 = \prod_{i=0}^{p-1} b(\upsilon_i) \frac{1}{(1+c_0\xi_i)^2} = \frac{1}{\prod_{i=0}^{p-1} b(\upsilon_i)} < 1$$

which establishes local asymptotic stability.

3 Examples

We give two examples of the use of Theorem 1 to analyze the evolutionary model (5)–(6) with specified dependences of the coefficient $b = b(v)$ on the phenotypic trait v. In the first example, the adaptive landscape defined by b (or by fitness $\ln b$) is unimodal with a global maximum. In the second example, the adaptive landscape is bimodal. In that example, oscillations are possible and Theorem 2 is applicable.

Example 1 A common assumption in evolutionary modeling is that vital parameters are normally distributive as functions of a phenotypic trait. If we take

$$b(v) = b_0 \exp\left(-\frac{v^2}{2w^2}\right), \quad b_0 > 0$$

then $v^e = 0$ is the only critical trait. Since

$$b''(0) = \frac{-b_0}{w^2} < 0$$

we find from Theorem 1 that if $\sigma^2 < 2w^2$ then

$(\hat{0}, 0)$ is GAS on $R^+ \times R^+$ if $b_0 < 1$ and unstable if $b_0 > 1$

$\left(\frac{b_0-1}{c_0}, 0\right)$ is GAS on $R^+ \times R^+$ if $b_0 > 1$.

Here the interval $U = R^+$ because $u^e = 0$ is globally attracting as an equilibrium of the (in this case simple linear) trait equation (6) $u_{t+1} = \left(1 - \sigma^2/w^2\right) u_t$. Note that the solutions of this equation are unbounded if $\sigma^2 > 2w^2$ and hence so are the orbits of the evolutionary model (5)–(6).

In Example 1, the adaptive landscape, i.e., the graph of fitness

$$\ln b\,(v) = \ln b_0 - \frac{v^2}{2w^2}$$

as a function the mean phenotypic trait v, is unimodal with a maximum at the critical trait $v = 0$. When orbits approach a survival equilibrium, i.e. when $b_0 > 0$ and $\sigma^2 < w^2$, the evolution of fitness $\ln b\,(u_t)$ along orbits of the model (5)–(6) tends to its maximum value $\ln b_0$. In this case, the trait $u^e = 0$ is said to be an evolutionarily stable strait (ESS) since it is located at the global maximum of fitness [11].

In the next example, we illustrate the application of Theorems 1 and 2 in a case when the adaptive landscape has multiple peaks and the evolutionary model can have multiple attractors.

Example 2 An example of a two peaked adaptive landscape is provided by the fitness function $\ln b\,(v)$ with

$$b\,(v) = b_0 \exp\left(-v^2\left(3v^2 - 2v - 3\right)\right), \quad b_0 > 0. \tag{19}$$

In this case,

$$\ln b\,(u) = \ln b_0 - u^2\left(3u^2 - 2u - 3\right)$$

has three critical traits

$$u^e = -\frac{1}{2}, \quad 0, \quad 1.$$

$b\,(u)$ has a local maximum at $u^e = -1/2$, a local minimum at $u^e = 0$, and a global maximum at $u^e = 1$. See Fig. 1. Calculations show

$$\begin{aligned}
&b\,(-1/2) = b_0 e^{5/16} \text{ and } b''\,(-1/2) = -9b_0 e^{5/16} < 0\\
&b\,(0) = b_0 \text{ and } b''\,(0) = 6b_0 > 0\\
&b\,(1) = b_0 e^2 \text{ and } b''\,(1) = -18b_0 e^2 < 0.
\end{aligned}$$

Theorem 1 implies the stability results in Table 1 for the extinction and survival equilibria of the evolutionary Beverton-Holt equation (5)–(6) when b is given by (19). In each case the set U is the interval of attraction of u^e as an equilibrium of the trait equation (6), which is in this example, the scalar difference equation

$$u' = u + 6\sigma^2 u\,(2u + 1)\,(1 - u)\,. \tag{20}$$

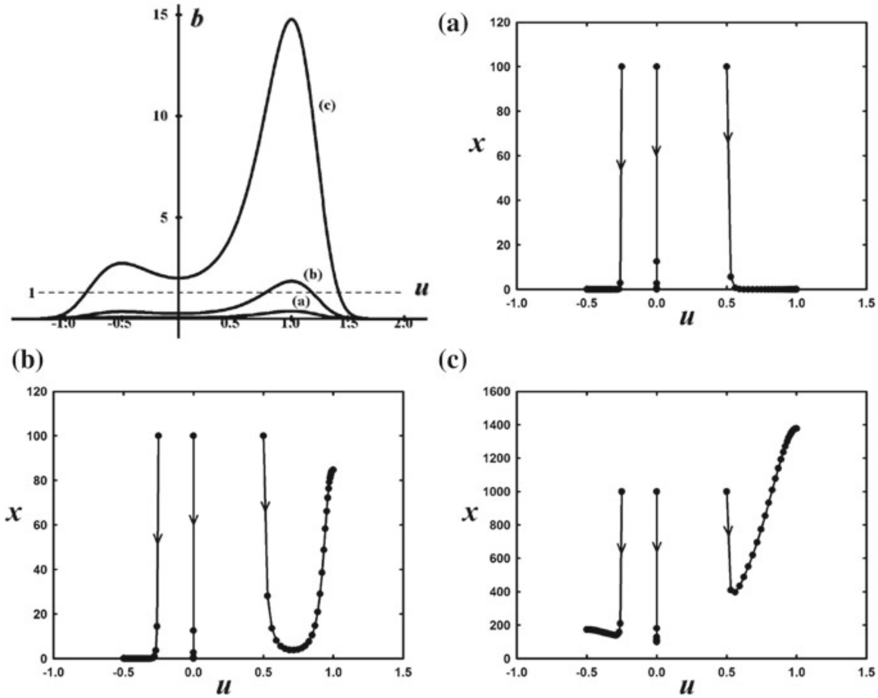

Fig. 1 The graph of $b(u)$ in (19) is shown for three values of b_0 which give rise to three different phase plane scenarios in Table 1. Phase plane plots for each case are shown, with three sample orbits, in plots (**a**), (**b**) and (**c**). In all cases $\sigma^2 = 0.01$ and $c_0 = 0.01$. **a** $b_0 = 1/20$. In this case, $b(-1/2) = e^{5/16}/20 < 1$ and $b(1) = e^2/20 > 1$ and there exists no survival equilibrium. All orbits approach an extinction equilibrium. **b** $b_0 = 1/4$. In this case, $b(-1/2) = e^{5/16}/4 < 1$ and $b(1) = e^2/4 > 1$ and there exists one survival equilibrium (7) with trait $u^e = 1$ which attracts orbits with $u_0 > 0$. Orbits with $u_0 \leq 0$ tend to an extinction equilibrium. **c** $b_0 = 2$. In this case, $b(-1/2) = 2e^{5/16} < 1$ and $b(1) = 2e^2 < 1$ and there exist three survival equilibria. Orbits with $u_0 > 0$ tend to the survival equilibrium (7) with trait $u^e = 1$. Orbits with $u_0 < 0$ tend to the survival equilibrium (7) with trait $u^e = -1/2$. Orbits with $u_0 = 0$ lie on the stable manifold of the unstable (saddle) survival equilibrium (7) with trait $u^e = 0$

When the survival equilibrium at $u^e = 1$ is stable, the trait $u^e = 1$ is an ESS since it is located at a global maximum of the fitness function. On the other hand, when the survival equilibrium at $u^e = -1/2$ is stable, the trait $u^e = -1/2$ is said to be evolutionarily convergent, but not an ESS since it does not yield a global maximum of the fitness function. It is possible, of course, for both survival equilibria at $u^e = 1$ and $u^e = -1/2$ to be stable, which occurs when $b_0 > e^{-5/16}$ and $\sigma^2 < 1/9$.

A cobwebbing analysis and a bifurcation diagram of trait equation (using the speed of evolution σ^2 as a bifurcation parameter) indicates what occurs when the slow evolution inequalities in Table 1 are violated. As σ^2 increases through the critical values 1/9 and 2/9, period doubling cascades to chaos occur as the corresponding survival equilibrium destabilizes. See Fig. 2. For any periodic cycle which results

Table 1 In the right column are the extinction and survival equilibria of the evolutionary Beverton-Holt equation (5)–(6) when b is given by (19). The middle and left columns show the criteria for the global stability and instability of each equilibrium. The set U is the interval of attraction of ue as an equilibrium of the trait equation (20)

Extinction equilibria	GAS on $R^+ \times U$	Unstable
$(0, -1/2)$	$b_0 < e^{-5/16}$ and $\sigma^2 < 2/9$	$b_0 > e^{-5/16}$ or $\sigma^2 > 2/9$
$(0, 0)$	never	always
$(0, 1)$	$b_0 < e^{-2}$ and $\sigma^2 < 1/9$	$b_0 > e^{-2}$ or $\sigma^2 > 1/9$
Survival equilibria (7)		
$\left(\frac{b_0 e^{5/16}-1}{c_0}, -1/2\right), b_0 > e^{-5/16}$	$\sigma^2 < 2/9$	$b_0 > e^{-5/16}$ and $\sigma^2 < 2/9$
$\left(\frac{b_0-1}{c_0}, 0\right), b_0 > 1$	never	always
$\left(\frac{b_0 e^2-1}{c_0}, 1\right), b_0 > e^{-2}$	$\sigma^2 < 1/9$	$\sigma^2 > 1/9$

from these bifurcations, Theorem 2 implies that there exists a periodic cycle of the evolutionary system (5)–(6) which is globally asymptotically stable on $R^+ \times U$ for some open neighborhood of the critical trait u^e.

4 Concluding Remarks

We considered the evolutionary Beverton-Holt model (2)–(3) under the assumption (4), which uncouples the trait equation (3) from the planar system (2)–(3). We proved the global stability criteria given in Theorems 1 and 2 for equilibria and for periodic cycles of the resulting system (5)–(6). The proofs of these theorems make use of the uncoupling of the equations, which produces a nonautonomous version of the Beverton-Holt equation (5) in which the coefficient $b = b(u_t)$ is driven by a solution of the scalar trait equation (6).

The dynamics of the evolutionary Beverton-Holt model (2)–(3) when the trait equation does not uncouple remains an interesting open question. For example, a common mathematical expression for the intraspecific competition coefficient $c = c(v - u)$ is [11]

$$c = c_0 \exp\left(-\frac{(v-u)^2}{2w_c^2}\right),$$

which assumes that the maximum competitive effect c_0 occurs when an individual's trait v equals the population mean trait u. If, more generally, the maximum competitive effect c_0 is dependent on the individual's trait v, then

$$c = c_0(v) \exp\left(-\frac{(v-u)^2}{2w_c^2}\right)$$

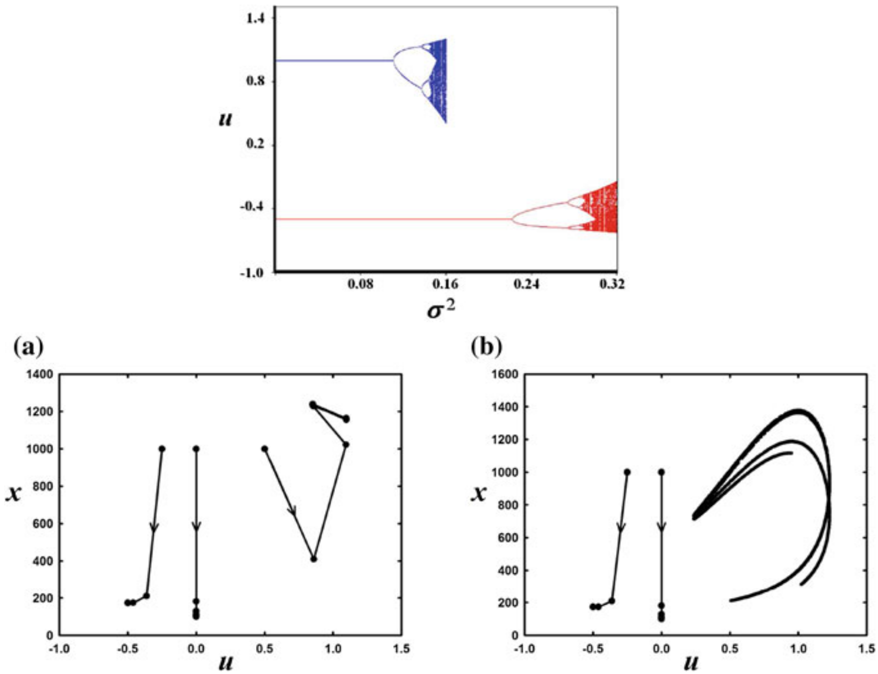

Fig. 2 The bifurcation diagram of the trait equation (20) shows a periodic doubling cascade originating at each of the survival equilibria (7) with $u^e = -1/2$ and 1 at their respective critical values of σ^2, namely, $1/9$ and $2/9$. **a** When the speed of evolution in the case of Fig. 1(**c**) is increased the $\sigma^2 = 0.12 > 1/9$, the survival equilibrium with $u_e = 1$ destabilizes, resulting in a stable period 2 cycle. **b** When the speed of evolution is further increased to $\sigma^2 = 0.17$, a chaotic attractor is reached through a period doubling cascade initiating at the survival equilibrium with $u^e = 1$

and the trait equation will no longer uncouple in the evolutionary Beverton-Holt model (2)–(3).

Acknowledgments J.M. Cushing was supported by NSF grant DMS 0917435.

Appendix

Lemma 1 *Consider the nonautonomous linear difference equation*

$$y_{t+1} = \alpha_t y_t + \beta_t \tag{21}$$

for $t = 0, 1, 2, \ldots$. *Assume* $\alpha_t, \beta_t \geq 0$ *and*

$$\lim_{t\to+\infty}\alpha_t=\alpha^e,\quad \lim_{t\to+\infty}\beta_t=\beta^e.$$

If $\alpha^e<1$ *then for all* $y_0>0$

$$\lim_{t\to+\infty}y_t=y^e:=\frac{\beta^e}{1-\alpha^e}.$$

Proof If we define $w_t:=y_t-y_e$, then

$$w_{t+1}=\alpha_t w_t+q_t$$

where

$$q_t:=\left(\alpha_t-\alpha^e\right)y^e+\beta_t-\beta^e$$

and hence

$$\lim_{t\to+\infty}q_t=0.$$

By induction

$$w_{t+1}=\left(\prod_{i=0}^{t}\alpha_i\right)w_0+\sum_{j=1}^{t}\left(\prod_{i=j}^{t}\alpha_i\right)q_{j-1}+q_t.$$

Let $\bar{\alpha}$ and $\bar{q}>0$ be upper bounds for the bounded sequences α_t and $|q_t|$ respectively. Choose a positive $\rho<1$. Since $\alpha^e<1$ and $\lim_{t\to+\infty}b_t=b^e$, for arbitrary $\varepsilon>0$, we can find a $T>0$ such that $t\geq T$ implies

$$0\leq\alpha_t\leq\rho,\quad |q_t|\leq\varepsilon\left(1-\rho\right).$$

For $t\geq T$ we have

$$\begin{aligned}|w_{t+1}|&\leq\left(\prod_{i=0}^{T}\alpha_i\right)\left(\prod_{i=T+1}^{t}\alpha_i\right)|w_0|+\sum_{j=1}^{T}\left(\prod_{i=j}^{t}\alpha_i\right)\left|q_{j-1}\right|\\&\quad+\sum_{j=T+1}^{t}\left(\prod_{i=j}^{t}\alpha_i\right)\left|q_{j-1}\right|+|q_t|\\&\leq\bar{\alpha}^{T+1}\rho^{t-T}\,|w_0|+\sum_{j=1}^{T}\left(\prod_{i=j}^{T}\alpha_i\right)\left(\prod_{i=T+1}^{t}\alpha_i\right)\left|q_{j-1}\right|\\&\quad+\sum_{j=T+1}^{t}\rho^{t-j+1}\varepsilon\left(1-\rho\right)+\varepsilon\left(1-\rho\right)\end{aligned}$$

$$|w_{t+1}| \leq \bar{\alpha}^{T+1}\rho^{t-T}\,|w_0| + \rho^{t-T}\bar{q}\sum_{j=1}^{T}\bar{\alpha}^{T-j+a} + \frac{1}{1-\rho}\varepsilon\,(1-\rho)\,.$$

Letting $t \to +\infty$ we obtain

$$\lim_{t\to+\infty}\sup|w_{t+1}| \leq \varepsilon.$$

Because $\varepsilon > 0$ is arbitrary, we conclude $\lim_{t\to+\infty}|w_{t+1}| = 0$.

References

1. Abrams, P.A.: Modelling the adaptive dynamics of traits involved in inter- and intraspecific interactions: an assessment of three methods. Ecol. Lett. **4**, 166–175 (2001)
2. Cushing, J.M.: Periodically forced nonlinear systems of difference equations. J. Differ. Equ. Appl. **3**, 547–561 (1998)
3. Cushing, J.M.: A bifurcation theorem for Darwinian matrix models. Nonlinear Stud. **17**(1), 1–13 (2010)
4. Dercole, F., Rinaldi, S.: Analysis of Evolutionary Processes: The Adaptive Dynamics Approach and Its Applications. Princeton University Press, Princeton (2008)
5. Lande, R.: Natural selection and random genetic drift in phenotypic evolution. Evolution **30**, 314–334 (1976)
6. Leslie, P.H.: Further notes on the use of matrices in population mathematics. Biometrika **35**, 213–245 (1948)
7. Pielou, E.C.: Mathematical Ecology. Wiley Interscience, New York (1967)
8. Poole, R.W.: An Introduction to Quantitative Ecology. McGraw-Hill Series in Population Biology, McGraw-Hill Book Company, New York (1974)
9. Skellam, J.F.: Studies in statistical ecology. 1. Spatial pattern. Biometrika **39**, 346–362 (1952)
10. Utida, S.: Damped oscillation of population density at equilibrium. Res. Popul. Ecol. **9**, 1–9 (1967)
11. Vincent, T.L., Brown, J.S.: Evolutionary Game Theory, Natural Selection, and Darwinian Dynamics. Cambridge University Press, Cambridge (2005)

The Periodic Decomposition Problem

Bálint Farkas and Szilárd Gy. Révész

Abstract If a function $f : \mathbb{R} \to \mathbb{R}$ can be represented as the sum of n periodic functions as $f = f_1 + \cdots + f_n$ with $f(x + \alpha_j) = f(x)$ $(j = 1, \ldots, n)$, then it also satisfies a corresponding nth order difference equation $\Delta_{\alpha_1} \ldots \Delta_{\alpha_n} f = 0$. The periodic decomposition problem asks for the converse implication, which may hold or fail depending on the context (on the system of periods, on the function class in which the problem is considered, etc.). The problem has natural extensions and ramifications in various directions, and is related to several other problems in real analysis, Fourier and functional analysis. We give a survey about the available methods and results, and present a number of intriguing open problems. Most results have already appeared elsewhere, while the recent results of [7, 8] are under publication. We give only some selected proofs, including some alternative ones which have not been published, give substantial insight into the subject matter, or reveal connections to other mathematical areas. Of course this selection reflects our personal judgment. All other proofs are omitted or only sketched.

Dedicated to Imre Z. Ruzsa on the occasion of his 60th birthday.

Bálint Farkas was supported in part by the Hungarian National Foundation for Scientific Research, Project #K-100461, Szilárd Gy Révész was supported in part by the Hungarian National Foundation for Scientific Research, Project #'s K-81658, K-100461, NK-104183, K-109789.

B. Farkas
Bergische Universität Wuppertal, Wuppertal, Germany
e-mail: farkas@math.uni-wuppertal.de

Sz. Gy. Révész (✉)
Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, Kuwait University, Kuwait, Kuwait
e-mail: szilard.revesz@renyi.hu; szilard@sci.kuniv.edu.kw

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_8

Keywords Periodic functions · Periodic decomposition · Difference equation · Almost periodic and mean periodic functions · Transformation invariant functions · Functions with values in a group · Operator semigroups

1 Introduction

Let $f : \mathbb{R} \to \mathbb{R}$ be a function with

$$f = f_1 + \dots + f_n, \qquad f_j(x + \alpha_j) = f_j(x) \quad \forall x \in \mathbb{R}, \quad j = 1, \dots, n, \tag{1}$$

where $\alpha_j \in \mathbb{R}$ are fixed real numbers. We call this an $(\alpha_1, \dots, \alpha_n)$*-periodic decomposition* of f. For $\alpha \in \mathbb{R}$ let Δ_α denote the (forward) *difference operator*

$$\Delta_\alpha : \mathbb{R}^\mathbb{R} \to \mathbb{R}^\mathbb{R}, \quad \Delta_\alpha g(x) := g(x + \alpha) - g(x).$$

Then the α_i-periodicity of f_i above means $\Delta_{\alpha_i} f_i = 0$. The difference operators commute, so

$$\Delta_{\alpha_1} \Delta_{\alpha_2} \dots \Delta_{\alpha_n} f = 0. \tag{2}$$

Problem 1.1 *(Ruzsa, 70s)* Does the converse implication "(2) ⇒ (1)" hold true?

Naturally, this question can be posed *in any given function class* $\mathcal{F} \subseteq \mathbb{R}^\mathbb{R}$.

Definition 1.2 Let $\mathcal{F} \subseteq \mathbb{R}^\mathbb{R}$ be a set of functions. With $n \in \mathbb{N}$, $n \geq 1$, and $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ given, the function class $\mathcal{F}$ is said to have the *decomposition property with respect to* $\alpha_1, \dots, \alpha_n$ if for each $f \in \mathcal{F}$ satisfying (2) a periodic decomposition (1) exists with $f_j \in \mathcal{F}$ $(j = 1, \dots, n)$. Furthermore, the function class $\mathcal{F}$ has the n*-decomposition property* if it has the decomposition property for every choice of $\alpha_1 \dots, \alpha_n \in \mathbb{R}$, and $\mathcal{F}$ has the *decomposition property* if it has the n-decomposition property for each integer $n \geq 1$.

Notice that we did not speak about uniqueness of decompositions. As we shall see uniqueness is an intriguing problem and in general cannot be expected. Note also that $\mathbb{R}^\mathbb{R}$ or $\mathrm{C}(\mathbb{R})$ (space of continuous functions) do *not* have the n-decomposition property for $n \geq 2$. Indeed, let $n = 2$ and $\alpha_1 = \alpha_2 = \alpha$. The identity function $\mathrm{id}(x) := x$ satisfies $\Delta_\alpha \Delta_\alpha \mathrm{id} = 0$, but it fails to be α-periodic. So the implication "(2) ⇒ (1)" fails. As a matter of fact, a function class containing the identity does not have the decomposition property.

The above choice for α_1, α_2 hides the nature of the problem a bit: The existence of periodic decompositions may depend on the system $\alpha_1, \dots, \alpha_n$ of prescribed periods. If we take $\alpha_1 = 1$ and $\alpha_2 = \sqrt{2}$ the arguments above do not work. And in fact, *if* α_1 *and* α_2 *are incommensurable* (i.e., $\alpha_1\mathbb{Z} \cap \alpha_2\mathbb{Z} = \{0\}$) then $f = \mathrm{id} : \mathbb{R} \to \mathbb{R}$ has a decomposition as $f = f_1 + f_2$, $\Delta_{\alpha_j} f_j = 0$.

Proposition 1.3 *Let $\alpha_1, \alpha_2 \in \mathbb{R}$ be incommensurable. Then each function $f : \mathbb{R} \to \mathbb{R}$ satisfying* (2) *can be written as $f = f_1 + f_2$, with f_1, f_2 being α_1 and α_2 periodic, respectively. That is, $\mathbb{R}^{\mathbb{R}}$ has the decomposition property with respect to any system of two incommensurable reals.*

Proof Using the *axiom of choice*, we can select one representative from each of the classes of the equivalence relation $x \sim y \Leftrightarrow x - y \in \alpha_1\mathbb{Z} + \alpha_2\mathbb{Z}$.

On each class we construct our f_j as follows. For the fixed class representative $y \in \mathbb{R}$ take $f_1(y + k\alpha_1 + m\alpha_2) := f(y + m\alpha_2)$ and $f_2(y + k\alpha_1 + m\alpha_2) := f(y + k\alpha_1) - f(y)$. Then f_j are α_j-periodic and by (2)

$$\begin{aligned} f(y + k\alpha_1 + m\alpha_2) &= f(y + m\alpha_2) + f(y + k\alpha_1) - f(y) \\ &= f_1(y + k\alpha_1 + m\alpha_2) + f_2(y + k\alpha_1 + m\alpha_2). \end{aligned}$$

This ends the construction of a periodic decomposition. □

Of course, the decomposition given in the preceding proof depends on the particular choice of the representatives for the equivalence classes, hence uniqueness cannot be expected. In fact, by adding and subtracting a function constant on $\alpha_1\mathbb{Z} + \alpha_2\mathbb{Z}$ to f_1 and f_2 respectively, we immediately obtain different decompositions. In Sect. 8 below we shall return to this matter. The above decomposition can be far worse than the function itself. E.g., $f = \mathrm{id}$ is continuous, while f_1 and f_2 are certainly not, for continuous periodic functions, hence also their sums, are necessarily bounded. That $f = \mathrm{id}$ does not even have a measurable decomposition, is proved in [34] by a somewhat involved argumentation.

In fact, no function with $\lim_{x\to\infty} f(x) = \infty$ can have a measurable periodic decomposition. To see this, let $\varepsilon, \eta > 0$ be arbitrarily fixed, and assume that f has a measurable decomposition (1). Then for each $j = 1, \ldots, n$, f_j must be bounded on $[0, \alpha_j]$ by some constant $K_j < \infty$ apart from an exceptional set $A_j \subseteq [0, \alpha_j]$ of Lebesgue measure $|A_j| < \eta$. Therefore, on any interval I of length ℓ (large), f is bounded by $K := K_1 + \cdots + K_n < \infty$ apart from an exceptional set $A \subseteq I$ of measure $|A| < (\lceil \ell/\alpha_1 \rceil + \cdots + \lceil \ell/\alpha_n \rceil)\eta < \varepsilon\ell$, if η is chosen small enough. So f is "locally almost bounded": for any $\varepsilon > 0$ there is $K < \infty$ such that on any sufficiently large interval I, $|\{x \in I : |f(x)| > K\}| < \varepsilon|I|$.

One would think that the bug here is with the axiom of choice, the huge number of "ugly", non-measurable functions, so that once a continuous function has a relatively nice—say, measurable—decomposition, then it must also have a continuous one. However, the contrary is true:

Proposition 1.4 (Keleti [24]) *There exists $f \in C(\mathbb{R})$ having measurable decomposition* (1) *but without a continuous periodic decomposition.*

For the proof see [23, Theorem 4.8].

We can also look for further immediate solutions of (2): For example polynomials of degree $m < n$ satisfy this difference equation. So, we can ask for *quasi-decompositions* with *periodic functions and polynomials*

$$f = p + f_1 + \cdots + f_n, \quad \text{with} \quad \Delta_{\alpha_j} f_j = 0 \quad \text{and} \quad \deg p < n \ \text{ a polynomial.} \qquad (3)$$

Theorem 1.5 (Ruzsa and Szegedy (unpublished)) *There exist continuous, unbounded solutions of* (2) *with* $\lim_{x\to\infty} f(x)/x = 0$.

As a consequence $\mathrm{C}(\mathbb{R})$ does not have the quasi-decomposition property either. For a discussion see [29, pp. 338–339]. It can be precisely described which functions in $\mathrm{C}(\mathbb{R})$ have continuous periodic quasi-decompositions (3).

Theorem 1.6 (Laczkovich and Révész [29]) *For a function* $f \in \mathrm{C}(\mathbb{R})$ *the existence of a quasi-decomposition* (3) *is equivalent to* (2) *together with the Whitney condition*

$$\delta_n(f) := \sup\Big\{\sum_{j=0}^{n}(-1)^j\binom{n}{j} f(x+jh) \ : \ x, h \in \mathbb{R}\Big\} < \infty.$$

Proof Notice first $\sum_{j=0}^{n}(-1)^{n-j}\binom{n}{j} f(x+jh) = \Delta_h^n f(x)$. Hence, if $f = p + f_1 + \cdots + f_n$ as in (3), then $\Delta_h^n p = 0$. Since f_j is α_j-periodic and continuous $\delta_n(f_j) \le 2^n \sup_{t\in[0,\alpha_j]} |f_j(t)|$. So that (3) implies both (2) and $\delta_n(f) < \infty$. Conversely, a result of Whitney [38] says that $\delta_n(f) < \infty$ entails that f can be approximated by a polynomial p of degree $\deg p < n$ within a bounded distance: $\|f - p\|_\infty < \infty$. Thus, for $g := f - p \in \mathrm{BC}(\mathbb{R})$ we have $\Delta_{\alpha_1} \ldots \Delta_{\alpha_n} g = 0$ and it remains to establish the decomposition property of $\mathrm{BC}(\mathbb{R})$, postponed to Sect. 4.1. □

2 Continuous Periodic Decompositions

In view of the foregoing discussion it is natural to pose the boundedness condition on the occurring functions and look at subclasses $\mathcal{F}$ of the space $\mathrm{BC}(\mathbb{R})$ of bounded continuous functions on $\mathbb{R}$. Note that if f has a continuous periodic decomposition it is *uniformly almost periodic* (alternatively, Bohr or Bochner almost periodic), i.e., the set

$$\big\{f(\cdot + t) : t \in \mathbb{R}\big\} \subseteq \mathrm{BC}(\mathbb{R})$$

of its translates is relatively compact with respect to the supremum norm $\|f\|_\infty := \sup_{x\in\mathbb{R}} |f(x)|$. Denote by $\mathrm{UAP}(\mathbb{R})$ the set of all such functions, which becomes a Banach space, actually a C^*-algebra, if endowed with the supremum norm and pointwise operations, see [2, Chap. 1]. Evidently, a solution of (2) in $\mathcal{F} \subseteq \mathrm{BC}(\mathbb{R})$ must be contained by $\mathrm{UAP}(\mathbb{R})$ if $\mathcal{F}$ has the decomposition property.

Proposition 2.1 *The space* UAP($\mathbb{R}$) *has the decomposition property.*

At this point, we give a proof only for the case of incommensurable periods to illustrate the use of Fourier analytic techniques. The complete proof will be given in Sect. 3 as a special case of a more general result, see Example 3.7.

Proof Suppose $\alpha_1, \dots, \alpha_n$ are incommensurable and let $f \in \mathrm{UAP}(\mathbb{R})$. Any $f \in \mathrm{UAP}(\mathbb{R})$ has a mean value

$$Mf := \lim_{t \to \infty} \frac{1}{t} \int_0^t f(s)\mathrm{d}s \in \mathbb{C}$$

by [2, Sect. 1.3], and M is a continuous linear functional on UAP($\mathbb{R}$). Moreover, for $\lambda \in \mathbb{R}$ the Fourier coefficients of f are defined as $a(\lambda) := M(f(s)\mathrm{e}^{-is\lambda})$ among which only countably many are nonzero, denote these by c_k and the corresponding "frequences" by λ_k. We say that f has the Fourier series $f \sim \sum_k c_k \mathrm{e}^{ix\lambda_k}$.

Let $\alpha \in \mathbb{R}$. In what follows M is understood with respect to the variable s and Δ_α with respect to the variable x. We have

$$\begin{aligned} c_k \Delta_\alpha\Big(\mathrm{e}^{ix\lambda_k}\Big) &= \Delta_\alpha\Big(M(f(s)\mathrm{e}^{-is\lambda_k})\mathrm{e}^{ix\lambda_k}\Big) \\ &= M\Big(f(s)\mathrm{e}^{-i(s-(x+\alpha))\lambda_k}\Big) - M\Big(f(s)\mathrm{e}^{-i(s-x)\lambda_k}\Big) \\ &= M\Big((f(s+\alpha) - f(s))\mathrm{e}^{-i(s-x)\lambda_k}\Big) = M\Big(\Delta_\alpha f(s)\mathrm{e}^{-is\lambda_k}\Big)\mathrm{e}^{ix\lambda_k}. \end{aligned}$$

So that the difference equation (2) implies $\Delta_{\alpha_1} \dots \Delta_{\alpha_n} c_k \mathrm{e}^{ix\lambda_k} = 0$. Since $c_k \neq 0$, this is only possible if $\lambda_k = 2\pi\ell/\alpha_j$ for some $\ell \in \mathbb{Z}$ and $j \in \{1, \dots, n\}$. Since $\alpha_1, \dots, \alpha_n$ are incommensurable there can be at most one such j.

On the other hand, by Sect. 1.8.6° in [2] $\frac{1}{N}\sum_{k=1}^N f(s+k\alpha_j)$ converges uniformly (in s) as $N \to \infty$ to an α_j-periodic continuous function f_j, whose (non-zero) Fourier coefficients are precisely those Fourier coefficients $a(\lambda)$ of f for which $\lambda \in (2\pi/\alpha_j)\mathbb{Z}$. We see therefore that $f_1 + \dots + f_n$ and f have the same Fourier coefficients, hence they coincide by Theorem I.4.7° in [2]. □

That is to say if we *a priori* know that f is uniformly almost periodic, then the difference equation (2) implies the periodic decomposition (1).

The next step is to deduce this almost periodicity. Let $\mu \in \mathrm{M}_c(\mathbb{R})$, i.e., a compactly supported finite (signed) Borel measure on $\mathbb{R}$, and let $f \in \mathrm{C}(\mathbb{R})$. Then

$$f * \mu(x) := \int_{\mathbb{R}} f(x-t)\mathrm{d}\mu(t)$$

defines a continuous function, the *convolution* of f and μ. The convolution of two measures $\mu, \nu \in \mathrm{M}_c(\mathbb{R})$ is defined by $f * (\mu * \nu) := (f * \mu) * \nu$ (for $f \in \mathrm{C}(\mathbb{R})$): As a continuous linear functional on the locally convex space $\mathrm{C}(\mathbb{R})$, $\mu * \nu$ is a compactly supported measure, i.e., $\mu * \nu \in \mathrm{M}_c(\mathbb{R})$. It is also easy to see that convolution is commutative and associative in $\mathrm{M}_c(\mathbb{R})$.

Now denote $\mu_\alpha := \delta_{-\alpha} - \delta_0$, where δ_β is the Dirac measure at $\beta \in \mathbb{R}$. Then $f * \mu_\alpha = f * (\delta_{-\alpha} - \delta_0) = \Delta_\alpha f$. With this Eq. (2) takes the form

$$f * (\mu_{\alpha_1} * \cdots * \mu_{\alpha_n}) = f * \big((\delta_{-\alpha_1} - \delta_0) * \cdots * (\delta_{-\alpha_n} - \delta_0)\big) = 0.$$

Definition 2.2 *(Schwartz* [35]*)* A function $f \in \mathrm{C}(\mathbb{R})$ is *mean periodic* if there exists a compactly supported Borel measure μ on $\mathbb{R}$ with $f * \mu = 0$. i.e. $\int_{-\infty}^{\infty} f(x-t)\mathrm{d}\mu(t) = 0$.

Let us recall from [21, p. 44] the following.

Proposition 2.3 (Kahane) *A bounded uniformly continuous mean periodic function is uniformly almost periodic.*

An immediate consequence of this and of Proposition 2.1 is the following.

Proposition 2.4 (Gajda [13]) *The Banach space* $\mathrm{BUC}(\mathbb{R})$ *has the decomposition property.*

Gajda proved this results with a different argument (using Banach limits) that can be easily extended to the case of translations on locally compact Abelian groups (see Corollary 7.2).

However, the result of Gajda for $\mathrm{BUC}(\mathbb{R})$ falls short of the complete truth, in the extent that it does not tell that a continuous function satisfying (2) is necessarily uniformly continuous, a fact that would imply even the decomposition property of the whole $\mathrm{BC}(\mathbb{R})$ itself.

No direct proof of the implication "$f \in \mathrm{BC}(\mathbb{R})$ & (2) $\Rightarrow f \in \mathrm{BUC}(\mathbb{R})$" is known, so the decomposition property of $\mathrm{BC}(\mathbb{R})$ lies deeper. In fact, to prove that a bounded continuous solution of (2) is uniformly continuous, we have no other known ways than this periodic decomposition result on $\mathrm{BC}(\mathbb{R})$ itself.

Before proceeding let us formulate the following more general question than Problem 1.1.

Problem 2.5 Let μ, ν (or $\mu_1, \ldots, \mu_n$) be given Borel measures of compact support on $\mathbb{R}$. Clearly, if

$$f = g + h \quad \text{with} \quad g, h \in \mathrm{C}(\mathbb{R}) \quad \text{such that} \quad g * \mu = 0,\ h * \nu = 0, \tag{4}$$

then $f * (\mu * \nu) = 0$. Find conditions, under which we have the converse implication: If $f \in \mathrm{C}(\mathbb{R})$, and $f * (\mu * \nu) = 0$, then (4) holds. Or find conditions on μ ensuring that a solution $f \in \mathrm{BC}(\mathbb{R})$ of $f * \mu = 0$ is almost periodic.

In this formulation we use no assumption on boundedness or uniform continuity. Clearly, then additional assumptions are needed. E.g. additional functional equations must also be satisfied? Spectra must be simple? Spectra of μ and ν should be distinct? Several variations may be considered.

Remark 2.6 In the problem above f is by default mean periodic. However, convergence of mean periodic Fourier expansions was shown only in a complicated, complex sense. Perhaps, recent developments in the Fourier synthesis and representation of mean periodic functions can be used, see Székelyhidi [37]. Then again, boundedness and uniform continuity could be of use by means of Proposition 2.3 of Kahane.

Wierdl [39] showed that the space BC($\mathbb{R}$) of bounded continuous functions has the 2-decomposition property. Subsequently, Laczkovich and Révész proved this for general n as the main result of [29], which was the first internationally published paper in this topic (but see also the preceding paper [28]).

Theorem 2.7 (Laczkovich and Révész [29]) *The Banach space* BC($\mathbb{R}$) *has the decomposition property.*

Although many generalizations and interpretations have since been described and various tools could be invoked depending on the setup, oddly enough this first non-trivial result could be covered by neither extensions. To date, we have no other proof than the essentially elementary yet tricky original argument. In Sect. 4.1 we present a proof of this result utilizing the operator theoretic approach to be developed next.

3 Generalizations to Linear Operators

For $\alpha \in \mathbb{R}$ the translation by α acts as a homeomorphism on $\mathbb{R}$. Consider the so-called *Koopman (or composition) operator*, in this case called the *shift operator*,

$$T_\alpha : \mathbb{R}^{\mathbb{R}} \to \mathbb{R}^{\mathbb{R}}, \quad T_\alpha f(x) := f(x+\alpha).$$

Observe that the solutions of the difference equation (2) form the subspace

$$\ker(T_{\alpha_1} - \mathrm{I}) \cdots (T_{\alpha_n} - \mathrm{I})$$

(where I denotes the identity operator), while the functions having a periodic decomposition (1) are the elements of

$$\ker(T_{\alpha_1} - \mathrm{I}) + \cdots + \ker(T_{\alpha_n} - \mathrm{I}).$$

Then Problem 1.1 can be rephrased so as whether the equality

$$\ker(T_{\alpha_1} - \mathrm{I}) \cdots (T_{\alpha_n} - \mathrm{I}) = \ker(T_{\alpha_1} - \mathrm{I}) + \cdots + \ker(T_{\alpha_n} - \mathrm{I}) \tag{5}$$

holds? Of course, one can restrict the question by considering linear subspaces of $\mathbb{R}^{\mathbb{R}}$ that are invariant under the occurring operators. The equality then means the decomposition property of $\mathcal{F}$. Or more generally one can ask the following:

Problem 3.1 Let E be a linear space and let $T_1, \ldots, T_n : E \to E$ be commuting linear operators. Find conditions such that

$$\ker(T_1 - \mathrm{I}) \cdots (T_n - \mathrm{I}) = \ker(T_1 - \mathrm{I}) + \cdots + \ker(T_n - \mathrm{I}). \tag{6}$$

Remark 3.2 For a system of pairwise commuting operators $T_1, \ldots, T_n$ the inclusion "$\ker(T_1 - \mathrm{I}) \cdots (T_n - \mathrm{I}) \supseteq \ker(T_1 - \mathrm{I}) + \cdots + \ker(T_n - \mathrm{I})$" trivially holds. This corresponds to the trivial implication "(1) $\Rightarrow$ (2)".

The first result in this direction is the following:

Theorem 3.3 (Laczkovich and Sz. Révész [30]) *Let X be a topological vector space and $T_1, \ldots, T_n$ be commuting continuous linear operators on X. Suppose that for every $x \in X$ and $j \in \{1, \ldots, n\}$ the closed convex hull of $\{T_j^m x : m \in \mathbb{N}\}$ contains a fixed point of T_j, that is*

$$\overline{\mathrm{conv}}\left\{T_j^m x : m \in \mathbb{N}\right\} \cap \ker(T_j - \mathrm{I}) \neq \emptyset.$$

Then (6) *holds.*

We shall give the proof of this theorem in a special case only, see Proposition 3.6, because that proof yields some extra information about the obtained decompositions. For the proof of the general statement we refer to [30]. For a Banach space E we denote by $\mathscr{L}(E)$ the space of bounded linear operators on E. Here are some consequences of the previous theorem:

Corollary 3.4 *Let E be a Banach space and let $T_1, \ldots, T_n \in \mathscr{L}(E)$ be commuting power bounded operators. Suppose an additional vector topology τ is given on E such that the unit ball* $\mathrm{B} := \{x \in E : \|x\| \leq 1\}$ *is τ-compact, and the operators T_j are τ-continuous. Then* (6) *holds.*

The proof is the application of the foregoing result and the Markov–Kakutani fixed point theorem (see, e.g., [5, Sect. 10.1]) to the closed convex hull $\overline{\mathrm{conv}}\{T_j^m x : m \in \mathbb{N}\}$, which was assumed to be τ-compact.

The above together with the Banach–Alaoglu theorem yields the following:

Proposition 3.5 *Let X be a normed space, $E := X^*$ and let $\tau := \sigma(X^*, X)$ be the weak* topology on X^*. If $T_1, \ldots, T_n \in \mathscr{L}(E)$ are commuting, power bounded weakly* continuous operators, then* (6) *holds.*

Let E be a Banach space. Suppose $T_1, \ldots, T_n \in \mathscr{L}(E)$ are power bounded, then the fixed point condition in Theorem 3.3 means precisely the mean ergodicity of $T_1, \ldots, T_n$, see [5, Theorem 8.20]. Recall that $T \in \mathscr{L}(E)$ is *mean ergodic* if

$$Px := \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} T^j x$$

exists for every $x \in E$. In this case the limit P is a bounded projection onto $\ker(T-\mathrm{I})$, the so-called *mean ergodic projection*, and one has $E = \mathrm{rg} P \oplus \ker P$ and $\ker P = \overline{\mathrm{rg}}(T - \mathrm{I})$, where rg and $\overline{\mathrm{rg}}$ stand for the range and the closure of the range of an operator, respectively, see [5, Sect. 8.4].

If T, S are commuting mean ergodic operators with mean ergodic projections P, Q, then $PS = SP$ (and $TQ = QT$), so that $PQ = QP$.

Proposition 3.6 *Let E be a Banach space and $T_1, \ldots, T_n \in \mathscr{L}(E)$ be commuting mean ergodic operators. Then the equality* (6) *holds.*

Proof Since the operators $T_1, \ldots, T_n$ commute, so do the mean ergodic projections $P_1, \ldots, P_n$, and actually all operators occurring in this proof commute with each other. A moment's thought explains that the *direct* decomposition

$$\begin{aligned} E &= \mathrm{rg} P_1 \oplus \mathrm{rg} P_2(\mathrm{I} - P_1) \oplus \cdots \oplus \mathrm{rg}\big(P_n(\mathrm{I} - P_{n-1}) \cdots (\mathrm{I} - P_1)\big) \\ &\quad \oplus \mathrm{rg}\big((\mathrm{I} - P_n)(\mathrm{I} - P_{n-1}) \cdots (\mathrm{I} - P_1)\big) \end{aligned}$$

is valid, i.e. for any $x \in E$ we can uniquely write $x = x_1 + \cdots + x_n + y$ with $x_i \in \mathrm{rg} P_i = \ker(T_i - \mathrm{I})$ and $y \in \mathrm{rg}(\mathrm{I} - P_1) \cdots (\mathrm{I} - P_n)$. Let now $x \in \ker(T_1 - \mathrm{I}) \cdots (T_n - \mathrm{I})$: then $(T_1 - \mathrm{I}) \cdots (T_n - \mathrm{I})y = 0$. It follows that $y \in \ker(T_1 - \mathrm{I}) \cdots (T_n - \mathrm{I}) \subseteq \ker(\mathrm{I} - P_1) \cdots (\mathrm{I} - P_n)$, thus $y \in \mathrm{rg}(\mathrm{I} - P_1) \cdots (\mathrm{I} - P_n) \cap \ker(\mathrm{I} - P_1) \cdots (\mathrm{I} - P_n)$. However, $(\mathrm{I} - P_1) \cdots (\mathrm{I} - P_n)$ is a projection, so from this $y = 0$ follows. ◻

Actually, the proof above and the result itself appears in [19] in a slightly more general form, and as a matter of fact even much earlier in [30]. None of the papers however formulated it by using the notion of mean ergodicity.

Example 3.7 Since shift operators T_α are all mean ergodic on $E = \mathrm{UAP}(\mathbb{R})$ we obtain a (complete) proof of Proposition 2.1. To see that T_α is mean ergodic it suffices to note that $\{T_\alpha^n : n \in \mathbb{N}\}$ is compact in the strong operator topology and to invoke [5, Theorem 8.20]; or alternatively one can use [2, Sect. 1.8.6°] as in the proof of Proposition 2.1 given for incommensurable periods.

Remark 3.8

(a) Notice, that $Q_j = P_j(\mathrm{I} - P_{j-1}) \cdots (\mathrm{I} - P_2)(\mathrm{I} - P_1)$ is a bounded projection on E. One trivially has $\|P_j\| \le \|T_j\|$ and $\|\mathrm{I} - P_j\| \le 1 + \|T_j\|$. If we suppose $\|T_j\| \le 1$ for $j = 1, \ldots, n$, then $\|Q_j\| \le 2^{j-1}$. The proof above yields that the decomposition obtained is actually

$$\begin{aligned} x &= P_1 x + P_2(\mathrm{I} - P_1)x + \cdots + P_n(\mathrm{I} - P_{n-1}) \cdots (\mathrm{I} - P_2)(\mathrm{I} - P_1)x \\ &= Q_1 x + Q_2 x + \cdots + Q_n x. \end{aligned}$$

Hence x has a decomposition $x = x_1 + \cdots + x_n$ with $x_j \in \ker(T_j - \mathrm{I})$ and

$$\max_{j=1,\dots,n} \|x_j\| \le 2^{n-1}\|x\|.$$

(b) If E is a Hilbert space, then the mean ergodic projections P_j are orthogonal, see [5, Theorem 8.6]. So that $\mathrm{I} - P_j$ is also an orthogonal, hence contractive, projection. This implies that $x \in \ker(T_1 - \mathrm{I}) \cdots (T_n - \mathrm{I})$ has a decomposition $x = x_1 + \cdots + x_n$ with $x_j \in \ker(T_j - \mathrm{I})$ and

$$\max_{j=1,\dots,n} \|x_j\| \le \|x\|.$$

(c) In the original setting of the decomposition problem Laczkovich and Révész have shown that on $E = \mathrm{BC}(\mathbb{R})$ with T_j being translations by a_j a function f satisfying (2) has a decomposition $f = f_1 + \cdots + f_n$ with

$$\max_{j=1,\dots,n} \|f_j\|_\infty \le 2^{n-2}\|f\|.$$

The estimate is sharp for $n = 2$, see [29].

Problem 3.9 Find the best constant C_n such that any $x \in \ker(T_1 - \mathrm{I}) \cdots (T_n - \mathrm{I})$ has some decomposition $x = x_1 + \cdots + x_n$ with $x_j \in \ker(T_j - \mathrm{I})$ and

$$\max_{j=1,\dots,n} \|x_j\| \le C_n\|x\|.$$

We saw $C_n \le 2^{n-1}$ in general, $C_n \le 2^{n-2}$ for translations on $\mathrm{BC}(\mathbb{R})$. Are these estimates sharp? Is it true that $C_n = 1$ for translations on $\mathrm{BC}(\mathbb{R})$ for every $n? \in \mathbb{N}$, $n \ge 1$? Under which conditions on E and/or $T_1, \cdots, T_n$ does $C_n = 1$ hold?

Example 3.10 It is a classical result that a power bounded operator on a reflexive Banach space E is mean ergodic. As a consequence, commuting power bounded operators on a reflexive Banach space E fulfill the conditions of Proposition 3.6, hence (6) holds true. See also [30, Corollary 2.6]

Definition 3.11 Let E be a Banach space, or, more generally, a topological vector space. We say that E has the *decomposition property with respect to the pairwise commuting operators* $T_1, \dots, T_n \in \mathscr{L}(E)$ if (6) holds. Moreover, if $\mathscr{A} \subseteq \mathscr{L}(E)$ and E has the decomposition property for each system of n pairwise commuting operators $T_1, \dots, T_n \in \mathscr{A}$, then E is said to have the *n-decomposition property with respect to $\mathscr{A}$*. Finally, if this holds for all $n \in \mathbb{N}$, then E is said to have the *decomposition property with respect to $\mathscr{A}$*.

So that e.g. Example 3.10 means that a reflexive Banach space has the decomposition property with respect to (commuting) power bounded operators. This new terminology shall not cause any ambiguity in connection with the decomposition property of function classes $\mathscr{F} \subseteq \mathbb{R}^{\mathbb{R}}$ (in Definition 1.2).

Remark 3.12 If 1 is not an eigenvalue of say T_1, then the questioned equality (6) reduces to $\ker(T_2 - \mathrm{I}) \dots (T_n - \mathrm{I}) = \ker(T_2 - \mathrm{I}) + \dots + \ker(T_n - \mathrm{I})$. That is to say the order n reduces to order $n-1$. In particular, if 1 is not a spectral value for every $T_1, \dots, T_n$, then (6) is satisfied trivially, both sides being $\{0\}$.

Note the following border-line feature of our subject matter. It is only interesting to look at cases when $\|T_1\| \geq 1, \dots, \|T_n\| \geq 1$ (since $\mathrm{I} - T$ is invertible for $\|T\| < 1$). On the other hand, if $T_1, \dots, T_n$ are power bounded and commute, we can equivalently renorm E by $|||x||| := \sup_{k_1,\dots,k_n \in \mathbb{N}} \left\| T_1^{k_1} \dots T_n^{k_n} x \right\|$, such that for the new norm each operator becomes a contraction. Hence in the end with the assumption $\|T_1\| = \dots = \|T_n\| = 1$ one loses no generality (for the particularly fixed power bounded operators $T_1, \dots, T_n$).

Recall that a Banach space E is called *m-quasi-reflexive* if E has codimension m in its bidual E^{**}.

Theorem 3.13 (Kadets and Shumyatskiy [20])

(a) A 1-quasi reflexive Banach space E has the 2-decomposition property with respect to any pair of commuting linear transformations S, T of norm 1.
(b) If E is m-quasi reflexive with $m > 1$, then there exist commuting linear transformations $S, T \in \mathscr{L}(E)$ of norm 1 such that E fails to have the 2-decomposition property with respect to S, T.

Also Kadets and Shumyatskiy proved the following:

Theorem 3.14 (Kadets and Shumyatskiy [19]) *Neither the space c_0 of null sequences, nor ℓ^1 has the 2-decomposition property with respect to operators of norm 1.*

See [19] for the proofs and for further information on averaging techniques which can be used in connection with the periodic decomposition problem. Several natural questions arise, see [20]:

Problem 3.15

1. Is it true that in a 1-quasi reflexive space E has the decomposition property with respect to any finite system of commuting operators of norm 1?
2. Does the 2-decomposition property with respect to contractions imply the n-decomposition property with respect to contractions?
3. Does the 2-decomposition property with respect to power bounded operators characterizes m-quasi reflexive Banach spaces with $m \leq 1$?

Let us finally remark that a recent result of Fonf et al. [12] states that a separable 1-quasi reflexive space can be equivalently renormed such that every contraction with respect to the new norm becomes mean ergodic. Also a classical result of theirs, see [11], is that a Banach space E is reflexive if (and only if) every power bounded operator is mean ergodic. These indicate the possible difficulty of Problem 3.15.

The original setting of the decomposition problem has a special feature, namely that the translation operators T_t on translation invariant subspaces E of $\mathbb{R}^{\mathbb{R}}$ form a one-parameter (semi)group of linear operators. In the rest of this section we shall study this aspect from a more general point of view. Given a Banach space E, a *one-parameter semigroup* T is a unital semigroup homomorphism $T : [0, \infty) \to \mathscr{L}(E)$, i.e., $T(t+s) = T(t)T(s)$ and $T(0) = \mathrm{I}$ are fulfilled for every $t, s \geq 0$. Whereas a *one-parameter group* defined analogously as a group homomorphism (into the group of invertible operators). On $\mathbb{R}^{\mathbb{R}}$ one can define the translation group by $T(t)f(x) = f(t+x)$, which is then, as said above, a one-parameter group.

Problem 3.16 Under which conditions does a Banach space E have the decomposition property with respect to operators $T_1, \ldots, T_n$ coming from a one-parameter (semi)group T as $T_j = T(t_j)$ for some $t_j > 0$, $j = 1, \ldots, n$?

A one-parameter (semi)group is called a C_0-*(semi)group* if it is strongly continuous, i.e., continuous into $\mathscr{L}(E)$ endowed with the strong (i.e., pointwise) operator topology. The translation group is not strongly continuous on the Banach space $\mathrm{B}(\mathbb{R})$ of bounded functions or on $\mathrm{BC}(\mathbb{R})$, but it is strongly continuous on the Banach space $\mathrm{BUC}(\mathbb{R})$ of bounded unformly continuous functions. A one-parameter (semi)group is called bounded if $\|T(t)\| \leq M$ for all $t \in [0, \infty)$ (or $t \in \mathbb{R}$). See [6] for the general theory.

Theorem 3.17 (Kadets and Shumyatskiy [20]) *Let T be a bounded C_0-group, and let $t_1, t_2 > 0$. Then*

$$\ker(T(t_1) - \mathrm{I})(T(t_2) - \mathrm{I}) = \ker(T(t_1) - \mathrm{I}) + \ker(T(t_2) - \mathrm{I}). \tag{7}$$

Translations on $\mathrm{BUC}(\mathbb{R})$ is a C_0-group of isometries, providing another proof of the 2-decomposition property of $\mathrm{BUC}(\mathbb{R})$, formulated in Proposition 2.4.

In general the idea is to find a closed subspace $F \subseteq E$ invariant under the semigroup operators $T(t)$, such that one can apply Proposition 3.6 to the restricted operators. Concerning the nature of the problem there is one immediate candidate for this subspace. In what follows T will be a fixed bounded C_0-semigroup. A vector $x \in E$ is called *asymptotically almost periodic* (with respect to the semigroup T) if the orbit $\{T(t)x : t \geq 0\}$ is relatively compact in E. Denote by $E_{\mathrm{aap}}$the collection of asymptotically almost periodic vectors, which is easily seen to be a closed subspace of E invariant under the semigroup operators. It can be proved that if T is a bounded C_0-group then for $x \in E_{\mathrm{aap}}$ one actually has also the relative compactness of the entire orbit $\{T(t)x : t \in \mathbb{R}\}$. The proof of Theorem 3.17 by Kadets and Shumyatskiy establishes the fact that $\ker(T(t_1) - \mathrm{I})(T(t_2) - \mathrm{I}) \subseteq E_{\mathrm{aap}}$.

The only known extensions/variations of the Kadets–Shumyatskiy result follow the same strategy (or some modifications of it) and are the following:

Theorem 3.18 (Farkas [8]) *Let E be a Banach space and let T be a bounded C_0-group. Suppose that E does not constain an isomorphic copy of the Banach space c_0 of null sequences. Then for every $n \in \mathbb{N}$ and $t_1, \ldots, t_n \in \mathbb{R}$ we have*

$$\ker(T(t_1)-I)\cdots(T(t_n)-I)=\ker(T(t_1)-I)+\cdots+\ker(T(t_n)-I). \qquad (8)$$

It is not surprising that Bohl–Bohr–Kadets type theorems (see [1, 18]) play an important role here. In this regard let us mention just the following:

Theorem 3.19 (Basit [1], Farkas [7]) *A separable Banach space E does not contain an isomorphic copy of c_0 if and only if for every $x \in E$, $T \in \mathscr{L}(E)$ invertible with T and T^{-1} both power bounded the following statements are equivalent:*

(i) $\{T^{n+1}x - T^n x : n \in \mathbb{N}\}$ *is relatively compact.*
(ii) $\{T^{n+m}x - T^n x : n \in \mathbb{N}\}$ *is relatively compact for some $m \in \mathbb{N}$, $m \geq 1$.*
(iii) $\{T^{n+m}x - T^n x : n \in \mathbb{N}\}$ *is relatively compact for all $m \in \mathbb{N}$.*
(iv) $\{T^n x : n \in \mathbb{N}\}$ *is relatively compact.*

The next class of C_0-semigroups for which the decomposition problem has positive solution is of those that are *norm-continuous at infinity*, including also *eventually norm-continuous* semigroups, see [31] or [6, Sect. 2.1] for these notions.

Theorem 3.20 (Farkas [8]) *Let T be a bounded C_0-semigroup that is norm-continuous at infinity. Then for all $n \in \mathbb{N}$ and $t_1, \ldots, t_n \geq 0$ (8) holds.*

Problem 3.21
1. Is the Kadets–Shumyatskiy theorem true for every n? I.e., can one drop the geometric assumptions on E from Theorem 3.18?
2. What about the case of C_0-semigroups? Can one get rid of the eventual norm-continuity in Theorem 3.20?
3. None of the above covers the decomposition property of $\mathrm{BC}(\mathbb{R})$. What can be said about one-parameter semigroups that are only strongly continuous with respect to some weaker topology on the Banach space E? Can one cover the decomposition property of $\mathrm{BC}(\mathbb{R})$ by some extension of the results for one-parameter semigroups?

4 Application of the Operator Theoretic Results

In this section we present some applications of the results in the foregoing section.

4.1 The decomposition property of BC($\mathbb{R}$)

We devote this subsection to the proof of Theorem 2.7. We slightly differ from the original proof of [29], in exploiting the previous results and in particular Proposition 2.1.

For $n = 1$ the statement is trivial, so we argue by induction. Suppose $f \in \mathrm{BC}(\mathbb{R})$ satisfies (2). We group the periods according to commensurability:

$$\{\alpha_1,\dots,\alpha_n\}=\{\alpha_1,\dots,\alpha_a\}\cup\{\beta_1,\dots,\beta_b\}\cup\dots\cup\{\rho_1,\dots,\rho_r\}.$$

Denote the *least common multiple* of these by $\alpha,\beta,\dots,\rho$, i.e., α is the non-negative generator of the cyclic group $\bigcap_{j=1}^{a}\alpha_j\mathbb{Z}$ etc. Then from (2) we obtain

$$\Delta_\alpha^a\dots\Delta_\rho^r f=0. \tag{9}$$

Lemma 4.1 *Let* $f\in \mathrm{B}(\mathbb{R})$ *(a bounded function) and* $\alpha\in\mathbb{R}\setminus\{0\}$*,* $n\in\mathbb{N}$*. If* $\Delta_\alpha^n f=0$*, then* $\Delta_\alpha f=0$.

Proof Obviously, it suffices to work out the proof for $n=2$. Let $g:=\Delta_\alpha f$. By condition, $\Delta_\alpha g=0$, so g is α-periodic. Therefore,

$$f(x+N\alpha)=f(x)+\sum_{i=0}^{N-1}\Delta_\alpha f(x+i\alpha)=f(x)+Ng(x),$$

thus f cannot be bounded if $g(x)\neq 0$. □

As a consequence, from (9) we obtain

$$\Delta_\alpha\dots\Delta_\rho f=0. \tag{10}$$

Hence in case $\alpha_1,\dots,\alpha_n$ are not all pairwise incommensurable then f is also a solution of a difference equation of order less than n. We can therefore apply the induction hypothesis providing that f has an $(\alpha,\dots,\rho)$-decomposition. So in particular $f\in \mathrm{UAP}(\mathbb{R})$, which space has the decomposition property in view of Proposition 2.1, and so we are done.
It remains to handle the case when $\alpha_1,\dots,\alpha_n$ are pairwise incommensurable. The crux of the proof is thus the following:

Lemma 4.2 *Let* $\alpha_1,\dots,\alpha_n$ *be pairwise incommensurable, and let* $f\in \mathrm{BC}(\mathbb{R})$ *satisfy* (2)*. Then* f *has an* $(\alpha_1,\dots,\alpha_n)$*-decomposition in* $\mathrm{BC}(\mathbb{R})$.

To prove this lemma it is natural to get rid of one period and reduce the situation to a difference equation of order $n-1$ by considering $g:=\Delta_{\alpha_n}f$, which then satisfies $\Delta_{\alpha_1}\dots\Delta_{\alpha_{n-1}}g=0$, and thus by the induction hypothesis, by Remark 3.8(a) and by Example 3.7

$$g=g_1+\dots+g_{n-1}\qquad(\Delta_{\alpha_j}g_j=0,\ j=1,\dots,n-1),$$

where $g_j=Q_jg$ for some bounded projection Q_j on $\mathrm{UAP}(\mathbb{R})$. If f were subject to the representation (1), then we could guess $\Delta_{\alpha_n}f_j=g_j$. So we try to "lift up" the g_j to some functions f_j with $\Delta_{\alpha_j}f_j=\Delta_{\alpha_j}g_j=0$ and $\Delta_{\alpha_n}f_j=g_j$. Suppose this works, we find such $f_j\in \mathrm{BC}(\mathbb{R})$. Then

$$f_n:=f-(f_1+\dots+f_{n-1})\in \mathrm{BC}(\mathbb{R}),$$

and $\Delta_{\alpha_n} f_n = g - (g_1 + \cdots + g_{n-1}) = 0$, so f has a decomposition (1). So it remains to show the possibility of a lift-up for any incommensurable periods.

Lemma 4.3 *Let* $g \in \mathrm{C}(\mathbb{R})$, *let* $\beta, \gamma \in \mathbb{R}$ *be incommensurable, and suppose* $\Delta_\beta g = 0$. *Then the following are equivalent:*

(i) There exists $K > 0$ *such that*

$$\Big|\sum_{i=0}^{k-1} g(x + i\gamma)\Big| < K \quad (\textit{for } x \in \mathbb{R}, k \in \mathbb{N}).$$

(ii) There is $h \in \mathrm{C}(\mathbb{R})$ *such that* $\Delta_\beta h = 0$ *and* $\Delta_\gamma h = g$.

Proof This is a special case of a well-known ergodic theory result, see [14, Theorem 14.11, p.135], as putting $Y := \mathbb{R}/\gamma\mathbb{Z}$, the homeomorphism $\Theta(x) := x + \beta$ mod γ has minimal orbit-closure Y for every x. □

To complete the proof of Theorem 2.7 we need to check that condition (i) in the preceding lemma is fulfilled. For $j \in \{1, \ldots, n-1\}$ the projection Q_j commutes with translations so that

$$\Big|\sum_{i=0}^{k-1} g_j(x + i\alpha_n)\Big| = \Big|\sum_{i=0}^{k-1} (Q_j g)(x + i\alpha_n)\Big|$$
$$= \Big|Q_j \sum_{i=0}^{k-1} g(x + i\alpha_n)\Big| = \Big|Q_j\big(f(x + k\alpha_n) - f(x)\big)\Big| \le 2\|Q_j\| \cdot \|f\|_\infty$$

for every $x \in \mathbb{R}$, $k \in \mathbb{N}$. The proof is hence complete.

4.2 Applications to L^p spaces

Let (X, Σ, μ) be a measure space. In this subsection our standing assumption is as follows:

Condition 4.4 *For* $j = 1, \ldots, n$ *let* $T_j : X \to X$ *be pairwise commuting measurable mappings such that* $\mu(T_j^{-1}(A)) \le \mu(A)$ *for every* $A \in \Sigma$.

Then the Koopman operators, denoted by the same letter and defined by

$$T_j f := f \circ T_j$$

are contractions on all of the spaces $\mathrm{L}^p(X, \Sigma, \mu)$. In particular the condition above is fulfilled if the T_js are measure-preserving, in which case the Koopman operators T_j become isometries on each of the L^p spaces.

For the reflexive range the next corollary of Proposition 3.6 is immediate:

Corollary 4.5 *Let* $1 < p < \infty$. *Under Condition 4.4 consider the Koopman operators* T_j *on* $\mathrm{L}^p(X, \Sigma, \mu)$. *Then* (6) *holds true.*

The same result is true for the case $p = 1$, but the proof is different since infinite dimensional L^1 spaces are non-reflexive. We remark however that if (X, Σ, μ) is finite, then the Koopman operators T_j are simultaneous L^1 and L^∞ contractions, so-called Dunford–Schwartz operators, that are known to be mean ergodic on L^1, see, e.g., [5, Sect. 8.4].

Proposition 4.6 (Laczkovich and Révész [30]) *Under Condition 4.4 consider the Koopman operators* T_j *on* $\mathrm{L}^1(X, \Sigma, \mu)$. *Then* (6) *holds true.*

We do not give the proof here, but note that the mean ergodicity of the operators can be replaced by an application of Birkhoff's pointwise ergodic theorem, see, e.g. [5, Chap. 11]. See [30] for the detailed proof.

The case of $p = \infty$ is more subtle. Let us recall the following notion.

Definition 4.7 A measure space (X, Σ, μ) is called *localizable* if the dual of the Banach space $\mathrm{L}^1(X, \Sigma, \mu)$ is $\mathrm{L}^\infty(X, \Sigma, \mu)$ (with the usual identification).

As a matter of fact, the original definition of Segal (see [36, Sect. 5]) was different, but is equivalent to the one above. Known examples of localizable measure spaces include:

Example 4.8

1. σ-finite measure spaces,
2. (X, Σ, μ) with X a set $\Sigma = \mathscr{P}(X)$ the power set, μ the counting measure,
3. (X, Σ, μ) purely atomic,
4. (X, Σ, μ), X a locally compact group, Σ the Baire algebra, μ a (left/right) Haar measure.

Hence, in all of these cases the results below apply. In particular if one considers commuting left- (or right) translations on some locally compact group G, then the respective Koopman operators will satisfy (6). Note that the left and the right Haar measures are absolutely continuous with respect to each other, so we can fix each and any of them for our considerations below.

Theorem 4.9 (Laczkovich and Révész [30]) *Let* (X, Σ, μ) *be a localizable measure space, and suppose that for the pairwise commuting measurable mappings* $T_j : X \to X$ $(j = 1, \dots, n)$ *the push-forward measures* $\mu \circ T_j^{-1}$ *are all absolutely continuous with respect to* μ. *Then for the Koopman operators* T_j *on* $\mathrm{L}^\infty(X, \Sigma, \mu)$ (6) *holds true.*

The proof relies on the fact that under the conditions of localizability of (X, Σ, μ) and absolute continuity of the push-forward measures, the operators T_j will be weak* continuous on $\mathrm{L}^\infty(X, \Sigma, \mu)$ hence one can apply Proposition 3.5. For the details see [30].

Problem 4.10 Can one drop the localizability assumption?

Corollary 4.11 (Gajda [13], Laczkovich and Révész [30]) *The space* $\mathrm{B}(X)$ *of bounded functions on a set* X *has the decomposition property with respect to any system of commuting Koopman operators.*

This follows from Theorem 4.9 and from Example 4.8(2) above. The proof of Gajda uses Banach limits, see also Sect. 7. To sum up we have:

Corollary 4.12 *The Banach spaces* $\mathrm{L}^p(\mathbb{R})$ *(*$1 \leq p \leq \infty$*, Lebesgue measure) have the decomposition property.*

Of course, 0 is the only periodic function in $\mathrm{L}^p(\mathbb{R})$ if $p < \infty$, hence the message of the previous result is that (2) has 0 as the only L^p-solution if $p < \infty$. This follows also from a more general result of Edgar and Rosenblatt [4, Corollary 2.7] stating that the translates of a function $0 \neq f \in \mathrm{L}^p(\mathbb{R}^d)$, $p < 2d/(d-1)$ are linearly independent.

4.3 More Spaces with the Decomposition Property

Proposition 4.13 (Laczkovich and Révész [30]) *The following spaces of real-valued functions on* $\mathbb{R}$ *have the decomposition property:*

(a) $\mathrm{BV}^1(\mathbb{R}) := \{f : f \in \mathrm{B}(\mathbb{R}) \text{ with unif. bdd. variation on } [x, x+1], x \in \mathbb{R}\}$
(b) $\mathrm{Lip}_b(\mathbb{R}) := \{f : f \text{ is bounded and Lipschitz continuous}\}$
(c) $\mathrm{Lip}_b^k(\mathbb{R}) := \{f : f \in \mathrm{BC}(\mathbb{R}) \text{ } k \text{ times differentiable with } f^{(k)} \text{ Lipschitz}\}$

The cases (a) and (b) can be handled by introducing an appropriate norm turning the spaces under consideration into Banach spaces, then by noting that the unit ball is compact for the pointwise topology. Hence Theorem 3.3 is applicable. Details are in [30]. Part (c) relies on the following result interesting in its own right:

Proposition 4.14 (Laczkovich and Révész [30]) *Let* $\mathcal{F} \subseteq \mathrm{C}(\mathbb{R})$ *be a function class with the property that whenever* $f \in \mathcal{F}$ *and* $c \in \mathbb{R}$ *then* $f + c \in \mathcal{F}$*. Let* $k \in \mathbb{N}$ *and define*

$$\mathcal{G} := \{f : f \in \mathrm{BC}(\mathbb{R}) \text{ is } k \text{ times differentiable with } f^{(k)} \in \mathcal{F}\}.$$

If the function class $\mathcal{F}$ *has the decomposition property so does* $\mathcal{G}$*.*

Problem 4.15 There are several interesting Banach function spaces. Which of them do have the decomposition property? Just take your favorite non-reflexive translation invariant Banach function space on $\mathbb{R}$. Does it have the decomposition property? Denote by $\mathrm{L}_p^1(\mathbb{R})$ the set of functions with

$$\|f\|_{1,p} := \sup_{x \in \mathbb{R}} \left(\int_x^{1+x} |f(t)|^p \mathrm{d}t \right)^{1/p} < \infty,$$

and by $S^p(\mathbb{R})$ the closure of trigonometric polynomials in this norm. The elements of $S^p(\mathbb{R})$ are called *Stepanov almost periodic functions*, see [2]. Does the Banach space $L^1_p(\mathbb{R})$ have the decomposition property? If the answer were affirmative it would follow that $f \in L^1_p(\mathbb{R})$ and (2) imply that $f \in S^p(\mathbb{R})$. (This is because periodic functions belong to $S^p(\mathbb{R})$.) So, is an $L^1_p(\mathbb{R})$ solution of (2) Stepanov almost periodic?

5 Results for Arbitrary Transformations

Treating the periodic decomposition problem for various classes $\mathcal{F}$ of real functions a natural approach would be to split the question into two. That is first looking for a periodic decomposition into arbitrary periodic functions with the given periods, and then investigating whether the existence of such arbitrary decomposition entails the existence of a decomposition within the function class $\mathcal{F}$. In this section we address the first question, which can be actually done in a far more general setup. We consider this problem interesting in its own right, even if we already know that for some important function classes (e.g., for $C(\mathbb{R})$, see the paragraph after Proposition 1.3) the answer to the second question is in the negative.

Let X be a non-empty set. The decomposition problem can be formulated in *the whole space of functions* $\mathbb{R}^X$ with respect to *arbitrary commuting transformations* in X^X. To do that to a self map $T : X \to X$, called *transformation*, we associate the Koopman operator (denoted by the same letter) $Tf := f \circ T$, and the *T-difference operator* $\Delta_T f := Tf - f$. A function f satisfying $\Delta_T f = 0$ is then called *T-invariant*. A *$(T_1, \dots, T_n)$-invariant decomposition* of some function f is a representation

$$f = f_1 + \cdots + f_n, \qquad \text{where} \qquad \Delta_{T_j} f_j = 0 \quad (j = 1, \dots, n). \tag{11}$$

For pairwise commuting transformations T_i the functional equation

$$\Delta_{T_1} \dots \Delta_{T_n} f = 0 \tag{12}$$

is evidently necessary for the existence of invariant decompositions. On the example of translations on $\mathbb{R}$ we saw that it is not sufficient. Now in this general setting our basic question sounds:

Problem 5.1 Give necessary and sufficient conditions, containing (12), in order to have some $(T_1, \dots, T_n)$-invariant decomposition (11). Or give restrictions either on the transformations or on X (but not on the function class $\mathbb{R}^X$) such that (12) becomes also sufficient.

More precisely, we focus on complementary conditions, functional equations, on the functions, which they must satisfy in case of existence of an invariant decomposition (11) and which equations will also imply existence of such a decomposition.

Difference equations (of higher order) and/or inequalities occur here naturally, as is also suggested by the appearance of the Whitney condition in Theorem 1.6.

Further necessary conditions can be easily obtained. Indeed, as the transformations commute, (12) implies

$$\Delta_{T_1^{k_1}} \dots \Delta_{T_n^{k_n}} f = 0 \qquad (\forall k_1, \dots, k_n \in \mathbb{N}). \tag{13}$$

Now the major difficulties come from the following features:

1. The transformations T_j may not be invertible.
2. The "mix-up" of transformations can be completely irregular: $T^5 S^3 x = T^7 S^2 x$ for some $x \in X$ and nothing similar for other points $y \in X$.
3. Functions on X lack any structure beyond the obvious linear one—no boundedness, continuity, measurability, compatibility with underlying structure of X, nothing—so not much theoretical mathematics but pure combinatorics can be invoked.

For two transformations, i.e., $n = 2$, the answer is completely known:

Theorem 5.2 (Farkas and Révész [10]) *Let X be a non-empty set, let $S, T : X \to X$ be commuting transformations, and let $f \in \mathbb{R}^X$. The following are equivalent:*

(i) There exists a decomposition $f = g+h$, with g and h being S- and T-invariant, respectively.

(ii) $\Delta_S \Delta_T f = 0$, and if for some $x \in X$ and $k, n, k', n' \in \mathbb{N}$ the equality

$$T^k S^n x = T^{k'} S^{n'} x \tag{14}$$

holds, then

$$f(T^k x) = f(T^{k'} x).$$

(iii) $\Delta_S \Delta_T f = 0$, and if for some $x \in X$ and $k, n, k', n' \in \mathbb{N}$ (14) holds, then

$$f(S^n x) = f(S^{n'} x).$$

Of course, the equivalence of (ii) and (iii) is due to symmetry, if one knows that any one of them is equivalent to (i). We do not give the proof (see [10]), but mention an idea that will be useful also below. First we partition the set X with respect to an *equivalence relation*: $x, y \in X$ are equivalent if there exist $k, n, k', n' \in \mathbb{N}$ such that $T^k S^n x = T^{k'} S^{n'} y$. X splits into equivalence classes $X/\sim$, from which *by the axiom of choice* we choose a representation system. Obviously, it is enough to define g and h on each of these equivalence classes. Indeed, for $x \in X$ the elements x, Tx and Sx are all equivalent, so the invariance of the desired functions g, h is decided already in the common equivalence class. So the task is now reduced to defining the functions g and h on a fixed, but arbitrary equivalence class.

For general $n \in \mathbb{N}, n \geq 2$ the following difference equation type necessary conditions can be found:

Condition (*) *For every* $N \leq n$, *disjoint* N *-term partition* $B_1 \cup B_2 \cup \cdots \cup B_N = \{1, 2, \ldots, n\}$, *distinguished elements* $h_j \in B_j$ $(j = 1, \ldots, N)$*, indices* $0 < k_j, l_j, l'_j \in \mathbb{N}$, $(j = 1, \ldots, N)$ *and* $z \in X$ *once the conditions*

$$T_{h_j}^{k_j} T_i^{l_i} z = T_i^{l'_i} z \qquad \text{for all } i \in B_j \setminus \{h_j\}, \text{for all } j = 1, \ldots, N \tag{15}$$

are satisfied, then

$$\Delta_{T_{h_1}^{k_1}} \ldots \Delta_{T_{h_N}^{k_N}} f(z) = 0. \tag{16}$$

Theorem 5.3 (Farkas and Révész [10]) *Let* $T_1, \ldots, T_n$ *be commuting transformations of* X *and let* f *be a real function on* X*. In order to have a* $(T_1, \ldots, T_n)$*-invariant decomposition* (11) *of* f *Condition* $(*)$ *is necessary.*

If the blocks B_j are all singletons the condition (15) is empty, so (16) expresses exactly (13). In particular, Condition $(*)$ contains (12).
For $n = 3$ transformations Condition $(*)$ is not only necessary but also sufficient for the existence of invariant decompositions.

Theorem 5.4 (Farkas and Révész [10]) *Suppose that* T_1*,* T_2 *and* T_3 *commute and that the function* f *satisfies Condition* $(*)$*. Then* f *has a* (T_1, T_2, T_3)*-invariant decomposition.*

Again the proof is combinatorially involved, so let us just state one main ingredient, the "lift-up lemma" corresponding to Lemma 4.3 above. It is proved itself in a series of lemmas, which we do not detail here.

Lemma 5.5 *Let* T, S *be commuting transformations of* X *and let* $g : X \to \mathbb{R}$ *be a function satisfying* $\Delta_S g = 0$*. Then there exists a function* $h : X \to \mathbb{R}$ *satisfying both* $\Delta_S h = 0$ *and* $\Delta_T h = g$ *if and only if for every* $x \in X$ *it holds*

$$\sum_{i=0}^{k-1} g(T^i x) = 0 \quad \textit{whenever} \quad T^k S^l x = S^{l'} x \textit{ with some } k, l, l' \in \mathbb{N}. \tag{17}$$

Problem 5.6 Is Condition $(*)$ equivalent to (11) for all $n \in \mathbb{N}$ $(n \geq 4)$?

5.1 Unrelated Transformations

If the orbits of the transformations show no recurrence then a satisfactory answer can be given. The relevant notion is the following.

Definition 5.7 We call two commuting transformations S, T on X *unrelated* if $T^n S^k x = T^m S^l x$ can occur only if $n = m$ and $k = l$. In particular, then neither of the two transformations can have any cycles in their orbits, nor do their joint orbits have any recurrence.

If all pairs T_i and T_j ($1 \le i \ne j \le n$) are unrelated, then Condition (∗) degenerates, as in (15) we necessarily have that all blocks B_j are singletons. Hence Condition (∗) reduces merely to (13) or, equivalently, to (12).

Theorem 5.8 (Farkas and Révész [10]) *Suppose the transformations $T_1, \ldots, T_n$ are pairwise commuting and unrelated. Then the difference equation* (12) *is equivalent to the existence of some invariant decomposition* (11).

Proof Only sufficiency is to be proved. We argue by induction. The cases of small n are obvious. Let $F := \Delta_{T_{n+1}} f$. Then F satisfies a difference equation of order n, hence by the inductive hypothesis we can find an invariant decomposition of F in the form $F = F_1 + \cdots + F_n$, where $\Delta_{T_j} F_j = 0$ for $j = 1, \ldots, n$. Since the transformation are unrelated, condition (17) in Lemma 5.5 is void, and therefore the "lift-ups" f_j with $\Delta_{T_j} f_j = 0$, $\Delta_{T_{n+1}} f_j = F_j$ exist for all $j = 1, \ldots, n$. Therefore, $f_{n+1} := f - (f_1 + \cdots + f_n)$ provides a function satisfying $\Delta_{T_{n+1}} f_{n+1} = F - (F_1 + \cdots + F_n) = 0$. Thus a required decomposition of f is established. □

5.2 Invertible Transformations

When the transformations T_j are invertible, the situation simplifies somewhat. Denote by $G \subseteq X^X$ the group generated by $T_1, \ldots, T_n$. As before, we work on equivalence classes, now *orbits* $O := \{Tx : T \in G\}$ for some $x \in X$, under the action of the transformation group G. Given a group G denote by $\langle a \rangle$ the cyclic group generated by a i.e., $\langle a \rangle := \{a^n : n \in \mathbb{Z}\}$, and for $H \subseteq G$ let $[H] := \bigcap_{h \in H} \langle h \rangle$.

Condition(**) For all orbits O of G, for all partitions

$$B_1 \cup B_2 \cup \cdots \cup B_N = \left\{T_1 \mid_O, T_2 \mid_O, \ldots, T_n \mid_O\right\}$$

and any element $S_j \in [B_j]$, $j = 1, \ldots, N$, we have that

$$\Delta_{S_1} \ldots \Delta_{S_N} f \mid_O = 0 \quad \text{holds.} \tag{18}$$

The next is the main result in this setting:

Theorem 5.9 (Farkas et al. [9]) *Let $T_1, \ldots, T_n$ be pairwise commuting invertible transformations on a set X. Let $f : X \to \mathbb{R}$ be any function. Then f has a $(T_1, T_2, \ldots, T_n)$-invariant decomposition* (11) *if and only if it satisfies Condition (**).*

The proof relies on a variant of Lemma 5.5.

6 Decompositions on Groups

Let us see some consequences of the result in the foregoing section. Let G be a group, and let $a_1, \dots, a_n \in G$. Consider the actions of $a_1, \dots, a_n$ on G as left multiplications. For a function $f : G \to \mathbb{R}$ we introduce the *left a-difference operator* $\Delta_a f(x) := f(ax) - f(x)$. The function f is called *left a-invariant* (or left a-periodic) if $\Delta_a f = 0$. Since the actions are transitive we get:

Corollary 6.1 *Let G be a group and $a_1, \dots, a_n \in G$ pairwise commuting. Then a function $f : G \to \mathbb{R}$ decomposes into a sum of left a_j-invariant functions, $f = f_1 + \dots + f_n$, if and only if for all partitions $B_1 \cup B_2 \cup \dots \cup B_N = \{a_1, \dots, a_n\}$ and for each element $b_j \in [B_j]$*

$$\Delta_{b_1} \dots \Delta_{b_N} f = 0.$$

In a torsion free Abelian group A for $B \subseteq A$ the generator of the cyclic group $[B]$ is uniquely determined (up to taking inverse). In [10] we called this (maybe two) element(s) the *least common multiple* of the elements in B. For instance, with this terminology we have that the least common multiple of 1 and $\sqrt{2}$ in the group $(\mathbb{R}, +)$ is 0. Then we have the next result:

Corollary 6.2 *Let A be a torsion free Abelian group and $a_1, \dots, a_n \in A$. A function $f : A \to \mathbb{R}$ decomposes into a sum of a_j-periodic functions, $f = f_1 + \dots + f_n$, if and only if for all partitions $B_1 \cup B_2 \cup \dots \cup B_N = \{a_1, \dots, a_n\}$ and b_j being the least common multiple of the elements in B_j one has*

$$\Delta_{b_1} \dots \Delta_{b_N} f = 0. \tag{19}$$

If we take $A = \mathbb{R}$ and $\alpha_1, \dots, \alpha_n$ incommensurable we obtain the following.

Corollary 6.3 (Mortola and Peirone [32], Farkas and Révész [10]) *Let $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ be incommensurable. Then a function $f : \mathbb{R} \to \mathbb{R}$ satisfies* (2) *if and only if it has periodic decomposition* (1).

The above results remain true if one considers functions with values in torsion free groups Γ. The proof of the following is the same as for Theorem 5.9 with the new aspect that taking averages in Γ requires some additional care.

Theorem 6.4 (Farkas et al. [9]) *Let A, Γ be torsion free Abelian groups and $a_1, \dots, a_n \in A$. A function $f : A \to \Gamma$ decomposes into a sum of a_j-periodic functions $f_j : A \to \Gamma$, $f = f_1 + \dots + f_n$ if and only if for all partitions $B_1 \cup B_2 \cup \dots \cup B_N = \{a_1, \dots, a_n\}$ and b_j being the least common multiple of the elements in B_j one has* (19).

Let A be a torsion free Abelian group. By the previous theorem for $\Gamma = \mathbb{R}$ and for $\Gamma = \mathbb{Z}$, we obtain that for a function $f : A \to \mathbb{Z}$ the existence of a real-valued periodic decomposition and the existence of an integer-valued periodic decomposition are both equivalent to the same condition.

Corollary 6.5 *If an integer-valued function f on a torsion free Abelian group A decomposes into the sum of a_j-periodic real-valued functions for some $a_1, \ldots, a_n$, then f decomposes into the sum of a_j-periodic integer-valued ones.*

There are examples showing that one cannot get rid of the torsion freeness of A in Corollary 6.5 or Theorem 6.4, see [9].

Note that in crystallography and other applications, reconstruction or at least unique identification of integer-valued functions or characteristic functions of sets from various (partial) information concerning their Fourier transform are rather important. This also motivates the interest in integer-valued periodic decompositions or decompositions with values within a subgroup. In turn, support of a Fourier transform can reveal the existence of a periodic decomposition, see e.g. [27, 2.7 and 2.8], or the analogous idea of the proof for Proposition 2.1. For more about this see [27] and the references therein.

7 Actions of Semigroups

Let X be a non-empty set and let $T : X \to X$ be an arbitrary mapping. If a function $f : X \to \mathbb{R}$ is invariant under T, i.e., $\Delta_T f = 0$, then it is evidently invariant under each iterate T^n of T for $n \in \mathbb{N}$. Given commuting mappings $T_1, \ldots, T_n : X \to X$ consider the generated semigroups

$$S_j := \left\{T_j^n : n \in \mathbb{N}\right\}. \tag{20}$$

The corresponding semigroup of the Koopman operators on $\mathbb{R}^X$ is denoted by $\mathscr{S}_j$. (Recall that we use the same symbol T for the Koopman operator of $T \in X^X$.) For a subset $\mathscr{A}$ of linear operators on $\mathbb{R}^X$ we introduce the notations $\ker \mathscr{A} := \bigcap_{A \in \mathscr{A}} \ker A$. Then the equality

$$\ker(T_1 - \mathrm{I}) \cdots (T_n - \mathrm{I}) = \ker(T_1 - \mathrm{I}) + \cdots + \ker(T_n - \mathrm{I}) \tag{21}$$

is easily seen to be equivalent to

$$\ker(\mathscr{S}_1 - \mathrm{I}) \cdots (\mathscr{S}_n - \mathrm{I}) = \ker(\mathscr{S}_1 - \mathrm{I}) + \cdots + \ker(\mathscr{S}_n - \mathrm{I}). \tag{22}$$

In what follows we study this equality when $\mathscr{S}_j$ are general, not necessarily cyclic, semigroups.

Let S be a discrete semigroup with unit element, and let S_j, $j = 1, \ldots, n$ unital subsemigroups of S that all act on the non-empty set X (from the left), the unit acting as the identity. Suppose furthermore $st = ts$ for all $s \in S_j$ and $t \in S_i$ with $i \neq j$ (the actions of different S_js are commuting).

Theorem 7.1 (Farkas [8]) *Suppose that for $j = 1, \ldots, n$ the unital semigroups S_j on the set X are (right-)amenable and that the actions of the different S_j are commuting. Denote by $\mathscr{S}_j$ the semigroups of the Koopman operators. Then* (22) *holds in the space* $\mathrm{B}(X)$. *Furthermore, if X is uniform (topological) space and the action of S_j on X is uniformly equicontinuous, then* (22) *holds in the space* $\mathrm{BUC}(X)$.

This result and its proof generalizes those of Gajda [13], who used Banach limits (i.e., amenability of $\mathbb{Z}$ or $\mathbb{N}$) to establish the above for $\mathbb{Z}$ and $\mathbb{N}$ actions, i.e., for semigroups as in (20). The next consequence immediately follows.

Corollary 7.2 (Gajda [13]) *Let A be a locally compact Abelian group, and let $a_1, \ldots, a_n \in A$. Then* (21) *holds in* $\mathrm{BUC}(\mathbb{R})$ *for T_j being the shift operator by a_j. In particular* $\mathrm{BUC}(\mathbb{R})$ *has the decomposition property.*

Let us finally return to the purely linear operator setting on an arbitrary Banach space E. A subsemigroup $\mathscr{S} \subseteq \mathscr{L}(E)$ of bounded linear operators is called *mean ergodic* if the closed convex hull $\overline{\mathrm{conv}}(\mathscr{S}) \subseteq \mathscr{L}(E)$ contains a zero element P, i.e., $PT = P = TP$ for every $T \in \mathscr{S}$. In this case P is a projection, called the *mean ergodic projection* of $\mathscr{S}$, and it holds (see [33])

$$E = \mathrm{rg}P \oplus \mathrm{rg}(\mathrm{I} - P) \quad \text{with} \quad \mathrm{rg}P = \ker(\mathscr{S} - \mathrm{I})\ .$$

Theorem 7.3 (Farkas [8]) *Let $\mathscr{S}_1, \mathscr{S}_2, \ldots, \mathscr{S}_n \subseteq \mathscr{L}(E)$ be mean ergodic operator semigroups and suppose that $ST = TS$ whenever $T \in \mathscr{S}_i$, $S \in \mathscr{S}_j$ with $i \neq j$. Then* (22) *holds.*

Since an operator T is mean ergodic if and only if the semigroup $\{T^n : n \in \mathbb{N}\}$ is mean ergodic, the previous result contains Proposition 3.6. Moreover, the obvious modification of Theorem 3.3 (using fixed points in the closed convex hull) for this semigroup setting is easily proved, but this we will not pursue here. Furthermore, the analogue of Corollary 3.4 can be formulated for amenable semigroups instead of cyclic ones, where of course one applies Day's fixed point theorem, see [3], instead of the one of Markov and Kakutani.

Problem 7.4 Does the space $\mathrm{BC}(A)$ of bounded and continuous functions, where A is a locally compact Abelian group, has the decomposition property with respect to translations? If A is compact or discrete or $A = \mathbb{R}$, this is so by the previous results. What about $A = \mathbb{R}^2$?

8 Further Results

We briefly touch upon topics that, regrettably, could not be covered in detail.

First we take a second glimpse at the original problem.

Theorem 8.1 (Natkaniec and Wilczyński [34]) *Let $\alpha_1, \ldots, \alpha_n \in \mathbb{R} \setminus \{0\}$ be incommensurable. A function $f : \mathbb{R} \to \mathbb{R}$ has a decomposition* (1) *with $f_1, \ldots, f_n$ Darboux functions if and only if* (2) *holds.*

See [34] for the proof where also the decomposition property of Marczewski measurable functions is studied for incommensurable periods. It is also shown that the identity is not the sum of periodic functions having the Baire property. For classes of measurable functions we have, e.g., the following.

Theorem 8.2 (Keleti [26]) *None of the following classes $\mathcal{F}$ have the decomposition property:*

(a) $\mathcal{F} = \{f : f : \mathbb{R} \to \mathbb{Z},\ f \in \mathrm{L}^\infty(\mathbb{R})\}$,
(b) $\mathcal{F} = \{f : f : \mathbb{R} \to \mathbb{Z}$ *is bounded measurable*$\}$,
(c) $\mathcal{F} = \{f : f : \mathbb{R} \to \mathbb{R}$ *is a.e. integer-valued and* $f \in \mathrm{L}^\infty(\mathbb{R})\}$,
(d) $\mathcal{F} = \{f : f : \mathbb{R} \to \mathbb{Z}$ *is a.e. integer-valued, bounded and measurable*$\}$.

For more information on measurable decompositions see also [23–25]. Next we turn to integer-valued decompositions on Abelian groups. We mention only three exemplary results from [22]:

Theorem 8.3 (Károlyi et al. [22])

(a) Suppose $f : \mathbb{Z} \to \mathbb{Z}$ has an $(\alpha_1, \ldots, \alpha_n)$-periodic decomposition into real-valued functions with $a_j \in \mathbb{Z}$. Then it has an $(\alpha_1, \ldots, \alpha_n)$-periodic integer-valued decomposition.
(b) For $\alpha_1, \ldots, \alpha_n \in \mathbb{Z}$ the class of $\mathbb{Z} \to \mathbb{Z}$ functions has the decomposition property.
(c) Let A be a torsion-free Abelian group. Then the class of bounded $A \to \mathbb{Z}$ functions has the decomposition property if and only if A is isomorphic to an additive subgroup of $\mathbb{Q}$.

For a proof and for an abundance of further information we refer to [22], and remark that part c) above implies that the class of bounded and integer-valued functions does not have the decomposition property known also from Theorem 8.2, see also [22, Corollary 3.4].
Finally, we discuss some aspects of uniqueness of decompositions. Of course, one cannot expect uniqueness in the original setting, since appropriate constant functions can be added to the summands in (1) not affecting the validity of (2). If one restricts to certain function classes then only this trivial procedure can produce different decompositions (for incommensurable periods).

Theorem 8.4 (Laczkovich and Révész [30]) *For incommensurable periods a periodic decomposition in $\mathrm{L}^\infty(\mathbb{R})$ of a function $f \in \mathrm{L}^\infty(\mathbb{R})$ is unique up to additive constants.*

In the original setting of the decomposition problem, i.e., in $\mathbb{R}^\mathbb{R}$ the situation is somewhat more complicated. E.g. consider $n = 2$, $f = f_1 + f_2$ with $\Delta_{a_j} f_j = 0$,

$j = 1, 2$. Let h be a not identically 0 function that is both α_1- and α_2-periodic. Then $f = (f_1 + h) + (f_2 - h)$ is a different decomposition.

In general two decompositions $f = g_1 + \cdots + g_n$ and $f = f_1 + \cdots + f_n$ with $\Delta_{\alpha_j} g_j = \Delta_{\alpha_j} f_j = 0$ $j = 1, \ldots, n$ are called *essentially the same* if there are functions $h_{ij} \in \mathbb{R}^{\mathbb{R}}$ for $i, j = 1, \ldots, n$ with $h_{ii} = 0$, $h_{ij} = -h_{ji}$, $\Delta_{\alpha_i} h_{ij} = 0$, $\Delta_{\alpha_j} h_{ij} = 0$ such that for all $j = 1, \ldots, n$ one has $f_j - g_j = \sum_{i=1}^{n} h_{ij}$. Note that for incommensurable periods $\alpha_i/\alpha_j \notin \mathbb{Q}$ we necessarily have h_{ij} =constant on each equivalence class of $\mathbb{R}$ (for the equivalence relation as in the paragraph after Theorem 5.2), whence in presence of continuity on the whole real line.

Essential uniqueness of decomposition depends very much on the periods:

Theorem 8.5 (Harangi [17]) *For $\alpha_1, \ldots, \alpha_n \in \mathbb{R} \setminus \{0\}$ the following assertions are equivalent:*

(i) If any three numbers $\alpha_i, \alpha_j, \alpha_k$ from $\alpha_1, \ldots, \alpha_n$ are pairwise linearly independent over $\mathbb{Q}$, then they are linearly independent over $\mathbb{Q}$.
(ii) Any two $(\alpha_1, \ldots, \alpha_n)$-periodic decomposition of a function f are essentially the same, i.e., the decomposition is essentially unique.
(iii) If a function $f : \mathbb{R} \to \mathbb{Z}$ has an $(\alpha_1, \ldots, \alpha_n)$-periodic decomposition into bounded real-valued functions, then it has also one into bounded integer-valued functions.

See also [15], [17] or [16] for details and further directions.

We end this survey by posing the following problem:

Problem 8.6 Study the periodic decomposition problem for functions f on $\mathbb{R}$, or on an Abelian group, with values in $\mathbb{R}$ mod 1 (or in an Abelian group).

References

1. Bazit, R.B.: Generalization of two theorems of M.I. Kadets concerning the indefinite integral of abstract almost periodic functions. Matematicheskie Zametki **9**(3), 311–321 (1971)
2. Besicovitch, A.S.: Almost Periodic Functions. Dover Publications Inc., New York (1955)
3. Day, M.M.: Fixed-point theorems for compact convex sets. Ill. J. Math. **5**, 585–590 (1961)
4. Edgar, G.A., Rosenblatt, J.M.: Difference equations over locally compact abelian groups. Trans. Amer. Math. Soc. **253**, 273–289 (1979)
5. Eisner, T., Farkas, B., Haase, M., Nagel, R.: Operator Theoretic Aspects of Ergodic Theory. Graduate Text in Mathematics. Springer, Berlin (2013, In Press)
6. Engel, K.-J., Nagel, R.: One-Parameter Semigroups for Linear Evolution Equations, Graduate Texts in Mathematics, vol. 194. Springer, New York (2000)
7. Farkas, B.: A Bohl-Bohr-Kadets type theorem characterizing Banach spaces not containing c_0 (2013). (Submitted); arxiv:1301.6250
8. Farkas, B.: A note on the periodic decomposition problem for semigroups (2014). (Submitted); arxiv:1401.1226
9. Farkas, B., Harangi, V., Keleti, T.: Révész, Sz.: Invariant decomposition of functions with respect to commuting invertible transformations. Proc. Amer. Math. Soc. **136**(4), 1325–1336 (2008)

10. Farkas, B.: Révész, Sz.: Decomposition as the sum of invariant functions with respect to commuting transformations. Aequationes Math. **73**(3), 233–248 (2007)
11. Fonf, V.P., Lin, M., Wojtaszczyk, P.: Ergodic characterizations of reflexivity of banach spaces. J. Funct. Anal. **187**(1), 146–162 (2001)
12. Fonf, V.P., Lin, M., Wojtaszczyk, P.: A non-reflexive banach space with all contractions mean ergodic. Isr. J. Math. **179**, 479–491 (2010)
13. Gajda, Z.: Note on decomposition of bounded functions into the sum of periodic terms. Acta Math. Hungar. **59**(1–2), 103–106 (1992)
14. Gottschalk, W.H., Hedlund, G.A.: Topological Dynamics. Amer. Math. Soc. Colloquium Publications. American Mathematical Society, Providence, R.I. (1955)
15. Harangi, V.: Periodic decompositions of functions. Real Anal. Exchange (31st Summer Symposium Conference), 185–188 (2007)
16. Harangi, V.: Sums of periodic functions. Master's thesis, Eötvös Loránd Univeristy, Budapest (2007). (In Hungarian)
17. Harangi, V.: On the uniqueness of periodic decomposition. Fund. Math. **211**(3), 225–244 (2011)
18. Kadec, M.I.: The integration of almost periodic functions with values in a Banach space. Funkcional. Anal. i Priložen. **3**(3), 71–74 (1969)
19. Kadets, V.M., Shumyatskiy, B.M.: Averaging technique in the periodic decomposition problem. Mat. Fiz. Anal. Geom. **7**(2), 184–195 (2000)
20. Kadets, V.M., Shumyatskiy, B.M.: Additions to the periodic decomposition theorem. Acta Math. Hungar. **90**(4), 293–305 (2001)
21. Kahane, J.P.: Lectures on Mean Periodic Functions. Tata Institute, Bombay (1959)
22. Károlyi, G., Keleti, T., Kós, G., Ruzsa, I.Z.: Periodic decomposition of integer valued functions. Acta Math. Hungar. **119**(3), 227–242 (2008)
23. Keleti, T.: Difference functions of periodic measurable functions. Ph.D. thesis
24. Keleti, T.: On the differences and sums of periodic measurable functions. Acta Math. Hungar. **75**(4), 279–286 (1997)
25. Keleti, T.: Periodic decomposition of measurable integer valued functions. Real Anal. Exchange (31st Summer Symposium Conference), 189–193 (2007)
26. Keleti, T.: Periodic decomposition of measurable integer valued functions. J. Math. Anal. Appl. **337**(2), 1394–1403 (2008)
27. Keleti, T. and Kolountzakis, M.: On the determination of sets by their triple correlation in finite cyclic groups. Online J. Anal. Comb. 1, 4, 18 pp. (electronic) (2006)
28. Laczkovich, M., Révész, Sz: Periodic decompositions of functions. Real Anal. Exch. **13**(1), 126–128 (1987)
29. Laczkovich, M.: Révész, Sz.: Periodic decompositions of continuous functions. Acta Math. Hungar. **54**(3–4), 329–341 (1989)
30. Laczkovich, M.: Révész, Sz.: Decompositions into the sum of periodic functions belonging to a given Banach space. Acta Math. Hungar. **55**(3–4), 353–363 (1990)
31. Martinez, J., Mazon, J.M.: C_0-semigroups norm continuous at infinity. Semigroup Forum **52**(2), 213–224 (1996)
32. Mortola, S., Peirone, R.: The sum of periodic functions. Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat. (8) **2**(2), 393–396 (1999)
33. Nagel, R.: Mittelergodische Halbgruppen linearer Operatoren. Ann. Inst. Fourier (Grenoble) **23**(4), 75–87 (1973)
34. Natkaniec, T., Wilczyński, W.: Sums of periodic Darboux functions and measurability. Atti Sem. Mat. Fis. Univ. Modena **51**, 369–376 (2003)
35. Schwartz, L.: Théorie générale des fonctions moyenne-périodiques. Ann. Math. **2**(48), 857–929 (1947)
36. Segal, I.E.: Equivalences of measure spaces. Amer. J. Math. **73**, 275–313 (1951)
37. Székelyhidi, L.: On mean periodic functions. Aequationes Math. **58**(3), 203–213 (1999)
38. Whitney, H.: On functions with bounded nth differences. J. Math. Pures Appl. **9**(36), 67–95 (1957)
39. Wierdl, M.: Continuous functions that can be represented as the sum of finitely many periodic functions. Mat. Lapok **32**(1–3), 107–113 (1981/1984)

Existence of Periodic and Almost Periodic Solutions of Discrete Ricker Delay Models

Yoshihiro Hamaya

Abstract The aim of this article is to investigate the sufficient conditions for the existence of periodic and almost periodic solutions of a generalized Ricker delay model,

$$N(n+1) = N(n)\exp\{f(n, N(n - r(n)))\},$$

when f are periodic and almost periodic functions in n, respectively, which appears as a model for dynamics with single species in changing periodic and almost periodic environments, by applying the technique of boundedness and stability conditions which derives the fixed point theorems and uniformly asymptotically stable of solutions for above equation, respectively. Moreover, we consider the existence of an almost periodic solution of the case where f has the Volterra term with an infinite delay.

1 Introduction

For ordinary differential equations and functional differential equations, the existence of periodic and almost periodic solutions of systems has been studied by many authors. One of the most popular methods is to find Liapunov functions/functionals [1, 3, 7, 11, 14] for boundedness and stability conditions. For the periodic functional difference equation, the existence of uniform bounded and uniform ultimately bounded solutions imply the existence of a periodic solution, see [9], However, for an almost periodic equation, the boundedness of solutions does not necessarily imply the existence of an almost periodic solution even for scalar differential equations with no delay. Recently, He et al. [10] have shown the existence of periodic and almost periodic solutions for a non-autonomous scalar delay differential equation of modeling single species dynamics in a temporally changing

Y. Hamaya (✉)
Department of Information Science, Okayama University of Science,
1-1 Ridai-chyo Kitaku, Okayama 700-0005, Japan
e-mail: hamaya@mis.ous.ac.jp

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_9

environment. Their results extend the results of Gopalsamy [3] to a non-autonomous differential equation by using the 3/2 stability conditions [13] when its almost periodic case. To the best of our knowledge, there are no relevant results on the existence of periodic and almost periodic solutions for discrete Ricker models by means of our approach of discrete boundedness and stability theorems [16]. We emphasize that our results extend [4, 10] as delay discrete periodic and almost periodic cases. Equation (27) does not contain a delay independent stabilising negative feedback term and hence (27) require a different approach from that in the literature [11]. In this paper, we discuss the existence of periodic and almost periodic solutions for a generalized non-autonomous discrete Ricker type difference equations with finite delay and infinite delay in f, respectively.

In what follows, we denote by R real Euclidean space, Z is the set of integers, $Z^+ := [0, \infty)$ is the set of nonnegative integers and $|\cdot|$ will denote the Euclidean norm in R. For any discrete interval $I \subset Z := (-\infty, \infty)$, we denote by $BS(I)$ the set of all bounded functions mapping I into R, and set $|\phi|_I = \sup\{|\phi(s)| : s \in I\}$. We introduce an almost periodic function $f(n, \phi) : Z \times BS \to R$. where BS is an open set in R. After, this BS is defined by $I = [-h, 0]$ for some $h > 0$.

Definition 1 $f(n, \phi)$ is said to be almost periodic in n uniformly for $\phi \in BS$, if for any $\varepsilon > 0$ and any compact set K in BS, there exists a positive integer $L(\varepsilon, K)$ such that any interval of length $L(\varepsilon, K)$ contains an integer τ for which

$$|f(n + \tau, \phi) - f(n, \phi)| \le \varepsilon$$

for all $n \in Z$ and all $\phi \in K$. Such a number τ in above inequality is called an ε-translation number of $f(n, \phi)$.

In order to formulate a property of almost periodic functions, which is equivalent to the above definition, we discuss the concept of the normality of almost periodic functions. Namely, Let $f(n, \phi)$ be almost periodic in n uniformly for $\phi \in BS$. Then, for any sequence $\{h_k^{'}\} \subset Z$, there exists a subsequence $\{h_k\}$ of $\{h_k^{'}\}$ and function $g(n, \phi)$ such that

$$f(n + h_k, \phi) \to g(n, \phi) \tag{1}$$

uniformly on $Z \times K$ as $k \to \infty$, where K is a compact set in BS. There are many properties of the discrete almost periodic functions [2, 11], which are corresponding properties of the continuous almost periodic functions $f(t, x) \in C(R \times D, R)$ [cf. [14]]. We shall denote by $T(f)$ the function space consisting of all translates of f, that is, $f_\tau \in T(f)$, where

$$f_\tau(n, \phi) = f(n + \tau, \phi), \qquad \tau \in Z. \tag{2}$$

Let $H(f)$ denote the uniform closure of $T(f)$ in the sense of (2). $H(f)$ is called the hull of f. In particular, we denote by $\Omega(f)$ the set of all limit functions $g \in H(f)$ such that for some sequence $\{n_k\}$, $n_k \to \infty$ as $k \to \infty$ and $f(n + n_k, \phi) \to g(n, \phi)$

uniformly on $Z \times S$ for any compact subset S in BS. By (1), if $f : Z \times BS \to R$ is almost periodic in n uniformly for $\phi \in BS$, so is a function in $\Omega(f)$.

The following concept of asymptotic almost periodicity was introduced by Frechet in the case of continuous function (cf.[14]).

Definition 2 $u(n)$ is said to be asymptotically almost periodic if it is a sum of a almost periodic function $p(n)$ and a function $q(n)$ defined on $I^* = [a, \infty) \subset Z^+$ which tends to zero as $n \to \infty$, that is,

$$u(n) = p(n) + q(n).$$

$u(n)$ is asymptotically almost periodic if and only if for any sequence $\{n_k\}$ such that $n_k \to \infty$ as $k \to \infty$ there exists a subsequence $\{n_k\}$ for which $u(n + n_k)$ converges uniformly on n; $a \leq n < \infty$.

2 Preliminary Lemma

In this paper, we shall consider a discrete non-autonomous Ricker delay difference equation of the form

$$N(n+1) = N(n) \exp\{f(n, N(n - r(n)))\} \quad \text{for} \quad n \geq n_0, n_0 \in Z. \tag{3}$$

As example of the type (3), we provide the following example

$$N(n+1) = N(n) \exp\left\{a + bN^p(n-r) - cN^q(n-r)\right\}.$$

We set the following assumptions for Eq. (3).

(H_1) f is continuous at second term defined on $[n_0, \infty) \times R^+$ to R, r defined on $[n_0, \infty) \subset Z$ to Z^+ and $0 < r^l \leq r(n) \leq r^L < \infty$ for $n \geq n_0$, and some $n_0 \in Z$.

$(H_{1'})$ f is almost periodic function in n uniformly for ϕ and it's continuous at second term defined on $[n_0, \infty) \times R^+$ to R. r defined on $[n_0, \infty) \subset Z$ to Z^+ and r is almost periodic in n, $0 < r^l = \liminf_{n\to\infty} r(n) \leq r(n) \leq \limsup_{n\to\infty} r(n) = r^L < \infty$ for $n \geq n_0$, and some $n_0 \in Z$.

(H_2) There exist continuous function $F_i : R^+$ to R $(i = 1, 2)$ such that $F_1(y) \leq f(n, y) \leq F_2(y)$ for $(n, y) \in [n_0, \infty) \times R^+$.

(H_3) There exist $\xi_i > 0 (i = 1, 2)$ such that

$$\begin{cases} F_i(\xi_i) = 0, \ (i = 1, 2), \\ F_i(x) > 0 \ \text{ for } \ x \in [0, \xi_i), \\ F_i(x) < 0 \ \text{ for } \ x \in (\xi_i, \infty). \end{cases}$$

We assume that the initial conditions associated with (3) are as follows:

$$N(s) = \phi(s) \geq 0, \ \ s \in \left[n_0 - r^L, n_0\right], \phi(n_0) > 0 \ \text{and} \ \phi \ \text{defined on} \ \left[n_0 - r^L, n_0\right]. \tag{4}$$

By (3) and (H_3), we can see that solutions of (3) satisfy

$$\phi(n_0)\exp\left[\sum_{j=n_0}^{n-1} F_1(N(j-r(j)))\right] \le N(n) \le \phi(n_0)\exp\left[\sum_{j=n_0}^{n-1} F_2(N(j-r(j)))\right] \quad n > n_0.$$

We show the following key lemma in which the solution of (3) is permanence [8, 9].

Lemma 1 *If the assumptions* (H_2) *and* (H_3) *hold, then there exists an* $n_1 \ge n_0$ *such that any solution* $N(n)$ *of* (3) *satisfies*

$$N(n) \le M \equiv \xi_2 e^{\alpha r^L} \qquad \text{for } n \ge n_1. \tag{5}$$

where $\alpha = \max_{0\le x\le \xi_2} F_2(x)$. *Furthermore, if we assume that* (H_2) *and* (H_3), *and if*

$$F_1(x) \ge F_1(M) \qquad \text{for } 0 \le x \le M,$$

then there exists an $n_2 \ge n_1$ *such that*

$$N(n) \ge m \equiv \xi_1 e^{F_1(M) r^L} \qquad \text{for } n \ge n_2. \tag{6}$$

3 Periodic Models with Finite Delay

To construct the discrete type existence theorem of the periodic solution, we first consider the scalar general functional difference equation

$$x(n+1) = g(n, x_n). \tag{7}$$

Here x_n is the segment of $x(s)$ on $[n-h, n]$ shifted to $[-h, 0]$, where $h > 0$ is a fixed constant integer. For $h > 0$ and $\phi \in BS = BS([-h, 0])$, $|\phi| = \sup_{-h\le s\le 0}|\phi(s)|$. In (18), $g(n, \phi)$ is continuous for ϕ (second term) defined on $Z \times BS$ to R and it takes bounded sets into bounded sets. Moreover, it satisfies a local Lipschitz condition in ϕ and there is a T with $g(n+T, \phi) = g(n, \phi)$ whenever ϕ is also T-periodic. A solution through (n_0, ϕ) is denoted by $x(n_0, \phi)$ with value at being $x(n, n_0, \phi)$ and with $x(n_0, n_0, \phi) = \phi$. The solution of (7) is unique for the initial function ϕ (cf. [8, 9, 11]).

We consider the following fixed point theorems by Schauder and Horn [cf. [1]].

Theorem A *(Schauder) A continuous mapping* Q *of a compact convex nonempty subset* Y *in the Banach space* X *into itself has at least one fixed-point.*

Theorem B *(Horn) Let* $S_0 \subset S_1 \subset S_2$ *be convex subsets of the Banach space* X, *with* S_0 *and* S_2 *compact and* S_1 *open relative to* S_2. *Let* $Q : S_2 \to X$ *be a continuous* *function such that for some integer* $q > 0$,

$$(a)\ Q^j S_1 \subset S_2, \quad 1 \le j \le q-1$$

and

$$(b)\ Q^j S_1 \subset S_0, \quad q \le j \le 2q-1.$$

Then Q has a fixed point in S_0.

we now define the uniformly bounded and uniformly ultimately bounded of solutions for (7).

Definition 3 The solutions of (7) are uniformly bounded if for any $K > 0$, there exists a $B_1 = B_1(K) > 0$ such that $|\phi| \le K$ implies $|x(n, n_0, \phi)| < B_1$ for all $n \ge n_0$.

Definition 4 The solutions of (7) are uniform ultimately bounded for bound B, if there exists a $B > 0$ and if corresponding to any $K > 0$, there exists a $T^* = T^*(K) > 0$ such that $|\phi| \le K$ implies that $|x(n, n_0, \phi)| < B$ for all $n \ge n_0 + T^*$.

The Definitions 3 and 4 can employ to our finite delay difference Eq. (3) and so, Lemma 1 shows that the solution $N(n)$ of Eq. (3) is uniformly bounded and uniformly ultimately bounded under assumptions (H_1), (H_2) and (H_3). We show the following theorems by improving, as discrete case, the proof of theorems in [1].

Theorem 1 *Suppose that* $x(n+T)$ *is a solution of* (7) *whenever* $x(n)$ *is a solution of* (7). *If solutions of Eq.* (7) *are uniformly ultimately bounded for bound* B, *then it has a* $\bar{m}T$*-periodic solution for some positive integer* $\bar{m}$.

Theorem 2 *Suppose that* $x(n+T)$ *is a solution of* (7) *whenever* $x(n)$ *is a solution of* (7). *If solutions of Eq.* (7) *are uniformly bounded and uniformly ultimately bounded for bound* B, *then it has a* T*-periodic solution which is bounded by* B.

Proof Let $x(n) = x(n, 0, \phi)$ be the solution defined on Z^+ with initial time $n_0 = 0$. Since solution $x(n)$ of (7) is uniformly bounded, for $B > 0$ of uniformly ultimately bounded for bound B, there is a $B_1 > 0$ such that $|\phi| \le B$ implies that $|x(n, 0, \phi)| < B_1$. For this B_1, there exists a $B_2 > 0$ such that $|\phi| \le B_1 + 1$ implies that $|x(n, 0, \phi)| < B_2$. Also, since solution $x(n)$ is uniformly ultimately bounded, there is an integer m such that $|\phi| < B_1 + 1$ implies that $|x(n, 0, \phi)| < B$ for $n \ge mT - h$. By uniformly boundedness, for above B_2 there exists an $L > 0$ such that $|\phi| \le B_2$ implies that $|x(n+1, 0, \phi)| \le L$. Let

$$S_0 = \{\phi \in BS \,|\, |\phi| \le B, |\phi(u) - \phi(v)| \le L|u - v|\},$$
$$S_2 = \{\phi \in BS \,|\, |\phi| \le B_2, |\phi(u) - \phi(v)| \le L|u - v|\},$$

and

$$S_1 = \{\phi \in BS \,|\, |\phi| < B_1 + 1\} \cap S_2.$$

Then, the S_i, $(i = 0, 1, 2)$ are convex, and moreover S_0 and S_2 are compact by Ascoli's theorem. Furthermore, S_1 is open in S_2. Define $Q : S_2 \to BS$ by $\phi \in S_2$ implies $Q\phi = x(n+T, 0, \phi)$ for $-h \le n \le 0$. Now $x(n+T, 0, \phi)$ is a solution for $n \ge 0$ and its initial function is $Q\phi$. Hence,

$$x(n+T, 0, \phi) = x(n, 0, Q\phi) \tag{8}$$

by the uniqueness theorem. Next,

$$Q^2\phi = x(n+T, 0, Q\phi) \quad \text{for} \quad -h \le n \le 0$$

and $x(n+T, 0, Q\phi)$ is a solution with initial function $Q\phi$. Hence,

$$x(n+T, 0, Q\phi) = x\left(n, 0, Q^2\phi\right) \tag{9}$$

by uniqueness. Now in (8) let n be replaced by $n+T$ so that

$$x(n+2T, 0, \phi) = x(n+T, 0, Q\phi) = x\left(n, 0, Q^2\phi\right)$$

by (9). In general, for each integer $k > 0$, $x(n+kT, 0, \phi) = x(n, 0, Q^k\phi)$. By construction of S_1 and S_2 we have $Q^j S_1 \subset S_2$ for $1 \le j \le m$. By choice of m we have $Q^j S_1 \subset S_0$ for $j \ge m$. Also, $Q^j S_0 \subset S_1$ for all j. We can see that, by Horn's fixed point theorem, Q has a fixed point $\phi \in S_0$. Then, we conclude that $x(n, 0, \phi)$ is T-periodic solution of (7). This completes the proof.

Now, in our main theorem, we show that the existence of a periodic solution for the Eq. (3).

Theorem 3 *Under the assumptions* (H_1), (H_2) *and* (H_3), *if* $f(n, \phi)$ *is continuous in its second term, satisfies a local Lipschity condition in* ϕ *and periodic in* n *with period* T, *then Eq.* (3) *has a positive periodic solution of period* T.

4 Almost Periodic Models with Finite Delay

First, in (7), we assume that the $g(n, \phi)$ is almost periodic in n uniformly for $\phi \in BS$. In what follows, we need the following definitions of stability.

Definition 5 The bounded solution $x(n)$ of Eq. (7) is said to be;

(i) uniformly stable (in short, US) if for any $\varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$ such that if $n_0 \ge 0$, $|x_{n_0} - u_{n_0}| < \delta(\varepsilon)$, then $|x_n - u_n| < \varepsilon$ for all $n \ge n_0$, where $u(n)$ is a solution of (18) through (n_0, ψ) such that $u_{n_0}(s) = \psi(s)$ for all $-h \le s \le 0$.
(ii) uniformly asymptotically stable (in short, UAS) if it is US and if for any $\varepsilon > 0$ there exists a $\delta_0 > 0$ and a $T(\varepsilon) > 0$ such that if $n_0 \ge 0$, $|x_{n_0} - u_{n_0}| < \delta_0$, then

$|x_n - u_n| < \varepsilon$ for all $n \geq n_0 + T(\varepsilon)$, where $u(n)$ is a solution of (7) through (n_0, ψ) such that $u_{n_0}(s) = \psi(s)$ for all $-h \leq s \leq 0$. The δ_0 and the T in above are independent of n_0.

(iii) globally uniformly asymptotically stable (in short, GUAS) if it is US and if for any $\varepsilon > 0$ and $\alpha > 0$, there exists a $T(\varepsilon, \alpha) > 0$ such that if $n_0 \geq 0$, $|x_{n_0} - u_{n_0}| < \alpha$, then $|x_n - u_n| < \varepsilon$ for all $n \geq n_0 + T(\varepsilon, \alpha)$, where $u(n)$ is a solution of (7) through (n_0, ψ) such that $u_{n_0}(s) = \psi(s)$ for all $-h \leq s \leq 0$. The T in above is independent of n_0.

We consider the scalar delay difference equation as special case of (7).

$$\Delta x(n) = -a(n)x(n - r(n)), \tag{10}$$

where Δ is difference of $x(n)$, $a : Z^+ \to R^+$ and $r : Z^+ \to [0, h]$. We assume that $a(n)$ and $r(n)$ are almost periodic functions in n, and $0 < r^l = \liminf_{n\to\infty} r(n) \leq r(n) \leq \limsup_{n\to\infty} r(n) = r^L$.

The following lemma is derived from the results of Theorem 1.1 and 1.2 in [16], and so we omit the proof of Lemma 2.

Lemma 2 *If $a(n)$ and $r(n)$ are in R and*

$$\sum_{s=0}^{\infty} a(s) = \infty, \quad \lambda \equiv \limsup_{n\geq 0} \sum_{s=n-r(n)}^{n} a(s) \leq 1 + \frac{r^l + 2}{2\left(r^L + 1\right)},$$

then the zero solution $x(n) \equiv 0$ of (10) is uniformly stable. Moreover if $\lambda < 1 + \frac{r^l+2}{2(r^L+1)}$, then the zero solution of (10) is uniformly asymptotically stable.

We have the following theorem by Lemma 2.

Theorem 4 *Suppose that all the assumptions of Lemma 2 are satisfied. Let $\tilde{g}(n, z)$: $[n_0, \infty) \times R \to R^+$ be defined by*

$$\tilde{g}(n, z) = -\frac{\partial f(n, e^z)}{\partial z}. \tag{11}$$

We assume that there exists a function $G : [n_0, \infty) \to R^+$ such that

$$\sum_{s=n_0}^{\infty} \tilde{g}(s, x) = \infty, \qquad \text{and } \tilde{g}(n, x) \leq G(n) \text{ for } m \leq x \leq M$$

where m and M are defined by (5) and (6). If

$$\limsup_{n\to\infty} \sum_{s=n-r(n)}^{n} G(s) \leq 1 + \frac{r^l + 2}{2\left(r^L + 1\right)}, \tag{12}$$

then any positive solution of (3) *is uniformly asymptotically stable.*

Remark 1 From Theorem 2 in [9] and above Theorem 1, we can see that, under the assumptions of Theorem 1, if f is a periodic function in n with period $T > 0$, then (3) has a periodic solution with period T which is uniformly asymptotically stable. From Theorem 1, we obtain the following main theorem.

Theorem 5 *Under the assumptions* $(H_{1'})$, (H_2) *and* (H_3), *if* $f(n, \phi)$ *in* (3) *is continuous with its second term, satisfies a local Lipschitz condition in* ϕ *and an almost periodic function in* $n \in [n_0, \infty)$ *uniformly for* $\phi \in BS([-h, 0], R^+)$, *then Eq.* (3) *has a unique positive almost periodic solution which is globally uniformly asymptotically stable.*

Proof Let $N(n)$ denote a positive solution of (3) such that

$$0 < m \le N(n) \le M, \qquad \text{for } n \ge n_0,$$

where m and M are defined by (5) and (6). Let $\{n_j\}$ be a sequence of integer such that $n_j \to \infty$ as $j \to \infty$. Define $x_i(n)$ and $x_j(n)$ as follows,

$$e^{x_i(n)} = N(n + n_i), \qquad e^{x_j(n)} = N(n + n_j).$$

It follows from (3)

$$\Delta x_i = x_i(n+1) - x_i(n) = f\left(n + n_i, e^{x_i(n-r(n))}\right),$$
$$\Delta x_j = x_j(n+1) - x_j(n) = f\left(n + n_j, e^{x_j(n-r(n))}\right).$$

We let $y_{i,j}(n) = x_i(n) - x_j(n)$ and note that

$$\begin{aligned}\Delta y_{i,j}(n) &= y_{i,j}(n+1) - y_{i,j}(n)\\ &= f\left(n + n_i, e^{x_i(n-r(n))}\right) - f\left(n + n_j, e^{x_j(n-r(n))}\right)\\ &= f\left(n + n_i, e^{x_i(n-r(n))}\right) - f\left(n + n_i, e^{x_j(n-r(n))}\right) + h(n; n_i, n_j),\end{aligned} \tag{13}$$

where $h(n; n_i, n_j) = f\left(n + n_i, e^{x_j(n-r(n))}\right) - f\left(n + n_j, e^{x_j(n-r(n))}\right)$. By (11), we can rewrite (13) in the form

$$\Delta y_{i,j}(n) = -[\tilde{g}(n + n_i, \xi_{i,i}(n - r(n)))]y_{i,j}(n - r(n)) + h(n; n_i, n_j), \tag{14}$$

in which $\xi_{i,j}(n)$ lies between $N(n + n_i)$ and $N(n + n_j)$. By hypotheses and Lemma 2, the trivial solution of the linear equation

$$\Delta z(n) = -e(n)z(n - r(n)) \tag{15}$$

with $e(n) = \tilde{g}(n + n_i, \xi_{i,i}(n - r(n)))$ is uniformly asymptotically stable and hence the fundamental solution $W(n)$ associated with (15) satisfies an estimate of the type (see [1, 6])

$$\left|W(n)W^{-1}(s)\right| \leq C^* \eta^{\gamma(n-s)} \text{ for } \gamma \in (0, \infty), \;\; C^* \in [1, \infty), \eta \in (0, 1). \quad (16)$$

By the variation of constants formula [1], we have from (14) that

$$y_{i,j}(n) = W(n)W^{-1}(n_0)y_{i,j}(n_0) + \sum_{s=n_0}^{n-1} W(n)W^{-1}(s)h(s; n_i, n_j) \quad (17)$$

Using (16) and (17) we obtain

$$|y_{i,j}(n)| = C^* \eta^{\gamma(n-n_0)}|y_{i,j}(n_0)| + \sum_{s=n_0}^{n-1} C^* \eta^{\gamma(n-n_0)}|h(s; n_i, n_j)|, \;\; n \geq n_0. \quad (18)$$

From the almost periodic property of $f(n, \phi)$ in n uniformly for $\phi \in BS$, we can choose integers n_i^* and n_j^* large enough such that

$$|h(n; n_i, n_j)| = \left|f\left(n + n_i, e^{x_j(n-r(n))}\right) - f\left(n + n_j, e^{x_j(n-r(n))}\right)\right| < \frac{\varepsilon\gamma}{2C^*} \qquad \text{for } n_i \geq n_i^*, n_j \geq n_j^* \quad (19)$$

and for arbitrary $\varepsilon > 0$. Also, we have from $|y_{i,j}(n_0)| = |x(n_0 + n_i) - x(n_0 + n_j)|$ that

$$|y_{i,j}(n_0)| < \frac{\varepsilon}{2} \qquad \text{for } n_i \geq n_i^*, n_j \geq n_j^* \quad (20)$$

due to the uniformly convergence of the sequence $\{x(n + n_i)\}$ for n in compact subsets of $[n_0, \infty)$. Thus, it follows from (18), (19) and (20) that

$$|y_{i,j}(n)| = |x(n + n_i) - x(n + n_j)| < \varepsilon \qquad \text{for } n \in [n_0, \infty), n_i \geq n_i^*, n_j \geq n_j^*$$

and hence, $x(n)$ is asymptotically almost periodic in the sense that there exist x^* and $\bar{x}$ satisfying

$$x(n) = x^*(n) + \bar{x}(n) \quad (21)$$

where x^* is almost periodic on $[n_0, \infty)$ and $\bar{x}$ is on $[n_0, \infty)$ with $\bar{x}(n) \to 0$ as $n \to \infty$. We can now proceed as in [7] to show that the almost periodic part x^* in (21) is a solution of

$$\Delta x^*(n) = f\left(n, e^{x^*(n-r(n))}\right).$$

It follows that $N^*(n) = e^{x^*(n)}$ is an almost periodic solution of (3). The uniform asymptotic stability of $N^*(n)$ is an immediate consequence of Theorem that is, the proof is complete.

Conjecture *For the functional differential equation with finite delay, it is well known that the condition* (12) *and others is able to change the* $3/2$ *stability [cf.[13]]. Then, we have the following conjecture; it seems that if* (12) *is replacing*

$$\limsup_{n\to\infty} \sum_{s=n-r(n)}^{n} G(s) \le \frac{3}{2} + \frac{1}{2\left(r^L + 1\right)},$$

then the result of Theorem 4 is also right for (3). *However, we have no prove, yet.*

5 Examples

We first consider the following almost periodic delay model of the form

$$N(n+1) = N(n)\exp\left\{a(n) + b(n)N^p(n-r(n)) - c(n)N^q(n-r(n))\right\}, \quad (22)$$

where a, b, c and r are defined on $[n_0, \infty)$ with $a(n) > 0,\ b(n) \in R,\ c(n) > 0$ and $r(n) > 0$ are almost periodic functions in n, and

$$0 < p < q,\ \ 0 < r^l = \liminf_{n\to\infty} \le r(n) \le \limsup_{n\to\infty} r(n) = r^L, \quad (23)$$
$$0 < a^l = \liminf_{n\to\infty} a(n) \le a(n) \le \limsup_{n\to\infty} a(n) = a^L,$$

$$b^l = \liminf_{n\to\infty} b(n) \le b(n) \le \limsup_{n\to\infty} b(n) = b^L \text{ and} \quad (24)$$
$$0 < c^l = \liminf_{n\to\infty} c(n) \le c(n) \le \limsup_{n\to\infty} c(n) = c^L$$

for $n \ge n_0$. The original differential model of (22) is considered by Ladas and Qian [12]. Let

$$f(n, y) = a(n) + b(n)y^p - c(n)y^q$$

and

$$F_1(y) = a^l + b^l y^p - c^L y^q, \quad F_2(y) = a^L + b^L y^p - c^l y^q.$$

We note that $F_1(y) \le f(n, y) \le F_2(y)$ for $n \ge n_0,\ \ y \in [0, \infty)$. It is easy to see that all the assumptions (H_1), (H_2) and (H_3) are satisfied. We denote by y_* and y^* the unique positive solution of the equations $F_1(y) = 0$ and $F_2(y) = 0$, respectively. Then, any positive solution $N(n)$ of (22) satisfies eventually for all large n

$$m_1 \le N(n) \le M_1,$$

where

$$M_1 = \begin{cases} y^* e^{a^L r^L}, & b(n) \le 0, \\ y^* \exp\left[r^L \left(a^L + b^L \left(\frac{b^L p}{c^l q}\right)^{p/(q-p)} - c^l \left(\frac{b^L p}{c^l q}\right)^{q/(q-p)}\right)\right], & b(n) > 0 \end{cases}$$

and

$$m_1 = y_* \exp\left[r^L \left(a^l + b^l M_1^p - c^L M_1^q\right)\right]. \tag{25}$$

We obtain the result which is a generalization to the delay difference equation of that in [5]. According to the above result and Theorem 5, we can obtain the following corollary for Eq. (22) which has been studied recently [10] treated delay differential equations, and the conditions of the following corollary offer an improvement of this known results for differential equations.

Corollary 1 *Let $a(n)$, $b(n)$ and $c(n)$ be almost periodic functions. Suppose that (23) and (24) hold. If*

$$\limsup_{n\to\infty} \sum_{s=n-r(n)}^{n-1} \left[c(s) q M_1^q - b(s) p M_1^p\right] < 1 + \frac{r^l + 2}{2\left(r^L + 1\right)} \quad \textit{for } b(n) < 0$$

and

$$\limsup_{n\to\infty} \sum_{s=n-r(n)}^{n-1} \left[c(s) q M_1^q - b(s) p m_1^p\right] < 1 + \frac{r^l + 2}{2\left(r^L + 1\right)} \quad \textit{for } b(n) \ge 0$$

in which m_1 and M_1 are defined by (25), then (22) has a unique almost periodic solution which is global uniformly asymptotically stable.

Proof Let $N(n)$ be any positive solution of (22). Let $e^{u(n)} = N(n)$. Note that

$$g(n,u) = -\frac{\partial f(n,e^u)}{\partial u} = -b(n)pN^p + c(n)qN^q. \tag{26}$$

We have from above example that $N(n)$ satisfies

$$m_1 \le N(n) \le M_1,$$

eventually for all large n which, together with (26), implies that $g(n,u) \le G(n)$ where

$$G(n) = \begin{cases} c(n)qM_1^q - b(n)pM_1^p & \text{for } b(n) < 0 \\ c(n)qM_1^q - b(n)pm_1^p & \text{for } b(n) \ge 0. \end{cases}$$

Then, from Theorem 5, the conclusion follows.

6 Almost Periodic Models with Infinite Delay

We final consider the existence of an almost periodic solution of the following almost periodic Ricker type difference equation in the case where f of (3) has an infinite delay (cf. [4]);

$$N(n+1) = N(n)\exp\left\{a(n) - b(n)\sum_{s=0}^{\infty} K(s)N(n-s)\right\}, \tag{27}$$

where a and b are positive almost periodic functions in n, and K is nonnegative kernel on $[0,\infty)$ such that there exists a $\sigma > 0$ satisfying

$$\sum_{s=0}^{\infty} K(s) = 1, \quad \sum_{s=0}^{\infty} sK(s) \le \sigma < \infty \quad \text{and} \quad \sum_{s=0}^{\sigma} K(s) > 0. \tag{28}$$

Moreover, we assume that

$$\begin{aligned} 0 < a^l &= \liminf_{n\to\infty} a(n) \le a(n) \le \limsup_{n\to\infty} a(n) = a^L, \\ 0 < b^l &= \liminf_{n\to\infty} b(n) \le b(n) \le \limsup_{n\to\infty} b(n) = b^L. \end{aligned} \tag{29}$$

In the following we are concerned with positive solutions of (27) corresponding to initial conditions of the form

$$N(s) = \phi(s) \ge 0, \;\; s \in (-\infty, 0], \phi(0) > 0 \text{ and } \phi \in BS\left((-\infty,0], R^+\right). \tag{30}$$

The following lemma provides a priori upper estimate of any positive solution of (27).

Lemma 3 *If the assumptions* (28)*,* (29) *and* (30) *hold, then there exists a* $T^* > 0$ *such that any positive solution* $N(n)$ *of* (27) *satisfies*

$$N(n) \leq N^*,$$

where

$$N^* = \frac{a^L}{b^l \sum_{s=0}^{\sigma} K(s)} e^{a^L \sigma} \text{ for } n > T^*. \tag{31}$$

Furthermore, if

$$a^l > b^L N^*,$$

then there exist a $T_* \geq T^*$ *and an* $(N^* >) N_* > 0$ *such that*

$$N(n) \geq N_* \quad \text{for} \quad n \geq T_*.$$

Theorem 6 *Assume that* a, b *and* K *in* (27) *satisfy* (28)*,* (29) *and* (30)*.*

Then Eq. (27) *has an unique positive almost periodic solution say* $p(n)$ *such that any other positive solution* $N(n)$ *of* (27) *satisfies*

$$\lim_{n \to \infty} \{N(n) - p(n)\} = 0. \tag{32}$$

Proof For (27), we first introduce the change of variables

$$N(n) = \exp\{v(n)\}, \quad p(n) = \exp\{y(n)\}.$$

Then, (27) can be written as

$$v(n+1) - v(n) = a(n) - b(n) \sum_{s=-\infty}^{n} K(n-s) \exp\{v(s)\}. \tag{33}$$

We first consider Liapunov functional

$$V_1 = V_1(v(n), y(n)) = |v(n) - y(n)| + \sum_{s=0}^{\infty} K(s) \sum_{l=n-s}^{n-1} b(s+l) |\exp\{v(l)\} - \exp\{y(l)\}|,$$

where $y(n)$ and $v(n)$ are solutions of (33) which remains in bounded set $B := \{x \in R | N_* \leq x \leq N^*\}$. We have

$$
\begin{aligned}
\Delta V_1(v(n), y(n)) \le\ & |v(n+1) - v(n)| - |y(n+1) - y(n)| \\
& + \sum_{s=0}^{\infty} K(s)[b(s+n)|\exp\{v(n)\} - \exp\{y(n)\}| \\
& - b(n)|\exp\{v(n-s)\} - \exp\{y(n-s)\}|] \\
=\ & |a(n) - b(n)\sum_{s=0}^{\infty} K(s)\exp\{v(n-s)\}| - |a(n) \\
& - b(n)\sum_{s=0}^{\infty} K(s)\exp\{y(n-s)\}| \\
& + \sum_{s=0}^{\infty} K(s)[b(s+n)|\exp\{v(n)\} - \exp\{y(n)\}| \\
& - b(n)|\exp\{v(n-s)\} - \exp\{y(n-s)\}|] \\
\le\ & |a(n) - b(n)\exp\{v(n-s)\}| - |a(n) - b(n)\exp\{y(n-s)\}| \\
& + b(s+n)|\exp\{v(n)\} - \exp\{y(n)\}| - b(n)|\exp\{v(n-s)\} - \exp\{y(n-s)\}| \\
\le\ & -2b(n)|\exp\{v(n-s)\} - \exp\{y(n-s)\}| + b(s+n)|\exp\{v(n)\} - \exp\{y(n)\}|.
\end{aligned}
$$

Secondly we consider

$$
V_2 = V_2(v(n), y(n)) = 2|v(n+s) - y(n+s)|.
$$

Similar we can calculate

$$
\begin{aligned}
\Delta V_2(v(n), y(n)) &\le 2(|v(n+s+1) - v(n+s)| - |y(n+s+1) - y(n+s)|) \\
&\le -2b(s+n)|\exp\{v(n)\} - \exp\{y(n)\}|.
\end{aligned}
$$

We take $V = V_1 + V_2$. Then, we have

$$
\begin{aligned}
\Delta V(v(n), y(n)) \le\ & -2b(n)|\exp\{v(n-s)\} - \exp\{y(n-s)\}| \\
& - b(n+s)|\exp\{v(n)\} - \exp\{y(n)\}| \\
\le\ & -b(n+s)|\exp\{v(n)\} - \exp\{y(n)\}| \\
\le\ & -b^l|\exp\{v(n)\} - \exp\{y(n)\}|.
\end{aligned}
$$

From the mean value theorem, we have

$$
|\exp\{v(n)\} - \exp\{y(n)\}| = \exp\{\theta(n)\}|v(n) - y(n)|,
$$

where $\theta(n)$ lies between $v(n)$ and $y(n)$. Then, we have

$$
\Delta V(v(n), y(n)) \le -b^l D|v(n) - y(n)|,
$$

where set $D = \exp(N_*)$. Let be solutions $z(n)$ of (33) such that $N^* \geq z(n) \geq N_*$ for $n \geq T_*$. Thus $|v(n) - y(n)| \to 0$ as $n \to \infty$, and hence $V = V(v(n), y(n)) \to 0$ as $n \to \infty$. Moreover, we can show that $|v^*(n) - y^*(n)| \to 0$ as $n \to \infty$, where v^* and y^* are solutions in hull of (33) by the same argument as in [11]. By using similar Liapunov functional V^* of V, we can show that $V^* \to 0$ as $n \to \infty$. Note that this Liapunov functional V^* is a non-increasing functional on Z, and hence $V^* = V^*(v^*(n), y^*(n)) \equiv 0$. Thus, $v^*(n) = y^*(n)$ for all $n \in Z$. Therefore, each hull equation of (33) has a unique strictly positive bounded solution. By the equivalence between (27) and (33), it follows from results in [15] that (27) has an almost periodic solution $p(n)$ such that $N_* \leq p(n) \leq N^*$ for all $n \in Z$ and (32) yields. This proof is complete.

References

1. Burton, T.A.: Stability and Periodic Solutions of Ordinary and Functional Differential Equations. Academic Press, Orlando (1985)
2. Corduneanu, C.: Almost periodic discrete processes. Libertas Mathematica **2**, 159–169 (1982)
3. Gopalsamy, K.: Stability and Oscillations in Delay Differential Equations of Population Dynamics. Kluwer Academic, Dordrecht (1994)
4. Gopalsamy, K., He, X.Z., Wen, L.: Global attractivity and 'levei-crossings' in a periodic logistic integrodifferential equation. Math. Nachr. **156**, 25–44 (1992)
5. Gopalsamy, K., He, X.Z., Wen, L.: Global attractivity and oscillations in an almost periodic delay logistic equation. Nonlinear Times Digest **1**, 9–24 (1994)
6. Hale, J.K., Lunel, S.M.V.: Introduction to Functional Differential Equations. Springer, New York (1993)
7. Hamaya, Y.: Periodic solutions of nonlinear integrodifferential equations. Tohoku Math. J. **41**, 105–116 (1989)
8. Hamaya, Y.: Existence of almost periodic solutions of discrete Ricker delay models, Int. J. Diff. Eqn. (To appear)
9. Hamaya, Y., Tanaka, T.: Existence of periodic solutions of discrete Ricker delay models, Int. J. Math. Anal. To appear.
10. He, X.Z., Zhang, B.G., Gopalsamy, K.: Single species dynamics in changing environments. Dyn. Stab. Syst. **9**(4), 293–303 (1994)
11. Itokazu, T., Hamaya, Y.: Almost periodic solutions of prey-predator discrete models with delay, Adv. in Diff. Eqn. (2009). Article ID 976865 online
12. Ladas, G., Qian, C.: Oscillation and global stability in a delay logistic equation. Dyn. Stab. Syst. **9**, 153–162 (1994)
13. Yoneyama, T.: The 3/2 stability theorem for one-dimensional delay-differential equations with unbounded delay. J. Math. Anal. Appl. **165**, 133–143 (1992)
14. Yoshizawa, T.: Stability Theory and the Existence of Periodic Solutions and Almost Periodic Solutions. Springer, New York (1975)
15. Zhang, S., Zheng, G.: Almost periodic solutions of delay difference systems. Appl. Math. Comput. **131**, 497–516 (2002)
16. Zhou, Z., Zhang, Q.: Uniform stability of nonlinear difference systems. J. Math. Anal. Appl. **225**, 486–500 (1998)

Generalized Lagrange Identity for Discrete Symplectic Systems and Applications in Weyl–Titchmarsh Theory

Roman Šimon Hilscher and Petr Zemánek

Abstract In this paper we consider discrete symplectic systems with analytic dependence on the spectral parameter. We derive the Lagrange identity, which plays a fundamental role in the spectral theory of discrete symplectic and Hamiltonian systems. We compare it to several special cases well known in the literature. We also examine the applications of this identity in the theory of Weyl disks and square summable solutions for such systems. As an example we show that a symplectic system with the exponential coefficient matrix is in the limit point case.

1 Introduction

In this paper we consider a $2n$-dimensional discrete symplectic system

$$z_{k+1}(\lambda) = \mathbb{S}_k(\lambda)\, z_k(\lambda), \qquad (S_\lambda)$$

whose coefficient matrix $\mathbb{S}_k(\lambda) \in \mathbb{C}^{2n\times 2n}$ is analytic in the spectral parameter $\lambda \in \mathbb{C}$ in a neighborhood of 0 and satisfies a symplectic-type identity, i.e.,

$$\mathbb{S}_k(\lambda) = \sum_{j=0}^{\infty} \lambda^j \mathscr{S}_k^{[j]}, \quad \mathbb{S}_k^*(\lambda)\, \mathscr{J}\, \mathbb{S}_k(\bar{\lambda}) = \mathscr{J}, \quad \mathscr{J} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \qquad (1)$$

R. Šimon Hilscher (✉) · P. Zemánek
Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlářská 2, 61137 Brno, Czech Republic
e-mail: hilscher@math.muni.cz

P. Zemánek
e-mail: zemanekp@math.muni.cz

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_10

The superscript * denotes the conjugate transpose and $M^*(\lambda) := [M(\lambda)]^*$. For the applications we will in addition assume that a certain weight matrix $\Psi(\lambda) \in \mathbb{C}^{2n\times 2n}$ is positive semidefinite. The terminology "symplectic system" refers to the fact that $\mathbb{S}_k(\lambda)$ and the fundamental matrix of system (S_λ) are complex symplectic (also called conjugate symplectic or $\mathcal{J}$-unitary) when λ is real, i.e., they satisfy the identity $M^*\mathcal{J}M = \mathcal{J}$.

For convenience we write system (S_λ) as two n-dimensional equations with $z_k(\lambda) = (x_k^*(\lambda),\ u_k^*(\lambda))^*$ and $\mathbb{S}_k(\lambda) = \begin{pmatrix} \mathscr{A}_k(\lambda) & \mathscr{B}_k(\lambda) \\ \mathscr{C}_k(\lambda) & \mathscr{D}_k(\lambda) \end{pmatrix}$. System (S_λ) was studied in the literature in several special cases. In [2–4, 6, 9] the first equation in (S_λ) does not depend on λ and the second equation is linear in λ, which by [2, Remark 3(iii)] gives the form

$$z_{k+1}(\lambda) = \begin{pmatrix} \mathscr{A}_k & \mathscr{B}_k \\ \mathscr{C}_k - \lambda W_k \mathscr{A}_k & \mathscr{D}_k - \lambda W_k \mathscr{B}_k \end{pmatrix} z_k(\lambda), \tag{2}$$

where $\mathscr{S}_k := \begin{pmatrix} \mathscr{A}_k & \mathscr{B}_k \\ \mathscr{C}_k & \mathscr{D}_k \end{pmatrix}$ is complex symplectic, W_k is Hermitian, and $W_k \geq 0$. Note that system (2) covers also the classical second order Sturm–Liouville equation

$$-\Delta(R_k \Delta y_k(\lambda)) + Q_k y_{k+1}(\lambda) = \lambda W_k y_{k+1}(\lambda) \tag{3}$$

with Hermitian matrices $R_k, Q_k, W_k \in \mathbb{C}^{n\times n}$, invertible R_k, and $W_k > 0$, see Example 1. System (S_λ) with a general linear dependence on λ

$$z_{k+1}(\lambda) = (\mathscr{S}_k + \lambda \mathscr{V}_k)\, z_k(\lambda) \tag{4}$$

was studied in [18, 19], where the matrix $\mathscr{S}_k$ is complex symplectic, $\mathscr{V}_k^* \mathcal{J} \mathscr{S}_k$ is Hermitian and positive semidefinite, and $\mathscr{V}_k^* \mathcal{J} \mathscr{V}_k = 0$. In [15, 16] the linear Hamiltonian difference system

$$\Delta \begin{pmatrix} x_k(\lambda) \\ u_k(\lambda) \end{pmatrix} = \begin{pmatrix} A_k & B_k + \lambda W_k^{[1]} \\ C_k - \lambda W_k^{[2]} & -A_k^* \end{pmatrix} \begin{pmatrix} x_{k+1}(\lambda) \\ u_k(\lambda) \end{pmatrix} \tag{5}$$

is considered with the matrices $A_k, B_k, C_k, W_k^{[1]}, W_k^{[2]} \in \mathbb{C}^{n\times n}$, $\tilde{A}_k := (I - A_k)^{-1}$ exists, $B_k, C_k, W_k^{[1]}, W_k^{[2]}$ are Hermitian, and $W_k^{[1]} \geq 0$, $W_k^{[2]} \geq 0$. Upon expanding the forward difference in (5), we can verify that system (5) corresponds to a discrete symplectic system (S_λ) with quadratic dependence on λ. Another linear Hamiltonian system

$$\Delta \begin{pmatrix} x_k(\lambda) \\ u_k(\lambda) \end{pmatrix} = \lambda \mathcal{J} H_k \begin{pmatrix} x_{k+1}(\lambda) \\ u_k(\lambda) \end{pmatrix}, \qquad H_k := \begin{pmatrix} -C_k & A_k^* \\ A_k & B_k \end{pmatrix}, \tag{6}$$

with $H_k \in \mathbb{C}^{2n\times 2n}$ Hermitian and $\tilde{A}_k(\lambda) := (I - \lambda A_k)^{-1}$ is studied in [13, 14]. Upon expanding the latter inverse into a power series we get the analytic dependence on λ in the coefficient matrix of system (6). More generally, the linear Hamiltonian system corresponding to

$$\Delta \begin{pmatrix} x_k(\lambda) \\ u_k(\lambda) \end{pmatrix} = \mathscr{J}\,(H_k^{[0]} + \lambda H_k^{[1]}) \begin{pmatrix} x_{k+1}(\lambda) \\ u_k(\lambda) \end{pmatrix}$$

with Hermitian $H_k^{[0]}, H_k^{[1]} \in \mathbb{C}^{2n\times 2n}$ is considered in [5]. Finally, a discrete symplectic system (S_λ) with analytic dependence on λ and $\mathscr{S}_k^{[0]} = I$ is studied in [7]. The latter paper also motivated the present study.

All the above mentioned references are devoted to various results in the spectral theory of the corresponding system. As it is known, the Lagrange identity plays a fundamental role in these investigations. This identity connects the $\mathscr{J}$-skew-product of two solutions of system (S_λ) with the associated weight matrix $\Psi_k(\lambda)$. In this paper we prove a general form of the Lagrange identity for system (S_λ) with analytic dependence on $\lambda \in \mathbb{C}$ and calculate the corresponding weight matrix explicitly in terms of the coefficients of (S_λ). This result includes the Lagrange identities for the above mentioned special systems. As a consequence we obtain the $\mathscr{J}$-monotonicity of the fundamental matrix $\Phi_k(\lambda)$ of (S_λ), which is used in [7] for proving the Krein traffic rules for the eigenvalues of $\Phi_k(\lambda)$. We also investigate applications of the generalized Lagrange identity in the discrete Weyl–Titchmarsh theory. In particular, we show that under an appropriate Atkinson condition involving the weight matrix $\Psi_k(\lambda)$, the theory of eigenvalues, Weyl disks, and square summable solutions developed in [18, 19] for system (4) remains valid without any change also for system (S_λ) with the analytic dependence on λ.

2 Lagrange Identity

Consider system (S_λ) with complex $2n \times 2n$ matrix $\mathbb{S}_k(\lambda)$ such that (1) holds. The parameter $\lambda \in \mathbb{C}$ is restricted to $|\lambda| < \varepsilon$ for some $\varepsilon > 0$ ($\varepsilon = \infty$ is allowed), which bounds the region of convergence of $\mathbb{S}_k(\lambda)$ in (1) for all $k \in [0,\infty)_{\mathbb{Z}} := [0,\infty) \cap \mathbb{Z}$. It follows that the matrices $\mathscr{S}_k^{[j]}$, $j \in [0,\infty)_{\mathbb{Z}}$, satisfy the identities

$$\mathscr{S}_k^{[0]*} \mathscr{J} \mathscr{S}_k^{[0]} = \mathscr{J} \tag{7}$$

$$\sum_{j=0}^{m} \mathscr{S}_k^{[j]*} \mathscr{J} \mathscr{S}_k^{[m-j]} = 0, \qquad m \in \mathbb{N} \tag{8}$$

for all $k \in [0,\infty)_{\mathbb{Z}}$. We note that $|\det \mathscr{S}_k^{[0]}| = 1$, as the determinant of any complex symplectic matrix is a complex unit. The second identity in (1) implies that $\mathbb{S}_k(\lambda)$ is invertible and hence

$$\mathbb{S}_k^{-1}(\lambda) = -\mathscr{J}\,\mathbb{S}_k^*(\bar{\lambda})\,\mathscr{J} = -\sum_{j=0}^{\infty} \lambda^j\,\mathscr{J}\,\mathscr{S}_k^{[j]*}\,\mathscr{J}.$$

Remark 1 From $\mathbb{S}_k(\lambda)\,\mathbb{S}_k^{-1}(\lambda) = I$ we then obtain the identity $\mathbb{S}_k(\lambda)\,\mathscr{J}\,\mathbb{S}_k^*(\bar{\lambda}) = \mathscr{J}$ or equivalently

$$\mathscr{S}_k^{[0]}\mathscr{J}\mathscr{S}_k^{[0]*} = \mathscr{J}, \qquad \sum_{j=0}^{m} \mathscr{S}_k^{[j]}\mathscr{J}\mathscr{S}_k^{[m-j]*} = 0, \qquad m \in \mathbb{N}.$$

First we study the $\mathscr{J}$-skew product of two coefficient matrices with different values of the spectral parameter. This lemma gives a main tool for the proof of the Lagrange identity.

Lemma 1 *Assume* (7)–(8). *Then for every* $\lambda, \nu \in \mathbb{C}$ *with* $|\lambda| < \varepsilon$, $|\nu| < \varepsilon$,

$$\mathbb{S}_k^*(\lambda)\,\mathscr{J}\,\mathbb{S}_k(\nu) = \mathscr{J} + (\bar{\lambda} - \nu)\,\Omega_k(\bar{\lambda}, \nu),$$

$k \in [0, \infty)_{\mathbb{Z}}$, *where the* $2n \times 2n$ *matrix* $\Omega(\bar{\lambda}, \nu)$ *is defined by*

$$\Omega_k(\bar{\lambda}, \nu) := \sum_{m=0}^{\infty} \sum_{j=0}^{m} \bar{\lambda}^{m-j}\,\nu^j \sum_{l=0}^{j} \mathscr{S}_k^{[m-l+1]*}\mathscr{J}\mathscr{S}_k^{[l]}. \tag{9}$$

Moreover, for $\nu = \lambda$ *the matrix* $\Omega_k(\bar{\lambda}, \lambda)$ *is Hermitian for all* $k \in [0, \infty)_{\mathbb{Z}}$.

Proof We fix $|\lambda| < \varepsilon$, $|\nu| < \varepsilon$, and $k \in [0, \infty)_{\mathbb{Z}}$. The power series for $\mathbb{S}_k^*(\lambda)$ and $\mathbb{S}_k(\nu)$ converge absolutely, so that the terms in the product $\mathbb{S}_k^*(\lambda)\,\mathscr{J}\,\mathbb{S}_k(\nu)$ can be re-arranged to the separate powers of $\bar{\lambda}^{m-j}\nu^j$, that is,

$$\mathbb{S}_k^*(\lambda)\,\mathscr{J}\,\mathbb{S}_k(\nu) = \sum_{m=0}^{\infty} \sum_{j=0}^{m} \bar{\lambda}^{m-j}\,\nu^j\,\mathscr{S}_k^{[m-j]*}\mathscr{J}\mathscr{S}_k^{[j]}.$$

By using identity (8) for each $m \in \mathbb{N}$, we replace the term $\nu^m\,\mathscr{S}_k^{[0]*}\mathscr{J}\mathscr{S}_k^{[m]}$ by

$$-\nu^m \left(\mathscr{S}_k^{[m]*}\mathscr{J}\mathscr{S}_k^{[0]} + \mathscr{S}_k^{[m-1]*}\mathscr{J}\mathscr{S}_k^{[1]} + \cdots + \mathscr{S}_k^{[1]*}\mathscr{J}\mathscr{S}_k^{[m-1]}\right).$$

Thus, with the aid of (7) we get

$$\mathbb{S}_k^*(\lambda)\,\mathscr{J}\,\mathbb{S}_k(\nu) = \mathscr{J} + \sum_{m=1}^{\infty} \sum_{j=1}^{m} (\bar{\lambda}^j - \nu^j)\,\nu^{m-j}\,\mathscr{S}_k^{[j]*}\mathscr{J}\mathscr{S}_k^{[m-j]}.$$

Upon factoring $\bar{\lambda} - \nu$ out of each term $\bar{\lambda}^j - \nu^j$ and collecting the remaining products with the same powers of $\bar{\lambda}$ and ν, we get

$$\mathbb{S}_k^*(\lambda)\,\mathscr{J}\,\mathbb{S}_k(\nu) = \mathscr{J} + (\bar{\lambda} - \nu) \sum_{m=1}^{\infty} \sum_{j=1}^{m} \Bigg(\sum_{l=1}^{j} \bar{\lambda}^{j-l} \nu^{l-1} \Bigg) \nu^{m-j}\, \mathscr{S}_k^{[j]*} \mathscr{J} \mathscr{S}_k^{[m-j]}$$
$$= \mathscr{J} + (\bar{\lambda} - \nu)\, \Omega_k(\bar{\lambda}, \nu),$$

where $\Omega_k(\bar{\lambda}, \nu)$ is given in (9). Finally, for $\nu := \lambda$ we get by using $\mathscr{J}^* = -\mathscr{J}$ and identities (8) that the matrix $\Omega_k(\bar{\lambda}, \lambda)$ is Hermitian. This latter fact is also shown in [7, Proposition 1]. □

The following theorem provides the main result of this section. Its relationship to known discrete Lagrange identities in the literature is discussed in Sect. 3.

Theorem 1 (Lagrange identity) *Assume* (7)–(8) *and fix* $\lambda, \nu \in \mathbb{C}$ *with* $|\lambda| < \varepsilon$, $|\nu| < \varepsilon$. *For any two solutions* $z(\lambda)$ *and* $z(\nu)$ *of systems* (S_λ) *and* (S_ν)*, respectively, we have for all* $k \in [0, \infty)_{\mathbb{Z}}$

$$\Delta\big(z_k^*(\lambda)\,\mathscr{J}\,z_k(\nu)\big) = (\bar{\lambda} - \nu)\; z_k^*(\lambda)\,\Omega_k(\bar{\lambda}, \nu)\,z_k(\nu), \tag{10}$$

$$z_{k+1}^*(\lambda)\,\mathscr{J}\,z_{k+1}(\nu) = z_0^*(\lambda)\,\mathscr{J}\,z_0(\nu) + (\bar{\lambda} - \nu) \sum_{j=0}^{k} z_j^*(\lambda)\,\Omega_j(\bar{\lambda}, \nu)\,z_j(\nu). \tag{11}$$

Proof Given that $z_{k+1}(\lambda) = \mathbb{S}_k(\lambda)\,z_k(\lambda)$ and $z_{k+1}(\nu) = \mathbb{S}_k(\nu)\,z_k(\nu)$ for all $k \in [0, \infty)_{\mathbb{Z}}$, we obtain from Lemma 1 that

$$\Delta\big(z_k^*(\lambda)\,\mathscr{J}\,z_k(\nu)\big) = z_k^*(\lambda)\,\big[\mathbb{S}_k^*(\lambda)\,\mathscr{J}\,\mathbb{S}_k(\nu) - \mathscr{J}\,\big]\,z_k(\nu)$$
$$= (\bar{\lambda} - \nu)\; z_k^*(\lambda)\,\Omega_k(\bar{\lambda}, \nu)\,z_k(\nu).$$

The summation of (10) over the interval $[0, k]_{\mathbb{Z}}$ then yields (11). □

Motivated by Lemma 1, we define for $k \in [0, \infty)_{\mathbb{Z}}$ the Hermitian $2n \times 2n$ matrix

$$\Psi_k(\lambda) := \Omega_k(\bar{\lambda}, \lambda) = \sum_{m=0}^{\infty} \sum_{j=0}^{m} \bar{\lambda}^{m-j} \lambda^j \sum_{l=0}^{j} \mathscr{S}_k^{[m-l+1]*} \mathscr{J} \mathscr{S}_k^{[l]}. \tag{12}$$

The following identities show that $\Psi_k(\lambda)$ is the correct weight matrix for the spectral theory of system (S_λ), see the examples and applications in Sects. 3 and 4.

Corollary 1 *For every* $\lambda \in \mathbb{C}$ *with* $|\lambda| < \varepsilon$ *and* $k \in [0, \infty)_{\mathbb{Z}}$ *we have*

$$\Delta\big(z_k^*(\lambda)\,\mathscr{J}\,z_k(\lambda)\big) = -2i\,\mathrm{im}(\lambda)\,z_k^*(\lambda)\,\Psi_k(\lambda)\,z_k(\lambda), \tag{13}$$

$$z_{k+1}^*(\lambda)\,\mathscr{J}\,z_{k+1}(\lambda) = z_0^*(\lambda)\,\mathscr{J}\,z_0(\lambda) - 2i\,\mathrm{im}(\lambda)\sum_{j=0}^{k} z_j^*(\lambda)\,\Psi_j(\lambda)\,z_j(\lambda), \tag{14}$$

$$z_k^*(\lambda)\,\mathscr{J}\,z_k(\bar\lambda) = z_0^*(\lambda)\,\mathscr{J}\,z_0(\bar\lambda). \tag{15}$$

Remark 2 When $|\lambda| < \varepsilon$ and $\lambda \in \mathbb{R}$, we have

$$\Psi_k(\lambda) = \sum_{m=0}^{\infty}\sum_{j=0}^{m}\lambda^m \sum_{l=0}^{j} \mathscr{S}_k^{[m-l+1]*}\mathscr{J}\,\mathscr{S}_k^{[l]} = -\mathbb{S}_k^*(\lambda)\,\mathscr{J}\,\dot{\mathbb{S}}_k(\lambda) = \dot{\mathbb{S}}_k^*(\lambda)\,\mathscr{J}\,\mathbb{S}_k(\lambda),$$

where the dot denotes the derivative with respect to λ. The weight matrix

$$\mathscr{J}\,\dot{\mathbb{S}}_k(\lambda)\,\mathscr{J}\,\mathbb{S}_k^*(\lambda)\,\mathscr{J} = \mathbb{S}_k^{*-1}(\lambda)\,\Psi_k(\lambda)\,\mathbb{S}_k^{-1}(\lambda)$$

was used in [11, 17] in the oscillation theory of systems (S_λ) with general nonlinear dependence on λ.

3 Special Examples

In this section we show the connection of the generalized Lagrange identity in Theorem 1 to several special cases known in the literature. We also demonstrate that a positive definite weight matrix $\Psi_k(\lambda)$ can be obtained when $\mathbb{S}_k(\lambda)$ is quadratic in λ.

Example 1 In the simplest case, i.e., for the second order Sturm–Liouville difference equation (3), the Lagrange identity is

$$\Delta\big[y_k^*(\lambda)\,R_k\,\Delta y_k(\nu) - (\Delta y_k^*(\lambda))\,R_k\,y_k(\nu)\big] = (\bar\lambda - \nu)\;y_{k+1}^*(\lambda)\,W_k\,y_{k+1}(\nu),$$

see e.g. [1, Theorem 4.2.1] or [10, Theorem 2.2.3]. This can be seen from (10) and (9), in which $\varepsilon = \infty$, $x_k := y_k$, $u_k := R_k\Delta y_k$, $z_k = (x_k^*,\ u_k^*)^*$, and use the formula $x_{k+1} = x_k + R_k^{-1}u_k$. The coefficient matrix of (S_λ) is $\mathbb{S}_k(\lambda) := \mathscr{S}_k + \lambda\mathscr{V}_k$ with

$$\mathscr{S}_k := \mathscr{S}_k^{[0]} = \begin{pmatrix} I & R_k^{-1} \\ Q_k & I + Q_k R_k^{-1}\end{pmatrix}, \quad \mathscr{V}_k := \mathscr{S}_k^{[1]} = -\begin{pmatrix} 0 & 0 \\ W_k & W_k R_k^{-1}\end{pmatrix},$$
$$\Omega(\bar\lambda, \nu) = \mathscr{V}_k^*\mathscr{J}\,\mathscr{S}_k = \big(I,\ R_k^{-1}\big)^*\ W_k\ \big(I,\ R_k^{-1}\big) = \Psi_k(\lambda),$$

and $\mathscr{S}_k^{[j]} := 0$ for $j \geq 2$. Note that Eqs. (7) and (8) with $m \in \{1, 2\}$ hold, since R_k, Q_k, W_k are assumed to be Hermitian.

Example 2 Consider system (4) with general linear dependence on λ. In this case $\mathscr{S}_k^{[0]} := \mathscr{S}_k$, $\mathscr{S}_k^{[1]} := \mathscr{V}_k$, $\mathscr{S}_k^{[j]} := 0$ for $j \geq 2$, $\varepsilon = \infty$, and $\Omega_k(\bar{\lambda}, \nu) = \mathscr{V}_k^* \mathscr{J} \mathscr{S}_k = \Psi_k(\lambda)$ is constant in λ and Hermitian. Identities (7) and (8) with $m \in \{1, 2\}$ are

$$\mathscr{S}_k^* \mathscr{J} \mathscr{S}_k = \mathscr{J}, \quad \mathscr{S}_k^* \mathscr{J} \mathscr{V}_k + \mathscr{V}_k^* \mathscr{J} \mathscr{S}_k = 0, \quad \mathscr{V}_k^* \mathscr{J} \mathscr{V}_k = 0. \tag{16}$$

The Lagrange identity in (10) is, compare with [19, Theorem 2.6],

$$\begin{aligned} \Delta\big(z_k^*(\lambda)\, \mathscr{J}\, z_k(\nu)\big) &= (\bar{\lambda} - \nu)\ z_k^*(\lambda)\, \mathscr{V}_k^* \mathscr{J} \mathscr{S}_k\, z_k(\nu) \\ &= (\bar{\lambda} - \nu)\ z_{k+1}^*(\lambda)\ \mathscr{J}\, \mathscr{V}_k\, \mathscr{J}\, \mathscr{S}_k^* \mathscr{J}\, z_{k+1}(\nu). \end{aligned} \tag{17}$$

Observe that by (16) the matrix $\mathscr{V}_k$ is singular. Hence, $\Omega_k(\bar{\lambda}, \nu)$ and $\Psi_k(\lambda)$ are in this case singular as well. Moreover, $\det \mathbb{S}_k(\lambda) = \det \mathscr{S}_k$ and thus $|\det \mathbb{S}_k(\lambda)| = 1$.

Example 3 System (2) represents a special case of Example 2, namely

$$\begin{gathered} \mathscr{S}_k := \mathscr{S}_k^{[0]} = \begin{pmatrix} \mathscr{A}_k & \mathscr{B}_k \\ \mathscr{C}_k & \mathscr{D}_k \end{pmatrix}, \quad \mathscr{V}_k := \mathscr{S}_k^{[1]} = -\begin{pmatrix} 0 & 0 \\ W_k \mathscr{A}_k & W_k \mathscr{B}_k \end{pmatrix}, \\ \Omega_k(\bar{\lambda}, \nu) = \mathscr{V}_k^* \mathscr{J} \mathscr{S}_k = \big(\mathscr{A}_k,\ \mathscr{B}_k\big)^*\ W_k\ \big(\mathscr{A}_k,\ \mathscr{B}_k\big) = \Psi_k(\lambda). \end{gathered}$$

In this case the Lagrange identity in (10) or (17) has the form

$$\Delta\big(z_k^*(\lambda)\, \mathscr{J}\, z_k(\nu)\big) = (\bar{\lambda} - \nu)\ x_{k+1}^*(\lambda)\, W_k\, x_{k+1}(\nu), \tag{18}$$

where $z(\lambda) = (x^*(\lambda),\ u^*(\lambda))^*$ and $z(\nu) = (x^*(\nu),\ u^*(\nu))^*$. Identity (18) is used in [6, Lemma 2.3] and [4, Lemma 2.6].

Example 4 In this example we discuss the symplectic analogue (or a generalization to the symplectic system) of the linear Hamiltonian system (5), which was studied in [15, 16]. We take system (S_λ) with a special quadratic dependence on λ

$$z_{k+1}(\lambda) = \begin{pmatrix} \mathscr{A}_k & \mathscr{B}_k + \lambda \mathscr{A}_k W_k^{[2]} \\ \mathscr{C}_k - \lambda W_k^{[1]} \mathscr{A}_k & \mathscr{D}_k + \lambda \mathscr{C}_k W_k^{[2]} - \lambda W_k^{[1]} (\mathscr{B}_k + \lambda \mathscr{A}_k W_k^{[2]}) \end{pmatrix} z_k(\lambda), \tag{19}$$

where $W_k^{[1]}$ and $W_k^{[2]}$ are Hermitian. That is, $\mathbb{S}_k(\lambda) = \mathscr{S}_k + \lambda \mathscr{V}_k + \lambda^2 \mathscr{W}_k$ with $\varepsilon = \infty$,

$$\begin{gathered} \mathscr{S}_k := \mathscr{S}_k^{[0]} = \begin{pmatrix} \mathscr{A}_k & \mathscr{B}_k \\ \mathscr{C}_k & \mathscr{D}_k \end{pmatrix}, \quad \mathscr{W}_k := \mathscr{S}_k^{[2]} = \mathscr{J}\, \widetilde{W}_k \mathscr{S}_k\, \mathscr{J}\, \widehat{W}_k = \begin{pmatrix} 0 & 0 \\ 0 & -W_k^{[1]} \mathscr{A}_k W_k^{[2]} \end{pmatrix}, \\ \mathscr{V}_k := \mathscr{S}_k^{[1]} = \mathscr{J}\, \widetilde{W}_k \mathscr{S}_k + \mathscr{S}_k\, \mathscr{J}\, \widehat{W}_k = \begin{pmatrix} 0 & \mathscr{A}_k W_k^{[2]} \\ -W_k^{[1]} \mathscr{A}_k & \mathscr{C}_k W_k^{[2]} - W_k^{[1]} \mathscr{B}_k \end{pmatrix}, \\ \Omega_k(\bar{\lambda}, \nu) = \widehat{W}_k + (I - \bar{\lambda} \widehat{W}_k \mathscr{J})\, \mathscr{S}_k^*\, \widetilde{W}_k\, \mathscr{S}_k\, (I + \nu \mathscr{J}\, \widehat{W}_k), \end{gathered}$$

and $\mathscr{S}_k^{[j]} := 0$ for $j \geq 3$. The Hermitian $2n \times 2n$ matrices $\widetilde{W}_k := \mathrm{diag}\{W_k^{[1]}, 0\}$ and $\widehat{W}_k := \mathrm{diag}\{0, W_k^{[2]}\}$ are block diagonal. We can see that in this case $\Psi_k(\lambda) = \Omega_k(\bar{\lambda}, \lambda)$ is Hermitian but no longer constant in λ, as was the case in Examples 1–3. The above coefficients satisfy identities (7) and (8) with $m \in \{1, 2, 3, 4\}$, i.e.,

$$\mathscr{S}_k^* \mathscr{J} \mathscr{S}_k = \mathscr{J}, \qquad \mathscr{S}_k^* \mathscr{J} \mathscr{V}_k + \mathscr{V}_k^* \mathscr{J} \mathscr{S}_k = 0, \qquad \mathscr{V}_k^* \mathscr{J} \mathscr{W}_k + \mathscr{W}_k^* \mathscr{J} \mathscr{V}_k = 0$$
$$\mathscr{S}_k^* \mathscr{J} \mathscr{W}_k + \mathscr{V}_k^* \mathscr{J} \mathscr{V}_k + \mathscr{W}_k^* \mathscr{J} \mathscr{S}_k = 0, \qquad \mathscr{W}_k^* \mathscr{J} \mathscr{W}_k = 0.$$

The Lagrange identity in (10) now reads as

$$\Delta\big(z_k^*(\lambda) \mathscr{J} z_k(\nu)\big) = (\bar{\lambda} - \nu) \left[x_{k+1}^*(\lambda)\, W_k^{[1]}\, x_{k+1}(\nu) + u_k^*(\lambda)\, W_k^{[2]}\, u_k(\nu)\right], \tag{20}$$

where $z(\lambda) = (x^*(\lambda),\ u^*(\lambda))^*$ and $z(\nu) = (x^*(\nu),\ u^*(\nu))^*$. Identity (20) can be found in [16, Lemma 2.2]. We note that we can factorize $\mathbb{S}_k(\lambda)$ and $\Omega_k(\bar{\lambda}, \nu)$ as

$$\mathbb{S}_k(\lambda) = \begin{pmatrix} I & 0 \\ -\lambda W_k^{[1]} & I \end{pmatrix} \begin{pmatrix} \mathscr{A}_k & \mathscr{B}_k \\ \mathscr{C}_k & \mathscr{D}_k \end{pmatrix} \begin{pmatrix} I & \lambda W_k^{[2]} \\ 0 & I \end{pmatrix},$$
$$\Omega_k(\bar{\lambda}, \nu) = \begin{pmatrix} \mathscr{A}_k^* & 0 \\ \mathscr{B}_k^* + \bar{\lambda} W_k^{[2]} \mathscr{A}_k^* & I \end{pmatrix} \begin{pmatrix} W_k^{[1]} & 0 \\ 0 & W_k^{[2]} \end{pmatrix} \begin{pmatrix} \mathscr{A}_k & \mathscr{B}_k + \nu \mathscr{A}_k W_k^{[2]} \\ 0 & I \end{pmatrix}.$$

Therefore, $\det \mathbb{S}_k(\lambda) = \det \mathscr{S}_k$ and $|\det \mathbb{S}_k(\lambda)| = 1$ as in Example 2, and

$$\det \Psi_k(\lambda) = \det \Omega_k(\bar{\lambda}, \nu) = |\det \mathscr{A}_k|^2 \times \det W_k^{[1]} \times \det W_k^{[2]}. \tag{21}$$

Equation (21) shows that the determinant of the weight matrix $\Psi_k(\lambda)$ does not depend on λ. Moreover, $\Psi_k(\lambda)$ is invertible if and only if $\mathscr{A}_k$, $W_k^{[1]}$, $W_k^{[2]}$ are invertible. And in this case the matrix $\Psi_k(\lambda)$ is positive definite if and only if $W_k^{[1]}$ and $W_k^{[2]}$ are positive definite. However, an invertible (positive definite) weight matrix $\Psi_k(\lambda)$ can occur only when system (19) corresponds to a linear Hamiltonian system (5) with invertible (positive definite) $W_k^{[1]}$ and $W_k^{[2]}$, because in this case $\mathscr{A}_k = \tilde{A}_k$ is invertible. The other coefficients of (19) are then given by $\mathscr{B}_k = \tilde{A}_k B_k$, $\mathscr{C}_k = C_k \tilde{A}_k$, and $\mathscr{D}_k = C_k \tilde{A}_k B_k + I - A_k^*$, see [16, Formula (2.3)].

Example 5 Consider the linear Hamiltonian difference system (6), in which

$$Sco_k^{[0]} := I, \quad \mathscr{S}_k^{[1]} := \mathscr{J} H_k, \quad \mathscr{S}_k^{[j]} := \begin{pmatrix} A_k^j & A_k^{j-1} B_k \\ C_k A_k^{j-1} & C_k A_k^{j-2} B_k \end{pmatrix}, \quad j \geq 2,$$
$$\Omega_k(\bar{\lambda}, \nu) = D_k^*(\lambda)\, H_k\, D_k(\nu), \quad D_k(\lambda) := \begin{pmatrix} \tilde{A}_k(\lambda) & \lambda \tilde{A}_k(\lambda)\, B_k \\ 0 & I \end{pmatrix},$$

where $\tilde{A}_k(\lambda) := (I - \lambda A_k)^{-1}$, see [7, p. 5]. Therefore, the dependence of $\mathbb{S}_k(\lambda)$ on λ is analytic with $\varepsilon = \inf\{1/\mathrm{sprad}(A_k),\ k \in [0, \infty)_{\mathbb{Z}}\}$, provided this infimum is positive, where $\operatorname{sprad}(M) = \max\{|\mu|,\ \mu \text{ is an eigenvalue of } M\}$ denotes the spectral radius of M. The Lagrange identity in (10) is, compare with [13, Formula (9)],

$$\Delta\big(z_k^*(\lambda)\,\mathscr{J}\,z_k(\nu)\big) = (\bar{\lambda} - \nu)\,\begin{pmatrix} x_{k+1}(\lambda) \\ u_k(\lambda) \end{pmatrix}^* H_k \begin{pmatrix} x_{k+1}(\nu) \\ u_k(\nu) \end{pmatrix}.$$

Example 6 In [7, 8], the discrete symplectic system (S_λ) with $\mathscr{S}_k^{[j]} := (1/j!)\,R_k^j$ for $j \in [0,\infty)_{\mathbb{Z}}$ is studied, where $R_k \in \mathbb{C}^{2n\times 2n}$ satisfies $R_k^*\mathscr{J} + \mathscr{J}R_k = 0$. This means that the coefficient matrix $\mathbb{S}_k(\lambda)$ is of exponential type, i.e., $\varepsilon = \infty$ and

$$\mathbb{S}_k(\lambda) = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}\, R_k^j = \exp(\lambda R_k). \tag{22}$$

By [7, p. 6] or [8, Sect. 2], we then have

$$\Omega_k(\bar{\lambda},\nu) = \sum_{j=1}^{\infty} (-1)^j\, \frac{(\bar{\lambda}-\nu)^{2j-1}}{(2j)!}\,(R_k^*)^j\mathscr{J}R_k^j - \sum_{j=0}^{\infty} (-1)^j\, \frac{(\bar{\lambda}-\nu)^{2j}}{(2j+1)!}\,(R_k^*)^j\mathscr{J}R_k^{j+1}.$$

The Lagrange identity has the same form as in (10) with the above $\Omega_k(\bar{\lambda},\nu)$.

4 Weyl–Titchmarsh Theory

In this section we discuss the applications of the Lagrange identity from Theorem 1 in the Weyl–Titchmarsh theory for system (S_λ) with analytic dependence on λ. We assume that the Hermitian weight matrix $\Psi_k(\lambda)$ defined in (12) satisfies

$$\Psi_k(\lambda) \geq 0, \qquad k \in [0,\infty)_{\mathbb{Z}}. \tag{23}$$

In [19] we have recently developed the Weyl–Titchmarsh theory for system (4), i.e., for system (S_λ) with general linear dependence on λ. In this section we show that most of the results in [19] remain valid also for the analytic dependence on λ, when we modify the corresponding Atkinson-type condition to this more general setting. One of the crucial properties is that the fundamental matrix $\Phi_k(\lambda)$ of (S_λ) satisfies

$$\Phi_k^*(\lambda)\,\mathscr{J}\,\Phi_k(\bar{\lambda}) = \mathscr{J} \tag{24}$$

for all $k \in [0,\infty)_{\mathbb{Z}}$ whenever (24) holds at the initial point $k = 0$. Note that identity (24) now follows from (15) in Corollary 1. In the subsequent paragraphs we review the most important results, which are in particular connected to the theory of square summable solutions of (S_λ).

In [19] we identified the minimal requirements for the solutions of (S_λ) to satisfy the Atkinson condition. In this way we obtained the weak and strong Atkinson conditions, which are needed for different statements in the Weyl–Titchmarsh theory. For completeness we reformulate these conditions in the setting of this paper. Let $\Phi_k(\lambda) = (Z_k(\lambda),\ \widetilde{Z}_k(\lambda))$ be the partition of the fundamental matrix of system (S_λ) into $2n \times n$ solutions, which are given by the initial conditions $Z_0(\lambda) = \alpha^*$ and $\widetilde{Z}_0(\lambda) = -\mathscr{J}\alpha^*$ for some fixed $\alpha \in \mathbb{C}^{n\times 2n}$ with $\alpha\mathscr{J}\alpha^* = 0$ and $\alpha\alpha^* = I$. The solution $\widetilde{Z}(\lambda)$ is called the natural conjoined basis of (S_λ) and the spectral properties of the associated eigenvalue problem are formulated in terms of this natural conjoined basis. Let us fix for a moment an index $N \in [1,\infty)_{\mathbb{Z}}$.

Hypothesis 1 (*Finite weak Atkinson condition*) For all $\lambda \in \mathbb{C}$ with $|\lambda| < \varepsilon$ and every column $z(\lambda)$ of the natural conjoined basis $\widetilde{Z}(\lambda)$ of (S_λ) we assume that

$$\sum_{k=0}^{N} z_k^*(\lambda)\,\Psi_k(\lambda)\,z_k(\lambda) > 0. \tag{25}$$

If $\beta \in \mathbb{C}^{n\times 2n}$ with $\beta\mathscr{J}\beta^* = 0$ and $\beta\beta^* = I$ is also fixed, then we consider the symplectic eigenvalue problem

$$(S_\lambda), \quad k \in [0,N]_{\mathbb{Z}}, \quad \alpha\, z_0(\lambda) = 0, \quad \beta\, z_{N+1}(\lambda) = 0. \tag{26}$$

It follows as in [19, Theorem 2.8] that under Hypothesis 1 the eigenvalues of (26) are real, isolated, and they are characterized by $\det\ \beta\,\widetilde{Z}_{N+1}(\lambda) = 0$. The corresponding eigenfunctions are then of the form $\widetilde{Z}(\lambda)\,d$ with nonzero $d \in \operatorname{Ker}\beta\,\widetilde{Z}_{N+1}(\lambda)$.

The $M(\lambda)$-function for system (S_λ) is defined by $M_k(\lambda) := -[\beta\,\widetilde{Z}_k(\lambda)]^{-1}\beta\,Z_k(\lambda)$ and it satisfies the properties in [19, Lemma 2.10 and Theorem 2.13]. In particular, $M_k^*(\lambda) = M_k(\bar{\lambda})$ and $M_k(\lambda)$ is analytic in λ. Define the Weyl solution $\chi(\lambda, M)$ of (S_λ) corresponding to $M \in \mathbb{C}^{n\times n}$ and the Hermitian matrix function $\mathscr{E}(\lambda, M)$ by

$$\chi_k(\lambda, M) := \Phi_k(\lambda)\,\left(I,\ M^*\right)^*, \quad \mathscr{E}_k(\lambda, M) := i\,\delta(\lambda)\,\chi_k^*(\lambda, M)\,\mathscr{J}\,\chi_k(\lambda, M), \tag{27}$$

where $\delta(\lambda) := \operatorname{sgn}\operatorname{im}(\lambda)$. The Weyl disk $D_k(\lambda)$ and the Weyl circle $C_k(\lambda)$ are then defined as the sets

$$D_k(\lambda) := \left\{M \in \mathbb{C}^{n\times n},\ \mathscr{E}_k(\lambda, M) \le 0\right\}, \quad C_k(\lambda) := \left\{M \in \mathbb{C}^{n\times n},\ \mathscr{E}_k(\lambda, M) = 0\right\}.$$

It follows that the results in [19, Sect. 3] regarding the Weyl disks and Weyl circles hold exactly in the same form, but now under the following assumption.

Hypothesis 2 (*Infinite weak Atkinson condition*) There exists $N_0 \in \mathbb{N}$ such that for all $\lambda \in \mathbb{C}$ with $|\lambda| < \varepsilon$ every column $z(\lambda)$ of $\widetilde{Z}(\lambda)$ satisfies (25) with $N = N_0$.

We summarize the main properties of the Weyl disks in the following.

Theorem 2 *Let $\lambda \in \mathbb{C} \setminus \mathbb{R}$ with $|\lambda| < \varepsilon$ and suppose that (23) and Hypothesis 2 hold. Then for every $k \geq N_0 + 1$ the Weyl disk and Weyl circle satisfy*

$$D_k(\lambda) = \big\{P_k(\lambda) + R_k(\lambda)\, V R_k(\bar{\lambda}),\ V \in \mathbb{C}^{n\times n},\ V^*V \leq I\big\},$$
$$C_k(\lambda) = \big\{P_k(\lambda) + R_k(\lambda)\, U R_k(\bar{\lambda}),\ U \in \mathbb{C}^{n\times n},\ U^*U = I\big\},$$

where the center $P_k(\lambda)$ and the matrix radius $R_k(\lambda)$ are defined by

$$P_k(\lambda) := -\mathscr{H}_k^{-1}(\lambda)\, \mathscr{G}_k(\lambda), \qquad R_k(\lambda) := \mathscr{H}_k^{-1/2}(\lambda) \tag{28}$$

with $\mathscr{H}_k(\lambda)$ and $\mathscr{G}_k(\lambda)$ given by $\mathscr{H}_k(\lambda) := i\,\delta(\lambda)\, \widetilde{Z}_k^*(\lambda)\, \mathscr{J}\, \widetilde{Z}_k(\lambda)$ and $\mathscr{G}_k(\lambda) := i\,\delta(\lambda)\, \widetilde{Z}_k^*(\lambda)\, \mathscr{J}\, Z_k(\lambda)$. Moreover, the Weyl disks $D_k(\lambda)$ are closed, convex, and $D_{k+1}(\lambda) \subseteq D_k(\lambda)$ for all $k \geq N_0 + 1$.

Proof The proof follows the same arguments as in [19, Theorem 3.8]. We note that by (14) in Corollary 1 we have

$$\mathscr{H}_k(\lambda) = 2\,|\text{im}(\lambda)| \sum_{j=0}^{k-1} \widetilde{Z}_j^*(\lambda)\, \Psi_j(\lambda)\, \widetilde{Z}_j(\lambda). \tag{29}$$

This shows that under Hypothesis 2 the matrices $\mathscr{H}_k(\lambda)$ are Hermitian and positive definite for $k \geq N_0 + 1$, so that the center $P_k(\lambda)$ and the matrix radius $R_k(\lambda)$ are well defined. □

The properties of the Weyl disks $D_k(\lambda)$ in Theorem 2 and formula (29) imply that for $k \to \infty$ there exists the limiting Weyl disk $D_+(\lambda) := \bigcap_{k \geq N_0+1} D_k(\lambda)$, which is closed and convex and which satisfies

$$D_+(\lambda) = \big\{P_+(\lambda) + R_+(\lambda)\, V R_+(\bar{\lambda}),\ V \in \mathbb{C}^{n\times n},\ V^*V \leq I\big\},$$

where the limiting center and the limiting matrix radius are complex $n \times n$ matrices

$$P_+(\lambda) := \lim_{k\to\infty} P_k(\lambda), \qquad R_+(\lambda) := \lim_{k\to\infty} R_k(\lambda) \geq 0,$$

compare with [19, Theorem 3.9 and Corollary 3.11]. The elements of the limiting Weyl disk are characterized in the following result.

Theorem 3 *Let $\lambda \in \mathbb{C} \setminus \mathbb{R}$ with $|\lambda| < \varepsilon$ and suppose that (23) and Hypothesis 2 hold. The matrix $M \in \mathbb{C}^{n\times n}$ belongs to $D_+(\lambda)$ if and only if*

$$\sum_{k=0}^{\infty} \chi_k^*(\lambda, M)\, \Psi_k(\lambda)\, \chi_k(\lambda, M) \leq \frac{\text{im}(M)}{\text{im}(\lambda)}, \tag{30}$$

where $\chi(\lambda, M)$ is the Weyl solution of (S_λ) corresponding to M defined in (27).

Proof The proof follows by applying identity (14) in Corollary 1 to $\mathscr{E}_k(\lambda, M)$, see also [19, Corollary 3.12]. □

We now discuss the number of square summable solutions of (S_λ) with analytic dependence on λ. As the weight matrix $\Psi_k(\lambda)$ now depends on λ, we define for $\lambda \in \mathbb{C}$ with $|\lambda| < \varepsilon$ the semi-inner product and the semi-norm

$$\langle z, \tilde{z} \rangle_{\Psi(\lambda)} := \sum_{k=0}^{\infty} z_k^* \, \Psi_k(\lambda) \, \tilde{z}_k, \quad \|z\|_{\Psi(\lambda)} := \sqrt{\langle z, z \rangle_{\Psi(\lambda)}} = \left(\sum_{k=0}^{\infty} z_k^* \, \Psi_k(\lambda) \, z_k \right)^{1/2},$$

and the corresponding space of all square summable sequences with respect to $\Psi(\lambda)$

$$\ell^2_{\Psi(\lambda)} := \big\{ \{z_k\}_{k=0}^{\infty}, \ z_k \in \mathbb{C}^{2n}, \ \|z\|_{\Psi(\lambda)} < \infty \big\}. \tag{31}$$

Observe that the space $\ell^2_{\Psi(\lambda)}$ now also depends on λ. However, in some special cases this space can be taken independent on λ, as it is shown for systems (2), (3), (4) in Examples 1–3. Also, in view of (20) in Example 4 we may consider for systems (19) or (5) the space

$$\ell^2_{W^{[1]}, W^{[2]}} := \left\{ \{z_k = (x_k^*, u_k^*)^*\}_{k=0}^{\infty}, \ \sum_{k=0}^{\infty} \big(x_{k+1}^* \, W_k^{[1]} x_{k+1} + u_k^* \, W_k^{[2]} u_k \big) < \infty \right\},$$

which does not depend on λ. Given the space $\ell^2_{\Psi(\lambda)}$ in (31), its subspace of all square summable solutions of (S_λ) is denoted by

$$\mathcal{N}(\lambda) := \big\{ z \in \ell^2_{\Psi(\lambda)}, \ z = \{z_k\}_{k=0}^{\infty} \text{ solves} (\mathrm{S}_\lambda) \big\}.$$

Under assumption (23) and Hypothesis 2, the result in Theorem 3 implies that $n \le \dim \mathcal{N}(\lambda) \le 2n$ for all $\lambda \in \mathbb{C} \setminus \mathbb{R}$ with $|\lambda| < \varepsilon$, see also [19, Theorem 4.2]. The two extreme cases are then called as the limit point case when $\dim \mathcal{N}(\lambda) = n$, and the limit circle case when $\dim \mathcal{N}(\lambda) = 2n$. The cases when $\dim \mathcal{N}(\lambda)$ is between $n+1$ and $2n-1$ are called intermediate. It follows that the results in [19, Theorem 4.2, Corollary 4.15] hold for system (S_λ) with analytic dependence on λ in exactly the same form under the appropriate weak or strong Atkinson type condition. We summarize the main result regarding the number of linearly independent square summable solutions of (S_λ) in the following theorem.

Theorem 4 *Let $\lambda \in \mathbb{C} \setminus \mathbb{R}$ with $|\lambda| < \varepsilon$ and suppose that (23) and Hypothesis 2 hold. Then system (S_λ) has exactly $n + \operatorname{rank} R_+(\lambda)$ linearly independent square summable solutions, i.e.,*

$$\dim \mathcal{N}(\lambda) = n + \operatorname{rank} R_+(\lambda),$$

where $R_+(\lambda)$ is the matrix radius of the limiting Weyl disk $D_+(\lambda)$.

Proof We refer to the proof of [19, Theorem 4.9] for the details. □

As a consequence of Theorem 4 we obtain the characterization of the limit point case and limit circle case for system (S_λ) with analytic dependence on λ in terms of the limiting matrix radius $R_+(\lambda)$.

Corollary 2 *Let $\lambda \in \mathbb{C} \setminus \mathbb{R}$ with $|\lambda| < \varepsilon$ and suppose that (23) and Hypothesis 2 hold. System (S_λ) is in the limit point case if and only if $R_+(\lambda) = 0$, and in this case $D_+(\lambda) = \{P_+(\lambda)\}$ and $D_+(\bar\lambda) = \{P_+(\bar\lambda)\}$. System (S_λ) is in the limit circle case if and only if $R_+(\lambda)$ is invertible.*

In a similar way, the results in [18] regarding the Weyl–Titchmarsh theory for system (S_λ) with jointly varying endpoints remain valid also for the analytic dependence on λ, when we assume the following finite or infinite strong Atkinson type condition.

Hypothesis 3 (*Finite strong Atkinson condition*) For all $\lambda \in \mathbb{C}$ with $|\lambda| < \varepsilon$, every nontrivial solution $z(\lambda)$ of (S_λ) satisfies (25).

Hypothesis 4 (*Infinite strong Atkinson condition*) There exists $N_0 \in \mathbb{N}$ such that for all $\lambda \in \mathbb{C}$ with $|\lambda| < \varepsilon$ every nontrivial solution $z(\lambda)$ of (S_λ) satisfies (25) with $N = N_0$.

We illustrate the Weyl–Titchmarsh theory of system (S_λ) with analytic dependence on λ by the following interesting example.

Example 7 In this example we show that the discrete symplectic system

$$z_{k+1}(\lambda) = \exp(\lambda \mathscr{J})\, z_k(\lambda). \tag{32}$$

is in the limit point case for every $\lambda \in \mathbb{C} \setminus \mathbb{R}$. Moreover, we calculate the unique $2n \times n$ solution (up to an invertible multiple) of (32) whose columns lie in $\ell^2_{\Psi(\lambda)}$ and thus form a basis of $\mathcal{N}(\lambda)$. System (32) is an exponential symplectic system from Example 6, where $\varepsilon = \infty$, $\mathbb{S}_k(\lambda) := \exp(\lambda \mathscr{J})$ is given in (22) with $R_k := \mathscr{J}$. This matrix satisfies the conditions $R_k^* \mathscr{J} + \mathscr{J} R_k = 0$ in Example 6, so that by (22) we have $\mathbb{S}_k(\lambda) = \exp(\lambda \mathscr{J}) = (\cos\lambda)\, I + (\sin\lambda)\, \mathscr{J}$. Note that this matrix does not depend on k.

For simplicity we perform the calculations below in the scalar case, i.e., for $n = 1$. The general case follows with the same arguments upon multiplication by the $n \times n$ or $2n \times 2n$ identity matrices at appropriate places. The fundamental matrix $\Phi_k(\lambda)$ of (32) with $\Phi_0(\lambda) = I$ is given by

$$\Phi_k(\lambda) = \exp(k\lambda \mathscr{J}) = (\cos k\lambda)\, I + (\sin k\lambda)\, \mathscr{J} = \begin{pmatrix} \cos k\lambda & \sin k\lambda \\ -\sin k\lambda & \cos k\lambda \end{pmatrix}$$

for every $k \in [0, \infty)_{\mathbb{Z}}$. Since $\Phi_0(\lambda) = I$, we take $\alpha := (1,\ 0)$, so that $\alpha \mathscr{J} \alpha^* = 0$ and $\alpha\alpha^* = 1$ are satisfied. It follows that the natural conjoined basis of (32) is determined by the second column of $\Phi(\lambda)$, i.e., $\widetilde{Z}_k(\lambda) = ((\cos k\lambda)^*,\ (\sin k\lambda)^*)^*$. Since the powers of $\mathscr{J}$ repeat in a cycle of length four, the weight matrix $\Psi_k(\lambda) = \Omega(\bar{\lambda}, \lambda)$ is given in Example 6 as (we substitute $x := \mathrm{im}(\lambda)$)

$$\begin{aligned}\Psi_k(\lambda) &= \frac{1}{2x}\sum_{j=0}^{\infty}\frac{(2x)^{2j+1}}{(2j+1)!}\,I + \frac{1}{2x}\sum_{j=1}^{\infty}\frac{(2x)^{2j}}{(2j)!}\,i\mathscr{J} = \frac{\sinh\,2x}{2x}\,I + \frac{\cosh\,2x-1}{2x}\,i\mathscr{J}\\ &= \frac{\sinh\,x}{x}\left[(\cosh\,x)\,I + (\sinh\,x)\,i\,\mathscr{J}\right] = \frac{\sinh\,x}{x}\begin{pmatrix}\cosh x & i\sinh\,x\\ -i\sinh x & \cosh\,x\end{pmatrix} > 0,\end{aligned}$$

where we used the formulas for hyperbolic functions $\sinh\,2x = 2\sinh\,x\,\cosh\,x$, $\cosh\,2x = 2\cosh^2 x - 1$, and the identity $\cosh^2 x - \sinh^2 x = 1$. By the definition of $\mathscr{H}_k(\lambda)$ and $\mathscr{G}_k(\lambda)$ in Theorem 2,

$$\begin{aligned}\mathscr{H}_k(\lambda) &= i\,\delta(\lambda)\,(\sin k\bar{\lambda}\,\cos k\lambda - \cos k\bar{\lambda}\,\sin k\lambda) = \sinh\,(2k\,|\mathrm{im}(\lambda)|),\\ \mathscr{G}_k(\lambda) &= -i\,\delta(\lambda)\,(\sin k\bar{\lambda}\,\sin k\lambda + \cos k\bar{\lambda}\,\cos k\lambda) = -i\,\delta(\lambda)\,\cosh\,(2k\,\mathrm{im}(\lambda)),\end{aligned}$$

where we used the identities $\cosh\,x = \cos\,ix$ and $i\sinh\,x = \sin\,ix$ relating the hyperbolic and trigonometric functions. The same value for $\mathscr{H}_k(\lambda)$ is of course obtained from formula (29) after some calculations. Therefore, Hypothesis 2 is satisfied with $N_0 = 1$, and by (28)

$$P_k(\lambda) = i\,\coth\,(2k\,\mathrm{im}(\lambda)), \qquad R_k(\lambda) = 1/\sqrt{\sinh\,(2k\,|\mathrm{im}(\lambda)|)}.$$

The center and radius of the limiting disk $D_+(\lambda)$ are then

$$P_+(\lambda) = \lim_{k\to\infty} P_k(\lambda) = i\,\delta(\lambda), \qquad R_+(\lambda) = \lim_{k\to\infty} R_k(\lambda) = 0,$$

so that system (32) is in the limit point case for every $\lambda \in \mathbb{C}\setminus\mathbb{R}$. From Corollary 2 and Theorem 3 we obtain that $\dim\mathcal{N}(\lambda) = 1$, and the space $\mathcal{N}(\lambda)$ of square integrable solutions of system (32) is generated by the Weyl solution

$$\begin{aligned}\chi_k(\lambda, P_+(\lambda)) &= \Phi_k(\lambda)\begin{pmatrix} I\\ P_+(\lambda)\end{pmatrix} = \begin{pmatrix}\cos\,k\lambda + i\,\delta(\lambda)\,\sin\,k\lambda\\ -\sin\,k\lambda + i\,\delta(\lambda)\,\cos\,k\lambda\end{pmatrix}\\ &= \begin{pmatrix}1\\ i\,\delta(\lambda)\end{pmatrix} e^{i\,\delta(\lambda)\,k\lambda},\end{aligned}$$

for which (we again substitute $x := \mathrm{im}(\lambda)$)

$$\begin{aligned}\|\,\chi(\lambda, P_+(\lambda))\,\|^2_{\Psi(\lambda)} &= \sum_{k=0}^{\infty}\chi_k^*(\lambda, P_+(\lambda))\,\Psi_k(\lambda)\,\chi_k(\lambda, P_+(\lambda))\\ &= \frac{2\sinh\,x}{x}\times[\cosh\,x - \delta(\lambda)\,\sinh\,x]\times\sum_{k=0}^{\infty}e^{-2\,|x|k}\end{aligned}$$

$$= \frac{2\sinh x}{x} \times [\cosh x - \delta(\lambda)\, \sinh x] \times \frac{1}{1 - e^{-2|x|}} = \frac{1}{|x|}.$$

This shows that $\| \chi(\lambda, P_{+}(\lambda)) \|_{\Psi(\lambda)} = 1/\sqrt{|\mathrm{im}(\lambda)|} < \infty$, and so indeed we have $\chi(\lambda, P_{+}(\lambda)) \in \ell^2_{\Psi(\lambda)}$ for every $\lambda \in \mathbb{C} \setminus \mathbb{R}$. On the other hand, we also have

$$\| \widetilde{Z}(\lambda) \|^2_{\Psi(\lambda)} = \sum_{k=0}^{\infty} \widetilde{Z}^*_k(\lambda)\, \Psi_k(\lambda)\, \widetilde{Z}_k(\lambda) \overset{(29)}{=} \frac{1}{2\,|\mathrm{im}(\lambda)|} \lim_{k\to\infty} \mathscr{H}_k(\lambda)$$
$$= \frac{1}{2\,|\mathrm{im}(\lambda)|} \lim_{k\to\infty} \sinh(2k\,|\mathrm{im}(\lambda)|) = \infty,$$

so that $\tilde{Z}(\lambda) \notin \ell^2_{\Psi(\lambda)}$. Thus, again we get that $\dim \mathcal{N}(\lambda) = 1$. Similarly, in arbitrary dimension n we get that the n columns of the Weyl solution $\chi(\lambda, P_{+}(\lambda))$ are linearly independent and they belong to $\ell^2_{\Psi(\lambda)}$, while the n columns of the natural conjoined basis $\widetilde{Z}(\lambda)$ are linearly independent and they do not belong to $\ell^2_{\Psi(\lambda)}$. Hence, $\dim \mathcal{N}(\lambda) = n$ and system (32) is in the limit point case for all $\lambda \in \mathbb{C} \setminus \mathbb{R}$.

Finally, as a consequence of (14) we obtain the $\mathscr{J}$-monotonicity of the fundamental matrix of (S_λ). We recall the terminology from [12, p. 7] saying that a matrix $M \in \mathbb{C}^{2n\times 2n}$ is $\mathscr{J}$-nondecreasing if $i\,M^*\mathscr{J}M \geq i\mathscr{J}$, and it is $\mathscr{J}$-nonincreasing if $i\,M^*\mathscr{J}M \leq i\mathscr{J}$. Similarly we define the corresponding notions of a $\mathscr{J}$-increasing and $\mathscr{J}$-decreasing matrix. These concepts are used in [12] to study the stability zones for continuous time periodic linear Hamiltonian systems. In a similar way, such stability zones are studied in [13, 14] for discrete linear Hamiltonian systems (6) and in [7] for discrete symplectic systems (S_λ) with $\mathscr{S}^{[0]}_k = I$.

Corollary 3 *Fix $\lambda \in \mathbb{C}$ with $|\lambda| < \varepsilon$ and assume (23). Let $\Phi(\lambda)$ be a fundamental matrix of system (S_λ) such that $\Phi_0(\lambda)$ is complex symplectic, i.e., $\Phi^*_0(\lambda)\,\mathscr{J}\,\Phi_0(\lambda) = \mathscr{J}$. Then for every $k \in [0,\infty)_{\mathbb{Z}}$ the matrix $\Phi_k(\lambda)$ is $\mathscr{J}$-nondecreasing or $\mathscr{J}$-nonincreasing depending on whether $\mathrm{im}(\lambda) > 0$ or $\mathrm{im}(\lambda) < 0$. Moreover, under Hypothesis 4 the $\mathscr{J}$-monotonicity of $\Phi_k(\lambda)$ is strict for $k \geq N_0 + 1$.*

Proof By applying (14) to the fundamental matrix $\Phi_k(\lambda)$ we get

$$i\,\Phi^*_k(\lambda)\,\mathscr{J}\,\Phi_k(\lambda) - i\mathscr{J} = 2\,\mathrm{im}(\lambda) \sum_{j=0}^{k-1} \Phi^*_j(\lambda)\,\Psi_j(\lambda)\,\Phi_j(\lambda). \tag{33}$$

By (23), the sum in (33) is nonnegative, so that $\Phi_k(\lambda)$ is $\mathscr{J}$-nondecreasing when $\mathrm{im}(\lambda) > 0$, and it is $\mathscr{J}$-nonincreasing when $\mathrm{im}(\lambda) < 0$. Moreover, under Hypothesis 4 the sum in (33) is positive definite for $k \geq N_0+1$, so that $\Phi_k(\lambda)$ is $\mathscr{J}$-increasing when $\mathrm{im}(\lambda) > 0$, and it is $\mathscr{J}$-decreasing when $\mathrm{im}(\lambda) < 0$. □

Acknowledgments The first author was supported by the Czech Science Foundation under grant P201/10/1032. The second author was supported by the Program of "Employment of Newly Graduated Doctors of Science for Scientific Excellence" (grant number CZ.1.07/2.3.00/30.0009) co-financed from European Social Fund and the state budget of the Czech Republic.

References

1. Atkinson, F.V.: Discrete and Continuous Boundary Problems. Academic Press, London (1964)
2. Bohner, M., Došlý, O., Kratz, W.: An oscillation theorem for discrete eigenvalue problems. Rocky Mt. J. Math. **33**(4), 1233–1260 (2003)
3. Bohner, M., Došlý, O., Kratz, W.: Sturmian and spectral theory for discrete symplectic systems. Trans. Amer. Math. Soc. **361**(6), 3109–3123 (2009)
4. Bohner, M., Sun, S.: Weyl–Titchmarsh theory for symplectic difference systems. Appl. Math. Comput. **216**(10), 2855–2864 (2010)
5. Clark, S., Gesztesy, F.: On Weyl–Titchmarsh theory for singular finite difference Hamiltonian systems. J. Comput. Appl. Math. **171**(1–2), 151–184 (2004)
6. Clark, S., Zemánek, P.: On a Weyl–Titchmarsh theory for discrete symplectic systems on a half line. Appl. Math. Comput. **217**(7), 2952–2976 (2010)
7. Došlý, O.: Symplectic difference systems with periodic coefficients: Krein's traffic rules for multipliers. Adv. Differ. Equ. **2013**(85), 13 (2013)
8. Došlý, O.: Symplectic difference systems: natural dependence on a parameter. In: Proceedings of the International Conference on Differential Equations, Difference Equations, and Special Functions (Patras, 2012), Adv. Dyn. Syst. Appl., **8**(2), 193–201 (2013)
9. Došlý, O., Kratz, W.: Oscillation theorems for symplectic difference systems. J. Differ. Equ. Appl. **13**(7), 585–605 (2007)
10. Jirari, A.: Second order Sturm-Liouville difference equations and orthogonal polynomials. Mem. Amer. Math. Soc. **113**(542), 138 (1995)
11. Kratz, W., Šimon Hilscher, R.: A generalized index theorem for monotone matrix-valued functions with applications to discrete oscillation theory. SIAM J. Matrix Anal. Appl. **34**(1), 228–243 (2013)
12. Krein, M. G.: Foundations of the theory of λ-zones of stability of a canonical system of linear differential equations with periodic coefficients. In: Leifman, L.J. (ed.) Four Papers on Ordinary Differential Equations, Amer. Math. Soc. Transl. (2), Vol. 120, pp. 1–70, AMS (1983)
13. Răsvan, V.: Stability zones for discrete time Hamiltonian systems. Arch. Math. (Brno) **36**(5), 563–573 (2000)
14. Răsvan, V.: On stability zones for discrete-time periodic linear Hamiltonian systems. Adv. Differ. Equ. **2006**, Art. 80757, 1–13 (2006)
15. Shi, Y.: Spectral theory of discrete linear Hamiltonian systems. J. Math. Anal. Appl. **289**(2), 554–570 (2004)
16. Shi, Y.: Weyl–Titchmarsh theory for a class of discrete linear Hamiltonian systems. Linear Algebra Appl. **416**, 452–519 (2006)
17. Šimon Hilscher, R.: Oscillation theorems for discrete symplectic systems with nonlinear dependence in spectral parameter. Linear Algebra Appl. **437**(12), 2922–2960 (2012)
18. Šimon Hilscher, R., Zemánek, P.: Weyl disks and square summable solutions for discrete symplectic systems with jointly varying endpoints. Adv. Differ. Equ. **2013**(232), 18 (2013)
19. Šimon Hilscher, R., Zemánek, P.: Weyl–Titchmarsh theory for discrete symplectic systems with general linear dependence on spectral parameter. J. Differ. Equ. Appl **20**(1), 84–117 (2014)

Dynamic Selection Systems and Replicator Equations

Zdeněk Pospíšil

Abstract The dynamic replicator equation is inferred from a generalized Kolmogorov-type dynamic system that is called selection system. This way, both systems have the same dimension. The main result shows that the replicator equation is in a certain sense equivalent to a selection system of lower dimension. Corollaries demonstrating connections with known results are also presented. The equations are interpreted as models of biological evolution on different time scales. Hence, the results show a link between ecology and evolution at least on the level of mathematical models.

1 Introduction

There are two general theories dealing with biological evolution. One of them—the ecology—aims to describe interactions among population and their environment and consequent changes in population abundances. Suitable tools for this purpose are Kolmogorov-type systems of differential equations

$$u_i' = u_i r_i(u_1, u_2, \dots, u_m), \quad i = 1, 2, \dots, m,$$

and their discrete counterparts, see e.g. [10]. Here, r_i denotes growth rate of the ith population affected by all of the populations forming the community. The other one—the evolution theory in the strict sense—is interested in change of gene frequencies, traits or so one. These phenomena can be described by replicator equations

$$x_i' = x_i \big(f_i(x_1, x_2, \dots, x_n) - \Phi(x_1, x_2, \dots, x_n) \big), \quad i = 1, 2, \dots, n,$$

where f_i and Φ denote fitness of the ith species and an average fitness, respectively, that depend on structure of the modelled community, see e.g. [7].

Z. Pospíšil(✉)
Department of Mathematics and Statistics, Masaryk University, Kotlářská 2,
611 37 Brno, Czech Republic
e-mail: pospisil@math.muni.cz

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_11

A trivial but noteworthy fact is that ecology and evolution operate on different time scales. Although the previous sentence uses the word "time scale" in a common meaning, not in the technical one, the mentioned different passing of the ecological and evolutionary times hints to embody both mathematical ecology and evolution theory in the unified framework of dynamic equations, i.e. equations on time scales in the strict sense [1].

There is a huge literature on continuous and discrete equations of population dynamics (see e.g. [5, 6] and references therein), as well as on continuous and discrete replicator dynamics ([3, 9] and references therein). Hence, it may be useful to point out a general character common to all of them. This is the aim of the presented paper.

The subsequent section introduce a time scale form of the replicator equation. It is inferred from the selection system, i.e. from the system describing evolution of interacting populations. This part follows a terminology and ideas that G.P. Karev used for continuous systems [4]. The main result is presented in Sect. 3 and it consists in the proof of certain equivalence of dynamic replicator and selection systems. Theorem 2 generalizes a result of Hofbauer [2] obtained for a special continuous replicator equation and a Lotka-Volterra system. The replicator equation with linear fitness function is of a particular interest, it describes an evolutionary game dynamics introduced independently by Taylor and Jonker [11] and Schuster and Sigmund [8]. Hence, some corollaries of Theorem 2 for these systems are presented.

The notation used is standard. Vectors are denoted by bold italic letters, the symbol v_i or $(\boldsymbol{v})_i$ denote the ith entry of the vector $\boldsymbol{v}$. The symbol $\mathbb{R}_+$ denotes the set of non-negative reals, the sets

$$S_n = \left\{ \boldsymbol{x} \in \mathbb{R}_+^n : \sum_{i=1}^n x_i = 1 \right\}, \quad S_n^\circ = \left\{ \boldsymbol{x} \in S_n : \prod_{i=1}^n x_i > 0 \right\}, \quad \partial S_n = S_n \setminus S_n^\circ$$

are (probability) simplex, its relative interior and boundary, respectively. Throughout the paper the symbols σ, Δ, and μ denote the forward jump operator, the Hilger derivative and the graininess function on a considered time scale $\mathbb{T}$, respectively. The "regressive minus" operation is defined by the formula

$$a \ominus b = \frac{a-b}{1+\mu b};$$

for details of time scale calculus see the excellent monograph [1].

2 The Equations and Their Interpretation

The dynamic selection system consists of the equations

$$u_i^\Delta = u_i r_i(\boldsymbol{u}), \quad i = 1, 2, \ldots, m, \tag{1}$$

where $r_i : \mathbb{R}^m_+ \to \mathbb{R}$ are positively regressive rd-continuous functions such that the system with the initial condition

$$\boldsymbol{u}(t_0) = \boldsymbol{u}_0 \in \mathbb{R}^m_+ \tag{2}$$

has exactly one solution.

Proposition 1 *Let* $\boldsymbol{u}$ *be a solution of the problem* (1), (2)*. Then* $u_i(t_0) = 0$ *implies* $u_i(t) = 0$ *and* $u_i(t_0) > 0$ *implies* $u_i(t) > 0$ *for each* $t \in \mathbb{T}$ *and any* $i \in \{1, 2, \ldots, m\}$.

Proof The first statement follows from the assumed uniqueness of the solution. If $t_1 \in \mathbb{T}$ is right scattered and $u_i(t_1) > 0$, then

$$u_i^\sigma(t_1) = u_i(t_1)\Big(1 + \mu(t_1) r_i\big(\boldsymbol{u}(t_1)\big)\Big) > 0$$

by the supposed positive regressivity of the function r_i; if $t_1 \in \mathbb{T}$ is right (left) dense, then there exists a right (left) neighborhood of t_1 such that $u(t) > 0$ on it. Hence, the statement follows from the induction principle [1, Theorem 1.7]. □

The proposition shows that the selection system (1) can be considered as a deterministic model of an isolated biological community formed by interacting populations. Here, u_i is interpreted as a size (population density, biomass) of the ith population. The modeled community consists of constant number of populations; none of them goes to extinction in a finite time, neither immigration nor emigration is taken into account.

Let $\boldsymbol{g} : \mathbb{T} \times S_m \to \mathbb{R}^m$ be a map. We define the average function $\bar{g} : \mathbb{T} \times S_m \to \mathbb{R}$ by

$$\bar{g}(t, \boldsymbol{x}) = \sum_{i=1}^{m} x_i g_i(t, \boldsymbol{x}) = \boldsymbol{x}^\mathsf{T} \boldsymbol{g}(t, \boldsymbol{x}).$$

Theorem 1 *Let the vector function* $\boldsymbol{u} : \mathbb{T} \to \mathbb{R}^m$ *be a solution of the initial value problem for the selection system* (1), (2) *with* $\sum_{i=1}^{m} u_i(t_0) > 0$*. Put*

$$N(t) = \sum_{i=1}^{m} u_i(t), \qquad x_i(t) = \frac{u_i(t)}{N(t)}, \quad i = 1, 2, \ldots, m.$$

Then there are rd-continuous and positively regressive functions $f_i : \mathbb{T} \times S_m \to \mathbb{R}$*,* $i = 1, 2, \ldots, m$ *such that the vector function* $\boldsymbol{x}$ *solves the dynamic equations*

$$x_i^\Delta = x_i\big(f_i(t, \boldsymbol{x}) \ominus \bar{f}(t, \boldsymbol{x})\big), \quad i = 1, 2, \ldots, m. \tag{3}$$

Proof First note that the functions x_i are non-negative according to Proposition 1 and they satisfy

$$\sum_{i=1}^{m} x_i = \frac{1}{N}\sum_{i=1}^{m} u_i = 1,$$

hence, $\boldsymbol{x} \in S_m$. The linearity of the Hilger derivative operator yields

$$N^\Delta = \sum_{j=1}^{m} u_j^\Delta = \sum_{j=1}^{m} u_j r_j(\boldsymbol{u}), \quad \text{i.e. } N^\sigma = N + \mu N^\Delta = \sum_{j=1}^{m} u_j\big(1 + \mu r_j(\boldsymbol{u})\big).$$

Let the functions $f_i : \mathbb{T} \times S_m$ be defined by the formulae

$$f_i(t, \boldsymbol{x}) = r_i\big(N(t)\boldsymbol{x}\big), \quad i = 1, 2, \ldots, m.$$

Then the functions f_i are rd-continuous. Further,

$$1 + \mu(t) f_i(t, \boldsymbol{x}) = 1 + \mu(t) r_i\big(N(t)\boldsymbol{x}\big) = 1 + \mu(t) r_i\big(\boldsymbol{u}(t)\big) > 0,$$

hence the functions f_i are positively regressive. Now, we have

$$\begin{aligned} x_i^\Delta &= \left(\frac{u_i}{N}\right)^\Delta = \frac{u_i^\Delta N - u_i N^\Delta}{N N^\sigma} = \frac{1}{N^\sigma}\left(u_i r_i(\boldsymbol{u}) - u_i \sum_{j=1}^{m} \frac{u_j}{N} r_j(\boldsymbol{u})\right) \\ &= \frac{u_i}{\sum\limits_{j=1}^{m} u_j\big(1 + \mu r_j(\boldsymbol{u})\big)}\left(r_i(\boldsymbol{u}) - \sum_{j=1}^{n} \frac{u_j}{N} r_j(\boldsymbol{u})\right) \\ &= \frac{x_i}{\sum\limits_{j=1}^{m} x_j\big(1 + \mu r_j(N\boldsymbol{x})\big)}\left(r_i(N\boldsymbol{x}) - \sum_{j=1}^{m} x_j r_j(N\boldsymbol{x})\right) = x_i \frac{r_i(N\boldsymbol{x}) - \sum\limits_{j=1}^{m} x_j r_j(N\boldsymbol{x})}{1 + \mu \sum\limits_{j=1}^{m} x_j r_j(N\boldsymbol{x})}, \end{aligned}$$

since $\sum\limits_{j=1}^{m} x_j = 1$. The proof is complete. □

The theorem suggests a form of a dynamic equation describing an evolution of the community structure. We shall take an "autonomous form" of the Eq. (3).

Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ be an rd-continuous and positively regressive map. *The dynamic replicator equation* is the vector equation with the components

$$x_i^\Delta = x_i\left(f_i(\boldsymbol{x}) \ominus \bar{f}(\boldsymbol{x})\right), \quad i = 1, 2, \ldots, n. \tag{4}$$

Further, assume that all of the functions f_i are such that any initial value problem for the Eq. (4) with the initial condition

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 \tag{5}$$

possesses exactly one solution.

Proposition 2 *Let the vector function $\boldsymbol{x} : \mathbb{T} \to \mathbb{R}^n$ be the solution of the problem (4), (5). If $x_i(t_0) = 0$ for an index i then $x_i(t) = 0$ for all $t \in \mathbb{T}$. If $\boldsymbol{x}_0 \in S_n$, then $\boldsymbol{x}(t) \in S_n$ for all $t \in \mathbb{T}$.*

A simple consequence of the proposition is that $\boldsymbol{x}_0 \in S_n^\circ$ and $\boldsymbol{x}_0 \in \partial S_n$ imply $\boldsymbol{x}(t) \in S_n^\circ$ and $\boldsymbol{x}(t) \in \partial S_n$ for all $t \in \mathbb{T}$, respectively.

Proof The first statement follows immediately from the assumed uniqueness of the solution.

Let us observe now that the positive regressivity of the functions f_i yields positive regressivity of the function $\bar{f}$ on the set S_n. Indeed, for any $\boldsymbol{x} \in S_n$ and $t \in \mathbb{T}$, we have

$$1 + \mu(t)\bar{f}(\boldsymbol{x}) = \sum_{i=1}^{n} x_i + \mu(t)\sum_{i=1}^{n} x_i f_i(\boldsymbol{x}) = \sum_{i=1}^{n} x_i\big(1 + \mu(t) f_i(\boldsymbol{x})\big) > 0.$$

Consequently, the functions $f_i \ominus \bar{f}$ are positively regressive on S_n, since

$$1 + \mu(t)\frac{f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})}{1 + \mu(t)\bar{f}(\boldsymbol{x})} = \frac{1 + \mu(t) f_i(\boldsymbol{x})}{1 + \mu(t)\bar{f}(\boldsymbol{x})} > 0.$$

In a similar way as in the previous proposition, we can show that $x_i(t_0) \geq 0$ implies $x_i(t) \geq 0$ for all $t \in \mathbb{T}$.

Let $\boldsymbol{x}$ be a solution of the Eq. (4) and define the function $y : \mathbb{T} \to \mathbb{R}$ by the formula $y(t) = \sum_{i=1}^{n} x_i(t) - 1$. Then

$$\begin{aligned}
y^{\Delta}(t) &= \sum_{i=1}^{n} x_i(t)\Big(f_i\big(\boldsymbol{x}(t)\big) \ominus \bar{f}\big(\boldsymbol{x}(t)\big)\Big) = \sum_{i=1}^{n} x_i(t)\frac{f_i\big(\boldsymbol{x}(t)\big) - \bar{f}\big(\boldsymbol{x}(t)\big)}{1 + \mu(t)\bar{f}\big(\boldsymbol{x}(t)\big)} \\
&= \frac{1}{1 + \mu(t)\bar{f}\big(\boldsymbol{x}(t)\big)}\left(\sum_{i=1}^{n} x_i(t) f_i\big(\boldsymbol{x}(t)\big) - \bar{f}\big(\boldsymbol{x}(t)\big)\sum_{i=1}^{n} x_i(t)\right) \\
&= \frac{\bar{f}\big(\boldsymbol{x}(t)\big)}{1 + \mu(t)\bar{f}\big(\boldsymbol{x}(t)\big)}\left(1 - \sum_{i=1}^{n} x_i(t)\right) = -\frac{\bar{f}\big(\boldsymbol{x}(t)\big)}{1 + \mu(t)\bar{f}\big(\boldsymbol{x}(t)\big)}\, y(t) \\
&= \big[\ominus\bar{f}\big(t, \boldsymbol{x}(t)\big)\big]\, y(t).
\end{aligned}$$

That is, the function y solves the linear homogeneous dynamic equation

$$y^{\Delta} = \left[\ominus \bar{f}\big(\boldsymbol{x}(t)\big)\right] y$$

which yields $y(t) = y_0 \left[e_{\ominus \bar{f}(\boldsymbol{x}(\cdot))}(t, t_0)\right]$. Hence, if $y_0 = 0$ then $y(t) = 0$ for all $t \in \mathbb{T}$, cf. [1, Theorem 2.48]. Subsequently, if $\boldsymbol{x}_0 \in S_n$ then $\sum_{i=1}^{n} x_i(t) = 1$ for all $t \in \mathbb{T}$. That is, the second statement of the proposition holds. □

Proposition 2 enables one to interpret the solution of the Eq. (4) as an evolution of frequencies of some replicated—i.e. neither appearing nor vanishing—entities, e.g. genes, memes or traits present in a population or in a community. For brevity, we will use an accustomed terminology and call these entities "quasispecies". Then, the entries of vector $\boldsymbol{x}$ can be interpreted as frequencies of quasispecies forming a biological community and the replicator equation (4) as a model of evolving structure of it. The function f_i expresses a fitness of the i th quasispecies and the fitness depends on time and on structure of the community. Hence, the replicator equation precises (or specifies) the basic tenet of Darwinism that the evolutionary success of a (quasi)species depends on its fitness: the relative change of the i th quasispecies frequency equals the regressive difference of its fitness and of the average one.

3 Equivalence of Selection and Replicator Systems

Now, we are ready to prove the main result of the paper: the statement that the dynamic selection system and the replicator equation are in a certain sense equivalent.

Theorem 2 *Let the vector function $\boldsymbol{x} : \mathbb{T} \to S_n$ be the solution of the initial value problem (4), (5) with the initial value $\boldsymbol{x}_0 \in S_n^\circ$. Then there exist a time scale $\tilde{\mathbb{T}}$, one-to-one continuous maps $\varphi : \mathbb{T} \to \tilde{\mathbb{T}}$, $F : S_n^\circ \to (0, \infty)^{n-1}$, and positively regressive rd-functions $r_i : \tilde{\mathbb{T}} \to \mathbb{R}$, $i = 1, 2, \dots, n-1$ such that $F\big(\boldsymbol{x}(t)\big) = \boldsymbol{u}\big(\varphi(t)\big)$, where the vector function $\boldsymbol{u} : \tilde{\mathbb{T}} \to \mathbb{R}_+^{n-1}$ is the solution of the problem (1), (2) with $m = n - 1$ and $\boldsymbol{u}_0 = F(\boldsymbol{x}_0)$.*

Proof For $\boldsymbol{y} \in S_n^\circ$ define

$$F(\boldsymbol{y}) = \boldsymbol{v} \in \mathbb{R}_+^{n-1}, \quad \text{where} \quad v_i = \frac{y_i}{y_n}, \quad i = 1, 2, \dots, n-1. \tag{6}$$

The map F is obviously continuous on S_n°. Summation of the previous equalities gives

$$\sum_{j=1}^{n-1} v_j = \frac{1}{y_n} \sum_{j=1}^{n-1} v_j = \frac{1 - y_n}{y_n}, \quad \text{hence} \quad \frac{1}{y_n} = 1 + \sum_{j=1}^{n-1} v_j.$$

That is, the map F is invertible and $\boldsymbol{y} = F^{-1}(\boldsymbol{v})$ is given by the formulae

$$y_i = v_i \left(1 + \sum_{j=1}^{n-1} v_j\right)^{-1}, \quad i = 1, 2, \ldots, n-1, \qquad y_n = \left(1 + \sum_{j=1}^{n-1} v_j\right)^{-1}. \quad (7)$$

Let $\varphi : \mathbb{T} \to \mathbb{R}$ be defined by the integral

$$\varphi(t) = \int_{t_0}^{t} x_n(s) \Delta s \quad (8)$$

and put $\tilde{\mathbb{T}} = \{\varphi(t) : t \in \mathbb{T}\}$. Since $x_n(t) > 0$ for all $t \in \mathbb{T}$, the function φ is strictly increasing and, consequently, $\varphi : \mathbb{T} \to \tilde{\mathbb{T}}$ is bijection. The properties of the time scale integral imply that the function φ is rd-continuous. Hence, the set $\tilde{\mathbb{T}}$ is a time scale.

The graininess $\tilde{\mu}$ of the time scale $\tilde{\mathbb{T}}$ is given by

$$\tilde{\mu}(\tau) = \tilde{\mu}\left(\varphi^{-1}(t)\right) = \int_{t_0}^{\sigma(t)} x_n(s)\Delta s - \int_{t_0}^{t} x_n(s)\Delta s = \int_{t}^{\sigma(t)} x_n(s)\Delta s = \mu(t)x_n(t),$$

or briefly

$$\tilde{\mu} = \mu x_n. \quad (9)$$

Let the functions $u_i : \tilde{\mathbb{T}} \to \mathbb{R}, i = 1, 2, \ldots, n-1$ satisfy $\boldsymbol{u}(\tau) = F\Big(\boldsymbol{x}\big(\varphi^{-1}(\tau)\big)\Big)$, i.e.

$$u_i(\tau) = \frac{x_i\left(\varphi^{-1}(\tau)\right)}{x_n\left(\varphi^{-1}(\tau)\right)}, \quad i = 1, 2, \ldots, n-1. \quad (10)$$

Let $g : \tilde{T} \to \mathbb{R}$ be a differentiable function. For clarity, we denote the Hilger derivative on the time scale $\tilde{\mathbb{T}}$ by the symbol $\tilde{\Delta}$. The chain rule [1, Theorem 1.93] yields

$$(g \circ \varphi)^{\Delta}(t) = g^{\tilde{\Delta}}\big(\varphi(t)\big)\varphi^{\Delta}(t) = x_n(t) g^{\tilde{\Delta}}\big(\varphi(t)\big),$$

that is, for $\tau \in \tilde{\mathbb{T}}$ we have

$$g^{\tilde{\Delta}}(\tau) = \frac{1}{x_n\left(\varphi^{-1}(\tau)\right)} (g \circ \varphi)^{\Delta}\big(\varphi^{-1}(\tau)\big), \qquad \text{briefly} \quad g^{\tilde{\Delta}} = \frac{1}{x_n} g^{\Delta}.$$

Now, we can calculate

$$\big(f_i(\boldsymbol{x})\ominus\bar f(\boldsymbol{x})\big)-\big(f_n(\boldsymbol{x})\ominus\bar f(\boldsymbol{x})\big)=\frac{f_i(\boldsymbol{x})-\bar f(\boldsymbol{x})}{1+\mu\bar f(\boldsymbol{x})}+\frac{f_n(\boldsymbol{x})-\bar f(\boldsymbol{x})}{1+\mu\bar f(\boldsymbol{x})}=\frac{f_i(\boldsymbol{x})-f_n(\boldsymbol{x})}{1+\mu\bar f(\boldsymbol{x})},$$
$$x_n^\sigma\,\big(1+\mu\bar f(\boldsymbol{x})\big)=x_n\Big(1+\mu\,\big(f_n(\boldsymbol{x})\ominus\bar f(\boldsymbol{x})\big)\Big)\big(1+\mu\bar f(\boldsymbol{x})\big)=x_n\big(1+\mu f_n(\boldsymbol{x})\big),$$

and, subsequently according to (10), we have

$$u_i^{\tilde\Delta}=\frac{1}{x_n}\left(\frac{x_i}{x_n}\right)^{\Delta}=\frac{1}{x_n}\frac{x_i^\Delta x_n-x_ix_n^\Delta}{x_nx_n^\sigma}=\frac{x_i}{x_n}\frac{\big(f_i(\boldsymbol{x})\ominus\bar f(\boldsymbol{x})\big)-\big(f_n(\boldsymbol{x})\ominus\bar f(\boldsymbol{x})\big)}{x_n^\sigma}$$
$$=\frac{x_i}{x_n}\frac{f_i(\boldsymbol{x})-f_n(\boldsymbol{x})}{x_n\big(1+\mu f_n(\boldsymbol{x})\big)}.$$

Denoting $h_i(\boldsymbol{x})=f_i(\boldsymbol{x})/x_n$ and considering (9) we obtain

$$u_i^{\tilde\Delta}=\frac{x_i}{x_n}\frac{h_i(\boldsymbol{x})-h_n(\boldsymbol{x})}{1+\mu x_nh_n(\boldsymbol{x})}=u_i\frac{h_i\big(F^{-1}(\boldsymbol{u})\big)-h_n\big(F^{-1}(\boldsymbol{u})\big)}{1+\tilde\mu h_n\big(F^{-1}(\boldsymbol{u})\big)}$$
$$=u_i\Big[h_i\big(F^{-1}(\boldsymbol{u})\big)\ominus h_n\big(F^{-1}(\boldsymbol{u})\big)\Big],$$

where the "regressive minus" $\ominus$ is related to the time scale $\tilde{\mathbb{T}}$. Hence, the relations (7), (10) show that the functions $r_i:\mathbb{R}_+^{n-1}\to\mathbb{R}$, $i=1,2,\dots,n-1$ defined by the formulae

$$r_i(\boldsymbol{u})=H_i(\boldsymbol{u})\ominus H_n(\boldsymbol{u}),\quad\text{where}\quad H_i\big(\boldsymbol{u}\big)=\left(1+\sum_{j=1}^n u_j\right)f_i\big(F^{-1}(\boldsymbol{u})\big)\tag{11}$$

satisfy the statement of the theorem. □

We can reformulate the statement of Theorem 2—the systems (1) and (4) are qualitatively equivalent; the term "qualitative equivalence" means that there is a homeomorphism $S_n^\circ\to(0,\infty)^{n-1}$ that maps orbits of one equation to the orbits of the second one.

The proof of the theorem reveals that one can take the set

$$\tilde{\mathbb{T}}=\left\{\int_{t_0}^{t}x_n(s)\Delta s\,:\,t\in\mathbb{T}\right\}$$

as the time scale for the selection system (1) equivalent to the replicator equation (4). Then, the maps F, φ, and the functions r_i are defined by equalities (6), (8), and (11), respectively. The choice of x_n for the construction of the time scale $\tilde{\mathbb{T}}$ is not substantial, any other component of the vector function $\boldsymbol{x}$ solving the Eq. (4) is applicable as well.

If the time scale $\mathbb{T}$ is discrete, i.e. $\mu(t) > 0$ for all $t \in \mathbb{T}$, then the replicator equation (4) can be rewritten to the form

$$x_i^\sigma = x_i \frac{1 + \mu f_i(\boldsymbol{x})}{1 + \mu \bar{f}(\boldsymbol{x})}.$$

In particular, if $\mathbb{T} = \mathbb{Z}$, the replicator equation takes the form

$$x_i(t+1) = x_i(t) \frac{1 + f_i\big(\boldsymbol{x}(t)\big)}{1 + \bar{f}\big(\boldsymbol{x}(t)\big)}.$$

The replicator equation with linear fitness functions f_i, i.e. with the functions

$$f_i(\boldsymbol{x}) = \sum_{j=1}^{n} a_{ij} x_j = (\mathsf{A}\boldsymbol{x})_i ,$$

where A is a constant $n \times n$ matrix, is of a particular interest. It describes a game dynamics and the matrix A can be interpreted as a payoff matrix, cf. [3]. The functions f_i are positively regressive if

$$-\mu(t)\,(\mathsf{A}\boldsymbol{x})_i < 1, \quad i = 1, 2, \ldots, n$$

for all $t \in \mathbb{T}, \boldsymbol{x} \in S_n$. The condition is satisfied for any time scale if the matrix A is nonnegative; this constraint is not restrictive for payoff matrices at all.

The replicator equation with linear fitness functions takes the form

$$x_i^\Delta = x_i \left((\mathsf{A}\boldsymbol{x})_i \ominus \boldsymbol{x}^\mathsf{T} \mathsf{A}\boldsymbol{x} \right), \quad i = 1, 2, \ldots, n, \tag{12}$$

and it is equivalent to the selection system

$$u_i^\Delta = u_i \left[\left(\sum_{j=1}^{n-1} a_{ij} u_j + a_{in} \right) \ominus \left(\sum_{j=1}^{n-1} a_{nj} u_j + a_{nn} \right) \right] \quad i = 1, 2, \ldots, n-1. \tag{13}$$

Indeed, the term H_i in the equality (11) takes the form

$$H_i(\boldsymbol{u}) = \left(1 + \sum_{j=1}^{n-1} u_j \right) f_i\left(F^{-1}(\boldsymbol{u}) \right) = \left(1 + \sum_{j=1}^{n-1} u_j \right) \left[\mathsf{A} \frac{1}{1 + \sum\limits_{j=1}^{n-1} u_j} \begin{pmatrix} u_1 \\ \vdots \\ u_{n-1} \\ 1 \end{pmatrix} \right]_i$$

$$= \sum_{j=1}^{n-1} a_{ij} u_j + a_{in}.$$

We have to keep in mind that the time scale for the dynamic replicator equation (12) is different from the one for the dynamic selection system (13).

We finish with two corollaries of Theorem 2. The $n \times n$ matrix A with the entries a_{ij} is non-negative in both of them.

Corollary 1 (equivalence of replicator and Lotka-Volterra equations, [2]) *The differential replicator equation*

$$x_i' = x_i \left((\mathsf{A}\boldsymbol{x})_i - \boldsymbol{x}^{\mathsf{T}} \mathsf{A}\boldsymbol{x} \right), \quad i = 1, 2, \ldots, n.$$

is qualitatively equivalent to the Lotka-Volterra differential system

$$u_i' = u_i \left(c_i - (\mathsf{B}\boldsymbol{x})_i \right), \quad i = 1, 2, \ldots, n-1,$$

where $c_i = a_{in} - a_{nn}$, $b_{ij} = a_{nj} - a_{ij}$.

Corollary 2 *There exists a strictly increasing sequence* $\{\tau_k\}$ *of reals such that the difference replicator equation*

$$x_i(t+1) = \frac{1 + \left(\mathsf{A}\boldsymbol{x}(t)\right)_i}{1 + \boldsymbol{x}(t)^{\mathsf{T}} \mathsf{A}\boldsymbol{x}(t)}, \quad i = 1, 2, \ldots, n,$$

is qualitatively equivalent to the difference equation

$$u_i(\tau_{k+1}) = \frac{1 + (\Delta\tau_k)\left(\sum\limits_{j=1}^{n-1} a_{ij} u_j(\tau_k) + a_{in} \right)}{1 + (\Delta\tau_k)\left(\sum\limits_{j=1}^{n-1} a_{nj} u_j(\tau_k) + a_{nn} \right)} u_i(\tau_k), \quad i = 1, 2, \ldots, n-1.$$

There exists a strictly increasing sequence $\{t_k\}$ *of reals such that the difference equation*

$$u_i(\tau+1) = \frac{1 + a_{in} + \sum\limits_{j=1}^{n-1} a_{ij} u_j(\tau)}{1 + a_{nn} + \sum\limits_{j=1}^{n-1} a_{nj} u_j(\tau)} u_i(\tau), \quad i = 1, 2, \ldots, n-1,$$

is qualitatively equivalent to the difference replicator equation

$$x_i(t_{k+1}) = \frac{1 + (\Delta t_k)\big(\mathsf{A}\boldsymbol{x}(t_k)\big)_i}{1 + (\Delta t_k)\boldsymbol{x}(t_k)^{\mathsf{T}}\mathsf{A}\boldsymbol{x}(t_k)}, \quad i = 1, 2, \ldots, n.$$

Proof Multiplying the equation (13) by μ and adding x_i to the both sides of it, we obtain

$$u_i^\sigma = u_i + \mu u_i^\Delta = u_i \frac{1 + \mu\left(\sum_{j=1}^{n-1} a_{ij}u_j + a_{in}\right)}{1 + \mu\left(\sum_{j=1}^{n-1} a_{nj}u_j + a_{nn}\right)}.$$

□

Acknowledgments This research was supported by the grant No. GAP201/10/1032 of the Czech Grant Agency.

References

1. Bohner, M., Peterson, A.: Dynamic Equations on Time Scales. Birkhauser, Boston (2001)
2. Hofbauer, J.: On the occurrence of limit cycles in the Volterra-Lotka equation. Nonlinear Anal. **5**, 1003–1007 (1981)
3. Hofbauer, J., Sigmund, K.: Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge (2002)
4. Karev, G.P.: On mathematical theory of selection: continuous time population dynamics. J. Math. Biol. **60**, 107–129 (2010)
5. Murray, J.: Mathematical Biology I. Springer, Berlin (2002)
6. Murray, J.: Mathematical Biology II. Springer, Berlin (2003)
7. Nowak, M.A.: Evolutionary Dynamics. Harvard University Press, Cambridge (2006)
8. Schuster, P., Sigmund, K.: Coyness, philandering and stable strategies. Anim. Behavior **29**, 186–192 (1981)
9. Sigmund, K.: The Calculus of Selfishness. Princeton University Press, Princeton (2010)
10. Svirezhev, Y.M., Logofet, D.O.: Stability of Biological Communities. MIR, Moscow (1983)
11. Taylor, P.D., Jonker, L.: Evolutionary stable strategies and game dynamics. Math. Biosci. **40**, 145–156 (1978)

Asymptotic Equivalence of Difference Equations in Banach Space

Andrejs Reinfelds

Abstract Conjugacy technique is applied to analysis asymptotic equivalence of nonautonomous linear and semilinear difference equations in Banach space.

1 Introduction

The well-known Levinson's theorem [8] states that if the trivial solution of

$$x' = Ax \tag{1}$$

is uniformly stable, where $x \in \mathbb{R}^n$, and

$$\int_0^\infty |B(t)|\,dt < +\infty$$

then (1) and

$$x' = Ax + B(t)x$$

are asymptotically equivalent.

There are many studies in the literature which deal with asymptotic equivalence problem in the theory of differential equations, difference equations and equations on time scale at n-dimensional space $\mathbb{R}^n$. See in particular [1–11] and the references citied therein.

A. Reinfelds (✉)
Institute of Mathematics and Computer Science, Raiņa bulvāris 29, Rīga, Latvia

A. Reinfelds
Faculty of Physics and Mathematics, University of Latvia, Zeļu iela 8, Rīga, Latvia
e-mail: reinf@latnet.lv

Z. AlSharawi et al. (eds.), *Theory and Applications of Difference Equations and Discrete Dynamical Systems*, Springer Proceedings in Mathematics & Statistics 102,
DOI 10.1007/978-3-662-44140-4_12

In the present paper asymptotical equivalence of linear and semilinear nonautonomous difference equations is proved in Banach space. We prove that corresponding difference equations are conjugated and using this conjugacy we find sufficient conditions for asymptotical equivalence.

Consider the following difference equations in Banach space $\mathbf{E}$

$$x_1(t+1) = F_1(t, x_1(t)), \tag{2}$$

$$x_2(t+1) = F_2(t, x_2(t)). \tag{3}$$

Let us denote by $x_1(\cdot, s, x)\colon [s, +\infty) \to \mathbf{E}$ and by $x_2(\cdot, s, x)\colon [s, +\infty) \to \mathbf{E}$ the solutions of equations (2) and (3) satisfying the initial conditions $x_1(s, s, x) = x$ and $x_2(s, s, x) = x$ respectively.

Definition 1 Difference equations (2) and (3) are conjugate if there exists a homeomorphism $H(t, \cdot)\colon \mathbf{E} \to \mathbf{E}$ such that $|H(t, x)||x|^{-1}$ is uniformly bounded and

$$H(t, x_1(t, s, x)) = x_2(t, s, H(s, x)) \quad \text{for } t \geq s.$$

Definition 2 Difference equations (2) and (3) are asymptotically equivalent, if there exists a homeomorphism $H(s, \cdot)\colon \mathbf{E} \to \mathbf{E}$ it such that

$$\lim_{t \to +\infty} (x_2(t, s, H(s, x)) - x_1(t, s, x)) = 0.$$

Let us consider linear and semilinear difference equation in Banach space

$$x(t+1) = A(t)x(t) + f(t, x(t)) \tag{4}$$

$$x(t+1) = A(t)x(t) \tag{5}$$

where the map $A(t)\colon \mathbf{E} \to \mathbf{E}$ is invertible. We will assume that map $F\colon [s, +\infty) \times \mathbf{E} \to \mathbf{E}$ satisfy the Lipschitz condition

$$|f(t, x) - f(t, x')| \leq \gamma(t)|x - x'|$$

and (4) has the equilibrium point $x = 0$,

$$f(t, 0) = 0.$$

Let $X(t, s)$ $t, s \in \mathbb{Z}$ be the Cauchy evalutionary operator of (5), where $X(s, s) = I$ and I identity operator.

Our basic assumption is that

$$\Phi(s) = \sup_{r \geq s} \sum_{i=r}^{+\infty} |X(r, i+1)||X(i, r)|\gamma(i) < \infty \tag{6}$$

and

$$\lim_{s \to +\infty} \Phi(s) = 0. \tag{7}$$

Remark 1 If

$$|X(t, s)| \leq M < +\infty \text{ for all } t, s \in \mathbb{Z}$$

then conditions (6) and (7) reduces to condition

$$\sum_{i=s}^{+\infty} \gamma(i) < \infty.$$

Lemma 1 *If s is large enough, then the following estimate is valid*

$$\sum_{i=s}^{+\infty} |X(s, i+1)|\gamma(i)|x(i, s, x)| \leq \frac{\Phi(s)|x|}{1 - \Phi(s)}.$$

Proof The solution of (4) satisfying the initial condition $x(s, s, x) = x$ for $t > s$ is given by the formula

$$x(t, s, x) = X(t, s)x + \sum_{i=s}^{t-1} X(t, i+1) f(i, x(i, s, x)).$$

We have estimate

$$|X(s, t+1)|\gamma(t)|x(t, s, x)| \leq |X(s, t+1)|\gamma(t)|X(t, s)||x|.$$

$$+ |X(s, t+1)|\gamma(t) \sum_{i=s}^{t-1} |X(t, i+1)|\gamma(i)|x(i, s, x)|.$$

Summing up for t's with respect $t \geq s$, we obtain

$$\sum_{t=s}^{T} |X(s, t+1)|\gamma(t)|x(t, s, x)| \leq \sum_{t=s}^{T} |X(s, t+1)|\gamma(t)|X(t, s)||x|$$

$$+\sum_{t=s+1}^{T}|X(s,t+1)|\gamma(t)\sum_{i=s}^{t-1}|X(t,i+1)|\gamma(i)|x(i,s,x)|$$

$$\le \sum_{t=s}^{T}|X(s,t+1)|\gamma(t)|X(t,s)||x|$$

$$+\sum_{i=s}^{T}|X(s,i+1)\gamma(i)|x(i,s,x)|\sum_{t=i+1}^{T}|X(t,s)||X(s,t+1)|\gamma(t)$$

$$\le \sum_{t=s}^{+\infty}|X(s,t+1)|\gamma(t)|X(t,s)||x|$$

$$+\sum_{i=s}^{T}|X(s,i+1)\gamma(i)|x(i,s,x)|\sum_{t=s}^{+\infty}|X(t,s)||X(s,t+1)|\gamma(t)$$

Therefore

$$\sum_{i=s}^{T}|X(s,i+1)|\gamma(i)|x(i,s,x)| \le \sum_{i=s}^{+\infty}|X(s,i+1)|\gamma(i)|x(i,s,x)| \le \frac{\Phi(s)|x|}{1-\Phi(s)}.$$

Let us note that

$$|x(t,s,x)| \le |X(t,s)|\left(|x| + \sum_{i=s}^{+\infty}|X(s,i+1)|\gamma(i)|x(i,s,x)|\right) \le \frac{|X(t,s)||x|}{1-\Phi(s)}.$$

□

2 Conjugacy of Difference Equations

Let us consider the following semilinear difference equations in Banach space **E**

$$x_1(t+1) = A(t)x_1(t) + f_1(t, x_1(t)) \tag{8}$$

$$x_2(t+1) = A(t)x_2(t) + f_2(t, x_2(t)) \tag{9}$$

where

$$|f_1(t,x) - f_1(t,x')| \le \gamma(t)|x - x'|$$
$$|f_2(t,x) - f_2(t,x')| \le \gamma(t)|x - x'|$$

and

$$f_1(t,0) = 0, \quad f_2(t,0) = 0$$

Theorem 1 *Let $\Phi(s) < 1/2$. Then difference equations* (8) *and* (9) *are conjugate.*

Proof Consider the set of continuous maps

$$\mathfrak{M} = \left\{ h\colon [s,+\infty) \times \mathbf{E}, \mathbf{E}) \;\middle|\; \sup_{t,x} \frac{|h(t,x)|}{|x|} < +\infty \right\}$$

It is easy to see that $\mathfrak{M}$ is a Banach space with the norm

$$\|h\| = \sup_{t,x} \frac{|h(t,x)|}{|x|} < +\infty.$$

We will seek the map establishing the conjugacy of (8) and (9) in the form $H_1(t,x) = x + h_1(t,x)$. We examine the following functional equation

$$h_1(s,x) = \sum_{i=s}^{+\infty} X(s,i+1)(f_1(i,x_1(i,s,x)) - f_2(i,x_1(i,s,x) + h_1(i,x_1(i,s,x)))). \tag{10}$$

Let us consider the map $h_1 \mapsto \mathscr{L} h_1$, $h_1 \in \mathfrak{M}$ defined by the equality

$$\mathscr{L}h_1(s,x) = \sum_{i=s}^{+\infty} X(s,i+1)(f_1(i,x_1(i,s,x)) - f_2(i,x_1(i,s,x) + h_1(i,x_1(i,s,x)))).$$

First we obtain

$$|\mathscr{L}0| \le 2\sum_{i=s}^{+\infty} |X(s,i+1)|\gamma(i)|x_1(i,s,x)| \le \frac{2\Phi(s)|x|}{1-\Phi(s)}.$$

Then $\|\mathscr{L}(0)\| \le \frac{2\Phi(s)}{1-\Phi(s)}$. Next we get

$$|\mathscr{L}h_1(s,x) - \mathscr{L}h_1'(s,x)| \le \frac{\Phi(s)|x|}{1-\Phi(s)}\|h - h'\|.$$

We get that the map $\mathscr{L}$ is a contraction and consequently the functional equation (10) has a unique solution in $\mathfrak{M}$.

$$\begin{aligned}
h_1(t, x_1(t,s,x)) &= \sum_{i=t}^{+\infty} X(t,i+1)(f_1(i,x_1(i,s,x)) - f_2(i,x_1(i,s,x)+h_1(i,x_1(i,s,x))))\\
&= -\sum_{i=s}^{t-1} X(t,i+1)(f_1(i,x_1(i,s,x)) - f_2(i,x_1(i,s,x)+h_1(i,x_1(i,s,x))))\\
&\quad + \sum_{i=s}^{+\infty} X(t,i+1)(f_1(i,x_1(i,s,x)) - f_2(i,x_1(i,s,x)+h_1(i,x_1(i,s,x))))\\
&= \sum_{i=s}^{t-1} X(t,i+1) f_2(i,x_1(i,s,x)+h_1(i,x_1(i,s,x)))\\
&\quad - x_1(t,s,x) + X(t,s)(x+h_1(s,x)).
\end{aligned}$$

Consequently, we have

$$x_1(t,s,x) + h_1(t, x_1(t,s,x)) = x_2(t,s,x+h_1(s,x)).$$

Changing the roles of f_1 and f_2, we prove in the same way the existence of h_2 that satisfies the equality

$$x_2(t,s,x) + h_2(t, x_2(t,s,x)) = x_1(t,s,x+h_2(s,x))$$

Designing $H_1(t,x) = x + h_1(t,x)$, $H_2(t,x) = x + h_2(t,x)$, we get

$$\begin{aligned}
H_1(t, H_2(t, x_2(t,s,x))) &= x_2(t,s,H_1(s,H_2(s,x))),\\
H_2(t, H_1(t, x_1(t,s,x))) &= x_2(t,s,H_2(s,H_1(s,x))).
\end{aligned}$$

Taking into account uniqueness of mappings $H_2(t, H_1(t,\cdot)) - id$ and $H_1(t, H_2(t,\cdot)) - id$ in $\mathfrak{M}$ we have $H_2(t, H_1(t,\cdot)) = id$ and $H_1(t, H_2(t,\cdot)) = id$ and therefore $H_1(t,\cdot)$ is a homeomorphism establishing a conjugacy of the (8) and (9). □

3 Asymptotic Equivalence of Difference Equations

Theorem 2 *If*

$$\lim_{t\to+\infty} |X(t,s)|\Phi(t) = 0 \tag{11}$$

where

$$\Phi(s) = \sup_{r \geq s} \sum_{i=r}^{+\infty} |X(r, i+1)||X(i, r)|\gamma(i) < \infty$$

and

$$\lim_{s \to +\infty} \Phi(s) = 0$$

then difference equations (4) *and* (5) *are asymptotically equivalent.*

Proof Let $f_2(t, x) \equiv 0$, $f_1(t, x) = f(t, x)$. Then Theorem 1 implies that

$$X(t, s)(x + h_1(s, x)) - x_1(t, s, x) = h_1(t, x_1(t, s, x)).$$

We have the estimate

$$\begin{aligned} |h_1(t, x_1(t, s, x))| &= \left| \sum_{i=t}^{+\infty} X(t, i+1) f(i, x_1(i, s, x)) \right| \\ &\leq \sum_{i=t}^{+\infty} |X(t, i+1)| \gamma(i) |x_1(i, t, x_1(t, s, x))| \\ &\leq \frac{\Phi(t)|x_1(t, s, x)|}{1 - \Phi(t)} \leq \frac{\Phi(t)|X(t, s)||x|}{(1 - \Phi(t))(1 - \Phi(s))}. \end{aligned}$$

It follows that difference equations (4) and (5) are asymptotically equivalent. □

Acknowledgments This work has been partially supported by the grant No. 345/2012 of the Latvian Council of Science

References

1. Akhmet, M.U., Tleubergenova, M.A., Zafer, A.: Asymptotic equivalence of differential equations and asymptotically almost periodic solutions. Nonlinear Anal. **67**(6), 1870–1877 (2007)
2. Bodine, S., Lutz, D.A.: On asymptotic equivalence of perturbed linear systems of differential and difference equations. J. Math. Anal. Appl. **326**(2), 1174–1189 (2007)
3. Bohner, M., Lutz, D.A.: Asymptotic behavior of dynamic equations on time scales. J. Differ. Equ. Appl. **7**(1), 21–50 (2001)
4. Brauer, F.: Asymptotic equivalence and asymptotic behaviour of linear systems. Mich. Math, J. **9**, 33–43 (1962)
5. Choi, S.K., Koo, N.J., Ryo, H.S.: Asymptotic equivalence between two difference systems. Comput. Math. Appl. **45**(6–9), 1327–1337 (2003)
6. Elaydi, S.: Asyptotics for linear difference equations I: basic theory. J. Differ. Equ. Appl. **5**(6), 563–589 (1999)
7. Kaymakcalan, B, Mert, R, Zafer, A.: Asymptotic equivalence of dynamic systems on time scale, Discrete Contin. Dynam. Systems 2007 Suppl. 558–567 (2007)

8. Levinson, N.: The asymptotic nature of solutions of linear systems of differential equations. Duke Math. J. **15**, 111–126 (1948)
9. Medina, R., Pinto, M.: Dichotomies and asymptotic equivalence of nonlinear difference systems. J. Differ. Equ. Appl. **5**(3), 287–303 (1999)
10. Yakubovich, V.A.: On the asymptotic behavior of the solutions of a system of differential equations. Mat. Sb. **28**, 217–240 (1951)
11. Zafer, A.: On asymptotic equivalence of linear and quasilinear difference equations. Appl. Anal. **84**(9), 899–908 (2005)