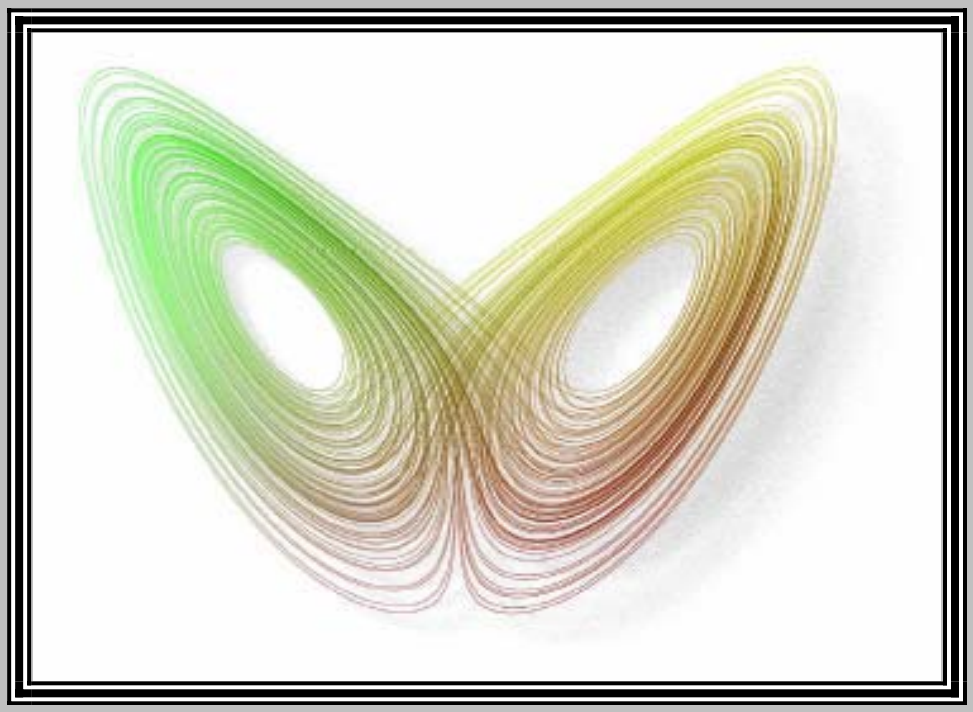
Volume. 1., 2012 ISSN 2253-0371

Annual Review of Chaos Theory, Bifurcations and Dynamical Systems



Annual Review of Chaos Theory, Bifurcations and Dynamical Systems

Aims and scope

Annual Review of Chaos Theory, Bifurcations and Dynamical Systems (*ARCTBDS*) (www.arctbds.com) is a multidisciplinary international peer reviewed journal of chaos theory, bifurcations and dynamical systems publishing high-quality articles quarterly. The primary objective of this review journal is to provide a forum for this multidisciplinary discipline of chaos theory, bifurcations and dynamical systems as well as general nonlinear dynamics. This review journal will be a periodical or series that is devoted to the publication of review (and original research) articles that summarize the progress in particular areas of chaos theory, bifurcations and dynamical systems during a preceding period. Also, the journal publishes original articles and contributions on the above topics by leading researchers and developers.

Subject Coverage

Annual Review of Chaos Theory, Bifurcations and Dynamical Systems (*ARCTBDS*) (www.arctbds.com) covers all aspects of chaos theory, bifurcations and dynamical systems. Topics of interest include, but are not limited to:

- Mathematical modeling, computational methods, principles and numerical simulations
- Chaos, bifurcation, nonlinear dynamical systems, complexity in nonlinear science and engineering and numerical methods for nonlinear differential equations.
- Fractals, pattern formation, solitons, coherent phenomena and nonlinear fluid dynamics.
- Control theory and stability and singularity on fundamental or applied studies.
- Real applications in all areas of science.

Papers can be oriented towards theory, algorithms, numerical simulations, applications or experimentation.

Readership and audience

Advanced undergraduates and graduate students in natural and human sciences and engineering such as physics, chemistry, biology or bioinformatics…etc; academics and practitioners in nonlinear physics and in various other areas of potential application; researchers, instructors, mathematicians, nonlinear scientists and electronic engineers interested in chaos, nonlinear dynamics and dynamical systems and all interested in nonlinear sciences.

Annual Review of Chaos Theory, Bifurcations and Dynamical Systems

Submission of Papers

Manuscripts should be in English and should be written in a LaTeX file only. All submissions should be sent electronically to the journal website:

http://www.easychair.org/conferences/?conf=arctbds

The other formats are not acceptable. Authors are advised to keep a copy of their manuscript since the journal cannot accept responsibility for lost copies. The submitted papers to this journal should not have been previously published nor be currently under consideration for publication elsewhere. All papers are refereed through a double blind process. There is no financial reward for reviewers and editors of this journal. These positions are purely voluntary. When the manuscript is accepted for publication, the authors agree to automatic transfer of the copyright to the journal. The manuscript should contain the following items: An informative title, author's name, address and e-mail, abstract, 3-5 keywords, text, conclusion, and references. Footnotes should be avoided if possible. References should be denoted in the text by numbers in square brackets, e.g. [10]. References should be complete, in the following style:

- **Papers:** Author(s) initials followed by last name for each author, "paper title," publication name, volume, inclusive page numbers, month and year. Authors should consult Mathematical Reviews for standard abbreviations of journal names..
- **Books:** Author(s), title, publisher, location, year, chapter or page numbers (if desired).
- **Each figure and table or any material** (i.e., Fig. 10, Table. 1, ...etc) should be mentioned in the text and numbered consecutively using Arabic numerals. Number each table consecutively using Arabic numerals. Type a brief title below each figure and table or any material. Figures should be submitted separately as encapsulated postscript (.eps) files.

For accepted papers, the author(s) will be asked to transfer copyright of the article to the journal. The manuscript will not be published until the **Copyright Transfer Form** is received.

Page proofs will be sent to the corresponding author. The proofs must be corrected and returned within three days of receipt. **There are no charges for publishing in this journal and each author will receive a PDF copy of his/her paper.**

Contribution Enquiries and Submitting

Editorial office (Algeria office): Annual Review of Chaos Theory, Bifurcations and Dynamical Systems (ARCTBDS) Dr. Zeraoulia Elhadj, Department of Mathematics, University of Tébessa, (12002), Algeria, e-mail: zeraoulia@mail.univ-tebessa.dz and zelhadj12@yahoo.fr.

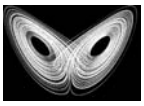
Annual Review of Chaos Theory, Bifurcations and Dynamical Systems

Annual Review of Chaos Theory, Bifurcations and Dynamical Systems

Annual Review of Chaos Theory, Bifurcations and Dynamical Systems

Volume. 1., 2012

Table of Contents



Annual Review of Chaos Theory, Bifurcations and Dynamical Systems
Vol. 1, (2012) 1-21, www.arctbds.com.

Dynamics in International Subsidy Games with Unit-elastic Demand

Akio Matsumoto
Department of Economics, Chuo University, 742-1,
Higashi-Nakano, Hachioji, Tokyo, 192-0393, Japan.
e-mail: akiom@tamacc.chuo-u.ac.jp.

Ferenc Szidarovszky
Department of Systems and Industrial Engineering
University of Arizona, Tucson, AZ, 85721-0020, USA.
e-mail: szidar@sie.arizona.edu.

Abstract

In this paper, we construct a three-country model with two governments and two firms and consider dynamic behavior of the sequential subsidy game in which the governments determine their optimal trade policies and then the firms determine their optimal outputs. We first show the existence of an optimal trade policy under realistic conditions. In the case of symmetric firms, the governments adopt periodic mixed trade policy (i.e., one government gives subsidy and the other levies tax in one period and then the governments interchange their policies in the next period) if the adjustment is naive, and the governments adopt a stable mixed policy if adaptive. In the case of asymmetric firms, a firm receives subsidy if its cost is lower and pays tax if higher. If the Cournot output point under the optimal subsidy is locally stable, then its dynamics can be periodic which is synchronized with the periodic trade policy. If it is locally unstable, then complex dynamics involving chaos emerges regardless of the cost difference.

Keywords: three country model, two stage game, "elastic, trade policy, unit-elatsic demand, Neimark-Sacker bifurcation.

Manuscript accepted February 16, 2012.

1 Introduction

Markets become imperfectly competitive due to many factors such as the small number of firms, the differentiated goods, the scale of economics, etc. In such an imperfectly competitive international market, the governments may be motivated to introduce trade policies like tariff, export subsidy and tax to increase national welfare of their countries. It has been demonstrated that an increase in a domestic export subsidy raises the domestic firm's output and its profit when the firms compete in a Cournot way [3]. It has been also demonstrated that an export tax can be optimal when the firms compete in a Bertrand way [4]. It is now well-known that the source of this sharp contrast comes from the difference in the assumption on the firms' strategic behavior (that is, the goods are strategic substitutes or strategic complements). It is also well-known that this behavioral difference relates to the curvature and the elasticity of the demand function. Recently, constructing a simplified version of the three-country model with two firms and two governments, [1] shows, among others, the following clear-cut results when the demand is unit-elastic:

(1) When the foreign government is passive, the optimal trade policy of the active domestic government is free trade if the production costs of the two firms are identical, an export subsidy if the home firm has lower cost and an export tax otherwise.

(2) When both governments are active, a continuum of policy equilibria exist if the production costs are identical and no policy equilibrium exists if they are different.

As a consequence of the second result, no dynamic consideration has been provided yet. In particular, it is not known yet how the optimal policy as well as the optimal outputs change over time and what kinds of changes might occur. The main purpose of this paper is twofold; to consider dynamics of the optimal trade policy and the associated optimal outputs under unit-elastic demand and to show that rich dynamics can be born when natural constraints are imposed on the government's policy selections. This paper complements [1] from a dynamic point of view. It is a continuation of [6] who focus mainly on the comparative statistic analysis of the similar model (i.e., the effects on the optimal outputs caused by a change in the trade policy of the domestic government). The dynamic model of outputs to be considered in this paper resembles nonlinear dynamic duopoly models, which have been extensively studied for the last twenty years. Comprehensive summary of the earlier work has been presented in [7]. More recent developments on this field are given in [2]. This paper also aims to apply the theoretical results obtained so far to the dynamic analysis in the framework of international economics. The paper is organized as follows. Section 2 presents a variant of the three-country model in which both governments are active. Section 3 considers policy dynamics and Section 4 analyzes output dynamics with the optimal trade policy. Section 5 gives concluding remarks.

2 Model

The model presented below is a variant of the three-country model. There are two countries with one firm in each of them, and these firms export their product to a third country. The outputs of the firms are denoted by x and y, and constant marginal costs of the two firms are denoted by c_1 and c_2, respectively. Competition in the third country is modeled

through a two-stage game. At the first stage, the governments hosting their firms choose subsidy rates, s_i for $i = 1, 2$, in order to maximize their welfare, taking the optimal behavior of the firms as given. At the second stage, the firms employ the quantity competition in a Cournot way and choose profit maximizing outputs, taking their governments' trade policies as given. Optimal subsides and optimal outputs are backwardly determined. In particular, we solve the profit maximization problems of the firms, given the levels of the subsidy in Section 2.1, then examine and solve the welfare maximization problems of the governments, given the optimal behavior of the firms in Section 2.2. In order to get a complete description of the dynamics of the subsidy game in the latter part of the paper, we will specify the best reply functions of the firms and those of the governments in this section.

2.1 Profit Maximization

Let the inverse demand function be unit-elastic,

$$P = \frac{1}{Q},$$

where Q is the total output, $Q = x + y$.[1] At the second stage in which the governments' subsidies are given, firm x and firm y choose outputs to maximize their profits defined by

$$\pi_1 = (P - (c_1 - s_1))x,$$

and

$$\pi_2 = (P - (c_2 - s_2))y.$$

The first-order conditions of the profit maximization are given by

$$\frac{\partial \pi_1}{\partial x} = \frac{y}{(x+y)^2} - (c_1 - s_1) = 0,$$

and

$$\frac{\partial \pi_2}{\partial y} = \frac{x}{(x+y)^2} - (c_2 - s_2) = 0,$$

where the second-order conditions are satisfied for any x and y that solve the first-order conditions. We call the production cost including the subsidy an *actual cost*. Although we will formally show later that the actual costs are non-negative, we suppose for the time being that subsidies are given as $c_x > 0$ and $c_y > 0$. From the first-order conditions, the explicit forms of the firms' best reply functions are derived as

$$\bar{r}_1(y) = \sqrt{\frac{y}{c_1 - s_1}} - y \tag{1}$$

and

$$\bar{r}_2(x) = \sqrt{\frac{x}{c_2 - s_2}} - x. \tag{2}$$

[1] See [5] that studies the same model with different demand, $P = Q^{-\lambda}$ and $\lambda \neq 1$.

Any intersection of the best reply functions determines a Cournot equilibrium at which we can solve for the output quantities:

$$x^C(s_1,s_2)=\frac{c_2-s_2}{((c_1-s_1)+(c_2-s_2))^2} \tag{3}$$

and

$$y^C(s_1,s_2)=\frac{c_1-s_2}{((c_1-s_1)+(c_2-s_2))^2}, \tag{4}$$

where superscript C is attached to variables associated with the Cournot point.[2] The Cournot outputs in (3) and (4) are substituted into the profit functions to obtain the Cournot profits:

$$\pi_1^C(s_1,s_2)=\left(\frac{c_2-s_2}{(c_1-s_1)+(c_2-s_2)}\right)^2 \tag{5}$$

and

$$\pi_2^C(s_1,s_2)=\left(\frac{c_1-s_2}{(c_1-s_2)+(c_2-s_1)}\right)^2. \tag{6}$$

Dividing (3) by (4) and (5) by (6) yields, after arranging terms, the ratios of outputs and profits,

$$\frac{x^C}{y^C}=\frac{c_2-s_2}{c_1-s_1}\gtreqless 1 \text{ according to } c_2-s_2\gtreqless c_1-s_1,$$

and

$$\frac{\pi_x^C}{\pi_y^C}=\left(\frac{x^C}{y^C}\right)^2\gtreqless 1 \text{ according to } x^C\gtreqless y^C.$$

These inequalities imply the following results on the optimal behavior of the firms: The firm with the lower actual cost produces more output and earns more profit than the firm with the higher actual cost.

2.2 Welfare Maximization

At the first stage of the sequential game, the governments determine the optimal levels of the subsidy so as to maximize the national welfare defined by

$$W_1(s_1,s_2)=\pi_1^C(s_1,s_2)-s_1x^C(s_1,s_2), \tag{7}$$

and

$$W_2(s_2,s_1)=\pi_2^C(s_2,s_1)-s_2y^C(s_2,s_1). \tag{8}$$

We derive the specific forms of the best reply functions of the governments and consider their characteristics in the policy space. Substituting x^C, $y^C, Q^C=x^C+y^C$ and

[2]Since $\bar{r}_1(y)$ and $\bar{r}_2(x)$ take mound-shaped curves starting at the origin, the curves intersects twice at $(0,0)$ and (x^C,y^C). The former is the trivial equilibrium point and the latter is the non-trivial equilibrium point. Our concern is on the non-trivial point and thus no further considerations are given to the trivial equilibrium point.

$P^C = (Q^C)^{-1}$ into (7) and (8) yields the explicit forms of the welfare functions of the governments,

$$W_1(s_1, s_2) = \frac{(c_2 - s_2 - s_1)(c_2 - s_2)}{((c_1 - s_1) + (c_2 - s_2))^2}, \tag{9}$$

and

$$W_2(s_2, s_1) = \frac{(c_1 - s_1 - s_2)(c_1 - s_1)}{((c_1 - s_1) + (c_2 - s_2))^2}. \tag{10}$$

The government of country 1 maximizes $W_1(s_1, s_2)$ with respect to s_1 and the government of country 2 maximizes $W_2(s_2, s_1)$ with respect to s_2. We can solve the first-order conditions to obtain the best reply functions:

$$r_1(s_2) = -s_2 + (c_2 - c_1) \text{ and } r_2(s_1) = -s_1 + (c_1 - c_2), \tag{11}$$

where the second-order conditions are satisfied. These functions are essentially the same as those derived by [1]. It is apparent that there is a continuum of optimal subsidies $s_1 + s_2 = 0$ for symmetric firms ($c_1 = c_2$) and no equilibrium exists for asymmetric firms ($c_1 \neq c_2$).

To avoid the indeterminacy of the optimal policy equilibrium in the case of asymmetric firms, we impose the following external upper and lower bound constraints on the levels of the optimal policy, s_i, taking account of the fact that the governments behave with control. The first constraint reflects the fact that the governments have the upper bound of the subsidy, due to their budget constraints. The second constraint takes account of the fact that the government does not levy such a strong export tax that might result in its firm to exit the market. Intuitively speaking, in choosing their policies, the governments neither take care of all the production costs nor take all of the profits.

Assumption 1. $s_i^L \leq s_i \leq s_i^U$ for $i = 1, 2$ where s_i^U is the upper bound of the subsidy level defined by $s_i^U = c_i$ and $s_i^L < 0$ is the lower bound of the subsidy level, which shows the upper bound of the export tax.

Under Assumption 1, the best reply function of the government of country 1 becomes piecewise linear with three segments:

$$\begin{cases} s_1^U & s_2 < s_2^u, \\ r_1(s_2) & s_2^u \leq s_2 \leq s_2^\ell, \\ s_1^L & s_2 > s_2^\ell, \end{cases}$$

where s_2^u and s_2^ℓ are defined by $r_1(s_2^u) = s_1^U$ and $r_1(s_2^\ell) = s_1^L$, respectively. In the same way, the best reply function of the government of country 2 is derived to be piecewise-linear with three segments:

$$\begin{cases} s_2^U & s_1 < s_1^u, \\ r_2(s_1) & s_1^u \leq s_1 \leq s_1^\ell, \\ s_2^L & s_1 > s_1^\ell, \end{cases}$$

where s_1^u and $s_1^\ell > 0$ are defined by $r_2(s_1^u) = s_2^U$ and $r_2(s_1^\ell) = s_2^L$, respectively.

An intersection of these modified best reply functions is a Nash equilibrium of the trade policy, (s_1^e, s_2^e). First of all, we should discuss the determination of the optimal

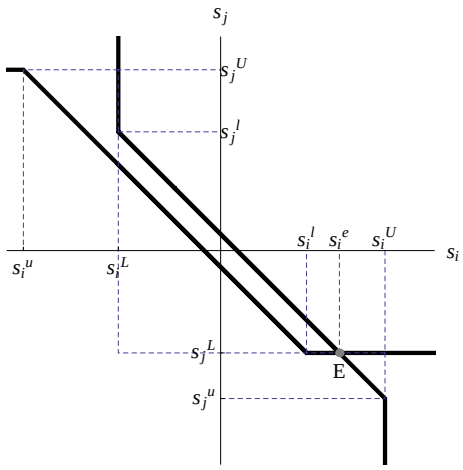


Figure 1: Determination of the optimal policy under $c_i < c_j$.

trade policy in the case where $c_i < c_j$ and $s_j^L \le s_j^u$. By definition of the piecewise linear best replies, it is clear that the optimal subsidy policy of firm i is $s_i^e = s_i^U$. From (3), (4) and Assumption 1 (i.e., $s_i^U = c_i$), the optimal output of firm 2 is zero if $c_1 < c_2$ and $s_2^L \le s_2^u$ whereas the optimal output of firm 1 is zero if $c_2 < c_1$ and $s_1^L \le s_1^u$. In both cases one of the firms will export nothing to the third country and the competition in the third country will be terminated. To confine our attention to the third-country model with active competition, we assume that s_j^u and s_j^L are given such that the following inequality holds.

Assumption 2. $s_j^L > s_j^u$ for $j = 1, 2$.

Notice that $s_j^u = c_j - 2c_i$ and $s_j^\ell = c_j - c_i - s_i^L$, so clearly $s_j^u < s_j^\ell$ for both firms. Assumption 2 requires that for $j = 1, 2$, $s_j^L > s_j^u$. In order to guarantee the existence of negative s_j^L bounds, we make the additional assumption:

Assumption 3. $c_j < 2c_i$ for $j = 1, 2$ and $i \ne j$.

Without losing generality, we can assume that in the case of asymmetric firms, $c_i < c_j$. First we show that $s_i^\ell < s_i^U$. Since $s_j^L > c_j - 2c_i > c_j - 2c_j = -c_j$, we have

$$s_i^\ell = c_i - c_j - s_j^L < c_i - c_j - (-c_j) = c_i = s_i^U.$$

However no such comparison can be made between s_i^L and s_i^ℓ. In order to guarantee that $s_i^L < s_i^\ell$, we make the following assumption:

Assumption 4. $s_i^L + s_j^L < c_i - c_j$.

Notice that this assumption holds if the marginal costs c_i and c_j are close to each other. The best reply functions are shown in Fig. 1 and from it we can conclude that the unique equilibrium is

$$s_i^e = c_j - c_i - s_j^L \text{ and } s_j^e = s_j^L \tag{12}$$

Hence we have the following:

Theorem 1 *Under Assumptions 1,2,3 and 4, if the firms are symmetric (i.e., $c_1 = c_2$), then there are infinitely many equilibria on the line $s_i = -s_j$ whereas if the firms are asymmetric with $c_i < c_j$, then $s_i^e = -s_j^L + (c_j - c_i) > 0$ and $s_j^e = s_j^L < 0$.*

3 Dynamics Analysis of Trade Policy

Theorem 1 guarantees the existence of a unique optimal trade policy. In this section, we consider the dynamic behavior of the trade policy in the policy space (s_1, s_2). To this end, we assume the following discrete time dynamic process of the policy selection:

$$\begin{cases} s_1' = (1-\alpha_1)s_1 + \alpha_1 R_1(s_2), \\ \\ s_2' = (1-\alpha_2)s_2 + \alpha_2 R_2(s_1), \end{cases} \tag{13}$$

where $'$ denotes the unit-time advancement operator, α_i is the adjustment coefficient with $0 < \alpha_i \leq 1$ and $R_1(s_2)$ and $R_2(s_1)$ are the best reply functions restricted to their intervals, $[s_2^L, s_2^U]$ and $[s_1^L, s_1^U]$,

$$R_1(s_2) = \begin{cases} s_1^L \text{ for } s_2^\ell \leq s_2 \leq s_2^U, \\ \\ r_1(s_2) \text{ for } s_2^L \leq s_2 < s_2^\ell, \end{cases}$$

and

$$R_2(s_1) = \begin{cases} s_2^L \text{ for } s_1^\ell \leq s_1 \leq s_1^U, \\ \\ r_2(s_1) \text{ for } s_1^L \leq s_1 < s_1^\ell, \end{cases}$$

3.1 Symmetric firms: $c_1 = c_2$

In the case of identical costs, we first perform some numerical simulations to examine the dynamic behavior of the governments. We then confirm analytically that the numerical results are robust, and finally summarize these results in Theorems 2 and 3. The numerical simulations are presented in Fig. 2 where the policy is best reply with naive expectation (i.e., $\alpha_i = 1$). The other case is given in Fig. 3 where the policy is adaptively adjusted (i.e., $0 < \alpha_i < 1$).

The feasible policy space is defined by the rectangle $S = [s_2^L, s_2^U] \times [s_1^L, s_1^U]$, which is divided into distinctive four parts by the horizontal and vertical lines, $s_1 = s_1^\ell$ and $s_2 = s_2^\ell$,

$$S_I = \{(s_1, s_2) \in S \mid s_1^\ell \leq s_1 \text{ and } s_2^\ell \leq s_2\},$$

$$S_{II} = \{(s_1, s_2) \in S \mid s_1 < s_1^\ell \text{ and } s_2^\ell \leq s_2\},$$

$$S_{III} = \{(s_1, s_2) \in S \mid s_1^\ell \leq s_1 \text{ and } s_2 < s_2^\ell\},$$

$$S_{IV} = \{(s_1, s_2) \in S \mid s_1 < s_1^\ell \text{ and } s_2 < s_2^\ell\}.$$

In Figs. 2(A) and 3(A), we select three initial points denoted as I_1, I_2 and I_3 in S_I, S_{II} and S_{III}, respectively and depict the three trajectories starting from these points. In

Figs. 2(B) and 3(B), which are enlargements of the region S_{IV}, we also select three initial points denoted as i_1, i_2 and i_3 inside S_{IV} and show three trajectories starting from these points. Simulations in Fig. 2(A) indicate that the trajectories converge to period-2 cycles when the trade policies are naively adjusted. The two periodic points are symmetric with respect to the line $s_1 + s_2 = 0$. On the other hand, the simulations shown in Fig. 3(A) indicate that the trajectories converge to stationary points on the line $s_1+s_2 = 0$ when the trade policies are adaptively adjusted. The first result on policy dynamics is summarized as follows:

Theorem 2 *If the firms are symmetric, then the naively adjusted process of the export trade policy (i.e., (13) with $\alpha_1 = \alpha_2 = 1$) gives rise to infinitely many stable period-2 cycles, and a trajectory starting from a point other than a stationary point converges to one of these cycles.*

Proof. We prove this statement with four steps. (I): It can be seen that $R_1(s_2) = s_1^L$ and $R_2(s_1) = s_2^L$ for all $(s_1, s_2) \in S_I$. By the identical cost assumption, $R_1(s_2^L) = s_1^\ell$ and $R_2(s_1^L) = s_2^\ell$ whereas $R_1(s_2^\ell) = s_1^L$ and $R_2(s_1^\ell) = s_2^L$ by the definitions of s_2^ℓ and s_1^ℓ. Thus any trajectory starting at a point inside S_1 converges to the period-2 cycle with periodic points, (s_1^L, s_2^L) and (s_1^ℓ, s_2^ℓ). The trajectory with the initial point I_1 in Fig. 3(A) is an example of this case. (II): Next take an initial point $(s_1^{II}, s_2^{II}) \in S_{II}$. Then the naive adjustment process conveys the point to $R_1(s_2^{II}) = s_1^L$ and $R_2(s_1^{II}) = -s_2^{II}$ and then $R_1(-s_2^{II}) = s_1^{II}$ and $R_2(s_1^L) = s_2^\ell$, that are bounced back to the point $(s_1^L, -s_2^{II})$. Thus any trajectory starting at a point $(s_1^{II}, s_2^{II}) \in S_{II}$ converges to the period-2 cycle with periodic points $(s_1^L, -s_2^{II})$ and (s_1^{II}, s_2^ℓ). The trajectory with the initial point I_2 is an example of this case. (III): In the same way, we can show that a trajectory starting at a point $(s_1^{III}, s_2^{III}) \in S_{III}$ converges to the period-2 cycle with periodic points, $(-s_2^{III}, s_2^L)$ and (s_1^ℓ, s_2^{III}). The trajectory with the initial point I_3 is an example of this case. (IV): We finally consider periodic behavior in Fig. 3(B) where the initial points are selected inside S_{IV}. Since the adjustment process in S_{IV} is given by

$$s_1' = -s_2 \text{ and } s_2' = s_1,$$

the governments expect the symmetric point $(-s_1^{IV}, -s_2^{IV})$ with respect to the $s_1 + s_2 = 0$ locus for any initial point (s_1^{IV}, s_2^{IV}). Taking $(-s_1^{IV}, -s_2^{IV})$ as given, the government expect (s_1^{IV}, s_2^{IV}) in the next time period. Thus any point $(s_1^{IV}, s_2^{IV}) \in S_{IV}$ and its symmetric point $(-s_1^{IV}, -s_2^{IV}) \in S_{IV}$ are period-2 points. Three period-2 cycles depicted in Fig. 2(B) are examples of this case. ■

We say that the governments take the *pure policy* if $s_1 s_2 > 0$, the *mixed policy* if $s_1 s_2 < 0$, and the *one-side free trade policy* if $s_1 s_2 = 0$ and $s_i \neq 0$ for $i = 1, 2$. Furthermore, we say that the governments take a pure subsidy policy if $s_1 > 0$ and $s_2 > 0$ and a pure tax policy if $s_1 < 0$ and $s_2 < 0$. In the mixed policy, one government pays subsidy and the other government charges tax. Figs. 2(A) and 2(B) are divided into four rectangles by the horizontal and vertical axes. The rectangles on the top-right and the bottom-left represent the set of points which generate pure policy (i.e., the pure subsidy policy and the pure tax policy). On the other hand the rectangles on the top-left and on the bottom-right represent the set of points which generate mixed policy. If a point is on the either

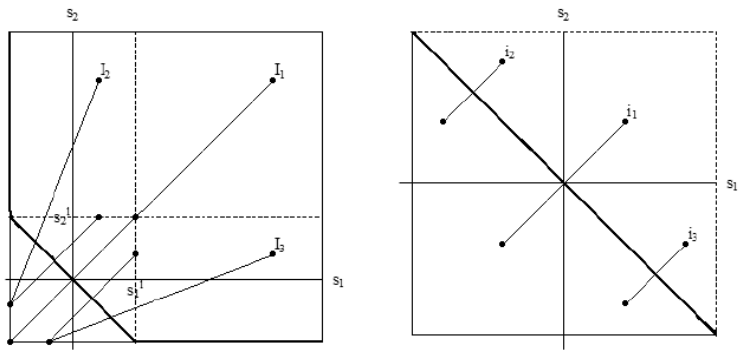


Figure 2: Coexistence of period-2 cycle under naive expectation.

axis, it represents the one-side free trade in which one government gives no subsidy and charges no tax and the other government either gives subsidy or charges tax.

Let us denote the period-2 points of the trade policy by $S^A = (s_1^A, s_2^A)$ and $S^B = (s_1^B, s_2^B)$. The period-2 cycle means that if the governments expect S^A, then S^B is realized; and if the governments expect S^B in the next period, then S^A is realized. Let us take the case of the pure trade policy. If both governments expect that their competitors take the same policy, for example export subsidies $s_1^A > 0$ and $s_2^A > 0$, then in the next period the process becomes the opposite pure policy, that is, export taxes $s_1^B < 0$ and $s_2^B < 0$. So the governments alternate between the pure subsidy policy and the pure tax policy. In contrast to this, if both governments expect that their competitors take the mixed policy, for example, government 1 adopts the subsidy policy $s_1^A > 0$ while government 2 takes the tax policy $s_2^A < 0$, then, in the next time period the process becomes the subsidy with $s_1^B = -s_2^A$ for government 1 and the tax with $s_2^B = -s_1^A$ for government 2. The governments therefore alternate between the mixed policies in which the values and the signs of the subsidy and the tax are interchanged. We summarize these results as a corollary of Theorem 2:

Corollary 1 *The trade policy has an initial point dependency: (1) If both governments expect that their competitors adopt the subsidy policy, then the tax policy is realized, and vice versa. (2) If a mixed policy is expected, then the opposite mixed policy is realized where the realized point of the subsidy or the tax is the mirror image of the expected point with respect to the -45° degree line.*

The optimal outputs associated to these periodic points are obtained by substituting the periodic points into (3) and (4): for $k = A, B$,

$$x^{Ck} = \frac{c_2 - s_2^k}{((c_1 - s_1^k) + (c_2 - s_2^k))^2} \text{ and } y^{Ck} = \frac{c_1 - s_1^k}{((c_1 - s_1^k) + (c_2 - s_2^k))^2}. \tag{14}$$

We now assume that the policy is adaptively adjusted (i.e., $\alpha_i < 1$) and the adjustment coefficients are the same (i.e., $\alpha_1 = \alpha_2 = \alpha$) for the sake of analytical simplicity. As can be seen in Fig. 3(A), any trajectory with an initial point inside $S_I \cup S_{II} \cup S_{III}$ sooner or latter enters S_{IV}. It is therefore sufficient for our purpose to consider the dynamics observed within S_{IV}. Our second result on policy dynamics is summarized as follows:

Theorem 3 *If the firms are symmetric, then the symmetric adaptive adjustment process of the export trade policy (i.e., (13) with $\alpha_1 = \alpha_2 < 1$) is stable and converges to a point on the line $s_1 + s_2 = 0$.*

Proof. Let us start with an initial point $(s_1^0, s_2^0) \in S_{IV}$. The optimal policy at the next period is determined by the adaptively adjusted process,

$$s_1^{'} = (1-\alpha)s_1^0 + \alpha(-s_2^0),$$

$$s_2^{'} = (1-\alpha)s_2^0 + \alpha(-s_1^0).$$

The line passing through these two points, (s_1^0, s_2^0) and (s_1', s_2'), is written as $s_2 = as_1 + b$ where the slope a and the vertical intercept b are

$$a = \frac{s_2' - s_2^0}{s_1' - s_1^0} = 1 \text{ and } b = s_2^0 - s_1^0.$$

It is clear that the adaptive process maps the optimal policy (s_1', s_2') to a point on the $s_2 = as_1 + b$ locus. That is, the trajectories of the optimal policies are controlled by

$$s_1^{'} = (1-\alpha)s_1 + \alpha(-s_1 - b),$$

$$s_2^{'} = (1-\alpha)s_2 + \alpha(-s_2 + b).$$

The adjusted processes are independent and governed by 1D difference equations with slopes less than unity in absolute value,

$$\left|\frac{ds_i^{'}}{ds_i}\right| = |1 - 2\alpha| < 1.$$

Hence the adjustment process is stable and a trajectory converges oscillating to the stationary state associated with the initial point (s_1^0, s_2^0),

$$s_1^e = -\frac{s_2^0 - s_1^0}{2} \text{ and } s_2^e = \frac{s_2^0 - s_1^0}{2}$$

which is the intersection of the $s_2 = -s_1$ and the $s_2 = as_1 + b$ curves. It is clear that $s_1^e \gtreqless 0$ and $s_2^e \lesseqgtr 0$ if $s_1^0 \gtreqless 0$ and $s_2^0 \lesseqgtr 0$. ■

The stationary point (s_1^e, s_2^e) is on the line $s_1 + s_2 = 0$. Fig. 3(B) shows that the trajectory starting from point i_1 converges to the origin, $s_1^e = s_2^e = 0$, the trajectory starting from point i_2 converges to point α with $s_1^\alpha < 0$, $s_2^\alpha > 0$ and $s_1^\alpha = -s_2^\alpha$, and the trajectory starting from i_3 converges to point β with $s_1^\beta > 0$, $s_2^\beta < 0$ and $s_1^\beta = -s_2^\beta$. This observation leads us to Corollary 2.

Corollary 2 *The optimal trade policy has an initial point dependency: if both governments start with the same initial expectations, then the free trade is materialized (i.e., $s_1^e = s_2^e = 0$); If they start with different initial expectations (i.e., $s_1^0 \neq s_2^0$), then the symmetric mixed trade policy is materialized:*

$$s_1^e = -\frac{s_2^0 - s_1^0}{2} \text{ and } s_2^e = -s_1^e.$$

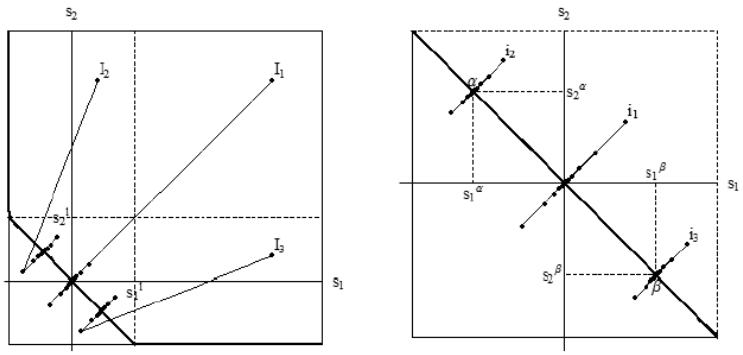


Figure 3: Coexistence of stable stationary points under adaptive expectation.

The corresponding stationary output values are obtained by substituting s_1^e and s_2^e into (3) and (4),

$$x^C == \frac{c_2 - s_2^e}{((c_1 - s_1^e) + (c_2 - s_2^e))^2} \text{ and } y^C = \frac{c_1 - s_1^e}{((c_1 - s_1^e) + (c_2 - s_2^e))^2}. \tag{15}$$

3.2 Asymmetric firms: $c_1 \neq c_2$

It is clear from Fig. 1 that the optimal policy is stable in the policy space if the firms are asymmetric, regardless of whether the policy is naively or adaptively adjusted. It is also clear that the optimal trade policy is mixed, which is summarized as follows:

Theorem 4 *If the firms are asymmetric, then the firm with lower production cost receives an export subsidy and the firm with higher production cost pays an export tax.*

The optimal output values are determined by substituting the optimal subsidies (12) into relations (3) and (4).

4 Dynamic Analysis of Output

In considering the output dynamics in the international subsidy game, we let the output variables in (1) and (2) construct the adaptively adjusted output process:

$$\begin{cases} x' = (1 - \beta_1)x + \beta_1 \left(\sqrt{\dfrac{y}{c_1 - s_1}} - y \right), \\ \\ y' = (1 - \beta_2)y + \beta_2 \left(\sqrt{\dfrac{x}{c_2 - s_2}} - x \right), \end{cases} \tag{16}$$

where β_i is the adjustment coefficient satisfying $0 < \beta_i \leq 1$, and s_1 and s_2 are governed by (13). As mentioned in the Introduction, the dynamic structure of system (16) resembles that of the nonlinear Cournot models extensively studied in [2-7] in which different subjects such as the emergence of complex dynamics involving chaos, multistability, the structure of the basin of attraction, delay dynamics, etc., are discussed. We skip the detailed examinations of system (16) and will apply these results to our dynamic analysis.

4.1 Symmetric firms: $c_1 = c_2$

As a benchmark, we take $\beta_1 = \beta_2 = 1$ and consider the output dynamics under naive expectation. Let us denote the actual cost ratio by

$$k = \frac{c_1 - s_1}{c_2 - s_2}.$$

It is already shown in [7] that loss of stability occurs when the actual cost ratio satisfies the following equation,[3]

$$\frac{(k-1)^2}{4k} = 1 \tag{17}$$

where the smaller solution is $3-2\sqrt{2}(\simeq 0.172)$ and the larger solution is $3+2\sqrt{2}(\simeq 5.828)$. It follows that if the actual cost ratio stays within the interval bounded by the smaller and larger solutions, then the Cournot point is stable. In the same way if the actual cost ratio falls outside the interval, it becomes locally unstable. We assume the stability of the Cournot point and examine the effects caused by the policy on output dynamics for a while. It is also shown in [7] that the dynamics is symmetric with respect to $k = 1$. In order to get new results, we confine our consideration to the case of $c_1 < c_2$ and assume that $c_1 = 1$ throughout the analysis for the sake of analytical convenience. It can be checked that the nonnegativity of the output trajectories is guaranteed when the actual cost ratio is at least $4/25(= 0.16)$.

If there is no policy lag in the sense that the firms receive the subsidies from the governments without any time delays, and the policy is naively adjusted, then the dynamics of the outputs and the subsidies are controlled by the dynamic equations

$$\begin{cases} x' = \sqrt{\dfrac{y}{c_1 - s_1}} - y, \\ y' = \sqrt{\dfrac{x}{c_2 - s_2}} - x, \\ s_1' = -s_2, \\ s_2' = -s_1, \end{cases} \tag{18}$$

where the domain of the policy dynamics is restricted to the rectangle, $[s_1^L, s_1^\ell] \times [s_2^L, s_2^\ell]$ for analytical simplicity.[4] According to Theorem 2, the optimal trade policy oscillates between two points, $S^A = (s_1^A, s_2^A)$ and $S^B = (s_1^B, s_2^B)$ where $s_1^A = -s_2^B$ and $s_2^A = -s_1^B$. We have already solved for the output quantities at each periodic point given in (14). Indeed, the Cournot outputs at point S^A are

$$x^{CA} = \frac{c_2 - s_2^A}{((c_1 - s_1^A) + (c_2 - s_2^A))^2} \text{ and } y^{CA} = \frac{c_1 - s_1^A}{((c_1 - s_1^A) + (c_2 - s_2^A))^2}$$

[3]This equation is obtained by setting the product of the derivatives of the best reply functions, $\bar{r}_1(y)$ and $\bar{r}_2(x)$, evaluated at the Cournot point equal to -1.

[4]As can be seen in Fig. 2(A), any trajectory starting at a point outside S_{IV} will enter S_{IV} after several iterations.

and the Cournot outputs at point S^B are

$$x^{CB} = \frac{c_2 - s_2^B}{((c_1 - s_1^B) + (c_2 - s_2^B))^2} \text{ and } y^{CB} = \frac{c_1 - s_1^B}{((c_1 - s_1^B) + (c_2 - s_2^B))^2}.$$

Since the output dynamics depends on the policy dynamics but not vice versa, we can be fairly certain that the output dynamics gives rise to a periodic cycle when the trade policy has a period-2 cycle. We specify the parameter values as $c_1 = c_2 = 1$, $S^A = (0.64, -0.4)$ and $S^B = (0.4, -0.64)$ and perform simulations. First of all, it can be pointed out that the output equilibria, $C^A = (x^{CA}, y^{CA})$ and $C^B = (x^{CB}, y^{CB})$, are locally asymptotically stable under these specifications if the trade policy is fixed since the cost ratios are greater than $3 - 2\sqrt{2}$,

$$k^A = \frac{c_1 - s_1^A}{c_2 - s_2^A} \simeq 0.366 \text{ and } k^B = \frac{c_1 - s_1^B}{c_2 - s_2^B} \simeq 0.257.$$

However, the trade policy is not fixed, it is switched from one periodic point to the other in every period. Fig. 4 reveals that the output dynamics is represented by a period-2 cycle, which is synchronized with the period-2 cycle of the optimal subsidy. Fig. 4(A) shows a return map. The best reply functions of firm 2 are illustrated as mound-shaped curves and the inner curve is shifted to the outer curve when the policy is switched from S^A to S^B. In the same way, the best reply functions of firm 1 are illustrated as two upward sloping curves and the shift from the left curve to the right is caused by the policy switching from S^A to S^B. Here C^a and C^b are the two periodic points of the output cycle. Fig. 4(B) depicts the time trajectory of output y. Contrary to our intuition, the periodic points of the output cycle are not the Cournot points denoted by C^A and C^B in Fig. 4(A). The reason is that dynamic equations of the outputs are switched from $(x', y') = (\bar{r}_1^A(y), \bar{r}_2^A(x))$ with S^A to $(x', y') = (\bar{r}_1^B(y), \bar{r}_2^B(x))$ with S^B at every iteration step where

$$\bar{r}_1^A(y) = \sqrt{\frac{y}{c_1 - s_1^A}} - y \text{ and } \bar{r}_2^A(x) = \sqrt{\frac{x}{c_2 - s_2^A}} - x$$

and

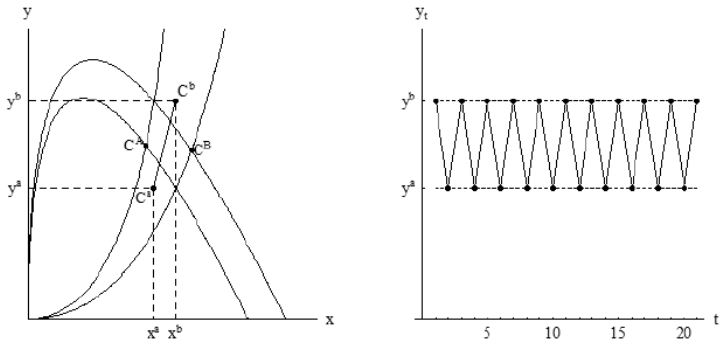
$$\bar{r}_1^B(y) = \sqrt{\frac{y}{c_1 - s_1^B}} - y \text{ and } \bar{r}_2^B(x) = \sqrt{\frac{x}{c_2 - s_2^B}} - x.$$

The periodic points $C^a = (x^a, y^a)$ and $C^b = (x^b, y^b)$ are the fixed points of the composite functions of $\bar{r}_i^A$ and $\bar{r}_i^B$ for $i = 1, 2$,

$$x^a = \bar{r}_1^A(\bar{r}_2^B(x^a)),\ y^a = \bar{r}_2^A(\bar{r}_1^B(y^a)),\ x^b = \bar{r}_1^B(\bar{r}_2^A(x^b)) \text{ and } y^b = \bar{r}_2^B(\bar{r}_1^A(y^b)).$$

We summarize these results as follows:

Theorem 5 *If the firms are symmetric, the trade policies and the outputs are naively adjusted and there is no policy lag, then the 4D dynamic system (18) gives rise to a period-2 cycle of the outputs which is synchronized with the period-2 cycle of the trade policies.*



(A) Return map (B) Time trajectory

Figure 4: Birth of a period-2 cycle.

We have assumed so far that there is no time lag in implementing the trade policy. However, the government policy usually works with a time lag since there is an inevitable delay between the subsidy decision and the actual payment which is the result of the political process. To examine the effect caused by the policy delay, a 10-period lag is introduced such that the trade policy is switched at every 10 periods. Simulation results are illustrated in Fig. 5 where we use the same parameter specifications with the only difference that the length of lag is changed to 10 from zero. In Figs. 5(A) and 5(B), the output dynamics shows the cyclic behavior in the following way: it fluctuates around x^{CA} for 10 periods and then around x^{CB} for the next 10 periods, after which it jumps back to the original cyclic behavior. We have already seen that x^{CA} and x^{CB} are locally stable if the trade policy is fixed. The cyclic behavior around each stationary point is a dumping oscillation. When the policy is changed in the middle of the converging process, the trajectory changes its direction and starts approaching a new equilibrium. As a result, a new dumping oscillation is born, which is again interrupted before arriving at the equilibrium by a change of the policy. A n-period time lag of the trade policy creates a period-$2n$ cycle. It fluctuates around one stationary point for n periods and then jumps to a neighborhood of the other stationary point when the policy is changed. Then it fluctuates around the new stationary state for the next n periods and jumps back to the previous neighborhood when the policy is changed again. This recursive process repeats itself. Dynamics with time lag can be summarized as follows:

Theorem 6 *If the firms are symmetric, the policies and the outputs are naively adjusted and there is a n-period policy lag, then the 4D dynamic system gives rise to a period-2 cycle of the trade policy and a period-*$2n$ *cycle of the output.*

If the policy is naively adjusted and the output is adaptively adjusted, then essentially the same dynamics will be observed. In other words, as far as the policy is periodically changed, the output dynamics is affected by these policy switching and exhibits periodic behavior no matter which formation of the expectations is selected. If the policy is adaptively adjusted, then the policy adjustment process converges to the stationary point according to Theorem 2. We have, however, qualitatively different output dynamics, as will be seen shortly. In this case we can suppose without loss of generality that the

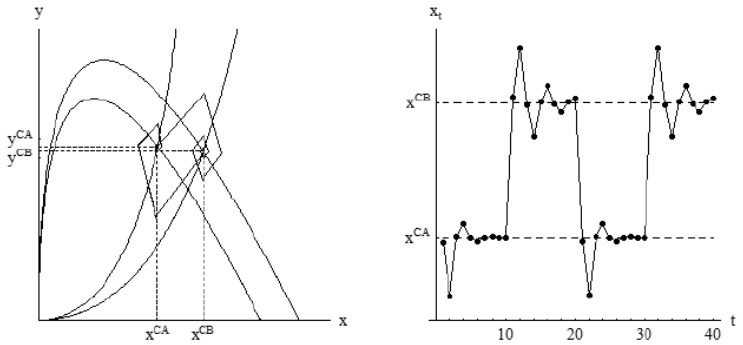


Figure 5: Birth of a period-20 cycle.

dynamic process of the trade policy is rapid and the firms receive the stationary values of the trade policy from the beginning of the output dynamic process. This assumption reduces the 4D dynamic system to the 2D output dynamic system (16) with $s_1 = s_1^e$, $s_2 = s_2^e$ and $s_1^e = -s_2^e$:

$$\begin{cases} x' = \sqrt{\dfrac{y}{c + s_2^e}} - y, \\ \\ y' = \sqrt{\dfrac{x}{c - s_2^e}} - x, \end{cases} \tag{19}$$

where $c = c_1 = c_2$.[5] The stability of system (19) depends on the actual cost ratio,

$$k = \frac{c + s_2^e}{c - s_2^e} \tag{20}$$

Given c, k increases from zero to infinity as s_2^e increases from $-c$ to c. In linear model, instability implies divergence. However this is not necessarily the case with nonlinear models because the nonlinearities may prevent unstable trajectories from globally diverging. We restrict our analysis to the unstable case, henceforth, to examine what dynamics the nonlinear output system can generate. In a linearized version of nonlinear output dynamic system (19), loss of stability occurs when the modulus of a pair of complex roots passes through 1,

$$\left| \frac{\partial x'}{\partial y} \frac{\partial y'}{\partial x} \right| = 1.$$

Following [7] or rearranging the above equation shows that stability of the output is violated if $k \leq 3 - 2\sqrt{2}$ and the output trajectories are non-negative as far as $k \geq 4/25$. Therefore, the output stationary state becomes locally unstable but the trajectories are nonnegative if

$$-\frac{21}{29}c < s_2^e < -\frac{\sqrt{2}}{2}c. \tag{21}$$

[5]It is possible to construct the dynamic system in terms of s_1^e. However, the results to be obtained are qualitatively the same.

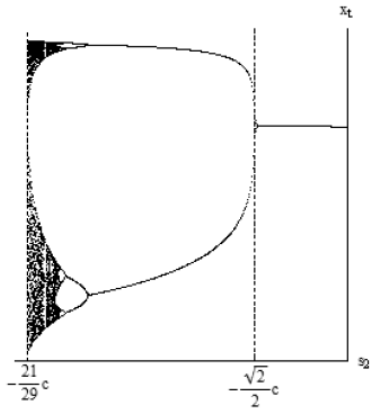


Figure 6: One-parameter bifurcation diagram with respect to s_2^e.

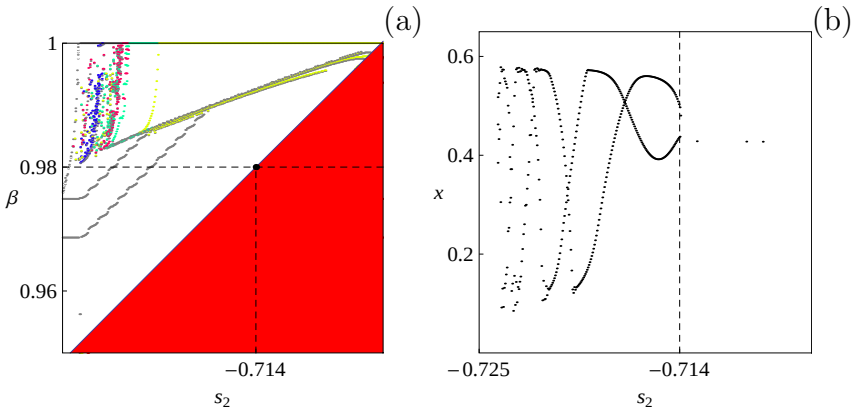


Figure 7: Bifurcation diagrams generated by system (22). (a) Two-parameter case. (b) One-parameter case.

Here $k = 3 - 2\sqrt{2}$ if s_2^e is equal to this upper bound which is the threshold of the loss of stability and will be called the *instability value* and $k = 4/25$ if s_2^e is equal to this lower bound, which is the threshold of loss of nonnegativity and will be called the *nonnegativity value*. When s_2^e decreases from the instability value to the nonnegativity value in the interval, the stationary state is destabilized and goes to chaos through a Neimark-Sacker cascade as shown in Fig. 6 where $c = 1$ is selected.

Now we replace the naive expectation formation with the adaptive expectation formation by taking $0 < \beta_1 < 1$ and $0 < \beta_2 < 1$. The output dynamic system is

$$\begin{cases} x' = (1-\beta_1)x + \beta_1\left(\sqrt{\dfrac{y}{c+s_2^e}} - y\right), \\ \\ y' = (1-\beta_2)y + \beta_2\left(\sqrt{\dfrac{x}{c-s_2^e}} - x\right). \end{cases} \tag{22}$$

This system is essentially the same as the one examined by Puu (2003). It is shown there

that loss of stability under the adaptive expectation occurs when the cost ratio and the speeds of adjustment satisfy the following relation:

$$\frac{(k-1)^2}{4k} = \frac{1}{\beta_1} + \frac{1}{\beta_2} - 1 \tag{23}$$

under which the determinant of the Jacobian matrix of the dynamic system (22) becomes unity. To examine what dynamics can be generated, we simulate system (22) when the adjustment speeds are identical ($\beta_1 = \beta_2 = \beta$). Taking the same parameter specifications as in Fig. 6, we present one- and two-parameter bifurcation diagrams in Fig. 7. The diagram in Fig. 7(A) has the optimal subsidy level to firm 2 (i.e., s_2^e) on the horizontal axis and the identical adjustment speed (i.e., β) on the vertical axis while the diagram in Fig. 7(B) plots the values of $x(t)$ against s_2, fixing the value of $\beta = 0.98$. The one-parameter diagram is obtained by increasing the value of s_2^e along the dotted line at $\beta = 0.98$ in the two-parameter diagram. Notice that the nonnegativity value of s_2 is $-21/29 \simeq -0.724$ and the instability value is $-\sqrt{2}/2 \simeq -0.707$. Solving (23) for β and substituting (20) provide the partition curve of the parameter space (s_2^e, β),

$$\beta = \frac{2(c - s_2^e)(c + s_2^e)}{c^2}. \tag{24}$$

Given c, for all (s_2^e, β) under the partition curve, the stability condition is satisfied and this stable region is colored in red. For all (s_2^e, β) above the curve, the stability condition is violated but the nonlinearities of the dynamics system may prevent diverging trajectories. The different colors of the regions correspond to the different periods of periodic cycles up to period 16. The gray regions indicate that the period of the cycle is larger than 16 or chaos emerges while trajectories become infeasible in the white region. Substituting 0.98 into (24) and solving the resultant equation for s_2 gives the threshold value $s_2 \approx -0.714$ for which the stationary output loses stability, as can be seen in Figs. 7(A) and (B). We summarize this result as follows:

Lemma 1 *If the firms are symmetric and the trade policy is adaptively and rapidly adjusted, then the adaptively adjusted output exhibits various dynamics ranging from a period-4 cycle to chaotic fluctuations, depending on the values of* (s_2^e, β).

4.2 Asymmetric firms: $c_1 \neq c_2$

We now turn our attention to the asymmetric firms with $c_1 < c_2$. We have already seen that the dynamic process of the trade policy is stable and converges to the optimal points given in (12),

$$s_1^e = -s_2^L + (c_2 - c_1) \text{ and } s_2^e = s_2^L,$$

where $0 > s_2^L > c_2 - 2c_1$ due to Assumption 2. The corresponding optimal outputs are obtained by substituting these optimal subsidies into the expressions in (15),

$$x^C = \frac{c_2 - s_2^L}{4c_1^2} \text{ and } y^C = \frac{(2c_1 - c_2) + s_2^L}{4c_1^2}.$$

If we assume that the implementation of the trade policy has no time lags, then the output dynamic system with $\beta_1 = \beta_2 = 1$ becomes

$$\begin{cases} x' = \sqrt{\dfrac{y}{c_1 - (-s_2^L + c_2 - c_1)}} - y, \\ \\ y' = \sqrt{\dfrac{x}{c_2 - s_2^L}} - x. \end{cases} \tag{25}$$

Substituting s_1^e and s_2^e into the actual cost ratio provides

$$k = \frac{2c_1 - c_2 + s_2^L}{c_2 - s_2^L}. \tag{26}$$

It follows that, given c_1 and c_2, the actual cost ratio decreases to zero from $(2c_1 - c_2)/c_2 > 0$ if s_2^L decreases to its lower bound $s_2^u = -(2c_1 - c_2)$ from zero. Depending on the value of s_2^L, the output dynamics can be destabilized. Furthermore, due to the nonlinearities of (16), the output dynamics can exhibit a rich dynamics if s_2^L is in the interval

$$c_2 - \frac{50}{29}c_1 < s_2^L < c_2 - \frac{1}{2 - \sqrt{2}}c_1. \tag{27}$$

The upper bound value and the lower bound value of s_2^L make the actual cost ratio equal to $3 - 2\sqrt{2}$ and $4/25$, respectively.

When the symmetric adaptive expectation formation is adopted (i.e., $\beta_1 = \beta_2 = \beta < 1$), the output dynamic system is

$$\begin{cases} x' = (1 - \beta)x + \beta \left(\sqrt{\dfrac{y}{c_1 - (-s_2^L + c_2 - c_1)}} - y \right), \\ \\ y' = (1 - \beta)y + \beta \left(\sqrt{\dfrac{x}{c_2 - s_2^L}} - x \right). \end{cases} \tag{28}$$

Solving (23) for β and substituting (26) yield the partition curve

$$\beta = \frac{2(c_2 - s_2^L)(2c_1 - c_2 + s_2^L)}{c_1^2} \tag{29}$$

which divides the (s_2^L, β)-space into two parts: stable region in the right to the curve and unstable region left.

It can be shown that dynamics generated by (19), respectively (22), are essentially the same as dynamics generated by (25), respectively (28). Introducing the new variables $C = c_1$ and $S = c_1 - c_2 + s_2^L$ reduces (25) and (28) to (19) and (22), respectively. One system can be transformed to the other through variable changes. Thus both systems are topologically conjugate to each other and generate qualitatively the same dynamics. In particular, the instability value and the nonnegativity value of (25) can be obtained from those values of (19) with C and S.

$$S = -\frac{21}{29}C \Longrightarrow s_2^L = c_2 - \frac{50}{29}c_1$$

and

$$S = -\frac{\sqrt{2}}{2}C \Longrightarrow s_2^L = c_2 - \frac{1}{2-\sqrt{2}}c_1.$$

Furthermore the partition line, (24) with C and S, can be transformed to (29).

$$\beta = \frac{2(C-S)(C+S)}{C^2} \Longrightarrow \beta = \frac{2(c_2 - s_2^L)(2c_1 - c_2 + s_2^L)}{c_1^2}.$$

The equivalence of the dynamic systems implies that (25) generates the same dynamics as illustrated in Fig. 6 with replacing the interval (21) with (27). Similarly, the output dynamics by (28) is the same as illustrated in Fig. 7 with replacing the partition curve (24) with (29). We can summarize these results as follows:

Lemma 2 *If the firms are asymmetric, then the naively adjusted dynamic system (25) starts the period-doubling bifurcation leading to chaos if s_2^L decreases from the instability value of the interval (27) to the nonnegative value whereas the adaptively adjusted dynamic system (28) generates complex dynamics involving chaos for (s_2^L, β) such as*

$$\beta > \frac{2(c_2 - s_2^L)(2c_1 - c_2 + s_2^L)}{c_1^2}.$$

Lemma 1 is concerned with the output dynamics of the symmetric firms while Lemma 2 is concerned with the output dynamics of the asymmetric firms. Notice that the results are essentially the same, hence, the production cost differences do not affect the asymptotic behavior of the unstable output dynamics if the trade policy is adaptively adjusted. These results can be summarized as follows:

Theorem 7 *If the optimal trade policy is asymptotically stable, then the output dynamic system generates the same dynamics regardless of the symmetry or asymmetry of the firms.*

The output dynamic system with adaptive expectation might lead to negative quantities or complex values in numerical simulations. To avoid such economically unfavorable phenomena, we preform the simulations with the modified output dynamic equations that take care of the nonnegativity of the output trajectories. The first modified system including the nonnegativity conditions is

$$\begin{cases} x' = (1-\beta)x + \beta Max\left[0, \sqrt{\dfrac{y}{c_1 + s_2^e}} - y\right], \\ \\ y' = (1-\beta)y + \beta Max\left[0, \sqrt{\dfrac{x}{c_2 - s_2^e}} - x\right]. \end{cases} \tag{30}$$

Taking $c_1 = c_2 = 1$, the two-parameter bifurcation diagram is shown in Fig. 8(A) in which the colored periodicity regions take "sausage-shaped". This is not the only way to prevent

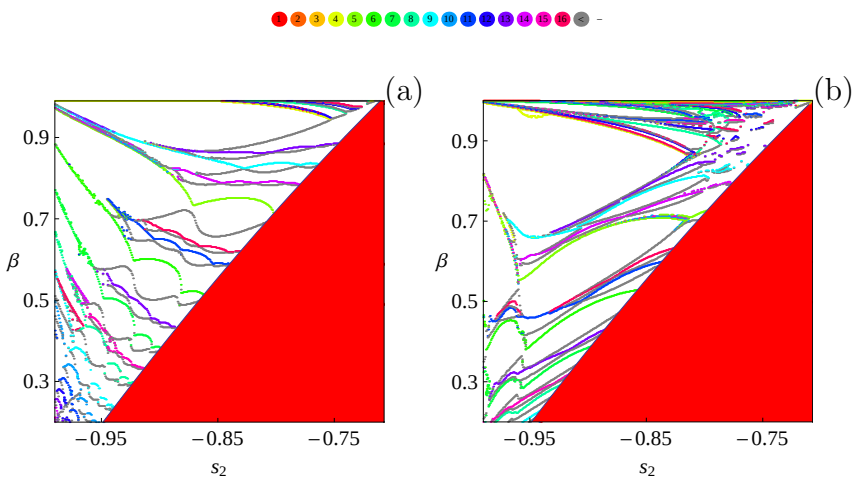


Figure 8: Two-parameter bifurcation diagrams with the nonnegativity conditions. (a) Dynamic system (30). (b) Dynamic system (31).

negative quantities. Following Yousefi (2002), we can use the alternative formulation

$$\begin{cases} x' = Max\left[0, (1-\beta)x + \beta\left(\sqrt{\dfrac{y}{c_1 + s_2^e}} - y\right)\right], \\ y' = Max\left[0, (1-\beta)y + \beta\left(\sqrt{\dfrac{x}{c_2 - s_2^e}} - x\right)\right]. \end{cases} \tag{31}$$

Fig. 8(B) shows the two-parameter bifurcation diagram generated by (31). Apparently there are many differences between these bifurcation diagrams, which imply that the asymptotic behavior of (30) is different from the asymptotic behavior of (31). However, our main finding that the adaptive systems can generate rich dynamics still holds.

5 Conclusion

In this paper, we construct a three-country model with two governments and two firms and consider the dynamic behavior of the sequential subsidy game in which the governments determine their optimal trade policies and then the firms determine their optimal outputs. We first deal with the governments' decision process from static and dynamic points of view. We find that the cost difference and the expectation formation of the governments are crucial in characterizing the optimal trade policy. In short, if the firms are symmetric, then there are infinitely many optimal policies (Theorem 1). A symmetric period-2 cycle of the trade policy emerges if naive adjustment process is adopted (Theorem 2) and a trajectory converges to one of the optimal policies if adaptive adjustment process is used (Theorem 3). If the firms are asymmetric, then a unique optimal policy exits and is asymptotically stable regardless of the expectation formations (Theorem 4). We then deal with the output dynamics and demonstrate that the expectation formation of the government matters but the cost difference does not matter. If the trade policy

is adaptively adjusted, then an output trajectory exhibits periodic cycle which is synchronized with the period-2 cycle of the optimal trade policy even if the Cournot output equilibrium is locally stable (Theorems 5 and 6). If the trade policy is adaptively adjusted, then complex output dynamics involving chaos emerges regardless of the expectation formation (Theorem 7). Finally it is worth mentioning that complex dynamics can be born under a small or even zero difference of the production costs in our model while much larger difference is required to generate chaotic dynamics in nonlinear duopoly models with isoelastic price function.

Acknowledgement: The authors are indebted to a referee for excellent suggestions and helpful comments. They highly appreciate financial supports from the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (C) 21530172), Chuo University (Joint Research Project 0981) and the Japan Economic Research Foundation. This paper was prepared when the first author visited the Department of Systems and Industrial Engineering of the University of Arizona. He appreciated its hospitality over his stay. The authors also want to acknowledge the encouragement and support by Kei Matsumoto for the research leading to this paper. The usual disclaimer applies.

References

[1] S, Bandyopadhyay, "Demand Elasticity, Asymmetry and Strategic Trade Policy," Journal of International Economics, 42, 167-177, 1997.

[2] G, Bischi, C, Chiarella, K, Kopel and F, Szidarovszky, *Nonlinear Oligopolies: Stability and Bifurcation*, Berlin/New York, Springer-Verlag, 2010.

[3] J, Brander and B, Spencer, "Export Subsidies and International Market Share Rivalry," Journal of International Economics, 18, 83-100, 1985.

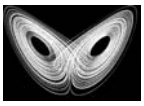
[4] J, Eaton and G, Grossman, "Optimal Trade and Industry Policy under Oligopoly," Quarterly Journal of Economics, 101, 383-403, 1986.

[5] A, Matsumoto and F, Szidarovszky, "A Little Help from My Friend: International Subsidy Games with Hyperbolic Demand," DP# 119, Institute of Economic Research, Chuo University (http://www2.chuo-u.ac.jp/keizaiken/discuss.htm), 2009.

[6] A, Matsumoto and N, Serizawa, "Strategic Trade Policy under Isoelastic Demand and Asymmetric Production Costs," Annals of Regional Science, 41, 525-543, 2007.

[7] Puu, T., *Attractors, Bifurcations and Chaos*, Berlin/Heidelberg/New York, Springer-Verlag, 2003.

[8] S, Yousefi, "Intrinsic Stability, Multi Stability and Instability," in *Oligopoly Dynamics*, T, Puu and I, Sushko (Eds.) 219-231, Berlin/Heidelberg/New York, Springer-Verlag, 2002.



Annual Review of Chaos Theory, Bifurcations and Dynamical Systems
Vol. 1, (2012) 22-27, www.arctbds.com.

On Piecewise Monotone Interval Maps and Periodic Points

Jose S. Cánovas

Universidad Politécnica de Cartagena.
Departamento de Matemática Aplicada y Estadística.
C/ Doctor Fleming sn. 30202 Cartagena (Spain).
e-mail: Jose.Canovas@upct.es

Abstract

In this short note, we find that a continuous piecewise monotone interval map f is chaotic in the sense of Li and Yorke if and only if f restricted to the set of its periodic points is not Lyapunov stable.

Keywords: Low dimensional dynamics, Li–Yorke chaos, periodic points.
2010 Mathematics Subject Classification codes: 26A18, 37E05.

Manuscript accepted March 01, 2012.

1 Introduction

Let $f : I = [0, 1] \to I$ be a continuous map. In the huge list of conditions which are equivalent to zero entropy of a continuous interval map (see e.g. [14]), there were three, (C2)–(C4), included for a long time:

(C1) The map f has zero topological entropy (see [1] for the definition).

(C2) The map $f|_{\mathrm{P}(f)}$ is Lyapunov stable (it has equicontinuous powers).

(C3) The set $\mathrm{R}(f)$ is a F_σ set.

(C4) The set $\mathrm{P}(f)$ is a G_δ set.

Recall that the orbit of a point $x \in I$ is given by the sequence $(f^n(x))$, where $f^1 = f$ and $f^n = f \circ f^{n-1}$ for $n > 1$. A point $x \in I$ is periodic provided there is $n \in \mathbb{N}$ such that $f^n(x) = x$. The smallest positive integer n holding this condition is called the period of x. The limit points of the orbit of x is called the ω–limit set of x under f, denoted by $\omega(x, f)$. A point x is called recurrent if $x \in \omega(x, f)$. Denote by $\mathrm{P}(f)$ and $\mathrm{R}(f)$ the sets of periodic and recurrent points of f, respectively. Recall that a subset A is a G_δ set provided it is equal to the intersection of a countable collection of open subsets. The set F is an F_σ if it is the countable union of closed sets.

At the beginning of XXI century, the equivalence among these properties were proved to be false. In [15] was proved that condition (C1) was not equivalent to (C2), although condition (C2) always implies (C1). A similar result was proved in [17] concerning conditions (C1) and (C3). Finally, in [16], the equivalence between (C1) and (C4) is disproved by proving that (C1) does not imply (C4), and in [11] has been recently proved that (C4) does not imply (C3).

If we think about these properties for a while, we see that conditions (C3) and (C4) are related to the topological structure of two sets from the topological dynamics of f. Property (C2) is a dynamical property itself, because states that the dynamics of $f|_{\mathrm{P}(f)}$ is quite simple. Let us point out that, recently, in [5] and [6] the dynamics of f has been studied from the set of periodic points of the map f.

The maps of the above mentioned counterexamples for (C3) and (C4) were obtained as functional limit of continuous maps and hence, they are not piecewise monotone. Recall that $f : I \to I$ is piecewise monotone if there is a partition $0 = x_0 < x_1 < ... < x_n = 1$ of I such that $f|_{(x_i, x_{i+1})}$ is monotone for $i = 0, ..., n-1$. This fact was not strange for conditions (C3) and (C4), because these equivalences are true for such kind of maps (see [17]). The counterexample on property (C2) was constructed by a so–called weakly unimodal map, which has two pieces of monotonicity. The aim of this paper is to go further and proving the following result.

Theorem 1 *Let $f : I \to I$ be a piecewise monotone continuous map. Then the map $f|_{\mathrm{P}(f)}$ is Lyapunov stable if and only if f is not chaotic in the sense of Li and Yorke.*

Recall that a continuous interval map f is chaotic in the sense of Li and Yorke (LY–chaotic) if there is an uncountable set $S \in I$ such that

$$0 = \liminf_{n\to\infty} |f^n(x) - f^n(y)| < \limsup_{n\to\infty} |f^n(x) - f^n(y)|,$$

for any $x, y \in S$, $x \neq y$. In addition, we say that f is LY–simple if for any $x \in I$ and any $\varepsilon > 0$, there is a periodic point y such that $\limsup_{n\to\infty} |f^n(x) - f^n(y)| < \varepsilon$.

Hence, we can add our main result to the following one. First, we recall that $\omega(f) = \cup_{x\in I}\omega(x, f)$. $\Omega(f)$ denotes the set of nonwandering points, that is, those points $x \in I$ such that for any $\varepsilon > 0$ there is $n > 0$ such that $(x - \varepsilon, x + \varepsilon) \cap f^n(x - \varepsilon, x + \varepsilon) \neq \emptyset$. $\mathrm{AP}(f)$ is the set of almost periodic points for which for any $\varepsilon > 0$ there is $k > 0$ such that $f^{kn}(x) \in (x - \varepsilon, x + \varepsilon)$ for any $n \geq 0$. For the definition of topological sequence entropy, which is an extension of topological entropy, see [10] or [9].

Corollary 1 *Let $f : I \to I$ be a piecewise monotone continuous map. The following statements are equivalent:*

(a) The map f is not LY–chaotic.

(b) The map f is LY–simple.

(c) $\omega(f) = \{x \in [0,1] : \lim_{n\to\infty} f^{2^n}(x) = x\}$.

(d) $\mathrm{AP}(f) = \omega(f)$.

(e) The map $f|_{\omega(f)}$ is Lyapunov stable.

(f) The map $f|_{\Omega(f)}$ is Lyapunov stable.

(g) The topological sequence entropy of f, $h_A(f)$, is zero for any increasing sequence of integers A.

(h) The map $f|_{\mathrm{P}(f)}$ is Lyapunov stable.

We remark that the properties (b)–(g) in Corollary 1 are equivalent to (a) without the assumption that f is piecewise monotone as one can see in the references [7], [9] and [8]. The equivalence between properties (a) and (h) comes from Theorem 1. On the other hand, there is an example in [7] proving that condition (h) cannot imply (a) without the assumption that f is piecewise monotone.

We will prove our main result in next section.

2 Proof of Theorem 1

Before proving our main result, we will show the following one, whose proof is immediate.

Proposition 2 *Let $f : [0,1] \to [0,1]$ be non LY–chaotic. Then $f|_{\mathrm{P}(f)}$ is Lyapunov stable.*

Proof. Since f is not chaotic, by [7], we have that $f|_{\omega(f)}$ is Lyapunov stable. The result follows because $\mathrm{P}(f) \subset \omega(f)$.□

Proof of Theorem 1. In view of Proposition 2, we just need to prove that if f is LY–chaotic, then $f|_{\mathrm{P}(f)}$ cannot be Lyapunov stable. Recall that a LY–chaotic map with zero topological entropy has an infinite ω–limit set $\omega(x,f)$ with the following properties (see [18]):

- There is a nested sequence of intervals $J_0 \supset J_1 \supset ... \supset J_n \supset ...$ such that $f^{2^n}(J_n) = J_n$ and
$$\omega(x,f) \subset \bigcap_{n\geq 0} \bigcup_{j=0}^{2^n-1} f^j(J_n).$$
- $\omega(x,f)$ contains two f–nonseparable points u, v, that is, for any $n \geq 1$, u and v are contained in the same periodic interval $f^j(J_n)$.

Now, we consider the set $\Sigma = \{0,1\}^{\mathbb{N}}$, and for any $\alpha \in \Sigma$ and $n \in \mathbb{N}$, let $\alpha|_n = (\alpha_1, ..., \alpha_n)$. We write $J_{\mathbf{0}|_n} = J_n$, where $\mathbf{0} = (0,0,...)$. Then, denote by $J_{\alpha|_n} = f^j(J_n)$ in such a way that $\alpha|_n = a_n^j(\mathbf{0}|_n)$, where $a_n(1,1,...,1) = \mathbf{0}|_n$ and $a_n(\alpha|_n) = \alpha|_n * 1$ for $\alpha|_n \neq (1,...,1)$, where $*$ denotes the operation which adds 1 to α_1; if $\alpha_1 + 1 = 1$, then $a_n(\alpha|_n) = (1, \alpha_2, ..., \alpha_n)$, if $\alpha_1 + 1 = 2$, then we put 0 in the first component and add 1 to α_2 and repeat this porcces untill α_j will be 1. For instance $a_3(1,1,0) = (0,0,1)$ and $a_3(0,1,1) = (1,1,1)$. Clearly, for $\alpha \in \Sigma$ and $n < m$, we have that $J_{\alpha|_m} \subset J_{\alpha|_n}$. Denote by $J_\alpha = \cap_{n\geq 1} J_{\alpha|_n}$

For a subinterval J, $|J|$ will be its length. Now, let $\delta > 0$. Let $\mathcal{A}_\delta = \{\alpha \in \not\prec : |J_\alpha| \geq \delta\}$. Now, we claim that there exists an $n_\delta \in \mathbb{N}$ such that for any $n \geq n_\delta$ it is held

- if $\alpha \in \mathcal{A}_\delta$ then $\max\{|J^+_{\alpha|_n}|, |J^-_{\alpha|_n}|\} < \delta$, where $J^+_{\alpha|_n}$ and $J^-_{\alpha|_n}$ are the right and left side subintervals of $J_{\alpha|_n} \setminus J_\alpha$.
- if $\theta \in \{0,1\}^n$ and $\alpha|_n \neq \theta$ for any $\alpha \in \mathcal{A}_\delta$ then $|J_\theta| < \delta$.

To prove our claim, let $\alpha \in \mathcal{A}_\delta$. Since $(J_{\alpha|_n})_{n=1}^\infty$ decreases to J_α, if n is large enough then $\max\{|J^+_{\alpha|_n}|, |J^-_{\alpha|_n}|\} < \delta$. Since $\mathcal{A}_\delta$ is finite we have $\max\{|J^+_{\alpha|_n}|, |J^-_{\alpha|_n}|\} < \delta$ for all $\alpha \in \mathcal{A}_\delta$ and all sufficient large n. Now, we show that if n is large enough then $|J_\theta| < \delta$ for any $\theta \in \{0,1\}^n$ with the property $\alpha|_n \neq \theta$ for all $\alpha \in \mathcal{A}_\delta$. Suppose the contrary. Then there are a strictly increasing sequence $(n_j)_{j=1}^\infty$ and sequences $\theta^j \in \{0,1\}^{n_j}$ such that $|J_{\theta^j}| \geq \delta$ and $\alpha|_j \neq \theta^{n_j}$ for any $\alpha \in \mathcal{A}_\delta$. Let x_j be the midpoint of J_{θ^j}. It is clearly not restrictive to assume that $(x_j)_{j=1}^\infty$ converges to some x and $|x_j - x| < \delta/2$ for any j. Since for any fixed n all intervals J_θ, $\theta \in \{0,1\}^n$, are pairwise disjoint, this means that each pair K_{θ^j} and $K_{\theta^{j+1}}$ has non-empty intersection, which clearly implies $J_{\theta^{j+1}} \subset J_{\theta^j}$ for any j and hence the existence of an $\alpha \in \Sigma$ with $\alpha|_{n_j} = \theta^j$ for any j. Due to the definition of the intervals J_{θ^j}, α cannot belong to $\mathcal{A}_\delta$. However, $J_\alpha = \bigcap_{n=1}^\infty J_{\alpha|_n} = \bigcap_{j=1}^\infty J_{\theta^j}$ so $|J_\alpha| \geq \delta$, a contradiction.

Now, fix $\varepsilon = |u - v|$. Since $\overline{\mathrm{P}}(f) = \omega(f)$ (cf. [4]), there are sequences of periodic points u_n and v_n which converge to u and v, respectively. Now, fix $\delta > 0$, $\delta < \varepsilon$, and $\mathcal{A}_\delta$ as before. There is $n_0 \in \mathbb{N}$ such that u_n and v_n are contained in $J^+_{\alpha|_n} \cup J^-_{\alpha|_n}$, where $\alpha \in \Sigma$ is such that $u, v \in J_\alpha$. Since u_n and v_n are periodic points, there is $\theta \in \{0,1\}^n$, $n \geq \max\{n_\delta, n_0\}$ such that $f^j(u_n)$ and $f^j(v_n)$ belong to J_θ for some $0 < j < 2^n$ and such that $|J_\theta| < \delta$. Hence $|f^j(u_n) - f^j(v_n)| < \delta$ and $|f^{2^n - j}(f^j(u_n)) - f^{2^n-j}(f^j(v_n))| > \varepsilon$, which proves that $f|_{\mathrm{P}(f)}$ cannot be Lyapunov stable.□

Remark 3 *Recall that a wandering interval of f is an interval whose iterates are pairwise disjoint and such that neither of the orbits of its points is attracted by any periodic orbit. In fact if a map $f \in C(I)$ of zero entropy is chaotic then it must possess a wandering interval (see e.g. [2]). In many "natural" maps (including all analytic ones) wandering intervals cannot exist [12] and then they cannot be LY–chaotic. So, we wonder about the validity of Theorem 1 under regularity conditions of f, for instance for C^1 maps.*

Acknowledgement: This paper has been 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] R. L, Adler, A. G, Konheim and M. H, McAndrew, "Topological entropy," Trans. Amer. Math. Soc, 114, 309–319, 1965.

[2] F, Balibrea and V, Jiménez López, "The measure of scrambled sets: a survey," Acta Univ. Mathaei Belii Nat. Sci. Ser. Ser. Math, 7, 3–11, 1999.

[3] L. S, Block and A. Coppel, *Lectures on one dimensional dynamics*, Lecture Notes in Mathematics, 1513, Springer–Verlag, Berlin, 1992.

[4] J. S, Cánovas and A. Linero, "Non-chaotic antitriangular maps," Appl. Gen. Topol, 6, 171–183, 2005.

[5] E. M, Coven and E, D'Aniello, "Chaos for maps of the interval via ω-limit points and periodic points," Atti Sem. Mat.Fis. Univ. Modena, 49, 523–530, 2001.

[6] E. D'Aniello and T. H, Steele, "Approximating ω-limit sets with periodic orbits," Aequationes Math, 75, 93–102, 2008.

[7] V. V, Fedorenko, A. N, Sharkovsky and J, Smítal, "Characterizations of weakly chaotic maps of the interval," Proc. Amer. Math. Soc, 110, 141–148, 1990.

[8] V. V, Fedorenko and J. Smítal, "Maps of the interval Ljapunov stable on the set of nonwandering points," Acta Math. Univ. Comenian, 60, 11–14, 1991.

[9] N, Franzová and J. Smítal, "Positive sequence topological entropy characterizes chaotic maps," Proc. Amer. Math. Soc, 112, 1083–1086,1991.

[10] T. N. T, Goodman, "Topological sequence entropy," Proc. London Math. Soc, 29, 331–350, 1974.

[11] Z, Kocan, "On some properties of interval maps with zero topological entropy," Aequationes Math, 76, 305–314, 2008.

[12] M, Martens, W, de Melo and S, van Strien, "Julia-Fatou-Sullivan theory for real one-dimensional dynamics," Acta Math, 168, 271–318, 1992.

[13] M, Misiurewicz and J, Smítal, "Smooth chaotic functions with zero topological entropy," Ergod. Th. and Dynam. Sys, 8, 421–424, 1988.

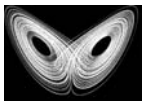
[14] A. N, Sharkovsky, S. F, Kolyada, A. G, Sivak and V. V, Fedorenko, *Dynamics of one–dimensional maps*, Kluwer Academic Publishers, 1997.

[15] P, Sindelárová, "A counterexample to a statement concerning Lyapunov stability," Acta Math. Univ. Comenian, (N.S.), 70, 265–268, 2001.

[16] P, Sindelárová, "A zero topological entropy map for which periodic points are not a G_δ set," Ergodic Theory Dynam. Systems, 22, 947–949, 2002.

[17] P, Sindelárová, "A zero topological entropy map with recurrent points not F_σ," Proc. Amer. Math. Soc, 131, 2089–2096, 2003.

[18] J, Smítal, "Chaotic functions with zero topological entropy," Trans. Amer. Math. Soc, 297, 269–282, 1986.



Annual Review of Chaos Theory, Bifurcations and Dynamical Systems
Vol. 1, (2012) 28-49, www.arctbds.com.

Velocity/Pressure Response Curve Characterization of the Arterial Path of the P3 Iliac Reflection During Hemorrhage Simulation

Martin C Baruch
Empirical Technologies Corporation, Charlottesville, Virginia, 22906, USA.
e-mail: mcbaruch@comcast.net

Darren ER Warburton
Cardiovascular Physiology Laboratory, University of British Columbia
Vancouver, British Columbia, V6T 1Z3, Canada.
e-mail: darrenwb@interchange.ubc.ca

Shannon SD Bredin
Cardiovascular Physiology Laboratory, University of British Columbia
Vancouver, British Columbia, V6T 1Z3, Canada.
e-mail: shannon.bredin@.ubc.ca

Anita Cote
Cardiovascular Physiology Laboratory, University of British Columbia
Vancouver, British Columbia, V6T 1Z3, Canada.
e-mail: acote@cscpacific.ca

David W. Gerdt
Empirical Technologies Corporation, Charlottesville, Virginia, 22906, USA.
e-mail: davidetc@firstva.com

Charles M. Adkins
Empirical Technologies Corporation, Charlottesville, Virginia, 22906, USA.
e-mail: cma.etc@comcast.net

Abstract

Background: Recent work suggests that pulse pressure may be a marker of temporal changes in central blood volume, facilitating the non-invasive detection hemorrhage and the onset of hemorrhagic shock. A new approach to tracking blood pressure, and pulse pressure specifically is presented that is based on a new form

of pulse pressure wave analysis called Pulse Decomposition Analysis (PDA). The premise of the PDA model is that the peripheral arterial pressure pulse is a superposition of five individual component pressure pulses, the first of which is due to the left ventricular ejection from the heart while the remaining component pressure pulses are reflections and re-reflections that originate from only two reflection sites within the central arteries. The hypothesis examined here is that the PDA parameter T13, the timing delay between the first and third component pulses, correlates with pulse pressure. Furthermore the comparison of experimental and model predictions provides insight into the pressure/velocity responses of the different arterial segments that the iliac reflection pulse (P3) traverses.

T13 was monitored along with blood pressure, as determined by an automatic cuff and another continuous blood pressure monitor, during the course of lower body negative pressure (LBNP) sessions involving four stages, -15 mmHg, -30 mmHg, -45 mmHg, and -60 mmHg, in fifteen subjects (average age: 24.4 years, SD: 3.0 years; average height: 168.6 cm, SD: 8.0 cm; average weight: 64.0 kg, SD: 9.1 kg).

Results: Statistically significant correlations between T13 and pulse pressure as well as the ability of T13 to resolve the effects of different LBNP stages were established. Experimental T13 values were compared with predictions of the PDA model. These predictions revealed distinctly different response characteristics for the three major arterial segments that constitute the path of the P3 component pulse. The LBNP interventions resulted in pulse pressure changes of up to 7.8 mmHg (SE = 3.49 mmHg) as determined by the automatic cuff. Corresponding changes in T13 were a shortening by -72 milliseconds (SE = 4.17 milliseconds). In contrast to the other two methodologies, T13 was able to resolve the effects of the two least negative pressure stages with significance set at $p < 0.01$.

Conclusions: The agreement of observations and measurements provides a preliminary validation of the PDA model regarding the origin of the arterial pressure pulse reflections. The proposed physical picture of the PDA model is attractive because it identifies the contributions of distinct reflecting arterial tree components to the peripheral pressure pulse envelope. The results presented here suggest that the model may also be able to provide insight into the dynamic response characteristics of certain arterial tree sections, with important implication for the assessment of their health.

Keywords: Hemorrhage, Arterial Pulse Analysis, Pulse Decomposition Analysis, Pulse Pressure.

Manuscript accepted December 3, 2011.

1 Introduction

The objective of this work is the introduction of a new form of pulse pressure wave analysis, called Pulse Decomposition Analysis (PDA). As evidenced by the number of different

commercial and academic efforts, pulse analysis already provides access to the tracking of some cardiovascular parameters, such as cardiac output, left ventricular ejection etc. that are otherwise accessible only via either invasive means or sophisticated but also cumbersome monitoring modalities. As the work presented here suggests, a further refined pulse analysis model holds the promise of providing 1. access to additional vascular parameters and, perhaps equally important, their dynamic response characteristics and 2. the tracking of blood pressure without the physically restrictively engaging, and consequently uncomfortable, approaches that current continuous blood pressure monitors employ. Of particular interest in the context of tracking blood pressure is pulse pressure, the motivation being that pulse pressure appears to be a sensitive as well as specific marker for the detection of hemorrhage, [1-2] which remains one of the leading causes of death on the battlefield as well as in civilian trauma cases while also being highly preventable if intervention can be implemented [3-4]. However, detecting progressive hemorrhage requires resolution of changes on the order of a few mmHg in pulse pressures of, normally, 35-50 mmHg. Given the separate and unequal uncertainties in determining systole and diastole, using the best brachial cuff techniques, [5] such determinations are by and large out of reach even in controlled environments.

In response we present here the PDA approach, which not only offers the capability to continuously track blood pressure but to also provide insight into the response characteristics of certain arterial tree segments. The analysis of the arterial pressure pulse has been the subject of many studies, with works whose results are still relevant today dating from the 1800s and the early 1900s, [6-8] as well as a significant body of work that has been published over the past 40 or so years [9-16]. PDA presents the extension of the findings of a number of studies that have utilized ballistocardiography and invasive central artery manometers to track mechanical events such as heart contractions and pressure pulse reflections in the central arterial tree, to the arterial periphery. These studies [17-19] have confirmed the existence of two major reflection sites in the central arteries. The first reflection site is the juncture between thoracic and abdominal aorta, which is marked by a significant decrease in diameter and a change in elasticity and the second site arises from the juncture between abdominal aorta and common iliac arteries. In what follows these reflection sites are respectively referred to as the *renal* and the *iliac* reflection site. A consequence of these reflection sites are reflected arterial pressure pulses that counter-propagate to the direction of the single arterial pressure pulse, due to left ventricular contraction, that gave rise to them. Referring to Fig. 1, the *downward* traveling primary pressure pulse # 1 gives rise to the *upward* traveling # 2 and # 3 pulses that are respectively due to the renal and the iliac reflection sites on which the # 1 pulse impinged.

As these reflected # 2 and # 3 pressure pulses reach the aortic arch, they will enter the subclavian arteries and head into the arterial periphery of the arm, following the # 1 pressure pulse that, besides traveling down the aorta, also entered the arm complex arteries. The # 2 and # 3 pulses will do so with certain time delays because of the *extra* traversal of the central arteries.

The # 2 pulse is commonly known as the *second systolic* peak. We refer to it as the *renal reflection* and it follows the primary ejection pulse (# 1) into the arterial periphery of the arm at delays of between 70-140 milliseconds. The pulse labeled # 3 in Fig. 1

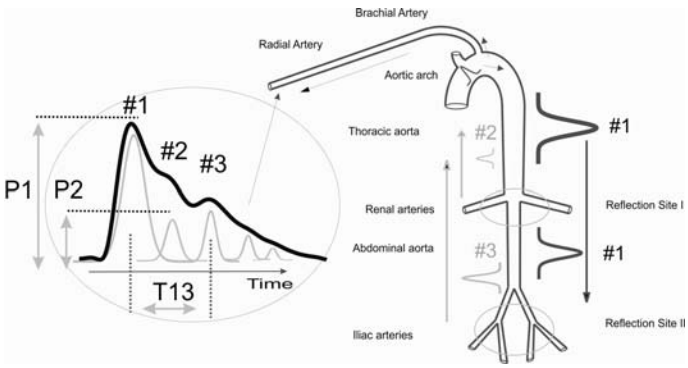


Figure 1: Sketch of the aorta/arm complex arterial system and its effect on the arterial pressure pulse line shape that is observed at the radial/digital artery. Two reflection sites, one at the height of the renal arteries, the other one in the vicinity of the iliac bifurcation, give rise to the reflected pulses (gray) that trail the primary left ventricular ejection (black). (from [31]).

is the much larger *iliac reflection*, which follows the # 1 pulse at delays of 180 to 400 milliseconds.

The described scenario succinctly explains the presence of three component pulses in the pressure pulse envelope that is observed in the arterial periphery of the arm, such as at the radial or digital arteries. In fact, there are additional component pulses. The presence of re-reflections between the central reflection sites has been previously suggested [20]. The physical picture is one where the iliac reflection pulse, in its travel up the aorta, re-reflects off the renal reflection site, and this re-re-reflection traveling downward, once again reflects off the iliac reflection site to follow the first three components pulses. With dramatically diminishing amplitude the scenario repeats for the fifth component pulse. These higher-order reflections are less relevant for quantitative analysis due to their poorer signal to noise characteristics and the fact that they are easily swallowed by the pulse envelope of the next cardiac cycle unless the heart rate is very low.

Based on these considerations the structure of the radial/digital arterial pressure pulse can be explained entirely by the interaction of the primary left ventricular ejection pressure pulse with two aortic reflection sites. We now hypothesize that it is possible to determine trends in aortic blood pressure through an analysis of the pulse envelope obtained in the arterial periphery of the arm.

The PDA model presented here analyzes the arterial pulse as observed on the lower arm by isolating, identifying and quantifying the temporal positions and amplitudes of the renal reflection pulse (# 2) and the iliac reflection pulse (# 3), each relative to the primary systolic pulse (# 1), within the pulse shape envelope of an individual cardiac cycle. The models predictions and experimental studies show that two pulse parameters are of particular importance. One parameter is the ratio of the amplitude of the renal reflection pulse (# 2) to that of the primary systolic pulse (# 1). These amplitudes, labeled P1 and P2, are indicated to the left of the arterial pressure pulse envelope. This parameter is herein referred to as the P2P1 ratio and it tracks changes in central beat-by-beat systolic pressure. The second parameter is the time difference between the arrival of

the primary systolic (# 1) pulse and the iliac reflection (# 3) pulse. This parameter is referred to as T13, as indicated in Fig. 1, and it tracks changes in arterial pulse pressure, also beat-by-beat.

It is the aim of this paper to validate the described arterial pressure pulse reflection scenario through the presentation of experimental data collected in the context of simulating central hemorrhage and its comparison with predictions of the PDA pulse propagation model. Specifically, our hypothesis is that the time delay between the primary component pulse (# 1) and the iliac reflection pulse (# 3), T13, correlates with pulse pressure.

We report here the results of monitoring the evolution of the PDA parameter T13 during the course of lower body negative pressure (LBNP) sessions. LBNP is an established technique used to physiologically stress the cardiovascular system. It has been used to simulate gravitational stress and hemorrhage, alter preload, and to manipulate baroreceptors [21]. LBNP was chosen for this project because it has been shown to be very effective at modulating pulse pressure, thereby providing a means to validate the equivalent PDA arterial pulse parameter, T13.

2 Patients and Methods

After IRB approval, tests of the CareTaker system, which is the hardware implementation of the PDA model that is described in more detail below, were performed at the Cardiovascular Physiology Laboratory of the University of British Columbia on fifteen healthy volunteers (average age: 24.4 years, SD: 3.0 years; average height: 168.6 cm, SD: 8.0 cm; average weight: 64.0 kg, SD: 9.1 kg) whose lower bodies, from the height of the iliac crest downwards, were subjected to increasingly negative pressures. A number of studies have demonstrated that it is possible to simulate significant internal hemorrhage using LBNP. Negative pressures of $10 - 20$ $mmHg$ correspond to 400 to 550 ml of central blood loss, $20 - 40$ $mmHg$ correspond to 500 to 1000 ml, and negative pressures in excess of -40 $mmHg$ correspond to blood losses exceeding 1000 ml [22].

The subjects were subjected to four stages of negative pressure, -15 $mmHg$, -30 $mmHg$, -45 $mmHg$, and $-60mm$ Hg, each stage lasting typically about 12 minutes. The blood pressure was monitored with an automatic cuff (BP TRU Automated Non-Invasive Blood Pressure Monitor (model BPM-100), VSM MedTech Devices Inc.) set to record blood pressures every three minutes, resulting in typically four readings per LBNP setting as well as an Ohmeda 2300 Finapres, and a pulse oximeter (Ohmeda Biox 3740 Pulse Oximeter, BOC Health Care) monitored oxygen saturation. The CareTaker system collected arterial pulse shapes beat-by-beat via a finger cuff attached to the central phalange of the middle digit. Four subjects became presyncopal and could not complete the -60 mmHg LBNP stage.

2.1 CareTaker Device

The hardware platform that provides the arterial pulse signal for the PDA algorithms analysis is the Care-Taker device (Empirical Technologies Corporation, Charlottesville, Virginia). It is a physiological sensing system whose three basic physical components are

a sensing pad such as a finger cuff that couples to an arterial pressure point, a pressure line that pneumatically telemeters the pulsations, and a custom-designed piezo-electric pressure sensor that converts the pressure pulsations, using transimpedance amplification, into a voltage signal that can be measured, digitized, transmitted and recorded. The coupling to the artery is accomplished using palpation coupling, such as at the radial artery, or approximate hydrostatic coupling, such as at the digital arteries. The completely self-contained device wirelessly transmits its signal representing the arterial pulse to a PC computer using the Bluetooth protocol. The device is not occlusive as it operates at a coupling pressure of about 40 $mmHg$. Another important characteristic of the device is that the signal it provides, sampled at 512 Hz, is the time derivative of the arterial pulse signal. The derivative provides significant signal to noise advantage and lowers the resolution requirements for digital acquisition of the signal because the derivative eliminates signal offsets. Because of the short time constants associated with its implementation, it also offers very short recovery times after signal disruptions. That is, in the absence of offsets due to the differentiation, the signal is always clamped to the signal base line, which in turn allows for increased amplification. Consequently the full digitization range of the analog to digital converter (A/D) can be used for the signal amplitude, as opposed to signal amplitude plus offsets.

2.2 The PDA model

The existence of two distinct central pressure pulse reflection sites make it is possible to propose a simple model of the arterial paths that the primary pulse and its reflections traverse and to compare its predictions with observations regarding the relative arrival times of the different components pulses. The models equations, which have been presented elsewhere [31], predict the time of arrival of each individual component pulse, subject to the total distance that the pulse has traveled and the pressure-dependent pulse propagation velocity in each arterial segment. The pressure dependence of the pulse propagation velocity is implemented using the Moens-Korteweg [23] equation, which relates pressure and velocity.

Another critical feature of the model is that R2, the renal reflection coefficient, is dependent on systolic pressure. The motivation for this is based on the following consideration. As discussed, the renal reflection (P2 pulse) originates at the junction between thoracic and abdominal aorta, a junction that is characterized by a significant change in arterial diameter. Since the thoracic aorta is the softest artery in the body, as evidence by the fact that it exhibits the lowest pulse pressure propagation velocities ($4 - 5\ m/s$) and much more extensible than the abdominal aorta, increasing peak pressure, or systole, will enlarge the diameter mismatch, giving rise to a more pronounced renal reflection pulse amplitude while falling systole will produce the opposite effect, an effect observed in manipulative experiments performed by Latham [17]. The critical insight then is that the amplitude of the renal reflection will increase relative to the amplitude of the primary systolic (P1 pulse) peak because, while both component pulses travel the arteries of the arm complex, and are therefore both subject to the pulse narrowing and heightening due to the taper and wall composition changes of the peripheral arteries, only the renal reflection will have sampled the pressure-induced aortic impedance mismatch changes. This

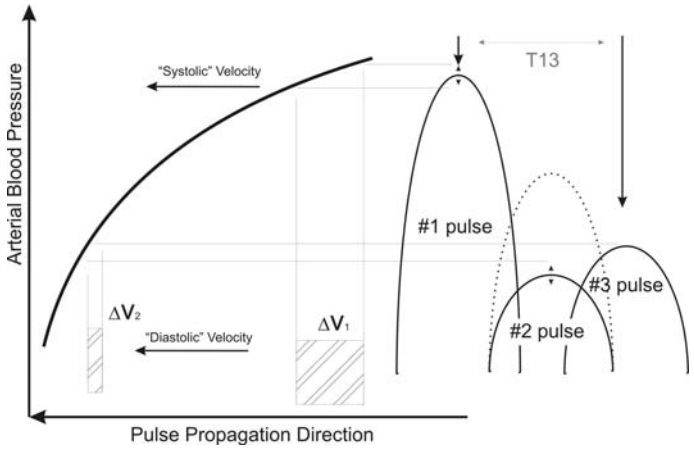


Figure 2: Relative amplitudes and velocity/pressure relationships of the component pulses. # 2's amplitude and propagation velocity can change significantly due to the pressure dependence of R2, the renal reflection coefficient. (from [31]).

establishes the motivation for taking the ratio of the amplitudes of the # 2 and the # 1 pulse, which is P2P1.

A similarly physical argument can be made for the difference in arrival times of the primary pulse (# 1) and the iliac reflection (# 3), or T13. The difference in the arrival times of the primary arterial pulse, that is the left ventricular ejection, and the iliac reflection pulse is determined by the differential velocities with which both pulses propagated along their arterial paths. In the case of the iliac reflection the path length is longer than that of the primary pulse by almost twice the length of the torso. More importantly, both pulses travel at different velocities because their pressure amplitudes are different. Specifically, the iliac reflection pulse amplitude, which is determined by the reflection coefficient of the iliac reflection site, is on the order of 40 percent of pulse pressure. This point is graphically made in Fig. 2, Both pulses therefore load the arterial wall differently during their arterial travel, as a result of which their propagation velocities are different. The second insight is that, because the pressure/velocity response curve is non-linear, a result known since the 1960s based on Anlikers work, [24] both pulses accelerate and decelerate at different rates as the pressure rises and falls. The primary pulse experiences the highest changes in velocity as a function of changes in blood pressure because it is subject to the steepest section of the pressure/ velocity response curve, while the iliac pulse, running at much lower pressure, changes velocity much more gradually. Changes in the time of arrival therefore then reflect changes in the differential arterial pressure that the two pulses experience. While this differential pressure is not exactly pulse pressure, that is the difference between the full pulse arterial pulse height and the diastolic pressure floor, it represents about 60-70 percent of it, assuming the previously stated iliac reflection coefficient.

Fig. 2, presents a graphic display of the relative amplitudes of the left ventricular ejection (# 1) and the trailing reflection pulses and their resulting relative positions on the pulse propagation velocity curve, which is highly pressure dependent. As a result the arrival times of the different pulses are highly pressure dependent, a point that is clarified

by Fig. 3, which presents the pulse travel times, from bottom to top, respectively, of the primary ejection pulse (# 1), the renal reflection (# 2), and the iliac reflection (# 3). The iliac pulses arrival time shortens only slightly with increasing pressure because its amplitude remains close to the diastolic pressure regime. The renal reflection peaks arrival time (middle) experiences significant non-linearity because the reflection coefficient, R2, is highly pressure dependent. The left ventricular ejection (# 1, bottom curve) has the highest amplitude and samples the steepest section of the pressure/velocity curve and is therefore most pressure dependent. Using *Youngs moduli* obtained from the literature and letting the model fit R2 as well as the velocities of the primary arterial path ways it is then possible to compare experimental data with model predictions.

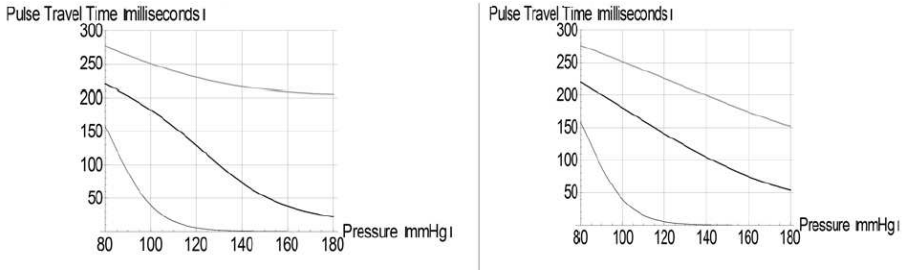


Figure 3: left graph (from [31]): Arrival times, as predicted by PDA model, of the three component pulses, from bottom up, # 1 (left ventricular ejection), # 2 (renal reflection), and # 3 (iliac reflection). Right graph: the effect of replacing the dynamic response of the renal reflection coefficient R2, shown in Fig. 12, with a constant reflection coefficient of 0.25.

Fig. 3, displays the fact that human arterial pathways, for the average height population we have studied, are generally very short relative the distances the arterial pulse traverses within a cardiac cycle. Typical arterial pulse propagation velocities range, for healthy and unstressed arteries, from $4 - 9\ m/s$. This fact influences particularly the arrival time of the # 1 pulse profoundly. In the lower pressure range, which is the pressure regime that was examined here, the # 1 pulse pulls away from the # 2 and # 3 reflection pulses, as evidenced by the fact that its arrival time shortens significantly faster with increasing pressure than the arrival times of # 2 and # 3. Consequently, in this pressure range, T13 would be expected to widen with increasing pressure and shorten with decreasing pressure. Fig. 3, therefore provides a quantitative basis for why T13 is hypothesized to be directly dependent on blood pressure changes in the blood pressure regime that was examined here.

As the pressure continues to increase, however, the arrival time of the primary # 1 pulse saturates as it runs out of arterial runway. Consequently further increases in arterial pulse propagation velocity do not result in a further shortening of the arrival time. Meanwhile the # 3 pulse continues to accelerate with increasing pressure, narrowing the T13 time delay in this high-pressure regime. The details of the pressure-dependent evolution of the arrival time curves are critically dependent on the choice of different velocity profiles for the different arterial sections, a point that is discussed later.

2.3 Pulse Decomposition Algorithm (PDA)

The algorithm that is based on the pulse analysis model just presented encompasses the following components:

1. A peak finder that identifies heartbeats in the derivative data stream.

2. A differentiator that produces the second derivative of the detected heart beat which is then used to find the inversions corresponding to the locations of the component pulses.

3. A digital integrator, implemented as a Bessel filter, that generates the integrated pulse wave form from the differentiated raw signal stream, and from which relative component pulse amplitudes are determined and 4. a low-pass filter that allows identification of the primary systolic peak. Furthermore the frequency content of the data stream is continuously analyzed in order to calculate signal to noise (S/N) figures of merit that determine whether signal fidelity is sufficiently high to permit peak detection and analysis.

The detection efficiency of the heart beats was typically on the order of 92 %, as evidenced by visual inspection of inter-beat spectra which readily reveal missed beats. Detection was typically poorest at the highest negative pressure ($-60\ mmHg$) because of significantly diminished pulse amplitude.

Once the temporal locations of the reflection component pulses and the systolic peak are identified, the T13 interval, the time delay between systolic (P1) and iliac peak (P3), is calculated. The P2P1 ratio is calculated using the amplitudes of the P2 peak and the systolic peak, in the integrated pulse spectrum. Detection efficiency of the component pulses was on the order of 90 %. Detection again was poorest at the highest negative pressure because of diminished pulse amplitude.

2.4 Statistical Analysis

We present regression coefficients between LBNP levels and pulse pressure responses of the three measurement systems. In order to compare relative sensitivities of the three systems to changes in pulse pressure we present results of different repeated measures ANOVA analysis, which were performed using the Minitab statistical software (Release 14, Minitab Ltd.). Data are presented as means +- SE unless specified otherwise.

3 Results

3.1 Comparison of Pulse Pressure Changes

In Fig. 4 we present an example of the evolution of arterial pressure pulse line shape changes for the 6 stages of an hour-long LBNP session (right-hand graph B) as well as the T13 trace for the entire session (left graph A). The subject in this case was a 31 y. female. The time evolution of the presented pulse line shapes is downward, starting at the top at atmospheric pressure, and ending with a pulse line shape obtained after

the LBNP chamber was vented from -60 $mmHg$ back to atmospheric pressure. Each pulse line shape represents a 10-pulse average. The dynamic range of the iliac and renal

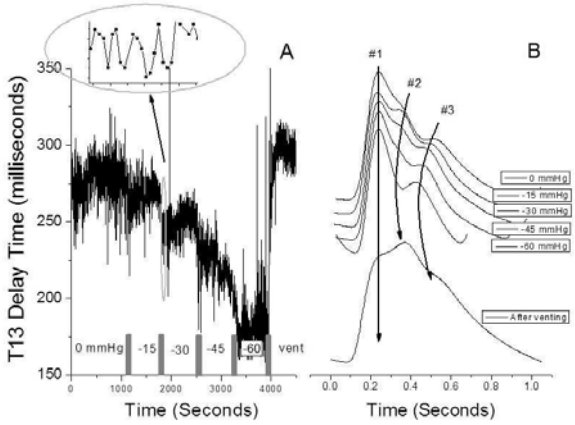


Figure 4: Relative amplitudes and velocity/pressure relationships of the component pulses. # 2's amplitude and propagation velocity can change significantly due to the pressure dependence of R2, the renal reflection coefficient. (from [31]).

peak positions is indicated by the downward sweeping arrows, while the position of the primary systolic peak (# 1) is indicated by the vertical solid arrow. The narrowing of the time interval between iliac and systolic component pulses with decreasing negative pressure is clearly visible. Furthermore, while the heart rate also changed, as indicated by the shortening inter-beat interval, it is clear that the rates of change for T13 and heart rate are different, i.e. the inter-beat interval narrowed faster than T13. A further point of interest is the shape of the arterial pressure pulse after venting, which in all subjects caused a significant rise in systolic blood pressure, as determined using the conventional blood pressure monitors. The pressure pulse line shape in question has the typical pulse shape associated with a positive augmentation index, which is defined as (height of # 2 pulse - height of # 1 pulse)/maximum overall amplitude [25]. A positive augmentation index is usually taken to be indicative of arterial aging, which, given the subjects age, is unlikely to be the case. This subjects pulse shape returned, along with normalizing systole, within minutes to the original line shape (top trace in Fig. 4B.)

While the results displayed in Fig. 4 exhibited a significant change in heart rate along with the change in T13, this was not a general observation. Figure 5 presents the results regarding inter-beat interval and T13 for subject # 9, a 24 y. male, who did not exhibit any appreciable change in heart rate until venting. The narrowing of T13 with decreasing negative pressure, however, matched those of all subjects.

Fig. 6 displays a representative side-by-side comparison of pulse pressures obtained with the automatic cuff (left graph) and the Finapres (center graph), as well as the evolution of the T13 parameter over the course of the LBNP session of subject # 3, (right graph). The general absence of a discernible trend in the readings of the cuff with progressing hypovolemia was typical for all data runs.

Fig. 7 presents comparative overall results for pulse pressures and T13 as a function of progressive decreasing negative pressure. Specifically, Fig. 7A presents the overall pulse

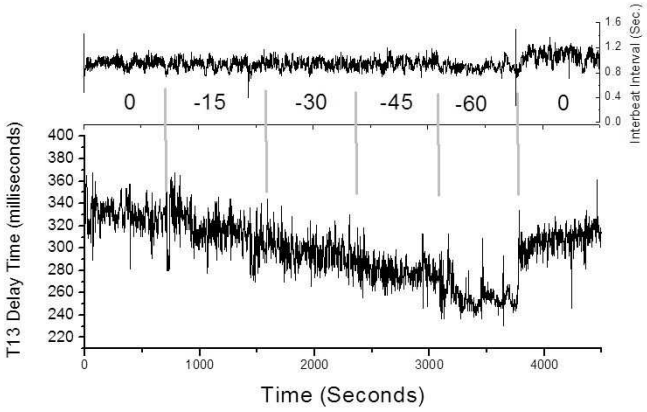


Figure 5: Temporal evolution of the inter-beat interval and T13 over the course of the LBNP session of subject # 9, 24 y. m., In this subject the narrowing of T13 was observed without any change in heart rate until venting. (from [31]).

pressure results of the automatic pressure cuff while Fig. 7B presents the overall results for T13. Fig. 7C. presents overall pulse pressure results for the Finapres.

The ability of the four measurement methods to resolve the effects of the different LBNP stages at a statistically significant level varied. While the PDA T13 parameter was able to resolve each of the four LBNP stages relative to atmospheric pressure, neither the Finapres nor the cuff were able to resolve the stages with the two least negative pressures (-15 & -30 $mmHg$), corresponding to the smallest changes in pulse pressure, with significance set at $p < 0.01$. Heart rate, as a detection modality for resolving the two least negative pressure LBNP stages, almost reached statistical significance, performing significantly better than the Finapres or the cuff. Table 1 presents the results of the ANOVA analysis.

Table 1: ANOVA: PDA, Finapres, Cuff versus LBNP (-15 & -30 mmHg)

Methodology	Significance
PDA - T13	0.00001
Finapres pulse pressure	0.636
Cuff pulse pressure	0.214
Heart rate (Finapres)	0.02

A receiver operating characteristic (ROC) analysis of the intra-subject ability of the four methodologies to resolve LBNP-induced differences of -15 $mmHg$ and atmospheric pressure, and -30 $mmHg$ and atmospheric pressure revealed similar differences. In Fig. 8 we present the results.

An important question is whether T13 is indeed a pulse pressure equivalent or whether it simply tracks the changes in heart rate referred to in Fig. 4 and Fig. 9 presents a comparison of pulse pressure, as measured with the cuff, as a function of T13 and heart rate, as measured with the Finapres, as a function of $T13$. While $T13$ correlates linearly

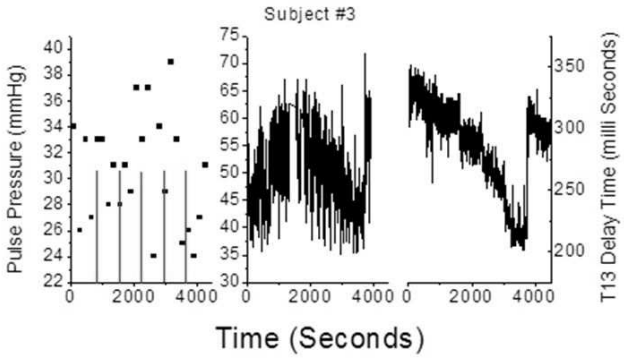


Figure 6: Comparison of the individual results for cuff-based pulse pressure (left graph), Finapres-based pulse pressure (center), and PDA-based T13 measurements, for subject # 3. The right panels present the simultaneously obtained T13 delay times between the primary left-ventricular ejection pulse and the iliac reflection pulse recorded on the subjects middle member of the middle digit. (from [31])).

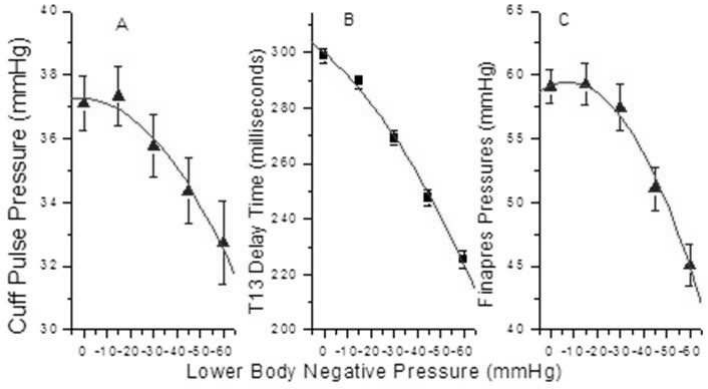


Figure 7: Overall results for pulse pressure obtained with the automatic cuff (graph A), the PDA pulse pressure-equivalent parameter T13 (B) and the Finapres (C). (from [31]).

with the pulse pressure determined using the cuff ($0.19\tilde{A} - T13$ (milliseconds) $+2.58$, $R2 = 0.98, p < 0.0001$) a second order model is required to obtain a correlation with heart rate ($-1.51 + T13$ (milliseconds) $\times\ 0.015 - 2.24526E - 5 \times T13E2$ (milliseconds)).

3.2 Comparison with model predictions

In Fig. 10 we present an overlay of the experimental results and the models predictions. The experimental data, all averages from 15 subjects, are the T13 values obtained from each LBNP stage as well as the corresponding pulse pressure values as determined with the Finapres. Since systole did not change appreciably for any of the subjects, we use the average value of 120 mmHg throughout. Consequently, as observed experimentally, changes in pulse pressure are driven entirely by changes in diastole. The most important aspect of the agreement between the model and data, as presented in Fig. 10, are the

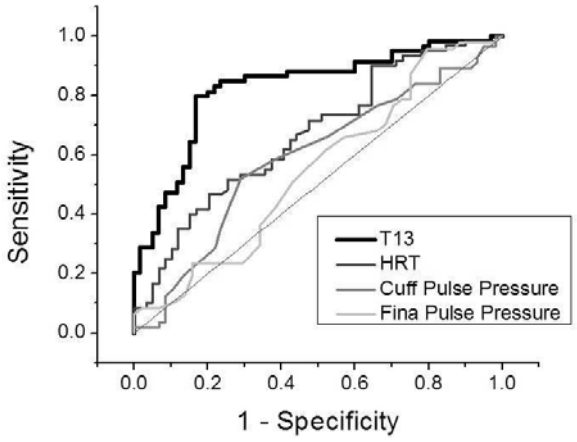


Figure 8: ROC analysis of the comparative intra-subject sensitivity/specificity of T13, heart rate, cuff pulse pressure, and Finapres pulse pressure to resolving the difference of -15 mmHg versus -30 mmHg LBNP relative to atmospheric pressure. Respectively the areas under the curve are 0.83, 0.66, 0.59, and 0.55. (from [31]).

arterial parameter assumptions that are required to achieve it. The single dominant factor that determines the response of T13 to pressure changes is the pressure/arterial pulse velocity response of the different arterial sections that the systolic pulse and its two central reflections traverse. Furthermore the range of relative pressure/velocity response curves that is possible, given the constraints of the experimental data, is very narrow. Fig. 11 presents the relative response curves of the three arterial segments that constitute the pulse pathways, the thoracic aorta, the abdominal aorta, and the arm complex arteries. Clearly the model at this stage uses a significant simplification of the arterial path sections and the response curves presented represent averages over these pathways. While more details will be introduced in future versions of the model the aim here is to demonstrate that the basic physical picture hypothesized by the PDA model matches observations. Fig. 12 displays the response characteristics of the reflection coefficient of the renal reflection site, as predicted by the model. While the starting values of the pressure/velocity curves for the arm complex arteries as well as the thoracic/abdominal aorta were based on published arterial pulse propagation velocities, [12] the results of the LBNP experiments provide an opportunity to deduce the relative dynamic response characteristics of the different sections, which are not readily available as they have not been the subject of research interests in a long time. In order to obtain the fit shown in Fig. 10, a different dynamic behavior of the arm complex relative to that of the central arteries had to be modeled. Specifically, while the arm complex arteries required a distinct exponential response characteristic, the simulated central arteries response, in the blood pressure range under consideration, was, in the case of the thoracic aorta, essentially linear and in the case of the abdominal aorta, of lower gain. These results are presented in Fig. 11. And it is this difference in dynamic response that enables the model to generate T13 curves whose slopes match those observed. In contrast, changes in starting values only shifted the family of curves in parallel up or down in but did not change the relative slopes of the curves.

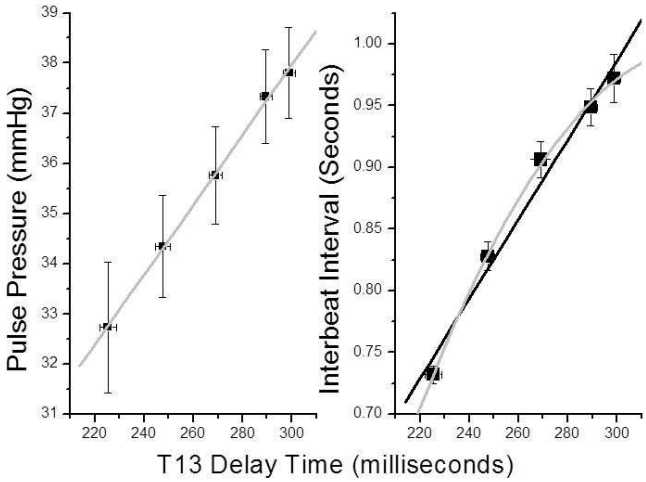


Figure 9: Functional comparison of T13 with pulse pressure obtained from cuff (left), and heart rate (right). (from [31]).

The other significant factor affecting the models prediction is the dynamic response curve of the renal reflection coefficient R2 as a function of pressure (Fig. 12). Not surprisingly, the dynamic response of the reflection coefficient becomes increasingly important at higher pressures. The effect is clearly seen in the modification of the arrival time curves of the component pulses, which are presented in the two graphs of Fig. 3. The left graph corresponds to the arrival times with the R2 having the pressure response characteristic shown in Fig. 12. The right graph displays the dramatically altered results, which would yield a very poor fit with the observed delay times, for a constant reflection coefficient of 0.25.

4 Discussion

Before discussing the implications of what appear to be statistically significant correlative results it is important to consider the hypothesis of whether they could be due to an unrelated experimental artifact, specifically the increasing abdominal compression with the increasing LBNP pressures that has been reported [26]. Two arguments can be made to refute this concern. If the increasing abdominal compression were to have progressively given rise to a new reflection site between renal and iliac reflection sites, a new reflected component pulse would have arisen between the # 2 and # 3 component pulses, progressively increasing with each LBNP-induced abdominal compression stage. Such additional central reflection sites have been observed by Kriz [20] in the context of aortic aneurysms. We have observed the resulting additional component pulses of such aortic aneurysms in the arterial periphery, a subject of future publication. However, as part of this work, in none of the LBNP stages and in none of the subjects studied were such additional component pulses observed.

A second argument is that the same response in T13 is observed in the case of actual

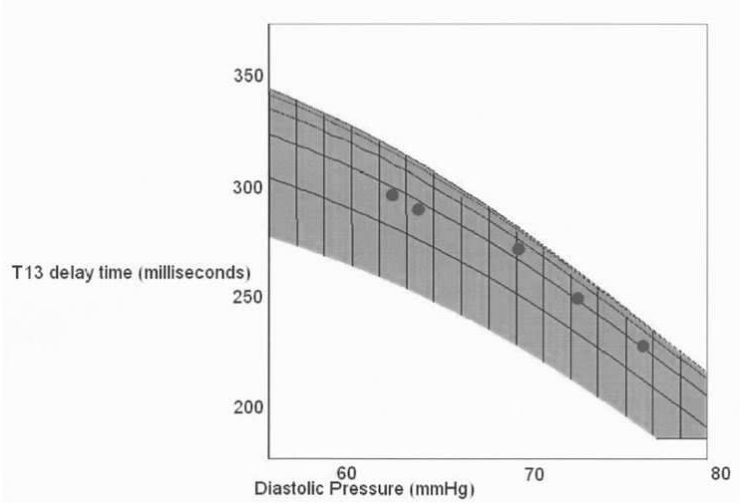


Figure 10: Overlap of PDA T13 prediction and LBNP study data as a function of diastolic pressure change. The line around which the data are grouped corresponds to systole = 120 mmHg. Adjacent lines correspond to +-5 mmHg. (from [31]).

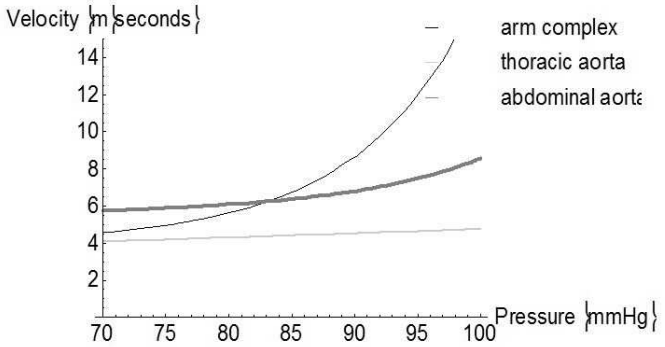


Figure 11: Arterial pulse velocity profiles as a function of pressure, based on fit to experimental results in Fig. 10.

hemorrhage, induced through blood donation of 1 *pint*. A just-completed study with 50 subjects, whose results will be published shortly, has confirmed the continuous decrease of T13 as the blood donation progressed. Moreover, the decrease observed in the blood donation experiment matched well the average decrease that was observed in the first stage ($-15\ mmHg$) of the LBNP results reported here, in line with the estimation of central blood loss for that stage [22].

The results presented therefore support the hypothesis that the time delay between the primary component pulse (# 1) and the iliac reflection pulse (# 3), T13, correlates with pulse pressure and provide a first milestone in the validation of the PDA model. A number of conclusions follow.

Any reflection sites in the arm complex arteries proximal to the radial/digital arteries will not affect the pulse line shape that is observed there because any pulse reflections due

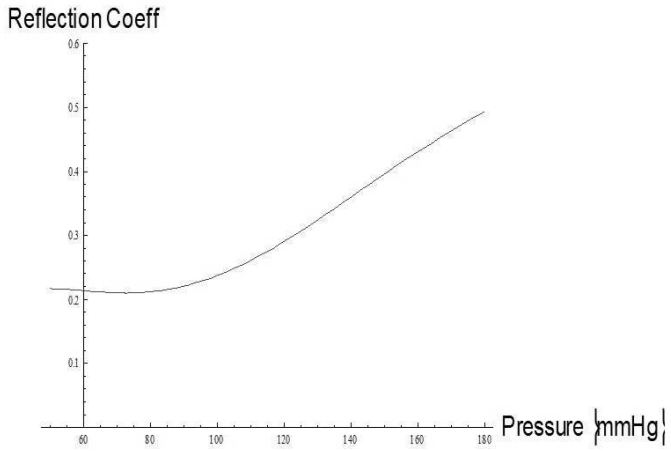


Figure 12: Pressure response of R2, reflection coefficient of the junction of thoracic/abdominal aortic sections at the height of the renal arteries.

to such reflection sites will travel away from them and back toward the central arteries. Their re-entry into the arm complex arteries could only be accomplished as re-reflections, with dramatically reduced amplitudes that would be masked by the primary renal and iliac reflections.

The hypothesis that the shape of the peripheral pressures pulse is predominantly determined by reflections in the aorta may seem surprising and will no doubt be gradually accepted. On the one hand maneuvers that are known to modify the thoracic pressure profile, such as valsalva that selectively modulates the renal reflection site and therefore the # 2 component pulse, [17] provide a ready method to demonstrate the critical importance of the central reflection sites. Alternatively manipulative experiments in the arterial periphery could be suggested to support or challenge the hypothesized physical picture. One possibility is partial occlusion of a femoral artery, with pulse monitoring distally on, for example, the pedal artery. Particularly since in the legs the # 2 and # 3 component pulses are diminished because they are re-reflections, it should be possible to observe the additional component pulse that would originate from the partial occlusion site, travel to the iliac reflection site and reflect there toward the distal monitoring site.

Movement of the occlusion site along the leg arteries should change the timing of the additional component pulse relative to the pulse envelope.

A relevant physiological phenomenon is pressure pulse amplification of the arterial periphery that is attributed to the taper of the arterial walls as well as their changing wall composition relative to the core arteries. The important realization here is that, while the arterial pressure pulse is temporally compressed and increased in amplitude, these changes to the pulse envelope are, in the absence of arterial dilation or constriction, static. This is the basis of using the validated transfer function approach [12] that uses pulse shapes obtained in the arterial periphery to predict central artery pulse shapes and blood pressures. It is the central artery dynamics that determine the peripheral pulse in the transfer function model. In the PDA model, which offers a concrete physical model instead of a generalized Fourier-based inverse filter, it is also the central artery dynamics that dominate the relationships between the components. Significant arterial dilation and

constriction does modify the component pulse relations, and this is an object of current study.

The differential pressure/velocity response curves predicted by the model for the different arterial sections that the iliac reflection pulse (# 3) traverses are not very surprising. The change-over from elastin, the main component of central artery walls [12] to the tougher collagen, which characterizes arterial wall content in the periphery, is well known. Consequently one would expect, particularly in the case of the thoracic aorta that has the slowest arterial pressure pulse propagation velocity (see starting value of 4.2 m/s), a much diminished increase in the velocity with rising pressure.

The regard to the changes in the arrival times of the component pulses (Fig. 3) with a change in the pressure response characteristic of the R2 (renal) reflection coefficient, this relation might at first appear surprising. The result is however less surprising after considering that the pressure response dynamic of R2 redistributes amplitude between the component pulses; in particular that it increases the P2 (renal) reflection pulse amplitude while reducing the P3 (iliac) pulse amplitude as the pressure rises. With lowered pressure amplitude, P3 travels slower than it would if it were not losing amplitude to P2 because of the pressure dependent R2 coefficient. The agreement with observed results consequently provides a rather compelling argument in support of the dynamic response of R2. The point was made earlier that the differential pressure between P1, the primary ejection pressure pulse, and P3, the iliac reflection pulse, is not exactly pulse pressure, that is the difference between the full pulse arterial height and the diastolic floor, but about $60-70$ percent of it. Consequently the question arises how useful the differential pressure, which is reflected in the time interval T13, is in approximating pulse pressure. As clarified by Fig. 2, T13 corresponds to the pressure amplitude difference between the primary ejection pulse (# 1) and the iliac reflection (# 3). From the graph one readily sees that this amplitude difference covers exactly the most important part of the response curve that relates pressure to velocity, this being the exponential response section. The amplitude of the (# 3) pulse is entirely in the diastolic regime, which is linear, as Anliker showed. [32] However, it is the exponential section that is responsible for 95 percent of the dynamic pressure/velocity response characteristics, the rest being simply a linear offset.

The head plays a much diminished role in regard to pressure pulse reflections that are observed at the arterial periphery of the arm. Arterial pressure pulse reflections that return via the carotid arteries will, upon entering the aortic arch and traveling along the descending thoracic, re-reflect off the reflection site in the vicinity of the renal arteries. Assuming a reflection coefficient on the order of 17 percent, the amplitude of such a re-reflection will be on the order of 3 percent of the primary peak amplitude, and consequently be masked by the much larger pressure primary pulse reflections # 2 and # 3.

The PDA model ties together recent related observations by others. The fact that the ratio of the amplitudes of the # 2 (P2) and # 1 (P1) pulses correlates with systolic pressure is not surprising in light of the results obtained by Takazawa et. al., [27] Takada et. al., [28] and Imanaga et al. [29] In Fig. 13, which displays the 10-pulse average of the arterial digital pressure pulse of a 21 y. male athlete, we also present the second derivative of the arterial pulse. Takazawa et. al. labeled the different inversions, *waves*, of this second derivative trace as indicated in Fig. 13. The results of several studies suggest

that the ratio d/a correlates with blood pressure, along with many other physiological parameters [27-28]. Comparative inspection of the two traces establishes that the waves **a** and **d** are temporally in very similar positions to P1 and P2, respectively. A parameter introduced by Bartolotto [25] that incorporates very similar definitions of P1 and P2 as the PDA model was also found to correlate with systolic pressure. This parameter is the augmentation index of the photoplethysmograph, PTG (AUGI), and it is defined as (P2-P1)/MA, where P2 and P1 are the amplitudes of the primary and second systolic peaks in the photoplethysmograph, respectively, and MA is the maximum envelope amplitude. These correlative results support the PDA model, which supplies the physical explanation both for the origin of the component pulses as well as why the correlation of the relative amplitudes of these component pulses with systolic pressure exists. As with the PDAs

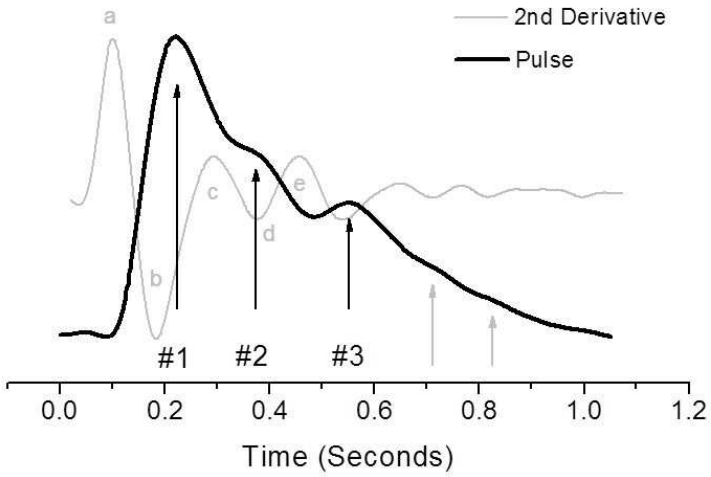


Figure 13: Arterial pressure pulse line shape (black trace, 10-pulse average), and its second derivative (gray trace), of a 21 y. male athlete collected at the middle phalange of the middle finger. The pulse line shape displays distinct pulsatile features labeled # 1, # 2, # 3. While the # 1 is the direct pass due to left ventricular ejection, the rest of the pulse envelope is due to arterial reflections. Indicative of distinct arterial pulse reflection sites is the fact that the reflected wave exhibits pulsatile components (# 2, # 3, and beyond) that feature comparable temporal extents as the primary ejection-related feature # 1. The inversions of the second derivative trace are labeled according to the convention introduced by Takada et. al. (from [31]).

P2P1 parameter, others have suggested measures that utilize the same time interval corresponding to T13 and have somewhat comparable physical interpretations. Millasseau [30] labels the time delay PPT in the digital volume pulse and suggests that it corresponds to the transit time of pressure waves from the root of the subclavian artery to the *apparent* site of reflection *in the lower body* and back to the subclavian artery. The reason for choosing the subclavian artery as a starting and ending point is however unclear, since the pressure wave does not originate there. If, on the other hand, the subclavian artery were to give rise to the # 3 pressure pulse as a reflection site, the amplitude of the iliac pulse would be dramatically lower than what is observed $(20-40)$ percent of the primary

peak) at the radial or digital artery. Succinctly put, the pulse would have traveled from the left ventricle to the subclavian artery, reflected there at some reflection site, then to travel to the iliac reflection site. It would return from there as a re-reflection pulse with commensurately much reduced amplitude, an unlikely scenario.

A significant benefit of measuring T13 over pulse pressure directly is its higher resolution and sensitivity.

The results indicate the equivalence of a change of about 200 milliseconds in T13 to a variation of about 8 mmHg in pulse pressure over the entire range of a simulated central blood loss in excess of 1 liter for this cohort of fit and relatively young subjects. The results therefore indicate that the PDA technology is capable of resolving small changes in pulse pressure, a feat that sphygmomanometers are not well suited for. Given the suggestion by others that pulse pressure can be considered as a surrogate for stroke volume and therefore as a means to track loss of blood volume in trauma patients, [1] the accurate monitoring of pulse pressure could be a vital component in predicting hemorrhagic shock.

The potential benefits of utilizing T13 in detecting small changes in pulse pressure, coupled with the small size of the wireless CareTaker hardware, which weighs on the order of $5\ oz$, and the fact that it tracks blood pressure at low coupling pressures, makes the system attractive for the monitoring of patients at risk for internal hemorrhage. A benefit of such field-based monitoring is that internal hemorrhage could be detected well before hemodynamic collapse, making timely intervention feasible.

Currently studies are underway to further validate the PDA model by simultaneously correlating intra-aortic blood pressure with the peripherally measured PDA parameters T13 and P2P1.

5 Conclusions

We have presented a new physical model of the propagation of the arterial pressure pulse and its reflections as well as a comparison of the predictions of the model with experimentally obtained pulse parameters and conventionally obtained blood pressures. The agreement of observations and measurements provides a preliminary validation of the model which in turn could provide a renewed impetus in the study of the human arterial pressure pulse. The model is based on few, physical, assumptions because it proposes that the structure of the pulse is due to it is readily identifiably arterial pulse reflection sites. As a result it is also readily testable.

Acknowledgement: This work was supported by the Office of Naval Research and the National Institutes of Health for portions of this research through contract N00014-09-M-0146 and grant 1R43HL087476-01A1.

References

[1] V. A, Convertino, W. H, Cooke , J. B, Holcomb, "Arterial pulse pressure and its association with reduced stroke volume during progressive central hypovolemia," J. Trauma, 61(3), 629-634, 2006.

[2] P, Leonetti, F, Audat, A, Girard, D, Laude, F, Lefrere, J. L, Elghozi, "Stroke volume monitored by modeling flow from finger arterial pressure waves mirrors blood volume withdrawn by phlebotomy," Clin Auton. Res, 14, 176-181, 2004.

[3] R. F, Bellamy, "The causes of death in conventional land warfare: implications for combat casualty care research," Mil Med, 149, 55-62, 1984.

[4] C. J, Carrico, J. B, Holcomb, I. H, Chaudry, "PULSE trauma work group. Post resuscitative and initial utility of life saving efforts. Scientific priorities and strategic planning for resuscitation research and lifesaving therapy following traumatic injury: report of the PULSE trauma work group," Acad Emerg Med, 9, 621-626, 2002.

[5] J. I, Davis, M. M, Band, "Peripheral blood pressure measurement is as good as applanation tonometry at predicting ascending aortic blood pressure," J. of Hyperten, 21, 571-576, 2003.

[6] M, Johannes, "Handbuch der Physiologie des Menschen fur Vorlesungen," 2 Vols. Coblenz: Verlag von J. Hôlscher.

[7] O, Frank, "Die Grundform des arteriellen Pulses. Erste Abhandlung," Mathematische Analyse, Z. Biol, 37, 483-526, 1899.

[8] P. H, Broemser, Uber die Messung des Schlagvolumens des Herzens auf unblutigem Weg," Zeitung fur Biologie, 90, 467-507, 1930.

[9] R, Busse, R. D, Bauer, A, Schabert, Y, Summa, P, Bumm, E, Wetterer, "The mechanical properties of exposed human common carotid arteries in vivo," Basic Res Cardiol, 74(5), 545-554, 1979.

[10] C. M, Quick, D. S, Berger, A, Noordergraaf, "Constructive and destructive addition of forward and reflected arterial pulsewaves," Am J Physiol Heart Circ Physiol, 280(4), H1519-27, 2001.

[11] M. G, Taylor, "Hemodynamics," Annual Review of Physiology, 35, 87-116, 1973.

[12] W. W, Nichols, M. F, O'Rourke, "McDonald's blood flow in arteries. Theoretical, experimental and clinical principles," London: Edward Arnold, 1999.

[13] S, Soderstrom, G, Nyberg, J, Ponten, J, Sellgren, M. F, O'Rourke, "Substantial equivalence between ascending aortic pressure waveforms and waveforms derived from radial pulse using a generalised transfer function?" [Abstract]. FASEB J, 12, 4131, 1998.

[14] M, Vyas, J. L, Izzor, Y, Lacourcière, J. M, Arnold, M. E, Dunlap, J. L, Amato, M.A, Pfeffer, G. F, Mitchell, "Augmentation index and central aortic stiffness in middle-aged to elderly individuals," Am J Hypertens, 20, 642-647, 2007.

[15] G. M, London, J, Blacher, B, Pannier, A. P, Guerin, S. J, Marchais, M. E, Safar, "Arterial wave reflections and survival in end-stage renal failure," Hypertension, 38, 434-438, 2001.

[16] M, Karamanoglu, D. E, Gallagher, A. P, Avolio, M. F, O'Rourke, "Functional origin of reflected pressure waves in a multibranched model of the human arterial system," Am. J. Physiol, 267 (Heart Circ. Physiol. 36), H1681-H1688, 1994.

[17] R. D, Latham, N, Westerhof, P, Sipkema, B. J, Rubal, P, Reuderink, J. P, Murgo, "Regional wave travel and reflections along the human aorta: a study with six simultaneous micromanometric pressures," Circulation, 72, 1257-1269, 1985.

[18] C. T, Ting, M. S, Chang, S. P, Wang, B. N, Chiang , F. C, Yin, "Regional pulse wave velocities in hypertensive and normotensive humans," Cardiovasc Res, 24(11), 865-872, 1990.

[19] J, Kriz, P, Seba, "Force plate monitoring of human hemodynamics," Nonlinear Biomed Phys, 2(1), 2008.

[20] J, Kriz, K, Martinik, P, Seba, V, Tosnerova, "Force Plate Measurement of Human Hemodynamics," http://cdsweb.cern.ch/record/855869/files/0507135.pdf?version=1.

[21] W. H, Cooke, J, Salinas, V. A, Convertino, "Heart rate variability and its association with mortality in pre-hospital trauma patients," J Trauma, 60, 363-370, 2006, discussion 370.

[22] W. H, Cooke, V. A, Convertino, "Heart Rate Variability and Spontaneous Baroreflex Sequences: Implications for Autonomic Monitoring During Hemorrhage," J. Trauma, Injury, Infection, and Critical Care, 58(4), 798-805, 2005.

[23] D. J, Korteweg, Uber die Fortpflanzungsgesschwindigkeit des Schalles in elastischen Rohren," Annals Phys Chem (NS), 5, 520-537, 1878.

[24] M, Anliker, M. B, Histand, E, Ogden, "Dispersion and attenuation of small artificial pressure waves in the canine aorta," Circ. Res, 23, 539-551, 1968.

[25] L. A, Bortolotto, J, Blacher, T, Kondo, K, Takazawa, M. E, Safar, "Assessment of vascular aging and atherosclerosis in hypertensive subjects: second derivative of photoplethysmogram versus pulse wave velocity," Am J Hypertens, 13, 165-171, 2000.

[26] H. G, Hinghofer-Szalkay, N, Goswami, A, Rôssler, E, Grasser, D, Schneditz, "Reactive hyperemia in the human liver," Am J Physiol Gastrointest Liver Physiol, 295(2), G332-7, 2008.

[27] K, Takazawa, N, Tanaka, M, Fujita, O, Matsuoka, T, Saiki, M, Aikawa, "Assessment of vasoactive agents and vascular aging by second derivative of the photoplethysmograph waveform," Hypertension, 32, 365-370, 1998.

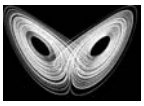
[28] H, Takada, K, Washino, J. S, Harrell, H, Iwata, "Acceleration plethysmography to evaluate aging effect in cardiovascular system," Med Progress Technol, 21, 205-210, 1997.

[29] I, Imanaga, H, Hara, S, Koyanagi, K, Tanaka, "Correlation between wave components of the second derivative of plethysmogram and arterial distensibility," Jpn Heart J, 39, 775-784, 1998.

[30] S. C, Millasseau, J. M, Ritter, K, Takazawa, P. J, Chowienczyk, "Contour analysis of the photoplethysmographic pulse measured at the finger," J. Hypertens, 24, 1449-1456, 2006.

[31] M. C, Baruch, D. E, Warburton, S. S, Bredin, A, Cote, D. W, Gerdt, C. M, Adkins, "Pulse Decomposition Analysis of the digital arterial pulse during hemorrhage simulation," Nonlinear Biomed Phys, 12, 5(1):1, 2011.

[32] M, Anliker, W. E, Moritz, E, Ogden, "Transmission characteristics of axial waves in blood vessels," J Biomech, 1(4), 235-246, 1968.



Annual Review of Chaos Theory, Bifurcations and Dynamical Systems
Vol. 1, (2012) 50-60, www.arctbds.com.

A Unified Piecewise Smooth Chaotic Mapping that Contains the Hénon and the Lozi Systems

Zeraoulia Elhadj
Department of Mathematics, University of Tébessa, (12002), Algeria.
e-mail: zeraoulia @ mail.univ-tebessa.dz, and zelhadj12 @ yahoo.fr.

J. C. Sprott
Department of Physics, University of Wisconsin, Madison, WI 53706, USA.
e-mail: sprott@physics.wisc.edu.

Abstract

In this paper we introduce a new piecewise smooth mapping of the plane as a unified discrete-time chaotic system that contains the original Hénon and Lozi systems as two extremes and other systems as a transition in between and that has robust homoclinic chaos over a portion of its key system parameters. Dynamical behaviors of the unified system are investigated in some detail.

Keywords: A unified piecewise smooth map, transition Hénon-like and Lozi-like chaotic attractors, robust homoclinic chaos.

Manuscript accepted November 12, 2011.

1 Introduction

Discrete mathematical models arise directly from experiment or by the use of the Poincaré section for the study of continuous models. Two of these models are the Hénon [1] and Lozi [2] maps given by, respectively:

$$H(x,y)=\begin{pmatrix} 1-ax^2+y \\ bx \end{pmatrix} \text{ and } L(x,y)=\begin{pmatrix} 1-a\left|x\right|+y \\ bx \end{pmatrix}. \tag{1}$$

The H mapping gives a chaotic attractor called the *Hénon attractor*, which is obtained for $a=1.4$ and $b=0.3$ as shown in Fig. 1(a). There are many papers that discuss the original Hénon and Lozi maps such as [3-6]. Moreover, it is possible to change the form

of the Hénon mapping H to obtain other chaotic attractors [2-7-8]. Applications of these maps include secure communications using the notions of chaos [11-12]. The Lozi map L is a 2-D invertible iterated map that gives a chaotic attractor called the *Lozi attractor*, which is obtained for $a = 1.4$ and $b = 0.3$ as shown in Fig. 1(b). It is therefore interesting to ask if there is a chaotic system that can unify these two chaotic systems and realize the continued transition from one to the other. This paper provides a positive answer to this question and reveals a surprising property of the transitional systems.

2 Robust chaos and its applications

Robust chaos is defined by the absence of periodic windows and coexisting attractors in some neighborhood of the parameter space, since the existence of these windows in some chaotic regions implies that small changes of the parameters would destroy the chaos. This effect implies the fragility of this type of chaos. Contrary to this situation, there are many practical applications such as in communications and spreading the spectrum of switch-mode power supplies to avoid electromagnetic interference [13-14] where it is necessary to obtain reliable operation in the chaotic mode and thus where robust chaos is required. A practical example can be found from electrical engineering to demonstrate robust chaos as shown in [10]. The occurrence of robust chaos in a smooth system is proved and discussed in [16] along with a general theorem and a practical procedure for constructing S-unimodal maps that generate robust chaos. This result is contrary to the hypothesis that robust chaos cannot exist in smooth systems [10]. On the other hand, many methods are used to search for a smooth and robust chaotic map, such as in [15], where a one-dimensional smooth map that generates robust chaos in a large domain of the parameter space is presented. In [17], simple polynomial unimodal maps that show robust chaos are constructed. Other methods are given in [16-18].

3 The unified chaotic system that contains the Hénon and the Lozi mappings

Since practical applications of chaos require the chaotic orbit to be robust, we introduce in this paper a new unified chaotic system that reduces to the original Hénon and Lozi systems [1-2] as two extremes and to other systems as a transition in between, and which has robust homoclinic chaos over a portion of its key system parameters. The proposed unified chaotic model is a piecewise smooth map of the plane defined by:

$$U(x,y) = \begin{pmatrix} 1 - 1.4 f_\alpha(x) + y \\ 0.3x \end{pmatrix}, \tag{2}$$

where $0 \leq \alpha \leq 1$ is the bifurcation parameter and the function f_α is given by:

$$f_\alpha(x) = \alpha |x| + (1-\alpha) x^2. \tag{3}$$

It is easy to remark that for $\alpha = 0$, one has the original Hénon map, and for $\alpha = 1$, one has the original Lozi map, and for $0 < \alpha < 1$, the unified chaotic map (2) is chaotic

with different kinds of attractors. The Lyapunov exponents and bifurcation diagram are shown in Fig. 2. We remark that the unified chaotic map (2) is a piecewise smooth map and due to the shape of the vector field U of the unified chaotic map (2), the plane can be divided into two regions denoted by:

$$D_1 = \left\{(x, y) \in \mathbb{R}^2 / \ x < 0\right\} \tag{4}$$

$$D_2 = \left\{(x, y) \in \mathbb{R}^2 / \ x > 0\right\}. \tag{5}$$

Let us define:

$$A = \left\{(x, y) \in \mathbb{R}^2 / \ x = 0\right\}, \tag{6}$$

which denotes a smooth curve that divides the phase plane into two regions D_1 and D_2, so that the unified chaotic map (2) can be rewritten as follow:

$$U(x, y) = \begin{pmatrix} \begin{cases} 1.4(\alpha - 1)x^2 + 1.4\alpha x + y + 1, & \text{if } (x, y) \in D_1 \\ 1.4(\alpha - 1)x^2 - 1.4\alpha x + y + 1, & \text{if } (x, y) \in D_2 \end{cases} \\ 0.3x \end{pmatrix}, \tag{7}$$

where in each of these regions the system (2) is a quadratic map. Notably, the unified system (2) has some special features and advantages as follows:

(1) System (2) is chaotic when $0 \leq \alpha \leq 1$.

(2) System (2) connects the Hénon and the Lozi maps and realizes the entire transition spectrum from one to the other.

(3) The control parameter α in system (2) reveals the evolution of dynamical behaviors from the Hénon to the Lozi attractors.

(4) System (2) has robust chaotic attractors for $0.493122734 \leq \alpha < 1$, while it is absent for $\alpha = 0$ and $\alpha = 1$.

4 Numerical simulations

In this section, the dynamical behaviors of the unified chaotic system (2) will be investigated numerically. For $0 \leq \alpha \leq 1$, the unified chaotic system has two kinds of chaotic orbits: Hénon-like chaotic attractors over the first portion of the interval $[0, 1[$ and a Lozi-like chaotic attractor over the second portion of the interval $]0, 1]$ as shown in Fig. 3(a) and (c). It seems that this phenomenon is related to the shape of the function f_α, where for values of α close to zero, the function f_α given in (3) behaves like the quadratic term x^2, while the values of α close to unity the function f_α behaves like the absolute value function $|x|$, as shown in Fig. 3(b) and (d). This explains the occurrence of the two kinds of chaotic attractors mentioned above.

5 A rigorous proof of the robustness of the homoclinic chaos

In this section, we begin by studying the existence of the fixed point of the U mapping in order to determine the associated normal form for the unified chaotic map (2), which

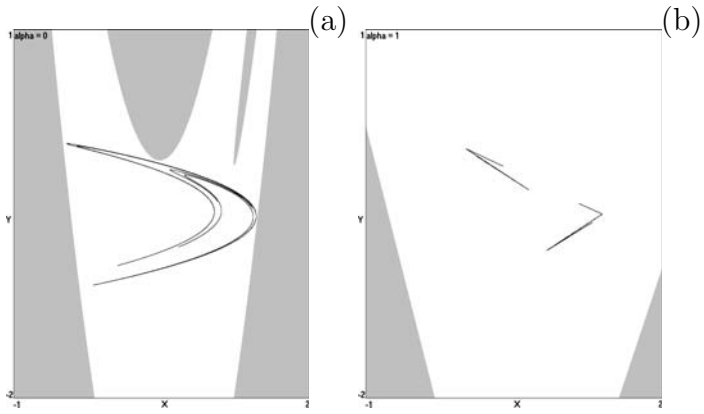


Figure 1: (a) The original Hénon chaotic attractor obtained from the H mapping with its basin of attraction (white) for $a = 1.4$ and $b = 0.3$. (b) The original Lozi chaotic attractor obtained from the L mapping with its basin of attraction (white) for $a = 1.4$ and $b = 0.3$.

permits us to prove rigorously the occurrence of robust homoclinic chaos, where we exclude the values $\alpha = 0$ and $\alpha = 1$ since both the Hénon and Lozi mapping are studied in detail in several works and in the references therein. We will show that if $0 \leq \alpha < 1$, then the unified chaotic map (2) has two fixed points given by:

$$P_1 = (x_1, 0.3x_1) \in D_1 \text{ and } P_2 = (x_2, 0.3x_2) \in D_2, \tag{8}$$

where

$$\begin{cases} x_1 = \frac{-0.7\alpha+0.35+\frac{\sqrt{-7.56\alpha+1.96\alpha^2+6.09}}{2}}{1.4(\alpha-1)} \\ \\ x_2 = \frac{0.7\alpha+0.35-\frac{\sqrt{-3.64\alpha+1.96\alpha^2+6.09}}{2}}{1.4(\alpha-1)}. \end{cases} \tag{9}$$

Obviously, the fixed points of the unified chaotic map (2) are the real solutions of the system:

$$1 - 1.4f_\alpha(x) + y = x \text{ and } y = 0.3x. \tag{10}$$

Hence one may easily obtain the two equations:

$$1.4(\alpha - 1)x^2 + (1.4\alpha - 0.7)x + 1 = 0 \text{ for } x < 0 \text{ and } y = 0.3x \tag{11}$$

$$1.4(\alpha - 1)x^2 - (1.4\alpha + 0.7)x + 1 = 0 \text{ for } x > 0 \text{ and } y = 0.3x. \tag{12}$$

If $0 \leq \alpha < 1$, then $1.4(\alpha - 1) < 0$, and the discriminant of the first equation of (11) is $-7.56\alpha + 1.96\alpha^2 + 6.09 > 0$. Thus, one can easily conclude that the only negative solution of the first equation of (11) is:

$$x_1 = \frac{-0.7\alpha + 0.35 + \frac{\sqrt{-7.56\alpha+1.96\alpha^2+6.09}}{2}}{1.4(\alpha - 1)} < 0. \tag{13}$$

On the other hand, the discriminant of the first equation of (12) is $-3.64\alpha + 1.96\alpha^2 + 6.09 > 0$ for all $0 \leq \alpha < 1$. Thus, one can easily conclude that the only positive solution of

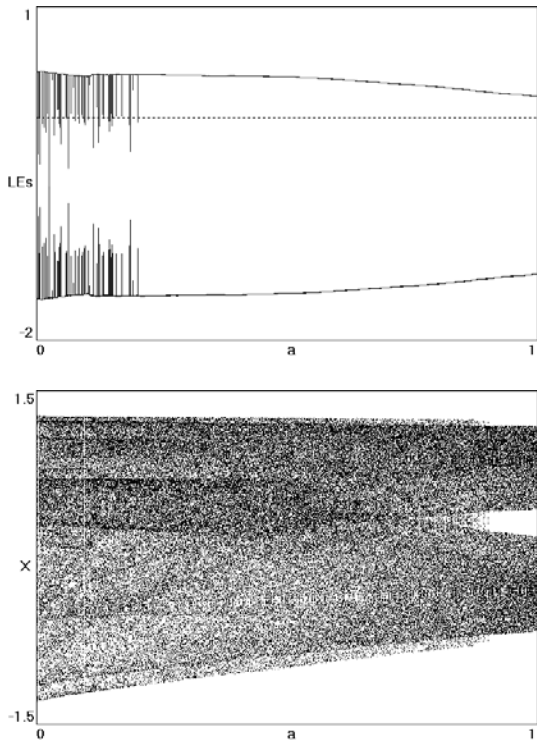


Figure 2: (a) Variation of the Lyapunov exponents of the unified map (2) for $0 \leq \alpha \leq 1$. (b) The bifurcation diagram of the unified chaotic map (2) for $0 \leq \alpha \leq 1$.

the first equation of (12) is:

$$x_2 = \frac{0.7\alpha + 0.35 - \frac{\sqrt{-3.64\alpha + 1.96\alpha^2 + 6.09}}{2}}{1.4\,(\alpha - 1)} > 0. \tag{14}$$

Finally, the unified chaotic map (2) has two simultaneous fixed points defined for $0 < \alpha < 1$ as $P_1 = (x_1, 0.3x_1) \in D_1$ and $P_2 = (x_2, 0.3x_2) \in D_2$.

The Jacobian matrix of the unified chaotic map (2) evaluated at a point (x, y) in the region D_1 is given by:

$$J_1\,(x, y) = \begin{pmatrix} 1.4\alpha - 2.8x + 2.8x\alpha & 1 \\ 0.3 & 0 \end{pmatrix}, \tag{15}$$

and at a point (x, y) in the region D_2 the Jacobian matrix is given by:

$$J_2\,(x, y) = \begin{pmatrix} 2.8x\alpha - 1.4\alpha - 2.8x & 1 \\ 0.3 & 0 \end{pmatrix}. \tag{16}$$

Thus, at P_1 one has:

$$J_1\,(P_1) = \begin{pmatrix} 0.7 + \sqrt{1.96\alpha^2 - 7.56\alpha + 6.09} & 1 \\ 0.3 & 0 \end{pmatrix}. \tag{17}$$

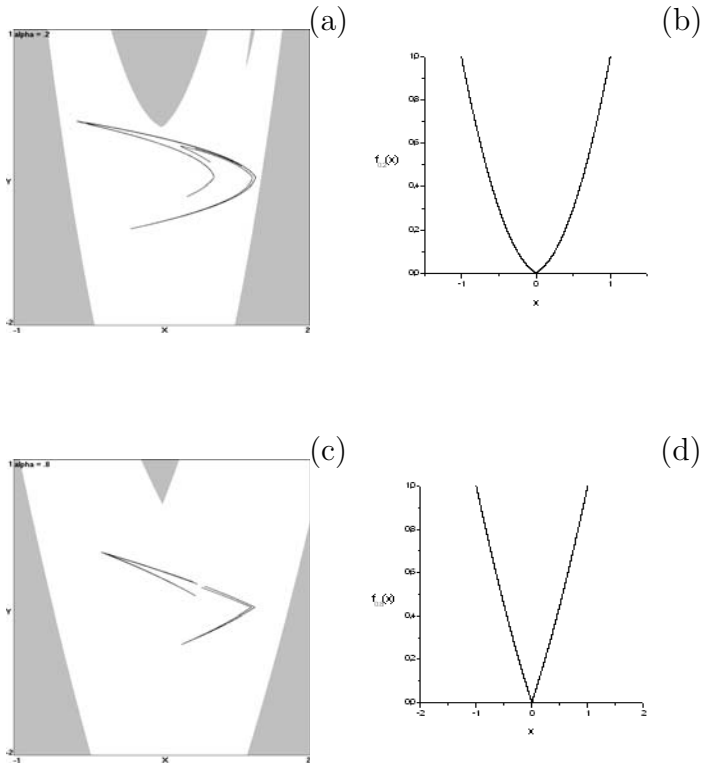


Figure 3: (a) The transition Hénon-like chaotic attractor obtained for the unified chaotic map (2) with its basin of attraction (white) for $\alpha = 0.2$. (b) The graph of the function $f_{0.2}$. (c) The transition Lozi-like chaotic attractor obtained for the unified chaotic map (2) with its basin of attraction (white) for $\alpha = 0.8$. (d) The graph of the function $f_{0.8}$.

The eigenvalues of $J_1(P_1)$ are

$$\begin{cases} \lambda_1 = \frac{\sqrt{1.96\alpha^2-7.56\alpha+6.09}+\sqrt{1.96\alpha^2-7.56\alpha+1.4\sqrt{1.96\alpha^2-7.56\alpha+6.09}+7.78}}{2} + 0.35 \\ \lambda_2 = \frac{\sqrt{1.96\alpha^2-7.56\alpha+6.09}-\sqrt{1.96\alpha^2-7.56\alpha+1.4\sqrt{1.96\alpha^2-7.56\alpha+6.09}+7.78}}{2} + 0.35, \end{cases} \tag{18}$$

and at P_2 one has:

$$J_2(P_2) = \begin{pmatrix} 0.7 - \sqrt{1.96\alpha^2 - 3.64\alpha + 6.09} & 1 \\ 0.3 & 0 \end{pmatrix}. \tag{19}$$

The eigenvalues of $J_2(P_2)$ are:

$$\begin{cases} \omega_1 = \frac{-\sqrt{1.96\alpha^2-3.64\alpha+6.09}+\sqrt{1.96\alpha^2-3.64\alpha-1.4\sqrt{1.96\alpha^2-3.64\alpha+6.09}+7.78}}{2} + 0.35, \\ \omega_2 = \frac{-\sqrt{1.96\alpha^2-3.64\alpha+6.09}-\sqrt{1.96\alpha^2-3.64\alpha-1.4\sqrt{1.96\alpha^2-3.64\alpha+6.09}+7.78}}{2} + 0.35, \end{cases} \tag{20}$$

In the case of two-dimensional piecewise smooth maps, it is possible to choose an appropriate coordinate transformation so that the choice of axis is independent of the parameter.

In so doing, the normal form of map (1) is given by [9]:

$$N(x,y)=\begin{cases}\begin{pmatrix}\tau_1 & 1\\ -\delta_1 & 0\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}+\begin{pmatrix}0\\ 1\end{pmatrix}\mu, \text{ if } x<0,\\ \\ \begin{pmatrix}\tau_2 & 1\\ -\delta_2 & 0\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}+\begin{pmatrix}0\\ 1\end{pmatrix}\mu, \text{ if } x>0,\end{cases} \tag{21}$$

where μ is a parameter, and $\tau_i, \delta_i, i=1,2$ are the traces and determinants of the corresponding matrices of the linearized map in the two subregion D_1 and D_2 evaluated at P_1 and P_2 respectively, and they are given by:

$$\begin{cases}\tau_1=0.7+\sqrt{1.96\alpha^2-7.56\alpha+6.09},\\ \tau_2=0.7-\sqrt{1.96\alpha^2-3.64\alpha+6.09},\\ \delta_1=\ \delta_2=-0.3,\end{cases} \tag{22}$$

It is shown in [10] that a robust homoclinic chaos (i.e. the existence of an infinity of homoclinic intersections between the two subregions D_1 and D_2) occurs in the piecewise smooth map of the form (21) when:

$$\begin{cases}\tau_1>1+\delta_1, \text{ and } \tau_2<-(1+\delta_2),\\ \delta_1<0, \text{ and } -1<\delta_2<0,\end{cases} \tag{23}$$

and the condition:

$$\frac{\lambda_1-1}{\tau_1-1-\delta_1}>\frac{\omega_2-1}{\tau_2-1-\delta_2}, \tag{24}$$

where the parameter range for boundary crisis is given by:

$$(\lambda_2-\tau_2)\lambda_1-\tau_1+\tau_2+\delta_1>0, \tag{25}$$

because $\delta_1=\delta_2$, where the inequality (25) determine the condition of stability of the chaotic attractor. However, if the first condition (24) is not satisfied, then the condition of existence of the chaotic attractor changes to:

$$\frac{\omega_2-1}{\tau_2-1-\delta_1}<\frac{(\tau_1-\delta_1-\lambda_2)}{(\tau_1-1-\delta_1)(\lambda_2-\tau_2)}, \tag{26}$$

because $\delta_1=\delta_2$. Finally, the formulas (18), (20), and (22), and the inequalities (23), (24), and (25), or the inequalities (23), (25), and (26) if they are satisfied, determine rigorously the region for the parameter α where the unified map (2) has robust homoclinic chaos.

6 Discussion

First, it is clear that the conditions of (23) are satisfied for all $0<\alpha<1$. Second, it is difficult to solve rigorously the conditions for existence of the chaotic attractor (24) or (26) and its condition for stability (25) since these inequalities contain complicated square formulas. Hence, we use numerical estimates of the portion of the range $0\leq\alpha<1$, for

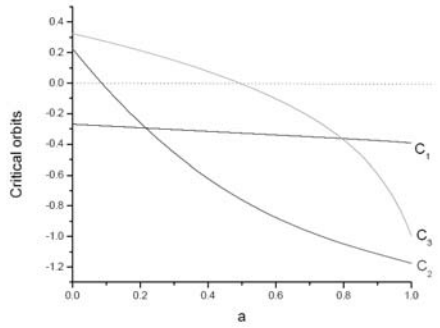


Figure 4: Critical curves corresponding to the conditions (24), (25) and (26).

which robust homoclinic chaos occurs in the unified piecewise smooth map (2). Here we exclude the value $\alpha = 0$, since there is no robust chaos in the Hénon map. We also exclude the value $\alpha = 1$, since both fixed points given in (9) are not defined for this value.

Second, let us consider the critical curves corresponding to the conditions (24), (25), and (26) as follows:

$$\begin{cases} C_1 : \frac{\lambda_1 - 1}{\tau_1 - 1 - \delta_1} - \frac{\omega_2 - 1}{\tau_2 - 1 - \delta_2} = 0, \\ C_2 : (\tau_2 - \lambda_2)\,\lambda_1 + \tau_1 - \tau_2 - \delta_1 = 0, \\ C_3 : \ \frac{\omega_2 - 1}{\tau_2 - 1 - \delta_2} - \frac{\delta_1(\tau_1 - \delta_1 - \lambda_2)}{(\tau_1 - 1 - \delta_1)(\delta_2\lambda_2 - \delta_1\tau_2)} = 0, \end{cases} \tag{27}$$

From Fig. 4 we remark that the curve (C_2) has an intersections with the axis $y = 0$, at $\alpha = 0.0866592234$, then conditions (24) holds for $\alpha \in [0, 0.0866592234]$, while the curve (C_1) does not hits the axis $y = 0$, then conditions (25) does not holds for all $0 \leq \alpha < 1$, and the curve (C_3) hits the axis $y = 0$ also one time at $\alpha = 0.493122734$, then condition (26) holds when $\alpha \in [0.493122734, 1[$, where the Newton method for finding roots of an algebraic equation was used with an error of 10^{-6}. Thus, the homoclinic chaos presented by the unified chaotic map (2) is robust not stable when $\alpha \in [0.493122734, 1[$, because the condition (25) does not hold in this interval. The chaotic attractor cannot be destroyed by small changes in the parameters, since small changes in the parameters can only cause small changes in the Lyapunov exponents. Hence, the percentage for the parameter $0 \leq \alpha < 1$, in which the map (2) converges to a robust chaotic attractor is approximately 50.6 88 percent, this result is also verified numerically by computing Laypunouv exponents and bifurcation diagram as shown in Fig. 2.

For $\alpha < 0.493122734$, the chaos is not robust in some ranges of the variable α, because there are numerous small periodic windows as shown in Figs. 5 (a), 5 (b) for example the period-8 window at $\alpha = 0.025$. Also, for $\alpha = 0.114$, there is some periodic windows. We remark, also the existence of some regions in the $\alpha-$line where the largest Lyapunouv exponent is positive, but this does not guaranty the unicity of the attractor, contrary in the case where $\alpha \in [0.493122734, 1[$, where there is guaranteed that the attractor is unique, due to the analytical expressions (23), (24), (25), and (26). When α approaches 0, there is a break of smoothness and the dynamics is too chaotic and presents some chaotic attractors very similar to the original Hénon attractor shown in Figs. 3 (a). Finally, it is interesting and surprising that the unified system (2) has such a property

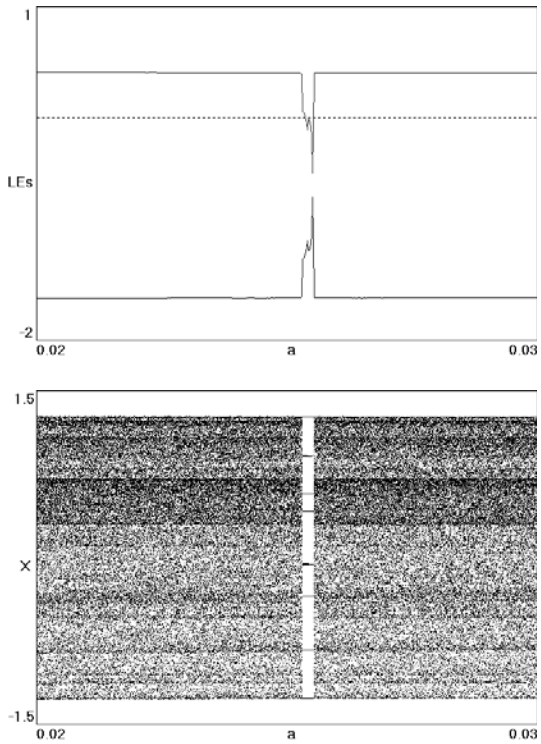


Figure 5: (a) Variation of the Lyapunov exponents of the unified chaotic map (2) for $0.02 \leq \alpha \leq 0.03$. (b) The bifurcation diagram of the unified chaotic map (2) for $0.02 \leq \alpha \leq 0.03$, showing a period-8 attractor obtained for $\alpha = 0.025$.

for an intermediate α while it was absent for $\alpha = 0$ or $\alpha = 1$ since the Hénon map is a *quasiattractor* and the Lozi map is a *Lorenz-type attractor*. These types of chaotic attractors have no robust homoclinic chaos over all portion of their key system parameters.

7 Conclusion

We have reported some results relevant to a new piecewise smooth 2-D discrete chaotic map as a unified chaotic system that contains the original Hénon and the Lozi systems as two extremes and other systems as a transition in between, and which has robust homoclinic chaos over a portion of its key system parameters, while this property is absent for the two systems at its extremes. Dynamics of piecewise continuous (smooth) mappings are a newly emerging area of research, due to the absence of continuity (smoothness), exist theories/methods in dynamical systems are not directly applicable, so new methods are needed for this important area.

References

[1] M, Hénon, "A two dimensional mapping with a strange attractor," Commun. Math. Phys. 50, 69-77, 1976.

[2] R, Lozi, "Un attracteur étrange du type attracteur de Hénon," Journal de Physique. Colloque C5, Supplément au n^0 8, 39, 9-10, 1978.

[3] F. R, Maorotto, "Chaotic behavior in the Hénon mapping," Com. Math. Phys, 68, 187-194, 1979.

[4] M, Misiurewicz, "Strange attractor for the Lozi-mapping, in Nonlinear dynamics," R.G. Heman (ed.) Annals of the New York academy of sciences 357, 348-358, 1980.

[5] M. Benedicks, and L. Carleson, "The dynamics of the Hénon maps," Ann. Math. 133, 1–25, 1991.

[6] Y, Cao, and Z, Liu, "Orientation-preserving Lozi map," Chaos, Solitons & Fractals, 9(11), 1857-1863, 1998.

[7] M. A, Aziz Alaoui, C, Robert and C, Grebogi, "Dynamics of a Hénon-Lozi map, " Chaos, Solitons & Fractals, 12(11), 2323-2341, 2001.

[8] E, Zeraoulia, "A new chaotic attractor from 2-D discrete mapping via border-collision period doubling scenario," Discrete dynamics in nature and society, Volume 2005, 235-238, 2005.

[9] S, Banerjee, and C, Grebogi, "Border collision bifurcations in two-dimensional piecewise smooth maps," Phy. Rev E, 59(4), 4052-4061, 1999.

[10] S, Banergee, J. A, York, and C. Grebogi, "Robust chaos," Phys. Rev. Lettres, 80(14), 3049-3052, 1998.

[11] J, Scheizer, and M, Hasler, "Multiple access communication using chaotic signals," -Proc. IEEE ISCAS'96, Atlanta, USA, 3,108-111, 1996.

[12] A, Abel, A, Bauer, K, Kerber, and W, Schwarz, "Chaotic codes for CDMA application," -Proc.ECCTD'97, 1, 306-314, 1997.

[13] J. M, Ottino, "The kinematics of mixing: stretching, chaos, and transport," Cambridge: Cambridge University Press, 1989.

[14] J. M, Ottino , F. J, Muzzion, M, Tjahjadi, J. G, Franjione, S. C, Jana , H. A, Kusch, "Chaos, symmetry, and self-similarity: exploring order and disorder in mixing processes," Science, 257, 754–760, 1992.

[15] M, Andrecut, and M. K, Ali, "Robust chaos in a smooth system," Inter. J. Modern Physics B, 15(2), 177-189, 2001.

[16] M, Andrecut and M. K, Ali, "On the occurrence of robust chaos in a smooth system," Modern Physics Letters B, 15(12-13), 391-395, 2001.

[17] P. E, Gabriel, "Robust chaos in polynomial unimodal maps," Int. J. Bifurcation. and Chaos, 14(7), 2431-2437, 2004.

[18] M. A, Jafarizadeh, and S, Behnia, "Hierarchy of Chaotic maps with an invariant measure and their compositions," J. Nonlinear. Math. Phy, 9(1), 26-41, 2002.