Mathematical Problems in Engineering

Theory, Methods, and Applications

Modeling Experimental Nonlinear Dynamics and Chaotic Scenarios

Modeling Experimental Nonlinear Dynamics and Chaotic Scenarios

Guest Editors: José Roberto Castilho Piqueira, Elbert E. Neher Macau, and Celso Grebogi

Contents

Editorial

Modeling Experimental Nonlinear Dynamics and Chaotic Scenarios

José Roberto Castilho Piqueira,[1] Elbert E. Neher Macau,[2] and Celso Grebogi[3]

[1] *Telecommunication and Control Engineering Department, Polytechnic School,*
   The University of São Paulo, 05508-970 São Paulo, Brazil
[2] *Laboratório Associado de Matemática Aplicada e Computação (LAC), Instituto Nacional de Pesquisas*
   Espaciais (INPE), São Josè dos Campos, 12227-010 São Paulo, Brazil
[3] *Institute for Complex Systems and Mathematical Biology, King's College, University of Aberdeen,*
   Aberdeen AB24 3UE, UK

Correspondence should be addressed to José Roberto Castilho Piqueira, piqueira@lac.usp.br

Received 25 August 2009; Accepted 25 August 2009

Thinking about nonlinearity in engineering areas, up to the 70s, was focused on intentionally built nonlinear parts in order to improve the operational characteristics of a device or system. Keying, saturation, hysteretic phenomena, and dead zones were added to existing devices increasing their behavior diversity and precision. In this context, an intrinsic nonlinearity was treated just as a linear approximation, around equilibrium points.

Approximation methods and Lyapunov stability theorems elegantly solved the dynamical behavior analysis problems and gave hints about how to synthesize systems with some desired properties. At that time, bifurcation and chaos were not present in engineers' daily life.

Chaotic dynamics was discovered at the end of the nineteenth century, by Henry Poincaré as a result from nonlinear phenomena subjected to qualitative changes in their behavior, provoked by bifurcations due to parameter variations. By studying the mathematical model of the circular restricted three-body problem, he was able to glimpse the chaotic motion that appears in complicated and apparently unpredictable trajectories that were close to periodic orbits, but spread fully in bounded regions of the phase space. Analyzing this motion, he concluded that "... it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible, and we have the fortuitous phenomenon."

In the last couple of decades of the twentieth century, the profound impact of chaos in science, engineering, and medicine was acknowledged. Fundamental experiments performed

in all areas of science allowed us to finally understand a realistic scenario in which the chaotic behavior not only occurs in Nature, but it is also a fundamental mechanism that Nature accounts for to mediate its phenomena. Thus, the chaotic behavior can be observed, for example, in the electrical activity from biological systems, in the transition of a fluid to turbulent motion, and in the motion of the moons of the giant planets.

Inspired on the rediscovering of the richness of nonlinear and chaotic phenomena, engineers started using analytical tools from "Qualitative Theory of Differential Equations," allowing more precise analysis and synthesis, in order to produce new vital products and services. Bifurcation Theory, Dynamical Systems, and Chaos started to be part of the mandatory set of tools for design engineers.

Consequently, an important interaction between Mathematics and Engineering started to gain importance and, in this scenario, the blending of three distinct methodological approaches, analytical, numerical, and experimental needs to be improved as follows.

(i) Numerical experiments that cleverly and precisely use tailored computer-based numerical simulations to give insights into problems that are analytically intractable so far.

(ii) Analytical methods to deal with nonlinear differential equations and to geometrically describe structures that arise as a consequences of the chaotic dynamics.

(iii) High precision experiments, which involve up-to-date measurement technology and procedures to capture the chaotic behavior in specific fields.

This special edition of the *Mathematical Problems in Engineering* aims to provide a picture of the importance of the Bifurcation Theory, relating it with nonlinear and chaotic dynamics for natural and engineered systems. Ideas of how this dynamics can be captured through precisely tailored real and numerical experiments and understanding by the combination of specific tools that associate dynamical system theory and geometric tools in a very clever, sophisticated, and at the same time simple and unique analytical environment are the subject of this issue, allowing new methods to design high precision devices and equipment.

Following these lines, this special issue contains 25 papers organized as follows. Concerning an overview of modeling nonlinear dynamics, there is one paper entitled "Modeling nonlinear dynamics and chaos: A review" by L. A. Aguirre and C. Letellier. Concerning an overview of the application of bifurcation theory to turbulence, there is one paper entitled "Flow around a slender circular cylinder: A case study on distributed Hopf bifurcation" by Aranha et al. Concerning applications of bifurcation theory to nonlinear mathematical and computation models, there are four papers entitled "Analysis of nonlinear dynamics for abrupt change of interphase structure in liquid-liquid mass transfer" by Zhang et al.; "Continuation of periodic solutions of dissipative and conservative systems— application to elastic pendulum" by Pavel Pokorny; "Dynamical analysis of an interleaved single inductor TITO switching regulator" by El Aroudi et al.; "Higher period stochastic bifurcation of nonlinear airfoil fluid-structure interaction" by J. A. S. Witteveen and H. Bijl. Concerning computational and numerical techniques of analysis of models, there are four papers entitled "The effect of spatial scale on predicting time series: A study on epidemiological system identification" by Monteiro et al.; "Macro- and micro-simulations for a sublimation growth of SiC single crystals" by J. Geiser and S. Irle; "Free vibration analysis of rectangular orthotropic membranes in large deflection" ZHENG et al.; "Modified Jacobian Newton iterative method: Theory and applications" by Juergen Geiser. Concerning

data analysis, there are three papers entitled "The staircase structure of the southern Brazilian continental shelf" by M. Baptista and L. A. Conti; "Solitons, peakons, and periodic cuspons of a generalized Degasperis-Procesi equation" by J. Zhou and L. Tian; "Chaotic patterns in aeroelastic signals" by F. Marques and R. M. G. Vasconcellos. Concerning synchronization and control, there are eight papers named "Limit cycle prediction based on evolutionary multi objective formulation" by Katebi et al.; "Intermittent behavior and synchronization of two coupled noisy driven oscillators" by Santos et al.; "Chaos synchronization between two different fractional systems of Lorenz family" by Ahmed Ezzat Matouk; "Diffusive synchronization of hyperchaotic Lorenz systems" by Ruy Barboza; "Synchronization of discrete-time chaotic systems in bandlimited channels" by Eisencraft et al.; "Trajectory sensitivity method and master-slave synchronization to estimate parameters of nonlinear systems" by Cari et al.; "Adaptive step-size control in simulation of diffusive CVD processes" by J. Geiser and C. Fleck; "A new mechanical model for particle transport by surface waves and applications" by Ragulskis et al. Concerning trans-disciplinary applications, there are four papers entitled "Nonlinear dynamics and chaos in a fractional-order HIV model" by H. Ye and Y. Ding; "Chaotic Image encryption design using Tompkins-Paige algorithm" by S. E. Borujeni and M. Eshghi; "An LPV fractional model for canal control" by Martinez-González et al.; "Flux reflection model of the ferroresonant circuit" by Milicevic et al.

As guest editors for this special issue, we wish to thank all those who submitted manuscripts for consideration. We also would like to thank the many individuals who served as referees. We hope that these articles can motivate and foster further scientific works that will allow for a continuous and better understand of the role of the chaotic dynamics in our world.

José Roberto Castilho Piqueira
Elbert E. Neher Macau
Celso Grebogi

Review Article

Modeling Nonlinear Dynamics and Chaos: A Review

Luis A. Aguirre[1] and Christophe Letellier[2]

[1] *Departamento de Engenharia Eletrônica da UFMG, Universidade Federal de Minas Gerais,*
  Avenida Antônio Carlos 6627, 31270-901 Belo Horizonte, MG, Brazil
[2] *CORIA UMR 6614, Université de Rouen, Avenue de l'Université, BP 12,*
  76801 Saint-Etienne du Rouvray Cedex, France

Correspondence should be addressed to Luis A. Aguirre, aguirre@cpdee.ufmg.br

This paper reviews the major developments of modeling techniques applied to nonlinear dynamics and chaos. Model representations, parameter estimation techniques, data requirements, and model validation are some of the key topics that are covered in this paper, which surveys slightly over two decades since the pioneering papers on the subject appeared in the literature.

1. Introduction

The field of nonlinear dynamics experienced a very quick and intense development in the last thirty years or so. To determine the starting point of any subject is very difficult simply because we all stand on somebody elses showlders in any "new" attempt we make. For the sake of presentation, and because we feel that this understanding is not absurd anyway, the origins of what we nowadays call the field of nonlinear dynamics can be traced back to the work of Henri Poincaré.

In his studies on nonlinear dynamical systems Poincaré figured out that, since no analytical solution to most of nonlinear systems can be obtained, the whole set of solutions can be investigated in the so-called phase space spanned by the set of variables required for a complete description of states of the system [1]. A few years later, while investigating the three-body problem, he observed that small perturbations can deeply affect the solution [2]. In order to investigate nonlinear dynamics, Poincaré introduced the concepts of phase portrait, Poincaré section, periodic orbit, return map, bifurcation, fixed point, and so on. Most of these concepts were used by Andronov's school in the 20 s [3] but the first representation in the phase space of a now called "chaotic solution" was due to the solution Edward Norton Lorenz [4]. The so-called Lorenz attractor thus represents the first chaotic attractor ever drawn.
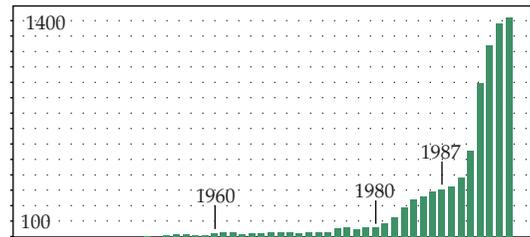
Figure 1: Number of papers published per year up to 1994. The search machine used was ISI Web of Science using Topic = ("nonlinear dynamics" OR chaos). Up to 1994 the total number of entries found was 8 400 (shown in this figure). Up to 2008 the total number of entries is 39 180 and only in 2008 over 2 700 papers were published within the topics nonlinear dynamics or chaos.

From the publication of Lorenz's paper up to the mid of the 70 s not many papers were published. Among those who contributed to the emergence of "chaos theory," we can quote Ruelle and Takens paper on turbulence [5], May about bifurcation [6], Rössler's new chaotic attractors [7], Li and Yorke's theorem for existence of chaos in map [8]. These contributions, published before 1980, popularized the word "chaos" as well as related techniques to investigate these new types of solutions.

An important turning point happened around 1980 (Figure 1). There can be little doubt that such a turning point was provoked, by the papers [9, 10]. The reason for this is quite simple. So far, in the investigation of a nonlinear system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, with $\mathbf{f} : \mathbb{R}^m \mapsto \mathbb{R}^m$, typically it was assumed that the entire state vector $\mathbf{x}$ was available. This constraint was clearly too restrictive in practical problems and consequently it was unclear what kind of role could the theory of nonlinear dynamics play in *real world problems*. From a suggestion by Ruelle, Packard and colleagues showed that it was possible to build a phase portrait for the original system using a scalar $s(t) = h(\mathbf{x})$, with $h : \mathbb{R}^m \mapsto \mathbb{R}$ [9]. Furthermore, Takens and Mañé mathematically proved under which conditions the phase space reconstructed using $s(t)$ was diffeomorphically equivalent to the one that would be obtained if the entire state vector $\mathbf{x}$ was available [10, 11]. A decade later, other important works on the subject were published [12, 13]. When such works came to light, hopes of using the "new theory of nonlinear dynamics" in practical problems were kindled, because now all that was required, in principle, and assuming an ideal measurement function was the recording of a single variable.

When, in the early 1980, it became accepted that a scalar-reconstructed attractor could be equivalent to the original one, a great deal of work was devoted to developing tools to quantitatively characterize chaotic attractors. In the following decade, geometric measures such as dimensions [14, 15], Lyapunov exponents [16] and entropies [17] were adapted, applied, and intensively investigated in the context of nonlinear dynamics. By the end of the 1980, there was a large set of works on these subjects as can be found in [18, 19].

Another turning point which can be detected in the number of papers published in the field of nonlinear dynamics and chaos happened in the end of the decade 1980 (Figure 1). Unlike the turning point in the beginning of the decade, which quite clearly can be related to fundamental papers on embedding theory, it is quite hard to relate the turning point in the late 1980 to a specific topic within the field. Most likely, such a turning point was an indication that several important topics started to be investigated as well as many applications started in various fields. Among such topics we are to find the *modeling of nonlinear dynamics and chaos*. To survey the main developments of this topic is the aim of the present paper.

2. Nonlinear System Identification

Before actually addressing how the modeling of nonlinear dynamics and chaos developed within the field, it is important to say that a significant amount of work was developed independently, and sometimes previously, in the field of nonlinear system identification. The basic goals were very similar to those that later were pursued by the nonlinear dynamics community but there were some important differences. Nonlinear system identification, which has its origins in the field of engineering, is usually concerned with nonautonomous systems, discrete-time models, disturbance modeling (this is vital to correctly deal with noisy data), and hardly ever was concerned with chaos. On the other hand, modeling nonlinear dynamics and chaos, with its origins in physics and applied mathematics, usually concerned with autonomous systems, very often considers continuous-time models, does not typically model disturbances, and is strongly focused on chaotic systems.

Having established such main distinctions between the two fields, let us point out that this paper is a survey of modeling techniques applied to nonlinear dynamics and chaos. Many such techniques have indeed been developed in the field of nonlinear system identification and will be mentioned in this survey only to the extent in which they were applied to modeling nonlinear dynamics and chaos. In the reminder of this section, for the sake of completion, we present some issues in nonlinear system identification which will be relevant in the discussion of modeling techniques applied to nonlinear dynamic and chaos. In [20] the authors also point out some conceptual differences between what they call stochastic and deterministic modeling.

By the end of the 1970 linear system identfication was well established. The typical problem was to build a model, usually a transfer function of the form

$$y(k) = \frac{B(q)}{A(q)} u(k),\tag{2.1}$$

from a set of possibly noisy data $y(k)$, $u(k)$, $k = 1, 2, \ldots, N$. In (2.1) $A(q)$ and $B(q)$ are polynomials in the backward shift operator q^{-1}, that is, $y(k-i) = q^{-i}y(k)$. A commonly used representation of (2.1) is the ARX (autoregressive model with exogeneous inputs) model

$$y(k) = a_1 y(k-1) + \cdots + a_{n_y} y(k - n_y) + b_1 u(k-1) + \cdots + b_{n_u} u(k - n_u),\tag{2.2}$$

where n_y and n_u are the maximum lags used. Noise is usually assumed to appear in the output $y(k)$. The presence and the type of noise usually are such that the standard least squares estimator will produce wrong results. This calls for disturbance modeling and some type of generalized least squares algorithm, which is no longer a linear estimator. The simples disturbance model which is the moving average (MA), which added to (2.2), yields the well-known ARMAX model. More general disturbance and system models have been proposed and studied in detail in standard texts such as [21, 22].

The approach to nonlinear system identification happened along different lines of action. One such line of action was to build so-called block-oriented models which consisted of a linear transfer function model (see (2.1)) either after or before a nonlinear *static* funtion $g(\cdot)$, such were the Hammerstein and Wiener models. Another line of action was the development of nonparametric models like the Volterra models [23]. The numerical algorithms involved in the development of such models were, in one way or another, related

to the correlation function. A glimps of the state-of-the-art in the development of block-oriented and Volterra models at the end of 1970 can be found in [24].

A third, and last one to be mentioned here, line of action was to rewrite (2.2) as

$$y(k) = f\left(y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, u(k-n_u)\right), \tag{2.3}$$

and to generalize for the case in which $f(\cdot)$ was a nonlinear function. Because the regressor variables in (2.3) were of the ARX type, and because $f(\cdot)$ was now allowed to be nonlinear, (2.3) was referred to as a NARX model. As before, in the case of noisy data, disturbance modeling is a must and, therefore, additional regressor terms of the MA (moving average) type have to be included in (2.3). This lead to the now well-known acronym NARMAX. The building of a NARMAX model is typically composed of three steps: (1) the choice of the class of models to be used for approximating $f(\cdot)$. Some model classes include: polynomial models [25], rational models [26], neural models [27], radial basis function models [28], and wavelet networks [29]. (2) The following step is to determine which regressor variables to use in (2.3), this is equivalent, although it is more general, to the choice of the embedding space in modeling nonlinear dynamics and chaos. For recent discussions and a survey on such techniques we refer the reader to [30–35]. (3) In model building the final step is to estimate whatever parameters $f(\cdot)$ may have. This step is far the easiest. Standard and robust techniques for accomplishing parameter are available [21, 22]. The important problem of model validation, which will be addressed later on, has more to do with model testing than model building.

3. Model Building for Nonlinear Dynamics and Chaos

After presenting a brief introduction (Section 1) and pointing out some standard problems in nonlinear system identification (Section 2), we are ready to start surveying the development of data-driven models for nonlinear dynamics and chaos. There are various possible ways of addressing this vast subject, and it is not clear which will turn out to be the most pedagogical one. We choose to start covering five of the earliest papers in the field and follow by discussing some of the mostly used model classes.

3.1. A Few Pioneering Works

In what follows, we begin by mentioning explicitly five pioneering works, the first four of which were of great influence; [36–40]. The reason for including [40] in this group is its early date. In the remainder of the section we will survey some of the major developments in the field. Together, these five papers have been quoted 2 719 times from their publication up to the end of 2008. (The search machine used was ISI Web of Science.)

3.1.1. The Local Linear Predictor

The first papers on modeling nonlinear dynamics and chaos seem to have appeared in 1987. One of them is related to local linear modeling, the rest is generally concerned with some aspect of global modeling. Because of its historical importance, in this subsection, the main points concerning the local linear predictor [36, 41] will be reviewed.

Suppose that a measured time series $y(1), y(2), \ldots, y(N)$ lies on a D-dimensional attractor of an n th-order deterministic dynamical system. The starting point obtains an embedding from the recorded data. A convenient, though not unique, representation is achieved by using *delay coordinates*, for which a delay vector has the following form:

$$\mathbf{y}(k) = \left[y(k) \;\; y(k - \tau) \;\; \cdots \;\; y(k - (d_e - 1)\tau) \right]^{\mathrm{T}}, \tag{3.1}$$

where d_e is the *embedding dimension* and τ is the *delay time*. Takens has shown that embeddings with $d_e > 2n$ will be faithful generically so that there is a smooth map $f : \mathbb{R}^{d_e} \mapsto \mathbb{R}$ such that

$$y(k + 1) = f(\mathbf{y}(k)) \tag{3.2}$$

for all integers k, and where the forecasting time T and τ are also assumed to be integers. Takens' theorem gives no indication as to how find. One of the first attempts to to build the map f from the data was the local linear predictor method, which started by partitioning the embedding space into neighbourhoods $\{\mathcal{U}_i\}_{i=1}^{N_n}$ within which the dynamics can be appropriately described by a linear map $g : \mathbb{R}^{d_e} \mapsto \mathbb{R}$ such that

$$y(k + 1) \approx g_i(\mathbf{y}(k)) \quad \text{for } \mathbf{y}(k) \in \mathcal{U}_i, \; i = 1, \ldots, N_n. \tag{3.3}$$

Several choices for g have been suggested in the literature such as linear polynomials [36, 42] which can be interpolated to obtain an approximation of the map f [43]. Simpler choices include *zeroth-order approximations*, also known as *local constant predictors* [36, 44], and a *weighted predictor* [45]. Linsay proposed to weigh the predictions according to information obtained from the local data [45]. In [46] the authors put forward the concept of *pseudofalse neighbors* which are left out from the construction of a local linear predictor to improve performance, and in [47] the authors consider the multivariate case.

A common difficulty of such approaches is that the data have to be separated into neighbourhoods. Thus given a point in the embedded space the closest neighbours to such a point must be found. It is well-known that for many methods most of the CPU time is spent in searching for close neighbours in the embedding space within the data [48] and that the effort required to accomplish this grows exponentially with the embedding dimension.

Another drawback has to do with model representation. The local linear predictor does not have a closed form, in other words, it is not a global model. This prevents using the obtained model in any kind of analytical investigation.

The local linear predictor has been applied to several real-time series including Wolf's sunspot numbers [20, 49], daily dollar exchange rates [50], R-R heart intervals [49], to mention a few.

3.1.2. Equations of Motion

In 1987 James Crutchfield and Bruce McNamara published a paper that became a reference in the field [37]. The title of this section is a reference to that paper rather to a particular class of models. It seems safe to say that this was the first journal paper in the field devoted to global modeling, as oposed to the "patchwork-models" produced by the local linear approach, and which the authors call the "atlas equation of motion procedure."

In [37] the authors consider linear-in-the-parameter models of the type $\dot{\vec{x}} = \vec{F}(\vec{x})$, with $\vec{F}(\vec{x}) = A^{\mathrm{T}}\vec{\phi}(\vec{x})$, where $\vec{\phi}$ is a vector of function basis and A is a vector of parameters to be estimated from data. Discrete-time models are also considered in the paper. The authors described their method without any particular class of function basis in mind. The parameter vector was estimated using a singular value decomposition implementation of least squares. Their method is general and quite basic: the embedding dimension and the number of basis functions used are varied over a set of alternatives. The choosen model is the one with smallest entropy $I(M)$. In their examples, that include the Hénon map, Duffing, and van der Pol oscillators, the function basis used was B-splines and discrete monomials.

In closing, it should be said that more than a method, Crutchfield and McNamara put forward a general philosophy for modeling chaotic data. With the exception of some issues that are specific to such data, their procedure does not differ substantially from works in the field of nonlinear system identification using linear-in-the-parameter models widely available at that time (see [34] for a survey).

3.1.3. Multivariable Functional Interpolation

The paper by David Broomhead and David Lowe is among the four pioneering works that we have selected to start surveying the main developments in modeling of nonlinear dynamics and chaos [38]. Such a paper is often quoted as being responsible for the proposition of using Radial Basis Function (RBF) models to describe nonlinear systems. Although in their paper the authors quote a review work by M. J. D. Powell entitled "Radial basis functions for multivariable interpolation: a review," it seems that the work by Powell did not consider dynamical systems but only static problems.

A typical model, as proposed by Broomhead and Lowe, can be written as

$$y(k) = b_0 + \sum_{i=1}^{p} \omega_i \phi \left(\|\mathbf{x}(k-1) - \mathbf{c}\| \right), \tag{3.4}$$

where p is the number of radial basis functions, ϕ, b_0, and w_i are the unknown parameters to be estimated, $\mathbf{c}$ are the centers, and $\|\cdot\|$ is usually the Euclidean norm. As it will be seen in the next section, Martin Casdagli also worked with RBF models, as pointed out by Broomwhead and Lowe: "We note that in this application our approach is very close to that of Casdagli who has applied radial basis functions to the construction of nonlinear maps from time series data. (Here the authors quote Casdagli's paper as "submitted to *Physica* D." That paper, that will be mentioned in the next section, was eventually published in 1989.) Unlike Casdagli who used strict interpolation, we will employ the least squares generalization." By "strict interpolation" the authors mean that the number of unknowns (parameters) and constraints (built from data) is the same (more on this in Section 3.1.4).

This paper is curious in some aspects. First of all, it is clearly devoted to machine learning for purposes of classification. Several pages are devoted to the problem of approximating the exclusive-OR function. The last examples in the paper seem to be an afterthought and consist of the modeling of the doubling map $x_{n+1} = 2x_n$ (modulo 1) and of the quadratic map $x_{n+1} = 4x_n(1 - x_n)$. Also, only three works are listed in the references which are related to modeling nonlinear dynamics and chaos: [36, 39] and a preprint by Lapedes and Farber [51]. (Such a manuscript was at the time submitted to Proceedings of the IEEE. It seems that this paper was never published as such, but has circulated as a technical

report. This report is cited by a number of early papers on modeling nonlinear dynamics and chaos, proving to have been quite influential.) It is also very curious that no paper or result on embedding theory was used by Broomhead and Lowe. Because both their examples use 1D maps, there was no need for any serious embedding. In summary, the contribution of [38] seems to be the suggestion of using RBF models for nonlinear dynamics, but for the general view of the embedding-modeling problem we should look into [37] as the pioneering work. The fusion of such papers, that is, the use of RBF models in an embedding environment to model and predict nonlinear dynamics, happened in the next paper to be considered.

3.1.4. Prediction of Chaotic Time Series

From the abstract of the paper we clearly read the objective of the paper: "numerical techniques are presented for constructing nonlinear predictive models directly from time series data" [39]. The basic motivation for such was to "convincingly distinguish low-dimensional chaos from randomness." In fact, Casdagli argues in favor of his model-based approach when comparing it with the use of dimension calculations and Lyapunov exponents to distinguish noise from chaos.

With respect to the model class, this paper is remarkable because it compared polynomial, rational, and RBF models, plus local linear predictions. Very few papers present results for more than one model class, a more recent exception to this rule is [52].

As for the use of "strict interpolation" (Broomhead and Lowe) or simply "inter-polants" (Casdagli) or "approximants" (Casdagli), the situation is not completely clear. The paper mentions both. When describing general techniques (polynomial and rational models), Casdagli mentions the least squares solution. However, when presenting the RBF model the only case considered is that of choosing the model structure so as to have an inverse problem with unique solution. At the end of that section, the least squares solution is mentioned again, but as "ongoing research." One of the critical problems of the "strict interpolation" approach is the total lack of robustness to noise. In fact, Casdagli does not consider any noise in the "global modeling" examples, but only in the "distinguishing noise from chaos" examples. In two occasions interesting remarks are made. In discussing modeling difficulties Casdagli says: "this does not appear to be due to numerical errors in the least-squares fitting and matrix inversion algorithms" [39, page 343]. This gives the impression that he was actually implementing both: the least-squares algorithm (to fit approximants) and the matrix inversion algorithm (to find the solution to the strict interpolation problem). The first was probably related to polynomial and rational models and the latter to RBF models. The second case, when discussing the use of RBF models for predicting invariant measures, the number of data points is quoted as $N = 25$ and the data were noise-free. This clearly suggests "strict interpolation."

The RBF models considered by Casdagli were of the form (compare with (3.4))

$$y(k) = \sum_{i=1}^{N-1} \omega_i \phi\left(\|\mathbf{x}(k-1) - \mathbf{c}\|\right) + \sum_{n=1}^{\hat{d}} \mu_n p_n(\mathbf{x}), \qquad (3.5)$$

where N is the number of data iterates, p_n is a basis of the space of polynomials of degree at most d from $\mathbb{R}^n$ to $\mathbb{R}$, with d known, and μ_n are parameters. The author points out that "frequently the polynomial term is not included," which is interesting in the light of other results, to be mentioned in Section 3.2.3.

Casdagli investigated the accuracy of short-term predictions based on model iteration and the *direct* method, which amounts to fitting a predictor with a single prediction horizon equivalent to the various one-step-ahead predictions of the *iterative* method. In the various cases investigated, the iterative method proved superior. A similar aim was pursued in the context of neural networks in [53], in the context of Volterra expansions in [54] and using a kernel-based approach in [43]. Finally, the author showed that the fitted RBF models were capable of reproducing some invariants, such as Poincaré sections and bifurcation diagrams. Although he only considered the noise-free case, a very significant step towards serious model validation was given. Perhaps the greatest lack in this relevant paper is that there is no mention on structure selection. Model instability under iteration is prematurely attributed to the model class and not to an unsuitable model structure.

By model class we mean the *type* of models used, as for instance polynomials, RBFs or neural networks. By model structure we mean the model *topology*, of a particular class. For instance, in the case of RBF models, the model structure is determined by the type and number of basis functions used, the lags used to compose the input space, and so on.

3.1.5. Construction of Differential Equations

The early paper by Cremers and Hübler had the clear objective to "obtain a concise description of an observed chaotic time sequence" [40]. From the beginning, the authors were clearly concerned with chaotic data. The model class they considered was a basis of Legendre polynomials, and in the introduction of the paper it was stated that the true objective was to estimate the parameters of a differential equation. This more restricted aim was confirmed later in [55].

The authors discuss a few aspects of embedding and the effect of dynamical noise. In their numerical examples no noise is considered. They mention sucessful estimation of the parameters for two systems: Lorenz and an autonomous version of the van der Pol oscillator, for which the following 4th-order approximation was used:

$$\dot{x}_j = \sum_{l,m=0}^{l+m<4} c_{j,l,m} P^l(x_1) P^m(x_2), \quad j = 1, 2, \tag{3.6}$$

where $c_{j,l,m}$ are the parameters to be estimated. The Legendre polynomials $P^l(x_1)$ and $P^m(x_2)$ must be evaluated at x_1 and x_2, in addition to the time derivatives of x_1 and x_2 at each point in state space considered. The derivatives were estimated numerically, but there is no clear indication as to how this was accomplished nor what would be the effect on such estimates of measurement noise (which they considered not). The authors justify the use of a polynomial expansion by stating "if nothing is known about the properties of the flow vector field, a fit by a polynom (sic) series is frequently favourable" [40, page 800]. This last remark results from a general theorem by Weierstrass stating that any analytical function can be approximated by an infinite polynomial expansion.

Breeden and Hübler later remarked that the 1987 paper also considered the discrete-time case, which is definitely not obvious. Then, so far the field of nonlinear dynamics is concerned, this paper seems to mark the beginning of building differential polynomial equations from data. Although no mention on important practical issues is made in this paper, such as structure selection, derivative and parameter estimation, it has the merit of having opened the way for a prosperous modeling class of techniques, to be surveyed in Section 3.2.1.

The citation relationships among these five pioneering papers are illustrated in Figure 2. In the center of the diagram the embedding papers [9, 10] appear as a common feature, with the exception of the paper [38] (see discussion in Section 3.1.3). This quotation diagram suggests that Cremers and Hübler were not aware of the other papers on model building. (Their paper was submitted in September 1986.) It is interesting to notice (not shown in the diagram) that Broomhead & Lowe and Casdagli mention the work of Lapedes and Farber (see [51]).

3.2. Model Classes

In the description to follow, the order of the model classes mentioned is somewhat arbitrary. We follow a rough chronological order of the first papers in each class.

3.2.1. Continuous-Time Polynomials

A number of papers appeared in the early nineties which had as a common goal fitting ordinary differential equations (ODEs) to observed data. As pointed out in Section 3.1.5, the paper [40] was pioneering in this field, although it did not deal with any practical issue. Unlike Cremers and Hübler, most papers that followed, instead of using a basis of Legendre polynomials, used "Taylor-series expansions." As an example, such an expansion for a 2D system is [56]

$$\dot{x} = f(x, y) = a_{00} + a_{10}x + a_{01}y + a_{11}xy + \cdots + a_{ij}x^i y^j + \cdots,$$

$$\dot{y} = g(x, y) = b_{00} + b_{10}y + b_{01}x + b_{11}xy + \cdots + b_{ij}x^j y^i + \cdots. \tag{3.7}$$

A similar representation was used by Gouesbet who started investigating under which conditions it was possible to reconstruct the vector field (X, Y, Z) from a single measurement $X = x$ of an original vector field (x, y, z) [57, 58]. The reconstructed system was expressed as

$$\dot{X} = Y,$$

$$\dot{Y} = Z, \tag{3.8}$$

$$\dot{Z} = F(X, Y, Z),$$

and both polynomial and rational expansions for F were investigated. The polynomial expansions of F used were equivalent to (3.7). Although no reconstruction from data was performed in that paper, the challenge was acknowledged and pursued shortly after [59]. In the last paper, the important issue of estimating the derivatives based on finite-difference schemes from data was considered. This was an important step towards dealing with real data but a key hurdle still had to be transposed: measurement noise. To do this was one of the key objectives in [60], where this global modeling technique finally reached maturity. As pointed out in an early paper "noise removal $(\cdots)$ is required because standard systems $(\cdots)$ rely on derivative evaluations" [58, page 1795], to deal with noise was to *remove* it from the data, not so much as for discrete-time modeling techniques which are made to cope with noise.

Other polynomial expansions are to be found in the literature. Giona and colleagues investigated the use of a basis of polynomials to approximate both continuous-time and
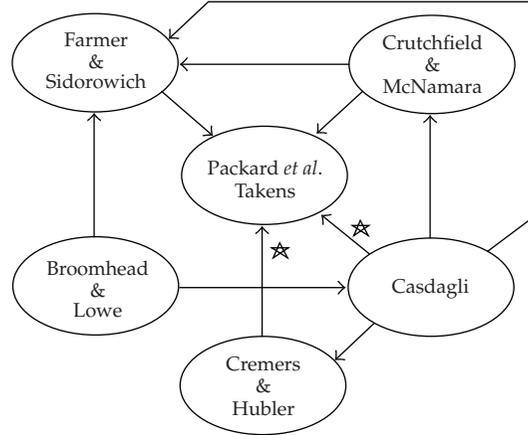
Figure 2: Quotation diagram. A → B indicates that A quotes B. The ⋆ indicates that only paper [10] was quoted. Also, Crutchfield and McNamara quoted Farmer and Sidorowich's paper as the technical report: Preprint LA-UR-87-1502.

discrete-time dynamical systems [61]. In both cases the system was approximated by a linear combination of such polynomial functions which are obtained "from the knowledge of the hierarchy of moments." The procedure is quite involved, and long data sets are required to estimate the parameters. In their examples, $N = 5 \times 10^4$ was used for the Hénon and Ikeda maps, and $N = 4 \times 10^5$ was used for the Lorenz system.

Practical implementation of global modelling using continuous-time model requires an integration scheme. For instance, the explicit Euler integration scheme

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \mathbf{F}\left(\mathbf{y}_n\right) \delta t, \tag{3.9}$$

where $\mathbf{y}_n = \mathbf{y}(n\delta t)$ with the time step may be used. It is known that the Euler is not very robust against time step change. This is why a Runge-Kutta scheme is most often used. Another alternative was proposed by [62]. They used an Adams-predictor-corrector scheme according to

$$\mathbf{y}_{n+1} = \mathbf{y}_n + \delta t \sum_{j=0}^{M} a_j^{(M)} \mathbf{F}\left(\mathbf{y}_{n+1-j}\right), \tag{3.10}$$

where $a_j^{(M)}$ are the implicit Adams predictor-corrector coefficients [62]. M designates the order of the corrector portion of the integration. The global model $\mathbf{F}$ has a polynomial form and is optimized with the help of a minimum description length criterion and the error function

$$\chi^2 = \frac{1}{2N\sigma^2} \sum_{n=1}^{N} \left[\mathbf{y}_{n+1} - \mathbf{y}_n - \delta t \sum_{j=0}^{M} a_j^{(M)} \sum_{i=0}^{N_p} K_i \phi_i(\mathbf{y}_{n+1-j}) \right]^2 \tag{3.11}$$

which is quadratic with respect to cœfficients K_i. This is thus a least square problem.

Structure selection issues for continuous-time polynomials have been discussed in [63, 64]. In [65] the authors use a priori knowledge of the driving force to help determining the model structure. When differential embeddings are used, the global model (3.8) is usually searched using a truncated polynomial expansion involving successive derivatives of the measured time series. In order to reduce the number of monomials, that is, the length of the model, it is possible to use a rational function whose monomials are selected according to an Ansatz library. Such a library contains canonical forms which can be mapped into an equivalent polynomial system whose variables are combinations of the measured time series and its successive derivatives. The canonical function F thus only contains selected terms. This reduced function allows to avoid numerical instabilities usually encountered with rational function. This procedure was introduced in [66] and improved in [67]. For instance a very accurate 7-term global model was obtained from the x-variable of the Lorenz system. A 26-term model was obtained from a copper electrodissolution experiments [68]. This model has to be compared to the 52-term model previously obtained in [69].

Other approximations for functions f and g in (3.7) include Volterra series expansions [70]. Examples of building global continuous-time models from real data can be found in a number of papers [69, 71–74].

3.2.2. Neural Networks

A neural network is a model class that resembles some aspects of a brain. Conventional simplifications made for perceptron models are: (i) to take only one hidden layer of nodes, (ii) to consider the output node linear, and (iii) to consider all the activation functions h of the hidden layer nonlinear are the same. A perceptron model with such features is illustrated in Figure 3(a) and can be described mathematically as

$$y(k) = b_o + \sum_{j=1}^{m} w_j^o h\left(b_j + \sum_{i=1}^{n} w_{ji}^h x_i(k-1)\right), \tag{3.12}$$

where w_{ji}^h indicates a weight (to be estimated) of the hidden layer that connects the ith input to the jth neuron of the hidden layer. w_j^o is the weight (to be estimated) of the jth hidden neuron output, b's are constants, called bias parameters, and the neuron *activation function* is h. Finally, $n = \dim(\mathbf{x})$ and m is the number of neurons in the hidden layer. The function shown in the right-hand side of (3.12) is often called *feed-forward* because there are no feedback loops *internal* to the network. It is important to notice that (3.12) is in the form $y(k) = f[\mathbf{x}(k-1)]$. Common choices for nonlinear activation functions are Gaussian, sigmoidal, and the hyperbolic tangent. In [75] they discuss the use of polynomials as activation functions for neural networks. An accessible introduction on the subject of neural networks applied to modeling chaotic systems is given in [76].

The first papers to build this kind of models from chaotic data seems to have been the references [78, 79]. (An earlier paper built and characterized a chaotic neural model with a single neuron but the network was not trained from data [80].) More recent and impressive results have been discussed in [81].

A rare study of bifurcation diagrams achieved by neural models has been presented in [82]. And an early work which reports on the improved performance of pruned networks is [83]. Neural networks have been applied in a number of instances, some examples include [84–86].

(a)

(b)

(c)
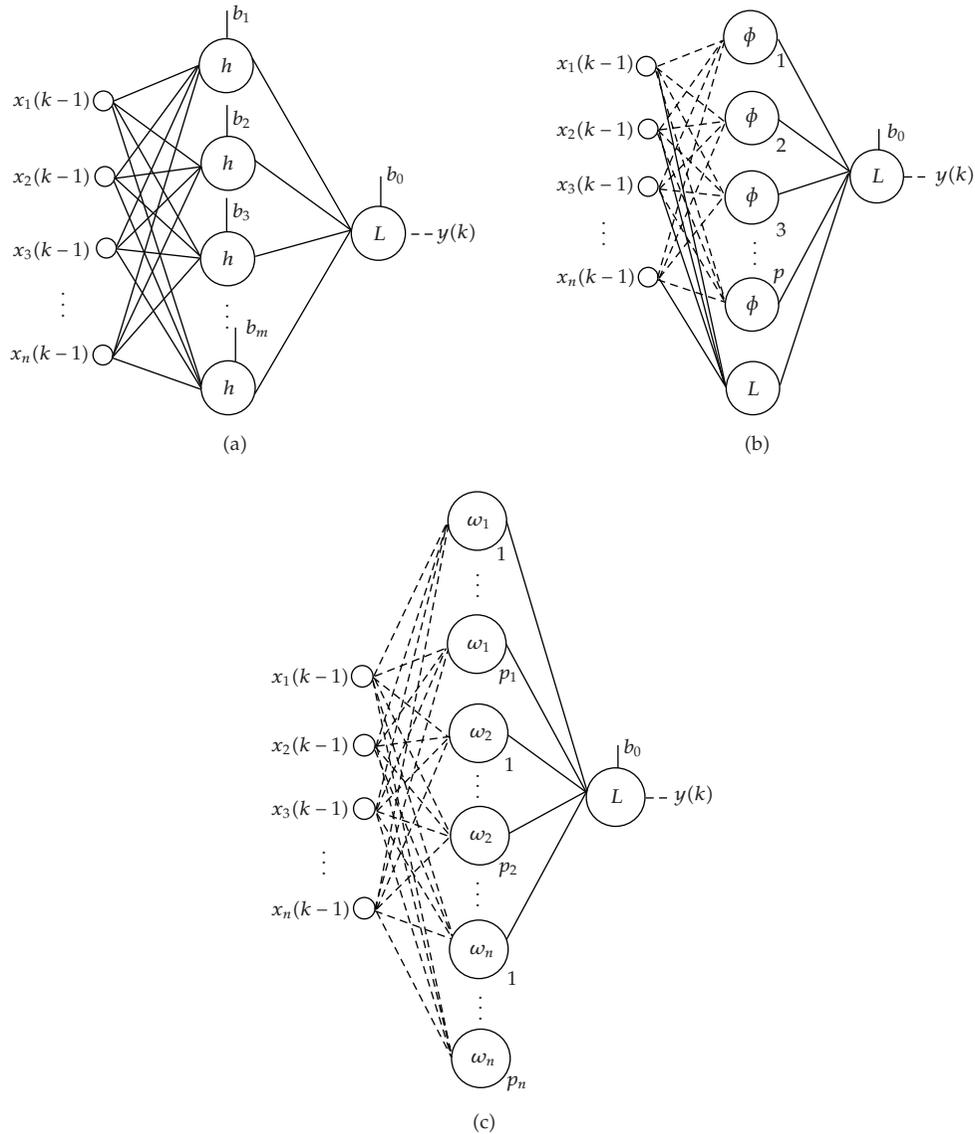
Figure 3: Schematic representation of some general model classes. (a) Typical perceptron model, (b) single-type basis function models, (c) multitype basis function models. The solid lines correspond to parameters that must be estimated. Dashed lines indicate absence of parameters. Therefore (a) is nonlinear in the parameters whilst (b) and (c) are linear in the parameters. Usually Φ depends on some parameters. When this is the case, such parameters are chosen beforehand. Figure reproduced from [77].

3.2.3. Radial Basis Function Models

The RBF model class is shown in Figure 3(b). One important difference with respect to the first class is that now there are no weights associated to the connections between the inputs and the nodes in the hidden layer. The ϕ—which is usually chosen as a radial function—is nonlinear and often depends on certain tuning parameters which are usually known when the weights associated to the connections indicated by solid lines (Figure 3(b)) are to be

estimated. As a consequence, the model is linear in the parameters and can be written thus (compare with (3.4) and (3.5))

$$y(k) = b_0 + \sum_{i}^{p} \omega_i \phi(\mathbf{x}(k-1)) + \sum_{i=1}^{n} a_i x_i (k-i),$$

(3.13)

where p is the number of radial basis functions. b_0, ω_i, and a_i are the unknown parameters to be estimated. Two of the first works to use this model class were mentioned in Sections 3.1.3 and 3.1.4 above. Work using this type of model has been recently surveyed in [87].

A number of papers have described the application of RBF networks in the modeling of nonlinear dynamical systems and chaos [88–91]. An accessible introduction on the subject of RBF models applied to modeling chaotic systems is given in [76].

3.2.4. Discrete-Time Polynomials

The model class illustrated in Figure 3(c) is quite similar to the preceding one. The main difference is that the basis functions are usually different, that is, $\omega_i \neq \omega_j$. Another important difference, not revealed in the figure, is that, whereas it is usually assumed in the second class that the input vector is uniform, in the third class such uniformity is not required (see Section 4.1). However, the main difference is that various basis functions are often used in Figure 3(c) in such a way as to enable matching different data features. One possible choice of basis functions ω_i is monomials of different degrees of nonlinearity in the range $1 \le m \le \ell$. Each mth-order term can contain a pth-order factor in $y(k - n_i)$ and a $(m - p)$th-order factor in $u(k - n_i)$ and is multiplied by a coefficient $c_{p,m-p}(n_1, \ldots, n_m)$ as follows:

$$y(k) = \sum_{m=0}^{\ell} \sum_{p=0}^{m} \sum_{n_1,n_m}^{n_y,n_u} c_{p,m-p}(n_1, \ldots, n_m) \prod_{i=1}^{p} y(k - n_i) \prod_{i=p+1}^{m} u(k - n_i),$$

(3.14)

where

$$\sum_{n_1,n_m}^{n_y,n_u} \equiv \sum_{n_1=1}^{n_y} \cdots \sum_{n_m=1}^{n_u},$$

(3.15)

and the upper limit is n_y if the summation refers to factors in $y(k - n_i)$ or n_u for factors in $u(k - n_i)$. In the case of noisy data $y(k - n_i)$, noise terms *must* be included in (3.14) to avoid the error-in-the-variables problem (see Section 5).

The choice of monomials as basis functions constrains the resulting models to those cases in which the dynamics underlying the data can be approximated by a linear combination of nonlinear monomials. For systems that are more strongly nonlinear, other basis functions should be preferred. On the other hand, the choice of monomial basis functions enables building models which are more information dependent than models for which all the basis functions are of the same type.

Discrete-time polynomial models have been used in a number of papers within the field of nonlinear dynamics and chaos. Bagarinao and colleagues used this model class to reconstruct bifurcation diagrams from data [92–95]. Data analysis and modeling applications

include [96–103]. Applications to real data—in the field of nonlinear dynamics—include [104–106].

3.2.5. Rational Models

Rational models are composed by the division of two polynomials such as (3.7) in the continuous time, and such as (3.14) in discrete time. It should be noticed at once that whereas *deterministic* polynomial models are linear-in-the-parameters, rational models, even deterministic, are nonlinear-in-the parameters. This calls for a series of cares in their estimation [107–109].

Rational models for continuous-time systems were considered in [57], although only analytically, as no model building was attempted in that paper. Rational models were estimated from data produced by an electronic oscillator in [110]. The practical difficulties with rational models seem to be revealed by the small number of papers that deal with such models in the context of nonlinear dynamics and chaos. At least for continuous-time model, rational functions lead to many numerical instabilities that are not trivial to remove. Nevertheless, such a problem is no longer observed when return maps are considered. Thus, Lorenz map and Ikeda map have been successfully reproduced with rational models [111].

3.2.6. Wavelet-Based Models

Wavelet models, wavelet networks, or just *wavenets* are similar to the linear-in-the-parameter models described in Sections 3.2.3 and 3.2.4. The main difference is that instead of radial basis functions or monomials, wavenets are composed of a linear combination of a set of wavelet functions $\psi_{a,b}(t)$ formed from a so-called wavelet prototype $\psi(t)$ by dilations and translations such as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \tag{3.16}$$

where $t, a, b \in \mathbb{R}$ and $a > 0$. As for RBF models, there are various choices of the wavelet function $\psi(t)$. There seems to be no clear guidance as to which type of function to choose. Since the various functions $\psi_{a,b}(t)$ used to compose the model are different (for different a and b), then the structure selection for this type of models resembles that of the discrete polynomials described in Section 3.2.4 because the user must chose not only the number of basis functions to use, but also the features of them (the as and bs).

One of the first papers to use wavenets in order to model a chaotic system was [29]. The authors used a mexican hat type wavelet function $\psi(t)$ to model various benchmark models. Optimized tensor product wavelet models were obtained for data from a vibrating string experiment in [112], and B-splines wavelet models for the nonautonomous Duffing oscillator were discussed in [113]. Wei and Billings investigated the use of structure selection algorithms for wavelet models in the context of chaotic systems [114].

3.2.7. Fuzzy Logic

Fuzzy models use internal variables which are linguistic. At a certain point of modeling the numerical variables must become linguistic by taking labels such as small negative large positive. The core of a fuzzy model then consists of a set of rules of the type: if x is large, then y is medium. Finally, the linguistic variables must be changed back to numerical variables.

Of course there are a number of subtelties which actually make this type of modeling work [115–118].

From the list of model classes surveyed in this paper, it seems to us that fuzzy logic models have been the least used to model nonlinear dynamics and chaos. One of the first papers on the subject seems to be [119]. A comparison of serveral techniques, including fuzzy models, in a problem of nonlinear prediction can be found in [120].

3.2.8. Input-Output Models

Input-output models are not a class of models but simply mean that the model class caters for the use of input signals. By input we mean external, time-dependent signal(s). Therefore an input-output model will be nonautonomous.

For the modeling of input-output models it is necessary to record not only the output, but also the input. As a general rule, global differential equations built from data are autonomous (for an exception, see [65] where the method applies to harmonically driven systems). In [121] the authors discuss the use of a bifurcation parameter (as an input) in the estimation of differential equations. However, in general, if inputs must be handled, then discrete-time models are more convenient.

There have been attempts to develop an embedding theory for input-output models [122, 123]. The feasibility of this theory and its benefits are not totally clear. Fortunately, there are other theoretical approaches to the input-output case [25, 124], and a vast number of examples illustrate the practical value of such models.

Nonautonomous chaotic models have been obtained in [65, 113, 125].

3.2.9. Equivalence between Continuous and Discrete-Time Polynomials

As surveyed above, there are various methods for building continuous-time models, typically in the form of ODEs. However, the data available is always discrete in time, and the simulation of ODEs is also carried out on digital computers. Therefore, at some stage some type of discretization of the ODEs must occur. In some methods such discretization is intentional [62, 126], however, in such cases the final aim was to have an ODE at the end of the day. What happens to the ODEs fixed-points after it is discretized was discussed in [127].

A more fundamental question concerns the choice of the integration step of numerical integration schemes. Given an ODE, say that produces a chaotic attractor, what happens to the attractor if the integration is varied? If it becomes too great, certainly the attractor will change, but will the "new" attractor still be some attractor of the original system? These questions have been addressed in [128]. In fact, so long as the frequency corresponding to the integration step is more than twice the highest frequency in the Fourier spectrum (Nyquist criterion), the "new" attractor is still topologically equivalent to an attractor solution to the set of ODEs. Only a displacement in the parameter space is induced by the integration step, which functions as a bifurcation parameter in some cases. It is only when the integration step is larger than the maximum established by Nyquist's criterion that the obtained attractor is spurious, that is, associated with numerical instabilities.

4. Structure Selection

In few words, the problem of structure selection is that of deciding *how many* and in some cases *which* function basis (see, e.g., Figure 3(c)) should be used to build a dynamical model

from data. This problem gains different tones depending on the model class considered but it is always present. Often in the context of networks, structure selection is known as "topology determination." This issue has been addressed in a number of papers [129–133]. In the case of fuzzy models, structure selection boils down to define how to partition the input space and to define the number of rules. Structure selection issues for fuzzy models have been investigated in [35, 134, 135]. Even for local linear models, structure selection would be the definition of the embedding and the number, size and location of neighborhoods, and so forth.

In the field of system identification, the issue of structure selection for nonlinear models gained much attention by the late 80 s. However, in the field of nonlinear dynamics and chaos this trend was delayed a few years. In the early (and sometimes late!) 90 s several papers simply assumed the structure known [55, 136–138]. A good example of what happens when structure selection is not considered during model building is provided in [139].

Soon the practical importance of model structure selection started to be recognized and dealt with in various ways and in varying degrees of success [62, 140, 141]. By the mid 90 s most authors became aware of the necessity of structure selection regardless of the model class chosen. However, the problems due to overparametrization were not well diagnosed, usually being attributed to an allegedly poor model class, poor data, and so on. A detailed study of the dynamicall effects of overparametrization on model attractors and bifurcation diagrams was produced in [142]. Other similar studies followed [143–147].

The key point in structure selection is to choose a model structure that is as simple as possible, but also sufficiently complex to capture the dynamics underlying the data. One of the easiests (and less efficient) methods for model selection is called zeroing-and-refitting [141] which consists of estimating the parameters for a large model—which is already a problem—and to eliminate those which are "sufficiently small." Of course, the "size" of a parameter will depend on the particular window of data used, on the noise variance, and on the particular type of basis function used, especially if the data are not normalized [148]. Therefore, zeroing-and-refitting does not generally work in practice [62].

An alternative and simple way of tackling this problem for nonlinear systems has been proposed in [149]. Instead of of deleting specific model terms, entire term clusters should be deleted, based on the behavior of the respective cluster parameters. Also, certain term clusters can be deleted to force symmetry [144, 150], which is necessary in some cases [151]. The concept of the nearest neighboors was used in [152] to determine the best input and output lags in NARX (nonlinear autoregressive with exogenous inputs) models.

One way of addressing the structure selection problem are to define some measure of complexity for a given model. In their paper Crutchfield and McNamara were concerned with quantifying and limiting the complexity of their models. They chose models that minimized the model entropy $I(M)$ and argue the importance of such a measure in the context of chaotic data [37]. The *maximum description length* (MDL) [153] has been used in several papers [132, 154]. Other ways of tackling the structure selection problem is to define a measure of quality for each regressor *in an orthogonal space* and then use such a criterion to select those regressors that are most relevant. This is the basic idea behind the *Error Reduction Ratio* (ERR) proposed in [155] and used in [156].

Unfortunately, there is no definite solution to the model structure selection problem so far. Situations in which the current methods fail abound. Brown and colleagues report failure of the MDL criterion [150], and Piroddi provides simple examples in which the ERR fails [33]. Fortunately, the benefits and successes outnumber the failures. New algorithms for structure selection are published at an amazing pace [30–35]. The authors of this survey are convinced

that the use of a priori information (see, e.g., [144, 151, 157]) will prove useful in addition to the new techniques.

4.1. Uniform and Nonuniform Embeddings

A key issue in modeling nonlinear dynamics is that of selecting an appropriate embedding space. In principle, this would include two stages: the choice of observables [158, 159] and the choice of embedding parameters [160]. In many practical situations, although the observable is determined before data acquisition, the embedding parameters—basically the embedding dimension and the delay time—can be determined by the user a posteriori. In [161] a method has been suggested to try to detect if a given variable is appropriate for global modeling. Using the nomenclature of Figure 3, the input vector, at the left-hand side of the models, determines the embedding space.

It has been duly pointed out that the problem of choosing an embedding in the context of model building is a bona fide stage of the modeling procedure [162]. Also, the uniform embedding defined by taking the elements of $\mathbf{y}(k)$ in (3.1) to be the coordinates is not necessarily optimal. Therefore, which elements (coordinates) should be chosen to compose $\mathbf{y}(k)$ is also part of the modeling problem. An optimal solution to such a problem might require an embedding space in which the "temporal distances" from one coordinate to another are not necessarily the same. The authors of [162] classify such embedding spaces as irregular. Despite this, to build discrete-time models using regular embeddings seems to be the rule rather than the exception in most of the literature.

To enable irregular embeddings (see Figure 3) it suffices not to connect certain input nodes to the middle layer. Here it is pointed out that the choice of particular basis functions ω_i is, in some cases equivalent to the choice of embedding coordinates which could turn out to be irregular. This is one of the advantages of using the ERR criterion to accomplish structure selection. In fact, (13) and (21) of [156] and (37) of [125] are some examples of models (automatically) built on irregular embeddings. In conclusion, it becomes clear that the modeling procedure followed in [125], for instance, includes the choice of embedding coordinates as a part of the modeling procedure, as pointed out in [162].

5. Parameter Estimation

Before, it is noted that synchronization has been used in parameter estimation problems [163–166].

One of the most commonly used algorithms for estimating unknowns in linear-in-the-parameters models is the least-squares algorithm or some generalization of it [34]. In the case of noisy measured data the least-squares estimator is biased due to the error-in-the-variables problem [167, 168] because such an estimator does not take into account the measurement errors [169]. Fortunately, there are well-established and robust algorithms for such situations which are mildly nonlinear [22]. Some of such algorithms solve the optimization problem in a higher-dimensional space thus successfully avoiding bias. Due to the unbiased estimators—some of which take into account the measurement errors—in the field of system identification, the error-in-the-variables problem only occurs when both input and output are noise contaminated. When only the output is noisy, the unbiased estimators successfully circumvent the error-in-the-variables problem.

All such algorithms are based in some norm of one-step-ahead prediction error. An alternative would be to minimize some norm of a k-step-ahead prediction error, with the

advantage of gaining robustness to noise. In such a case, however, the resulting optimization problem becomes strongly nonlinear and should be solved with adequate tools. The use of a back-propagation algorithm in this context was used in [170]. In [43], which also followed a local model procedure, the local maps were fitted to the data in such a way as to be consistent with k-step-ahead predictions. The authors of that paper reported that estimating parameters by least squares did not always yield good results. This should come as no surprise because they were fitting local maps and therefore required additional information to constrain parameter estimation. Global models are less sensitive to this type of problem.

In what concerns parameter estimation, an important difference between [59, 60] and [56] is that, whereas the former uses linear estimators—at the expense of having to estimate derivatives from data—the latter numerically integrates (3.7) and uses the observed data as they are. This procedure, which was called the *trajectory method*, results in an optimization problem which the authors solve using standard routines in the IMSL10 library. In [55], the authors provide more information on this estimation procedure and also consider the discrete-time case. The trajectory method in the context of RBF models is discussed in [76]. It should be noticed that [55, 56] do not consider modeling problems such as structure selection. The application of the trajectory method was illustrated in [171], and the use of the multiple shooting approach to estimated parameters of ordinary differential equations was discussed in [73, 136, 169].

For *neural networks* the weights and the bias terms are determined by optimization algorithms that search to minimize a cost function which usually depends on the difference between the given data and the network output. Because such models are nonlinear in the parameters, the optimization problem must be solved using some kind of nonlinear estimator, such as the back-propagation algorithm. In a recent study, different approaches to network training were compared [172].

Breeden and Packard have discussed the use of genetic algorithm and evolutionary programming for solving a number of optimization problems which occur in the modeling of nonlinear dynamics and chaos [173]. Other parameter estimation techniques include a generalized Gauss-Newton method for maximizing a likelihood function [74]; ridge regression or regularization [49, 89]. The multiple shooting approach has been used and discussed in a number of papers [136, 174]. Nonparametric estimation (a rare example in nonlinear dynamics) was provided in [175]. The use of regularization has been considered in [49, 76]. A specific procedure for estimating three parameters of a differential-delay equation was put forward in [176] and another method, specific to a certain class of one-dimensional maps, was put forward in [177]. A multiobjective estimator was discussed in [178]. Finally, a number of issues directly concerned with the estimation of dynamical invariants and indirectly related to the estimation of model parameters were discussed in [167].

6. Model Validation

The issue of model validation is vast. In order to cover such a wide subject in limited space, we will base this section on the paper [179] and refer the reader to [180] for a coverage of the field up to the beginning of the nineties.

Before actually starting to describe some results in the literature, a few remarks are in order. First and foremost, the challenge of model validation or of choosing among candidate models should take into account the intended use of the model. Hence, a model could be good for one type of applications and, nonetheless, perform poorly in another. In the context of this paper, the main concern is to assess the model dynamics. A different concern, though equally

valid, that would probably require a different approach would be to assess the forecasting capabilities of a model. Second, it should be realized that two similar though different problems are: (i) model validation, which usually aims at an *absolute* answer like: valid or not valid, (ii) model selection, which usually aims at a *relative* answer such as: model $\mathcal{M}_1$ is better than model $\mathcal{M}_2$ according to criterion $\mathcal{C}$. Finally, when it comes to model validation, the safest approach is to use many criteria, rather than just one.

Although very popular in other fields, the computation of various *measures of prediction errors* (one-step-ahead and free-run) in the case of nonlinear dynamical systems is not conclusive in what concerns the overall dynamics of the identified model [81, 132, 180], though it does convey much information on the forecasting capabilities of a model.

Subjective though it is, the *visual inspection of attractors* (or simply comparing the morphology of two time series) is quite common a way of assessing the quality of models [52, 60, 65, 72, 73, 78, 91, 119, 132, 162, 168, 171, 174, 181–183]. Such a procedure is not only subjective but also ineffective to discriminate between "close" models, that is, models with slight, but important, differences in their dynamical behavior. What renders this procedure subjective is the fact that no quantitative mechanism is used to compare how close are two reconstructed attractors. In this respect the work by Pecora and coworkers could be an alternative for determining how close are the original and the model attractors [184]. To the best of our knowledge the statistic measures that put forward in the mentioned paper have not yet been used in the context of model validation.

Still in relation to the visual inspection of attractors, it should be noticed that in many practical instances there is not much more that can be done consistently. For instance, in the case of slightly nonstationary data, to compare short-term predictions with the original data is basically the best that can be done. Building a model for which the free-run simulation approximates the original data in some sense is usually a nontrivial achievement. (In this respect we rather disagree with [185] who consider free-run simulation of models a trivial validity test.)

Other *attractor features* are still in common use when it comes to model validation. Among such features the following are frequently used: (some of such properties have been recently discussed in the context of model validation in [186]) Lyapunov exponents [43, 81, 97, 145, 147, 187, 188]; correlation dimension [78, 81, 132, 188]; location and stability of fixed-points [112, 125, 183]; (In particular, it has been shown that fixed-point stability of nonlinear models is consistent with breathing patterns found in real data [189].) Poincaré sections [81, 176]; geometry of attractors [85]; attractor symmetry [150, 151]; first-return maps on a Poincaré section [66, 190]; probability density functions of recurrence in state-space [191]; topological features such as linking numbers and unstable periodic orbits (UPOs) [192–194].

Meaningful validation can only be accomplished by taking into account the intended use of the model. A model that provides predictions consistent with the observed data will probably not be a good model to study, say, the sequence of bifurcations of the original system.

The use of model free-run simulations in the context of surrogate data analysis for model validation was suggested in [195]. For details on surrogate data analysis the reader is referred to [196, 197] and for achemical process application; see [198]. The main idea is to use estimated models to produce a large number of time series and to use some test statistic to try to assess if it is likely that the data could have been produced by a model (or models) such as those used to produce the surrogates. Of course, a key point in this procedure is the choice of the test statistic. For instance, suppose that the correlation dimension or the largest
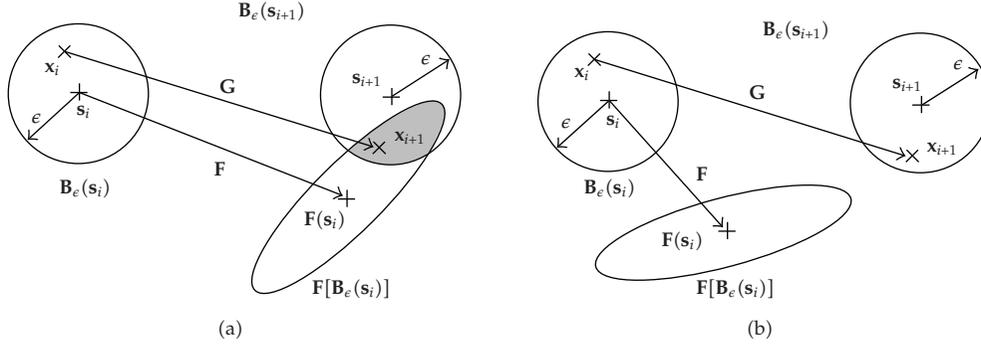
Figure 4: Graphical interpretation of (a) consistent prediction and (b) inconsistent prediction [200]. $\mathbf{G}$ and $\mathbf{F}$ are the "true" dynamics and model, respectively. $\mathbf{x}_i$ is the true state at time i and the observed state at that time is $\mathbf{s}_i$. $\mathbf{B}_\varepsilon(\mathbf{s}_i)$ is an observational uncertainty sphere of radius ε. In (a) the shaded region indicates that there is a subset of forecasts made with $\mathbf{F}$ that are indistinguishable from the "true" state $(\mathbf{x}_{i+1})$ within observational uncertainty. Figure reproduced from [179].

Lyapunov exponent is chosen to compare a set of free-run simulated data with the measured data. Suppose further that the null hypothesis is that the measured data is compatible with the estimated model. Even if we cannot reject the null hypothesis, that does not guarantee equivalence of *dynamical* behavior because quite different attractors may have similar indices [199]. This is also true for the correlation dimension.

A related method has been described in [200] and has been named consistent nonlinear dynamic (CND) testing. The main idea is depicted in Figure 4.

In the following two subsections we briefly discuss two other procedures which the authors have developed for model validation. The first one, following an early paper by Brown and colleagues [201], is based on the concept of synchronization. The second procedure, which is much more demanding, but also provide a much deeper insight into the model dynamics is based on topological analysis.

6.1. Synchronization

As done in Figure 4, let us denote the "true" dynamics by $\mathbf{G}$ and a given model by $\mathbf{F}$. In the present section it is assumed that the dynamics are continuous, that is [201],

$$\frac{d\mathbf{x}}{dt} = \mathbf{G}(\mathbf{x}),$$

$$\frac{d\mathbf{y}}{dt} = \mathbf{F}(\mathbf{y}) - \mathbf{E}(\mathbf{y} - \mathbf{x}),$$

$$(6.1)$$

where it has been assumed that $\dim[\mathbf{x}] = \dim[\mathbf{y}]$ and where the matrix $\mathbf{E}$ denotes the coupling between the true system and the model. The scheme illustrated in (6.1) will be referred to as dissipative synchronization or entrainment.

The rationale behind this procedure is as follows. Assume that the data $\mathbf{x}$ lies on a chaotic attractor. In many situations, provided $\mathbf{E}$ is adequately chosen and $\mathbf{G} = \mathbf{F}$, $\mathbf{y} \to \mathbf{x}$. That is, the model will synchronize to the system. If $\mathbf{G}$ and $\mathbf{F}$ differ slightly, the error $\mathbf{e} = \mathbf{y} - \mathbf{x}$ will not go to zero but will stay around the origin of the error space. The average distance to the origin of such a space will depend on $\mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{x})$. Therefore such a distance (which in

practice is a measure of quality of "synchronization") is a measure of how far the estimated model **F** is from the true dynamics **G** [201].

The difference $|\mathbf{x} - \mathbf{z}|$—where **z** is the model state vector *without* any driving force—is compared to $|\mathbf{x} - \mathbf{y}|$. If $|\mathbf{x} - \mathbf{y}|$ drops below a certain threshold (10^{-2} was used in [185]) and is "clearly" smaller than $|\mathbf{x} - \mathbf{z}|$, that is, $\mathbf{x} \approx \mathbf{y}$ and it is assumed that the model is synchronized to the data, therefore **F** should be sufficiently close to **G**.

As it often happens in the realm of model validation, this procedure also is highly subjective, since it requires an ad hoc threshold, mentioned in the previous paragraph. In what concerns dissipative synchronization, it is well-known that in many cases by increasing the strength of the coupling (matrix **E**) it is possible to force a greater degree of synchronization and, in some cases, even attain identical synchronization [201]. For instance, in [185] the authors found that for values of the coupling greater than 2, models of an electronic circuit would synchronize with the measured data. On the other hand, in [202] they have found a lower bound of 0.1 for the coupling strength in order to guarantee synchronization between the Rössler system and perturbed versions of the original equations. It therefore becomes clear that it is sometimes possible to synchronize even a poor model to the data so long as the coupling strength is made sufficiently large. In fact, is has been shown that even different systems can synchronize, at a rather high cost [203].

Therefore although the concept of synchronization could be useful in the context of model validation, it becomes apparent that some adjustments are required to render the procedure more practical. In [179] some steps were given in this direction, and it was shown using numerical examples that synchronization yielded similar results to the use of bifurcation diagrams for model validation [180] but at a much lower computational cost. Bifurcation diagrams in model validation were used in [86, 89].

6.2. Topological Analysis

A topological analysis usually starts by computing a first-return map to a Poincaré section of the phase portrait. This is indeed useful for extracting periodic orbits using a close return method although that other techniques can be used. In the best cases, the first-return map is made of n monotonic branches separated by an $n - 1$ critical point. The first-return map induces therefore a partition of the phase portrait in n zones. A symbol is associated with each branch.

Chaotic trajectories and the periodic orbits constituting their skeleton are thus encoded over the symbol set $\{0, 1, \ldots, n - 1\}$. Even symbols are associated with increasing branches. Increasing branches are preserving order and decreasing branches are reversing order. The phase portrait is thus divided in preserving order and reversing order strips. A preserving order strip presents an even number of half-turns while a reversing order strip represents an odd number of half-turns. The attractor can thus be schemed by a branched manifold—a template or a knot holder—on which periodic orbits can be drawn.

All topological properties are encoded in the template. Thus it is possible to extract topological invariants like linking numbers from a template construction. Linking number between two periodic orbits is the most often used topological invariant. It can be counted on a regular plane projection of the two periodic orbits. Each crossing (in the plane projection) between the two orbits is associated to ± 1 according to the third coordinate of each orbit segment. The linking number is then the half sum of the orientated crossings. The template is valid when all linking numbers it predicts are equal to those counted from the periodic orbits extracted from the attractor studied. Many details about topological analysis can be found in [204].

Topological analysis is surely the strongest validation method. Unfortunately, it is up-to-now restricted to 3D dyanmics. Topological validation of global models was performed for the copper electrodissolution experiments [69], a string experiment [193], a Belousov-Zhabotinskii reaction [205], and so forth.

7. Modeling with A Priori Knowledge

We start this section with a quotation from one of the pioneering papers of the field: "If extra information is available about a system in addition to a time series, such as explicit underlying partial differential equations or symmetries, then the inverse approach as presented here does not exploit such information. Consequently, for such systems it may be possible to improve significantly upon the inverse approach using other techniques" [39, page 354].

Despite opinions as the one just quoted, the problem of building models from data *and* additional information (sometimes referred to as a priori or *auxiliary*) has been postponed. In the context of linear models or nonlinear process models, some results are available [206–210]. However, as pointed out in [211] there seems to be less applications of gray-box modeling in the realm of nonlinear dynamics.

The general setting is to have the dynamical data *plus* some other source of information and to use both in building a model. The knowledge of the model equations is optimistic to be considered a priori information, and here it is assumed that structure selection techniques (Section 4) will be used to find out which basis functions should be used to compose the model. If all that is desired of the system is already available in the particular set of data used for model building, then granted that the model class is sufficiently general and that the model structure is adequate, the resulting model should be a sufficiently accurate representation of the system. In this case there is no motivation to use an additional piece of information. However, often in practice, for a number of different reasons—such as presence of noise, poor choice of observable, poor frequency content in the data, limited amplitude excursion, and the like—the available data either does not have all the desired information about the system or such information is difficult to obtain. In such cases it is conceivable that additional information is available and it is natural to enquire if it is possible to build the model. Thus, more often than not, if the data is of "good" quality and "sufficiently" complete, black-box modeling should be the practitioner's first choice.

In the realm of nonlinear dynamics, procedures have been put forward for building models using auxiliary information. A number of fixed-points were used in [144], the location of fixed points were used in [178]. Information about the harmonic driving were used to choose the model structure in [65]. Information about the symmetry was used to constrain not only the topology but also the parameter estimates of network and radial-basis function models [151]. In the last quoted paper, for instance, symmetry was imposed on a multilayer perceptron neural network in order to guarantee the pitchfork bifurcation—which is structurally unstable—when modeling the Duffing-Ueda oscillator. Topological features such as folding and tearing mechanisms [199, 212] in addition to the location and local eigenstructure of fixed points have been used in [213]. Lastly, information about the first period-doubling bifurcation diagram was used in [77].

8. Remaining Challenges and Conclusions

As early as 1987, the issue of modeling spatiotemporal nonlinear dynamical systems can be found in the literature [37]. Defining which type of a priori information is usable and then

learning how to use it for a given model class remain a challenge. In this difficult problem it is necessary to find adequate duplets of information and model class. To posses a priori information and not knowing how to use it for a given model class is useless. The reverse is equally true. The combination of models with different properties in an assemble approach [52] seems a promising line for future research in nonlinear dynamics and chaotic systems.

Over fifteen years ago Robert Gilmore, referring to model building, mentioned: "There are at present two distinct methods to model data. One is analytic, and has not been extensively used. The other is topological, and has been used even less" [214, page 515]. Although this paper did not aim at describing in detail the significant progress made in the field of "topological analysis," hopefully the reader did get the feeling that much has been done. On the other hand, concerning the "analytic methods" for modeling data, which was the main focus of this paper, it should become clear that quite a lot has been achieved since Gilmore's summary was written.

The view presented in this work, as any review paper, is strongly influenced by the authors' experience. A varied list of introductory and review articles was included. A rather short list of benchmark models and widely available data with indication of works that have used such models and data were provided. Although every effort has been made to survey such a wide subject in a comprehensive way, it goes without saying that a complete list of papers is outside the authors' reach. Nevertheless we are sure that the present survey will serve as a good starting point for future works in modeling nonlinear dynamics and chaos.

Appendix

A. Benchmark Models and Data Sets

As a helpful aid to those involved in developing new tools for modeling nonlinear dynamics and chaos, in this appendix we list some benchmark toy models and benchmark data sets (widely available) and point out papers that have discussed their modeling and analysis.

A.1. Maps

Logistic

The logistic map was originally investigated by Robert May [215, 216]. Papers that investigated the estimation of models from data produced by the logistic map include [38, 125].

Hénon

The logistic map was originally proposed in [217]. It was considered as a benchmark for modeling in [61, 125, 168, 218]. This map was modeled using a kernel estimation approach in [43], recurrent networks in [84], and a wavelet model in [114].

Ikeda

The Ikeda map was introduced in [219] and considered as a benchmark in [39, 61]. In [29] this map was modeled using wavelet networks, and in [132] results are reported for neural networks.

A.2. Autonomous Systems

van der Pol's Oscillator

The autonomous van der Pol oscillator settles to a limit cycle and therefore can be modeled. This oscillator has been considered in [119].

Lorenz

The Lorenz system was introduced in [4] and considered as a benchmark in [39, 76], although only model prediction errors are given. The polynomial and rational expansions of that system from a single observable were considered in [58], and the construction of ODEs from data was presented in [59, 60, 62]. This system was modeled using a kernel estimation approach in [43], in [125] it was modeled usign discrete-time polynomials, and in [29] it was modeled using wavelet networks.

Fuzzy models for this system were built in [119]. A piece-wise affine model for this system was developed and analyzed in [213], in a sense such a model bears resemblance with Takagi-Sugeno fuzzy models [116]. Unconstrained [85] neural networks and constrained RBF models [151] were also used in the modeling of this system.

Rössler

The well-known Rössler system was proposed in [220]. The polynomial and rational expansions of that system from a single observable were considered in [57], and the construction of ODEs from data was presented in [59, 60]. Fuzzy models for this system were built in [119, 125] it was modeled usign discrete-time polynomials. A piece-wise affine model for this system was developed and analyzed in [213], and a neural network was reported in [132].

Makey-Glass

The Makey-Glass delay-differential equation was introduced in [221] and considered as a benchmark in [39, 76, 125]. In [29] this system was modeled using wavelet networks, and in [84] the reader will find results with feedforward and recurrent neural networks.

A.3. Nonautonomous Systems

It must be pointed out that a nonautonomous oscillator or order n driven by a single frequency, say, an input of the type $\cos(\omega t)$ is, in practice, an autonomous system with dynamical order $n + 2$ [191]. In such cases the resulting models do not have an external variable $u(t)$ or $u(k)$ which can be *any* time-dependent signal.

Duffing's Oscillator

This oscillator was considered *for a fixed input* of the type $\cos(\omega t)$ in [37]. In [125] a model valid for any input was built from data. The estimated model and original oscillator have very similar bifurcation diagrams. Wavelet models with similar features were obtained in [113]. A close look at the pitchfork bifurcations will reveal that they are not (the symmetrical)

pitchforks but rather (the unsymmetrical) saddle-node bifurcations. Because symmetry is structurally unstable, any small deviation due to noise or "imperfect" estimation from data will destroy it, and with it the hope to reproduce a pitchfork bifurcation. However, by imposing symmetry it is possible to recover the correct bifurcation pattern. This has been done with neural network models for the Duffing oscillator in [151].

van der Pol's Oscillator

This oscillator appeared in [222]. An experimental implementation of this oscillator was considered *for a fixed input* of the type $\cos(\omega t)$ in [37]. In [125] a model valid for any input was built from data. The estimated model and original oscillator have very similar bifurcation diagrams. Wavelet models for this oscillator have been obtained in [114].

A.4. Benchmark Measured Time Series

Several papers in the literature illustrate the proposed techniques using measured time series. It would be impossible to list all of them. In this subsection, however, we list those time series that are most commonly used because they are widely available and therefore constitute typical benchmarks.

Santa Fé Time Series Competition Data

In 1994 the Santa Fé Institute promoted a time series prediction competition, which is described in [223]. Some measured and simulated time series were made available as benckmarks for the competition. The description of these data and the files can still be found at http://www-psych.stanford.edu/andreas/Time-Series/SantaFe.html.

The Laser data was considered in [81, 91]. The biomedical data set has been described in detail by Rigney and coworkers [224] in view of the Santa Fé Time Series Prediction and Analysis Competition [225]. Such data were considered in [189, 226–230].

Wolf's Sunspot Number Time Series

These data can be found at http://www.ngdc.noaa.gov/ and have been considered in [20, 49, 91, 106, 132, 231, 232]. In the last two references, the original data was transformed to a symetrical space in which the modeling is easier. The transformed data can be found at http://www.atomosyd.net/.

Copper Electrodissolution Data

These data can be found at http://www.atomosyd.net/spip.php?article40 and have been considered in [68, 69]. They were recorded by Zihao Fei and Jack Hudson. A 52-term global continuous-time model [69] and a 26-term model [68] were obtained from the data. A structure selection using an Ansatz library was used in the second case.

Canadian Lynx

These data results from the recrods of the Hudson Bay Company regarding the populations of lynx and hares. They can be found at http://www.atomosyd.net/spip.php?article39 and have been considered in [233]. They were recorded by Zihao Fei and Jack Hudson.

Electronic Oscillators

The electronic oscillator described in [234] was modeled and analyzed in [71, 73]. These data can be found at http://www.y2k.maths.ox.ac.uk/.

The so-called Chua's circuit [235] was modeled and analyzed in [236] using discrete-time polynomials, in [64] usign continuous-time polynomials and in [110] using rational models. Neural networks from simulated data of this circuit are described in [237]. These data can be found at http://www.cpdee.ufmg.br/~MACSIN/.

Acknowledgments

References

[1] H. Poincaré, "Sur les courbes définies par une équation différentielle," *Journal de Mathématiques Pures et Appliquées*, vol. 4, no. 1, pp. 167–244, 1885.

[2] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste*, vol. 3, Gauthier-Villard, Paris, France, 1899.

[3] A. A. Andronov, A. A. Vitt, and S. E. Khaikin, *Theory of Oscillators*, Pergamon Press, Oxford, UK, 1966, Traduction anglaise par F. Immirzi, Dover, New York, NY, USA, 1937.

[4] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of the Atmospheric Sciences*, vol. 20, no. 2, pp. 130–141, 1963.

[5] D. Ruelle and F. Takens, "On the nature of turbulence," *Communications in Mathematical Physics*, vol. 20, pp. 167–192, 1971.

[6] R. M. May, "Biological populations with nonoverlapping generations: stable points, stable cycles, and chaos," *Science*, vol. 186, no. 4164, pp. 645–647, 1974.

[7] O. E. Rössler, "An equation for continuous chaos," *Physics Letters A*, vol. 57, no. 5, pp. 397–398, 1976.

[8] T.-Y. Li and J. A. Yorke, "Period three implies chaos," *The American Mathematical Monthly*, vol. 82, no. 10, pp. 985–992, 1975.

[9] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.

[10] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980 (Coventry, 1979/1980)*, D. A. Rand and L. S. Young, Eds., vol. 898 of *Lecture Notes in Mathematics*, pp. 366–381, Springer, Berlin, Germany, 1981.

[11] R. Mañé, "On the dimension of the compact invariant sets of certain nonlinear maps," in *Dynamical Systems and Turbulence, Warwick 1980 (Coventry, 1979/1980)*, D. A. Rand and L. S. Young, Eds., vol. 898 of *Lecture Notes in Mathematics*, pp. 230–242, Springer, Berlin, Germany, 1981.

[12] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, "State space reconstruction in the presence of noise," *Physica D*, vol. 51, no. 1–3, pp. 52–98, 1991.

[13] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *Journal of Statistical Physics*, vol. 65, no. 3-4, pp. 579–616, 1991.

[14] J. D. Farmer, E. Ott, and J. A. Yorke, "The dimension of chaotic attractors," *Physica D*, vol. 7, no. 1–3, pp. 153–180, 1983.

[15] J. Theiler, "Estimating fractal dimension," *Journal of the Optical Society of America A*, vol. 7, no. 6, pp. 1055–1073, 1990.

[16] V. I. Oseledec, "A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems," *Transactions of the Moscow Mathematical Society*, vol. 19, pp. 179–210, 1968.

[17] A. N. Kolmogorov, "A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces," *Doklady Akademii Nauk SSSR*, vol. 119, no. 5, pp. 861–864, 1958.

[18] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. Sh. Tsimring, "The analysis of observed chaotic data in physical systems," *Reviews of Modern Physics*, vol. 65, no. 4, pp. 1331–1392, 1993.

[19] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, 2nd edition, 2004.

[20] M. Casdagli, D. Des Jardins, S. Eubank, et al., "Nonlinear modeling of chaotic time series: theory and applications," in *Applied Chaos*, J. H. Kim and J. Stringer, Eds., Wiley-Interscience Publication, chapter 15, pp. 335–380, John Wiley & Sons, New York, NY, USA, 1992.

[21] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control*, Prentice-Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1994.

[22] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, USA, 2nd edition, 1999.

[23] R. J. G. B. Campello, G. Favier, and W. C. do Amaral, "Optimal expansions of discrete-time Volterra models using Laguerre functions," *Automatica*, vol. 40, no. 5, pp. 815–822, 2004.

[24] S. A. Billings, "Identification of nonlinear systems—a survey," *IEE Proceedings D*, vol. 127, no. 6, pp. 272–285, 1980.

[25] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for nonlinear systems—part II: stochastic nonlinear systems," *International Journal of Control*, vol. 41, no. 2, pp. 329–344, 1985.

[26] S. A. Billings and Q. M. Zhu, "Rational model identification using an extended least-squares algorithm," *International Journal of Control*, vol. 54, no. 3, pp. 529–546, 1991.

[27] K. S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 4–27, 1990.

[28] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks*, vol. 2, no. 2, pp. 302–309, 1991.

[29] L. Cao, Y. Hong, H. Fang, and G. He, "Predicting chaotic time series with wavelet networks," *Physica D*, vol. 85, no. 1-2, pp. 225–238, 1995.

[30] L. Z. Guo, S. A. Billings, and D. Q. Zhu, "An extended orthogonal forward regression algorithm for system identification using entropy," *International Journal of Control*, vol. 81, no. 4, pp. 690–699, 2008.

[31] S. A. Billings and H. L. Wei, "An adaptive orthogonal search algorithm for model subset selection and non-linear system identification," *International Journal of Control*, vol. 81, no. 5, pp. 714–724, 2008.

[32] H.-L. Wei and S. A. Billings, "Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information," *International Journal of Modelling, Identification and Control*, vol. 3, no. 4, pp. 341–356, 2008.

[33] L. Piroddi, "Simulation error minimisation methods for NARX model identification," *International Journal of Modelling, Identification and Control*, vol. 3, no. 4, pp. 392–403, 2008.

[34] X. Hong, R. J. Mitchell, S. Chen, C. J. Harris, K. Li, and G. W. Irwin, "Model selection approaches for non-linear system identification: a review," *International Journal of Systems Science*, vol. 39, no. 10, pp. 925–946, 2008.

[35] F. P. Pach, A. Gyenesei, and J. Abonyi, "MOSSFARM: model structure selection by fuzzy association rule mining," *Journal of Intelligent and Fuzzy Systems*, vol. 19, no. 6, pp. 399–407, 2008.

[36] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Physical Review Letters*, vol. 59, no. 8, pp. 845–848, 1987.

[37] J. P. Crutchfield and B. S. McNamara, "Equations of motion from a data series," *Complex Systems*, vol. 1, no. 3, pp. 417–452, 1987.

[38] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Systems*, vol. 2, no. 3, pp. 321–355, 1988.

[39] M. Casdagli, "Nonlinear prediction of chaotic time series," *Physica D*, vol. 35, no. 3, pp. 335–356, 1989.

[40] J. Cremers and A. Hübler, "Construction of differential equations from experimental data," *Zeitschrift für Naturforschung A*, vol. 42, no. 8, pp. 797–802, 1987.

[41] J. D. Farmer and J. J. Sidorowich, "Exploiting chaos to predict the future and reduce noise," in *Evolution, Learning and Cognition*, Y. C. Lee, Ed., pp. 277–330, World Scientific, Teaneck, NJ, USA, 1988.

[42] M. Casdagli, "Chaos and deterministic versus stochastic nonlinear modelling," *Journal of the Royal Statistical Society. Series B*, vol. 54, no. 2, pp. 303–328, 1992.

[43] H. D. I. Abarbanel, R. Brown, and J. B. Kadtke, "Prediction in chaotic nonlinear systems: methods for time series with broadband Fourier spectra," *Physical Review A*, vol. 41, no. 4, pp. 1782–1807, 1990.

[44] M. B. Kennel and S. Isabelle, "Method to distinguish possible chaos from colored noise and to determine embedding parameters," *Physical Review A*, vol. 46, no. 6, pp. 3111–3118, 1992.

[45] P. S. Linsay, "An efficient method of forecasting chaotic time series using linear interpolation," *Physics Letters A*, vol. 153, no. 6-7, pp. 353–356, 1991.

[46] G. Xiaofeng and C. H. Lai, "Improvement of the local prediction of chaotic time series," *Physical Review E*, vol. 60, no. 5, pp. 5463–5468, 1999.

[47] L. Cao, A. Mees, and K. Judd, "Dynamics from multivariate time series," *Physica D*, vol. 121, no. 1-2, pp. 75–88, 1998.

[48] P. Grassberger, T. Schreiber, and C. Schaffrath, "Nonlinear time sequence analysis," *International Journal of Bifurcation and Chaos*, vol. 1, no. 3, pp. 521–547, 1991.

[49] D. Kugiumtzis, O. C. Lingjærde, and N. Christophersen, "Regularized local linear prediction of chaotic time series," *Physica D*, vol. 112, no. 3-4, pp. 344–360, 1998.

[50] L. Cao and A. S. Soofi, "Nonlinear deterministic forecasting of daily dollar exchange rates," *International Journal of Forecasting*, vol. 15, no. 4, pp. 421–430, 1999.

[51] A. Lapedes and R. Farber, "Nonlinear signal processing using neural networks: prediction and system modelling," submitted to Proceedings of the IEEE, Los Alamos Report LA-UR 87–2662, 1887.

[52] U. Parlitz, A. Hornstein, D. Engster, et al., "Identification of pre-sliding friction dynamics," *Chaos*, vol. 14, no. 2, pp. 420–430, 2004.

[53] J.-S. Zhang and X.-C. Xiao, "Predicting chaotic time series using recurrent neural network," *Chinese Physics Letters*, vol. 17, no. 2, pp. 88–90, 2000.

[54] J.-S. Zhang and X.-C. Xiao, "Predicting hyper-chaotic time series using adaptive higher-order nonlinear filter," *Chinese Physics Letters*, vol. 18, no. 3, pp. 337–340, 2001.

[55] J. L. Breeden and A. Hübler, "Reconstructing equations of motion from experimental data with unobserved variables," *Physical Review A*, vol. 42, no. 10, pp. 5817–5826, 1990.

[56] J. L. Breeden, F. Dinkelacker, and A. Hübler, "Noise in the modeling and control of dynamical systems," *Physical Review A*, vol. 42, no. 10, pp. 5827–5836, 1990.

[57] G. Gouesbet, "Reconstruction of the vector fields of continuous dynamical systems from numerical scalar time series," *Physical Review A*, vol. 43, no. 10, pp. 5321–5331, 1991.

[58] G. Gouesbet, "Reconstruction of vector fields: the case of the Lorenz system," *Physical Review A*, vol. 46, no. 4, pp. 1784–1796, 1992.

[59] G. Gouesbet and J. Maquet, "Construction of phenomenological models from numerical scalar time series," *Physica D*, vol. 58, no. 1–4, pp. 202–215, 1992.

[60] G. Gouesbet and C. Letellier, "Global vector-field reconstruction by using a multivariate polynomial L_2 approximation on nets," *Physical Review E*, vol. 49, no. 6, pp. 4955–4972, 1994.

[61] M. Giona, F. Lentini, and V. Cimagalli, "Functional reconstruction and local prediction of chaotic time series," *Physical Review A*, vol. 44, no. 6, pp. 3496–3502, 1991.

[62] R. Brown, N. F. Rulkov, and E. R. Tracy, "Modeling and synchronizing chaotic systems from time-series data," *Physical Review E*, vol. 49, no. 5, pp. 3784–3800, 1994.

[63] L. Le Sceller, C. Letellier, and G. Gouesbet, "Structure selection for global vector field reconstruction by using the identification of fixed points," *Physical Review E*, vol. 60, no. 2, pp. 1600–1606, 1999.

[64] L. A. Aguirre, U. S. Freitas, C. Letellier, and J. Maquet, "Structure-selection techniques applied to continuous-time nonlinear models," *Physica D*, vol. 158, no. 1–4, pp. 1–18, 2001.

[65] B. P. Bezruchko and D. A. Smirnov, "Constructing nonautonomous differential equations from experimental time series," *Physical Review E*, vol. 63, no. 1, Article ID 016207, 7 pages, 2000.

[66] C. S. M. Lainscsek, C. Letellier, and F. Schürrer, "Ansatz library for global modeling with a structure selection," *Physical Review E*, vol. 64, no. 1, Article ID 016206, 15 pages, 2001.

[67] M.-A. Boiron and J.-M. Malasoma, "Modélisation globale d'un système électrochimique," in *Compte-Rendus de la 8ème Rencontre du Non-Linéaire*, Non-Linéaire, Orsay, France, 2005.

[68] M.-A. Boiron and J.-M. Malasoma, "Modélisation globale d'un système électrochimique," in *Proceedings of the 5th Colloque on Temporal Chaos and Spatiotemporal Chaos*, Le Havre, France, December 2005.

[69] C. Letellier, L. Le Sceller, E. Maréchal, et al., "Global vector field reconstruction from a chaotic experimental signal in copper electrodissolution," *Physical Review E*, vol. 51, no. 5, pp. 4262–4266, 1995.

[70] A. D. Irving and T. Dewson, "Determining mixed linear-nonlinear coupled differential equations from multivariate discrete time series sequences," *Physica D*, vol. 102, no. 1-2, pp. 15–36, 1997.

[71] C. Letellier, G. Gouesbet, and N. F. Rulkov, "Topological analysis of chaos in equivariant electronic circuits," *International Journal of Bifurcation and Chaos*, vol. 6, no. 12B, pp. 2531–2555, 1996.

[72] C. Letellier, L. Le Sceller, G. Gouesbet, F. Lusseyran, A. Kemoun, and B. Izrar, "Recovering deterministic behavior from experimental time series in mixing reactor," *AIChE Journal*, vol. 43, no. 9, pp. 2194–2202, 1997.

[73] J. Timmer, H. Rust, W. Horbelt, and H. U. Voss, "Parametric, nonparametric and parametric modelling of a chaotic circuit time series," *Physics Letters A*, vol. 274, no. 3-4, pp. 123–134, 2000.

[74] W. Horbelt, J. Timmer, M. J. Bünner, R. Meucci, and M. Ciofini, "Identifying physical properties of a CO_2 laser by dynamical modeling of measured time series," *Physical Review E*, vol. 64, no. 1, Article ID 016222, 7 pages, 2001.

[75] K. H. Chon and R. J. Cohen, "Linear and nonlinear ARMA model parameter estimation using an artificial neural network," *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 3, pp. 168–174, 1997.

[76] S. Haykin and J. Príncipe, "Making sense of a complex world," *IEEE Signal Processing Magazine*, vol. 15, no. 3, pp. 66–81, 1998.

[77] L. A. Aguirre and E. C. Furtado, "Building dynamical models from data and prior knowledge: the case of the first period-doubling bifurcation," *Physical Review E*, vol. 76, no. 4, Article ID 046219, 13 pages, 2007.

[78] A. M. Albano, A. Passamante, T. Hediger, and M. E. Farrell, "Using neural nets to look for chaos," *Physica D*, vol. 58, no. 1–4, pp. 1–9, 1992.

[79] J. B. Elsner, "Predicting time series using a neural network as a method of distinguishing chaos from noise," *Journal of Physics A*, vol. 25, no. 4, pp. 843–850, 1992.

[80] K. Aihara, T. Takabe, and M. Toyoda, "Chaotic neural networks," *Physics Letters A*, vol. 144, no. 6-7, pp. 333–340, 1990.

[81] R. Bakker, J. C. Schouten, C. L. Giles, F. Takens, and C. M. van den Bleek, "Learning chaotic attractors by neural networks," *Neural Computation*, vol. 12, no. 10, pp. 2355–2383, 2000.

[82] H. M. Henrique, E. L. Lima, and J. C. Pinto, "A bifurcation study on neural network models for nonlinear dynamic systems," *Latin American Applied Research*, vol. 28, no. 3, pp. 187–200, 1998.

[83] R. A. Adomaitis, R. M. Farber, J. L. Hudson, I. G. Kevrekidis, M. Kube, and A. S. Lapedes, "Application of neural nets to system identification and bifurcation analysis of real world experimental data," in *Neural Networks: Biological Computers or Electronic Brains*, pp. 87–97, Springer, Paris, France, 1990.

[84] R. Gençay and T. Liu, "Nonlinear modeling and prediction with feedforward and recurrent networks," *Physica D*, vol. 108, no. 1-2, pp. 119–134, 1997.

[85] G. Boudjema and B. Cazelles, "Extraction of nonlinear dynamics from short and noisy time series," *Chaos, Solitons & Fractals*, vol. 12, no. 11, pp. 2051–2069, 2001.

[86] H. J. Kim and K. S. Chang, "A method of model validation for chaotic chemical reaction systems based on neural networks," *Korean Journal of Chemical Engineering*, vol. 18, no. 5, pp. 623–629, 2001.

[87] M. Small, *Applied Nonlinear Time Series Analysis*, vol. 52 of *World Scientific Series on Nonlinear Science. Series A*, World Scientific, Hackensack, NJ, USA, 2005.

[88] L. A. Smith, "Identification and prediction of low-dimensional dynamics," *Physica D*, vol. 58, no. 1–4, pp. 50–76, 1992.

[89] S. Ogawa, T. Ikeguchi, T. Matozaki, and K. Aihara, "Nonlinear modeling by radial basis function networks," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E79-A, no. 10, pp. 1608–1617, 1996.

[90] T. Miyano, S. Kimoto, H. Shibuta, K. Nakashima, Y. Ikenaga, and K. Aihara, "Time series analysis and prediction on complex dynamical behavior observed in a blast furnace," *Physica D*, vol. 135, no. 3-4, pp. 305–330, 2000.

[91] B. Pilgram, K. Judd, and A. Mees, "Modelling the dynamics of nonlinear time series using canonical variate analysis," *Physica D*, vol. 170, no. 2, pp. 103–117, 2002.

[92] E. Bagarinao Jr., T. Nomura, K. Pakdaman, and S. Sato, "Generalized one-parameter bifurcation diagram reconstruction using time series," *Physica D*, vol. 124, no. 1–3, pp. 258–270, 1998.

[93] E. Bagarinao Jr., K. Pakdaman, T. Nomura, and S. Sato, "Time series-based bifurcation diagram reconstruction," *Physica D*, vol. 130, no. 3-4, pp. 211–231, 1999.

[94] E. Bagarinao Jr., K. Pakdaman, T. Nomura, and S. Sato, "Reconstructing bifurcation diagrams from noisy time series using nonlinear autoregressive models," *Physical Review E*, vol. 60, no. 1, pp. 1073–1076, 1999.

[95] E. Bagarinao Jr., K. Pakdaman, T. Nomura, and S. Sato, "Reconstructing bifurcation diagrams of dynamical systems using measured time series," *Methods of Information in Medicine*, vol. 39, no. 2, pp. 146–149, 2000.

[96] M. Barahona and C.-S. Poon, "Detection of nonlinear dynamics in short, noisy time series," *Nature*, vol. 381, no. 6579, pp. 215–217, 1996.

[97] K. H. Chon, J. K. Kanters, R. J. Cohen, and N.-H. Holstein-Rathlou, "Detection of chaotic determinism in time series from randomly forced maps," *Physica D*, vol. 99, no. 4, pp. 471–486, 1997.

[98] L. A. Aguirre, B. O. S. Teixeira, and L. A. B. Tôrres, "Using data-driven discrete-time models and the unscented Kalman filter to estimate unobserved variables of nonlinear systems," *Physical Review E*, vol. 72, no. 2, Article ID 026226, 12 pages, 2005.

[99] N. Wessel, H. Malberg, R. Bauernschmitt, A. Schirdewan, and J. Kurths, "Nonlinear additive autoregressive model-based analysis of short-term heart rate variability," *Medical and Biological Engineering and Computing*, vol. 44, no. 4, pp. 321–330, 2006.

[100] D. A. Vaccari and H.-K. Wang, "Multivariate polynomial regression for identification of chaotic time series," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 13, no. 4, pp. 395–412, 2007.

[101] D. J. de Oliveira, C. Letellier, M. E. D. Gomes, and L. A. Aguirre, "The use of synthetic input sequences in time series modeling," *Physics Letters A*, vol. 372, no. 32, pp. 5276–5282, 2008.

[102] M. C. S. Coelho, E. M. A. M. Mendes, and L. A. Aguirre, "Testing for intracycle determinism in pseudoperiodic time series," *Chaos*, vol. 18, no. 2, Article ID 023125, p. 12, 2008.

[103] M. Lei and G. Meng, "The influence of noise on nonlinear time series detection based on Volterra-Wiener-Korenberg model," *Chaos, Solitons & Fractals*, vol. 36, no. 2, pp. 512–516, 2008.

[104] E. Floriani, T. Dudok de Wit, and P. Le Gal, "Nonlinear interactions in a rotating disk flow: from a Volterra model to the Ginzburg-Landau equation," *Chaos*, vol. 10, no. 4, pp. 834–847, 2000.

[105] M. E. D. Gomes, A. V. P. Souza, H. N. Guimarães, and L. A. Aguirre, "Evidences of determinism in HRV signals," *Chaos*, vol. 10, no. 2, pp. 398–410, 2000.

[106] L. A. Aguirre, C. Letellier, and J. Maquet, "Forecasting the time series of sunspot numbers," *Solar Physics*, vol. 249, no. 1, pp. 103–120, 2008.

[107] S. A. Billings and S. Chen, "Identification of nonlinear rational systems using a prediction-error estimation algorithm," *International Journal of Systems Science*, vol. 20, no. 3, pp. 467–494, 1989.

[108] Q. M. Zhu and S. A. Billings, "Recursive parameter estimation for nonlinear rational models," *Journal of Systems Engineering*, vol. 1, pp. 63–76, 1991.

[109] Q. M. Zhu and S. A. Billings, "Parameter estimation for stochastic nonlinear rational models," *International Journal of Control*, vol. 57, no. 2, pp. 309–333, 1993.

[110] M. V. Corrêa, L. A. Aguirre, and E. M. A. M. Menues, "Modeling chaotic dynamics with discrete nonlinear rational models," *International Journal of Bifurcation and Chaos*, vol. 10, no. 5, pp. 1019–1032, 2000.

[111] O. Ménard, C. Letellier, and G. Gouesbet, "Map modeling by using rational functions," *Physical Review E*, vol. 62, no. 5, pp. 6325–6331, 2000.

[112] D. Allingham, M. West, and A. I. Mees, "Wavelet reconstruction of nonlinear dynamics," *International Journal of Bifurcation and Chaos*, vol. 8, no. 11, pp. 2191–2201, 1998.

[113] S. A. Billings and D. Coca, "Discrete wavelet models for identification and qualitative analysis of chaotic systems," *International Journal of Bifurcation and Chaos*, vol. 9, no. 7, pp. 1263–1284, 1999.

[114] H. L. Wei and S. A. Billings, "Identification and reconstruction of chaotic systems using multiresolution wavelet decompositions," *International Journal of Systems Science*, vol. 35, no. 9, pp. 511–526, 2004.

[115] R. J. G. B. Campello and W. C. do Amaral, "A relational approach for complex system identification," *Controle & Automação*, vol. 10, no. 3, pp. 139–148, 1999.

[116] M. C. M. Teixeira and S. H. Zak, "Stabilizing controller design for uncertain nonlinear systems using fuzzy models," *IEEE Transactions on Fuzzy Systems*, vol. 7, no. 2, pp. 133–142, 1999.

[117] S. Guillaume, "Designing fuzzy inference systems from data: an interpretability-oriented review," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 3, pp. 426–443, 2001.

[118] R. Ballini and F. Gomide, "Heuristic learning in recurrent neural fuzzy networks," *Journal of Intelligent & Fuzzy Systems*, vol. 13, no. 2–4, pp. 63–74, 2002.

[119] H. L. Hiew and C. P. Tsang, "An adaptive fuzzy system for modeling chaos," *Information Sciences*, vol. 81, no. 3-4, pp. 193–212, 1994.

[120] A. A. P. Santos, N. C. A. da Costa Jr., and L. dos Santos Coelho, "Computational intelligence approaches and linear models in case studies of forecasting exchange rates," *Expert Systems with Applications*, vol. 33, no. 4, pp. 816–823, 2007.

[121] L. Le Sceller, C. Letellier, and G. Gouesbet, "Global vector field reconstruction including a control parameter dependence," *Physics Letters A*, vol. 211, no. 4, pp. 211–216, 1996.

[122] M. Casdagli, "A dynamical systems approach to modeling input-output systems," in *Nonlinear Modeling and Forecasting*, M. Casdagli and S. Eubank, Eds., pp. 265–281, Addison Wesley, New York, NY, USA, 1992.

[123] N. F. Hunter, "Application of nonlinear time-series models to driven systems," in *Nonlinear Modeling and Forecasting*, M. Casdagli and S. Eubank, Eds., pp. 467–491, Addison Wesley, New York, NY, USA, 1992.

[124] I. J. Leontaritis and S. A. Billings, "Input-output parametric models for nonlinear systems—part I: deterministic non-linear systems," *International Journal of Control*, vol. 41, no. 2, pp. 303–328, 1985.

[125] L. A. Aguirre and S. A. Billings, "Retrieving dynamical invariants from chaotic data using NARMAX models," *International Journal of Bifurcation and Chaos*, vol. 5, no. 2, pp. 449–474, 1995.

[126] J. Lu, J. Lü, J. Xie, and G. Chen, "Reconstruction of the Lorenz and Chen systems with noisy observations," *Computers & Mathematics with Applications*, vol. 46, no. 8-9, pp. 1427–1434, 2003.

[127] E. M. A. M. Mendes and S. A. Billings, "A note on discretization of nonlinear differential equations," *Chaos*, vol. 12, no. 1, pp. 66–71, 2002.

[128] E. M. A. M. Mendes and C. Letellier, "Displacement in the parameter space versus spurious solution of discretization with large time step," *Journal of Physics A*, vol. 37, no. 4, pp. 1203–1218, 2004.

[129] R. Reed, "Pruning algorithms—a survey," *IEEE Transactions on Neural Networks*, vol. 4, no. 5, pp. 740–747, 1993.

[130] A. C. Tsoia and A. Back, "Discrete time recurrent neural network architectures: a unifying review," *Neurocomputing*, vol. 15, no. 3-4, pp. 183–223, 1997.

[131] R. de Albuquerque Teixeira, A. P. Braga, R. H. C. Takahashi, and R. R. Saldanha, "Improving generalization of MLPs with multi-objective optimization," *Neurocomputing*, vol. 35, no. 1–4, pp. 189–194, 2000.

[132] M. Small and C. K. Tse, "Minimum description length neural networks for time series prediction," *Physical Review E*, vol. 66, no. 6, Article ID 066701, 12 pages, 2002.

[133] D. A. G. Vieira, J. A. Vasconcelos, and W. M. Caminhas, "Controlling the parallel layer perceptron complexity using a multiobjective learning algorithm," *Neural Computing & Applications*, vol. 16, no. 4-5, pp. 317–325, 2007.

[134] B. Feil, J. Abonyi, and F. Szeifert, "Model order selection of nonlinear input-output models—a clustering based approach," *Journal of Process Control*, vol. 14, no. 6, pp. 593–602, 2004.

[135] J. Madár, J. Abonyi, and F. Szeifert, "Genetic programming for the identification of nonlinear input-output models," *Industrial and Engineering Chemistry Research*, vol. 44, no. 9, pp. 3178–3186, 2005.

[136] E. Baake, M. Baake, H. G. Bock, and K. M. Briggs, "Fitting ordinary differential equations to chaotic data," *Physical Review A*, vol. 45, no. 8, pp. 5524–5529, 1992.

[137] G. L. Baker, J. P. Gollub, and J. A. Blackburn, "Inverting chaos: extracting system parameters from experimental data," *Chaos*, vol. 6, no. 4, pp. 528–533, 1996.

[138] V. S. Anishchenko, A. N. Pavlov, and N. B. Janson, "Global reconstruction in the presence of *a priori* information," *Chaos, Solitons & Fractals*, vol. 9, no. 8, pp. 1267–1278, 1998.

[139] G. Rowlands and J. C. Sprott, "Extraction of dynamical equations from chaotic data," *Physica D*, vol. 58, no. 1–4, pp. 251–259, 1992.

[140] A. I. Mees, "Parsimonious dynamical reconstruction," *International Journal of Bifurcation and Chaos*, vol. 3, no. 3, pp. 669–675, 1993.

[141] J. B. Kadtke, J. Brush, and J. Holzfuss, "Global dynamical equations and Lyapunov exponents from noisy chaotic time series," *International Journal of Bifurcation and Chaos*, vol. 3, no. 3, pp. 607–616, 1993.

[142] L. A. Aguirre and S. A. Billings, "Dynamical effects of overparametrization in nonlinear models," *Physica D*, vol. 80, no. 1-2, pp. 26–40, 1995.

[143] E. M. A. M. Mendes, *Identification of nonlinear discrete systems with intelligent structure detection*, Ph.D. thesis, University of Sheffield, Sheffield, UK, 1995.

[144] L. A. Aguirre and E. M. A. M. Mendes, "Global nonlinear polynomial models: structure, term clusters and fixed points," *International Journal of Bifurcation and Chaos*, vol. 6, no. 2, pp. 279–294, 1996.

[145] E. M. A. M. Mendes and S. A. Billings, "On identifying global nonlinear discrete models from chaotic data," *International Journal of Bifurcation and Chaos*, vol. 7, no. 11, pp. 2593–2601, 1997.

[146] E. M. A. M. Mendes and S. A. Billings, "On overparametrization of nonlinear discrete systems," *International Journal of Bifurcation and Chaos*, vol. 8, no. 3, pp. 535–556, 1998.

[147] E. M. A. M. Mendes and S. A. Billings, "An alternative solution to the model structure selection problem," *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, vol. 31, no. 6, pp. 597–608, 2001.

[148] L. A. Aguirre, "Some remarks on structure selection for nonlinear models," *International Journal of Bifurcation and Chaos*, vol. 4, no. 6, pp. 1707–1714, 1994.

[149] L. A. Aguirre and S. A. Billings, "Improved structure selection for nonlinear models based on term clustering," *International Journal of Control*, vol. 62, no. 3, pp. 569–587, 1995.

[150] R. Brown, V. In, and E. R. Tracy, "Parameter uncertainties in models of equivariant dynamical systems," *Physica D*, vol. 102, no. 3-4, pp. 208–226, 1997.

[151] L. A. Aguirre, R. A. M. Lopes, G. F. V. Amaral, and C. Letellier, "Constraining the topology of neural networks to ensure dynamics with symmetry properties," *Physical Review E*, vol. 69, no. 2, Article ID 026701, 11 pages, 2004.

[152] J. D. Bomberger and D. E. Seborg, "Determination of model order for NARX models directly from input-output data," *Journal of Process Control*, vol. 8, no. 5-6, pp. 459–468, 1998.

[153] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, vol. 15 of *Series in Computer Science*, World Scientific, Singapore, 1989.

[154] K. Judd and A. Mees, "On selecting models for nonlinear time series," *Physica D*, vol. 82, no. 4, pp. 426–444, 1995.

[155] S. A. Billings, S. Chen, and M. J. Korenberg, "Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator," *International Journal of Control*, vol. 49, no. 6, pp. 2157–2189, 1989.

[156] L. A. Aguirre and S. A. Billings, "Identification of models for chaotic systems from noisy data: implications for performance and nonlinear filtering," *Physica D*, vol. 85, no. 1-2, pp. 239–258, 1995.

[157] L. A. Aguirre, P. F. Donoso-Garcia, and R. Santos-Filho, "Use of a priori information in the identification of global nonlinear models—a case study using a buck converter," *IEEE Transactions on Circuits and Systems I*, vol. 47, no. 7, pp. 1081–1085, 2000.

[158] C. Letellier and L. A. Aguirre, "Investigating nonlinear dynamics from time series: the influence of symmetries and the choice of observables," *Chaos*, vol. 12, no. 3, pp. 549–558, 2002.

[159] C. Letellier, L. A. Aguirre, and J. Maquet, "How the choice of the observable may influence the analysis of nonlinear dynamical systems," *Communications in Nonlinear Science and Numerical Simulation*, vol. 11, no. 5, pp. 555–576, 2006.

[160] C. J. Cellucci, A. M. Albano, and P. E. Rapp, "Comparative study of embedding methods," *Physical Review E*, vol. 67, no. 6, Article ID 066210, 13 pages, 2003.

[161] D. A. Smirnov, B. P. Bezruchko, and Y. P. Seleznev, "Choice of dynamical variables for global reconstruction of model equations from time series," *Physical Review E*, vol. 65, no. 2, Article ID 026205, 7 pages, 2002.

[162] K. Judd and A. Mees, "Embedding as a modeling problem," *Physica D*, vol. 120, no. 3-4, pp. 273–286, 1998.

[163] U. Parlitz, "Estimating model parameters from time series by autosynchronization," *Physical Review Letters*, vol. 76, no. 8, pp. 1232–1235, 1996.

[164] A. Maybhate and R. E. Amritkar, "Use of synchronization and adaptive control in parameter estimation from a time series," *Physical Review E*, vol. 59, no. 1, pp. 284–293, 1999.

[165] C. Tao, Y. Zhang, G. Du, and J. J. Jiang, "Fitting model equations to time series using chaos synchronization," *Physics Letters A*, vol. 332, no. 3-4, pp. 197–206, 2004.

[166] U. S. Freitas, E. E. N. Macau, and C. Grebogi, "Using geometric control and chaotic synchronization to estimate an unknown model parameter," *Physical Review E*, vol. 71, no. 4, Article ID 047203, 4 pages, 2005.

[167] E. J. Kostelich, "Problems in estimating dynamics from data," *Physica D*, vol. 58, no. 1–4, pp. 138–152, 1992.

[168] L. Jaeger and H. Kantz, "Unbiased reconstruction of the dynamics underlying a noisy chaotic time series," *Chaos*, vol. 6, no. 3, pp. 440–450, 1996.

[169] H. U. Voss, J. Timmer, and J. Kurths, "Nonlinear dynamical system identification from uncertain and indirect measurements," *International Journal of Bifurcation and Chaos*, vol. 14, no. 6, pp. 1905–1933, 2004.

[170] J. M. Fullana, M. Rossi, and S. Zaleski, "Parameter identification in noisy extended systems: a hydrodynamic case," *Physica D*, vol. 103, no. 1–4, pp. 564–575, 1997.

[171] P. Perona, A. Porporato, and L. Ridolfi, "On the trajectory method for the reconstruction of differential equations from time series," *Nonlinear Dynamics*, vol. 23, no. 1, pp. 13–33, 2000.

[172] P. Connally, K. Li, and G. W. Irwin, "Prediction- and simulation-error based perceptron training: solution space analysis and a novel combined training scheme," *Neurocomputing*, vol. 70, no. 4–6, pp. 819–827, 2007.

[173] J. L. Breeden and N. P. Packard, "A learning algorithm for optimal representation of experimental data," *International Journal of Bifurcation and Chaos*, vol. 4, no. 2, pp. 311–326, 1994.

[174] J. Timmer, "Modeling noisy time series: physiological tremor," *International Journal of Bifurcation and Chaos*, vol. 8, no. 7, pp. 1505–1516, 1998.

[175] H. U. Voss, A. Schwache, J. Kurths, and F. Mitschke, "Equations of motion from chaotic data: a driven optical fiber ring resonator," *Physics Letters A*, vol. 256, no. 1, pp. 47–54, 1999.

[176] B. P. Bezruchko, A. S. Karavaev, V. I. Ponomarenko, and M. D. Prokhorov, "Reconstruction of time-delay systems from chaotic time series," *Physical Review E*, vol. 64, no. 5, Article ID 056216, 6 pages, 2001.

[177] D. A. Smirnov, V. S. Vlaskin, and V. I. Ponomarenko, "Estimation of parameters in one-dimensional maps from noisy chaotic time series," *Physics Letters A*, vol. 336, no. 6, pp. 448–458, 2005.

[178] E. G. Nepomuceno, R. H. C. Takahashi, G. F. V. Amaral, and L. A. Aguirre, "Nonlinear identification using prior knowledge of fixed points: a multiobjective approach," *International Journal of Bifurcation and Chaos*, vol. 13, no. 5, pp. 1229–1246, 2003.

[179] L. A. Aguirre, E. C. Furtado, and L. A. B. Tôrres, "Evaluation of dynamical models: dissipative synchronization and other techniques," *Physical Review E*, vol. 74, no. 6, Article ID 066203, 16 pages, 2006.

[180] L. A. Aguirre and S. A. Billings, "Validating identified nonlinear models with chaotic dynamics," *International Journal of Bifurcation and Chaos*, vol. 4, no. 1, pp. 109–125, 1994.

[181] E. Bagarinao and S. Sato, "Algorithm for vector autoregressive model parameter estimation using an orthogonalization procedure," *Annals of Biomedical Engineering*, vol. 30, no. 2, pp. 260–271, 2002.

[182] A. Garulli, C. Mocenni, A. Vicino, and A. Tesi, "Integrating identification and qualitative analysis for the dynamic model of a lagoon," *International Journal of Bifurcation and Chaos*, vol. 13, no. 2, pp. 357–374, 2003.

[183] K. Rodríguez-Vázquez and P. J. Fleming, "Evolution of mathematical models of chaotic systems based on multiobjective genetic programming," *Knowledge and Information Systems*, vol. 8, no. 2, pp. 235–256, 2005.

[184] L. M. Pecora, T. L. Carroll, and J. F. Heagy, "Statistics for mathematical properties of maps between time series embeddings," *Physical Review E*, vol. 52, no. 4, pp. 3420–3439, 1995.

[185] R. Brown, N. F. Rul'kov, and E. R. Tracy, "Modeling and synchronizing chaotic systems from experimental data," *Physics Letters A*, vol. 194, no. 1-2, pp. 71–76, 1994.

[186] G. Gouesbet, S. Meunier-Guttin-Cluzel, and O. Ménard, "Global reconstructions of equations of motion from data series, and validation techniques, a review," in *Chaos and Its Reconstruction*, G. Gouesbet, S. Meunier-Guttin-Cluzel, and O. Ménard, Eds., pp. 1–160, Nova Science, New York, NY, USA, 2003.

[187] D. Coca and S. A. Billings, "Identification of coupled map lattice models of complex spatio-temporal patterns," *Physics Letters A*, vol. 287, no. 1-2, pp. 65–73, 2001.

[188] S. Ishii and M.-A. Sato, "Reconstruction of chaotic dynamics by on-line EM algorithm," *Neural Networks*, vol. 14, no. 9, pp. 1239–1256, 2001.

[189] L. A. Aguirre and Á. V. P. Souza, "Stability analysis of sleep apnea time series using identified models: a case study," *Computers in Biology and Medicine*, vol. 34, no. 4, pp. 241–257, 2004.

[190] J. Maquet, C. Letellier, and L. A. Aguirre, "Scalar modeling and analysis of a 3D biochemical reaction model," *Journal of Theoretical Biology*, vol. 228, no. 3, pp. 421–430, 2004.

[191] O. Ménard, C. Letellier, J. Maquet, L. Le Sceller, and G. Gouesbet, "Analysis of a nonsynchronized sinusoidally driven dynamical system," *International Journal of Bifurcation and Chaos*, vol. 10, no. 7, pp. 1759–1772, 2000.

[192] C. Letellier, L. Le Sceller, P. Dutertre, G. Gouesbet, Z. Fei, and J. L. Hudson, "Topological characterization and global vector field reconstruction of an experimental electrochemical system," *Journal of Physical Chemistry*, vol. 99, no. 18, pp. 7016–7027, 1995.

[193] N. B. Tufillaro, P. Wyckoff, R. Brown, T. Schreiber, and T. Molteno, "Topological time-series analysis of a string experiment and its synchronized model," *Physical Review E*, vol. 51, no. 1, pp. 164–174, 1995.

[194] C. Letellier and G. Gouesbet, "Topological characterization of reconstructed attractors modding out symmetries," *Journal de Physique II*, vol. 6, no. 11, pp. 1615–1638, 1996.

[195] M. Small and K. Judd, "Comparisons of new nonlinear modeling techniques with applications to infant respiration," *Physica D*, vol. 117, no. 1–4, pp. 283–298, 1998.

[196] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D*, vol. 58, no. 1–4, pp. 77–94, 1992.

[197] J. Timmer, "Power of surrogate data testing with respect to nonstationarity," *Physical Review E*, vol. 58, no. 4, pp. 5153–5156, 1998.

[198] J. P. Barnard, C. Aldrich, and M. Gerber, "Identification of dynamic process systems with surrogate data methods," *AIChE Journal*, vol. 47, no. 9, pp. 2064–2075, 2001.

[199] C. Letellier, T. D. Tsankov, G. Byrne, and R. Gilmore, "Large-scale structural reorganization of strange attractors," *Physical Review E*, vol. 72, no. 2, Article ID 026212, 12 pages, 2005.

[200] P. E. McSharry and L. A. Smith, "Consistent nonlinear dynamics: identifying model inadequacy," *Physica D*, vol. 192, no. 1-2, pp. 1–22, 2004.

[201] R. Brown, N. F. Rul'kov, and E. R. Tracy, "Modeling and synchronizing chaotic systems from time-series data," *Physical Review E*, vol. 49, no. 5, pp. 3784–3800, 1994.

[202] C. Letellier, O. Ménard, and L. A. Aguirre, "Validation of selected global models," in *Modeling and Forecasting Financial Data: Techniques of Nonlinear Dynamics*, A. S. Soofi and L. Cao, Eds., pp. 283–302, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.

[203] C. Sarasola, F. J. Torrealdea, A. d'Anjou, and M. Graña, "Cost of synchronizing different chaotic systems," *Mathematics and Computers in Simulation*, vol. 58, no. 4–6, pp. 309–327, 2002.

[204] R. Gilmore and M. Lefranc, *The Topology of Chaos: Alice in Stretch and Squeezelan*, Wiley-Interscience, New York, NY, USA, 2002.

[205] C. Letellier, J. Maquet, L. Le Sceller, G. Gouesbet, and L. A. Aguirre, "On the non-equivalence of observables in phase-space reconstructions from recorded time series," *Journal of Physics A*, vol. 31, no. 39, pp. 7913–7927, 1998.

[206] H. J. A. F. Tulleken, "Grey-box modelling and identification using physical knowledge and Bayesian techniques," *Automatica*, vol. 29, no. 2, pp. 285–308, 1993.

[207] E. Eskinat, S. H. Johnson, and W. L. Luyben, "Use of auxiliary information in system identification," *Industrial and Engineering Chemistry Research*, vol. 32, no. 9, pp. 1981–1992, 1993.

[208] T. A. Johansen, "Identification of non-linear systems using empirical data and prior knowledge—an optimization approach," *Automatica*, vol. 32, no. 3, pp. 337–356, 1996.

[209] M. V. Corrêa, L. A. Aguirre, and R. R. Saldanha, "Using steady-state prior knowledge to constrain parameter estimates in nonlinear system identification," *IEEE Transactions on Circuits and Systems I*, vol. 49, no. 9, pp. 1376–1381, 2002.

[210] L. A. Aguirre, M. F. S. Barroso, R. R. Saldanha, and E. M. A. M. Mendes, "Imposing steady-state performance on identified nonlinear polynomial models by means of constrained parameter estimation," *IEE Proceedings: Control Theory and Applications*, vol. 151, no. 2, pp. 174–179, 2004.

[211] P. E. Rapp, T. I. Schmah, and A. I. Mees, "Models of knowing and the investigation of dynamical systems," *Physica D*, vol. 132, no. 1-2, pp. 133–149, 1999.

[212] G. Byrne, R. Gilmore, and C. Letellier, "Distinguishing between folding and tearing mechanisms in strange attractors," *Physical Review E*, vol. 70, no. 5, Article ID 056214, 9 pages, 2004.

[213] G. F. V. Amaral, C. Letellier, and L. A. Aguirre, "Piecewise affine models of chaotic attractors: the Rössler and Lorenz systems," *Chaose*, vol. 16, no. 1, Article ID 013115, 14 pages, 2006.

[214] R. Gilmore, "Summary of the second workshop on measures of complexity and chaos," *International Journal of Bifurcation and Chaos*, vol. 3, no. 3, pp. 491–524, 1993.

[215] R. M. May, "Deterministic models with chaotic dynamics," *Nature*, vol. 256, no. 5514, pp. 165–166, 1975.

[216] R. M. May, "Simple mathematical models with very complicated dynamics," *Nature*, vol. 261, no. 5560, pp. 459–467, 1976.

[217] M. Hénon, "A two-dimensional mapping with a strange attractor," *Communications in Mathematical Physics*, vol. 50, no. 1, pp. 69–77, 1976.

[218] M. Xu, G. Chen, and Y.-T. Tian, "Identifying chaotic systems using Wiener and Hammerstein cascade models," *Mathematical and Computer Modelling*, vol. 33, no. 4-5, pp. 483–493, 2001.

[219] K. Ikeda, "Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system," *Optics Communications*, vol. 30, no. 2, pp. 257–261, 1979.

[220] O. E. Rössler, "An equation for continuous chaos," *Physics Letters A*, vol. 57, no. 5, pp. 397–398, 1976.

[221] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, 1977.

[222] B. van der Pol and M. van der Mark, "The heartbeat considered as a relaxation oscillation, and an electrical model of the heart," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, Series 7*, vol. 6, pp. 763–775, 1928, Pl 10–12.

[223] N. A. Gershenfeld and A. A. Weigend, "The future of time series: learning and understanding," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. A. Weigend and N. A. Gershenfeld, Eds., pp. 1–70, Addison-Wesley, New York, NY, USA, 1994.

[224] D. R. Rigney, A. L. Goldberger, W. C. Ocasio, Y. Ichimaru, G. B. Moody, and R. G. Mark, "Multi-channel physiological data: description and analysis," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. A. Weigend and N. A. Gershenfeld, Eds., pp. 105–129, Addison-Wesley, New York, NY, USA, 1994.

[225] A. A. Weigend and N. A. Gershenfeld, *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley, New York, NY, USA, 1994.

[226] M. Casdagli and A. A. Weigend, "Exploring the continuum between deterministic and stochastic modeling," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. A. Weigend and N. A. Gershenfeld, Eds., pp. 347–366, Addison-Wesley, New York, NY, USA, 1994.

[227] D. Kaplan, "A geometrical statistic for detecting deterministic dynamics," in *Time Series Prediction: Forecasting the Future and Understanding the Past*, A. A. Weigend and N. A. Gershenfeld, Eds., pp. 415–428, Addison-Wesley, New York, NY, USA, 1994.

[228] M. Paluš, "Detecting nonlinearity in multivariate time series," *Physics Letters A*, vol. 213, no. 3-4, pp. 138–147, 1996.

[229] L. A. Aguirre, V. C. Barros, and Á. V. P. Souza, "Nonlinear multivariable modeling and analysis of sleep apnea time series," *Computers in Biology and Medicine*, vol. 29, no. 3, pp. 207–228, 1999.

[230] L. Cao and A. Mees, "Deterministic structure in multichannel physiological data," *International Journal of Bifurcation and Chaos*, vol. 10, no. 12, pp. 2767–2780, 2000.

[231] C. S. M. Lainscsek, F. Schürrer, and J. B. Kadtke, "A general form for global dynamical data models for three-dimensional systems," *International Journal of Bifurcation and Chaos*, vol. 8, no. 5, pp. 899–914, 1998.

[232] C. Letellier, L. A. Aguirre, J. Maquet, and R. Gilmore, "Evidence for low dimensional chaos in sunspot cycles," *Astronomy & Astrophysics*, vol. 449, no. 1, pp. 379–387, 2006.

[233] J. Maquet, C. Letellier, and L. A. Aguirre, "Global models from the Canadian lynx cycles as a direct evidence for chaos in real ecosystems," *Journal of Mathematical Biology*, vol. 55, no. 1, pp. 21–39, 2007.

[234] N. F. Rul'kov, A. R. Volkovskiĭ, A. Rodríguez-Lozano, E. del Río, and M. G. Velarde, "Mutual synchronization of chaotic self-oscillators with dissipative coupling," *International Journal of Bifurcation and Chaos*, vol. 2, no. 3, pp. 669–676, 1992.

[235] T. Matsumoto, L. O. Chua, and M. Komuro, "The double scroll," *IEEE Transactions on Circuits and Systems*, vol. 32, no. 8, pp. 797–818, 1985.

[236] L. A. Aguirre, G. G. Rodrigues, and E. M. A. M. Mendes, "Nonlinear identification and cluster analysis of chaotic attractors from a real implementation of Chua's circuit," *International Journal of Bifurcation and Chaos*, vol. 7, no. 6, pp. 1411–1423, 1997.

[237] B. Cannas, S. Cincotti, M. Marchesi, and F. Pilo, "Learning of Chua's circuit attractors by locally recurrent neural networks," *Chaos, Solitons & Fractals*, vol. 12, no. 11, pp. 2109–2115, 2001.

Review Article

Flow Around a Slender Circular Cylinder: A Case Study on Distributed Hopf Bifurcation

J. A. P. Aranha, K. P. Burr, I. C. Barbeiro, I. Korkischko, and J. R. Meneghini

Nucleus of Dynamics and Fluids (NDF), Mechanical Engineering, University of Sao Paulo, Sao Paulo, Brazil

Correspondence should be addressed to J. A. P. Aranha, japaran@usp.br

This paper presents a short overview of the flow around a slender circular cylinder, the purpose being to place it within the frame of the distributed Hopf bifurcation problems described by the Ginzburg-Landau equation (GLE). In particular, the chaotic behavior superposed to a well tuned harmonic oscillation observed in the range Re > 270, with Re being the Reynolds number, is related to the defect-chaos regime of the GLE. Apparently new results, related to a Kolmogorov like length scale and the *rms* of the response amplitude, are derived in this defect-chaos regime and further related to the experimental *rms* of the lift coefficient measured in the range Re > 270.

1. Flow Around a Circular Cylinder: An Overview

Let a cylinder with a circular cross section in the plane $\mathbf{x} = (x, y)$ exposed to an incident flow $U\mathbf{i}$; if d is the circle diameter and ν the kinematic fluid viscosity, the Reynolds number is defined by $\mathrm{Re} = Ud/\nu$. Assuming a unit system where $\rho = U = d = 1$, with ρ being the fluid density, the two-dimensional (2D) velocity and pressure fields, respectively $\{\mathbf{u}(\mathbf{x}, t); p(\mathbf{x}, t)\}$, satisfy the Navier-Stokes equations,

$$
\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} - \frac{1}{\mathrm{Re}} \nabla^2 \mathbf{u} + \nabla p = \mathbf{0};
$$

$$
\nabla \cdot \mathbf{u} = 0, \quad \left(\nabla = \mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} \right),
$$

(1.1)

Re = 102

Re = 161

(a)   (b)
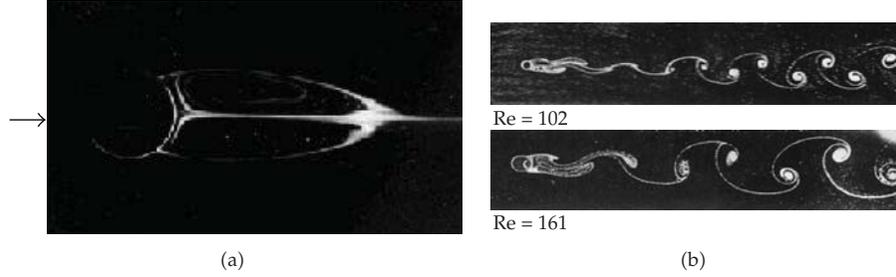
Figure 1: Flow around circular cylinder: (a) *steady state regime*, Re = 40 < 46.5; (b) *stable 2D limit cycles*, 46.5 < Re < 180 (Re = 102 above, Re = 161, below). (source: Van Dyke [1]).

and the boundary conditions (∂V_c: cylinder cross section)

$$\mathbf{u}(\mathbf{x},t)|_{\mathbf{x}\in\partial V_c} = \mathbf{0};$$

$$\lim_{\|\mathbf{x}\|\to\infty}\left\{\mathbf{u}(\mathbf{x},t);p(\mathbf{x},t)\right\} = \{U\mathbf{i};0\}. \tag{1.2}$$

Equations (1.1) and (1.2) has a *steady solution* $\mathbf{u}_s(\mathbf{x})$ that is however *stable* only in the range Re < 46.5; for Re > 46.5 a *limit cycle solution*, oscillating with the *Strouhal frequency* $\omega_s \approx U/d$, is observed. This limit cycle is stable in the 2D context—namely, if the perturbation is restricted to the plane $\mathbf{x} = (x,y)$—in a large range of Reynolds numbers and Figure 1 shows typical flow visualizations in the steady (Re = 40) and limit cycles regimes (Re = 102;161).

The periodic flow in the limit cycle regime can be expanded in its harmonic components by Fourier series decomposition; namely, if $\mathbf{u}(\mathbf{x},t) = u(\mathbf{x},t)\mathbf{i} + v(\mathbf{x},t)\mathbf{j}$ is the flow field then

$$u(\mathbf{x},t) = u_o(\mathbf{x}) + \sum_{n=1}^{\infty}\left[u_{n,c}(\mathbf{x})\cdot\cos(n\omega_s t) + u_{n,s}(\mathbf{x})\cdot\sin(n\omega_s t)\right];$$

$$v(\mathbf{x},t) = v_o(\mathbf{x}) + \sum_{n=1}^{\infty}\left[v_{n,c}(\mathbf{x})\cdot\cos(n\omega_s t) + v_{n,s}(\mathbf{x})\cdot\sin(n\omega_s t)\right]. \tag{1.3}$$

The time average $\mathbf{u}_o(\mathbf{x}) = u_o(\mathbf{x})\mathbf{i} + v_o(\mathbf{x})\mathbf{j}$ of $\mathbf{u}(\mathbf{x},t)$ is a flow field *symmetric* with respect to the x-axis, with $u_o(\mathbf{x})$ being an *even* function of $y(u_o(x,y) = u_o(x,-y))$ and $v_o(\mathbf{x})$ an *odd* one ($v_o(x,y) = -v_o(x,-y)$); the first harmonic $\mathbf{u}_1(\mathbf{x}) = u_1(\mathbf{x})\mathbf{i} + v_1(\mathbf{x})\mathbf{j}$ is an *anti-symmetric* field, with $\{u_1(x,y) = -u_1(x,-y); v_1(x,y) = v_1(x,-y)\}$ and, as a rule, one can show for a circular cylinder that the even harmonics are symmetric and the odd ones anti-symmetric: the *in-line force* (drag) depends thus only on the *even modes* while the *transverse force* (lift) depends only on the *odd modes*. Figure 2 displays, for Re = 100, the functions $\{(u_{1,c}(\mathbf{x}),v_{1,c}(\mathbf{x})); (u_{2,c}(\mathbf{x}),v_{2,c}(\mathbf{x}))\}$ obtained both from numerical simulation of the 2D flow and from PIV measurement of an actual flow experiment. Both plots are visually very similar and confirm the symmetry/anti-symmetry behavior quoted above; furthermore, the agreement between them is also quantitative: defining the *normalized mode amplitude* by the ratio $A_n(\text{Re}) = \max|u_n(\mathbf{x})|/\max|u_o(\mathbf{x})|$, Figure 3 displays the function $A_n(\text{Re})$ determined numerically in the range $60 \le \text{Re} \le 600$ and also the same value obtained experimentally at Re = 100.
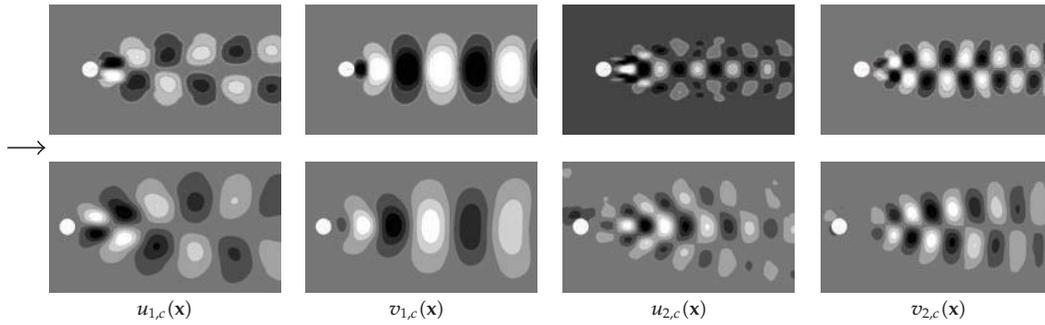
Figure 2: Harmonic decomposition of the 2D velocity field—Re = 100. *Above:* numerical simulation; *below:* experiments (PIV). (Source: Barbeiro & Korkischko (2008)—NDF).
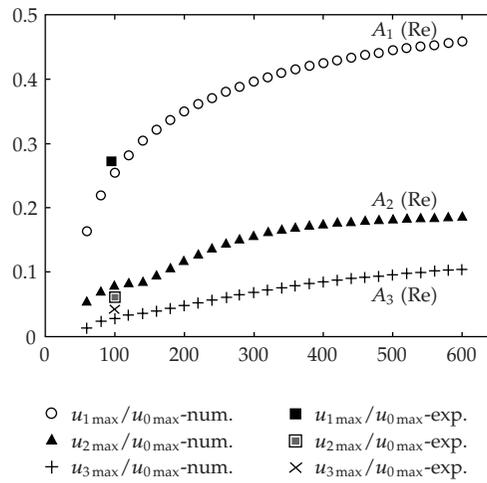
Figure 3: Amplitude $A_n(\text{Re}) = = \max|u_n(\mathbf{x})|/\max|u_o(\mathbf{x})|$ of the nth mode. Numerical computation (for Re = 100, also shown PIV computation) (source: Barbeiro & Korkischko (2008)—NDF).

It must be observed also that these numerical results indicate a *hierarchy* between the modes amplitudes and a *square root singularity* near $\mathbf{R}_{c1} \cong 46.5$, namely

$$A_n(\text{Re}) \cong \mathcal{O}(\varepsilon^n);$$

$$A_1(\text{Re}) \cong \mathcal{O}(\varepsilon) \quad \text{with } \varepsilon(\text{Re}) \cong 0.45\sqrt{1 - \frac{\mathbf{R}_{c1}}{\text{Re}}}, \tag{1.4}$$

typical of a *Hopf supercritical bifurcation*. This point will be elaborated in the following; in fact, if one writes (1.3) in the complex form,

$$\mathbf{u}_n(\mathbf{x}) = \frac{1}{2}\left(\mathbf{u}_{n,c}(\mathbf{x}) - i\mathbf{u}_{n,s}(\mathbf{x})\right);$$

$$\mathbf{u}(\mathbf{x},t) = \mathbf{u}_o(\mathbf{x}) + \sum_{n=1}^{\infty}\left(\mathbf{u}_n(\mathbf{x}) \cdot e^{in\omega_s t} + (*)\right), \quad \left(\lim_{\|\mathbf{x}\|\to\infty}\left\{\begin{array}{c}\mathbf{u}_o(\mathbf{x}) \\ \mathbf{u}_n(\mathbf{x}); \ n > 0\end{array}\right\} = \left\{\begin{array}{c}U\mathbf{i} \\ \mathbf{0}\end{array}\right\}\right), \tag{1.5}$$

placing (1.5) into (1.1) and separating the harmonic parcels $\{\exp(in\omega_s t); n = 0, 1, 2, \ldots\}$ one obtains the sequence of problems,

$$\mathbf{n} = \mathbf{0} : (\mathbf{u}_o \cdot \nabla)\mathbf{u}_o - \frac{1}{\mathrm{Re}}\nabla^2\mathbf{u}_o + \nabla p_o = \mathbf{f}_o(\mathbf{x})$$

$$= -\sum_{n=1}^{\infty}\left[(\mathbf{u}_n \cdot \nabla)\mathbf{u}_n^* + (\mathbf{u}_n^* \cdot \nabla)\mathbf{u}_n\right] \cong \mathcal{O}(\varepsilon^2);$$

$$\mathbf{n} = \mathbf{1} : i\omega_s\mathbf{u}_1 + \left[(\mathbf{u}_o \cdot \nabla)\mathbf{u}_1 + (\mathbf{u}_1 \cdot \nabla)\mathbf{u}_o\right] - \frac{1}{\mathrm{Re}}\nabla^2\mathbf{u}_1 + \nabla p_1 = \mathbf{f}_1(\mathbf{x})$$

$$= -\sum_{n=1}^{\infty}\left[(\mathbf{u}_{n+1} \cdot \nabla)\mathbf{u}_n^* + (\mathbf{u}_n^* \cdot \nabla)\mathbf{u}_{n+1}\right] \cong \mathcal{O}(\varepsilon^3); \tag{1.6}$$

$$\mathbf{n} = \mathbf{2} : 2i\omega_s\mathbf{u}_2 + \left[(\mathbf{u}_o \cdot \nabla)\mathbf{u}_2 + (\mathbf{u}_2 \cdot \nabla)\mathbf{u}_o\right] - \frac{1}{\mathrm{Re}}\nabla^2\mathbf{u}_2 + \nabla p_2 = \mathbf{f}_2(\mathbf{x})$$

$$= -(\mathbf{u}_1 \cdot \nabla)\mathbf{u}_1 - \sum_{n=1}^{\infty}\left[(\mathbf{u}_{n+2} \cdot \nabla)\mathbf{u}_n^* + (\mathbf{u}_n^* \cdot \nabla)\mathbf{u}_{n+2}\right] \cong \mathcal{O}(\varepsilon^2); \quad (\nabla \cdot \mathbf{u}_n = 0),$$

the first one, that determines $\mathbf{u}_o(\mathbf{x})$, being nonlinear and the remaining ones linear, as usual in an asymptotic expansion. In fact, if higher order terms in the "small parameter" ε are disregarded, one may express, to leading order, the field $\mathbf{u}_1(\mathbf{x})$ in the form

$$\mathbf{u}_1(\mathbf{x}) = a_{\mathrm{av}} \cdot \mathbf{e}_{\mathrm{av}}(\mathbf{x}) + \mathcal{O}(\varepsilon^3);$$

$$i\omega_s\mathbf{e}_{\mathrm{av}} + \left[(\mathbf{u}_o \cdot \nabla)\mathbf{e}_{\mathrm{av}} + (\mathbf{e}_{\mathrm{av}} \cdot \nabla)\mathbf{u}_o\right] - \frac{1}{\mathrm{Re}}\nabla^2\mathbf{e}_{\mathrm{av}} + \nabla p_{\mathrm{av}} = \mathbf{0}, \quad (\nabla \cdot \mathbf{e}_{\mathrm{av}} = 0), \tag{1.7}$$

with $\{\lambda_{\mathrm{av}} = i\omega_s; \mathbf{e}_{\mathrm{av}}(\mathbf{x})\}$ being the eigenvalue—eigenvector of the homogeneous problem defined in (1.7): this is consistent with a numerical result due to Barkley [2], stating that the *averaged flow* $\mathbf{u}_o(\mathbf{x})$ is *marginally stable* (*Real* $\lambda_{\mathrm{av}} = 0$) with respect to 2D perturbation.

But this is not enough for the present purpose: the final goal is to solve the tridimensional (3D) problem for a *slender* cylinder having to solve basically the 2D cross section problem: besides the obvious economy in the degrees of freedom needed in the numerical computation, the 2D flow is well organized (*laminar*) while the 3D one is chaotic (*turbulent*), as it will be seen later in this paper. As discussed in Aranha [3], the flow around a slender cylinder can be asymptotically approximated by the *Ginzburg-Landau equation* but one needs then, first of all, to express the harmonic mode $\mathbf{u}_1(\mathbf{x}) \cdot \exp(i\omega_s t)$ in the form $a(t) \cdot \mathbf{e}(\mathbf{x})$, as in (1.7), with $a(t)$ satisfying *Landau's equation*

$$\frac{da}{dt} - \sigma a + \mu(1 - ic_3)|a|^2 a = 0;$$

$$\varepsilon = \sqrt{\frac{\sigma}{\mu}}, \quad (\lambda = \sigma + i\omega). \tag{1.8}$$

The hope is that such $a(t) \cdot \mathbf{e}(\mathbf{x})$, with the related eigenvalue-eigenvector $\{\lambda(\mathrm{Re}) = \sigma + i\omega; \mathbf{e}(\mathbf{x}; \mathrm{Re})\}$, coalesce with the standard *Hopf bifurcation* in the limit $\varepsilon \to 0$ ($\mathrm{Re} \to \mathbf{R}_{c1}$), while

recovering the *Fourier series expansion* (1.5) when ε^2 is "small" but finite in the range $\text{Re} \gg \mathbf{R}_{c1}$. This double requisite obliges to look for a *basic stationary flow* that is neither the *steady solution* $\mathbf{u}_s(\mathbf{x})$, useless far from the bifurcation, nor the *averaged flow* $\mathbf{u}_o(\mathbf{x})$, always marginally stable and so unsuited to describe a Hopf bifurcation.

A clue is given by the following observation: the *steady state solution* $\mathbf{u}_s(\mathbf{x}; \text{Re})$, that becomes unstable at $\text{Re} = \mathbf{R}_{c1}$, satisfies the *homogeneous* 0th-order equation (1.6); with $\mathbf{f}_o(\mathbf{x})$ defined in (1.6), if one considers instead of $\mathbf{u}_s(\mathbf{x}; \text{Re})$ the field

$$(\widehat{\mathbf{u}}_o \cdot \nabla)\widehat{\mathbf{u}}_o - \frac{1}{\text{Re}}\nabla^2\widehat{\mathbf{u}}_o + \nabla\widehat{p}_o = \mathbf{f}_o(\mathbf{x}; \text{Re}) - \Delta\mathbf{f}_o(\mathbf{x}; \text{Re}) \cong \mathcal{O}(\varepsilon^4);$$

$$\Delta\mathbf{f}_o(\mathbf{x}; \text{Re}) = -\left[(\mathbf{u}_1^* \cdot \nabla)\mathbf{u}_1 + (\mathbf{u}_1 \cdot \nabla)\mathbf{u}_1^*\right] \cong \mathcal{O}(\varepsilon^2), \quad (\widehat{\mathbf{u}}_o - \mathbf{u}_s \cong \mathcal{O}(\varepsilon^4)),$$

(1.9)

the stability of this flow coalesce with the one related to the *steady state solution* $\mathbf{u}_s(\mathbf{x}; \text{Re})$ with an error of order ε^4 in the limit $\sigma \to 0$ ($\text{Re} \to \mathbf{R}_{c1}$): in Hopf bifurcation one has $\sigma(\text{Re}) = \alpha \cdot (1 - \mathbf{R}_{c1}/\text{Re}) \approx \varepsilon^2$ when $\text{Re} \to \mathbf{R}_{c1}$ and this relation is recovered if the field defined in (1.9) is used, instead of the standard steady state solution $\mathbf{u}_s(\mathbf{x}; \text{Re})$, as the basic field.

At $\mathbf{R}_{c1}$ one has strictly $\mathbf{u}_s(\mathbf{x}; \mathbf{R}_{c1}) \equiv \mathbf{u}_o(\mathbf{x}; \mathbf{R}_{c1})$; however, as Re increases the *steady state solution* $\mathbf{u}_s(\mathbf{x}; \text{Re})$ presents a bulbous region in the wake, similar to the one indicated in Figure 1 but with a length increasing linearly with Re: the difference between $\mathbf{u}_s(\mathbf{x}; \text{Re})$ and $\mathbf{u}_o(\mathbf{x}; \text{Re})$ becomes enormous in the range $\text{Re} > 200$, in despite of the fact that the forcing term $\mathbf{f}_o(\mathbf{x})$ in the problem that defines $\mathbf{u}_o(\mathbf{x}; \text{Re})$ be small, of order ε^2. This apparent paradox is in fact due to an extreme susceptibility of the *steady flow* $\mathbf{u}_s(\mathbf{x}; \text{Re})$ to the influence of "small forces", either applied directly, as $\mathbf{f}_o(\mathbf{x})$, or else indirectly, as the constraint forces that appear on the outer contour of the finite domain used in the numerical computation; for example, to determine numerically $\mathbf{u}_s(\mathbf{x})$ at $\text{Re} = 600$ with reasonable accuracy one needs to discretize a circle with radius $1000d$: only then the "small constraint forces" in the outer circle becomes small enough to not impair convergence. In the other hand, the presence of the small forcing term $\mathbf{f}_o(\mathbf{x})$ seems to regularize the problem, since then the *time average* field $\mathbf{u}_o(\mathbf{x}; \text{Re})$ is robust: it can be easily determined numerically, without any major concern about the region size to be discretized, and it also changes weakly with the Reynolds number.

In the asymptotic solution that leads to Landau's equation (1.8) terms of order ε^4 are ignored and the fields $\{\widehat{\mathbf{u}}_o; \widehat{p}_o\}$ can be thus determined by solving the *regular* linear system

$$\left.\begin{array}{l} \widehat{\mathbf{u}}_o = \mathbf{u}_o - \delta\mathbf{u}; \\ \widehat{p}_o = p_o - \delta p; \end{array}\right\} \implies \left[(\mathbf{u}_o \cdot \nabla)\delta\mathbf{u} + (\delta\mathbf{u} \cdot \nabla)\mathbf{u}_o\right] - \frac{1}{\text{Re}}\nabla^2(\delta\mathbf{u}) + \nabla(\delta p) = \Delta\mathbf{f}_o \cong \mathcal{O}(\varepsilon^2),$$

(1.10)

where the term $(\delta\mathbf{u} \cdot \nabla)\delta\mathbf{u} \cong \mathcal{O}(\varepsilon^4)$ was disregarded; notice that the linear operator (1.10) is *regular* since its eigenvalue λ with largest real part is given by $\{\text{Real }\lambda = 0; \text{Imag }\lambda = \omega_s \cong \mathcal{O}(1)\}$, see (1.7).

The eigenvalue-eigenvector $\{\lambda(\text{Re}) = \sigma + i\omega; \mathbf{e}(\mathbf{x}; \text{Re})\}$ corresponding to the basic flow defined in (1.9) is solution of the problem

$$\lambda\mathbf{e} + \left[(\widehat{\mathbf{u}}_o \cdot \nabla)\mathbf{e} + (\mathbf{e} \cdot \nabla)\widehat{\mathbf{u}}_o\right] - \frac{1}{\text{Re}}\nabla^2\mathbf{e} + \nabla p_e = \mathbf{0};$$

$$\lambda = \sigma + i\omega \quad \text{with } \sigma(\text{Re}) > 0 \quad \text{if } \text{Re} > \mathbf{R}_{c1}, \quad (\nabla \cdot \mathbf{e} = 0),$$

(1.11)

and since $\{\widehat{\mathbf{u}}_o - \mathbf{u}_o; \widehat{p}_o - p_o\} \cong \mathcal{O}(\varepsilon^2)$, comparing (1.7) to (1.11) one obtains

$$\{\mathbf{e} - \mathbf{e}_{\mathrm{av}}; p_e - p_{\mathrm{av}}\} \cong \mathcal{O}(\varepsilon^2);$$

$$\sigma + i(\omega - \omega_s) \cong \mathcal{O}(\varepsilon^2). \tag{1.12}$$

Observing that the harmonic components $\{\mathbf{u}_n(\mathbf{x}); n > 0\}$ in (1.5) tend to zero as $\|\mathbf{x}\| \to \infty$, and so it does the functions $\{\mathbf{f}_o(\mathbf{x}); \Delta\mathbf{f}_o(\mathbf{x})\}$, one considers now the solution of

$$\frac{\partial\widehat{\mathbf{u}}}{\partial t} + (\widehat{\mathbf{u}} \cdot \nabla)\widehat{\mathbf{u}} - \frac{1}{\mathrm{Re}}\nabla^2\widehat{\mathbf{u}} + \nabla\widehat{p} = \mathbf{f}_o(\mathbf{x}; \mathrm{Re}) - \Delta\mathbf{f}_o(\mathbf{x}; \mathrm{Re}) \cong \mathcal{O}(\varepsilon^4);$$

$$\nabla \cdot \widehat{\mathbf{u}} = 0, \tag{1.13}$$

satisfying the same boundary conditions (1.2). The *steady state solution* of (1.13), defined in (1.9), becomes unstable for $\mathrm{Re} > \mathbf{R}_{c1}$, the only *unstable mode* being given by $\{\lambda(\mathrm{Re}) = \sigma + i\omega; \mathbf{e}(\mathbf{x}; \mathrm{Re})\}$, solution of (1.11); the solution of (1.13) in the unstable range $\mathrm{Re} > \mathbf{R}_{c1}$ can be thus expressed by means of the standard asymptotic series

$$\widehat{\mathbf{u}}(\mathbf{x}, t) = \widehat{\mathbf{u}}_o(\mathbf{x}) + \left[a(t) \cdot \mathbf{e}(\mathbf{x})e^{i\omega t} + (*)\right] + \left[|a(t)|^2 \cdot \mathbf{u}_{20}(\mathbf{x}) + \left(a^2(t) \cdot \mathbf{u}_{22}(\mathbf{x})e^{2i\omega t} + (*)\right)\right]$$

$$+ \left[|a(t)|^2 a(t) \cdot \mathbf{u}_{31}(\mathbf{x})e^{i\omega t} + a^3(t) \cdot \mathbf{u}_{33}(\mathbf{x})e^{3i\omega t}\right] + \mathcal{O}(\varepsilon^4);$$

$$\widehat{p}(\mathbf{x}, t) = \widehat{p}_o(\mathbf{x}) + \left[a(t) \cdot p_e(\mathbf{x})e^{i\omega t} + (*)\right] + \left[|a(t)|^2 \cdot p_{20}(\mathbf{x}) + \left(a^2(t) \cdot p_{22}(\mathbf{x})e^{2i\omega t} + (*)\right)\right] \tag{1.14}$$

$$+ \left[|a(t)|^2 a(t) \cdot p_{31}(\mathbf{x})e^{i\omega t} + a^3(t) \cdot p_{33}(\mathbf{x})e^{3i\omega t}\right] + \mathcal{O}(\varepsilon^4);$$

$$a \cong \mathcal{O}(\varepsilon); \qquad \frac{da}{dt} \cong \mathcal{O}(\varepsilon^3); \qquad \{(\widehat{\mathbf{u}}_o, \widehat{p}_o); (\mathbf{e}, p_e); (\mathbf{u}_{\alpha\beta}, p_{\alpha\beta})\} \cong \mathcal{O}(1),$$

where (*) stands for the complex conjugate of the expression in the left and, as usual, the *mode amplitude* $a(t)$ is assumed to change slowly in time, the slow time being proportional to the amplitude square.

By placing (1.14) into (1.13) and separating terms of like orders in ε, a sequence of *linear problems* is obtained, allowing to compute the fields $\{\mathbf{u}_{\alpha\beta}(\mathbf{x}); p_{\alpha\beta}(\mathbf{x})\}$. Details will be omitted here but two points must be commented. First, the operator that determines $\mathbf{u}_{31}(\mathbf{x})$ is exactly the one defined in (1.11) and it is thus singular: the solvability condition (*Fredholm alternative*) of this $\mathbf{u}_{31}$-equation leads to *Landau's equation* (1.8); second, for future reference, the field $\mathbf{u}_{20}(\mathbf{x})$ is solution of the equation

$$\left[(\mathbf{u}_o \cdot \nabla)\mathbf{u}_{20} + (\mathbf{u}_{20} \cdot \nabla)\mathbf{u}_o\right] - \frac{1}{\mathrm{Re}}\nabla^2\mathbf{u}_{20} + \nabla p_{20} = \mathbf{f}_{20};$$

$$\mathbf{f}_{20} = -\left[(\mathbf{e}^* \cdot \nabla)\mathbf{e} + (\mathbf{e} \cdot \nabla)\mathbf{e}^*\right], \tag{1.15}$$

where the relation $(\widehat{\mathbf{u}}_o - \mathbf{u}_o) \cdot |a(t)|^2 \mathbf{u}_{20} \cong \mathcal{O}(\varepsilon^4)$ was used.

The 2D systems (1.1) and (1.13) have both the same singularity at $\mathbf{R}_{c1} \cong 46.5$ and are both regular in "all range" $\mathrm{Re} > \mathbf{R}_{c1}$, a result numerically confirmed by Henderson [4] up

to Re = 1000; since one system differ from the other only by a forcing term of order ε^4, one should have asymptotically

$$\mathbf{u}(\mathbf{x},t) = \widehat{\mathbf{u}}(\mathbf{x},t) + \mathcal{O}(\varepsilon^4);$$

$$p(\mathbf{x},t) = \widehat{p}(\mathbf{x},t) + \mathcal{O}(\varepsilon^4),$$

(1.16)

a result that will be explored next. In fact, recalling that $\{\widehat{\mathbf{u}}_o - \mathbf{u}_o; \widehat{p}_o - p_o\} \cong \mathcal{O}(\varepsilon^2)$ one has, with the help of (1.16),

$$\left.\begin{array}{l} \mathbf{u}(\mathbf{x},t) = \mathbf{u}_o(\mathbf{x}) + \left[\mathbf{u}_1(\mathbf{x})e^{i\omega_s t} + (*)\right] + \mathcal{O}(\varepsilon^2); \\ \widehat{\mathbf{u}}(\mathbf{x},t) = \mathbf{u}_o(\mathbf{x}) + \left[a(t) \cdot \mathbf{e}(\mathbf{x})e^{i\omega t} + (*)\right] + \mathcal{O}(\varepsilon^2); \end{array}\right\} \Longrightarrow \mathbf{u}_1(\mathbf{x})e^{i\omega_s t} = a(t) \cdot \mathbf{e}(\mathbf{x})e^{i\omega t} + \mathcal{O}(\varepsilon^3).$$

(1.17)

Two results can be derived directly from the latter equality (see also (1.9) and (1.15)),

$$\left|a(t)\right|^2 \mathbf{f}_{20}(\mathbf{x}) = \Delta \mathbf{f}_o(\mathbf{x}) + \mathcal{O}(\varepsilon^4) \Longrightarrow \left|a(t)\right|^2 \mathbf{u}_{20}(\mathbf{x}) = \delta \mathbf{u}(\mathbf{x}) + \mathcal{O}(\varepsilon^4);$$

$$a(t) = |a| \cdot e^{i(\omega_s - \omega)t},$$

(1.18)

and thus it follows from (1.14) that the asymptotic solution of (1.13), based on the *Landau's equation* (1.8), recovers the observed 2D periodic (limit cycle) solution (1.5) in the range Re $\gg$ $\mathbf{R}_{c1}$, with an error of order ε^4. Or in short: Landau's equation (1.8), strictly valid in a close neighborhood of a *Hopf supercritical bifurcation*, can be extended in the present flow problem to the range Re $\gg$ $\mathbf{R}_{c1}$, where a neat periodic solution persists.

Finally, once the 2D numerical solution $\mathbf{u}(\mathbf{x},t)$ is determined and its harmonic components $\{\mathbf{u}_n(\mathbf{x}); n = 0,1,2,3\}$ are computed, the *unstable mode* $\{\lambda = \sigma + i\omega; \mathbf{e}(\mathbf{x}); p_e(\mathbf{x})\}$ and the coefficients $\{\mu; c_3\}$ of *Landau's equation* can be easily estimated, as elaborated below. In fact, using the approximations and the normalization of the mode $\mathbf{e}(\mathbf{x})$,

$$\mathbf{e}(\mathbf{x}) = \frac{\mathbf{u}_1(\mathbf{x})}{|a|} + \mathcal{O}(\varepsilon^2);$$

$$p_e(\mathbf{x}) = \frac{p_1(\mathbf{x})}{|a|} + \mathcal{O}(\varepsilon^2);$$

(1.19)

$$\int_S \|\mathbf{e}(\mathbf{x})\|^2 dS = 1 \Longrightarrow |a| = \left(\int_S \|\mathbf{u}_1(\mathbf{x})\|^2 dS\right)^{1/2} + \mathcal{O}(\varepsilon^2),$$

multiplying (1.11) by $\mathbf{e}^*(\mathbf{x})$, integrating by parts and using $\nabla \cdot \mathbf{e}^* = 0$, one obtains

$$\lambda = \sigma + i\omega = -\int_S \left[(\widehat{\mathbf{u}}_o \cdot \nabla)\mathbf{e} + (\mathbf{e} \cdot \nabla)\widehat{\mathbf{u}}_o\right] \cdot \mathbf{e}^* dS - \frac{1}{\mathrm{Re}}\int_S (\nabla \mathbf{e} : \nabla \mathbf{e}^*) dS,$$

(1.20)

with $\nabla \mathbf{e} : \nabla \mathbf{e}^* = \nabla e_x \cdot \nabla e_x^* + \nabla e_y \cdot \nabla e_y^*$. Using again the approximation $\mathbf{e}_{av} \cong \mathbf{u}_1(\mathbf{x})/|a| \cong \mathbf{e}$, with the same error ε^2 in (1.7), the following identity can be derived,

$$i\omega_s = -\int_S \left[(\mathbf{u}_o \cdot \nabla)\mathbf{e} + (\mathbf{e} \cdot \nabla)\mathbf{u}_o \right] \cdot \mathbf{e}^* dS - \frac{1}{\mathrm{Re}} \int_S (\nabla \mathbf{e} : \nabla \mathbf{e}^*) dS, \qquad (1.21)$$

and subtracting one expression from the other, while using $\widehat{\mathbf{u}}_o = \mathbf{u}_o - \delta\mathbf{u}$, one obtains

$$\sigma + i(\omega - \omega_s) \cong \int_S \left[(\delta\mathbf{u} \cdot \nabla)\mathbf{e} + (\mathbf{e} \cdot \nabla)\delta\mathbf{u} \right] \cdot \mathbf{e}^* dS, \qquad (1.22)$$

with $\delta\mathbf{u}(\mathbf{x}; \mathrm{Re}) \cong \mathcal{O}(\varepsilon^2)$ being solution of the regular linear system (1.10).

Notice that (1.22) reaffirms, as it should, the estimated orders in (1.12) and placing the $a(t)$ defined in (1.18) into Landau's equation (1.8) one has

$$\mu \cong \frac{\sigma}{|a|^2};$$
$$c_3 \cong \frac{\omega_s - \omega}{\sigma}. \qquad (1.23)$$

Summarizing: through the 2D simulation one obtains $\{\mathbf{u}(\mathbf{x}, t); p(\mathbf{x}, t)\}$ and from the Fourier expansion in the harmonics of the observed frequency ω_s one determines the *averaged flow* $\{\mathbf{u}_o(\mathbf{x}; \mathrm{Re}); p_o(\mathbf{x}; \mathrm{Re})\}$ and the *first harmonic* $\{\mathbf{u}_1(\mathbf{x}; \mathrm{Re}); p_1(\mathbf{x}; \mathrm{Re})\}$ defined in (1.5). Solving the linear system (1.10) the field $\delta\mathbf{u}(\mathbf{x}; \mathrm{Re})$ can be computed and so the coefficients of Landau's equation using (1.19), (1.22), and (1.23). The gain in this extra computation is certainly marginal in the context of the 2D problem; however, as it will be discussed in the following sections, Landau's equation is the basis of the 3D Ginzburg-Landau equation (GLE) and with it one can possible predict an asymptotic approximation of the 3D behavior without having to resort to a 3D numerical computation of the flow field. In this context, the proposed approximation is similar to existing "slender body theories" in applied mechanics, as for example the *Lifting Line Theory* in the Aerodynamics of slender wings: in all of them one takes profit of the body slenderness to correct asymptotically the 2D solution. But before one addresses this 3D extension of Landau's equation it is worth to mention some general features of the actual 3D flow around a slender cylinder.

2. Features of the 3D Flow Around a Slender Cylinder

The 2D flow around a slender 3D cylinder is *unstable* with respect to *3D-perturbation* for Re > 190 and this is instability, known experimentally for a long time, has only recently been verified theoretically in a comprehensive numerical study done by Henderson [4]. The plot of the Strouhal number $\mathrm{St} = f_s d/U (f_s = \omega_s/2\pi)$ as a function of Re, see Figure 4, portrays this instability in a very clear way and Henderson [4] has shown that the bifurcation at $\mathbf{R}_{c2} \cong 190$ is *subcritical* while a second one at $\mathbf{R}_{c3} \cong 260$ is *supercritical*. The curve St(Re) presents a hysteretic behavior in the range $180 < \mathrm{Re} \leq 260$, where two competing solutions, corresponding to two distinct attractors, can appear depending on the initial conditions; as
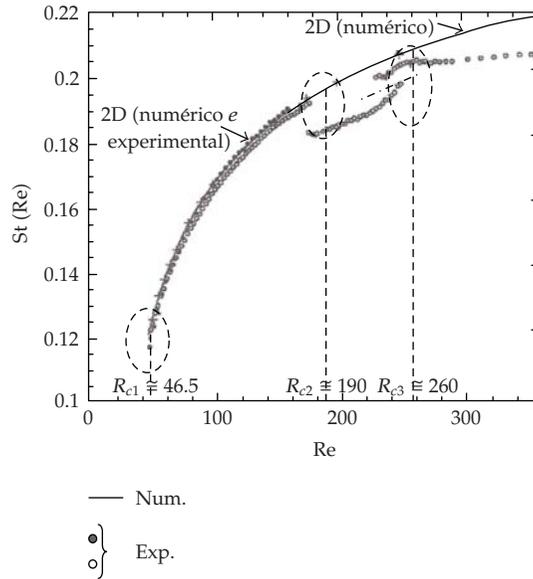
Figure 4: Strouhal number St $= f_s d/U(f_s = \omega_s/2\pi)$ as a function of Re. (Bifurcations points $\mathbf{R}_{C1}; \mathbf{R}_{C2}$ and $\mathbf{R}_{C3}$—source: Henderson [4].)
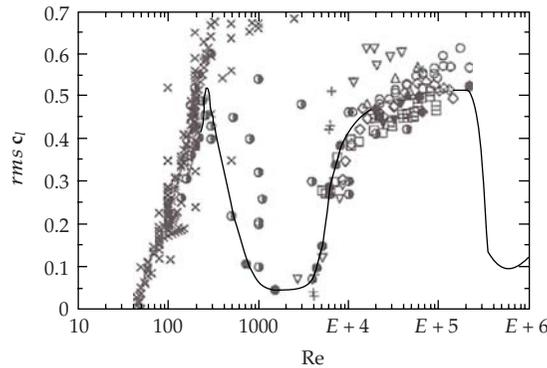
Figure 5: *Lift coefficiente: rms* $\mathbf{c}_l \times$ Re (Norberg [5]). (x: numeric; [●; ◇; ←; ○; ∇]: experiments)

usual in a subcritical bifurcation, the presence of the two attractors can be detected a little before the critical point $\mathbf{R}_{c2} \cong 190$.

In the range Re > 260 the Strouhal frequency changes weakly with Re and the flow pattern presents a well tuned frequency immersed in a chaotic (turbulent) background. In this range of Reynolds numbers the most conspicuous experimental result is, certainly, the *"lift crisis"* observed by Norberg [5] and briefly commented below.

The *sectional* transverse (lift) force $l(z,t)$ was measured by Norberg [5] in the range $250 < \text{Re} < 10\,000$ and the *rms* of the *lift coefficient* $\mathbf{c}_l(z,t) = l(z,t)/((1/2)\rho U^2 d)$ was plotted as a function of Re, the obtained result being shown in Figure 5. The "lift crisis" corresponds to the sharp drop of *rms* $\mathbf{c}_l$ at Re ≈ 260, reaching a minimum at Re ≈ 1000 of about 20% of the 2D value and there remaining up to Re ≈ 5000, where the value of *rms* $\mathbf{c}_l$ starts a slow recovering.

The behavior is similar to the well known "drag crisis" in the range $10^5 < \text{Re} < 10^6$, although even sharper, and it should be also related to the chaotic (turbulent) flow observed when $\text{Re} > 260$. The main purpose in the present paper is to indicate that the Ginzburg-Landau Equation (GLE) has the potential ability to recover Norberg's "lift crisis" and this point will be addressed next.

3. Ginzburg-Landau Equation in the Defect Chaos Regime

The 2D *unstable mode* $a(t) \cdot \mathbf{e}(\mathbf{x})\exp(i\omega t)$ is triggered by a random perturbation distributed along the cylinder's span and one should expect, as a consequence, a certain phase-lag of this mode in the z-direction: the amplitude a must then change with the span coordinate z, namely, $a = a(z,t)$. The z-dependence of the mode amplitude should modify the 2D Landau's equation by a parcel proportional to a z-derivative of a and observing that there is no preferred z-direction this derivative should be *even* in z: the obvious choice here is to take the *second derivative* $\partial_{zz}a$. This is perhaps the simplest argument to introduce the *Ginzburg-Landau Equation* (GLE),

$$\frac{\partial a}{\partial t} - \sigma a - \gamma(1 + ic_1)\frac{\partial^2 a}{\partial z^2} + \mu(1 - ic_3)|a|^2 a = 0; \quad \{\sigma; \gamma; \mu\} > 0, \tag{3.1}$$

as done by Ginzburg in 1950 in his joint study with Landau on super-conductivity, see Ginzburg [6]. It was introduced then as a *phenomenological model*, namely, as an equation that emulates the overall behavior of an observed phenomenon, and as such has been used in Physics, see Aranson and Kramer [7], to analyze a class of problems related to a *distributed Hopf bifurcation*; the flow around a slender cylinder is just an example of it.

In this context, the GLE was first proposed as a *phenomenological model* by Abarède and Monkewitz [8] and studied by Monkewitz and co-authors in several papers; particularly interesting is the work by Monkewitz et al. [9] where some subtle aspects of the flow are theoretically predicted and confirmed experimentally. These works were restricted, however, to the range $\text{Re} < 160$, within the stable range of the 2D periodic flow, and the purpose here is to extend it to the unstable regime $\text{Re} > \mathbf{R}_{c2} \cong 190$.

Normalizing time, space and amplitude by using $\{t \leftarrow \sigma t;\ z \leftarrow (\sigma/\gamma)^{1/2}z;\ a \leftarrow (\sigma/\mu)^{1/2}a\}$ the same equation (3.1) is obtained with $\sigma = \gamma = \mu = 1$: the behavior of the GLE depends only on the *dispersion coefficients* $\{c_1; c_3\}$ and it is not difficult to show, via Fourier Transform of the perturbed equation, that the 2D solution becomes unstable with respect to 3D perturbation when $c_1 \cdot c_3 > 1$; incidentally, this *stability condition* is usually called the *Benjamin-Feir condition*, in honour of a stability study in *water waves* done by these authors, see Benjamin and Feir [11]. In Figure 6 it is shown the results of a comprehensive numerical study done by Shraiman et al. [10] in the unstable region of the dispersion plane (c_1, c_3). It discloses, at first, two distinct chaotic regimes: one, very mild, called "phase chaos", is characterized by a "turbulence" superposed on the uniform 2D phase and restricted to a small strip on the unstable region $c_1 \cdot c_3 > 1$; the other, very energetic and covering the remaining of the unstable region, called "defect chaos", is related directly to the amplitude size $|a|$: the *"defects"* are the points in time-space plane (z,t) where the amplitude is null and the iso-phases either stop or bifurcate at them, see the plots of the iso-phases in the detached figures in Figure 6. Shraiman et al. [10] also observed a thin strip, coined *bi-chaotic*, close to
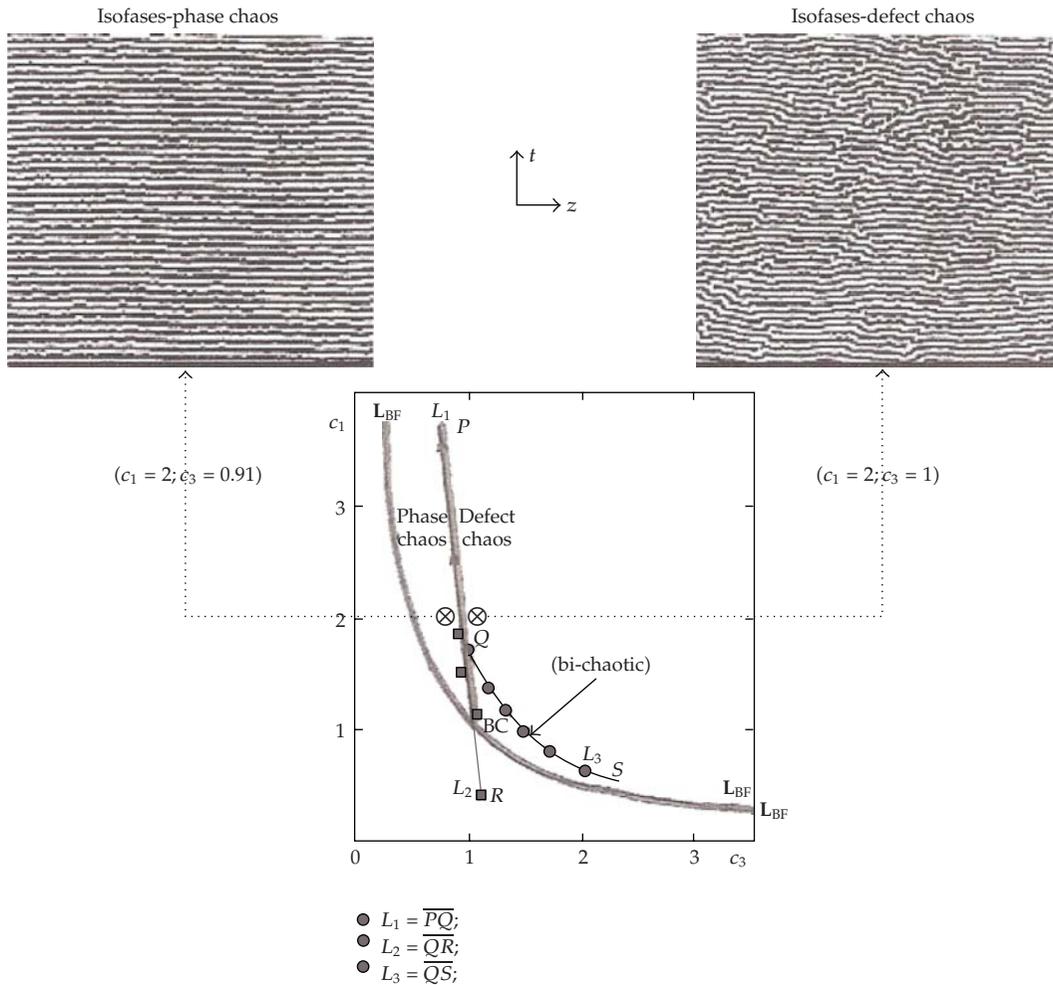
Figure 6: Behavior of GLE in the unstable region ($c_1 c_3 > 1$). $\mathbf{L}_{BF}$: $c_1 c_3 = 1$. *"Phase chaos"* and *"defect chaos"* regimes and corresponding iso-phases. In the *bi-chaotic* region GLE has two attractors with distinct frequencies. (Source: Shraiman et al. [10].)

the threshold curve $c_1 \cdot c_3 = 1$ and in fact penetrating a little into the stable region $c_1 \cdot c_3 < 1$, where the GLE has two chaotic attractors.

It seems then that the GLE, with recognized predictive ability in the *stable range* (Re < 190 or $c_1 \cdot c_3 < 1$), may be useful also in the *unstable range* (Re > 190 or $c_1 \cdot c_3 > 1$) since, as in the flow problem, it presents a bi-chaotic behavior in the vicinity of the threshold point (Re ≈ 190 or $c_1 \cdot c_3 \approx 1$) and a chaotic one when Re ≫ 190 or $c_1 \cdot c_3 \gg 1$. The difficulty here is first of all operational, since it seems awkward to adjust the parameters of the *phenomenological* GLE to the empirical data of the now chaotic flow, and also conceptual in some sense, once it is understood that GLE can model the problem just in the vicinity of Hopf bifurcation but not far from it, although Monkewitz et al. [9] used GLE to model properly the flow problem at a Re almost three times larger than critical Reynolds $\mathbf{R}_{c1} \cong 46.5$.

However, as seen in the first section, *Landau's equation* can be extended far beyond bifurcation and the GLE can be obtained as an *asymptotic approximation* of the 3D flow related
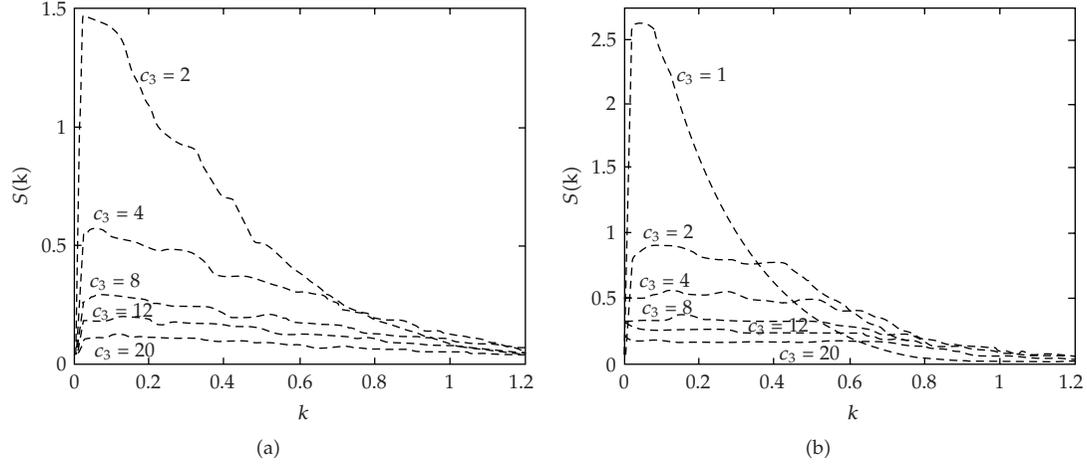
Figure 7: Wavenumber spectrum for several points in the dispersion plane $(c_1; c_3)$. $\sigma = \gamma = \mu = 1$: (a) $c_1 = 1$; (b) $c_1 = 4$. (Burr (2007) —NDF)

to (1.13), allowing one to determine the curve $(c_3(\mathrm{Re}); c_1(\mathrm{Re}))$ representing the flow problem. This computation was not done yet, however, only scarce results are available; meanwhile, it seems interesting to check whether or not GLE has the potential ability to recover the main features of the observed 3D flow. For example, one certainly should expect that the curve $(c_3(\mathrm{Re}); c_1(\mathrm{Re}))$ crosses the *Benjamin-Feir curve* $c_1 \cdot c_3 = 1$ at $\mathrm{Re} \approx 190$, penetrating after the defect chaos regime trough the bi-chaotic region, corresponding to the hysteretic behavior in the range $180 < \mathrm{Re} \leq 260$ observed in Figure 4; as Re rises above 260 the curve $(c_1(\mathrm{Re}); c_2(\mathrm{Re}))$ must go even deeper into the defect chaos regime and, in particular, *Norberg's lift crisis* must be predicted if the GLE approximation is consistent. But the transverse (lift) force is due to the odd harmonics, and so it is proportional to $a(z, t)$: the sharp drop in *rms* $\mathbf{c}_l$ must be related, in the GLE context, to a sharp drop in *rms* $|a(z, t)|$. The purpose here is to discuss this point while revealing some interesting aspects of the GLE in the *defect chaos regime*, that may have an interest in itself.

As in a turbulent flow regime, the chaotic solution in the "defect chaos" regime is characterized by a cascade of length scales k^{-1} limited below by a "Kolmogorov scale" $(k_{(\mathrm{kol})})^{-1}$, where the dissipated power, proportional to $\gamma \cdot |\partial_z a|^2$, is of order of the power given by the instability, proportional to $\sigma \cdot |a|^2$; it follows that

$$k_{(\mathrm{kol})} \approx \sqrt{\frac{\sigma}{\gamma}} \quad \text{or} \quad k_{(\mathrm{kol})}|_{\sigma=\gamma=1} \approx 1. \tag{3.2}$$

Equation (3.1) was integrated in the region $0 \leq z \leq l = 1000$ in the time interval $18000 \leq t \leq 20000$, using the periodic boundary condition $\{a(0, t) = a(l, t); \partial_z a(0, t) = \partial_z a(l, t)\}$. Figure 7 shows the wavenumber spectrum $S(k)$ of $a(z, t)$ for pairs of values $(c_3; c_1)$ and it is clear that the energy is almost exhausted in the region $k > k_{\mathrm{kol}} \approx 1$. This behavior was observed in all numerical experiments in the grid $\{1 \leq c_3 \leq 20; 1 \leq c_1 \leq 20; c_3 \cdot c_1 > 1\}$.
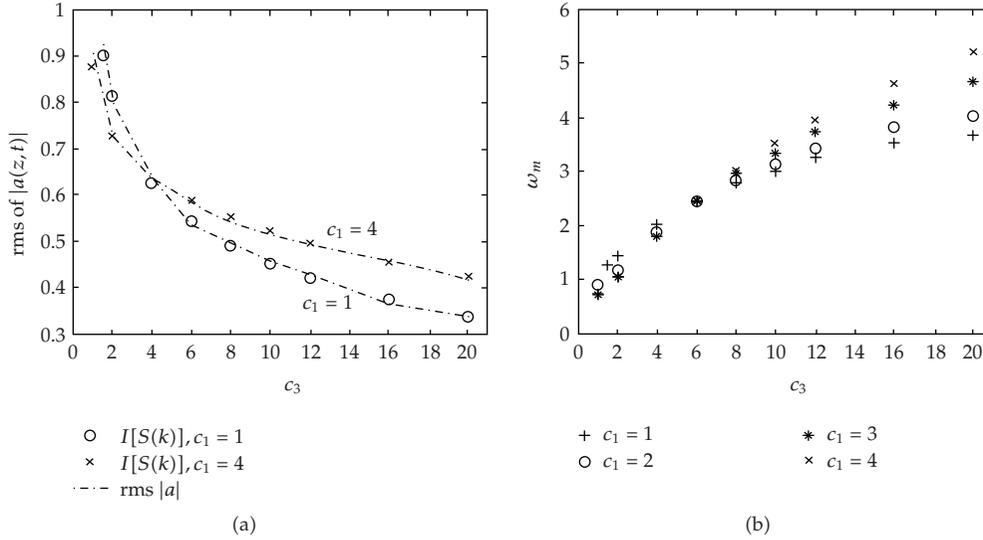
Figure 8: (a) Comparison between (3.3) and *rms* $|a|$; (b) Averaged frequency ω_m. (Burr (2007)—NDF)

The intensity of the response can be also estimated by the wavenumber spectrum integral,

$$I[S(k)] = \int_0^\infty S(k)\,dk = \frac{1}{l}\int_0^l |a(z,t)|^2 dz, \tag{3.3}$$

and in Figure 8(a) the values of $I[S(k)]$ and *rms* $|a(z_m,t)|$, $z_m = (1/2)l$, are plotted again for several points in the "dispersion plane" $(c_3; c_1)$. The almost exact agreement between the two plots indicates that the random signal $a(z,t)$ is *weakly stationary*, namely

$$\frac{d}{dt}\int_0^l |a(z,t)|^2 dz \cong l \cdot \frac{d}{dt}\left(rms|a(z_m,t)|\right) = 0, \quad \left(z_m = \frac{1}{2}l\right). \tag{3.4}$$

Figure 8(a) shows that the *rms* of $|a(z,t)|$ decreases monotonically with c_3, kept c_1 constant, but when c_3 is constant it increases with c_1, also monotonically in the range $c_3 > 4$. This behavior can be inferred from an identity of the GLE. In fact, if (3.1) is multiplied by a^* and integrated in the interval $0 \leq z \leq l$, one obtains, after using the periodicity of the boundary conditions and the weak stationary condition (3.4), the identities

$$
\begin{aligned}
\text{(i)} \quad & -\int_0^l |a|^2 dz + \int_0^l \left|\frac{\partial a}{\partial z}\right|^2 dz + \int_0^l |a|^4 dz = 0; \\[2mm]
\text{(ii)} \quad & \int_0^l \left(\frac{\partial a}{\partial t}a^* - \frac{\partial a^*}{\partial t}a\right)dz + 2ic_1\int_0^l \left|\frac{\partial a}{\partial z}\right|^2 dz - 2ic_3\int_0^l |a|^4 dz = 0.
\end{aligned}
\tag{3.5}
$$

Now, if $a(z,t) = |a(z,t)| \cdot \exp[i\varphi(z,t)]$ and introducing the average frequency ω_m by the expression

$$\omega_m = \left(\frac{1}{l}\int_0^l \frac{\partial\varphi}{\partial t} \cdot |a|^2 dz\right)/I[S(k)],\tag{3.6}$$

one obtains from (3.5)

$$\chi^2 = \frac{m_4}{m_2} = \frac{1+(\omega_m/c_1)}{1+(c_3/c_1)}; \qquad m_k = \frac{1}{l}\int_0^l |a|^k dz.\tag{3.7}$$

This relation was obtained under the weak stationary assumption (3.4) and it seems reasonable to assume that the intensity of *rms* $|a(z,t)|$ can be gauged by χ; notice, in particular, that χ is monotonically increasing with c_1 when $\omega_m/c_3 < 1$ and decreasing with c_3 increasing, in accordance to the observed in Figures 8(a), 8(b) for *rms* $|a(z,t)|$. From the relation $\chi \cong$ *rms* $|a(z,t)|$ it follows also the asymptotic relations

$$\begin{aligned}&\text{(i)}\quad \lim_{c_1\to\infty}(rms|a|)_{c_3\cong\mathcal{O}(1)} \cong \mathcal{O}(1);\\[6pt]&\text{(ii)}\quad \lim_{c_3\to\infty}(rms|a|)_{c_1\cong\mathcal{O}(1)} \cong \mathcal{O}\left(\sqrt{\frac{c_1+\omega_m}{c_3}}\right).\end{aligned}\tag{3.8}$$

The expression (ii) in (3.8) can be related to the *Kolmogorov scale* (3.2). In fact, lets recall, first of all, a standard result: by assuming an *harmonic wave solution* $a(z,t) = |a| \cdot \exp[i(k \cdot z + \omega \cdot t)]$ of the GLE (3.1) one obtains the *dispersion relation*,

$$\omega = c_3 \cdot |a|^2 - c_1 \cdot k^2, \quad (\sigma = \gamma = \mu = 1)\tag{3.9}$$

depending on the "dispersion coefficients" $(c_1; c_3)$. For a "random wave" one may take $(rms\, a(z,t))^2$ in the place of $|a|^2$ in (3.9) and if $k = k_{(kol)} \approx 1$ one obtains, with the help of (ii) in (3.8), $\omega \approx \omega_m$, or in short: the averaged frequency defined in (3.6) is the "Kolmogorov frequency scale" of the random signal $a(z,t)$ in the limit $c_3 \to \infty$, kept c_1 constant; in this limit ω_m tends to a bounded value $\omega_\infty(c_1)$, see Figure 8(b). The data of Figure 8(a) confirm, in the limit $\{c_3 \to \infty; c_1 = 1\}$, the asymptotic behavior *rms* $a(z,t) \cong \rho \cdot c_3^{-1/2}$ with $\rho \approx (c_1+\omega_m)^{1/2}$; in reality, $\rho \cong 1.49$ from the data of Figure 8(a) while $(c_1+\omega_m)^{1/2} \cong 2.17$ from Figure 8(b).

One expects then that *rms* $|a(z,t)|$ diminishes monotonically with increasing c_3, a result confirmed by the direct evaluation of *rms* $|a(z,t)|$ in the dispersion plane (c_3, c_1), see Figure 9; notice that expression (i) in (3.8) is also recovered, a result consistent with the "phase chaos" regime identified in Figure 6.

In the flow problem, the linear dispersion coefficient c_1 is not expected to change too much with Re but c_3, defined by the ratio $(\omega_s - \omega)/\sigma$, see (1.23), apparently does: the difference $(\omega_s - \omega)$ is small but fairly constant while σ appears to drop sharply for Re above 100, the ratio $(\omega_s - \omega)/\sigma = c_3$ becoming very large then: as it was seen, if $c_3 \gg 1$ then *rms* $|a| \ll 1$ and thus *rms* $\mathbf{c}_l \ll 1$. This result must be confirmed by a more refined numerical solution but it indicates, anyway, the ability of the GLE to predict *Norberg's "lift*
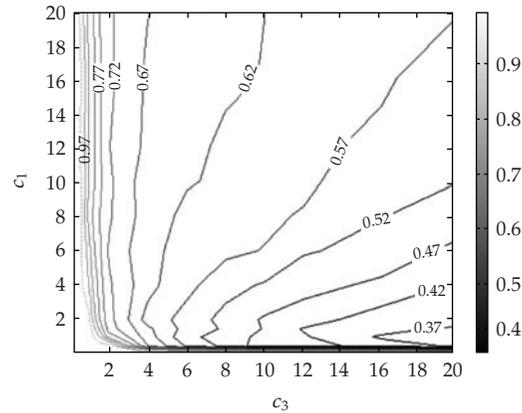
Figure 9: Contour lines of $rms\ a(z,t)$ in the dispersion plane $(c_3; c_1)$. (Burr (2007)—NDF)

crisis." Or, in other words, if the actual curve $(c_3(\text{Re}); c_1(\text{Re}))$ in fact penetrates the unstable range through the bi-chaotic region at $\text{Re} \approx \mathbf{R}_{c2} \cong 190$ and $c_3(\text{Re})$ increases rapidly with Re, then the GLE, together with the asymptotic expansion (1.14), defines in fact a *reduced* Navier-Stokes Equation for the flow around a slender cylinder, the practical importance of it being commented below.

4. Conclusion

In this paper the possibility to solve asymptotically the flow around a slender cylinder using a 2D computation and the Ginzburg-Landau equation to obtain the 3D correction was elaborated, stressing the regime above $\text{Re} \cong 190$, where three dimensionality has a marked influence. Although one must wait more refined numeric results to reach a definitive conclusion, the qualitative behavior of GLE in the range $c_3 \cdot c_1 > 1$ matches very well the most important qualitative features of the flow around a slender cylinder in the range $\text{Re} > 190$; as already discussed in Monkewitz et al. [9] , the matching between both is impressive in the range $\text{Re} < 160$.

A practical problem where the present study may be relevant is related to the *fatigue analysis* of "risers" (vertical ducts) in the offshore oil production systems, essential to assure the safe operation of these systems during its projected life: risers are exposed to ocean currents and oscillate transversally in the elastic modes with natural frequencies close to the flow's Strouhal frequency, the related cyclic stress causing fatigue of the material. The use of a 3D Navier-Stokes code to obtain practical answers is, however, completely out of question in the present stage of development, not only due to computer time needed, but also for the lack of confidence in the numerical results of the enormous discrete system related to it. The *reduced* Navier-Stokes equation, represented by the GLE, opens an opportunity to a feasible and relatively cheap computation: in it, the complex coupling between the *incoming flow* and the *riser's elasticity* can be represented by a coupled set of equations—one of them being the (extended) GLE, the other representing the riser's elastic behavior—both depending only on the space variable along the riser's span, turning the discrete model orders of magnitude smaller. This is the main motivation to study this problem at NDF.

Acknowledgments

References

[1] M. van Dyke, *An Album of Fluid Mechanics*, The Parabolic Press, Stanford, Calif, USA, 1982.

[2] D. Barkley, "Linear analysis of the cylinder wake mean flow," *Europhysics Letters*, vol. 75, no. 5, pp. 750–756, 2006.

[3] J. A. P. Aranha, "Weak three dimensionality of a flow around a slender cylinder: the Ginzburg-Landau equation," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, vol. 26, no. 4, pp. 355–367, 2004.

[4] R. D. Henderson, "Nonlinear dynamics and pattern formation in turbulent wake transition," *The Journal of Fluid Mechanics*, vol. 352, pp. 65–112, 1997.

[5] C. Norberg, "Fluctuating lift on a circular cylinder: review and new measurements," *Journal of Fluids and Structures*, vol. 17, no. 1, pp. 57–96, 2003.

[6] V. L. Ginzburg, *On Superconductivity and Superfluidity*, Nobel Lecture, Springer, Berlin, Germany, 2003.

[7] I. S. Aranson and L. Kramer, "The world of the complex Ginzburg-Landau equation," *Reviews of Modern Physics*, vol. 74, no. 1, pp. 99–143, 2002.

[8] P. Abarède and P. A. Monkewitz, "A model for the formation of oblique shedding and "chevron" patterns in cylinder wake," *Physics of Fluids A*, vol. 4, no. 4, pp. 744–756, 1992.

[9] P. A. Monkewitz, C. H. K. Williamson, and G. D. Miller, "Phase dynamics of Kármán vortices in cylinder wakes," *Physics of Fluids*, vol. 8, no. 1, pp. 91–96, 1996.

[10] B. I. Shraiman, A. Pumir, W. van Saarlos, P. C. Hohenberg, H. Chaté, and M. Holen, "Spatiotemporal chaos in the one-dimensional complex Ginzburg-Landau equation," *Physica D*, vol. 57, no. 3-4, pp. 241–248, 1992.

[11] T. B. Benjamin and J. E. Feir, "The desintegration of wave trains on deep water—part I: theory," *The Journal of Fluid Mechanics*, vol. 27, no. 3, pp. 417–430, 1967.

Research Article

Analysis of Nonlinear Dynamics for Abrupt Change of Interphase Structure in Liquid–Liquid Mass Transfer

DongXiang Zhang,[1] Min Sun,[2] and Jing Li[2]

[1] *School of Chemical Engineering and the Environment, Beijing Institute of Technology,*
  Beijing 100081, China
[2] *College of Applied Science, Beijing University of Technology, Beijing 100124, China*

Correspondence should be addressed to Jing Li, leejing@bjut.edu.cn

As a liquid-liquid system is far from equilibrium state, the phase thickness is variable when mass transfer process with chemical reaction occurs in interphase zone, and a dispersible transitional layer called the interphase dispersed zone (IDZ) is formed. The IZD model composed of thermodynamically instable O/W or W/O microemulsion has reasonably explained enormous experimental phenomena in nonlinear mass transfer. To forecast the possible parameter ranges of IDZ process and abrupt change of liquid-liquid mass transfer rate, the dynamic characteristics of a molecular diffusion model are considered in this paper. We applied the bifurcation theory of planar dynamical system, Laplace transform, and maple software to investigate the model, and obtain different phase portraits of the system in different regions. The results obtained will play an important directive role in the study of IDZ model.

1. Introduction

Liquid-liquid mass transfer and reaction mass transfer processes are widely used in many industries such as nuclear energy, hydrometallurgy, pharmaceutical, and chemical. In research on production process control and techniques, we found that interphase processes are main control procedures for these liquid-liquid systems. In mass transfer process, the interphase zone always evolves with time, accompanying with chemical reaction, hyedromchanics instability, automatic dispersion, formation of interphase layer, absorption and coalescence of dispersed particles, and formation of interphase layer structure, etc. [1–3].

Traditional dynamic research methods for interphase mass transfer kinetics in liquid-liquid systems use concentration difference to express the driving force for mass transfer,

and assumed that each phase is unchangeable even in the interface and interphase layers have different thicknesses. However, such assumption is unable to explain the bifurcation of mass transfer rate in liquid-liquid systems under the effect of mechanic field, electric field, magnetic field, gravitational field, and sound field, etc. [1–7]. Abrupt change of mass transfer rate has great effect on theoretical research and production practice and becomes an important content of theoretical research and engineering security control. To overcome the limit in traditional understanding, Tarasov et al. have developed the physical model of IDZ through analyzing actual process phenomena[1–3]. Taking the mass transfer process in liquid-liquid stripping system as an example, Zhang holds that, due to the effect of coordinate chemical binding between extractant and extract as well as intermolecular function between coordination compound and solvent molecules, when the extracted ion carries the extractant and solvent particles dispersing near the inorganic interphase layer in the process of transferring from organic to inorganic phase, the extractant and the solvent after separating from the ion can form an O/W microemulsion near the inorganic interphase zone under the competition of water molecules in the inorganic phase of other stripping agents. Similarly, the extract in the inorganic phase in liquid-liquid extraction system can also form a W/O microemulion near the organic interphase layer. Under the effect of the thermal movement of molecule and concentration driving force, the two types of microemulsion come into being dynamically and will change dynamically during mass transfer process. Moreover, the O/W microemulsion in the inorganic interphase and the W/O microemulsion near the organic interphase layer can come into being simultaneously in the interphase layer of the mass transfer process. It is inferable that a bicontinuous-phase microemulsion can also be produced between the O/W microemulsion and the W/O microemulsion in the interphase layer [8]. This process is generally known as mass transfer microemulsion process in liquid-liquid system interface. This model for nonhomogenous mass transfer or reaction mass transfer process in the interface is called the model of IDZ[3].

In the reactor shown in Figure 1, we have obtained the molecular diffusion equation for interphase mass transfer $\partial c/\partial t = D(\partial^2 c/\partial x^2)$ and the equation describing interphase mass transfer process $\partial c/\partial t = D(\partial^2 c/\partial x^2) + r$ through equilibrium calculations by carrying out experiments for extraction and stripping of inorganic acid and transitional metal compounds. When a small-range of external electrostatic field was applied in the system, the bifurcation phenomenon was found in interphase mass transfer rate (refer to Figure 2).

When using any amine extractant to extract hydrochloric acid, there also appears a sudden change of the interphase mass transfer velocity when the initial concentration of hydrochloric acid is adjusted a little [3]. The phenomenon of the sudden change of the interphase mass transfer velocity caused by the microranged concentration adjustment cannot be explained with Fick's Law of diffusion. To coordinate the contradiction, Zhang and Trasov gave an explanation that along with the transfer of the key components in the phase interface areas, the water phase and organic phase form an IDZ structure, that is, water-in-oil emulsion, which gives rise to changes in interphase structure [2, 8].

Koltsova et al. have simplified the mass transfer interhpase layer into noncontinuous fluid film mode with interspace, and studied the IDZ process by using fractal dynamic model $\partial^\alpha c_\alpha/\partial t^\alpha = D_\alpha(\partial^2 c_\alpha/\partial x^2) + kc_\alpha^m$ $(0 < \alpha < 1)$ through simulation of the impetus of nonequilibrium thermodynamics [6, 7]. Since the effect of chemical reaction rate kc_α^m in these systems is much less than that of diffusion, we have used $\partial^\alpha c_\alpha/\partial t^\alpha = D_\alpha(\partial^2 c_\alpha/\partial x^2)$ for quantitative research on the change of interphase layer thickness and dispersive characteristics of medium induced by applying mechanical function in the interphase layer zone (refer to Figures 3 and 4) [6].
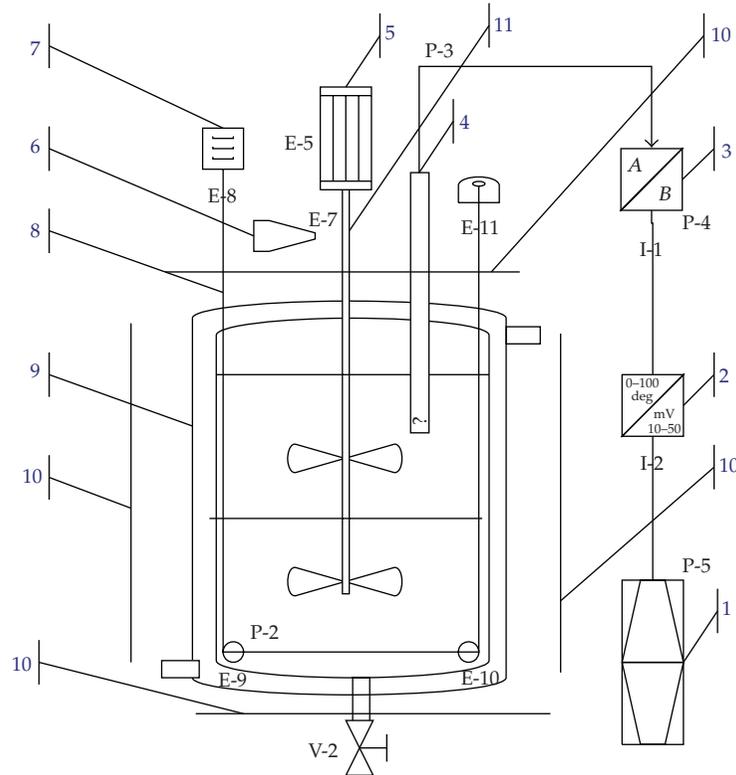
Figure 1: The improved constant interface cell 1-computer; 2-Digital switch-device; 3-concentration detector; 4-sensor; 5-constant speed motor; 6-speed device; 7-reciprocating device; 8-ribbon; 9-heat preserved water-jacket; 10-electric fields 11-stirrers.

Numerous researches show that the change of internal factors and the effect of external fields in mass transfer process can cause non-linear interphase molecular diffusion and abrupt change of interphase structure in liquid-liquid mass transfer. This also makes those parameter ranges in liquid-liquid mass transfer rate that may change abruptly a basic theoretical issue that has aroused common attention of researchers.

This paper studies the dynamic characteristics and phase diagram of molecular model $\partial c/\partial t = D(\partial^2 c/\partial x^2)$ in liquid layer. The possibility of abrupt change of the reaction mass transfer process $\partial c/\partial t = D(\partial^2 c/\partial x^2) + r$ is forecast through the basic mass transfer process $\partial c/\partial t = D(\partial^2 c/\partial x^2)$ for liquid-liquid molecular diffusion.

The bifurcation theory of the plane polynomial vector fields plays an important role in the study of nonlinear dynamic system. By using the bifurcation theory, Li et al. [9] are considered a parametrically and externally excited mechanical system. Li et al. [10] investigated the rotor-AMB system with time-varying stiffness of single degree of freedom and found that there exist, respectively, at least 17, 19, 21, and 22 limit cycles in the system under four different control conditions. The dynamic characteristics of a molecular diffusion model are considered in this paper, we applied the bifurcation theory of planar dynamical system, laplace transform, and maple software to investigate the model, we obtain different phase portraits of the system in different regions. The results obtained will play an important directive role in the study of IDZ model.
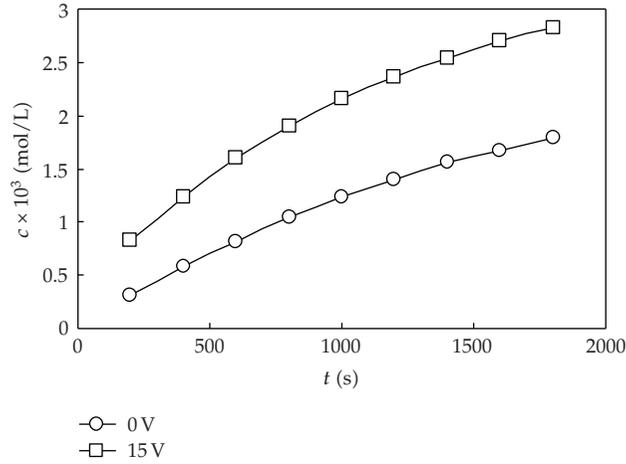
Figure 2: The changing of Cu^{2+} concentration in inorganic-phase along with the mass transfer process.
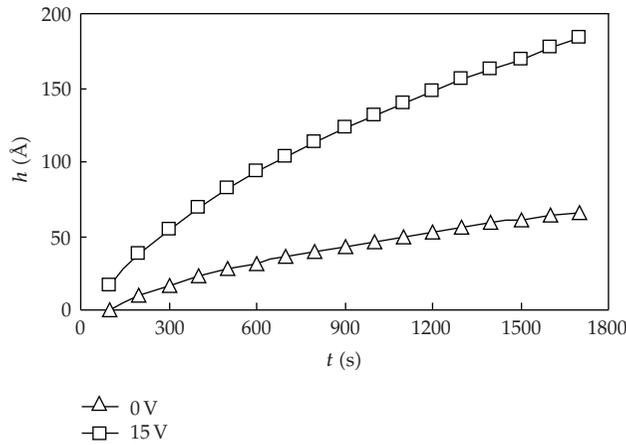
Figure 3: Curves of the interphase layer film thickness evolving with time contrasting with external electrostatic field.

2. Laplace Transform and the Averaged Equation

Based on the mathematical model of double film molecular diffusion, the one-dimensional second-order differential equation with thickness as h for controlling molecular diffusion in phase interface can be considered as $\partial c/\partial t = D(\partial^2 c/\partial x^2)$, where t refers to time, $c(t, x)$ refers to the concentration of the transferring composition at the time t and position x, and D refers to molecular diffusion coefficient. Unit of quantities in the diffusion coefficient is $[m^2/s]$. To find the molecular diffusion coefficient the formula in diluted solutions [11] can be used: $D = 7.4 \times 10^{-12}((\beta M)^{1/2} T/\mu v^{3/5})$, where M refers to molecular weight of solvent, T refers to temperature and its units is K, μ refers to coefficient of dynamic viscosity and its units is mPa.s, and β refers to coalescence parameter of solvent molecules.
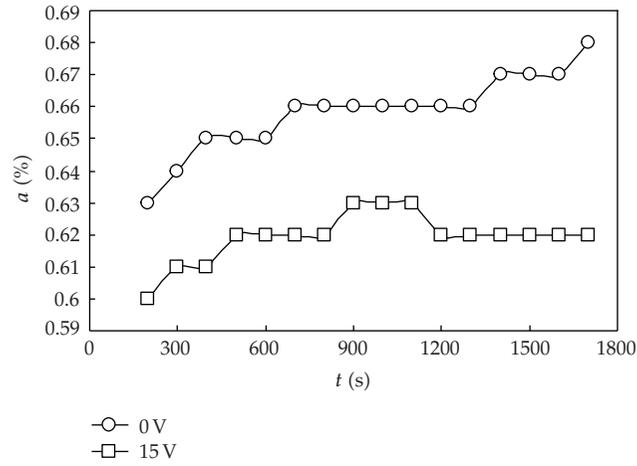
Figure 4: Changing curves of the interphase fractal dimension against time through comparing function of the external electrostatic field.

By using the Laplace transform to the following model as to t:

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2}, \tag{2.1}$$

we get

$$\bar{c}(p, x) = \int_0^T c(t, x)e^{-pt}dt. \tag{2.2}$$

From (2.2) we can obtain

$$\bar{c}_{xx} = \int_0^T c_{xx}e^{-pt}dt. \tag{2.3}$$

Simplify (2.2) we get

$$
\begin{aligned}
-p\bar{c}(p, x) &= \int_0^T c(t, x)de^{-pt} \\
&= c(T, x)e^{-pt} - c(0, x) - \int_0^T e^{-pt}c_t dt,
\end{aligned} \tag{2.4}
$$

Substitute $c_t = Dc_{xx}$ to the above

$$
\begin{aligned}
-p\bar{c}(p, x) &= c(T, x)e^{-pt} - c(0, x) - \int_0^T Dc_{xx}e^{-pt}dt \\
&= c(T, x)e^{-pt} - c(0, x) - D\bar{c}_{xx}.
\end{aligned} \tag{2.5}
$$

So

$$\overline{c}_{xx} = \frac{c(T, x)e^{-pt} - c(0, x) + p\overline{c}(p, x)}{D}. \tag{2.6}$$

Let

$$f(\overline{c}) = \frac{c(T, x)e^{-pt} - c(0, x) + p\overline{c}(p, x)}{D} = \frac{D_T x(\overline{c})}{D}. \tag{2.7}$$

Using the Taylor series expansion to $f(x(\overline{c}))$ at the point $\overline{c} = 0$

$$f(x(\overline{c})) = f(x(0)) + f'(x(0))\overline{c} + \frac{f''(x(0))}{2}\overline{c}^2 + \cdots. \tag{2.8}$$

At the same time, we introduce Taylor series expansion to the right side of (2.7) at the point $\overline{c} = 0$:

$$\frac{D_T x(\overline{c})}{D} = \frac{1}{D}(D_T(x(0))) + D'_T(x(0))\overline{c} + \frac{1}{2}D''_T\left(x(0)\overline{c}^2\right) + \cdots. \tag{2.9}$$

From (2.8) and (2.9), based on the two sides coefficient of $\overline{c}^i$ is equal, we can get

$$\overline{c}^0 : f(x(0)) = \frac{1}{D}D_T(x(0)),$$

$$\overline{c}^1 : f'(x(0)) = \frac{1}{D}D'_T(x(0)),$$

$$\overline{c}^2 : f''(x(0)) = \frac{1}{D}D''_T(x(0)), \tag{2.10}$$

$$\vdots$$

$$\overline{c}^i : f^{(i)}(x(0)) = \frac{1}{D}D_T^{(i)}(x(0)).$$

$$\vdots$$

So,

$$\overline{c}_{xx} = f(x(\overline{c}))$$
$$= \frac{1}{D}D_T(x(0)) + \frac{1}{D}D'_T(x(0))\overline{c} + \frac{1}{D}D''_T(x(0))\overline{c}^2 + \cdots + \frac{1}{D}D_T^i(x(0))\overline{c}^i + \cdots. \tag{2.11}$$

Substitute $\overline{c}$ for y and x for τ for the convenience of our study, we get $y_{\tau\tau} = f(y)$.

For any given initial values $D_T^i(x(0))$, $i \in N$, we can find the corresponding $f(y)$. Let $D_T^i(x(0)) = 0$, $i \in N$ and $i \neq 1,3,5$, $D_T'(x(0)) = D$, $D_T'''(x(0)) = -aD$ and $D_T^5(x(0)) = D/16$, we can obtain $f(y) = y - ay^3 + (1/16)y^5$.

Let $dy/d\tau = x + bx^3$, so $dx/d\tau = y_{\tau\tau}/(1 + 3bx^2) = f(y)/(1 + 3bx^2)$.

If we introduce the nonlinear transform $d\tau/d\xi = -(1 + 3bx^2)$, then

$$\frac{dx}{d\xi} = \frac{dx}{d\tau}\frac{d\tau}{d\xi} = f(y),$$

$$\frac{dy}{d\xi} = \frac{dy}{d\tau}\frac{d\tau}{d\xi} = -x - 4bx^3 - 3b^2x^5. \tag{2.12}$$

By means of the nonlinear transform, (2.6) is equivalent to the averaged equation

$$\frac{dx}{d\xi} = y - ay^3 + \frac{1}{16}y^5,$$

$$\frac{dy}{d\xi} = -x - 4bx^3 - 3b^2x^5, \tag{2.13}$$

where a and b are parameters, $a > 1/2$, $b < 0$.

The Hamiltonian function of system (2.13) is

$$H(x, y) = \frac{y^2}{2} - \frac{ay^4}{4} + \frac{y^6}{96} + \frac{x^2}{2} + bx^4 + \frac{b^2}{2}x^6. \tag{2.14}$$

3. Characteristic of the Singular Points

The singular points of system (2.13) are

$$A_{00}(0,0), \quad A_{01}(0,\pm y_1), \quad A_{02}(0,\pm y_2), \quad A_{10}(\pm x_1,0), \quad A_{11}(\pm x_1,\pm y_1),$$

$$A_{12}(\pm x_1,\pm y_2), \quad A_{20}(\pm x_2,0), \quad A_{21}(\pm x_2,\pm y_1), \quad A_{22}(\pm x_2,\pm y_2), \tag{3.1}$$

where $x_1 = \sqrt{-1/b}$, $x_2 = \sqrt{-1/3b}$, $y_1 = 2\sqrt{2a + \sqrt{4a^2 - 1}}$, $y_2 = 2\sqrt{2a - \sqrt{4a^2 - 1}}$.

For a singular point of system (2.13), its characteristic is determined by the eigenvalue of the polynomial

$$p(\lambda) = \lambda^2 - G(x,y), \tag{3.2}$$

where

$$G(x,y) = \left(1 - 3ay^2 + \frac{5}{16}y^4\right)\left(-1 - 12bx^2 - 15b^2x^4\right). \tag{3.3}$$

By the bifurcation theory of the plane polynomial vector fields, for a singular point (x_i, y_j) of a planar Hamiltonian system, if $G(x_i, y_j) > 0$, then this singular point is a saddle point; if $G(x_i, y_j) < 0$, then it is a center point; if $G(x_i, y_j) = 0$, the singular point is a cusp [12].

Based on the bifurcation theory of the plane polynomial vector fields, it is found that the singular points $A_{00}(0,0)$, $A_{01}(0,\pm y_1)$, $A_{10}(\pm x_1, 0)$, $A_{11}(\pm x_1, \pm y_1)$, $A_{22}(\pm x_2, \pm y_2)$ are the centers, and the singular points $A_{02}(0, \pm y_2)$, $A_{12}(\pm x_1, \pm y_2)$, $A_{20}(\pm x_2, 0)$, $A_{21}(\pm x_2, \pm y_1)$ are the saddle points.

From (2.14), we can obtain the value of the Hamiltonian function at the singular point of system (2.13), respectively,

$$
\begin{aligned}
&h_0^c = H(\pm x_1, \pm y_1) = s + t, \quad h_1^c = H(\pm x_1, 0) = 0, \quad h_2^s = H(\pm x_1, \pm y_2) = s - t, \\
&h_3^s = H(\pm x_2, \pm y_1) = s + t - \frac{2}{27b}, \qquad h_4^s = H(\pm x_2, 0) = -\frac{2}{27b}, \\
&h_5^s = H(\pm x_2, \pm y_2) = s - t - \frac{2}{27b}, \qquad h_6^c = H(0,0) = 0, \\
&h_7^c = H(0, \pm y_1) = s + t, \qquad h_8^s = H(0, \pm y_2) = s - t,
\end{aligned}
\tag{3.4}
$$

where $s = 4a - (32a^3/3)$, $t = (4 - 16a^2)\sqrt{4a^2 - 1}/3$.

4. Bifurcation Parameter and Dynamic Analysis

In order to compare h_2^s, h_3^s, h_4^s, we consider these curves:

$$
(C_1) : h_2^s = h_3^s, \qquad (C_2) : h_2^s = h_4^s, \qquad (C_3) : h_3^s = h_4^s.
\tag{4.1}
$$

These three curves and $a = 1/2$ divided the plane into some different bifurcation regions as shown in Figure 5.

Where the region I is $1/2 < a < \sqrt{3}/3$ and $2/27(t - s) < b < 0$, the region II is $a > \sqrt{3}/3$, and $1/27t < b < 0$. The region III is $a > \sqrt{3}/3$ and $2/27(t - s) < b < 1/27t$.

Case I

We consider the parameter condition of $UP = (a, b) = (0.6, -0.05) \in II$, the phase portrait of system (2.13) as shown in Figure 6.

So, $h_0^c = -0.293150$, $h_1^c = 0$, $h_2^s = 0.485150$, $h_3^s = 1.188330$, $h_4^s = 1.481481$, $h_5^c = 1.966632$, $h_6^c = 0$, $h_7^c = -0.293150$, $h_8^s = 0.485150$. We can obtain $h_0^c < h_1^c < h_2^s < h_3^s < h_4^s < h_5^c$, $h_0^c = h_7^c$, $h_1^c = h_6^c$, $h_2^s = h_8^s$.

Figure 7 demonstrates the changing process of the phase portraits as h varied under the condition UP.

There are nine different families $\{\Gamma_j^h\}$, $(j = 1, 2, \ldots, 9)$ of closed orbits for system (2.13) as the variable h changes, which are described as follows:$\{T_{1i}^h\}$, $h \in (h_0^c, h_2^s)$, four families closed orbits enclosing the singular points $(\pm x_1, \pm y_1)$;$\{T_{2i}^h\}$, $h \in (h_0^c, h_2^s)$: two families closed orbits enclosing the singular points $(0, \pm y_1)$;$\{T_{3i}^h\}$, $h \in (h_1^c, h_2^s)$: two families closed orbits enclosing the singular points $(\pm x_1, \pm y_1)$; $\{T_4^h\}$, $h \in (h_1^c, h_2^s)$: one family closed orbit
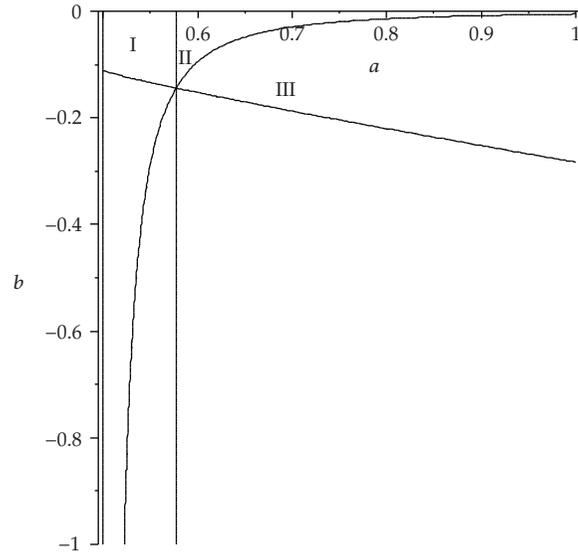
Figure 5: The bifurcation sets of the system (2.13).
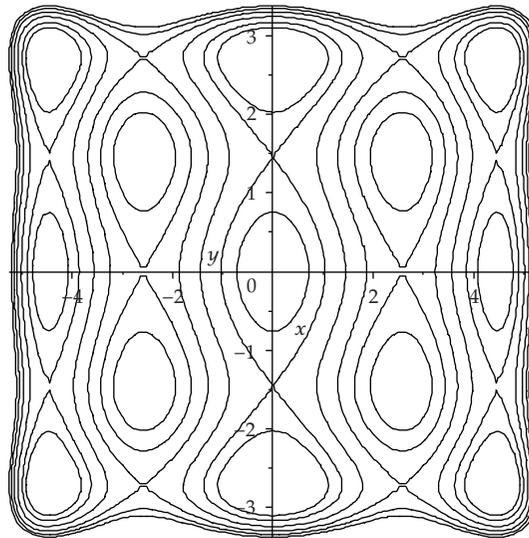
Figure 6: The phase portrait of system (2.13) under the condition **UP**.

enclosing the singular point $(0, 0)$; $\{T^h_{5i}\}$, $h \in (h^s_2, h^s_3)$: two families closed orbits enclosing the singular points $(\pm x_1, \pm y_1)$, $(\pm x_1, \pm y_2)$ and $(\pm x_1, 0)$; $\{T^h_6\}$, $h \in (h^s_2, h^s_3)$: one family closed orbit enclosing the singular points $(0, 0)$, $(0, \pm y_1)$ and $(0, \pm y_2)$; $\{T^h_{7i}\}$, $h \in (h^s_3, h^s_4)$: two families closed orbits enclosing the singular points $(\pm x_2, \pm y_2)$, $(\pm x_2, 0)$; $\{T^h_{8i}\}$, $h \in (h^s_4, h^c_5)$: four families closed orbits enclosing the singular points $(\pm x_2, \pm y_2)$; $\{T^h_9\}$, $h \in (h^s_3, \infty)$: the global closed orbit enclosing all the singular points.

Notice that as h increases, the periodic orbits T_7 and T_8 contract inwards all other periodic orbits expand outwards.

(a) $h = h_0^c$

(b) $h_0^c < h < h_1^c$

(c) $h = h_1^c$

(d) $h_1^c < h < h_2^s$

(e) $h = h_2^s$

(f) $h_2^s < h < h_3^s$

(g) $h = h_3^s$

(h) $h_3^s < h < h_4^s$

(i) $h = h_4^s$

(j) $h_4^s < h < h_5^c$

(k) $h = h_5^c$
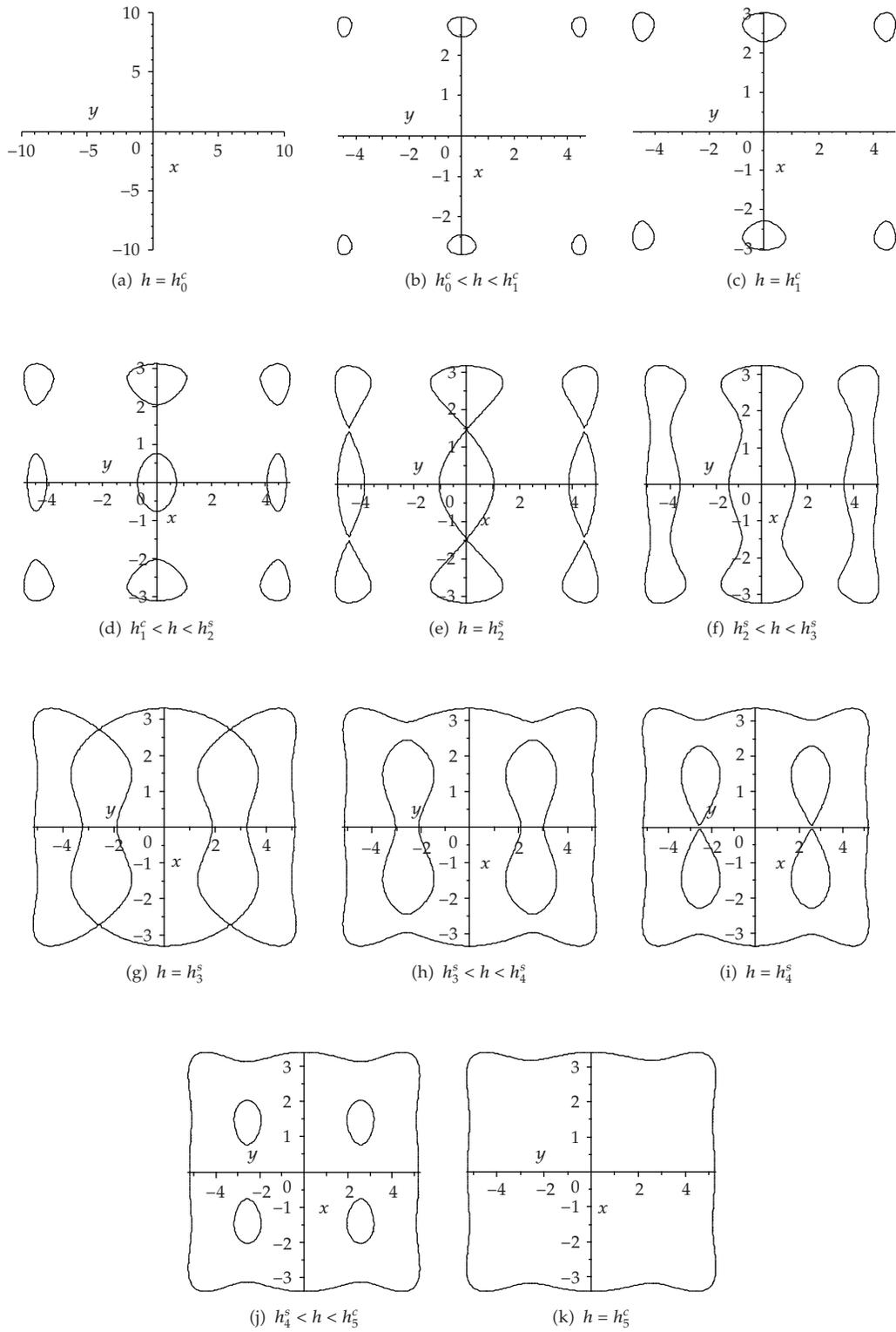
Figure 7: The changing process of the phase portraits as h varied under the condition **UP.**
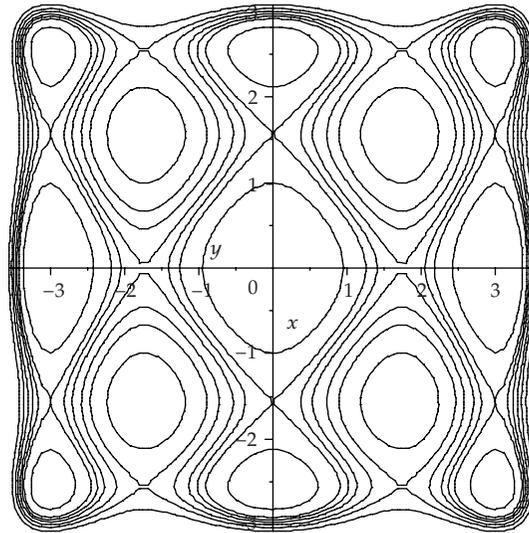
Figure 8: Phase portraits of system (2.13) in the region I.
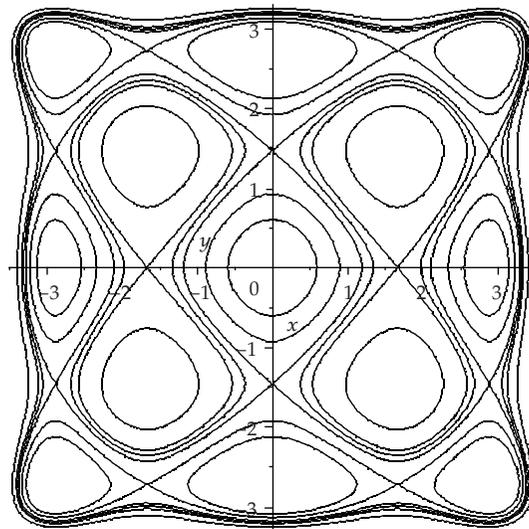
Figure 9: Phase portraits of system (2.13) in the region III.

The phase portraits are topologically equivalent to the Figure 6 when the parameters $(a, b) \in II$.

Case II

When a through the line $a = \sqrt{3}/3$ from right to left that means the region II changes into the region I, the bifurcation happened. Figure 8 show the phase portraits of system (2.13) in the region I.

The phase portraits are topologically equivalent to Figure 8 when the parameters $(a, b) \in I$.

Case III

When b through the curven $b = 1/27t$, $t = (4 - 16a^2)\sqrt{4a^2 - 1}/3$ from up to down that means the region II changes into the region III, the bifurcation happened. Figure 9 shows the phase portraits of system (2.13) in the region III.

The phase portraits are topologically equivalent to Figure 9 when the parameters $(a, b) \in III$.

Based on the above analysis, we know that the phase portraits are different from each other because there are in different regions.

5. Conclusions

The dynamic characteristics of a molecular diffusion model are considered in this paper, by using the bifurcation theory of planar dynamical system, Laplace transform, and maple software. First the model equation was transformed to the averaged equation by means of Laplace transform, and we obtain that when bifurcation parameters (a, b) through the line $a = \sqrt{3}/3$ and the curve $b = 1/27t$ the interphase structure will change suddenly, and then we get the phase portraits of system (2.13) in different regions. The results obtained enriches the dynamic research contents of the Liquid-liquid system, and have important theory values and apply values for further study of the IDZ model as well as to find the change regions in which the mass transfer rate maybe change suddenly.

Acknowledgments

References

[1] V. V. Tarasov and D. X. Zhang, "Interfacial layer in nonequilibrium liquid-liquid systems," *Doklady Physical Chemistry*, vol. 350, no. 4–6, p. 271, 1996.

[2] D. X. Zhang, T. Shiyu, X. Rongshu, and V. V. Tarasov, "Mass transfer characteristic of interfacial layer activated by periodic mechanism in liquid-liquid system," *Journal of Chemical Industry and Engineering*, vol. 51, no. 1, pp. 108–114, 2000 (Chinese).

[3] D. Zhang and V. V. Tarasov, "Interfacial phenomena in liquid-liquid systems: new explanation of some peculiarities of mass transfer with chemical reactions," *Journal of Chemical Industry and Engineering*, vol. 52, no. 8, pp. 701–707, 2001 (Chinese).

[4] D. X. Zhang and V. V. Tarasov, "Electrostatic field effects on mass exchange of oxalic acid and copper di-2-ethylhexyl phosphate," *Russian Journal of Physical Chemistry A*, vol. 82, no. 3, pp. 506–511, 2008.

[5] D. X. Zhang, A.-M. Li, X. Xu, and V. V. Tarasov, "Effect of surfactants and static electric field on characteristics of liquid-liquid mass transfer kinetics," *Transaction of Beijing Institute of Technology*, vol. 28, no. 3, pp. 267–270, 2008.

[6] V. A. Vasilenko, E. M. Kol'tsova, V. V. Tarasov, Z. D. Xiang, and L. S. Gordeev, "Methods of nonequilibrium thermodynamics for studying and simulating mass transfer processes in liquid-liquid systems," *Theoretical Foundations of Chemical Engineering*, vol. 41, no. 5, pp. 500–505, 2007.

[7] D. X. Zhang and E. M. Koltsova, "Non-linear dynamics of mass transfer process with an interphase fractal structure in liquid-liquid system," *Journal of Physics: Conference Series*, vol. 96, no. 1, Article ID 012187, 8 pages, 2008.

[8] D. X. Zhang, Y. Z. Liao, Y. Tu, and H. S. Li, "Preparation of ultra fine copper oxide powder using precipitation on the interfaces of microemulsions," *Transaction of Beijing Institute of Technology*, vol. 25, no. 5, pp. 466–470, 2005.

[9] J. Li, S. F. Miao, and W. Zhang, "Analysis on bifurcations of multiple limit cycles for a parametrically and externally excited mechanical system," *Chaos, Solitons & Fractals*, vol. 31, no. 4, pp. 960–976, 2007.

[10] J. Li, Y. Tian, W. Zhang, and S. F. Miao, "Bifurcation of multiple limit cycles for a rotor-active magnetic bearings system with time-varying stiffness," *International Journal of Bifurcation and Chaos*, vol. 18, no. 3, pp. 755–778, 2008.

[11] V. V. Kapharov, *Fundamentals of Mass Transfer*, High School, Moscow, Russia, 1972.

[12] Z. D. Zhang and Q. S. Bi, "Bifurcation analysis of traveling wave solutions of a generalized Camassa-Holm equation," in *Proceedings of the 7th Conference on Nonlinear Dynamics*, Nanjing, China, October 2004.

Research Article

Continuation of Periodic Solutions of Dissipative and Conservative Systems: Application to Elastic Pendulum

P. Pokorny

Department of Mathematics, Prague Institute of Chemical Technology, Technicka 5,
166 28 Prague, Czech Republic

Correspondence should be addressed to P. Pokorny, pavel.pokorny@vscht.cz

Continuation is an efficient algorithm for finding solutions of systems of nonlinear algebraic equations where the solutions form a one-dimensional continuum. Such systems arise naturally when investigating equilibrium points and periodic solutions of ordinary differential equations with one parameter. Continuation of isolated periodic solutions of dissipative systems is a well-established technique. Less attention has been devoted to continuation of periodic solutions of conservative systems, where periodic solutions typically form a one-parameter family. To specify a single periodic solution, additional condition must be considered. However, this gives an over-determined system, which has no solution when working with approximate numerical values. We propose a simple algorithm which solves this difficulty by using singular value decomposition of the Jacobian matrix. This algorithm is applied to the conservative model of elastic pendulum. A branch of periodic solutions with constant energy is found which is born by the period doubling bifurcation of vertical oscillations.

1. Motivation

When investigating the model of elastic pendulum [1], we have found a period doubling bifurcation when vertical oscillation loses its stability and a new branch of periodic solution is generated. To follow this branch of periodic solutions for varying parameter of the system and for constant energy, we developed a simple algorithm for continuation of periodic solutions in conservative systems. As we believe this algorithm will be useful for investigation of other systems as well we present the problems and solutions related to continuation of periodic solutions of conservative systems in the first half of this paper. The second half describes the results of this algorithm when applied to the model of the elastic pendulum.

Basic introduction to dynamical systems is [2]; principles of continuation can be found in [3]. To follow the second part, we invite the reader to check [1], although detailed knowledge of the second part is not necessary for understanding the algorithm.

2. Continuation of Equilibrium Points

For clarity we sketch the main idea behind continuation of equilibrium points of ODE's. A nice introduction can be found in [3]. Consider a system of ordinary differential equations with one parameter

$$\frac{dx}{dt} = f(x,p), \quad x \in R^n,\ p \in R, \tag{2.1}$$

where f is a smooth function of both x and p. A point x_E is called an equilibrium point of the system (2.1) if $f(x_E, p) = 0$, that is, $x(t) = x_E$ is a constant solution. For a system with one parameter p we want to investigate the dependence $x_E(p)$. The brute force approach would be to choose a fixed value $\tilde{p}$ of the parameter p and then to solve a system of n equations for n unknowns

$$f(x, \tilde{p}) = 0. \tag{2.2}$$

This could be done for various values $\tilde{p}$, possibly from an equidistant grid. A good candidate for the solver is Newton's method.

Instead of solving these problems independently, it would be convenient to use the solution for the previous value of the parameter as the initial approximation for the next (close) value of the parameter.

Continuation makes use of the smoothness of f even more. First, we introduce

$$y = (x,p) \in R^{n+1}, \qquad F(y) = f(x,p). \tag{2.3}$$

Then the equation to solve is

$$F(y) = 0, \tag{2.4}$$

where $F : R^{n+1} \to R^n$. We assume that the solutions form a curve in R^{n+1}.

Continuation is a predictor—corrector method. Given one point y_o on the curve, we first find the prediction y_p, and then we find its correction y_c. Then we iterate this procedure. As a result we find a series of points approximating the curve in R^{n+1}. As the corrector, we can use Newton's method after we add one more equation

$$(y - y_p) \cdot (y_p - y_o) = 0 \tag{2.5}$$

to get a system of $n + 1$ equations for $n + 1$ unknowns. The geometrical meaning of this additional equation is that we want to find a point y on the curve such that the correction $y - y_p$ is perpendicular to the predicted displacement $y_p - y_o$.

3. Continuation of Periodic Solutions

3.1. Previous Results

Continuation of periodic orbits of dissipative systems is a well-established technique [3–5]. Less attention has been devoted to continuation of periodic orbits of conservative systems. In [6] a conservative system is embedded in a one-parameter family of dissipative systems by adding a small gradient perturbation term to the vector field in such a way that a periodic orbit can only exist when the perturbation is zero. This has been generalized to systems having more than one first integrals in [7].

We use a different approach. It can be summarized as follows. To identify one periodic solution, additional equations are needed. This gives an overdetermined system, which has no solution when working with approximate values. Singular value decomposition is used to find the correction that minimizes the residue. This paper describes the principles, problems, and remedy of continuation of periodic solutions.

3.2. Dissipative Systems

Before we go to conservative systems, we describe continuation of periodic solutions of dissipative systems.

First, we rescale the right-hand side of the ODE. If the equation

$$\frac{dx}{dt} = f(x, p), \quad x \in R^n, \ p \in R \tag{3.1}$$

has a periodic solution with the least period $T > 0$, then

$$\frac{dx}{dt} = g(x, p, T), \tag{3.2}$$

where

$$g(x, p, T) = Tf(x, p), \tag{3.3}$$

has a periodic solution with the least period 1. For other than periodic solutions T may be any fixed positive constant.

Let us denote the flow of (3.2) by $\phi(t, x, p, T)$, that is, the solution $x(t)$ of (3.2), satisfying the initial condition

$$x(0) = x_{\text{ic}}, \tag{3.4}$$

can be written as

$$x(t) = \phi(t, x_{\text{ic}}, p, T). \tag{3.5}$$

To find a periodic solution, that is, a solution satisfying

$$\phi(1, x, p, T) - x = 0, \tag{3.6}$$

we have $n + 2$ unknowns. To use Newton's method for the correction we need to add two more equations:

$$(x - x_o) \cdot g(x_o, p, T) = 0 \tag{3.7}$$

(which is called the phase condition) and, as stated before,

$$(y - y_p) \cdot (y_p - y_o) = 0, \tag{3.8}$$

where $y = (x, p, T) \in R^{n+2}$.

The geometrical meaning of (3.7) is that we want the change $x - x_o$ to be perpendicular to the tangent vector of the trajectory. This is a natural condition because any point on the periodic trajectory satisfies (3.6) and we want to remove this degeneracy.

For an isolated periodic solution, which is a typical case for dissipative ODE's, the system of $n + 2$ algebraic equations (3.6), (3.7), and (3.8) for $n + 2$ unknowns x, p, T can be solved by Newton's method. Unfortunately this is not a typical case for conservative systems which we discuss in the next section.

3.3. Conservative Systems

3.3.1. Constant Parameter

In this subsection we assume that the parameter p is fixed and we leave it out.

In a typical case, periodic solutions to a conservative system form a one parameter family. We can choose one solution by adding one more equation

$$E(x, \dot{x}) = E_0, \tag{3.9}$$

namely, the condition of constant total energy. However, then we get a numerically ill-conditioned system.

This may be illustrated by the simplest possible case—the harmonic oscillator

$$\ddot{x} + x = 0, \quad x \in R. \tag{3.10}$$

The general solution is (up to a time shift)

$$x = a \cos t. \tag{3.11}$$

The constant a (which is arbitrary) is related to the total energy

$$E(x, \dot{x}) = \frac{x^2}{2} + \frac{\dot{x}^2}{2} = \frac{a^2}{2}, \tag{3.12}$$

which is conserved. In polar coordinates r, φ the system can be expressed as

$$\dot{r} = 0, \qquad \dot{\varphi} = 1. \tag{3.13}$$

The flow of this system is (omitting p and T)

$$\phi(t, r_0, \varphi_0) = (r_0, \varphi_0 + t), \tag{3.14}$$

and the condition for periodic solution

$$\phi(T, r, \varphi) - (r, \varphi) = (0, 2\pi) \tag{3.15}$$

is

$$0 = 0, \qquad T = 2\pi. \tag{3.16}$$

The condition of the total energy being E_0 gives the equation

$$r = \sqrt{2E_0}, \tag{3.17}$$

and the phase condition results in

$$\varphi = \varphi_0. \tag{3.18}$$

Now, the system (3.16)–(3.18) has a unique solution, and there seems to be no problem. There is none indeed when we can solve it analytically and exactly. However, in most cases we have to solve it only approximately and numerically using Newton's method.

When we rewrite the system (3.16)–(3.18) in single line

$$\mathcal{F}(T, r, \varphi) = (0, 0, 0, 0), \tag{3.19}$$

where

$$\mathcal{F}(T, r, \varphi) = \left(0, T - 2\pi, r - \sqrt{2E_0}, \varphi - \varphi_0\right), \tag{3.20}$$

then the Jacobi matrix of partial derivatives of $\mathcal{F}$, used in Newton's method, is

$$
J = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.21}
$$

(here we differentiate with respect to T, r, φ, resp.). When evaluated numerically, J will have all entries nonzero and approximate values. Then we get an overdetermined system which has no solution in a typical case. Singular value decomposition is a good approach to such systems. It is explained in the next section.

3.3.2. Singular Value Decomposition

Any $m \times n$ matrix A, whose number of rows m is greater than or equal to the number of columns n, can be written (see [8]) as the product of three matrices: an $m \times n$ column-orthogonal matrix U, an $n \times n$ diagonal matrix W with positive or zero elements on the diagonal (the singular values), and the transpose of an $n \times n$ orthogonal matrix V:

$$
A = U \cdot W \cdot V^T. \tag{3.22}
$$

If the system of equations

$$
A \cdot x = b \tag{3.23}
$$

has a unique solution, then it can be found as

$$
x = V \cdot W^{-1} \cdot U^T \cdot b. \tag{3.24}
$$

Here W^{-1} is the inverse to W if it exists (which is the case when all the entries w_{ii} on the diagonal of the matrix W are nonzero). In general the matrix W^{-1} is a diagonal matrix with entries that are computed by the following formula:

$$
(W^{-1})_{ii} = \begin{cases} \dfrac{1}{w_{ii}}, & \text{if } w_{ii} \neq 0 \\ 0, & \text{if } w_{ii} = 0. \end{cases} \tag{3.25}
$$

If there is no solution of (3.23) which is a typical case for an overdetermined system, then (3.24) gives the least squares solution, that is, the solution that minimizes the norm of the residue

$$
A \cdot x - b. \tag{3.26}
$$

Even in cases when (3.23) has infinitely many solutions, then (3.24) gives a reasonable result: the solution with the least norm.

In numerical computation we find the singular value decomposition of the given matrix A, and then to find W^{-1} we treat values w_{ii} in the matrix W below a certain threshold (say 10^{-6}) as zero. This gives a robust algorithm.

Note on dimensions of the matrices: there is another possible choice of the dimensions of the matrices in the decomposition (3.22), namely, U of type $m \times m$, W of type $m \times n$ and V of type $n \times n$ as before. This choice is more convenient if $m < n$. In our case, when $m > n$, this choice means that we add $m - n$ zero rows to the matrix W and we add $m - n$ columns to the matrix U, but these extra columns are insignificant, because they are multiplied by zero. Up to this formal difference, the two choices are equivalent.

3.3.3. Varying Parameter

Now, when we described the difficulty and the remedy with periodic solutions of conservative systems with a fixed parameter, we want to build the final algorithm for continuation of periodic solutions of conservative systems with one varying parameter.

Remember that adding one parameter in the system

$$f(x, p) = 0 \tag{3.27}$$

when searching equilibrium points, we appended the parameter p to the state vector x to get $y = (x, p) \in R^{n+1}$, and we had to add the orthogonality condition (2.5), saying that the correction should by orthogonal to the prediction.

For continuation of periodic solutions of dissipative systems, we append not only the parameter p but also the unknown period T to get $y = (x, p, T) \in R^{n+2}$. Then we had to add not only the orthogonality condition (3.8) but also the phase condition (3.7) to get the system of $n + 2$ equations for $n + 2$ unknowns to solve in the corrector step (by Newton's method).

For continuation of periodic solutions of conservative systems, we have a three-dimensional continuum of solutions of

$$\phi(1, x, p, T) - x = 0 \tag{3.28}$$

in a typical case. To pick out a single point we need three additional conditions: the orthogonality condition for $y = (x, p, T)$

$$(y - y_p) \cdot (y_p - y_0) = 0, \tag{3.29}$$

the phase condition

$$(x - x_0) \cdot g(x_0, p, T) = 0, \tag{3.30}$$

and one more extra condition. A possible choice for this extra condition is the constant energy condition

$$E(y) - E_0 = 0, \tag{3.31}$$

where E_0 is a fixed constant, possibly depending on the parameter p. A general form of this condition can be written as

$$\psi(E(y), p) = 0, \tag{3.32}$$

where ψ is a smooth function.

With these three conditions however we get an overdetermined system which can be conveniently solved by the singular value decomposition in each corrector step.

If we write (3.28)–(3.31) in a single line as

$$\mathcal{F}(y) = 0, \tag{3.33}$$

where $\mathcal{F} : R^{n+2} \rightarrow R^{n+3}$, which stands for $n + 3$ equations for $n + 2$ unknowns, then in each corrector step we can solve (3.33) by Newton-like iterations

$$y_{k+1} = y_k - \Delta y_k, \tag{3.34}$$

where Δy_k is the solution of $n + 3$ equations for $n + 2$ unknowns

$$J(y_k) \cdot \Delta y_k = \mathcal{F}(y_k) \tag{3.35}$$

with

$$J_{ij} = \frac{\partial \mathcal{F}_i}{\partial y_j}. \tag{3.36}$$

As we have shown above, J is not a regular square matrix, so we solve (3.35) by the singular value decomposition: first, we decompose J into

$$J = U \cdot W \cdot V^T \tag{3.37}$$

and then we find

$$\Delta y_k = V \cdot W^{-1} \cdot U^T \cdot \mathcal{F}(y_k), \tag{3.38}$$

where, as described above, we compute W^{-1} by replacing large values w_{ii} on the diagonal with w_{ii}^{-1} while replacing low values with 0.

3.3.4. Damping

It is safer to replace (3.34) with

$$y_{k+1} = y_k - q\Delta y_k, \tag{3.39}$$

where the damping factor q is found in the following way. We start with $q = 1$; we compute y_{k+1} and $\mathcal{F}(y_{k+1})$. Then we compare the norms of residues; if $|\mathcal{F}(y_{k+1})| < |F(y_k)|$, we are done. If not, we repeatedly decrease q (by a factor, say, 2) and we recompute y_{k+1} and $\mathcal{F}(y_{k+1})$ until we get a lower residue. If this is not achieved in a small number of iterations (say, 20) then a better prediction must be done with a smaller step size.

3.3.5. No Threshold Needed

Two more technical questions remain open. First, when computing W^{-1}, large values w_{ii} are replaced with w_{ii}^{-1} and low values are replaced with 0. How to choose the level to decide whether a value is large or low? Instead of choosing one fixed level, we use a different approach. We sort the singular values w_{ii} in decreasing order. First we replace the lowest singular value with zero, we compute Δy_k and then y_{k+1} and $|\mathcal{F}(y_{k+1})|$. Then we replace the two lowest singular values and compute $|\mathcal{F}(y_{k+1})|$ again. This is done for all possible choices and the case giving the lowest residue $|\mathcal{F}(y_{k+1})|$ is used.

3.3.6. How Many Iterations

The other question is how many iterations (3.34) should be done. Initially the residue $|\mathcal{F}(y_{k+1})|$ decreases with k until the noise level (given by the round-off error and by the error of the methods used) is reached. Then the residue attains similar values. This suggests to compute further iterations until the residue fails to decrease.

To illustrate this algorithm, we apply it to the model of elastic pendulum in the next section.

4. Elastic Pendulum

Elastic pendulum is a simple model that exhibits a wide and surprising range of dynamic phenomena [9–15]. The first known publication about the elastic pendulum is the paper [16] by Vitt and Gorelik. Deterministic chaos has been observed in numerical simulation and presented in the form of Poincaré section, autocorrelation function, Lyapunov exponent and power spectrum in [17–19]. The bifurcation diagram of a plane elastic pendulum is sketched in [20]. The most thorough treatment of small amplitude oscillation of both plane and space elastic pendulum can be found in the works by Peter Lynch [21–24].

Consider a pendulum consisting of a point bob of mass m_B suspended on a homogeneous elastic spring of mass m_S with the elasticity constant k; see Figure 1. Assume a homogeneous gravitational field $(0, -g)$.

In [1] we derive the equations of motion

$$\ddot{x}_1 = \left(\frac{1}{\sqrt{x_1^2 + x_2^2}} - 1 \right) x_1,$$

$$\ddot{x}_2 = \left(\frac{1}{\sqrt{x_1^2 + x_2^2}} - 1 \right) x_2 - p. \tag{4.1}$$
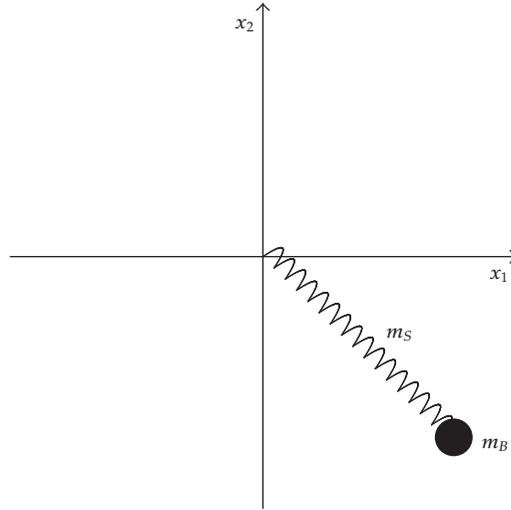
Figure 1: Heavy spring elastic pendulum consists of a point mass m_B attached to a spring with mass m_S which is fixed at the other end point.

This system has the only parameter

$$p = \frac{(m_B + m_S/2)g}{k\ell_0}, \tag{4.2}$$

where ℓ_0 is the length of the unloaded spring.

In [1] we also derive the stability condition for the vertical oscillations

$$x_1 = 0, \qquad x_2 = -1 - p + a \sin t, \tag{4.3}$$

where a is the amplitude of the vertical oscillation given by initial conditions. The vertical oscillation of the heavy spring elastic pendulum with the relative amplitude a is stable (see [1]) for

$$|a| < |a_C(p)|, \tag{4.4}$$

where

$$a_C(p) \doteq (3p)^{2/3} - 1. \tag{4.5}$$

The error of this estimate of $a_C(p)$ is less than 0.005 for $0 < p < 0.6$.

Figure 2 shows the regions in the p-a plane with stable (light gray, or blue in color version) and unstable (dark gray, or red in color version) vertical oscillation with amplitude a and the parameter p. For p and a sufficiently small, the vertical oscillations are stable. For a fixed $a = 0.1$ let us observe what happens when we increase p. When p crosses the critical value $p_1 \doteq (1/3)(1 - a)^{3/2} \doteq 0.2846$ the vertical oscillations (with the period 2π) lose their
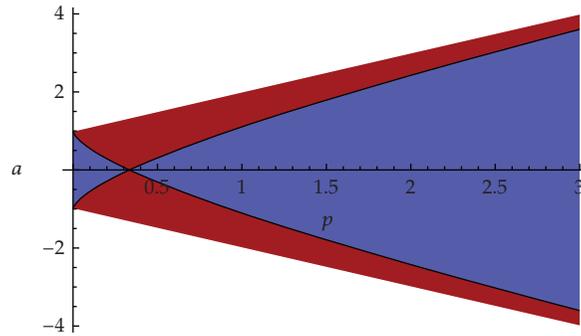
Figure 2: The stable region (light gray, or blue in color version) and the unstable region (dark gray, or red in color version) for the vertical oscillation with amplitude a and the external field p.
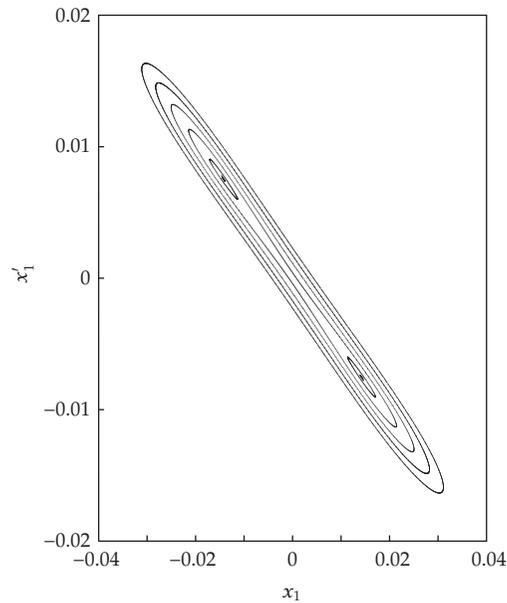
Figure 3: Poincaré section can be used to find the initial approximation of the periodic orbit, which is represented by an isolated point while surrounding closed curves correspond to invariant tori.

stability, one eigenvalue of the Jacobi matrix leaves the unit circle through the point -1 in the complex plane, and a period doubling bifurcation gives rise to a family of periodic oscillations (with the period close to 4π). Another period doubling bifurcation exists when decreasing the parameter p for the critical value $p_2 \doteq (1/3)(1+a)^{3/2} \doteq 0.3846$. We want to find the branch of these periodic oscillations.

As there is a continuum of periodic solutions for one value of the parameter, we want to pick out one of these periodic solutions by specifying the energy. The total energy of the system consists of the potential and of the kinetic energy

$$E(x_1, x_2, \dot{x}_1, \dot{x}_2) = E_{\text{pot}}(x_1, x_2) + E_{\text{kin}}(\dot{x}_1, \dot{x}_2), \tag{4.6}$$
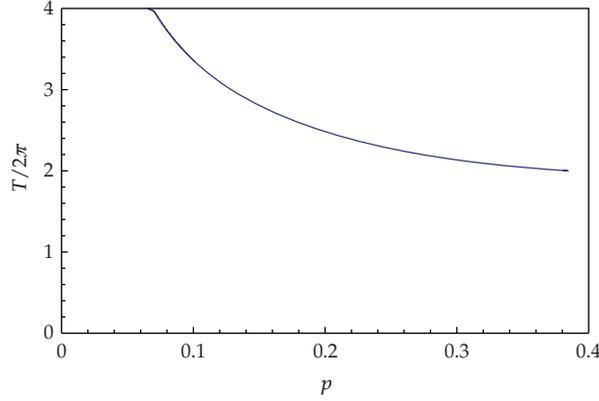
Figure 4: The period T of periodic oscillation (as a function of the parameter p) changes from 4π for high p up to 8π for low p.

where

$$E_{\text{pot}} = \frac{1}{2}\left(\sqrt{x_1^2 + x_2^2} - 1\right)^2 + px_2, \qquad E_{\text{kin}} = \frac{1}{2}\left(\dot{x}_1^2 + \dot{x}_2^2\right). \tag{4.7}$$

We are interested in periodic solutions with the total energy

$$E(x_1, x_2, \dot{x}_1, \dot{x}_2) = E_0, \tag{4.8}$$

where E_0 is the energy of the vertical oscillations (4.3):

$$E_0 = E(0, -1 - p + a\sin t, 0, a\cos t) = \frac{1}{2}\left(a^2 - p^2 - 2p\right). \tag{4.9}$$

To find the initial approximation of one periodic solution, we choose a parameter value $p < p_2$ and $a = 0.1$ Then we can use numerical integration and Poincaré section given by the horizontal plane $x_2 = -1 - p$ going through the only stable equilibrium point of the system; see Figure 3. For various initial conditions we get various closed invariant curves densely filled by the intersections of the trajectory with the Poincaré section. From this picture we can estimate the coordinates x_1 and $\dot{x}_1$. From the position of the Poincaré section we know that $x_2 = -1 - p$ and we compute $\dot{x}_2$ from the known total energy. We estimate the period to be $T = 4\pi$ because the period of vertical oscillations before the period doubling bifurcation is 2π. Finally we choose the initial prediction vector (determining the direction in which the curve will be traveled) $v_0 = (0, 0, 0, 0, -1, 0)$ meaning that we want to start with decreasing parameter p. Now we can start the continuation. The results are discussed in the next section.

5. Results

In this section we illustrate the above algorithm for continuation of periodic solutions by the Hamiltonian model of elastic pendulum. We choose the fixed energy corresponding to the amplitude of vertical oscillation with amplitude $a = 0.1$. When decreasing the parameter p,
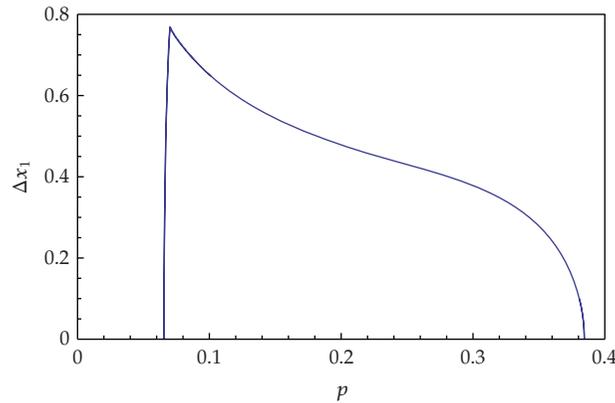
Figure 5: The range $\Delta x_1 = x_{1\max} - x_{1\min}$ of periodic oscillation as a function of the parameter p.
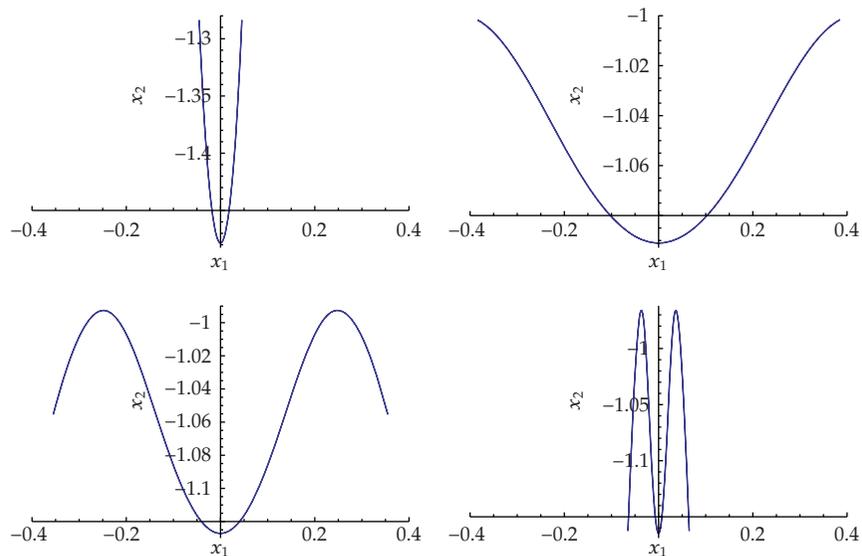
Figure 6: The projection of the trajectory to the x_1-x_2 plane for four values of the decreasing parameter p. Top left: For p, slightly below the critical value $p_2 \doteq 0.3846$ the period doubling bifurcation gives rise to the new oscillation with twice the period and with a small x_1 range. Top right: For decreasing p the cup-like shape of the trajectory widens and the top parts start to bend. Bottom left: For still lower p values the outer parts of the trajectory bend even more and the period grows. Bottom right: Finally the period is close to 8π which is four times the period of the vertical oscillation. For $p = p_3 \doteq 0.0655$ our periodic oscillation merges with the vertical oscillation.

the vertical oscillation becomes unstable for the critical value $p_2 \doteq 0.3846$. Period doubling bifurcation gives rise to a one parameter family of periodic solutions that exist down to the parameter value $p_3 \doteq 0.0655$. We use the above algorithm for continuation of this branch of periodic solutions.

Figure 4 shows how the period T of the periodic solution grows with decreasing parameter p from the value $T = 4\pi$ just after the period doubling bifurcation up to the value $T = 8\pi$.

Figure 5 shows for the same parameter values how the range of x_1 coordinates (the horizontal displacement of the bob) grows and then decreases down to zero.

When decreasing the parameter p the shape of the trajectory in the x_1-x_2 plane changes in an interesting way. This can be illustrated by plotting the trajectory for a few parameter values see Figure 6. First, as the vertical oscillation with period 2π loses its stability a new periodic solution is born with twice the period and a small range of x_1. As the parameter p goes down, the range of x_1 grows but the cup-like shape of the oscillation remains. For a still decreasing parameter p the top parts of the trajectory begin to bend and then they turn down. The period T grows. For parameter values close to $p_3 \doteq 0.0655$ the trajectory is four times folded and the range of x_1 decreases to zero. The period is now close to 8π which is four times the period of vertical oscillation. Finally our periodic solution merges with the vertical oscillation.

6. Conclusion

The difference between continuation of periodic solutions of dissipative and conservative systems was described. For dissipative systems periodic solutions are generically isolated. To specify a single point on the trajectory we need to add the orthogonality condition and the phase condition. Then continuation of periodic solutions can be done by the predictor—corrector algorithm. For conservative systems, however, periodic solutions form one-dimensional continuum parametrized by the energy. To specify a single solution we must add one more equation. When working with approximate numerical values, this yields an overdetermined system in the corrector step. To solve this problem we suggest to use singular value decomposition. This gives a simple algorithm that was tested by the Hamiltonian model of elastic pendulum. Using stability analysis from [1] we knew that a period doubling bifurcation exists for this model for a known parameter value. In this bifurcation the vertical oscillation loses its stability and a branch of new periodic solutions is generated. In the present paper we used the suggested algorithm to compute this new branch of periodic solutions for varying parameter p. It turns out that the new periodic solution starts with twice the period of the vertical oscillation, it gets wider, it bends and folds, its period grows to 8π which is four times the period of the vertical oscillation, and finally the new oscillation merges with the vertical oscillation again.

Acknowledgments

References

[1] P. Pokorny, "Stability condition for vertical oscillation of 3-dim heavy spring elastic pendulum," *Regular and Chaotic Dynamics*, vol. 13, no. 3, pp. 155–165, 2008.

[2] M. W. Hirsch, S. Smale, and R. L. Devaney, *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, vol. 60 of *Pure and Applied Mathematics*, Elsevier/Academic Press, Amsterdam, The Netherlands, 2nd edition, 2004.

[3] R. Seydel, "Tutorial on continuation," *International Journal of Bifurcation and Chaos*, vol. 1, no. 1, pp. 3–11, 1991.

[4] E. J. Doedel, R. C. Paffenroth, A. R. Champneys, et al., "AUTO2000: continuation and bifurcation software for ordinary differential equations," Tech. Rep., Department of Computer Science, Concordia University, Montreal, Canada, 2000.

[5] M. Marek and I. Schreiber, *Chaotic Behaviour of Deterministic Dissipative Systems*, Cambridge University Press, Cambridge, UK, 1991.

[6] J.-A. Sepulchre and R. S. MacKay, "Localized oscillations in conservative or dissipative networks of weakly coupled autonomous oscillators," *Nonlinearity*, vol. 10, no. 3, pp. 679–713, 1997.

[7] F. J. Muñoz-Almaraz, E. Freire, J. Galán, E. Doedel, and A. Vanderbauwhede, "Continuation of periodic orbits in conservative and Hamiltonian systems," *Physica D*, vol. 181, no. 1-2, pp. 1–38, 2003.

[8] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 2nd edition, 1992.

[9] M. G. Olsson, "Why does a mass on a spring sometimes misbehave?" *American Journal of Physics*, vol. 44, no. 12, pp. 1211–1212, 1976.

[10] T. E. Cayton, "The laboratory spring-mass oscillator: an example of parametric instability," *American Journal of Physics*, vol. 45, no. 8, pp. 723–732, 1977.

[11] M. G. Rusbridge, "Motion of the sprung pendulum," *American Journal of Physics*, vol. 48, no. 2, pp. 146–151, 1980.

[12] E. Breitenberger and R. D. Mueller, "The elastic pendulum: a nonlinear paradigm," *Journal of Mathematical Physics*, vol. 22, no. 6, pp. 1196–1210, 1981.

[13] H. M. Lai, "On the recurrence phenomenon of a resonant spring pendulum," *American Journal of Physics*, vol. 52, no. 3, pp. 219–223, 1984.

[14] B. A. Anicin, D. M. Davidovic, and V. M. Babovic, "On the linear theory of the elastic pendulum," *European Journal of Physics*, vol. 14, no. 3, pp. 132–135, 1993.

[15] D. M. Davidović, B. A. Aničin, and V. M. Babović, "The libration limits of the elastic pendulum," *American Journal of Physics*, vol. 64, no. 3, pp. 338–342, 1996.

[16] A. Vitt and G. Gorelik, "Oscillations of an elastic pendulum as an example of the oscillations of two parametrically coupled linear systems," *Journal of Technical Physics*, vol. 3, no. 2-3, pp. 294–307, 1933, translated from Russian by Lisa Shields with an Introduction by Peter Lynch.

[17] H. N. Núñez-Yépez, A. L. Salas-Brito, C. A. Vargas, and L. Vicente, "Onset of chaos in an extensible pendulum," *Physics Letters A*, vol. 145, no. 2-3, pp. 101–105, 1990.

[18] R. Cuerno, A. F. Rañada, and J. J. Ruiz-Lorenzo, "Deterministic chaos in the elastic pendulum: a simple laboratory for nonlinear dynamics," *American Journal of Physics*, vol. 60, no. 1, pp. 73–79, 1992.

[19] R. Carretero-Gonzalez, H. N. Nunez-Yepez, and A. L. Salas-Brito, "Regular and chaotic behaviour in an extensible pendulum," *European Journal of Physics*, vol. 15, no. 3, pp. 139–148, 1994.

[20] S. V. Kuznetsov, "The motion of the elastic pendulum," *Regular and Chaotic Dynamics*, vol. 4, no. 3, pp. 3–12, 1999.

[21] P. Lynch, "Resonant motions of the three-dimensional elastic pendulum," *International Journal of Non-Linear Mechanics*, vol. 37, no. 2, pp. 345–367, 2002.

[22] P. Lynch, "The swinging spring: a simple model of atmospheric balance," in *Large-Scale Atmosphere-Ocean Dynamics, Vol. II*, J. Norbury and I. Roulstone, Eds., pp. 64–108, Cambridge University Press, Cambridge, UK, 2002.

[23] D. D. Holm and P. Lynch, "Stepwise precession of the resonant swinging spring," *SIAM Journal on Applied Dynamical Systems*, vol. 1, no. 1, pp. 44–64, 2002.

[24] P. Lynch and C. Houghton, "Pulsation and precession of the resonant swinging spring," *Physica D*, vol. 190, no. 1-2, pp. 38–62, 2004.

Research Article

Dynamical Analysis of an Interleaved Single Inductor TITO Switching Regulator

Abdelali El Aroudi,[1] Vanessa Moreno-Font,[2] and Luis Benadero[2]

[1] *Departament d'Enginyeria Electrònica, Elèctrica i Automàtica (DEEEA), Universitat Rovira i Virgili, Tarragona, Spain*
[2] *Departament de Física Aplicada, Universitat Politècnica de Catalunya, Barcelona, Spain*

Correspondence should be addressed to Abdelali El Aroudi, abdelali.elaroudi@urv.cat

We study the dynamical behavior of a single inductor two inputs two outputs (SITITO) power electronics DC-DC converter under a current mode control in a PWM interleaved scheme. This system is able to regulate two, generally one positive and one negative, voltages (outputs). The regulation of the outputs is carried out by the modulation of two time intervals within a switching cycle. The value of the regulated voltages is related to both duty cycles (inputs). The stability of the whole nonlinear system is therefore studied without any decoupling. Under certain operating conditions, the dynamical behavior of the system can be modeled by a piecewise linear (PWL) map, which is used to investigate the stability in the parameter space and to detect possible subharmonic oscillations and chaotic behavior. These results are confirmed by numerical one dimensional and two-dimensional bifurcation diagrams and some experimental measurements from a laboratory prototype.

1. Introduction

Switching power converters are widely used in the power management area, due to their high efficiency, low cost and small size [1]. In most of the applications, these systems are used in situations where there is a need to stabilize an output voltage to a desired constant value. However, there are other applications where more than one output must be controlled. For instance, the portable equipments in modern vehicles such mobile phones, MP3 players, PDAs, GPS usually include a variety of loads such as LCD displays, memories, microprocessors, Universal Series Bus (USB) and Hard Disk Drives (HDD). These loads require different operating voltages and load currents and are powered by the rechargeable batteries through DC-DC converters. To make the system run life be longer and its size smaller, more and more system designers are focusing on improving the system power conversion efficiency with advanced power converters topologies. The traditional solution of

using independent converters, one for each output, has the shortcomings of higher number of switches and magnetics components. Besides this option and other dual DC-DC converter configurations [2–6], single inductor multiple input multiple output (SIMIMO) DC-DC converters are, in general, convenient solutions for these low power applications.

Recently, a great variety of complex nonlinear behaviors are studied in simple power electronic circuits like DC-DC converters [7–19]. These studies, which are based on obtaining accurate mathematical models, have allowed a deep understanding of their fundamental properties and dynamic behavior. Obtaining these mathematical models is a traditional challenge for the power electronics engineers and there exist in the recent literature many efforts devoted to this research area (see [2–14] and references therein). The main drawback of the models obtained is their complexity which prevents the obtaining of clear system design criteria in terms of the parameters. Pioneer references about these topics can be found in the works reported in [8, 9], where the first studies about the existence of bifurcations and chaotic behavior in buck converters controlled by PWM are done. We can also mention other related works in this research field in the recent literature [10–13] for the buck converter under voltage mode control and [14–16] for different types of power electronic converters.

Nowadays, there are many works dealing with nonlinear behavior in elementary stand alone [19] and other more complex power electronics circuits such as paralleled DC-DC converters [20], multicell and multilevel converters [21–23] and also for an example of single inductor two inputs two outputs (SITITOs) converters [17].

In this paper, more insights into the modeling and analysis of an SITITO interleaved converter are presented. After presenting the switched model, a systematic approach is described to obtain a simple PWL map that can be used to predict accurately the fast scale (switching) dynamics of the system. The model will be used to get some analytical conditions for stability of the system and for optimizing its performance.

The rest of the paper is organized as follows. Section 2 deals with the description of the SITITO DC-DC converter and the basis for the interleaving control circuit is presented. Then in Section 3, the general mathematical model is given. Under the assumption of perfect output regulation, the general PWL map is derived. Section 4 will deal with stability analysis by using this PWL map. We will use the derived map to get some analytical expressions for stability conditions and to draw some bifurcation curves of the system. In Section 5 we deal with numerical simulations from the PWL map and we obtain some one-dimensional and two-dimensional bifurcation diagrams of the system by using this map. Experimental validation of our analytical and simulation results is presented in Section 6. Finally, some conclusions are given in the last section.

2. Single Inductor TITO DC-DC Regulator

2.1. Power Stage Circuit

The schematic diagram in Figure 1 shows a DC-DC converter with a single inductor for two outputs v_P and v_N with opposite polarity. Both loads, considered here as equivalent resistances R_N and R_P, can be powered from the power source V_{in} by generating a sequence of command signals driving switches S_A and S_B.

2.2. Pulse Width Modulation Strategy with Interleaving

The control strategy for the SITITO topology of Figure 1, which is used in this paper and was proposed in [18], is shown in Figure 2.
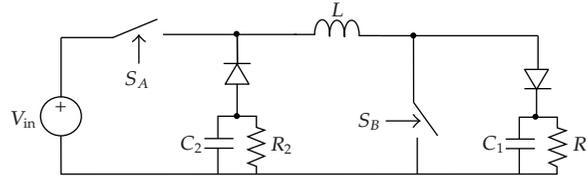
Figure 1: Schematic diagram of a single inductor DC-DC converter with positive v_P and negative v_N output voltages.
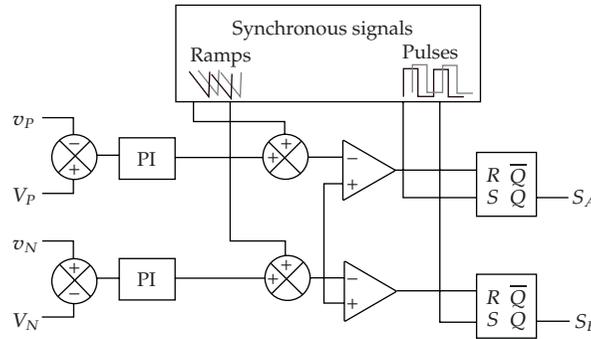
Figure 2: Control strategy of single inductor two outputs converter of Figure 1 based on interleaved pulse width modulation.

This interleaved current mode control works as follows: from the set of input-output errors, two dynamical references, which are associated to each of the channels, are obtained. Each of these reference signals is added to an appropriate phase-shifted (interleaved) compensating ramp v_r^j, so that in the steady state regime, a sequence of time intervals, once per channel, will be produced. In this way, the ramp signal v_r^j, $j \in \{A, B\}$) corresponding to channel j is shifted by a phase shift ϕ_j. Based on current mode control, the general idea is to obtain a current reference I_r^j for each channel (output) and to use two comparators. The particular reference I_r^j is formed by a single compensating ramp and a combination of the output of the PI error filtering blocks. This signal is to be compared to the inductor current i_L. It is demonstrated in [17] that in order to achieve stable behavior, each of the channels should use just the output of the PI corrector coming from the other one. That is switch S_A, which controls the time interval during which current is delivered to the negative output, must be driven accordingly to the error produced by the positive output. Likewise, switch S_B, which affects the positive output, is driven by the error of the negative output. The regulation is possible because of the crossed effect produced when considering the whole system (power stage plus control).

3. Closed Loop Mathematical Modeling

3.1. Switched Model for the SITITO Converter

The switched model of the system can be summarized in the set of five equations (3.1)–(3.5), each of them providing the evolution of the corresponding state variable [17]. The various configurations of the system are taken into account through the binary control signals u_A
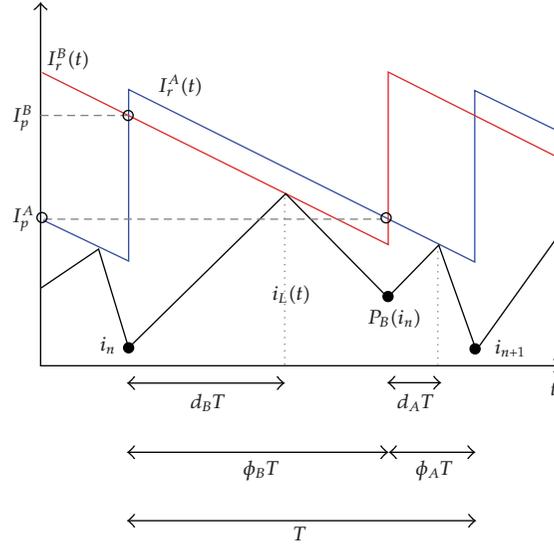
Figure 3: Control signals for the SITITO interleaved DC-DC converter.

and u_B imposed by the state of the switches. This model can be easily obtained by applying standard Kirchhoff's voltage law to the circuit. In doing so, we obtain:

$$\frac{dv_P}{dt} = \frac{1}{C_P}\left(i_L(1 - u_B) - \frac{v_P}{R_P}\right), \tag{3.1}$$

$$\frac{dv_N}{dt} = \frac{1}{C_N}\left(-i_L(1 - u_A) - \frac{v_N}{R_N}\right), \tag{3.2}$$

$$\frac{ds_P}{dt} = \frac{1}{\tau_P}(V_P - v_P), \tag{3.3}$$

$$\frac{ds_N}{dt} = \frac{1}{\tau_N}(v_N - V_N), \tag{3.4}$$

$$\frac{di_L}{dt} = \frac{1}{L}((u_B - 1)v_P + (1 - u_A)v_N - r_L i_L + V_{\text{in}} u_A), \tag{3.5}$$

where v_P is the voltage across the capacitor C_P, v_N is the voltage across the capacitor C_N, i_L is the current through the inductor L whose equivalent series resistance is r_L. R_P and R_N are the resistive loads for both outputs. s_P and s_N are the integral variables, being τ_P, τ_N, V_P and V_N the respective time constants and voltage references. u_A and u_B are the driving binary signal which are 1 if the corresponding switch is "on" and 0 if the switch is "off". To help with comprehension of the action of the control, relevant control signals and parameters are shown in Figure 3. In this figure, I_r^A and I_r^B are the reference currents that can be expressed as:

$$I_r^A(t) = \frac{1}{r_S}\left(g_P(V_P - v_P + s_P) + v_r^A(t)\right),$$

$$I_r^B(t) = \frac{1}{r_S}\left(g_N(v_N - V_N + s_N) + v_r^B(t)\right), \tag{3.6}$$

being $v_r^A(t)$ and $v_r^B(t)$ are the ramp signals defined as following:

$$v_r^A(t) = V_u - (V_u - V_l) \bmod \left(\frac{t}{T}, 1\right),$$

$$v_r^B(t) = V_u - (V_u - V_l) \bmod \left(\frac{t}{T} - \phi_B, 1\right),$$

(3.7)

where $\bmod(\cdot)$ stands for the modulo function and $\phi_A = 1 - \phi_B$.

The first subset of (3.1) and (3.2) refers to the dynamics of the voltage outputs v_P and v_N. Additional subset (3.3) and (3.4) deals with two equations for each of the integral terms s_P and s_N and (3.5) refers to the inductor current dynamics. During each switching period the switch S_A (resp., S_B) is closed for a time duration $d_A T$ (resp., $d_B T$). The duty ratio for each "on" sub-interval (d_A or d_B), which is determined by the action of its comparator, can be expressed implicitly in terms of the state variables as follows:

$$I_r^A(t) - i_L = 0,$$

$$I_r^B(t) - i_L = 0.$$

(3.8)

Obtaining d_A and d_B requires solving (3.8). It should be noted that these conditions should apply to the equations of $I_r^A(t)$ and $I_r^B(t)$ in the configuration before the asynchronous switching instant, and if they are not feasible and $i_L(T) < I_r^j(\phi_j T)$ then d_j is set to ϕ_j whereas it is set to 0 if $i_L(T) > I_r^j(\phi_j T)$. By combining (3.1)–(3.5) with (3.8), we obtain the closed loop switched continuous time model that can be used for computer simulations. Figure 4 depicts the waveforms of the state variables of the system of the above described model from numerical simulation. Note that the output voltages are well regulated to their respective references ($V_P = 3\,\text{V}$ and $V_N = -15\,\text{V}$) and also that the integral variables are almost constant.

3.2. Piecewise Linear Map

In this section we will try to give a one dimensional map that can capture the fast-scale dynamic assuming that the slow dynamics due to the external feedback loops are stable. The complete discrete modeling of the dynamics of a SITITO converter can be described by a five dimensional nonlinear piecewise smooth map. Under the assumption that all outputs are well regulated to their desired references, this map can be simplified by a one-dimensional map in which only the discrete time inductor current $i_n := i_L(nT)$ remains a state variable. This assumption is valid if the period of the modulating signal T is much lower than the time constants or periods (in case of oscillators) of every operating topology and the PI controller is designed to provide a stable slow dynamics. In this situation, the ripples of all capacitor voltages are small and consequently, these variables can be approximated by their averaged values, which are forced by the PI controller to be equal to the reference voltages (Figure 4). The state variables s_P and s_N can also be considered constant whose values can be well obtained by using the averaged model [17]. Due to the different operating modes that the system can present (d_A and d_B nonsaturated and saturated), the one-dimensional map has

(a)

(b)

(c)
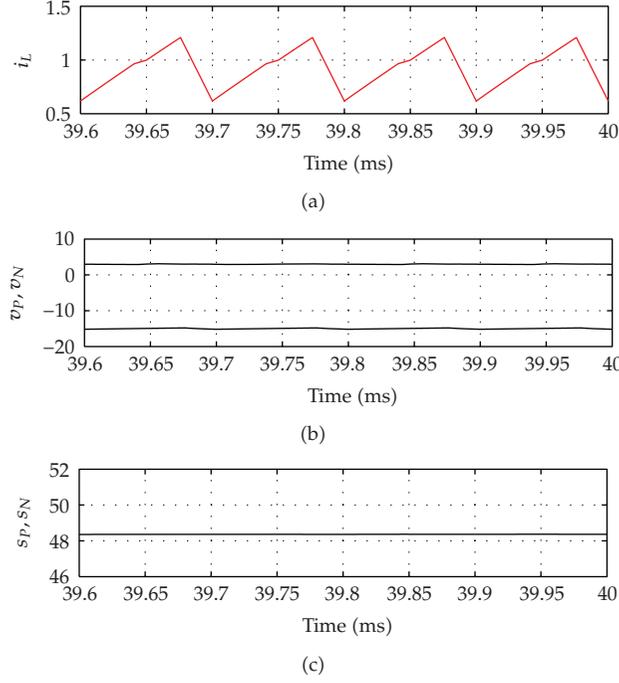
Figure 4: Waveforms of the state variables of the interleaved SITITO converter. $V_{\text{in}} = 6$ V, $V_P = 3$ V, $V_N = -15$ V, $V_u = 1$ A. Other parameter values are those of Table 1.

different forms. Independently of the operating mode, it can be shown that the map can be written in the following form:

$$i_{n+1} = P(i_n) = P_A \circ P_B(i_n), \tag{3.9}$$

where P_j is the local sub-mapping defined by:

$$P_j(x) = \begin{cases} x + m_{\text{on}}\phi_j T, & \text{if } d_j > \phi_j, \\ x + m_{\text{on}}d_j T + m_{\text{off}}^j(\phi_j - d_j)T, & \text{if } d_j < \phi_j, \end{cases} \tag{3.10}$$

where $j \in \{A, B\}$ in the subscript case and $j \in \{P, N\}$) in the superscript case, x is the state variable at the beginning of phase j, and m_{on} and m_{off}^j are the approximated slopes of the inductor current during the "on" and the "off" configuration of both phases, which are given by:

$$m_{\text{on}} = \frac{V_{\text{in}} - r_L I_L}{L},$$

$$m_{\text{off}}^P = \frac{V_{\text{in}} - V_P - r_L I_L}{L}, \tag{3.11}$$

$$m_{\text{off}}^N = \frac{V_N - r_L I_L}{L}.$$

This model applies if $P_j(x) > 0$. In the case that $P_j(x) < 0$, $P(x)$ is set to zero and the converter is said to work in the discontinuous conduction mode (DCM). This last mode is taken into account in all numerical simulation. In the expressions of m_{on}, m_{off}^P and m_{off}^N the term $r_L i_L$, which appears in the inductor current state equation, was approximated by $r_L I_L$, where the parameter I_L is obtained by means of the averaged model and it is given by [17]:

$$I_L = \frac{V_{\text{in}}}{2r_L} + \sqrt{\frac{V_{\text{in}}^2}{4r_L^2} - \frac{1}{r_L}\left(\frac{V_P^2}{R_P} + \frac{V_N(V_N - V_{\text{in}})}{R_N}\right)}. \tag{3.12}$$

The duty cycles d_A and d_B are obtained by assuming that the inductor current is linear during each charging phase. Using this assumption gives the following expressions:

$$d_A = \frac{I_p^A - i_n^A}{(m_{\text{on}} - m_r)T},$$
$$d_B = \frac{I_p^B - i_n^B}{(m_{\text{on}} - m_r)T}, \tag{3.13}$$

where $i_n^B = i_n$ and $i_n^A = P_B(i_n)$. I_p^A and I_p^B are respectively the value of $I_r^A(t)$ and $I_r^B(t)$ at the beginning of phases A and B and

$$m_r = -\frac{V_u - V_l}{r_S T} \tag{3.14}$$

is the slope of the ramp signals v_r^A and v_r^B. To obtain I_p^A and I_p^B, we consider that the inductor current is PWL and that the duty cycles corresponding to both phases are given from the averaged model by:

$$D_A = \phi_A - \left|\frac{V_N}{R_N I_L}\right|,$$
$$D_B = \phi_B - \left|\frac{V_P}{R_P I_L}\right|. \tag{3.15}$$

The values of I_p^A and I_p^B, which do not depend on i_n are, therefore, obtained using the following area condition:

$$\int_o^T i_L(t)dt = I_L T. \tag{3.16}$$

(a)

(b)
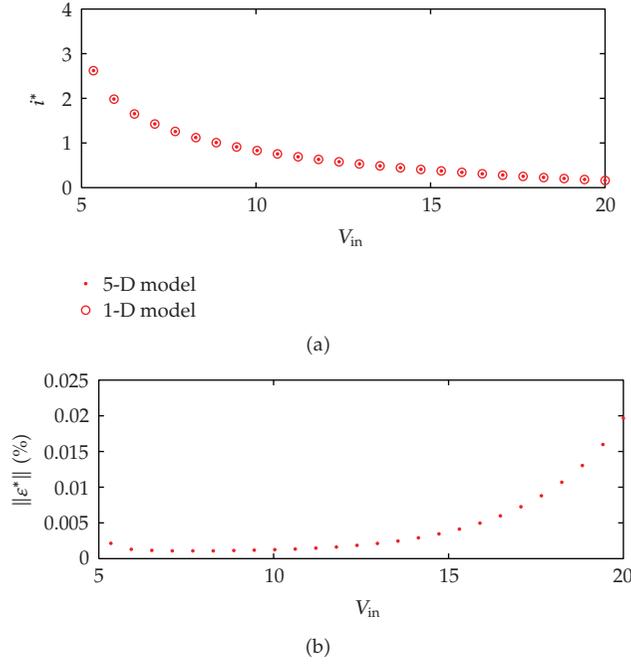
Figure 5: The fixed point i^* obtained from the 1D approximated model and the 5D exact model.

The fixed point i^* is obtained by forcing that $i_{n+1} = i_n = i^*$. This gives the following expression for i^*:

$$i^* = I_L + \left(\left(\left(\frac{(D_A^2 + D_B^2)}{2} - \phi_A D_A - 2\phi_B D_B \right) m_{\text{on}} \right. \right.$$

$$+ \left(2\phi_B D_B - \frac{3\phi_B^2}{2} + \frac{D_B^2}{2} \right) m_{\text{off}}^A \tag{3.17}$$

$$\left. \left. + \left(\phi_A D_A - \frac{D_A^2}{2} - \frac{\phi_B^2}{2} \right) m_{\text{off}}^A \right) T. \right.$$

This expression is then used to calculate I_p^A and I_p^B. The one-dimensional map has a single nonsaturated fixed point i^* which corresponds to the one periodic orbit of the switching system. Figure 5 gives the evolution of i^* when the input voltage varies. The result is obtained both numerically by using the exact 5D discrete time model and analytically by using (3.17). As it can be observed, the results match very well and the maximum error $\|\epsilon\|_{\text{max}}$ is less than 0.6%.

Figure 6 shows the cobweb plot of the map P for different values of V_P. Note the richness of behaviors that the map can exhibit. Namely, in Figure 6(a) we can see the coexistence of two different attractors of the system. The system will evolves toward an attractor or toward another depending on the starting point (initial condition). In Figure 6(b), dead beat phenomenon is observed. In this case, the system makes one and only one cycle
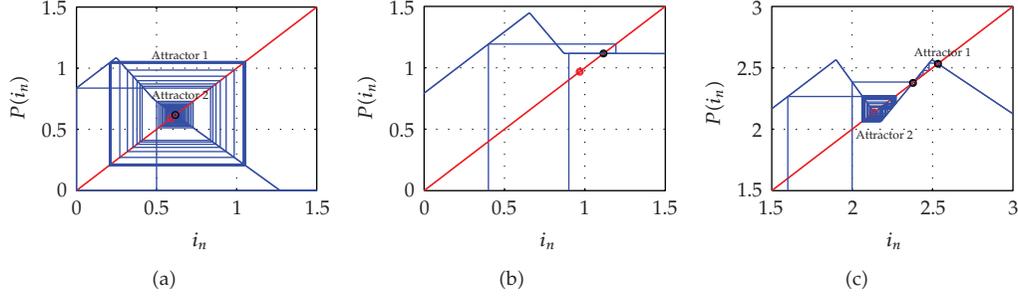
Figure 6: Cobweb plot of the map P for different values of V_P. (a) $V_P = 3\,\text{V}$, (b) $V_P = 8.27\,\text{V}$, (c) $V_P = 14\,\text{V}$. $V_u = 0.5\,\text{A}$.

to arrive to the fixed point once the trajectory hits the horizontal line defining the PWL map. Figure 6(c) presents the behavior of the system after a pitchfork bifurcation and its associated coexistence of attractors.

4. Stability Analysis from the PWL Map

4.1. The Stability Index

A sufficient condition for stability is that the absolute value of the derivative λ of the map P evaluated at the fixed point i^* is smaller than 1. It can be shown that this derivative is:

$$\lambda = \frac{(m_{\text{off}}^P - m_r)(m_{\text{off}}^N - m_r)}{(m_{\text{on}} - m_r)^2}, \tag{4.1}$$

and in terms of parameters of the system, λ becomes:

$$\lambda = \frac{(V_{\text{in}} - V_P - r_L I_L + (L(V_u - V_l)/T))(V_N - r_L I_L + (L(V_u - V_l)/r_S T))}{(V_{\text{in}} - r_L I_L + (L(V_u - V_l)/r_S T))^2}. \tag{4.2}$$

It can be observed that λ is a function of all parameters of the system. Let us define the following set of parameters:

$$\kappa = \frac{L(V_u - V_l)}{r_S T},$$

$$\alpha = r_L I_L - V_{\text{in}} + V_P,$$

$$\beta = r_L I_L + V_N,$$

$$\gamma = r_L I_L - V_{\text{in}}. \tag{4.3}$$
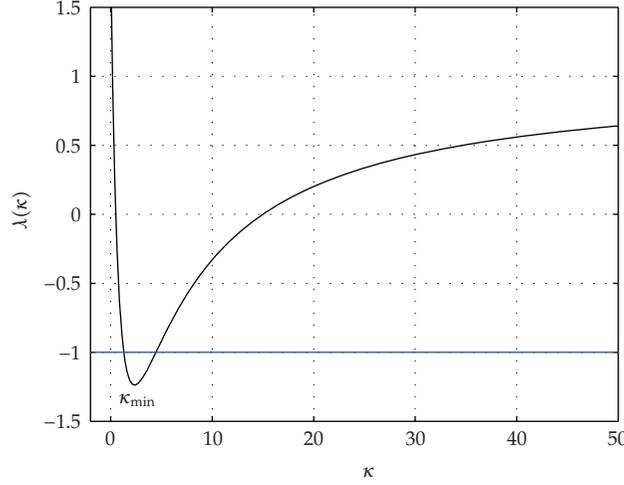
Figure 7: The stability index as function of κ. $\alpha = 0.5$, $\beta = 5$ and $\gamma = -2$.

Note that in the above expressions $\kappa > 0$ and $V_{\text{in}} > 0$. With these definitions the stability index λ becomes

$$\lambda(\kappa) = \frac{(\kappa - \alpha)(\kappa - \beta)}{(\kappa - \gamma)^2}. \tag{4.4}$$

This index should be between -1 and 1 in order to get a stable behavior. If this is not fulfilled, the system may present many nonlinear phenomena in the form of bifurcations, subharmonic oscillations and chaotic behavior. Considering the term $r_L I_L \ll V_{\text{in}}$ which is the case in practical circuits, and taking into account that under this assumption, the parameters fulfill conditions: $\beta > 0$, $\gamma < 0$, and $\alpha > \gamma$, the function $\lambda(\kappa)$ has the following properties (see Figure 7).

(i) The asymptotic value of λ is $\lim_{\kappa \to \infty} \lambda(\kappa) = 1$.

(ii) Although the index $\lambda(\kappa)$ has a singular point at $\kappa = \gamma$, this is not reachable provided that κ must be positive.

(iii) The index $\lambda(\kappa)$ has a minimum value at $\kappa_{\min}$ $((d\lambda/d\kappa)(\kappa_{\min}) = 0)$

$$\kappa_{\min} = \frac{-\beta\gamma + \alpha\gamma - 2\alpha\beta}{-\alpha - \beta + 2\gamma}. \tag{4.5}$$

4.2. Nonsmooth Flip Bifurcation ($\lambda(\kappa) = -1$) Curves

Flip bifurcation can appear for a nonlinear system if for a certain set of parameter values one have $\lambda(\kappa) = -1$. For our system, this will be possible if

$$\kappa_{\min} = \frac{-\beta\gamma + \alpha\gamma - 2\alpha\beta}{-\alpha - \beta + 2\gamma} < -1. \tag{4.6}$$

Then, the critical values of κ where the bifurcation flip can appear are given by the expression:

$$\kappa_{\text{flip}} = \frac{1}{4}\left(\alpha + \beta + 2\gamma \pm \sqrt{(\alpha - \beta)^2 - 4(\alpha - \gamma)(\beta - \gamma)}\right). \tag{4.7}$$

Notice that this bifurcation will be given only if κ_{flip} is a real positive number, that is,

$$(\alpha - \beta)^2 - 4(\alpha - \gamma)(\beta - \gamma) > 0 \quad \text{or} \quad \alpha + \beta + 2\gamma > \sqrt{(\alpha - \beta)^2 - 4(\alpha - \gamma)(\beta - \gamma)}. \tag{4.8}$$

4.3. Nonsmooth Pitchfork ($\lambda(\kappa) = 1$) Bifurcation Curves

If for a certain choice of parameter values one have $\lambda(\kappa) = 1$, then a nonsmooth pitchfork bifurcation occurs. The critical value of κ for which this bifurcation takes place is given by the expression:

$$\kappa_p = \frac{\alpha\beta - \gamma^2}{\alpha + \beta - 2\gamma}. \tag{4.9}$$

This bifurcation will actually appear if κ is positive, that is,

$$\alpha\beta > \gamma^2 \tag{4.10}$$

which will be never reached if $\alpha < 0$. Therefore, this bifurcation cannot appear if $V_{\text{in}} > V_P$ (see Figures 6(a) and 6(b)) and it is possible if $V_{\text{in}} < V_P$ (see Figure 6(c)).

4.4. Dead-Beat Response ($\lambda(\kappa) = 0$) Curves

A dead beat response to a small disturbance in the state of the system can be achieved by setting $\lambda(\kappa_{\text{db}}) = 0$ (Figure 6(b)). This implies the following relation between the parameters:

$$\kappa_{\text{db}} = \alpha \quad \text{or} \quad \kappa_{\text{db}} = \beta, \tag{4.11}$$

which corresponds to the following choice of circuit parameters:

$$\frac{L(V_u - V_l)}{r_S T} = V_P - V_{\text{in}} + r_L I_L \quad \text{or} \quad \frac{L(V_u - V_l)}{r_S T} = r_L I_L - V_N. \tag{4.12}$$

As it can be observed, this condition depends on parasitic parameters and therefore it will be difficult to meet in practical circuit. However it will be always the optimal choice of circuit parameters to get the fastest response. Figure 8 shows the different possible bifurcation curves in the plane (α, κ). The dead-beat response curves ($\lambda = 0$) is also shown in the same figure.
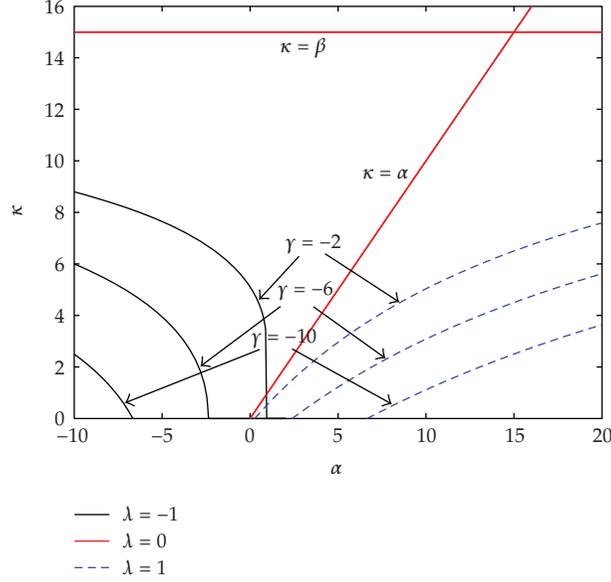
Figure 8: The bifurcation curves in the plane (α, κ) for different values of γ. $\beta = 15$.

Table 1: Parameter values used in numerical simulations.

Parameter	Value	Parameter	Value
V_{in}	Varying	V_l	0
L	$640\,\mu\text{H}$	V_u	Varying
r_L	$0.7\,\Omega$	V_P	Varying
C_N	$45\,\mu\text{F}$	V_N	$-15\,\text{V}$
C_P	$45\,\mu\text{F}$	τ_A	$200\,\mu\text{s}$
R_N	$68\,\Omega$	τ_B	$200\,\mu\text{s}$
R_P	$33\,\Omega$	g_P	$0.02\,\Omega^{-1}$
$f_s = 1/T$	$10\,\text{kHz}$	g_N	$0.02\,\Omega^{-1}$
r_S	$1\,\Omega$	$\phi_A = \phi_B$	$1/2$

5. Bifurcation Behavior from Numerical Simulation

After undergoing a bifurcation, the period τ of a possible periodic orbit of the nonlinear system is an integer multiple of the period of the forcing ramp signal, that is, $\tau = nT$, where $n = 1, 2, \ldots$. Such an operation of the converter will be referred to as an n-periodic orbit.

5.1. Bifurcations from the PWL Map

Let us consider the circuit of Figure 1 with the control scheme of Figure 2 and the values of parameters shown in Table 1. The bifurcation parameter used here is the upper value of the ramp voltage V_u. From mathematical analysis of the PWL map, the boundary of stability in terms of the parameters can be obtained. Two critical values, corresponding to flip curves, are
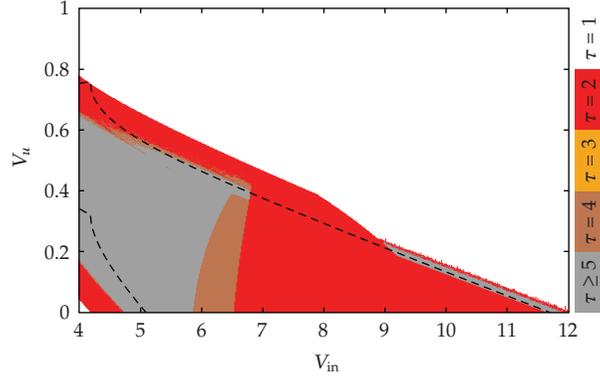
Figure 9: Two-dimensional bifurcation diagram taking V_u and V_{in} as bifurcation parameter. Initial conditions are selected in the vicinity of the DC averaged state I_L. $V_P = 3$ V.

derived for the amplitude of the ramp voltage V_u in terms of other bifurcation parameters. These critical values are given by:

$$V_{u,cri,1} = r_L I_L + \frac{r_S T \left(V_P - 3V_{in} - V_N - \sqrt{(V_P)^2 + 6V_P(V_N - V_{in}) + (V_N - V_{in})^2} \right)}{4L},$$

$$V_{u,cri,2} = r_L I_L + \frac{r_S T \left(V_P - 3V_{in} - V_N + \sqrt{(V_P)^2 + 6V_P(V_N - V_{in}) + (V_N - V_{in})^2} \right)}{4L}.$$

(5.1)

Note that theoretically there exist two different critical values of V_u for which stability is lost. However some times one of the critical values is virtual because it is complex or negative. One way to avoid instability and subharmonic oscillations is by making both critical values virtual. This can be achieved if the following condition is fulfilled:

$$(V_P)^2 + 6V_P(V_N - V_{in}) + (V_N - V_{in})^2 < 0. \tag{5.2}$$

Figure 9 shows the two dimensional bifurcation diagram obtained from numerical simulations by using the PWL discrete time model derived in the previous section. The analytical expressions giving conditions for stability of 1-periodic orbits are also plotted in the same figure (black dashed lines). As it can be observed, theoretical analysis and numerical simulations are in good agreement on the stability boundary of the 1-periodic orbits. It can be shown by mathematical analysis that the 1-periodic and the 2-periodic orbits have the same stability boundary although their domains of existence are different. This means that these two orbits will coexist in a certain zone of the parameter space and depending on the initial conditions the final state could be a 1-periodic or a 2-periodic orbit. In order to put more clear this coexisting attractors phenomenon, the one-dimensional bifurcation diagrams and their corresponding Lyapunov exponents are plotted in Figure 10 for different values of the input voltage and choosing differently the initial conditions. The hysteresis phenomenon, due to the coexistence of attractors within a certain interval of the bifurcation parameter, is reflected
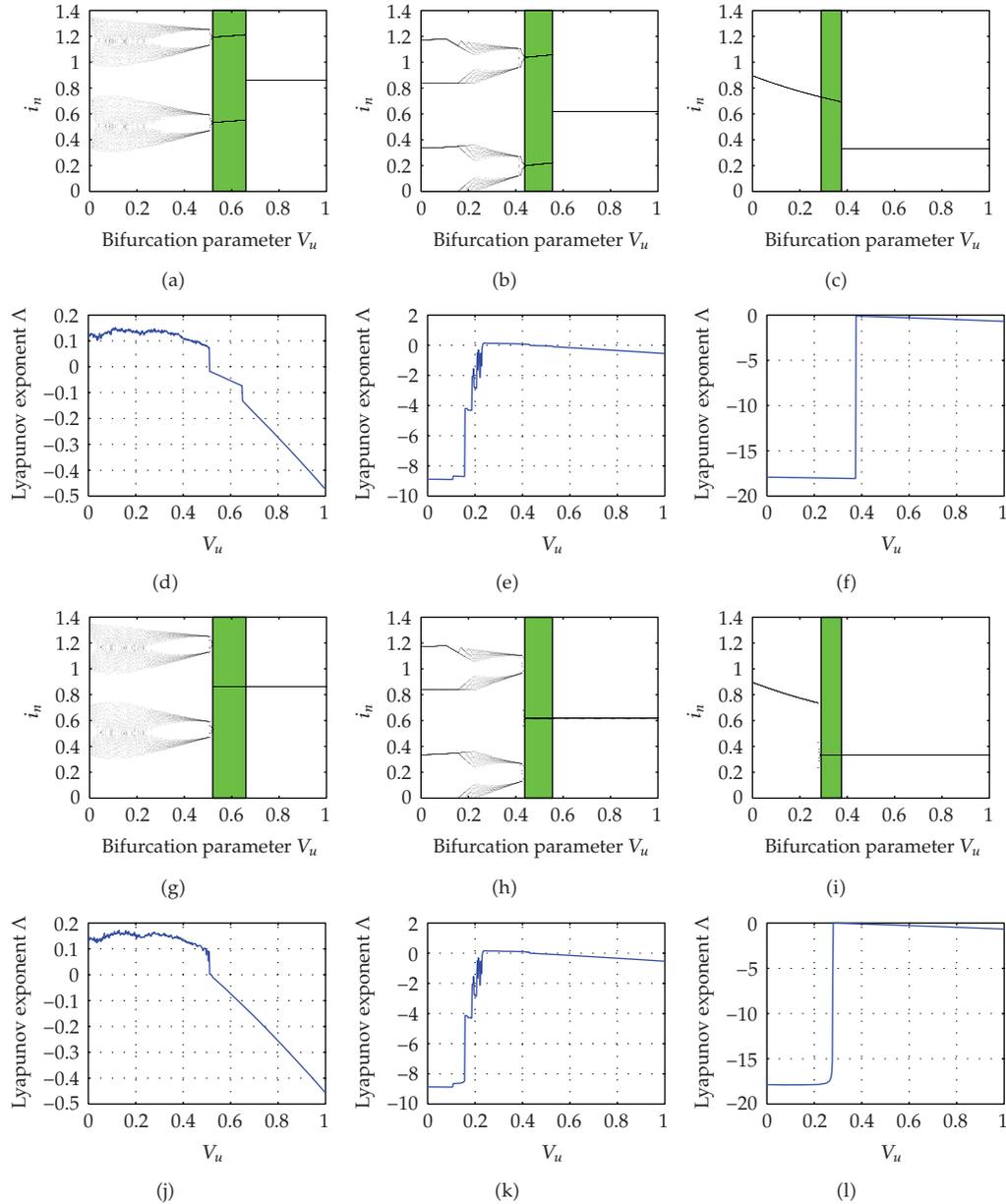
Figure 10: One dimensional bifurcation diagram and Lyapunov exponent taking V_u as bifurcation parameter for different values of V_{in}. (a)–(f) initial conditions are selected in the vicinity of I_L. (g)–(l) initial conditions are selected in the vicinity of i^*. From left to right $V_{in} = 5, 6, 8$ V. Shaded area shows hysteresis interval whose width decreases when V_{in} increases.

as a shaded area in this figure. The width of this interval decreases when V_{in} increases. The basin of attraction of an attractor, which is defined as the set of points in the state space such that initial conditions chosen in this set dynamically evolve to this attractor, can be obtained by sweeping the initial i_0 condition in a certain interval of the bifurcation parameter. The evolution of the basins of attraction of the different possible attractors in terms of V_u, are
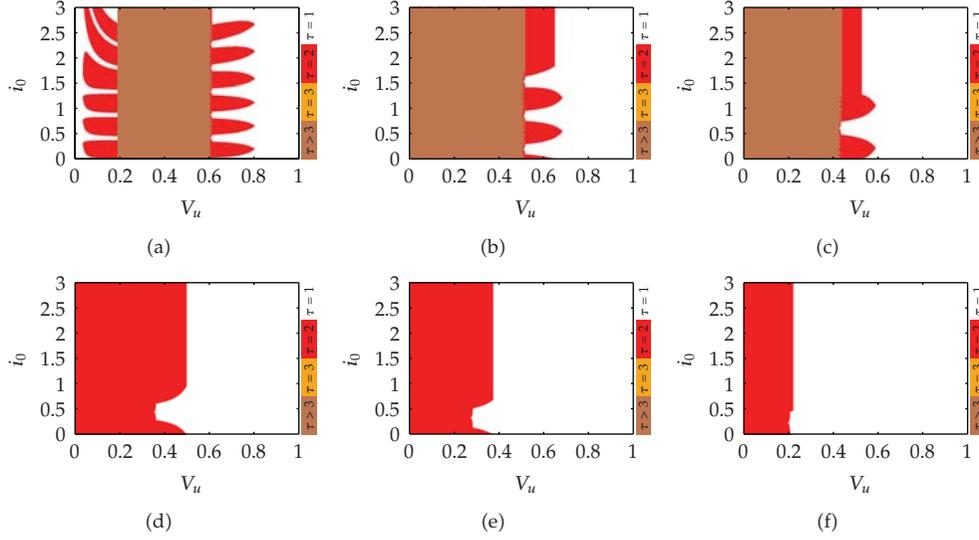
Figure 11: Evolution of the basins of attractions in terms of V_u for different values of V_{in} (a) $V_{in} = 4$ V, (b) $V_{in} = 5$ V (c) $V_{in} = 6$ V, (d) $V_{in} = 7$ V, (e) $V_{in} = 8$ V and (f) $V_{in} = 9$ V.

shown in Figure 11 for different values of V_{in}. In Figures 9 and 11, the period of the orbit was color coded. Only five colors were used to determine the different periodic modes that the system can present. Using more colors does not change the results because the system does not present stable periodic orbit with period bigger than four in this case. White color represent the stable zone (1-periodic). Zones of higher periods and chaotic behavior are also shown as parameters V_u and V_{in} vary. Particularly, we observe from Figure 9 for example that when $V_{in} = 6$ V and as V_u decreases the inductor current is 1-periodic then at a critical value of V_u becomes 2-periodic and suddenly becomes chaotic after a border collision bifurcation at another critical value of V_u. As V_u decreases further, the inductor current becomes 4-periodic. This is in a good concordance with the one-dimensional bifurcation diagram and the evolution of the Lyapunov exponent in Figure 10. The Lyapunov exponent is a measure of average amount of contraction or expansion of a trajectory near a periodic orbit and it is defined as follows:

$$\Lambda = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \ln \left| \frac{dP(i_n)}{di_n} \right|. \tag{5.3}$$

Due to the nonsmooth nature of the system, the map P is PWL and its derivative $dP(i_n)/di_n$ is piecewise constant. Therefore the Lyapunov exponent can undergo jumps at some bifurcation points.

5.2. Bifurcations from the Exact Switched Model

The above results were derived from the PWL map of the SITITO converter where v_P an v_N and the integral terms s_P and s_N are assumed constant. Therefore, it is necessary to check how these results are close to the actual system behavior when this assumption is not made

(a)

(b)
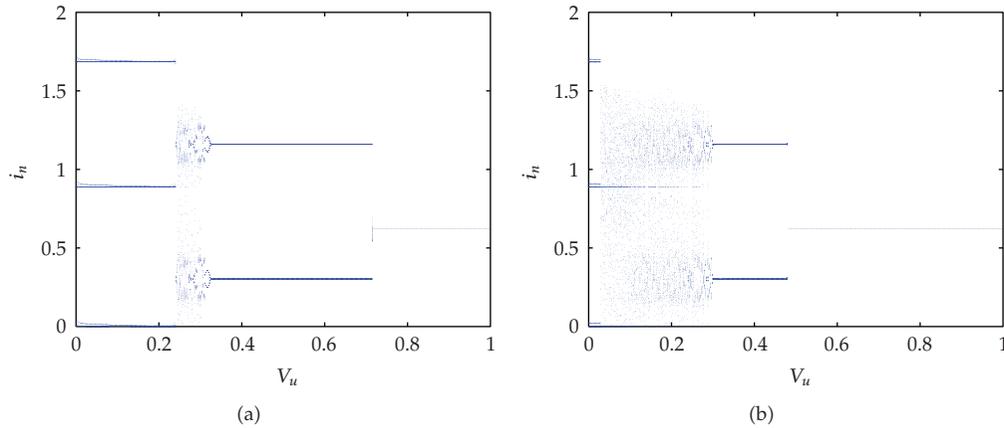
Figure 12: One dimensional bifurcation diagram from the 5D switched model taking V_u as bifurcation parameter for $V_{\text{in}} = 6$ V.

(a)

(b)
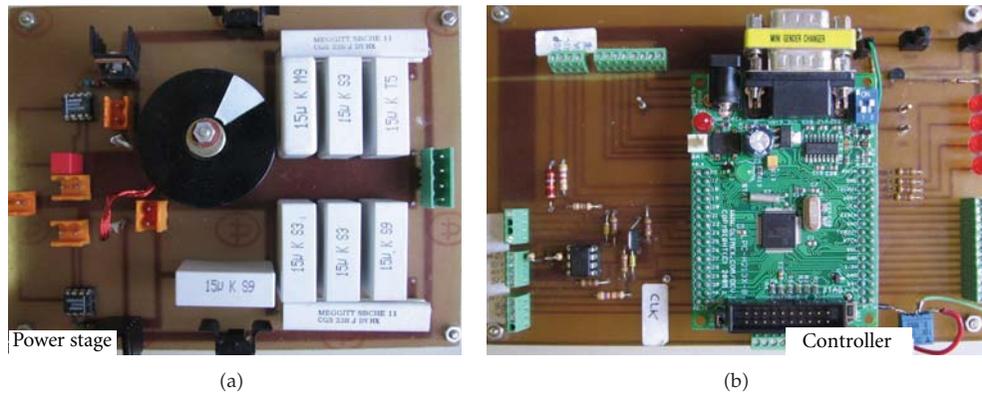
Figure 13: Experimental prototype of a SITITO DC-DC converter. (a) Power stage circuit. (b) PWM Interleaved controller.

and the output voltage ripples are taken into account. Figure 12 shows two representative bifurcation diagrams of the system from simulation of the circuit diagram using the switched model, in which the parameter V_u was varied as before. The parameter V_{in} is fixed to 6 V and all other parameter values and details are the same as for the above subsection. In one case the bifurcation parameter is increased while in another case it is decreased. In this way, initial conditions are chosen differently. Comparing Figures 12(a) and 12(b) from one side with Figures 10(b) and 10(h) from the other side, the first aspect in which the actual system dynamical behavior is different from that derived from the PWL map is that in the zone of small values of V_u which is not used in practical circuits. Another difference is the parameter value for which the first period doubling occur is a little bit different. Thus there is a range of parameters where the 2-periodic is stable for the exact model while the PWL map predict 1-periodic stable. Apart from this aspect, the prediction of the simplified model regarding transition from periodicity to chaos and coexisting attractors is enough accurate in almost all the parameter range from $V_u \approx 0.3$ V to $V_u = 1$ V.

Measure time: 14:36:35   TB A: 100 μs
Measure date: 20/02/2009   TR: EXT+AC
CH1: 200V/DIV DC   PT: 50

(a)

Measure time: 13:02:26   TB A: 100 μs
Measure date: 20/02/2009   TR: EXT+AC
CH1: 200V/DIV DC   PT: 50

(b)

(c)

(d)
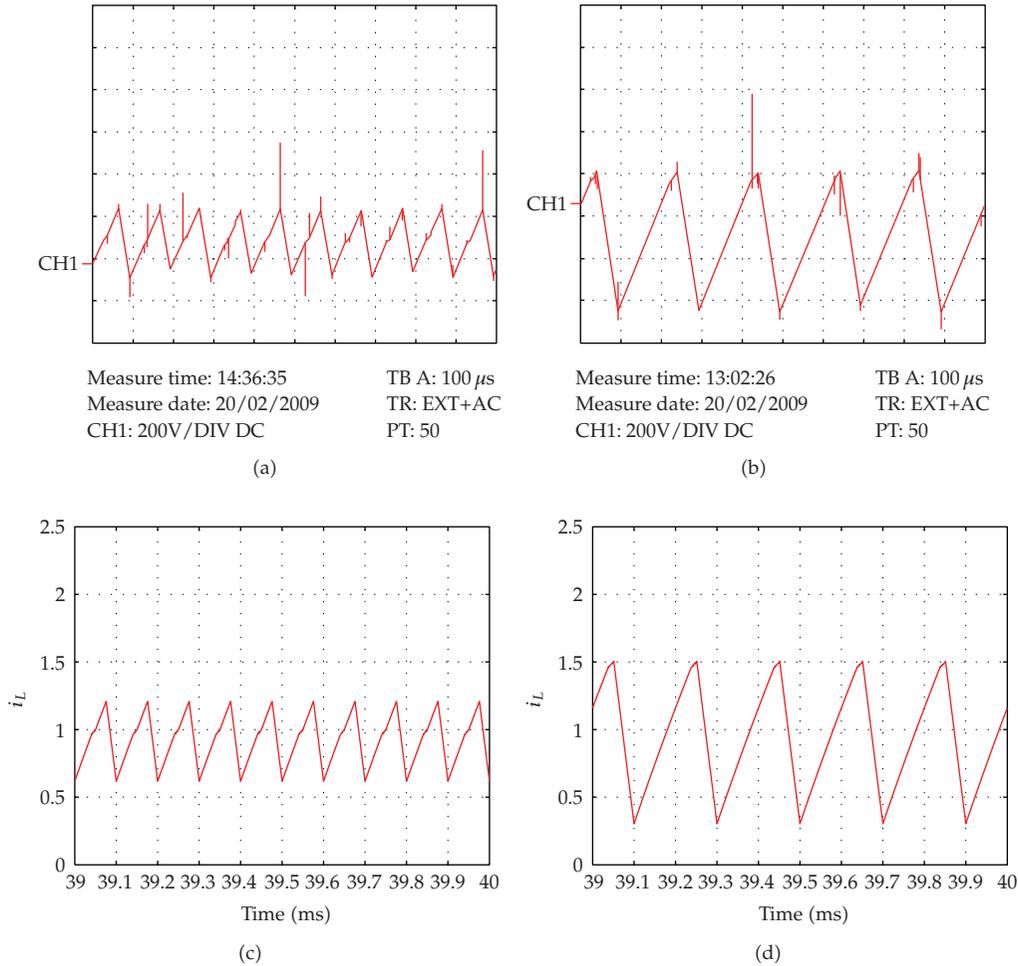
Figure 14: Experimental waveforms of the inductor current and their corresponding numerical simulations from the switched model showing normal periodic behavior (left $V_u = 1$ V), subharmonic oscillations (2-periodic) (right $V_u = 0.6$ V) after losing the stability of the 1-periodic orbit by a nonsmooth flip bifurcation. Other parameters are from Table 1.

6. Experimental Results

A prototype of a SITITO DC-DC converter (Figure 13) has been built in order to confirm our analytical results and numerical simulations. The nominal values of parameters are the same than previous sections. The inductor current is sensed using the LEM current probe PR30. The control signals are obtained and processed by a microcontroller. This device provides the command signals which are buffered by two dedicated MOS drivers from MAXIM (MAX626), then they are applied to the MOSFET IRF9Z34S (p-channel) and IRL530N (n-channel). Diodes in Figure 1 are Schottky barrier type (6CWQ04FN). The inductor current i_L is represented with a proportional voltage with a factor of 0.5 V/A. The phases ϕ_A and ϕ_B are selected slightly different from $(1/2)$ (0.55 and 0.45) to avoid some noise effects at the switching instants. From the point of view of dynamics, the ranges for some parameters have been chosen accordingly with our technical resources, but the results can

be rescaled, for instance, under higher switching frequencies and lower inductance, unless parasitic components effects become relevant. Figure 14 shows the experimental waveforms of the inductor current for the normal periodic behavior and subharmonic oscillations. Numerical simulations from the switched model are also shown for comparison. As it can be observed, except some noise at the switching instants, there is a good agreement between the experimental and the numerical results.

7. Conclusions

A single inductor two inputs two outputs (SITITO) DC-DC converter under current mode control with an interleaved PWM control strategy is considered in this paper. Its dynamics is described by a nonlinear modeling approach. The ripple, mainly the associated to the inductor current, and saturation effects can cause instabilities in the form of subharmonic oscillations and chaotic behavior. An expression for a simple PWL map is derived to obtain accurate information about the dynamics of the system. This map can predict many nonlinear behaviors such as bifurcations, chaos, coexisting attractors and the associated hysteresis phenomenon. Numerical simulations and experimental measurements from a laboratory prototype confirm the theoretical predictions.

Acknowledgments

References

[1] R. W. Erickson and D. Maksimovic, *Fundamentals of Power Electronics*, Springer, New York, NY, USA, 2001.

[2] I. Dénes and I. Nagy, "Two models for the dynamic behaviour of a dual-channel buck and boost DC-DC converter," *Electromotion*, vol. 10, no. 4, pp. 556–561, 2003.

[3] H.-P. Le, C.-S. Chae, K.-C. Lee, G.-H. Cho, S.-W. Wang, and G.-H. Cho, "A single-inductor switching DC-DC converter with 5 outputs and ordered power-distributive control," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 12, pp. 2706–2714, 2007.

[4] J. Hamar and I. Nagy, "Asymmetrical operation of dual channel resonant DC-DC converters," *IEEE Transactions on Power Electronics*, vol. 18, no. 1, part 1, pp. 83–94, 2003.

[5] R. Barabás, B. Buti, J. Hamar, and I. Nagy, "Control characteristics, simulation and test results of a dual channel DC-DC converter family," in *Proceedings of Power Electronics Electrical Drives Automation and Motion Conference (SPEEDAM '04)*, Capri, Italy, June 2004.

[6] Y. Xi and P. K. Jain, "A forward converter topology with independent and precisely regulated multiple outputs," *IEEE Transactions on Power Electronics*, vol. 18, no. 2, pp. 648–658, 2003.

[7] S. Banerjee and G. C. Verghese, *Nonlinear Phenomena in Power Electronics: Bifurcations, Chaos, Control, and Applications*, Wiley-IEEE Press, New York, NY, USA, 2001.

[8] D. C. Hamill and D. J. Jeffries, "Subharmonics and chaos in a controlled switched-mode power converter," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 8, pp. 1059–1061, 1988.

[9] J. H. B. Deane and D. C. Hamill, "Instability, subharmonics, and chaos in power electronic systems," *IEEE Transactions on Power Electronics*, vol. 5, no. 3, pp. 260–268, 1990.

[10] E. Fossas and G. Olivar, "Study of chaos in the buck converter," *IEEE Transactions on Circuits and Systems I*, vol. 43, no. 1, pp. 13–25, 1996.

[11] S. Banerjee, M. S. Karthik, G. Yuan, and J. A. Yorke, "Bifurcations in one-dimensional piecewise smooth maps—theory and applications in switching circuits," *IEEE Transactions on Circuits and Systems I*, vol. 47, no. 3, pp. 389–394, 2000.

[12] Z. T. Zhusubaliyev and E. Mosekilde, *Bifurcations and Chaos in Piecewise-Smooth Dynamical Systems*, vol. 44 of *World Scientific Series on Nonlinear Science. Series A*, World Scientific, River Edge, NJ, USA, 2003.

[13] S. Banerjee and K. Chakrabarty, "Nonlinear modeling and bifurcations in the boost converter," *IEEE Transactions on Power Electronics*, vol. 13, no. 2, pp. 252–260, 1998.

[14] M. di Bernardo and C. K. Tse, "Chaos in power electronics: an overview," in *Chaos in Circuits and Systems*, G. Chen and T. Ueta, Eds., vol. 11, chapter 16, pp. 317–340, World Scientific, New York, NY, USA, 2002.

[15] C. K. Tse and M. Di Bernardo, "Complex behavior in switching power converters," *Proceedings of the IEEE*, vol. 90, no. 5, pp. 768–781, 2002.

[16] L. Benadero, A. El Aroudi, G. Olivar, E. Toribio, and E. Gómez, "Two-dimensional bifurcation diagrams. Background pattern of fundamental DC-DC converters with PWM control," *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 13, no. 2, pp. 427–451, 2003.

[17] L. Benadero, R. Giral, A. El Aroudi, and J. Calvente, "Stability analysis of a single inductor dual switching DC-DC converter," *Mathematics and Computers in Simulation*, vol. 71, no. 4–6, pp. 256–269, 2006.

[18] L. Benadero, V. Moreno-Font, A. El Aroudi, and R. Giral, "Single inductor multiple outputs interleaved converters operating in CCM," in *Proceedings of the 13th International Power Electronics and Motion Control Conference (EPE-PEMC '08)*, pp. 2115–2119, Poznan, Poland, September 2008.

[19] A. El Aroudi, M. Debbat, R. Giral, G. Olivar, L. Benadero, and E. Toribio, "Bifurcations in DC-DC switching converters: review of methods and applications," *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, vol. 15, no. 5, pp. 1549–1578, 2005.

[20] H. H. C. Iu and C. K. Tse, "Bifurcation behavior in parallel-connected buck converters," *IEEE Transactions on Circuits and Systems I*, vol. 48, no. 2, pp. 233–240, 2001.

[21] A. El Aroudi, M. Debbat, and L. Martínez-Salamero, "Poincaré maps modeling and local orbital stability analysis of discontinuous piecewise affine periodically driven systems," *Nonlinear Dynamics*, vol. 50, no. 3, pp. 431–445, 2007.

[22] B. Robert and A. El Aroudi, "Discrete time model of a multi-cell DC/DC converter: non linear approach," *Mathematics and Computers in Simulation*, vol. 71, no. 4–6, pp. 310–319, 2006.

[23] A. El Aroudi, B. G. M. Robert, A. Cid-Pastor, and L. Martínez-Salamero, "Modelling and design rules of a two-cell buck converter under a digital PWM controller," *IEEE Transactions on Power Electronics*, vol. 23, no. 2, pp. 859–870, 2008.

Research Article

Higher Period Stochastic Bifurcation of Nonlinear Airfoil Fluid-Structure Interaction

Jeroen A. S. Witteveen and Hester Bijl

*Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1,
2629HS Delft, The Netherlands*

Correspondence should be addressed to Jeroen A. S. Witteveen, j.a.s.witteveen@tudelft.nl

The higher period stochastic bifurcation of a nonlinear airfoil fluid-structure interaction system is analyzed using an efficient and robust uncertainty quantification method for unsteady problems. The computationally efficient numerical approach achieves a constant error with a constant number of samples in time. The robustness of the method is assured by the extrema diminishing concept in probability space. The numerical results demonstrate that the system is even more sensitive to randomness at the higher period bifurcation than in the first bifurcation point. In this isolated point in parameter space the clear hierarchy of increasing importance of the random nonlinearity parameter, initial condition, and natural frequency ratio, respectively, even suddenly reverses. Disregarding seemingly less important random parameters based on a preliminary analysis can, therefore, be an unreliable approach for reducing the number of relevant random input parameters.

1. Introduction

It is widely know that the behavior of nonlinear dynamical systems is highly sensitive to small variations. Examples of significant effects of varying initial conditions and model parameters in time-dependent problems can be found in many branches of science and engineering. In turbulence modeling and nonlinear stability theory of transition it is recognized that uncertainty in the initial conditions has a substantial effect on the long-term solution [1–3]. The inherent sensitivity of meteorological and atmospheric models for weather prediction results in a rapid loss of simulation accuracy over time [4, 5]. Stochastic parameters also affect the voltage oscillations in the electric circuit of a nonlinear transistor amplifier [6]. In this paper, the aeronautical application of the effect of randomness on the bifurcation of a nonlinear aeroelastic wing structure is analyzed. Physical uncertainties are

encountered in this kind of fluid-structure systems due to varying atmospheric conditions, wear and tear, and production tolerances affecting material properties and the geometry. Compared to the deterministic case the stochastic bifurcation can lead to an earlier onset of unstable flutter behavior, which can cause fatigue damage and structural failure.

Fluid-structure interaction systems can be modeled deterministically using detailed finite-element method (FEM) structural discretizations and high-fidelity unsteady computational fluid dynamics (CFD) simulations. This computationally highly intensive approach is usually too expensive for performing many deterministic simulations required in a flutter analysis study. In flutter analysis the structure is, therefore, usually modeled by rigid airfoil mass-spring systems for wing structures and by plate equations for plate-like designs [7, 8]. Structural nonlinearity is then modeled by a cubic nonlinear spring, since structures commenly behave as a cubic stiffness hardening spring [9]. In this framework the flow forces are taken into account in the governing structural equations by source terms prescribed by aerodynamic models. These simplifications are even more frequently used in the stochastic analysis of aeroelastic systems, since each additional random input parameter contributes to the dimensionality of the parameter domain under consideration.

The stochastic bifurcation behavior of aeroelastic systems has previously been studied using perturbation techniques, Monte Carlo simulation, Polynomial Chaos formulations, and a range of other numerical and analytical methods. The perturbation approach [10] has been used by Poirion to obtain a first-order approximation of the flutter probability of a bending-torsion structural model, see [11, 12]. A second moment perturbation-based stochastic finite-element method has been applied by Liaw and Yang [13] to determine the effect of uncertainties on panel flutter. The vibration of a hydrofoil in random flow has been considered using a stochastic pertubation approach by Carcaterra et al. [14]. Monte Carlo simulations [15] have, for example, been used by Lindsley et al. [16, 17] to study the periodic response of nonlinear plates under supersonic flow subject to randomness. Poirel and Price [18] have studied random bending-torsion flutter equations with turbulent flow conditions and a linear structural model using also a Monte Carlo-type approach. The stochastic postflutter behavior of limit cycle oscillations has been studied by Beran et al. [19] using Monte Carlo sampling.

Other uncertainty quantification methodologies have, for example, been employed in an investigation of nonlinear random oscillations of aeroservoelastic systems by Poirion [20] using random delay modeling of control systems. Choi and Namachchivaya [21] have used nonstandard reduction through stochastic averaging in nonlinear panel flutter under supersonic flow subject to random fluctuations in the turbulent boundary layer to find response density functions. De Rosa and Franco [22] have predicted the stochastic response of a plate subject to a turbulent boundary layer using numerical and analytical approaches. Frequency domain methods have been considered for solving linear stochastic operator equations by Sarkar and Ghanem [23].

Polynomial Chaos methods [24–28] are, in general, computationally efficient alternatives for the detailed and quantitative probabilistic modeling of physical uncertainties by Monte Carlo simulation. However, in dynamic simulations the Polynomial Chaos method usually requires a fast increasing expansion order to maintain a constant accuracy in time. This leads to a fast increasing sample size in the more practical nonintrusive Polynomial Chaos formulations [29–33], which are based on the polynomial interpolation of an in general small number of samples. Resolving the asymptotic stochastic effect in a postflutter analysis using nonintrusive Polynomial Chaos can, however, lead to a very high number of required deterministic simulations. This effect is especially profound in problems with an oscillatory

solution in which the frequency of the response is affected by the random parameters [3, 34]. Pettit and Beran [35] have demonstrated that the Polynomial Chaos expansion is subject to energy loss in representing the periodic response of a bending-torsion flutter model for long integration times. They have also found that the wavelet based Wiener-Haar expansion of Le Maître et al. [36] loses its accuracy less rapidly. Also other multielement Polynomial Chaos formulations have been shown to postpone the resolution problems [37]. Millman et al. [38] have proposed a Fourier-Chaos expansion for oscillatory responses in application to a bending-torsion flutter problem subjected to Gaussian distributions. The effectivity of intrusive and nonintrusive Polynomial Chaos methods has been compared for a single-degree-of-freedom pitching airfoil stall flutter system in [39, 40].

Another special Polynomial Chaos formulation for oscillatory problems was recently also developed to maintain a constant accuracy in time with a constant polynomial order [41, 42]. The nonintrusive approach is based on normalizing the oscillatory samples in terms of their phase. The uncertainty quantification interpolation of the samples is then performed at constant phase, which eliminates the effect of frequency differences on the increase of the required sample size [43]. The method is proven to result in a bounded error as function of the phase with a constant number of samples for periodic responses and under certain conditions also in a bounded error in time [44]. The formulation was also extended to multifrequency responses of continuous structures by using a wavelet decomposition preprocessing step [45]. Application of the method to an elastically mounted airfoil showed that this fluid-structure interaction system is sensitive to small variations at the bifurcation from a stable solution to a period-1 limit cycle oscillation [43]. A period-1 motion refers to an oscillation that repeats itself after a 2π-orbit around a fixed point in phase space.

In this paper the latter uncertainty quantification approach is employed to analyze the stochastic higher period bifurcation of an aeroelastic airfoil with nonlinear structural stiffness. It is demonstrated that the fluid-structure interaction system is even more sensitive to randomness at the higher period bifurcation than in the previously considered first bifurcation point. The resulting general mathematical formulation of the uncertainty quantification problem is given in Section 2. The efficient and robust uncertainty quantification method for unsteady problems based on extrema diminishing interpolation of oscillatory samples at constant phase used to resolve the stochastic bifurcation behavior numerically is introduced in Section 3. As is common in stochastic flutter analysis, the aeroelastic system is modeled by a two-dimensional rigid airfoil with two degrees of freedom in pitch and plunge, and cubic nonlinear spring stiffness. The aerodynamic loads are computed using an aerodynamic model as described in Section 4. Randomness is introduced in terms of three random parameters in the system and its initial conditions. The effect of uncertainty in the ratio of natural pitch and plunge frequencies is resolved in Section 5. A random nonlinear spring parameter is considered in Section 6. The effect of randomness in the initial condition of the pitch angle is investigated in Section 7. The main findings are summarized in Section 8.

2. Mathematical Formulation of the Uncertainty Quantification Problem

Consider a dynamical system subject to $n_{\mathbf{a}}$ uncorrelated second-order random input parameters $\mathbf{a}(\omega) = \{a_1(\omega), \ldots, a_{n_{\mathbf{a}}}(\omega)\} \in A$ with parameter space $A \in \mathbb{R}^{n_{\mathbf{a}}}$, which governs an oscillatory response $u(\mathbf{x}, t, \mathbf{a})$

$$\mathcal{L}(\mathbf{x}, t, \mathbf{a}; u(\mathbf{x}, t, \mathbf{a})) = S(\mathbf{x}, t, \mathbf{a}), \tag{2.1}$$

with operator $\mathcal{L}$ and source term S defined on domain $D \times T \times A$, and appropriate initial and boundary conditions. The spatial and temporal dimensions are defined as $\mathbf{x} \in D$ and $t \in T$, respectively, with $D \subset \mathbb{R}^d$, $d = \{1, 2, 3\}$, and $T = [0, t_{\max}]$. A realization of the set of outcomes Ω of the probability space $(\Omega, \mathcal{F}, P)$ is denoted by $\omega \in \Omega = [0, 1]^{n_a}$, with $\mathcal{F} \subset 2^\Omega$ the σ-algebra of events and P a probability measure.

Here we consider a nonintrusive uncertainty quantification method l which constructs a weighted approximation $w(\mathbf{x}, t, \mathbf{a})$ of response surface $u(\mathbf{x}, t, \mathbf{a})$ based on n_s deterministic solutions $v_k(\mathbf{x}, t) \equiv u(\mathbf{x}, t, \mathbf{a}_k)$ of (2.1) for different parameter values $\mathbf{a}_k \equiv \mathbf{a}(\omega_k)$ for $k = 1, \ldots, n_s$. The samples $v_k(\mathbf{x}, t)$ can be obtained by solving the deterministic problem

$$\mathcal{L}(\mathbf{x}, t, \mathbf{a}_k; v_k(\mathbf{x}, t)) = \mathcal{S}(\mathbf{x}, t, \mathbf{a}_k), \tag{2.2}$$

for $k = 1, \ldots, n_s$, using standard spatial discretization methods and time marching schemes. A nonintrusive uncertainty quantification method l is then a combination of a sampling method g and an interpolation method h. Sampling method g defines the n_s sampling points $\{\mathbf{a}_k\}_{k=1}^{n_s}$ and returns the deterministic samples $\mathbf{v}(\mathbf{x}, t) = \{v_1(\mathbf{x}, t), \ldots, v_{n_s}(\mathbf{x}, t)\}$. Interpolation method h constructs an interpolation surface $w(\mathbf{x}, t, \mathbf{a})$ through the n_s samples $\mathbf{v}(\mathbf{x}, t)$ as an approximation of $u(\mathbf{x}, t, \mathbf{a})$. We are eventually interested in an approximation of the probability distribution and statistical moments $\mu_{u_i}(\mathbf{x}, t)$ of the output $u(\mathbf{x}, t, \mathbf{a})$, which can be obtained by sorting and weighted integration of $w(\mathbf{x}, t, \mathbf{a})$:

$$\mu_{u_i}(\mathbf{x}, t) \approx \mu_{w_i}(\mathbf{x}, t) = \int_A w(\mathbf{x}, t, \mathbf{a})^i f_{\mathbf{a}}(\mathbf{a}) \mathrm{d}\mathbf{a}. \tag{2.3}$$

This information can be used for reducing design safety factors and robust design optimization, in contrast to reliability analysis in which the probability of failure is determined [46].

3. An Efficient Uncertainty Quantification Method for Unsteady Problems

The efficient uncertainty quantification formulation for oscillatory responses based on interpolation of scaled samples at constant phase is developed in Section 3.2. The robust extrema diminishing uncertainty quantification method based on Newton-Cotes quadrature in simplex elements employed in the unsteady approach is first presented in the next section.

3.1. Robust Extrema Diminishing Uncertainty Quantification

A multielement uncertainty quantification method l evaluates integral (2.3) by dividing parameter space A into n_e non-overlapping simplex elements $A_j \subset A$:

$$\mu_{w_i}(\mathbf{x}, t) = \sum_{j=1}^{n_e} \int_{A_j} w(\mathbf{x}, t, \mathbf{a})^i f_{\mathbf{a}}(\mathbf{a}) \mathrm{d}\mathbf{a}. \tag{3.1}$$

Here we consider a multielement Polynomial Chaos method based on Newton-Cotes quadrature points and simplex elements [47]. A piecewise polynomial approximation
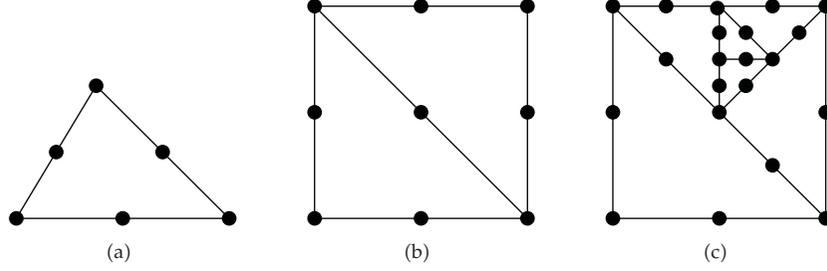
Figure 1: Discretization of two-dimensional parameter space A using 2-simplex elements and second-degree Newton-Cotes quadrature points given by the dots.

$w(\mathbf{x}, t, \mathbf{a})$ is then constructed based on n_s deterministic solutions $v_{j,k}(\mathbf{x}, t) = u(\mathbf{x}, t, \mathbf{a}_{j,k})$ for the values of the random parameters $\mathbf{a}_{j,k}$ that correspond to the $\tilde{n}_s$ Newton-Cotes quadrature points of degree d in the elements A_j:

$$\mu_{w_i}(\mathbf{x}, t) = \sum_{j=1}^{n_e} \sum_{k=1}^{\tilde{n}_s} c_{j,k} v_{j,k}(\mathbf{x}, t)^i, \tag{3.2}$$

where $c_{j,k}$ is the weighted integral of the Lagrange interpolation polynomial $L_{j,k}(\mathbf{a})$ through Newton-Cotes quadrature point k in element A_j:

$$c_{j,k} = \int_{A_j} L_{j,k}(\mathbf{a}) f_{\mathbf{a}}(\mathbf{a}) d\mathbf{a}, \tag{3.3}$$

for $j = 1, \ldots, n_e$ and $k = 1, \ldots, \tilde{n}_s$. Here, second-degree Newton-Cotes quadrature with $d = 2$ is considered in combination with adaptive mesh refinement in probability space, since low-order approximations are more effective for approximating response response surfaces with singularities. The initial discretization of parameter space A for the adaptive scheme consists of the minimum of $n_{e_{\text{ini}}} = n_a!$ simplex elements and $n_{s_{\text{ini}}} = 3^{n_a}$ samples, see Figure 1. The example of Figure 1 for two random input parameters can geometrically be extended to higher dimensional probability spaces. The elements A_j are adaptively refined using a refinement measure ρ_j based on the largest absolute eigenvalue of the Hessian H_j, as a measure of the curvature of the response surface approximation in the elements, weighted by the probability f_j contained by the elements

$$f_j = \int_{A_j} f_{\mathbf{a}}(\mathbf{a}) d\mathbf{a}, \tag{3.4}$$

with $\sum_{j=1}^{n_e} f_j = 1$. The stochastic grid refinement is terminated when convergence measure δ_{n_e} is smaller than a threshold value $\delta_{n_e} < \overline{\delta}$ where

$$\delta_{n_e} = \max\left(\frac{\left\| \mu_{w_{\lfloor n_e/2 \rfloor}}(\mathbf{x}, t) - \mu_{w_{n_e}}(\mathbf{x}, t) \right\|_\infty}{\left\| \mu_{w_{n_e}}(\mathbf{x}, t) \right\|_\infty}, \frac{\left\| \sigma_{w_{\lfloor n_e/2 \rfloor}}(\mathbf{x}, t) - \sigma_{w_{n_e}}(\mathbf{x}, t) \right\|_\infty}{\left\| \sigma_{w_{n_e}}(\mathbf{x}, t) \right\|_\infty} \right), \tag{3.5}$$

with $\mu_w(\mathbf{x}, t)$ and $\sigma_w(\mathbf{x}, t)$ the mean and standard deviation of $w(\mathbf{x}, t, \omega)$, or when a maximum number of samples $\bar{n}_s$ is reached. Convergence measure δ_{n_e} can be extended to include also higher statistical moments of the output.

In elements where the quadratic second-degree interpolation results in an extremum other than in a quadrature point, the element is subdivided into $\tilde{n}_e = 2^{n_a}$ subelements with a linear first-degree Newton-Cotes approximation of the response without performing additional deterministic solves. It is proven in [44] that the resulting approach satisfies the extrema diminishing (ED) robustness concept in probability space

$$\min_A(w(\mathbf{a})) \geq \min_A(u(\mathbf{a})) \wedge \max_A(w(\mathbf{a})) \leq \max_A(u(\mathbf{a})), \quad \forall u(\mathbf{a}), \tag{3.6}$$

where the arguments $\mathbf{x}$ and t are omitted for simplicity of the notation. The ED property leads to the advantage that no non-zero probabilities of unphysical realizations can be predicted due to overshoots or undershoots at discontinuities in the response. Due to the location of the Newton-Cotes quadrature points the deterministic samples are also reused in successive refinements and the samples are used in approximating the response in multiple elements.

3.2. Efficient Uncertainty Quantification Interpolation at Constant Phase

Polynomial Chaos methods usually require a fast increasing number of samples with time to maintain a constant accuracy. Performing the uncertainty quantification interpolation of oscillatory samples at constant phase instead of at constant time results, however, in a constant accuracy with a constant number of samples. Assume, therefore, that solving (2.2) for realizations of the random parameters $\mathbf{a}_k$ results in oscillatory samples $v_k(t) = u(\mathbf{a}_k)$, of which the phase $v_{\phi_k}(t) = \phi(t, \mathbf{a}_k)$ is a well-defined monotonically increasing function of time t for $k = 1, \ldots, n_s$.

In order to interpolate the samples $\mathbf{v}(t) = \{v_1(t), \ldots, v_{n_a}(t)\}$ at constant phase, first, their phase as function of time $\mathbf{v}_\phi(t) = \{v_{\phi_1}(t), \ldots, v_{\phi_{n_a}}(t)\}$ is extracted from the deterministic solves $\mathbf{v}(t)$. Second, the time series for the phase $\mathbf{v}_\phi(t)$ are used to transform the samples $\mathbf{v}(t)$ into functions of their phase $\hat{\mathbf{v}}(\mathbf{v}_\phi(t))$ according to

$$\hat{v}_k(v_{\phi_k}(t)) = v_k(t), \tag{3.7}$$

for $k = 1, \ldots, n_s$, see Figure 2. Third, the sampled phases $\mathbf{v}_\phi(t)$ are interpolated to the function $w_\phi(t, \mathbf{a})$:

$$w_\phi(t, \mathbf{a}) = h(\mathbf{v}_\phi(t)), \tag{3.8}$$

as approximation of $\phi(t, \mathbf{a})$. Finally, the transformed samples $\hat{\mathbf{v}}(\mathbf{v}_\phi(t))$ are interpolated at a constant phase $\varphi \in w_\phi(t, \mathbf{a})$ to

$$\hat{w}(\varphi, \mathbf{a}) = h(\hat{\mathbf{v}}(\varphi)). \tag{3.9}$$
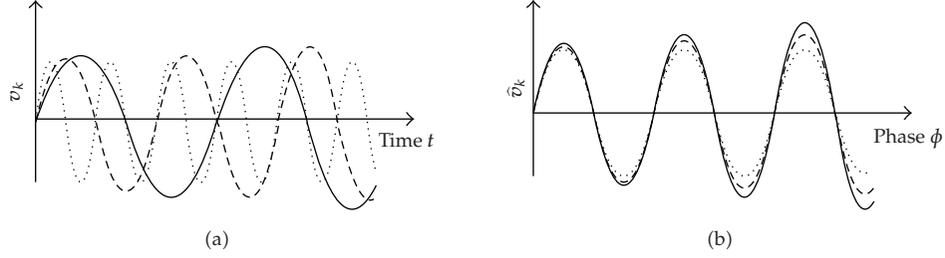
Figure 2: Oscillatory samples as function of time and phase.

Repeating the latter interpolation for all phases $\varphi \in w_\phi(t, \mathbf{a})$ results in the function $\widehat{w}(w_\phi(t, \mathbf{a}), \mathbf{a})$. The interpolation $\widehat{w}(w_\phi(t, \mathbf{a}), \mathbf{a})$ is then transformed back to an approximation in the time domain $w(t, \mathbf{a})$ as follows:

$$w(t, \mathbf{a}) = \widehat{w}(w_\phi(t, \mathbf{a}), \mathbf{a}). \tag{3.10}$$

The resulting function $w(t, \mathbf{a})$ is an approximation of the unknown response surface $u(t, \mathbf{a})$ as function of time t and the random parameters $\mathbf{a}(\omega)$. The actual sampling g and interpolation h is performed using the extrema diminishing uncertainty quantification method l based on Newton-Cotes quadrature in simplex elements described in the previous section.

This uncertainty quantification formulation for oscillatory responses is proven to achieve a bounded error $\widehat{\varepsilon}(\varphi, \mathbf{a}) = |\widehat{w}(\varphi, \mathbf{a}) - \widehat{u}(\varphi, \mathbf{a})|$ as function of phase φ for periodic responses according to

$$\widehat{\varepsilon}(\varphi, \mathbf{a}) < \delta, \quad \forall \varphi \in \mathbb{R}, \ \mathbf{a} \in A, \tag{3.11}$$

where δ is defined by

$$\widehat{\varepsilon}(\varphi, \mathbf{a}) < \delta, \quad \forall \varphi \in [0, 1], \ \mathbf{a} \in A. \tag{3.12}$$

The error $\varepsilon(t, \mathbf{a}) = |w(t, \mathbf{a}) - u(t, \mathbf{a})|$ is also bounded in time under certain conditions, see [44].

The phases $\mathbf{v}_\phi(t)$ are extracted from the samples based on the local extrema of the time series $\mathbf{v}(t)$. A trial and error procedure identifies a cycle of oscillation based on two or more successive local maxima. The selected cycle is accepted if the maximal error of its extrapolation in time with respect to the actual sample is smaller than a threshold value $\overline{\varepsilon}_k$ for at least one additional cycle length. The functions for the phases $\mathbf{v}_\phi(t)$ in the whole time domain T are constructed by identifying all successive cycles of $\mathbf{v}(t)$ and linear extrapolation to $t = 0$ and $t = t_{\max}$ before and after the first and last complete cycle, respectively. The phase is normalized to zero at the start of the first cycle and a user-defined parameter determines whether the sample is assumed to attain a local extremum at $t = 0$. The interpolation at constant phase is restricted to the time domain that corresponds to the range of phases that is reached by all samples in each of the elements. If the phase $v_{\phi_k}(t)$ cannot be extracted from one of the samples $v_k(t)$ for $k = 1, \ldots, n_s$, then uncertainty quantification interpolation h is directly applied to the time-dependent samples $\mathbf{v}(t)$.
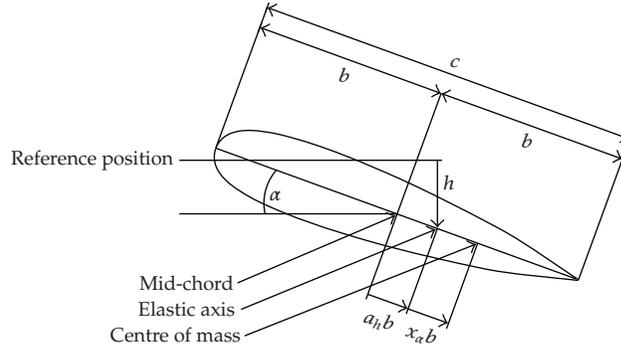
Figure 3: The two-degree-of-freedom airfoil flutter model.

4. A Nonlinear Airfoil Fluid-Structure Interaction System

The nonlinear airfoil flutter model used to simulate the airfoil fluid-structure interaction system is given in Section 4.1. The deterministic bifurcation behavior is briefly considered in Section 4.2.

4.1. A Two-Degree-of-Freedom Airfoil Flutter Model

The two-degree-of-freedom model for the pitch and plunge motion of an airfoil used here, see Figure 3, has also been studied deterministically, for example, by Lee et al. [48] and stochastically using Fourier Chaos by Millman et al. [38]. The aeroelastic equations of motion with cubic restoring springs in both pitch and plunge are given in [48] by

$$
\xi'' + x_\alpha \alpha'' + 2\zeta_\xi \frac{\overline{\omega}}{U^*} \xi' + \left(\frac{\overline{\omega}}{U^*}\right)^2 \left(\xi + \beta_\xi \xi^3\right) = -\frac{1}{\pi\mu} C_{\mathrm{L}}(\tau),
$$

$$
\frac{x_\alpha}{r_\alpha^2} \xi'' + \alpha'' + 2\frac{\zeta_\alpha}{U^*} \alpha' + \frac{1}{U^{*2}} \left(\alpha + \beta_\alpha \alpha^3\right) = \frac{2}{\pi\mu r_\alpha^2} C_{\mathrm{M}}(\tau),
$$

(4.1)

where $\alpha(\tau)$ is the pitch angle and $\xi(\tau) = h/b$ is the nondimensional version of the plunge displacement h of the elastic axis, with $b = c/2$ the half-chord, and initial conditions $\alpha(0) = \alpha_0$ and $\xi(0) = \xi_0$. The nonlinear spring constants in plunge and pitch are, respectively, β_ξ and β_α. Equivalently, the viscous damping coefficients are ζ_ξ and ζ_α. The ratio of natural frequencies is given by $\overline{\omega} = \omega_\xi/\omega_\alpha$, where ω_ξ and ω_α are the natural frequencies of the uncoupled plunging and pitching modes, respectively. The mass ratio μ is defined as $m/\pi\rho b^2$, with m the airfoil mass, and ρ the air density. The radius of gyration about the elastic axis is r_α, where elastic axis is located at a distance $a_{\mathrm{h}}b$ from the mid-chord point, and the mass center is located at a distance $x_\alpha b$ from the elastic axis. The bifurcation parameter is the ratio of time scales of the structure and the flow defined as $U^* = U/(b\omega_\alpha)$, with U the free stream velocity. The primes denote differentiation with respect to nondimensionalized time $\tau = Ut/b$. The expressions

for the aerodynamic force and moment coefficients, $C_L(\tau)$ and $C_M(\tau)$ are given by Fung [49] as

$$
C_L(\tau) = \pi \left(\xi'' - a_h \alpha'' + \alpha' \right) + 2\pi \left\{ \alpha(0) + \xi'(0) + \left[\frac{1}{2} - a_h \right] \alpha'(0) \right\} \phi(\tau)
$$

$$
+ 2\pi \int_0^\tau \phi(\tau - \sigma) \left[\alpha'(\sigma) + \xi''(\sigma) + \left(\frac{1}{2} - a_h \right) \alpha''(\sigma) \right] d\sigma,
$$

$$
C_M(\tau) = \pi \left(\frac{1}{2} + a_h \right) \left\{ \alpha(0) + \xi'(0) + \left(\frac{1}{2} - a_h \right) \alpha'(0) \right\} \phi(\tau) \tag{4.2}
$$

$$
+ \pi \left(\frac{1}{2} + a_h \right) \int_0^\tau \phi(\tau - \sigma) \left\{ \alpha'(\sigma) + \xi''(\sigma) + \left(\frac{1}{2} - a_h \right) \alpha''(\sigma) \right\} d\sigma
$$

$$
+ \frac{\pi}{2} a_h \left(\xi'' - a_h \alpha'' \right) - \left(\frac{1}{2} - a_h \right) \frac{\pi}{2} \alpha' - \frac{\pi}{16} \alpha'',
$$

where $\phi(\tau)$ is the Wagner function

$$
\phi(\tau) = 1 - \psi_1 e^{-\varepsilon_1 \tau} - \psi_2 e^{-\varepsilon_2 \tau}, \tag{4.3}
$$

with the constants $\psi_1 = 0.165$, $\psi_2 = 0.335$, $\varepsilon_1 = 0.0455$, and $\varepsilon_2 = 0.3$ given by Jones [50]. Based on (4.1) to (4.3), the following set of first-order ordinary differential equations for the motion of the airfoil is derived in [48]

$$
x_1' = x_2,
$$

$$
x_2' = \frac{c_0 H - d_0 P}{d_0 c_1 - c_0 d_1},
$$

$$
x_3' = x_4,
$$

$$
x_4' = \frac{-c_1 H + d_1 P}{d_0 c_1 - c_0 d}, \tag{4.4}
$$

$$
x_5' = x_1 - \varepsilon_1 x_5,
$$

$$
x_6' = x_1 - \varepsilon_2 x_6,
$$

$$
x_7' = x_3 - \varepsilon_1 x_7,
$$

$$
x_8' = x_3 - \varepsilon_2 x_8,
$$

with

$$
P = c_2 x_4 + c_3 x_2 + c_4 x_3 + c_5 x_3^3 + c_6 x_1 + c_7 x_5 + c_8 x_6 + c_9 x_7 + c_{10} x_8 - f(\tau),
$$
$$
H = d_2 x_2 + d_3 x_1 + d_4 x_1^3 + d_5 x_4 + d_6 x_3 + d_7 x_5 + d_8 x_6 + d_9 x_7 + d_{10} x_8 - g(\tau), \tag{4.5}
$$
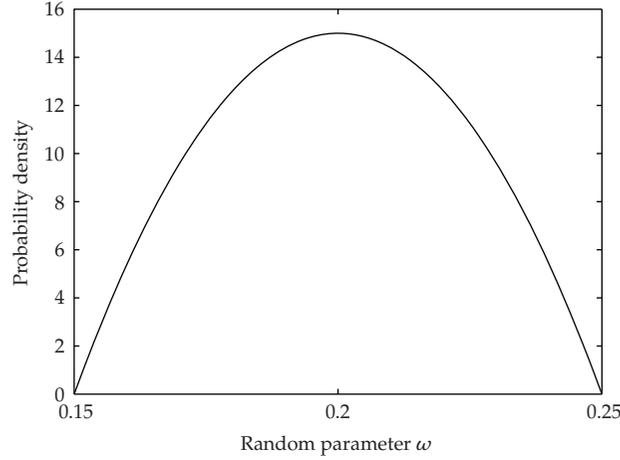
Figure 4: Input probability density function for the ratio of natural frequencies $\overline{\omega}$.

where a vector $\{x_i\}_{i=1}^8$ of new variables is defined as

$$x_1 = \alpha, \qquad x_2 = \alpha', \qquad x_3 = \xi, \qquad x_4 = \xi',$$

$$x_5 = w_1, \qquad x_6 = w_2, \qquad x_7 = w_3, \qquad x_8 = w_4,$$

$$w_1 = \int_0^\tau e^{-\epsilon_1(\tau-\sigma)}\alpha(\sigma)d\sigma,$$

$$w_2 = \int_0^\tau e^{-\epsilon_2(\tau-\sigma)}\alpha(\sigma)d\sigma, \tag{4.6}$$

$$w_3 = \int_0^\tau e^{-\epsilon_1(\tau-\sigma)}\xi(\sigma)d\sigma,$$

$$w_4 = \int_0^\tau e^{-\epsilon_2(\tau-\sigma)}\xi(\sigma)d\sigma.$$

Following [48], the solution is determined numerically until $\tau = 2000$ using the explicit fourth order Runge-Kutta method with a time step of $\Delta\tau = 0.1$, which is approximately $1/256$ of the smallest period. The other parameter values are chosen to be $\mu = 100$, $a_h = -0.5$, $x_\alpha = 0.25$, $r_\alpha = 0.5$, $\xi_0 = 0$, $\beta_\xi = 0$, and $\zeta_\alpha = \zeta_\xi = 0$ as in [48].

The system parameters that are assumed to be uncertain are $\overline{\omega}$, β_α, and α_0. The randomness of these three parameters is described by a symmetric unimodal beta distribution with $\beta_1 = \beta_2 = 2$ to limit their parameter range to a finite domain with vanishing probability at the interval boundaries. The ratio of natural frequencies $\overline{\omega}$ has a mean of $\mu_{\overline{\omega}} = 0.2$ and an interval of $\mu_{\overline{\omega}} \in [0.15; 0.25]$. The input probability density function for $\overline{\omega}$ is shown as an example in Figure 4. For a hard spring model with $\beta_\alpha > 0$ the system exhibits a stable limit cycle oscillation beyond the first bifurcation point [51]. The mean of the nonlinear stiffness parameter $\beta_\alpha(w)$ is chosen to be $\mu_{\beta_\alpha} = 100$ to limit the pitch angle α to the domain in which the aerodynamic model is valid, and the interval is set to $\beta_\alpha(w) \in [90, 110]$. The initial condition α_0 has an interval of $\alpha_0 \in [9, 11]$ degrees around mean $\mu_{\alpha_0} = 10°$. The resulting coefficients
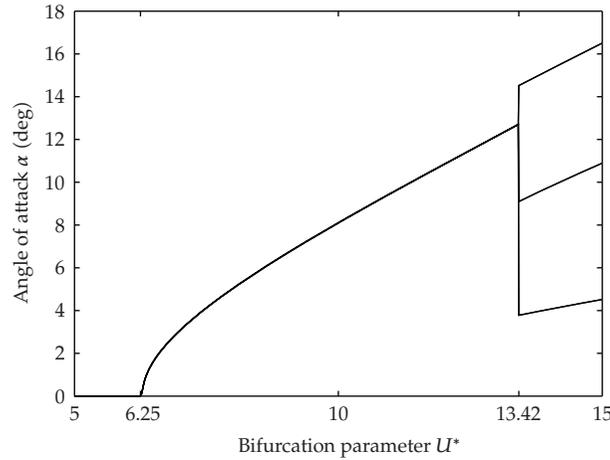
Figure 5: Deterministic bifurcation plot of the nonlinear airfoil flutter system.

of variation for the random parameters are $cv_{\overline{\omega}}$ = 11.2%, cv_{β_α} = 4.48%, and cv_{α_0} = 4.48%. The effect of the random parameters $\overline{\omega}$, β_α, and α_0 is analyzed in Sections 5 to 7. First the deterministic bifurcation behavior is explored in the next section.

4.2. Deterministic Bifurcation Behavior

The deterministic bifurcation plot for the aeroelastic system given by (4.1) to (4.3) is shown in Figure 5 as function of bifurcation parameter $U^* \in [5, 15]$ in terms of the angles of attack α for which $\alpha' = 0$ in the asymptotic range. In what follows the first deterministic bifurcation point of $U^* = 6.25$ the system response is a decaying oscillation to $\alpha = 0$. At $U^* = 6.25$ the system exhibits a supercritical Hopf bifurcation to a stable period-1 limit cycle oscillation with an increasing amplitude for increasing U^*. At the second bifurcation point $U^* = 13.42$ the response shows an abrupt bifurcation to a higher period limit cycle oscillation. The oscillation amplitude continues to increase beyond the second bifurcation point.

Three typical time histories of pitch α and plunge ξ in the three different regimes are given in Figure 6 as function of nondimensional time τ for $U^* = \{5, 10, 15\}$. At $U^* = 5$ both α and ξ are decaying oscillations to the stable fixed point $(\alpha, \xi) = (0, 0)$. A period-1 oscillation can be identified for α and ξ at $U^* = 10$. For $U^* = 15$ the angle of attack α exhibits a higher period oscillation with a higher amplitude, while the plunge deflection ξ maintains a period-1 oscillation. The mean, standard deviation, and probability distribution of the more interesting pitch degree of freedom α is, therefore, considered in the following stochastic flutter analysis. The plunge ξ is used to extract the phase of the oscillation.

5. Random Natural Frequency Ratio $\overline{\omega}(\omega)$

First the effect of randomness in the ratio of natural frequencies $\overline{\omega}$ is resolved. The results are presented in terms of the time histories of the mean $\mu_\alpha(\tau)$ and standard deviation $\sigma_\alpha(\tau)$ of the pitch angle α in Figure 7. The bifurcation of the system is illustrated in Figure 8 by the response surface of α as function of the random parameter $\overline{\omega}$ at $\tau = 2000$. In Figure 9

(a)

(b)

(c)
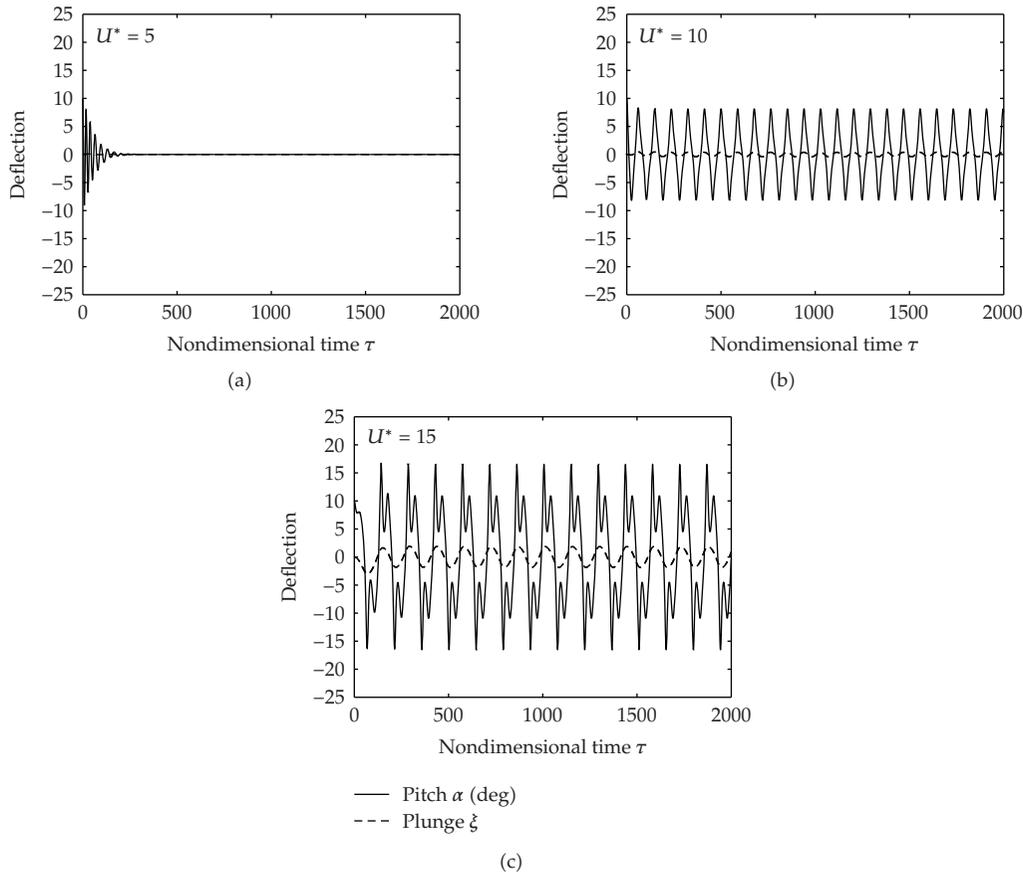
Figure 6: Time histories of pitch $\alpha(\tau)$ and plunge $\xi(\tau)$ in the three regimes of the deterministic airfoil flutter system.

the P-bifurcation behavior of the probability density function (PDF) of α also at $\tau = 2000$ is given. In all three figures the bifurcation parameter values $U^* = \{6.25; 10; 13.42; 15\}$ are considered. This corresponds to the first ($U^* = 6.25$) and second ($U = 13.42$) deterministic bifurcation point, and the period-1 ($U^* = 10$) and higher period ($U^* = 15$) regime. The case of $U^* = 5$ also considered in the previous section is not shown here, since the system response in the prebifurcation domain is equal to the trivial solution. The required number of sampling points n_s in the stochastic simulations for the different values of U^* is established after performing an convergence study which is summarized as an example in Tables 1–4. The results are compared to converged Monte Carlo reference solutions based on $n_s = 10^3$ samples.

For $U^* = 6.25$ the mean μ_α of the pitch angle shows a decaying oscillation to zero and the standard deviation approaches the steady asymptotic value of $\sigma_\alpha = 0.423$ after an initial increase from the deterministic initial condition in Figure 7(a). The decaying mean is caused by a combination of decaying and periodic realizations as can be concluded from the response surface of Figure 8(a). The non-zero asymptotic value of the standard deviation also indicates that due to the randomness in $\overline{\omega}$ the system is already stochastically bifurcated in the deterministic bifurcation point $U^* = 6.25$. The onset of the stochastic bifurcation occurs,

(a)

(b)

(c)

(d)
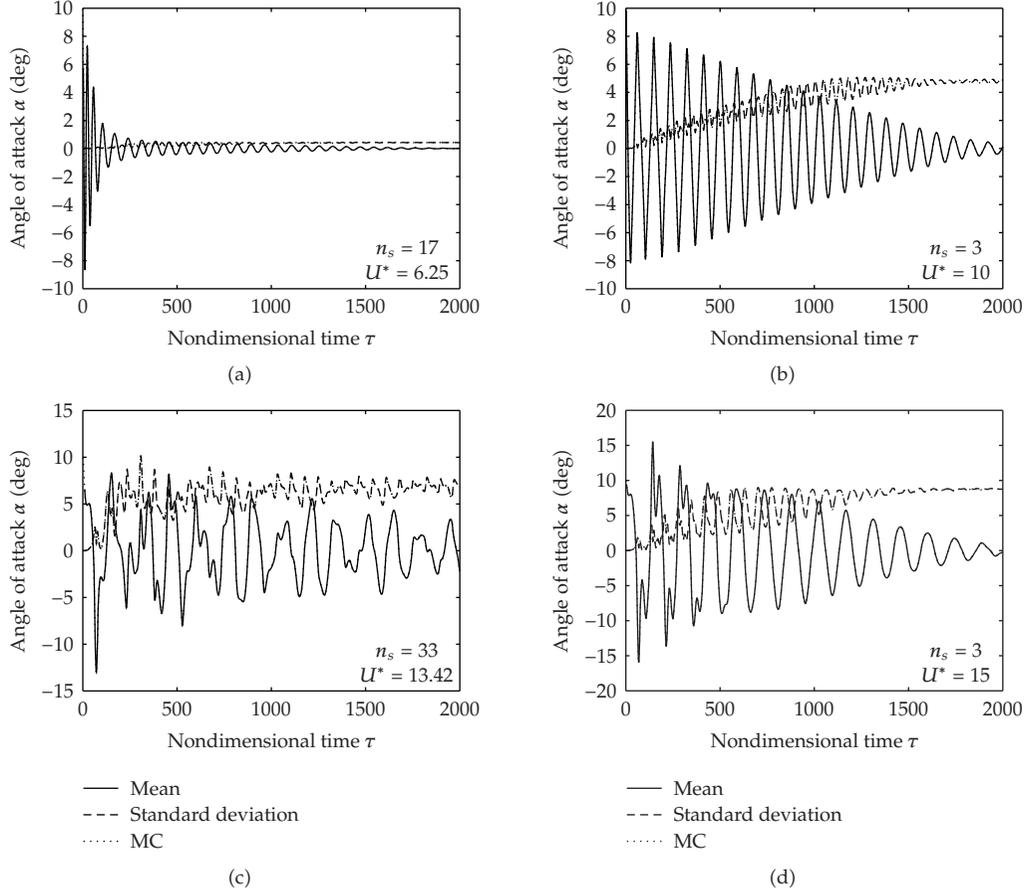
Figure 7: Time histories of the mean $\mu_\alpha(\tau)$ and standard deviation $\sigma_\alpha(\tau)$ of the pitch angle α due to the random frequency ratio $\overline{\omega}$.

Table 1: Relation between convergence measure δ_{n_e} and L_∞ error ε_{L_∞} for the mean μ_α and standard deviation σ_α at $U^* = 6.25$.

| n_e | n_s | Mean μ_α | | Standard deviation σ_α | |
		conv. δ_{n_e}	error ε_{L_∞}	conv. δ_{n_e}	error ε_{L_∞}
1	3	—	$1.515 \cdot 10^{-1}$	—	$1.533 \cdot 10^{0}$
2	5	$1.151 \cdot 10^{-1}$	$4.115 \cdot 10^{-2}$	$7.117 \cdot 10^{-1}$	$4.293 \cdot 10^{-1}$
4	9	$3.334 \cdot 10^{-2}$	$8.088 \cdot 10^{-3}$	$3.190 \cdot 10^{-1}$	$1.097 \cdot 10^{-1}$
8	17	$6.332 \cdot 10^{-3}$	$1.768 \cdot 10^{-3}$	$6.915 \cdot 10^{-2}$	$5.138 \cdot 10^{-2}$

therefore, at a lower value of the bifurcation parameter than in the deterministic case. As a consequence a deterministic flutter analysis predicts a later start of unstable behavior by neglecting the variability in system parameters, which can lead to disastrous effects by defining the flight envelope based on a too optimistic deterministic flutter boundary.

In the period-1 regime at $U^* = 10$ the mean μ_α exhibits a decaying oscillation due the fully periodic response. The resulting frequency differences lead to increasing phase differences in time and increasingly to realizations of opposite sign, which cancel each other

(a)

(b)

(c)

(d)
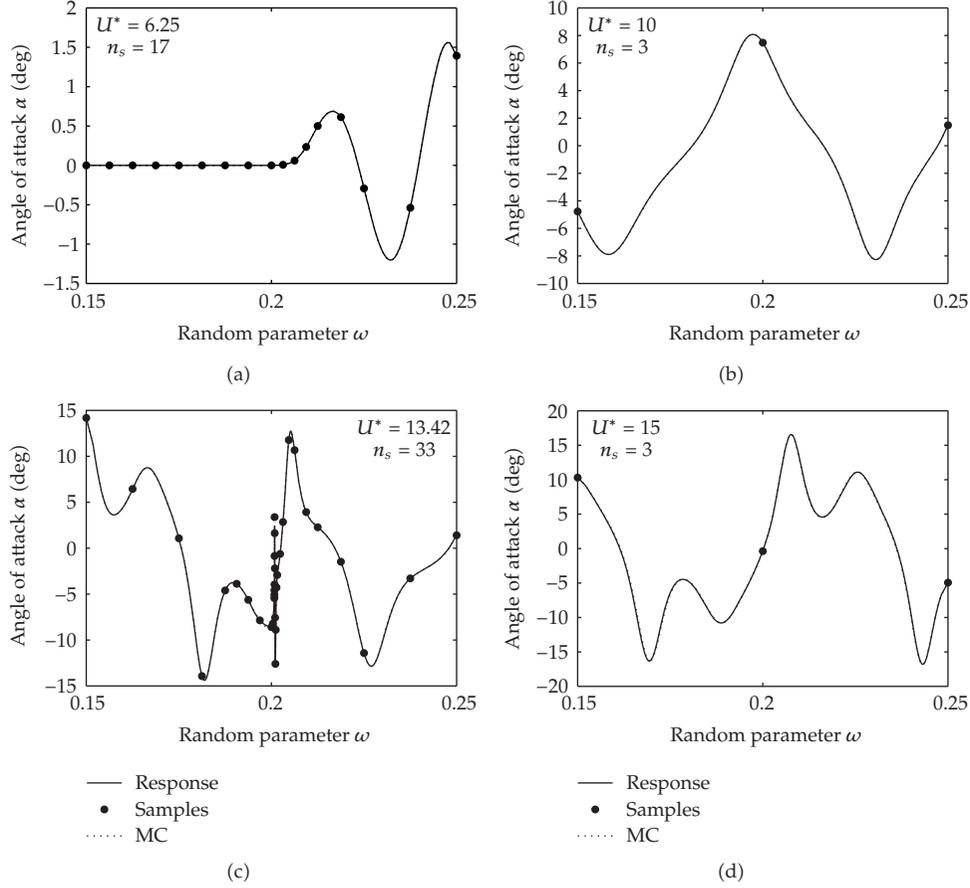
Figure 8: Response surface of the pitch angle α at $\tau = 2000$ as function of the random frequency ratio $\overline{\omega}$.

Table 2: Relation between convergence measure δ_{n_e} and L_∞ error ε_{L_∞} for the mean μ_α and standard deviation σ_α at $U^* = 10$.

n_e	n_s	Mean μ_α		Standard deviation σ_α	
		conv. δ_{n_e}	error ε_{L_∞}	conv. δ_{n_e}	error ε_{L_∞}
1	3	—	$1.677 \cdot 10^{-3}$	—	$1.743 \cdot 10^{-3}$
2	5	$1.814 \cdot 10^{-3}$	$2.141 \cdot 10^{-4}$	$1.791 \cdot 10^{-3}$	$4.635 \cdot 10^{-4}$
4	9	$2.533 \cdot 10^{-4}$	$1.092 \cdot 10^{-4}$	$4.536 \cdot 10^{-4}$	$2.335 \cdot 10^{-4}$
8	17	$1.318 \cdot 10^{-4}$	$1.237 \cdot 10^{-4}$	$2.291 \cdot 10^{-4}$	$1.983 \cdot 10^{-4}$

resulting in a decaying mean pitch. The standard deviation reaches a significantly higher steady asymptotic value of $\sigma_\alpha = 4.8$ due to the increased amplitudes of the limit cycle oscillation at higher values of U^*. The effect of $\overline{\omega}$ on the frequency of the response can be derived from the oscillatory response surface of Figure 8(b). The deterministic oscillation period shape of α shown in Figure 6(b) can also be recognized in the shape of the response surface.

At the second deterministic bifurcation point $U^* = 13.42$ the mean μ_α and standard deviation σ_α show an irregular behavior with only a slowly decaying mean and a large

(a)

(b)

(c)

(d)
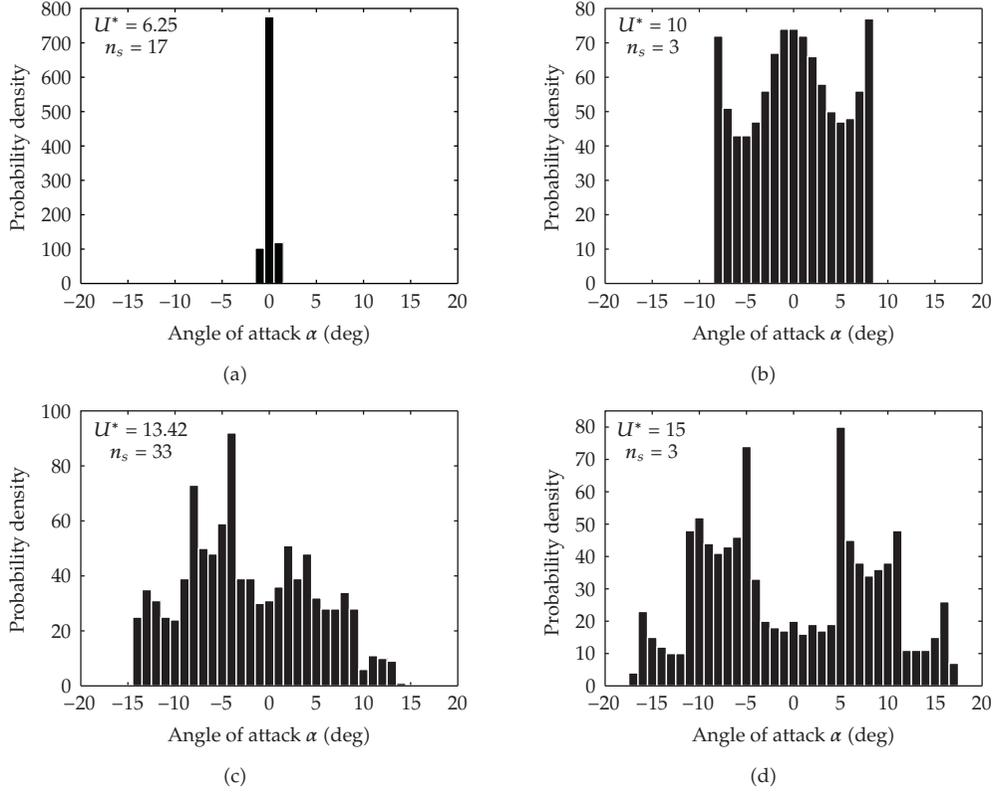
Figure 9: Bifurcation of the probability density function of the pitch angle α due to the random frequency ratio $\overline{\omega}$.

Table 3: Relation between convergence measure δ_{n_e} and L_∞ error ε_{L_∞} for the mean μ_α and standard deviation σ_α at $U^* = 13.42$.

n_e	n_s	Mean μ_α		Standard deviation σ_α	
		conv. δ_{n_e}	error ε_{L_∞}	conv. δ_{n_e}	error ε_{L_∞}
1	3	—	$3.176 \cdot 10^{-1}$	—	$3.224 \cdot 10^{-1}$
2	5	$2.666 \cdot 10^{-1}$	$1.905 \cdot 10^{-1}$	$3.389 \cdot 10^{-1}$	$2.117 \cdot 10^{-1}$
4	9	$1.679 \cdot 10^{-1}$	$1.024 \cdot 10^{-1}$	$1.685 \cdot 10^{-1}$	$1.629 \cdot 10^{-1}$
8	17	$5.335 \cdot 10^{-2}$	$1.136 \cdot 10^{-1}$	$7.391 \cdot 10^{-2}$	$1.272 \cdot 10^{-1}$
16	33	$1.103 \cdot 10^{-1}$	$6.984 \cdot 10^{-3}$	$1.253 \cdot 10^{-1}$	$1.060 \cdot 10^{-2}$

asymptotic standard deviation of approximately $\sigma_\alpha = 7$. This is a result of the discontinuity in the response of α in Figure 8(c) caused by the deterministic bifurcation present at $\mu_{\overline{\omega}} = 0.2$. On the left and the right of the discontinuity at $\overline{\omega} = 0.2$ the higher period and period-1 shape function can be recognized in the response, respectively, which suggests a subcritical Hopf bifurcation as function of $\overline{\omega}$. For $U^* = 15$ in the higher period regime μ_α and σ_α give again a decaying oscillation and a steady asymptotic value of $\sigma_\alpha = 8.8$, respectively. The time histories of μ_α and σ_α are initially more complex than in the period-1 regime of $U^* = 10$ due to the higher period behavior of the realizations. In the response surface of Figure 8(d) the deterministic higher period shape of α shown Figure 6(c) can again be identified.

Table 4: Relation between convergence measure δ_{n_e} and L_∞ error ε_{L_∞} for the mean μ_α and standard deviation σ_α at $U^* = 15$.

n_e	n_s	Mean μ_α		Standard deviation σ_α	
		conv. δ_{n_e}	error ε_{L_∞}	conv. δ_{n_e}	error ε_{L_∞}
1	3	—	$7.315 \cdot 10^{-3}$	—	$7.503 \cdot 10^{-3}$
2	5	$6.319 \cdot 10^{-3}$	$9.960 \cdot 10^{-4}$	$7.444 \cdot 10^{-3}$	$1.619 \cdot 10^{-3}$
4	9	$9.979 \cdot 10^{-4}$	$2.598 \cdot 10^{-4}$	$1.615 \cdot 10^{-3}$	$3.905 \cdot 10^{-4}$
8	17	$2.636 \cdot 10^{-4}$	$1.675 \cdot 10^{-4}$	$3.608 \cdot 10^{-4}$	$3.812 \cdot 10^{-4}$

The required number of samples n_s used in the stochastic simulations depends significantly on the value of bifurcation parameter U^*. In Tables 1–4 the convergence δ_{n_e} and the error ε_{L_∞} with respect to the Monte Carlo reference solutions $\mu_{\alpha_{MC}}(\tau)$ and $\sigma_{\alpha_{MC}}(\tau)$ are given. The convergence measure δ_{n_e} used for μ_α and σ_α separately is defined by (3.5) and the L_∞ error ε_{L_∞} is defined for μ_α as

$$\varepsilon_{L_\infty} = \frac{\left\| \mu_\alpha(\tau) - \mu_{\alpha_{MC}}(\tau) \right\|_\infty}{\left\| \mu_{\alpha_{MC}}(\tau) \right\|_\infty} \tag{5.1}$$

and equivalently for σ_α. The method is highly efficient in the periodic regimes, in which $n_s = 3$ samples is already sufficient to match the Monte Carlo results based on $n_s = 10^3$ samples. This holds even for the oscillatory response surface in the higher period case of $U^* = 15$. At the deterministic bifurcation points the adaptive method robustly captures the singularity in the response surface by automatically refining near the bifurcation in probability space.

The resulting bifurcation behavior of the PDF of α at $\tau = 2000$ is shown in Figure 9. At $U^* = 6.25$ the PDF is already bifurcated from a delta function in the stochastic pre-bifurcation domain to a unimodal PDF with the highest probability at $\alpha = 0°$. The PDF develops into a multimodal distribution with peaks at $\alpha = \pm 8$ due to the oscillatory behavior of the response at $U^* = 10$. The multimodal PDF evolves further into a distribution with 6 peaks at approximately $\alpha = \{\pm 5, \pm 11, \pm 16\}$ due to the higher period motion at $U^* = 15$. At the second deterministic bifurcation point $U^* = 13.42$ the PDF is in an intermediate state between the approximately symmetric multimodal distributions of $U^* = 10$ and $U^* = 15$.

The stochastic behavior of the system is also shown in Figure 10 for the three random parameters in terms of the bifurcation of the maximum standard deviation $\sigma_{\alpha_{max}}$ in the asymptotic range defined as

$$\sigma_{\alpha_{max}} = \max_{\tau \in [1500, 2000]} (\sigma_\alpha(\tau)). \tag{5.2}$$

It can be seen in Figure 10(a) that the first bifurcation of the maximum standard deviation $\sigma_{\alpha_{max}}$ starts at an earlier location than the first deterministic bifurcation. In the period-1 regime in between the two deterministic bifurcations $\sigma_{\alpha_{max}}$ gradually increases due to the increasing limit cycle oscillation amplitude in combination with the random frequency. At the second deterministic bifurcation point the standard deviation reaches a local maximum of $\sigma_{\alpha_{max}} = 8.0$ and it continues to increase at a higher rate beyond $U^* = 13.42$.

(a)

(b)

(c)

• Maximum standard deviation
—— MC
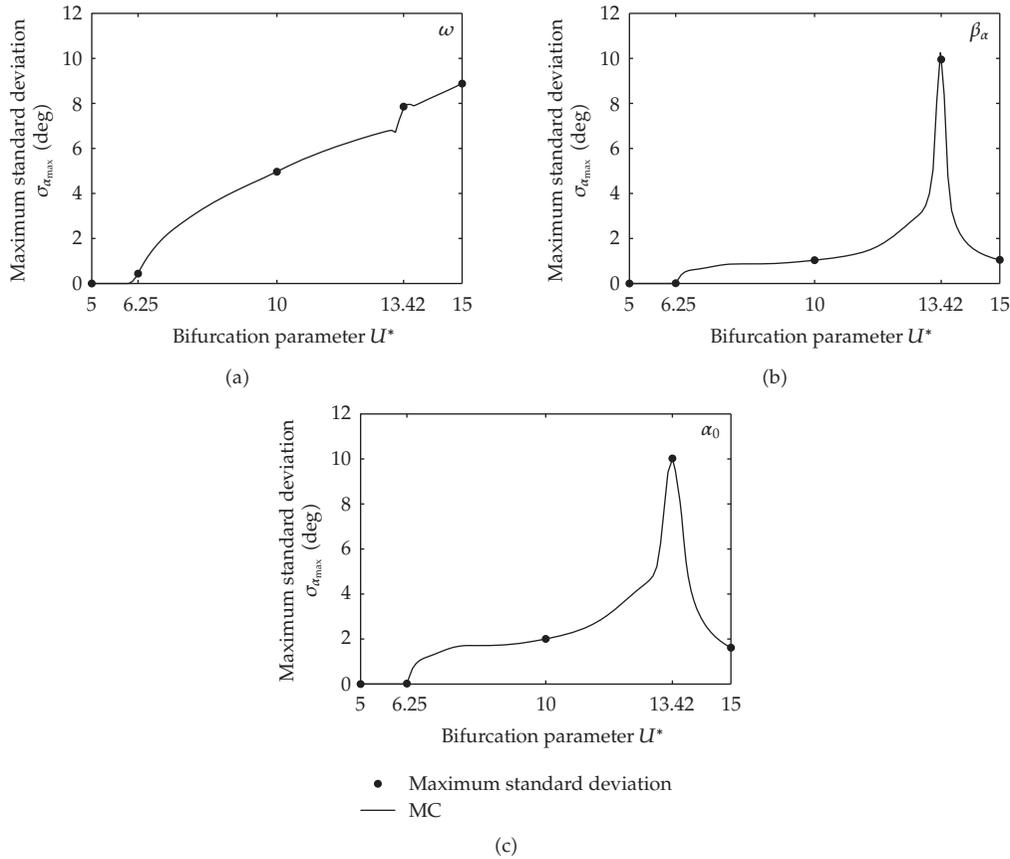
Figure 10: Bifurcation of the maximum of the standard deviation $\sigma_{\alpha_{max}}$ for the three random parameters.

6. Random Nonlinearity Parameter $\beta_\alpha(\omega)$

Next the effect of a random nonlinearity parameter for the pitch degree of freedom β_α on the stochastic behavior of the system is considered. The results for the mean and standard deviation, the response surface, and the PDF are shown in Figures 11–13. The required number of samples n_s in the simulations for random β_α is again determined based on convergence studies.

For $U^* = 6.25$, $U^* = 10$, and $U^* = 15$ the random parameter β_α has a qualitatively different effect on the system than $\overline{\omega}$. Both the mean μ_α and standard deviation σ_α decay in this case to zero for $U^* = 6.25$, which suggests that randomness in β_α does not lead to an earlier bifurcation. For both $U^* = 10$ and $U^* = 15$ the mean shows an oscillatory behavior, which closely resembles the deterministic time histories of Figures 6(b) and 6(c). The standard deviation has for these two cases a low constant value of approximately $\sigma_\alpha = 0.7$ and $\sigma_\alpha = 0.3$, respectively. The response surfaces of Figures 12(b) and 12(d) are also nonoscillatory, which indicates that β_α has little effect on the oscillation frequency. It can be concluded that randomness in β_α has for these values of U^* a small effect of the system behavior. This can be understood from the fact that the nonlinearity parameter has only a significant effect on the limit cycle oscillation amplitude. It can, therefore, be expected that a

(a)

(b)

(c)

(d)

— Mean
--- Standard deviation
······ MC

— Mean
--- Standard deviation
······ MC
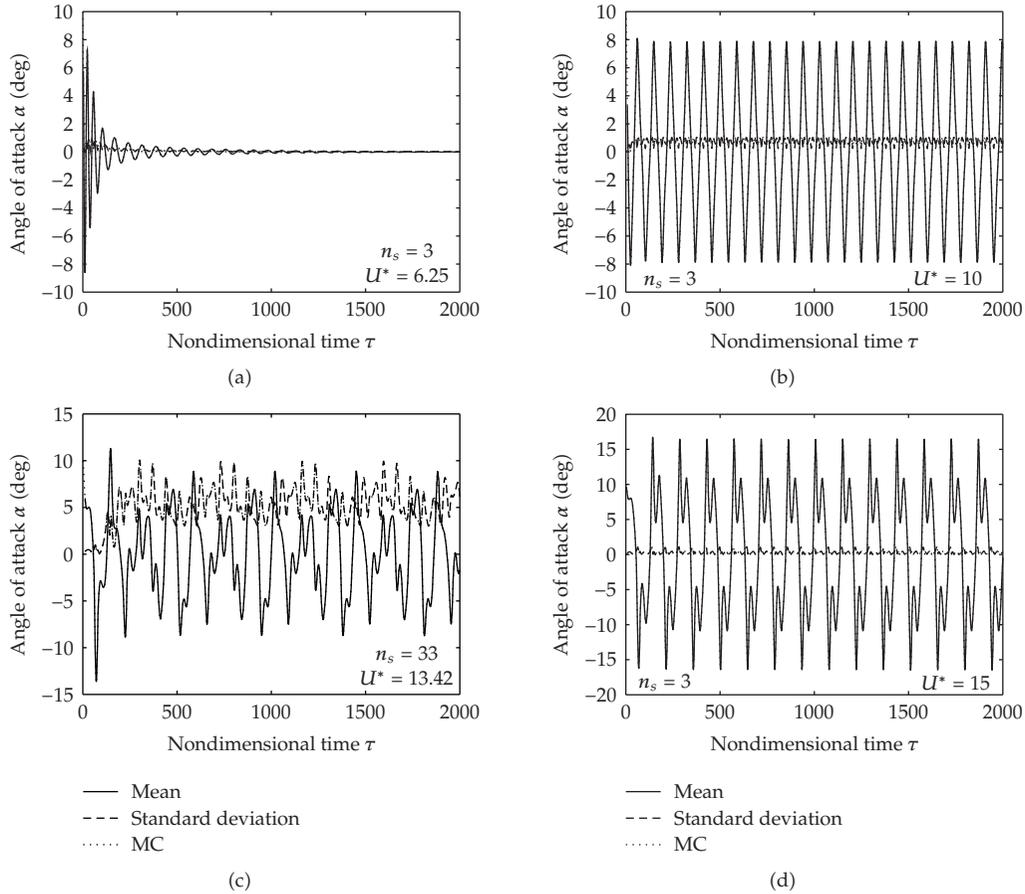
Figure 11: Time histories of the mean $\mu_\alpha(\tau)$ and standard deviation $\sigma_\alpha(\tau)$ of the pitch angle α due to the random nonlinearity parameter β_α.

random β_α has a small effect in the prebifurcation domain, and that the effect in the periodic regimes is constant in time.

However, the effect of random β_α is significant in the second deterministic bifurcation point $U^* = 13.42$. The stochastic system shows at $U^* = 13.42$ an irregular behavior with a non-decaying mean μ_α and a large standard deviation oscillating around approximately $\sigma_\alpha = 7$, which is comparable to the results for random $\overline{\omega}$. The sudden large effect of β_α is caused by the discontinuity in the response at $\mu_{\beta_\alpha} = 100$. So, even the randomness in parameter β_α, which has in general a small effect on the response, becomes important at the conditions of the second deterministic bifurcation point $U^* = 13.42$. The adaptive method resolves also this discontinuous response accurately and the other response surface approximations require again only $n_s = 3$ deterministic simulations to match the Monte Carlo results.

The PDF in Figure 13 is also significantly distorted from the unimodal input distribution at $U^* = 13.42$ only. For $U^* = 6.25$ the histogram shows a delta function PDF, which indicates that the stochastic bifurcation for random β_α has not yet started in the first deterministic bifurcation point. This observation is confirmed by the bifurcation of the maximum standard deviation $\sigma_{\alpha_{max}}$ in Figure 10(b). The stochastic bifurcation of $\sigma_{\alpha_{max}}$

(a)

(b)

(c)

(d)
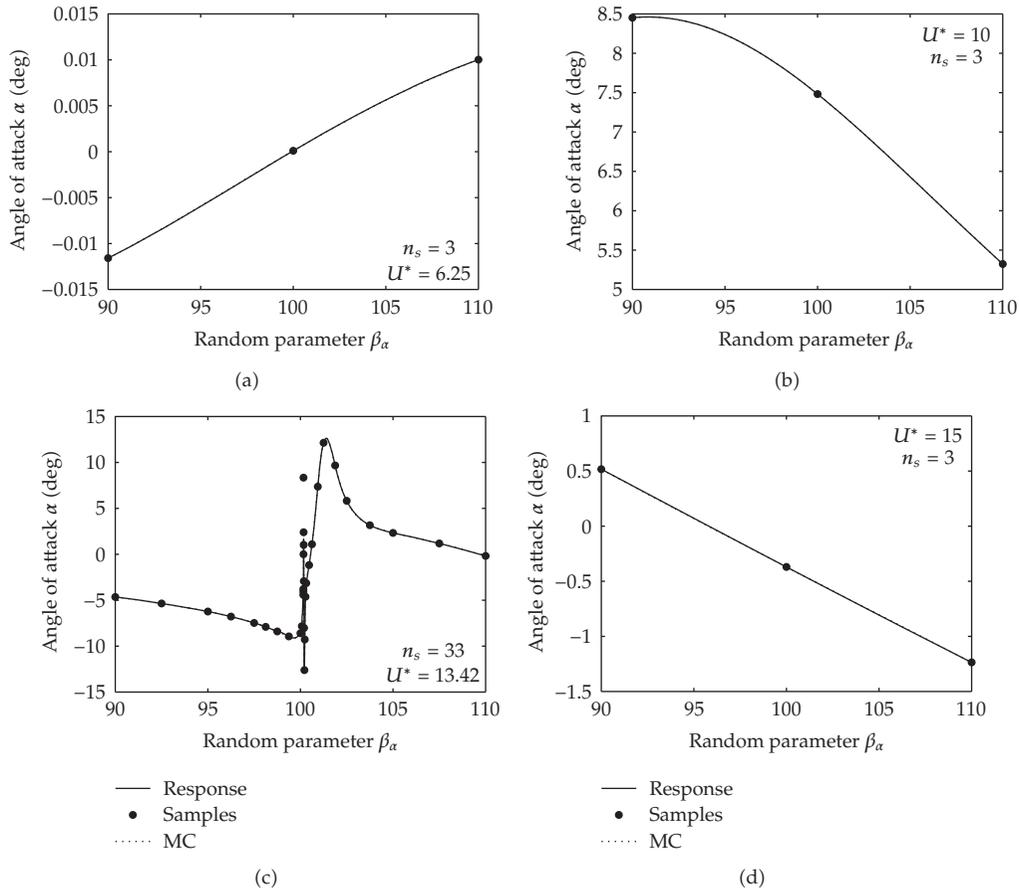
Figure 12: Response surface of the pitch angle α at $\tau = 2000$ as function of the random nonlinearity parameter β_α.

coincides with the location of the first deterministic bifurcation, which suggests that β_α has no effect on the value of U^* at which the unstable behavior starts, since bifurcation is initially a linear phenomenon. On the other hand, the nonlinearity parameter does have an effect on the limit cycle oscillation amplitude in the period-1 regime between $U^* = 6.25$ and $U^* = 13.42$, where the standard deviation is approximately constant at a value of $\sigma_{\alpha_{max}} = 1$. In accordance with the previous results the standard deviation reaches a maximum in the deterministic bifurcation point of $\sigma_{\alpha_{max}} = 10.3$. Beyond $U^* = 13.42$ the standard deviation drops to the value $\sigma_{\alpha_{max}} = 1$ of the period-1 domain. Whether the response is period-1 or higher period does, therefore, not affect the influence of β_α on the oscillation amplitude.

7. Random Initial Condition $\alpha_0(\omega)$

The results for randomness in the pitch initial condition α_0 are given in Figures 14–16. The mean μ_α shows a decaying oscillation to zero for $U^* = 6.25$ and periodic oscillations for $U^* = 10$ and $U^* = 15$, which closely resemble the deterministic results of Figure 6. The standard deviation σ_α also decays to zero for $U^* = 6.25$, and oscillates around only $\sigma_\alpha = 1.4$ and $\sigma_\alpha = 0.5$

(a)

(b)

(c)

(d)
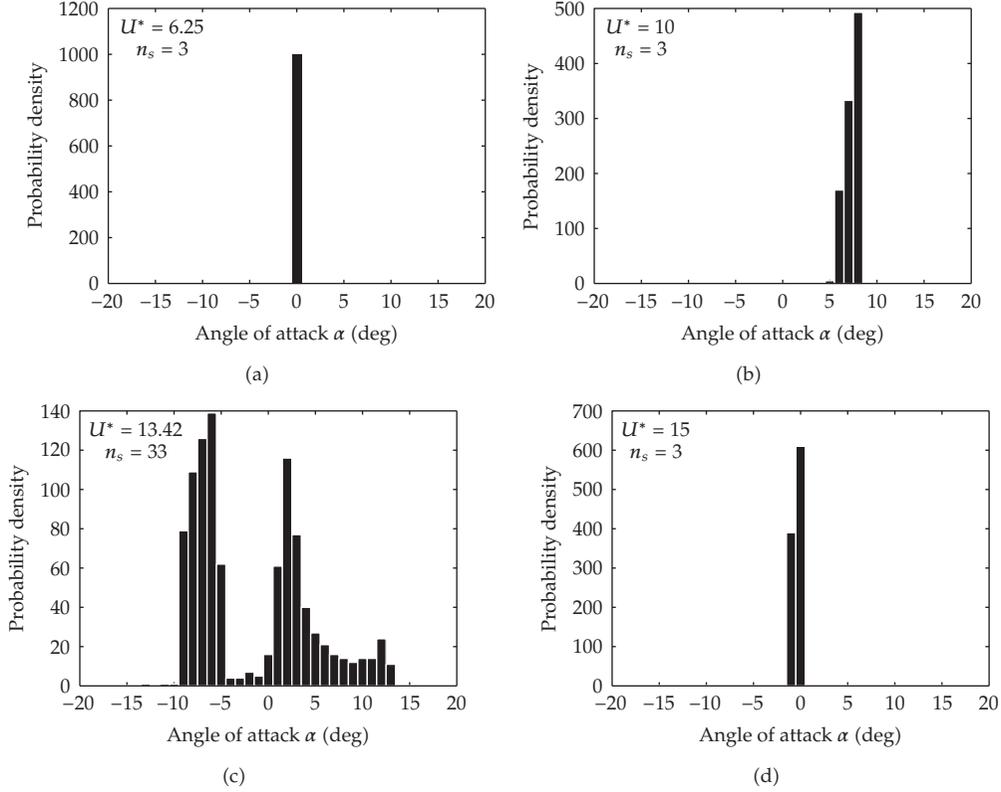
Figure 13: Bifurcation of the probability density function of the pitch angle α due to the random nonlinearity parameter β_α.

at $U^* = 10$ and $U^* = 15$, respectively. For $U^* = 13.42$ the mean and standard deviation shows again a sudden irregular behavior with the standard deviation oscillating around $\sigma_\alpha = 7$ due to the discontinuity in the response surface of Figure 15(c). The PDF of α of Figure 16 shows also a multimodal character for $U^* = 13.42$ only.

The bifurcation of the maximum standard deviation in Figure 10(c) gives in the largest part of the bifurcation parameter domain a two times higher value of $\sigma_{\alpha_{\max}}$ than for random β_α with the identical input coefficient of variation of $\mathrm{cv}_{\beta_\alpha} = 4.48\%$. The system is, therefore, twice as sensitive to randomness in α_0 than to random β_α. The random initial condition results actually in a variation of the initial phase of the realizations. Phase differences in the response have a large effect on the stochastic behavior as we have observed for random $\overline{\omega}$. However, the phase differences do not increase in time for random α_0 in contrast to the case with randomness in the ratio of natural frequencies. The gradually increasing $\sigma_{\alpha_{\max}}$ due to random $\overline{\omega}$ is, therefore, in the majority of the bifurcation parameter range larger than the effect of random α_0. This effect is not only caused by the larger input coefficient of variation for $\overline{\omega}$ but also mainly by the increasing phase differences in time.

However, in the second deterministic bifurcation point $U^* = 13.42$ the maximum standard deviation peaks for random α_0 at a higher value of $\sigma_{\alpha_{\max}} = 10.0$ compared to random $\overline{\omega}$. At $U^* = 13.42$ the effect of randomness in the parameters β_α and α_0 on $\sigma_{\alpha_{\max}}$ is, therefore, larger than that of random $\overline{\omega}$, while in the rest of the bifurcation domain the parameters β_α, α_0,
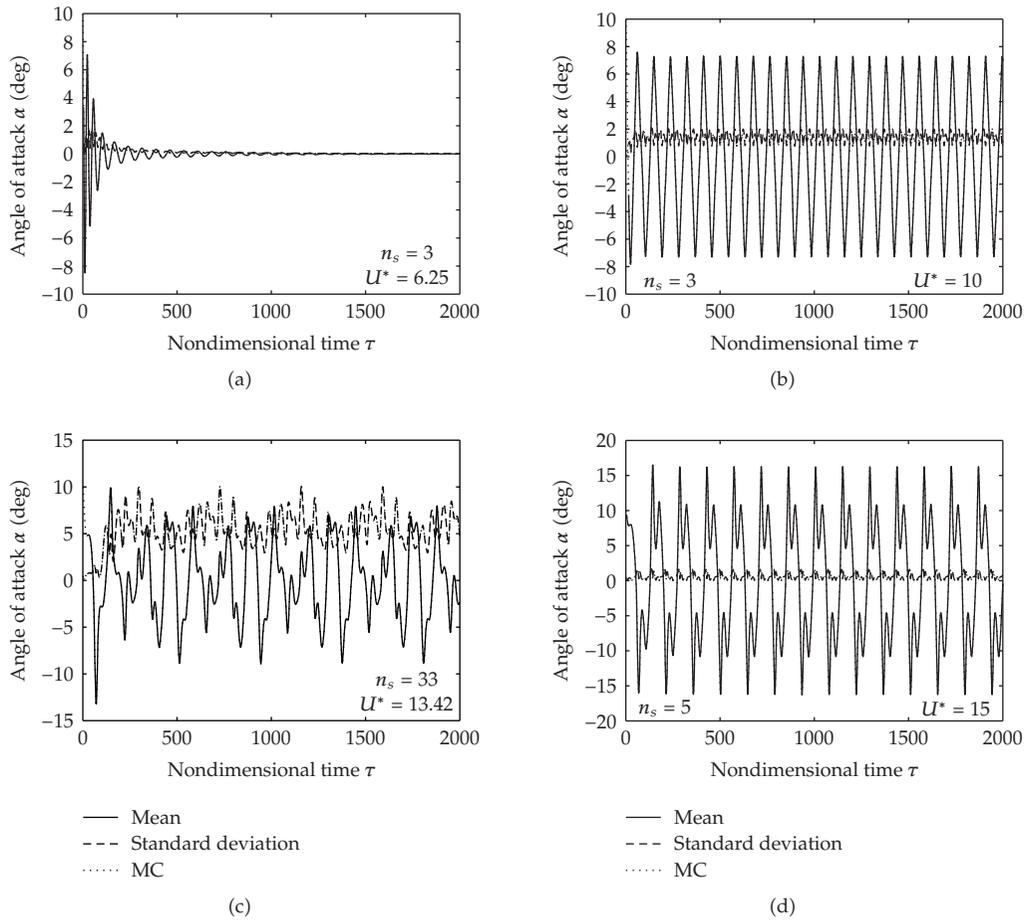
Figure 14: Time histories of the mean $\mu_\alpha(\tau)$ and standard deviation $\sigma_\alpha(\tau)$ of the pitch angle α due to the random initial condition α_0.

and $\overline{\omega}$, respectively, have a clear hierarchy of increasing importance. Even while $\overline{\omega}$ has twice the coefficient of variation compared to the other parameters, at the second deterministic bifurcation point the singularity results in larger variation in the response surface for β_α and α_0.

In deterministically already highly computationally intensive problems subject to a large number of random input parameters, the actual uncertainty analysis is usually performed for a subset of the most important random input parameters only, which is selected based on preliminary results for a limited number of parameter settings. The current results should warn the reader that this can be a dangerously unreliable approach, since the importance of the random input parameters can highly depend on the chosen bifurcation parameter value. In isolated points in parameter space a clear relative importance of the random parameters can even suddenly reverse, such that none of the parameters can be disregarded in advance. A multidimensional treatment of the combined effect of multiple random parameters in stochastic aeroelastic applications will, therefore, be considered in future work.

(a)

(b)

(c)

(d)
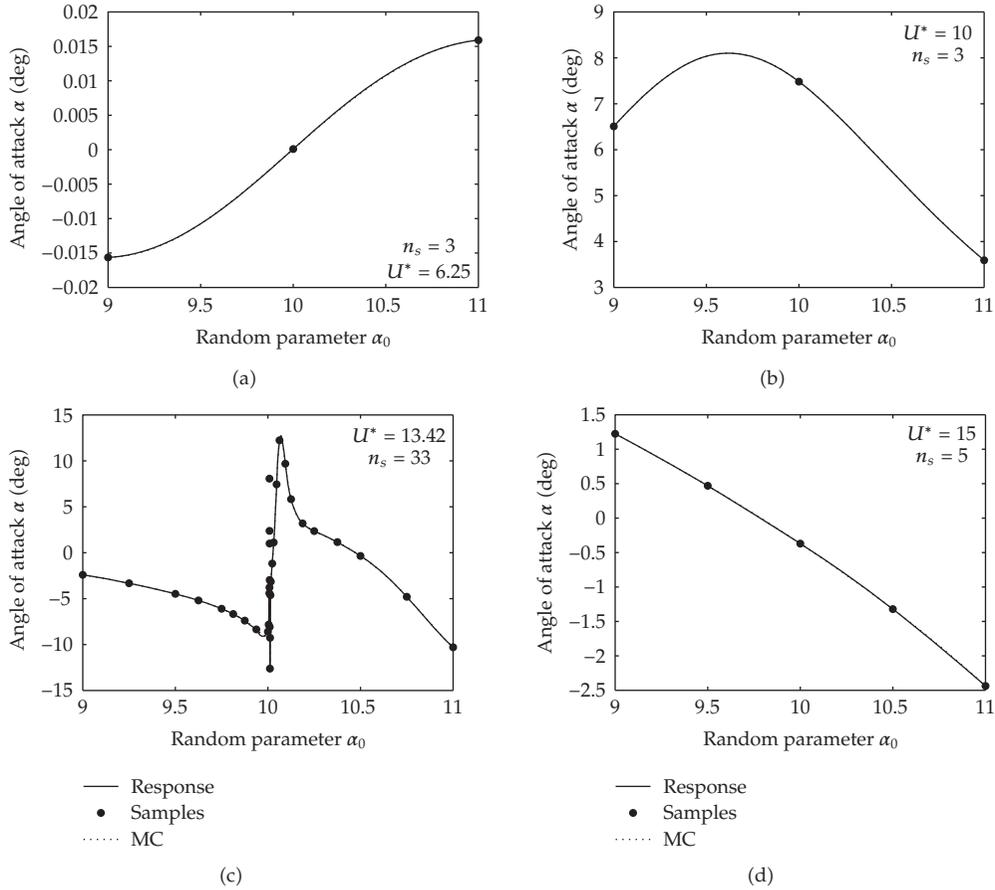
Figure 15: Response surface of the pitch angle α at $\tau = 2000$ as function of the random initial condition α_0.

8. Conclusions

The higher period stochastic bifurcation of a nonlinear airfoil flutter model is studied numerically. The fluid-structure interaction model consists of a two-degree-of-freedom rigid airfoil with cubic nonlinear springs and an aerodynamic model to determine the fluid loads in pitch and plunge. The employed uncertainty quantification method for unsteady problems is robust and efficient due to the extrema diminishing interpolation of oscillatory samples at constant phase.

The effect on the time history of the pitch angle α is considered for randomness in the ratio of natural pitch and plunge frequencies $\overline{\omega}$, a nonlinear spring parameter β_α, and the initial condition of the pitch angle α_0. The random natural frequency ratio $\overline{\omega}$ affects the frequency of the response, which results in a gradual increase of the maximum standard deviation of the pitch angle in the asymptotic range to $\sigma_{\alpha_{\max}} = 8.0°$ in the second deterministic bifurcation point of $U^* = 13.42$. The output variability also starts to increase from the trivial solution at an earlier position compared to the first deterministic bifurcation point $U^* = 6.25$.

The effect of uncertainty in the nonlinear stiffness parameter β_α is approximately constant beyond $U^* = 6.25$ at $\sigma_{\alpha_{\max}} = 1$ due to its effect on the limit cycle oscillation amplitude.

(a)

(b)

(c)

(d)
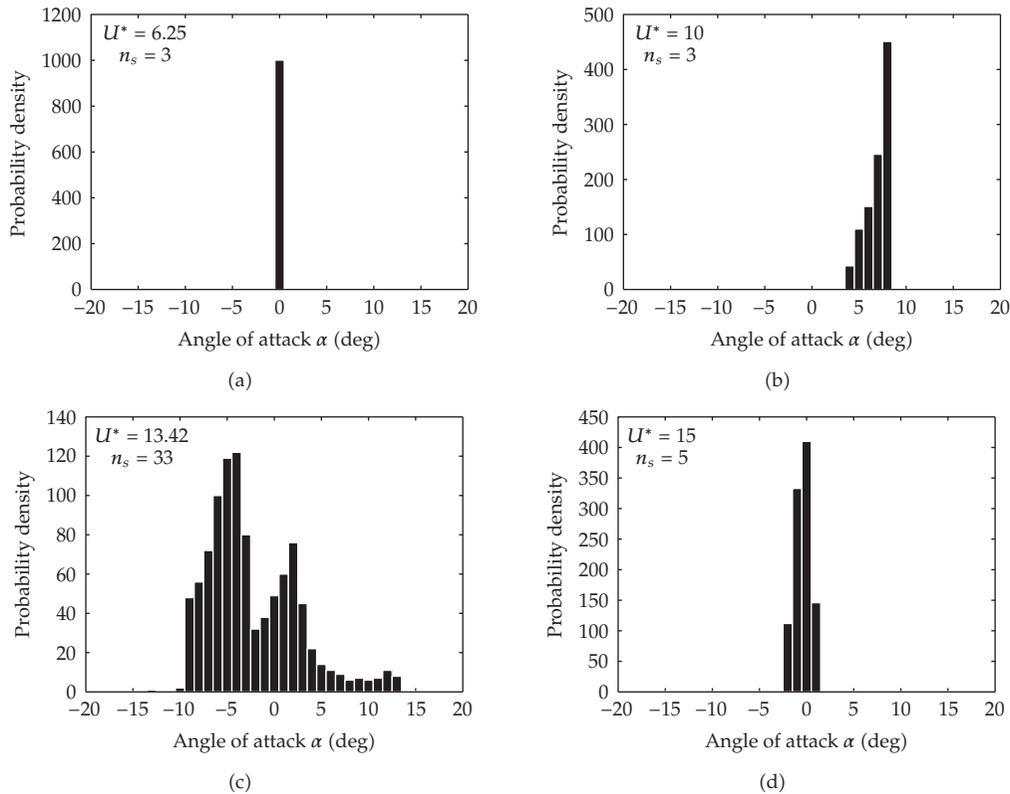
Figure 16: Bifurcation of the probability density function of the pitch angle α due to the random initial condition α_0.

At the second deterministic bifurcation point $U^* = 13.42$ random β_α results in a sudden peak in the output standard deviation of $\sigma_{\alpha_{max}} = 10.3$. The random initial condition α_0 reaches its maximum output standard deviation of $\sigma_{\alpha_{max}} = 10.0$ also in the second bifurcation point. In the rest of the bifurcation domain α_0 results approximately in a two times higher output randomness than β_α due to its effect on the phase of the response.

Despite the largest effect of $\overline{\omega}$ in the majority of the bifurcation domain, β_α and α_0 are the most important sources of randomness at the second deterministic bifurcation point. This is caused by the larger variance in the response surface for β_α and α_0 due to the singularity at $U^* = 13.42$. Reducing the number of random input parameters based on preliminary results for a limited number of parameter settings can, therefore, give unreliable results, since the order of relative parameter importance can reverse in isolated singular points in parameter space.

Acknowledgments

References

[1] R. H. Kraichnan, "Direct-interaction approximation for a system of several interacting simple shear waves," *Physics of Fluids*, vol. 6, no. 11, pp. 1603–1609, 1963.

[2] S. A. Orszag and L. R. Bissonnette, "Dynamical properties of truncated Wiener-Hermite expansions," *Physics of Fluids*, vol. 10, no. 12, pp. 2603–2613, 1967.

[3] R. Rubinstein and M. Choudhari, "Uncertainty quantification for systems with random initial conditions using Wiener-Hermite expansions," *Studies in Applied Mathematics*, vol. 114, no. 2, pp. 167–188, 2005.

[4] E. Lorenz, "Deterministic nonperiodic flow," *Journal of the Atmospheric Sciences*, vol. 20, pp. 130–141, 1963.

[5] C. Y. Shen, T. E. Evans, and S. Finette, "Polynomial-chaos applied to Lorenz's model for quantification of growth of initial uncertainties," in *Proceedings of the 5th European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS '08)*, Venice, Italy, June 2008.

[6] R. Pulch, "Polynomial chaos for linear DAEs with random parameters," preprint BUWAMNA 08/01, Bergische Universität Wuppertal, Wuppertal, Germany, 2008, http://www.math.uni-wuppertal.de/org/Num/Files/amna_08_01.pdf.

[7] E. F. Sheta, V. J. Harrand, D. E. Thompson, and T. W. Strganac, "Computational and experimental investigation of limit cycle oscillations of nonlinear aeroelastic systems," *Journal of Aircraft*, vol. 39, no. 1, pp. 133–141, 2002.

[8] D. M. Tang and E. H. Dowell, "Comparison of theory and experiment for nonlinear flutter and stall response of a helicopter blade," *Journal of Sound and Vibration*, vol. 165, no. 2, pp. 251–276, 1993.

[9] B. H. K. Lee, S. J. Price, and Y. S. Wong, "Nonlinear aeroelastic analysis of airfoils: bifurcation and chaos," *Progress in Aerospace Sciences*, vol. 35, no. 3, pp. 205–334, 1999.

[10] M. Kleiber and T. D. Hien, *The Stochastic Finite Element Method. Basic Perturbation Technique and Computer Implementation*, John Wiley & Sons, Chichester, UK, 1992.

[11] R. Lind and M. J. Brenner, "Robust flutter margin analysis that incorporates flight data," Tech. Rep. NASA/TP-1998-206543, NASA, Moffett Field, Calif, USA, 1998, http://www.nasa.gov/centers/dryden/pdf/88570main_H-2209.pdf.

[12] C. L. Pettit, "Uncertainty in aeroelasticity analysis, design and testing," in *Engineering Design Reliability Handbook*, E. Nicolaidis and D. Ghiocel, Eds., CRC Press, Boca Raton, Fla, USA, 2004.

[13] D. G. Liaw and H. T. Y. Yang, "Reliability and nonlinear supersonic flutter of uncertain laminated plates," *AIAA Journal*, vol. 31, no. 12, pp. 2304–2311, 1993.

[14] A. Carcaterra, D. Dessi, and F. Mastroddi, "Hydrofoil vibration induced by a random flow: a stochastic perturbation approach," *Journal of Sound and Vibration*, vol. 283, no. 1-2, pp. 401–432, 2005.

[15] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, Methuens Monographs on Applied Probability and Statistics, Methuen, London, UK, 1964.

[16] N. J. Lindsley, P. S. Beran, and C. L. Pettit, "Effects of uncertainty on nonlinear plate aeroelastic response," in *Proceedings of the 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Denver, Colo, USA, April 2002, paper no. AIAA 2002-1271.

[17] N. J. Lindsley, C. L. Pettit, and P. S. Beran, "Nonlinear plate aeroelastic response with uncertain stiffness and boundary conditions," *Structure & Infrastructure Engineering*, vol. 2, no. 3-4, pp. 201–220, 2006.

[18] D. Poirel and S. J. Price, "Random binary (coalescence) flutter of a two-dimensional linear airfoil," *Journal of Fluids and Structures*, vol. 18, no. 1, pp. 23–42, 2003.

[19] P. S. Beran, C. L. Pettit, and D. R. Millman, "Uncertainty quantification of limit-cycle oscillations," *Journal of Computational Physics*, vol. 217, no. 1, pp. 217–247, 2006.

[20] F. Poirion, "On some stochastic methods applied to aeroservoelasticity," *Aerospace Science and Technology*, vol. 4, no. 3, pp. 201–214, 2000.

[21] S. Choi and N. S. Namachchivaya, "Stochastic dynamics of a nonlinear aeroelastic system," *AIAA Journal*, vol. 44, no. 9, pp. 1921–1931, 2006.

[22] S. De Rosa and F. Franco, "Exact and numerical responses of a plate under a turbulent boundary layer excitation," *Journal of Fluids and Structures*, vol. 24, no. 2, pp. 212–230, 2008.

[23] A. Sarkar and R. Ghanem, "Mid-frequency structural dynamics with parameter uncertainty," *Computer Methods in Applied Mechanics and Engineering*, vol. 191, no. 47-48, pp. 5499–5513, 2002.

[24] I. Babuška, R. Tempone, and G. E. Zouraris, "Galerkin finite element approximations of stochastic elliptic partial differential equations," *SIAM Journal on Numerical Analysis*, vol. 42, no. 2, pp. 800–825, 2004.

[25] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer, New York, NY, USA, 1991.

[26] J. A. S. Witteveen and H. Bijl, "A monomial chaos approach for efficient uncertainty quantification in nonlinear problems," *SIAM Journal on Scientific Computing*, vol. 30, no. 3, pp. 1296–1317, 2008.

[27] J. A. S. Witteveen and H. Bijl, "Efficient quantification of the effect of uncertainties in advection-diffusion problems using polynomial chaos," *Numerical Heat Transfer, Part B*, vol. 53, no. 5, pp. 437–465, 2008.

[28] D. Xiu and G. E. Karniadakis, "The Wiener-Askey polynomial chaos for stochastic differential equations," *SIAM Journal on Scientific Computing*, vol. 24, no. 2, pp. 619–644, 2002.

[29] I. Babuška, F. Nobile, and R. Tempone, "A stochastic collocation method for elliptic partial differential equations with random input data," *SIAM Journal on Numerical Analysis*, vol. 45, no. 3, pp. 1005–1034, 2007.

[30] S. Hosder, R. W. Walters, and R. Perez, "A non-intrusive polynomial chaos method for uncertainty propagation in CFD simulations," in *Proceedings of the 44th AIAA Aerospace Sciences Meeting*, pp. 10649–10667, Reno, Nev, USA, January 2006, paper no. AIAA-2006-891.

[31] G. J. A. Loeven and H. Bijl, "Probabilistic collocation used in a two-step approach for efficient uncertainty quantification in computational fluid dynamics," *Computer Modeling in Engineering & Sciences*, vol. 36, no. 3, pp. 193–212, 2008.

[32] L. Mathelin, M. Y. Hussaini, and T. A. Zang, "Stochastic approaches to uncertainty quantification in CFD simulations," *Numerical Algorithms*, vol. 38, no. 1–3, pp. 209–236, 2005.

[33] M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio, "Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection," *Combustion and Flame*, vol. 132, no. 3, pp. 545–555, 2003.

[34] C. L. Pettit and P. S. Beran, "Effects of parametric uncertainty on airfoil limit cycle oscillation," *Journal of Aircraft*, vol. 40, no. 5, pp. 1217–1229, 2004.

[35] C. L. Pettit and P. S. Beran, "Spectral and multiresolution Wiener expansions of oscillatory stochastic processes," *Journal of Sound and Vibration*, vol. 294, no. 4, pp. 752–779, 2006.

[36] O. P. Le Maître, O. M. Knio, H. N. Najm, and R. G. Ghanem, "Uncertainty propagation using Wiener-Haar expansions," *Journal of Computational Physics*, vol. 197, no. 1, pp. 28–57, 2004.

[37] X. Wan and G. E. Karniadakis, "Long-term behavior of polynomial chaos in stochastic flow simulations," *Computer Methods in Applied Mechanics and Engineering*, vol. 195, no. 41–43, pp. 5582–5596, 2006.

[38] D. R. Millman, P. I. King, and P. Beran, "Airfoil pitch-and-plunge bifurcation behavior with Fourier chaos expansions," *Journal of Aircraft*, vol. 42, no. 2, pp. 376–384, 2005.

[39] S. Sarkar, J. A. S. Witteveen, A. Loeven, and H. Bijl, "Effect of uncertainty on the bifurcation behavior of pitching airfoil stall flutter," *Journal of Fluids and Structures*, vol. 25, no. 2, pp. 304–320, 2009.

[40] J. A. S. Witteveen, S. Sarkar, and H. Bijl, "Modeling physical uncertainties in dynamic stall induced fluid-structure interaction of turbine blades using arbitrary polynomial chaos," *Computers and Structures*, vol. 85, no. 11–14, pp. 866–878, 2007.

[41] J. A. S. Witteveen, A. Loeven, S. Sarkar, and H. Bijl, "Probabilistic collocation for period-1 limit cycle oscillations," *Journal of Sound and Vibration*, vol. 311, no. 1-2, pp. 421–439, 2008.

[42] J. A. S. Witteveen and H. Bijl, "An unsteady adaptive stochastic finite elements formulation for rigid-body fluid-structure interaction," *Computers and Structures*, vol. 86, no. 23-24, pp. 2123–2140, 2008.

[43] J. A. S. Witteveen and H. Bijl, "An alternative unsteady adaptive stochastic finite elements formulation based on interpolation at constant phase," *Computer Methods in Applied Mechanics and Engineering*, vol. 198, no. 3-4, pp. 578–591, 2008.

[44] J. A. S. Witteveen and H. Bijl, "A TVD uncertainty quantification method with bounded error applied to transonic airfoil flutter," *Communications in Computational Physics*, vol. 6, pp. 406–432, 2009.

[45] J. A. S. Witteveen and H. Bijl, "Effect of randomness on multi-frequency aeroelastic responses resolved by unsteady adaptive stochastic finite elements".

[46] R. E. Melchers, *Structural Reliability: Analysis and Prediction*, Ellis Horwood Series in Civil Engineering, Ellis Horwood, Chichester, UK, 1987.

[47] J. A. S. Witteveen, A. Loeven, and H. Bijl, "An adaptive stochastic finite elements approach based on Newton-Cotes quadrature in simplex elements," *Computers and Fluids*, vol. 38, no. 6, pp. 1270–1288, 2009.

[48] B. H. K. Lee, L. Y. Jiang, and Y. S. Wong, "Flutter of an airfoil with a cubic nonlinear restoring force," in *Proceedings of the 39th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit and AIAA/ASME/AHS Adaptive Structures Forum*, pp. 237–257, Long Beach, Calif, USA, April 1998.

[49] Y. Fung, *An Introduction to Aeroelasticity*, Dover, New York, NY, USA, 1969.

[50] R. T. Jones, "The unsteady lift of a wing of finite aspect ratio," NACA Report 681, National Advisory Committee for Aeronautics, Langley, Va, USA, 1940.

[51] B. H. K. Lee and P. LeBlanc, "Flutter analysis of a two-dimensional airfoil with cubic nonlinear restoring force," Tech. Rep. NAE-AN-36 NRC-25438, National Research Council of Canada, Ottawa, Canada, 1986.

Research Article

The Effect of Spatial Scale on Predicting Time Series: A Study on Epidemiological System Identification

L. H. A. Monteiro,[1,2] D. N. Oliveira,[1] and J. G. Chaui-Berlinck[3]

[1] *Pós-graduação em Engenharia Elétrica, Escola de Engenharia, Universidade Presbiteriana Mackenzie, Rua da Consolação, n.896, 01302-907 São Paulo, SP, Brazil*

[2] *Departamento de Engenharia de Telecomunicações e Controle, Escola Politécnica, Universidade de São Paulo, Avendia Professor Luciano Gualberto, travessa 3, n.380, 05508-900 São Paulo, SP, Brazil*

[3] *Departamento de Fisiologia, Instituto de Biociências, Universidade de São Paulo, Rua do Matão, travessa 14, n.321, 05508-900 São Paulo, SP, Brazil*

Correspondence should be addressed to L. H. A. Monteiro, luizm@mackenzie.br

A susceptible-infective-recovered (SIR) epidemiological model based on probabilistic cellular automaton (PCA) is employed for simulating the temporal evolution of the registered cases of chickenpox in Arizona, USA, between 1994 and 2004. At each time step, every individual is in one of the states S, I, or R. The parameters of this model are the probabilities of each individual (each cell forming the PCA lattice) passing from a state to another state. Here, the values of these probabilities are identified by using a genetic algorithm. If nonrealistic values are allowed to the parameters, the predictions present better agreement with the historical series than if they are forced to present realistic values. A discussion about how the size of the PCA lattice affects the quality of the model predictions is presented.

1. Introduction

In control engineering, it is fundamental to identify the system to be controlled in order to determine the best strategy for achieving the intended goals. The task of identifying a system is to determine the best model able of describing the dynamical behavior of such a system from measured data (e.g., [1]). Mathematical models in terms of differential equations for mechanical, electrical, thermal, and chemical systems are commonly obtained from well-known physical laws or conservation principles. For instance, the model of a mechanical system can be derived from the Newton's laws of motion; the model of an electrical system can be deduced from the Maxwell's equations of electromagnetism; models for chemical and thermal systems usually obey principles of energy and mass conservation. Biological systems,

however, do not present a clear starting point for obtaining the corresponding mathematical models. In fact, models for such systems are usually empirical, and the identification and the controller synthesis are not usually simple. Moreover, scaling problems pose additional difficulties to accomplish these tasks.

Scaling problems are ubiquitous in nature and prominent in biology. However, such problems are not easily treated, or even recognized, in a plethora of cases. As put by Haldane in the beginning of the 20th century, "The most obvious differences between different animals are differences in size, but for some reason the zoologists have paid singularly little attention to them" (cited in [2]). Interestingly, the same kind of "overlooking" attention is found in many ecological and epidemiological contexts. The question that arises is whether we are prone to obtain reliable identification of systems if this bias is maintained.

There are several approaches to scaling problems, from dimensional analysis to fractal dimensions (e.g., [3, 4]). These approaches are extremely relevant in engineering problems. For instance, the benefits of dimensional analysis and similitude principles both in designing prototypes and in modeling general principles within given conditions are above argument (e.g., [5]). However, ecological/epidemiological problems have a subtle "dimensional architecture." If system identification can be obtained in these problems, then one of the basic goals of scaling approaches would be fulfilled, namely, "the design of (relevant) experiments and the proper organization of the results" ([5, page 227]; we put the brackets in that phrase). In other words, there should be expected an important gain both in basic sciences and in economical aspects.

Here, we report a numerical study revealing the influence of scale on parametric identification of an epidemiological model used for predicting the temporal evolution of the number of chickenpox cases registered by the Arizona Department of Health Services [6], USA, between 1994 and 2004. The model is based on probabilistic cellular automaton (PCA), where each cell of the lattice corresponds to an individual, which is in one of three states: S, I, or R. The state S represents the individual that is susceptible and therefore subjected to this contagious disease; the state I is related to the individual that is infective and hence can transmit the disease for susceptible ones; the state R is associated to the recovered individual. The numerical values of the probabilities concerning the transitions among these three states define this SIR model.

In Section 2, the epidemiological model is introduced. In Section 3, the genetic algorithm (GA) employed for identifying the probabilities of state transitions in this SIR model is described. In Section 4, the numerical results are shown. In Section 5, a discussion about scaling problems in ecological and epidemiological systems is presented.

2. SIR Model Based on PCA

In our cellular automaton (CA) model [7–9], the population lives in a square matrix formed by $n \times n = n^2$ cells (individuals) with periodic boundary conditions. At each time step t, there is a probability $P_i(v)$ of a S-cell being infected, where P_i depends on the number v of infective neighbors and it obeys the constraints: $P_i(0) = 0$ (i.e., a susceptible individual can contract the disease only if $v > 0$) and if $v_1 < v_2$, then $P_i(v_1) < P_i(v_2)$ (i.e., $P_i(v)$ is a monotone increasing function of v). Each I-cell has probability P_c per time step of becoming cured and probability P_d per time step of dying because of the disease. At each iteration, infective and recovered cells may die for other causes with probability P_n. When individuals die, susceptible ones replace them. Therefore, the total number of individuals n^2 remains

constant, a common assumption in epidemiological models (e.g., [10]). The states of all cells are simultaneously updated at each time step t.

The coupling topology influences the dynamical behavior of biological, electronic, or social networks (e.g., [9, 11, 12]). In studies on spreading diseases using CA, the contact network among individuals (which defines the neighborhood of each cell in the lattice) can be considered as regular in a first approximation. Hence, the von Neumann (e.g., [13]) and Moore (e.g., [7, 8, 14]) neighborhoods are commonly employed. The von Neumann neighborhood of radius r of a cell is formed by the cells orthogonally surrounding such a cell until the distance r (i.e., if $r = 1$, the von Neumann neighborhood comprises the 4 closest neighbors: left, right, up, and down); the Moore neighborhood of radius r of a cell is constituted by all cells pertaining to the square matrix of size $2r+1$ centered in this cell (i.e., if $r = 2$, the Moore neighborhood consists of the 24 cells pertaining to the square matrix of size 5 centered in such a cell).

With this model, we intend to fit the data shown in Table 1 by finding appropriate values for $P_i(v)$, P_c, P_d, and P_n able of reproducing the number of chickenpox cases recorded in each year, in the State of Arizona, USA, during 11 years. The values of these probabilities are found by using a genetic algorithm (GA).

3. GA Used for Identifying the SIR Model

There have been published several works on identification of transition rules of CA by using GA (e.g., [15–19]).

Here each generation of our GA is composed by 15 chromosomes (15-candidate solutions), where the length of the chromosome depends on the kind and on the radius of the neighborhood. For instance, when the von Neumann neighborhood of $r = 1$ is employed, each chromosome is formed by 7 genes (the values of $P_i(1)$, $P_i(2)$, $P_i(3)$, $P_i(4)$, P_c, P_d, P_n), when the Moore neighborhood of $r = 1$ is used, each one consists of 11 genes (the values of $P_i(1)$, $P_i(2)$, ..., $P_i(8)$, P_c, P_d, P_n). Here, we investigate both neighborhoods for $r = 1$ and 2.

The fitness F of every chromosome is evaluated by simulations with the PCA model. The value of F is high if the numbers of infective individuals in the PCA lattice are close to the data composing the historical time series. The fitness function is written as $F = 1/(F_1 F_2)$, where F_1 is defined here by

$$F_1 = \sum_{q=1}^{N} \sum_{j=2}^{m} |x_j - x_{j,q}^{\text{PCA}}|, \tag{3.1}$$

and F_2 by

$$F_2 = \sum_{q=1}^{N} \sum_{j=2}^{m} \frac{\left|\sigma(x_j - x_{j-1}) - \sigma(x_{j,q}^{\text{PCA}} - x_{j-1,q}^{\text{PCA}})\right|}{2}, \tag{3.2}$$

where σ is a signal function: $\sigma(y) = -1$ if $y = -1$ and $\sigma(y) = +1$; otherwise, m is the total number of years composing the time series (here $m = 11$ according to Table 1); j labels each year of the epidemiological series. Thus, $j = 2$ corresponds to the second year (1995), and $j = 1$ is the initial year (corresponding to 1994; in fact, the number of infective cases in this year is

Table 1: Registered cases of chickenpox in Arizona between 1994 and 2004.

Year	Number of cases (scale 1:1)	Number of cases (scale 1:20)
1994	6783	339
1995	2658	133
1996	3319	166
1997	1987	99
1998	1673	83
1999	960	48
2000	1522	76
2001	951	47
2002	606	30
2003	1620	81
2004	1091	54

the initial condition of the PCA simulations). x_j is the number of historical cases registered in the jth year; $x_{j,q}^{\text{PCA}}$ is the number of infective individuals obtained in the PCA model in the jth year in the qth simulation. Every chromosome is evaluated N times (here $N = 4$) in lattices $n \times n$, where the numerical values of the initial conditions are always the same, but the corresponding geographical distributions of susceptible, infective, and recovered individuals may be altered from one simulation to another.

The function F_1 is the sum of the modules of the differences between the data and the number of cases obtained in N simulations by using the probabilities corresponding to that chromosome; thus, F_1 gives the total error between the historical series and the series produced by simulating N times the disease spreading in the lattice. The higher F_1, the lower F.

The function F_2 is the number of segments of the time series, where the registered cases and the predicted values present contrary tendencies of evolution in N simulations. For instance, if in two consecutive years the numbers in Table 1 decreased but the numbers of infective cases in the qth simulation increased (i.e., if both curves have slopes with different signals), then $|\sigma(x_j - x_{j-1}) - \sigma(x_{j,q}^{\text{PCA}} - x_{j-1,q}^{\text{PCA}})|/2 = 1$; if they have the same tendency of variation, this term is zero. The higher F_2, the lower F.

An initial population of chromosomes is randomly generated, and the value g of their genes must always be real numbers between zero and one, that is, $g \in [0, 1]$, because they represent probabilities. Also the genes representing $P_i(v)$ must obey the constraint: $P_i(v_1) < P_i(v_2)$ for $v_1 < v_2$. The best seven chromosomes are carried over to the next GA generation without alteration. The other eight chromosomes (in order to complete 15) are obtained from the current generation by applying crossover and mutation operations.

Crossover is performed by selecting a single point in two randomly picked chromosomes (the parents) and swapping $1/3$ of the genes between them. Thus, two new chromosomes (the children) are created. The number of child chromosomes produced by crossover is seven.

Every parent chromosome has 50% of chance of suffering mutation and producing a child chromosome. Every gene composing such a chromosome has 50% of chance of changing its value. If a gene is selected to suffer mutation, then a number between -0.3 and $+0.3$ is added to the value g of this gene (if after this genetic operation $g > 1$, then we impose $g = 1$; if $g < 0$, then we impose $g = 0$ in order to get $g \in [0, 1]$).

After applying these genetic operations for producing new chromosomes, the values of P_i in such chromosomes are reordered (if necessary) because $P_i(v)$ must be a monotone increasing function of v.

To choose eight chromosomes for the next GA generation, the pool of parents and children is organized in a crescent order in terms of F (excluding the seven fitter candidate solutions). Then, eight chromosomes are selected by applying eight times the expression:

$$p = \rho(l\sqrt{r}), \tag{3.3}$$

where p is the position of the chromosome in such an ordering that will be chosen for composing the next generation; $\rho(y)$ is the highest integer number lesser than or equal to y; l is the length of this ordering (the total number of candidate solutions excluding those seven elite chromosomes); $r \in [0, 1]$ is a random number. The idea of using $\sqrt{r}$ instead of r is to favor the selection of chromosomes in the end of this ordering (with higher value of F). The total number of generations of this GA is 25.

Notice that the value of F_2 does not depend on the lattice size n; it only depends on the number of data of the historical series and on the number of simulations. Thus, $0 \le F_2 \le (m - 1)N$. However, F_1 is dependent on n. In fact, $0 \le F_1 \le (m - 1)n^2N$. In order to compare the quality of predictions obtained in PCA lattices with different sizes, we propose a metric E called efficiency written as $E = (E_1 + E_2)/2$. The value of E_2 is obtained from F_2 by $E_2 = 1 - [F_2/(N(m - 1))]$, and the value of E_1 is obtained from F_1 by

$$E_1 = \frac{1}{2} + \frac{N\overline{F}_1 - F_1}{2N\overline{F}_1}, \tag{3.4}$$

where $\overline{F}_1$ is given by

$$\overline{F}_1 = \sum_{j=2}^{m} |x_j - \overline{x}|, \tag{3.5}$$

and $\overline{x}$ is the average value of the historical series excluding the first datum; thus,

$$\overline{x} = \frac{1}{(m - 1)} \sum_{j=2}^{m} x_j. \tag{3.6}$$

Notice that if $F_1 = 0$ (i.e., the numbers of infective individuals generated in the PCA lattice of size $n \times n$ in all N simulations are equal to the numbers of registered cases converted to this spatial scale), then $E_1 = 100\%$; if $F_1 = N\overline{F}_1$ (i.e., the predictions in N simulations give a value of F_1 corresponding to the average deviation $\overline{F}_1$ presented by the historical series), then $E_1 = 50\%$.

Because $0 \le E_2 \le 100\%$ and usually $0 \le E_1 \le 100\%$, then $0 \le E \le 100\%$. The efficiency E is calculated only for the fitter chromosome of the last GA generation. In the first phrase of this paragraph, "usually" means that, at the end of the evolutionary process, the value of F_1 corresponding to the fitter chromosome commonly obeys the constraint $0 \le F_1 \le 2N\overline{F}_1$
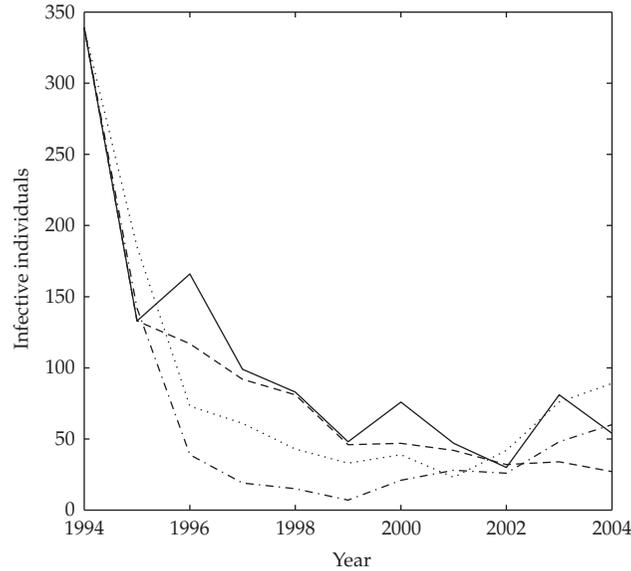
Figure 1: Historical series (solid line) and simulations in the spatial scale 1:20 with $E = 68.2\%$ (dashed line), $E = 50.2\%$ (dotted line), and $E = 36.4\%$ (dash-dotted line).

(hence $0 \leq E_1 \leq 100\%$). In fact, this always happened in our simulations. Obviously, the higher the value of E, the better the PCA predictions.

4. Results

About 5 million people live in Arizona. For a lattice of size $n = 250$, the cases of chickenpox must be multiplied by $n^2/(5 \times 10^6) = 1/20$ as shown in the third column of Table 1, in order to convert the historical data to this spatial scale.

Each time step of the PCA is equivalent to 2 months. Thus, to predict the number of infective individuals from one year to another, the PCA must be iterated by 6 time steps.

At first, we performed simulations considering von Neumann and Moore neighborhoods of $r = 1$ or 2, and 6 time steps per year. After 25 GA generations, the fitter chromosome was obtained with Moore neighborhood of $r = 1$, and it was formed by the following genes: P_i varies from 17.4% (for $v = 1$) to 75.5% (for $v = 8$), $P_c = 86.4\%$, $P_d = 4.4\%$, $P_n = 89.5\%$. The efficiency of this candidate solution was $E = 68.2\%$. Observe that the numbers obtained for P_d and P_n are far from being realistic because these values mean that more than 90% of infective and recovered individuals die and are replaced by susceptible ones at each two months. The value of P_n should be about 0.1% because about 0.1% of the population die at each two months [20]. The usual death rate of chickenpox is 10 per 100 000 cases [21]; hence P_d should be about 0.01%. The value of P_c found by the GA is small, since more than 99% of infected individuals become cured in 2 months. In spite of these values being unrealistic, this set of probabilities can reasonably approximate the historical series as shown in Figure 1 (see the solid and the dashed lines).

In order to obtain realistic values for the state transition probabilities, the limit $P_n \leq 1\%$ was imposed to the evolutionary algorithm. In this case, the fitter chromosome (after 25 GA
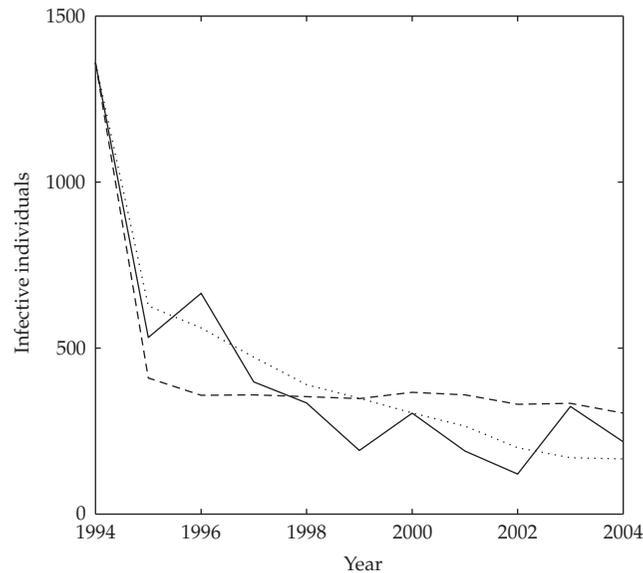
Figure 2: Historical series (solid line) and simulations in the spatial scale 1:5 with $E = 61.6\%$ (dashed line) and $E = 70.5\%$ (dotted line).

generations) was found with Moore neighborhood of $r = 2$, and its efficiency was $E = 50.2\%$, a worst result than the previous one. Figure 1 also exhibits a temporal evolution (see the dotted line) of the number of infective individuals simulated by employing the PCA model with the corresponding chromosome.

Then, more modifications on the GA were accomplished. Now each time step of the PCA can correspond to 1 or 2 months (thus, 12 or 6 time steps are necessary to complete 1 year, resp.). And additional limits were imposed to the following genes: $0.001\% \leq P_d, P_n \leq 1\%$, and $P_c \geq 70\%$. With such modifications, the efficiency of the better candidate solution decreases to $E = 36.4\%$, and it was obtained for von Neumann neighborhood of $r = 2$ and 6 times steps per year. The values of P_i vary from 45.2% to 98.9%, $P_c = 70.0\%$, $P_d = 0.25\%$, and $P_n = 1\%$. These values are still unrealistic. Figure 1 presents a temporal evolution of the PCA with these probabilities (see the dash-dotted line).

In the simulations reported above, $n = 250$ and consequently the spatial scale is 1:20 (one cell of the lattice corresponds to 20 individuals living in Arizona). For $n = 1000$, the scale is 1 : 5. By using this new spatial scale and the additional limits to the death and cure probabilities, the best solution, with $E = 61.6\%$, was found with von Neumann neighborhood of $r = 2$ and 12 time steps per year. The values of P_i vary from 1.0% to 88.7%, $P_c = 99.99\%$, $P_d = 0.001\%$, and $P_n = 0.001\%$. A temporal evolution of the infective cases in PCA is presented in Figure 2 (see the dashed line).

In the scale 1:5, if those additional limits on P_c, P_d, and P_n are removed, the best chromosome was found with Moore neighborhood of $r = 1$ and 6 time steps per year. It has the highest efficiency found in all numerical experiments, $E = 70.5\%$; however, the probabilities are again unrealistic. For instance, this chromosome presents $P_n = 100\%$ (an absurd value, of course). Figure 2 also illustrates the temporal evolution of the corresponding PCA. Observe the good agreement between the historical series and a PCA simulation (see the dotted line).

5. Discussion

We found that the size of the PCA lattice affects the quality of the predictions of our epidemiological model, after being identified by using a GA. The best set of state transition probabilities, with $E = 70.5\%$, was obtained in the scale 1:5. By imposing realistic constraints to the values of these probabilities, the efficiency of the best solution presented $E = 61.6\%$. In the scale 1:20, the best solution respecting such constraints dropped to $E = 36.4\%$.

Here, we had to deal with the problem of scaling the lattice of the CA in relation to the number of individuals, healthy and infective ones. The most obvious first approach is to maintain the ratios from the original data, that is, a linear relation of size. In this sense, space is completely embedded in the number of individuals, which, ultimately, becomes the size reference frame. One potential problem in doing this is that the neighborhood (either Moore or von Neumann) maintains, therefore, its absolute size. Let us exemplify this.

Epidemiological models based on differential equations can be used for describing the spreading of infectious diseases (e.g., [7–10]). These equations are a mean-field approximation for PCA if the three groups (S, I, and R) are homogeneously distributed in space. A parameter called basic reproduction number, R_0, is obtained from the infectiveness, mortality, and recovery rate constants (e.g., [7–10]). The value of R_0 identifies the outbreak or the extinction of the disease (usually $R_0 < 1$ represents extinction and $R_0 > 1$ represents outbreak, e.g., [7–10]). It was suggested that "to make the R_0's equivalent," the transmission probability of the local model should be the global (mean-field) probability times the ratio population size to neighborhood size ([22, page 116]). However, as it can be seen, our results indicate that such a linear proportionality may not hold, at least to estimate reliable parameters from real data.

Turner et al. [23] studied how spatial scale interferes with the identification of estimators in landscapes indexes as diversity and dominance. They concluded that both quantitative and qualitative changes in the measurements would be present depending on how the scale is defined. Later, in the same lines, Dungan et al. [24] called into attention that the various spatial scaling measures (as grain, resolution, range) are not interchangeable and suggested avoiding the term scale. As the present study showed, this shortcoming in space is readily present when size is an inherited variable of individuals.

DeAngelis and Petersen [25] focused in a predation model, where salmons migrate through predators and concluded that the size resolution of the cells has a crucial importance in the predicted outcome. The authors emphasize that fewer cells (therefore, a coarse spatial resolution) over predict the mortality rate. Moreover, irrespectively to the spatial resolution, the models give different predictions than the mean-field ones. On the other hand, Pascual et al. [26] concluded that despite variations in the predicted rates due to scaling problems, the functional forms are quite the same as the well-mixed (mean-field) models.

The discussion of system identification in ecological/epidemiological studies and scaling issues is still an open question. Our investigation goes in the same direction as many previous ones (see above), that is, that size matters. We add to the query the problem of interchanging space with individuals (and their neighborhoods). Therefore, if modeling is concerned with real predictions much more than simple data fitting, then a judicious evaluation of the underlying scale should be performed as a first step. It seems that these ecological/epidemiological problems need their particular definitions of size, whether this means space itself or number of individuals or a combination of both. In other words, the geometry for scaling these systems is still awaiting the due attention.

Acknowledgment

References

[1] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Upper Saddle River, NJ, USA, 1999.

[2] W. A. Calder III, "Scaling of physiological processes in homeothermic animals," *Annual Review of Physiology*, vol. 43, pp. 301–322, 1981.

[3] A. Bejan, *Shape and Structure: From Engineering to Nature*, Cambridge University Press, Cambridge, UK, 2000.

[4] G. I. Barenblatt, *Scaling*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, UK, 2003.

[5] A. Alexandrou, *Principles of Fluid Mechanics*, Prentice-Hall, Upper Saddle River, NJ, USA, 2001.

[6] Arizona Department of Health Services, Reported Cases of Notifiable Diseases by Year (1994–2004), October 2008, http://www.azdhs.gov/phs/oids/stats/pdf/t3_cases1994-2004.pdf.

[7] L. H. A. Monteiro, H. D. B. Chimara, and J. G. Chaui-Berlinck, "Big cities: shelters for contagious diseases," *Ecological Modelling*, vol. 197, no. 1-2, pp. 258–262, 2006.

[8] L. H. A. Monteiro, J. B. Sasso, and J. G. Chaui-Berlinck, "Continuous and discrete approaches to the epidemiology of viral spreading in populations taking into account the delay of incubation time," *Ecological Modelling*, vol. 201, no. 3-4, pp. 553–557, 2007.

[9] P. H. T. Schimit and L. H. A. Monteiro, "On the basic reproduction number and the topological properties of the contact network: an epidemiological study in mainly locally connected cellular automata," *Ecological Modelling*, vol. 220, no. 7, pp. 1034–1042, 2009.

[10] J. D. Murray, *Mathematical Biology. I. An Introduction*, vol. 17 of *Interdisciplinary Applied Mathematics*, Springer, New York, NY, USA, 3rd edition, 2002.

[11] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

[12] J. R. C. Piqueira, M. Q. Oliveira, and L. H. A. Monteiro, "Synchronous state in a fully connected phase-locked loop network," *Mathematical Problems in Engineering*, vol. 2006, Article ID 52356, 12 pages, 2006.

[13] S. Yakowitz, J. Gani, and R. Hayes, "Cellular automaton modeling of epidemics," *Applied Mathematics and Computation*, vol. 40, no. 1, pp. 41–54, 1990.

[14] M. A. Fuentes and M. N. Kuperman, "Cellular automata and epidemiological models with spatial dependence," *Physica A*, vol. 267, no. 3, pp. 471–486, 1999.

[15] M. Mitchell, P. T. Hraber, and J. P. Crutchfield, "Revisiting the edge of chaos: evolving cellular automata to perform computations," *Complex Systems*, vol. 7, no. 2, pp. 89–130, 1993.

[16] S. A. Billings and Y. Yang, "Identification of probabilistic cellular automata," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 33, no. 2, pp. 225–236, 2003.

[17] N. Ganguly, P. Maji, B. K. Sikdar, and P. P. Chaudhuri, "Design and characterization of cellular automata based associative memory for pattern recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 34, no. 1, pp. 672–679, 2004.

[18] P. P. B. de Oliveira, J. C. Bortot, and G. M. B. Oliveira, "The best currently known class of dynamically equivalent cellular automata rules for density classification," *Neurocomputing*, vol. 70, no. 1–3, pp. 35–43, 2006.

[19] S. Suzuki and T. Saito, "Synthesis of desired binary cellular automata through the genetic algorithm," in *Proceedings of the 13th International Conference on Neural Information Processing (ICONIP '06)*, vol. 4234 of *Lecture Notes in Computer Science*, pp. 738–745, Hong Kong, October 2006.

[20] Central Intelligence Agency, *The 2008 World Factbook*, U.S. Government Printing Office, Washington, DC, USA, 2008.

[21] H. Rawson, A. Crampin, and N. Noah, "Deaths from chickenpox in England and Wales 1995–7: analysis of routine mortality data," *British Medical Journal*, vol. 323, no. 7321, pp. 1091–1093, 2001.

[22] E. Holmes, "Basic epidemiological concepts in a spatial context," in *Spatial Ecology*, D. Tilman and P. Kareiva, Eds., pp. 111–136, Princeton University Press, Princeton, NJ, USA, 1997.

[23] M. G. Turner, R. V. O'Neill, R. H. Gardner, and B. T. Milne, "Effects of changing spatial scale on the analysis of landscape pattern," *Landscape Ecology*, vol. 3, no. 3-4, pp. 153–162, 1989.

[24] J. L. Dungan, J. N. Perry, M. R. T. Dale, et al., "A balanced view of scale in spatial statistical analysis," *Ecography*, vol. 25, no. 5, pp. 626–640, 2002.

[25] D. L. DeAngelis and J. H. Petersen, "Importance of the predator's ecological neighborhood in modeling predation on migrating prey," *Oikos*, vol. 94, no. 2, pp. 315–325, 2001.

[26] M. Pascual, P. Mazzega, and S. A. Levin, "Oscillatory dynamics and spatial scale: the role of noise and unresolved pattern," *Ecology*, vol. 82, no. 8, pp. 2357–2369, 2001.

Research Article

Macro- and Microsimulations for a Sublimation Growth of SiC Single Crystals

Jürgen Geiser[1] and Stephan Irle[2]

[1] *Humboldt University of Berlin, Unter den Linden 6, 10099 Berlin, Germany*
[2] *Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan*

Correspondence should be addressed to Jürgen Geiser, geiser@mathematik.hu-berlin.de

The numerous technical applications in electronic and optoelectronic devices, such as lasers, diodes, and sensors, demand high-quality silicon carbide (SiC) bulk single crystal for industrial applications. We consider an SiC crystal growth process by physical vapor transport (PVT), called modified Lely method. We deal with a model for the micro- and macroscales of the sublimation processes within the growth apparatus. The macroscopic model is based on the heat equation with heat sources due to induction heating and nonlocal interface conditions, representing the heat transfer by radiation. The microscopic model is based on the quantum interatomic potential and is computed with molecular dynamics. We study the temperature evolution in the apparatus and reflect the growth behavior of the microscopic model. We present results of some numerical simulations of the micro- and macromodels of our growth apparatus.

1. Introduction

The motivation for this study comes from the technical demand to simulate a crystal growth apparatus for SiC single crystals. The single crystals are used as a high-valued and expensive material for optoelectronics and electronics (cf. [1]). We concentrate on a deterministic model for simulating crystal growth; alternative models are discussed with comprehensive probabilistic modeling (see [2]).

The silicon carbide (SiC) bulk single crystals are produced by a growth process through physical vapor transport (PVT), called modified Lely method. The modeling for the thermal processes within the growth apparatus is done in [3, 4]. In this paper, we propose one step more in the modeling of the macroscopic and microscopic parts. The idea is to exchange results from the macroscopic to the microscopic scale to obtain a feedback to control the growth process of the SiC bulk. Here the benefits are an acceleration of solving interactive growth processes of the crystal with their underlying temperature in the apparatus. Using

only standard codes, which are decoupled, a simple parameter exchange of temperature and pressure in the deposition region cannot resolve the growth problem accurately. We propose a first framework of a combined model, which is based on the authors' knowledge of a novel work and a first approach to a coupled solver method.

2. Macroscopic Model: Heat-Flux

In the following, we discuss the macroscopic model, which is based on continuum equations for the heat-flux.

2.1. Mathematical Model

The underlying equations of the model are given as follows.

(a) In this work, we assume that the temperature evolution inside the gas region Ω_g can be approximated by considering the gas as pure argon (Ar). The reduced heat equation is

$$\rho_g \partial_t U_g - \nabla \cdot (\kappa_g \nabla T) = 0, \tag{2.1}$$

$$U_g = z_{Ar} R_{Ar} T, \tag{2.2}$$

where T is the temperature, t is the time, and U_g is the internal energy of the argon gas. The parameters are given as ρ_g being the density of the argon gas, κ_g being the thermal conductivity, z_{Ar} being the configuration number, and R_{Ar} being the gas constant for argon.

(b) The temperature evolution inside the region of solid materials Ω_s (e.g., inside the silicon carbide crystal, silicon carbide powder, graphite, and graphite insulation) is described by the heat equation

$$\rho_s \partial_t U_s - \nabla \cdot (\kappa_s \nabla T) = f, \tag{2.3}$$

$$U_s = \int_0^T c_s(S) \, dS, \tag{2.4}$$

where ρ_s is the density of the solid material, U_s is the internal energy, κ_s is the thermal conductivity, and c_s is the specific heat.

The equations hold in the domains of the respective materials and are coupled by interface conditions, for example, requiring the continuity for the temperature and for the normal components of the heat flux on the interfaces between opaque solid materials. On the boundary of the gas domain, that is, on the interface between the solid material and the gas domain, we consider the interface condition

$$\kappa_g \nabla T \cdot \mathbf{n}_g + R - J = \kappa_s \nabla T \cdot \mathbf{n}_g, \tag{2.5}$$

where $\mathbf{n}_g$ is the normal vector of the gas domain, R is the radiosity, and J is the irradiosity. The irradiosity is determined by integrating R along the whole boundary of the gas domain

(cf. [5]). Moreover, we have

$$R = E + J_{\text{ref}},$$

$$E = \sigma \epsilon T^4 \quad \text{(Stefan-Boltzmann equation)}, \qquad (2.6)$$

$$J_{\text{ref}} = (1 - \epsilon) J,$$

where E is the radiation, J_{ref} is the reflexed radiation, ϵ is the emissivity, and σ is the Boltzmann radiation constant.

The density of the heat source induced by the induction heating is determined by solving Maxwell's equations. We deal with these equations under the simplifying assumption of an axisymmetric geometry, axisymmetric electromagnetic fields, and a sinusoidal time dependence of the involved electromagnetic quantities, following [6]. The considered system and its derivation can be found in [3, 4, 7].

In this paper, we focus on the discretization and material properties, which are important for realistic simulations. Our underlying software tool WIAS-HiTNIHS (cf. [4]) allows us a flexibility in the grid generation and for the material parameters.

In the next section, we describe the used discretization.

2.2. Discretization

For the discretization of the heat equation (diffusion equation), we apply the implicit Euler method in time and the finite volume method for the space discretization (cf. [3, 4, 8]). We consider a partition $\tau = (\omega_i)_{i \in I}$ of Ω such that, for $m \in \{s, g\}$ (with s solid, g gas) and $i \in I$, $\omega_{m,i} := \omega_i \cap \Omega_m$ defines either a void subset or a nonvoid, connected, and open polyhedral subset of Ω. By integrating the corresponding heat equation (2.1) or (2.3) over $\omega_{m,i}$, we derive the following nonlinear equations for the temperature variables,

$$\rho_m \int_{\omega_{m,i}} \left(U_m(T^{n+1}) - U_m(T^n) \right) r \, dx - \Delta t^{n+1} \int_{\partial \omega_{m,i}} \kappa_m(T^{n+1}) \nabla T^{n+1} \cdot \mathbf{n}_{\omega_{m,i}} r \, ds = \Delta t^{n+1} \int_{\omega_{m,i}} f_m r \, dx,$$

$$(2.7)$$

where the time interval is $\Delta t^{n+1} = t^{n+1} - t^n$. The temperature is given as $T^{n+1} = T(t^{n+1}, x)$, where x represents cylindrical coordinates. For the right-hand sides, we demand $f_s := f \geq 0$ and $f_g = 0$.

More details of the discretization and of dealing with the interface conditions are presented in [3, 4, 9, 10].

In the next section, the properties of the materials in the crystal growth apparatus are described.

2.3. Material Properties

For the technical realization of the apparatus, we implement the axisymmetric geometry given in [11], which is presented in Figure 1. Furthermore, the properties of the materials are specified in [3, 9, 12].

Within the following specific material functions and parameters for the processes, the thermal conductivity κ is given in W/(m K), the electrical conductivity σ_c is given in 1/(Ohm m), the mass density ρ is given in kg/m^3, the specific heat c_{sp} is given in J/(K kg), the temperature T is given in K, and the relative gas constant R_{Ar} is given in J/(K kg). Further, the emissivity ϵ and relative magnetic permeability μ are given dimensionless.

(i) For the gas phase (argon), we have

$$\kappa_{Ar}(T) = \begin{cases} 1.83914\,10^{-4}\,T^{0.800404} & T \le 500, \\ -7.128738 + 6.610288\,10^{-2}\,T - 2.440839\,10^{-4}\,T^2 \\ +4.497633\,10^{-7}\,T^3 - 4.132517\,10^{-10}\,T^4 + 1.514463\,10^{-13}\,T^5 & 500 \le T \le 600, \\ 4.1944\,10^{-4}\,T^{0.671118} & 600 \ge T, \end{cases} \tag{2.8}$$

where $\sigma_{c,Ar} = 0.0$, $\rho_{Ar} = 3.73\,10^{-3}$, $\mu_{Ar} = 1.0$, $z_{Ar} = 3/2$, $R_{Ar} = 2.081308\,10^{-2}$.

(ii) For graphite felt insulation, we have

$$\kappa_{Ins}(T) = \begin{cases} 8.175\,10^{-2} + 2.485\,10^{-4}\,T & T \le 1473, \\ -1.1902\,10^2 + 0.346838\,T - 3.9971\,10^{-4}\,T^2 + 2.2830\,10^{-7}\,T^3 \\ -6.46047\,10^{-11}\,T^4 + 7.2549\,10^{-15}\,T^5 & 1473 \le T \le 1873, \\ -0.7447 + 7.5\,10^{-4}\,T & 1873 \ge T, \end{cases} \tag{2.9}$$

where $\epsilon_{Ins} = 0.2$, $\sigma_{c,Ins}(T) = 2.45\,10^2 + 9.82\,10^{-2}\,T$, $\rho_{Ins} = 170.00$, $\mu_{Ins} = 1.00$, $c_{sp,Ins} = 2100.00$.

(iii) For the graphite, we have

$$\kappa_{Graphite}(T) = 37.715 \exp\left(-1.96\,10^{-4}\,T\right),$$

$$\epsilon_{Graphite}(T) = \begin{cases} 0.67 & T \le 1200, \\ 3.752 - 7.436\,10^{-3}\,T + 6.4163\,10^{-6}\,T^2 \\ -2.3366\,10^{-9}\,T^3 - 3.0833\,10^{-13}\,T^4 & 1200 \le T \le 2200, \\ 0.79 & 2200 \ge T, \end{cases} \tag{2.10}$$

where $\sigma_{c,Graphite} = 10^4$, $\rho_{Graphite} = 1750.0$, $\mu_{Graphite} = 1.0$, $c_{sp,Graphite}(T) = 1/(4.411\,10^2\,T^{-2.306} + 7.97\,10^{-4}\,T^{-0.0665})$.

(iv) For the SiC crystal, we have $\kappa_{SiC-C}(T) = \exp(9.892 + (2.498\,10^2)/T - 0.844\,\ln(T))$, $\epsilon_{SiC-C} = 0.85$, $\sigma_{c,SiC-C} = 10^5$, $\rho_{SiC-C} = 3140.0$, $\mu_{SiC-C} = 1.0$, $c_{sp,SiC-C}(T) = 1/(3.91\,10^4\,T^{-3.173} + 1.835\,10^{-3}\,T^{-0.117})$.

(v) For the SiC powder, we have $\kappa_{SiC-P}(T) = 1.452\,10^{-2} + 5.47\,10^{-12}\,T^3$, $\epsilon_{SiC-P} = 0.85$, $\sigma_{c,SiC-P} = 100.0$, $\rho_{SiC-P} = 1700.0$, $\mu_{SiC-P} = 1.0$, $c_{sp,SiC-P} = 1000.0$.

The functions are programmed in our flexible software package WIAS-HiTNIHS.

In the next section, we discuss the microscopic model.
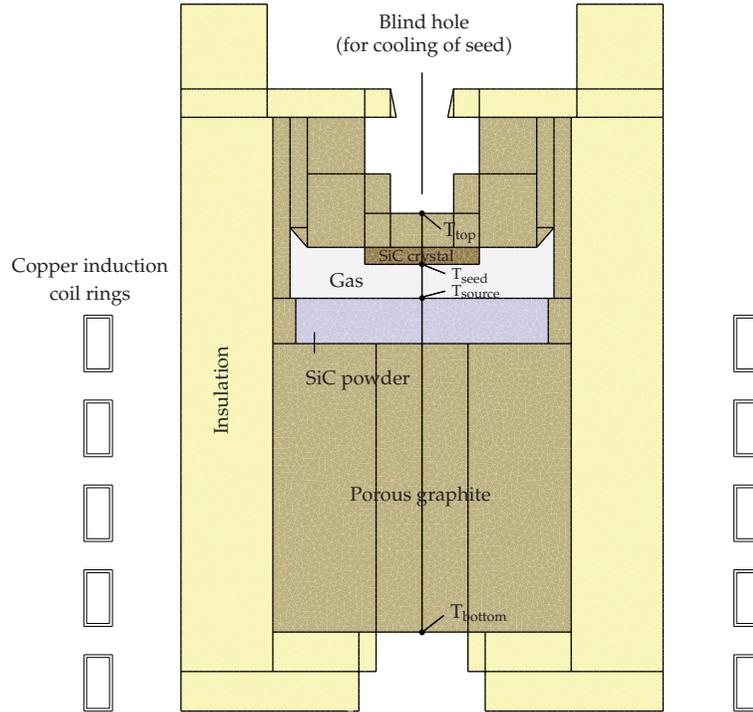
Figure 1: The growth apparatus' dimensions: $r_{\min} = 0$, $r_{\max} = 8.4$ cm, $z_{\min} = 0$, $z_{\max} = 25.0$ cm, the coil rings' dimensions: $r_{\min} = 4.2$ cm, $r_{\max} = 5.2$ cm, $z_{\min} = 0$, $z_{\max} = 14.0$ cm.

2.4. Coupling Method for Macroscopic and Microscopic Models: Operator Splitting

Often simple coupling via the parameters (e.g., target-temperature and growth velocity of the bulk) is enough for the problem.

Here we propose a new idea of coupling the model equations together, on the one hand the heat equations and on the other hand the kinetic equations for molecules.

For a first idea, we deal with abstract operators, which include the heat- and the kinetics equations.

Using our two standard codes of the macro- and micromodels, we could implement a coupled model, by a so-called iterative operator-splitting method. Such a proposed method couples the two physical processes of the thermal situation in the growth apparatus and their important geometrical differences at the deposition layer with the kinetic molecular model. The benefits are a numerical algorithm, that exchanged the underlying operators of the thermal situation and the kinetic molecular situation, which are computed by each software code independently and coupled via an iterative solver step; see a detailed coupling analysis in [13].

In the following algorithm, an iteration method is proposed, with fixed splitting discretization step-size τ.

Due to the underlying multiscale problem of kinetics and heat processes, we have to solve fine time scales of kinetic equations and coarse time scales for heat equations. On a time

interval $[t^n, t^{n+1}]$ that is sufficiently small to yield accurate kinetics, we solve the following subproblems consecutively for $i = 0, 2, \ldots, 2m$ (cf. [14, 15]):

$$\frac{\partial c_i(t)}{\partial t} = Ac_i(t) + Bc_{i-1}(t), \quad \text{with } c_i(t^n) = c^n, \ i = 1, 2, \ldots, j, \tag{2.11}$$

$$\frac{\partial c_i(t)}{\partial t} = Ac_{i-1}(t) + Bc_i(t), \quad \text{with } c_{i+1}(t^n) = c^n, \ i = j + 1, j + 2, \ldots, m, \tag{2.12}$$

where we assume that the operator A has a large time scale (macroscopic model) and B has a small time scale (microscopic model). Further $c_0(t^n) = c^n$, $c_{-1} = 0$, and c^n are initialization and starting conditions.

In the following, we give an overview to the accuracy of the method, which is given in the convergence and the rate of the convergence.

Theorem 2.1. *Let us consider the abstract Cauchy problem in a Banach space* $\mathbf{X}$:

$$\partial_t c(t) = Ac(t) + Bc(t), \quad 0 < t \leq T,$$
$$c(0) = c_0, \tag{2.13}$$

where $A, B, A+B : \mathbf{X} \to \mathbf{X}$ *are given linear operators being generators of the* C_0-*semigroup and* $c_0 \in \mathbf{X}$ *is a given element. Then the iteration process* (2.11)-(2.12) *is convergent. The rate of convergence is of higher order and given as* $\mathcal{O}(\tau_n^{2m})$, *where the iterations are* $i = 1, 3, \ldots, 2m + 1$.

The proof is given in [15].

In the next subsection, we present the methods for the microscopic model.

3. Microscopic Model: Quantum Chemical Molecular Dynamics (QM/MD) of SiC Condensation (Methodology)

The density-functional tight-binding (DFTB) method is employed as the quantum inter-atomic potential in our molecular dynamics (MD) simulations, using atomic and diatomic parameters obtained from density functional theory; see [16]. DFTB is an approximate density functional theory method based on the tight binding approach, and utilizes an optimized minimal LCAO Slater-type all-valence basis set in combination with a two-center approximation for Hamiltonian matrix elements. Parameter sets for Si-Si and Si-C were taken from [17]. Energies and gradients are evaluated direct (on the fly) during the dynamic simulation. As in our previous simulations of carbon cap [18] and subsequent nanotube formation [19] on the C- and Si-faces of SiC(000-1) surfaces during sublimation evaporation, we have not included charge- or spin-polarization in the present work. Further, we will consider in a next model electrokinetic effect on heat transfer in parallel-plate microchannels, hydrodynamic focusing effects, and nanoeffect as done in [20–23].

For time propagation we employed a velocity Verlet integrator with a time step of 1.209 fs (50 atomic units) and used a Nose-Hoover chain thermostat to generate a canonical ensemble for target temperature T_t; see [24]. The thermostat was employed uniformly in the reaction system.
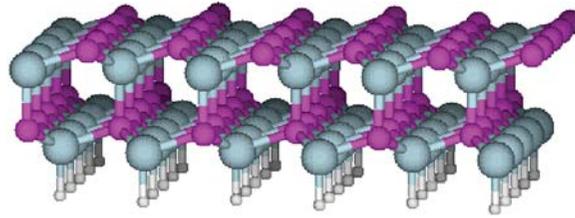
Figure 2: Optimized geometry of the C-face of the (000-1) SiC surface as initial starting point for QM/MD simulations. Blue spheres correspond to silicon atoms, purple spheres correspond to carbon atoms, and white spheres correspond to hydrogen atoms terminating the slab model in bulk direction. The model is the unit cell used in periodic boundary calculations with infinite surface extension.

Regarding the atomistic structure of the employed surface model systems, we have chosen the C-face of the same square SiC(000-1) slab unit cell as in our previous study, [19] consisting of two SiC layers terminated by hydrogen atoms to mimic bulk effect in the direction away from the surface. Periodic boundary conditions were employed with a unit cell size of 1000 Å in the direction perpendicular to the surface and 16.0 Å and 15.4 Å in the other two surface-directions to achieve two-dimensional slab periodicity. The geometry optimized structure of this surface model is shown in Figure 2.

During MD simulations, the movements of hydrogen terminating atoms were frozen. Using such an approach, we have effectively introduced a steep temperature gradient from the deepest bulk-side SiC layer to the atoms lying above on the surface. The slab model was then annealed at T_t = 2000 K for 1.20 picoseconds, and 3 structures were selected as initial starting geometries at t = 0.60 picosecond (trajectory A50), t = 0.72 picosecond (trajectory A60), and t = 0.86 picosecond (trajectory A80). In the vicinity of the surface, 10 SiC molecules were randomly distributed in the gas phase. Since these molecules are nonbonded to the surface, they are subsequently thermostated at T_t. Gas phase molecules approaching the surface will experience immediate cooling, which will drive the condensation process during these simulations.

In the microscopic model, we can derive the growth rate v of the seed surface in dependence on temperature and pressure. Based on these growth rates, we can adapt the geometry for the macroscopic model. Such modification helps to give more accurate temperature differences in the macroscopic model and understand the growth process.

In the next section, we present results of our numerical experiments.

4. Numerical Experiments

We present in the following our macro- and microscopic simulations, where the microscopic simulations take into account the target temperature of the macroscopic model.

4.1. Macroscopic Model: Simulation of the Temperature Field in the Apparatus

For the numerical results, we apply the parameter functions in Section 2.3. We consider the geometry shown in Figure 1, using a constant total input power of 10 kW (cf. [11]). The numerical experiments are performed using the software WIAS-HiTNIHS (cf. [4]) based

Special temperature-differences for transient
computations for grid-width 0.00001
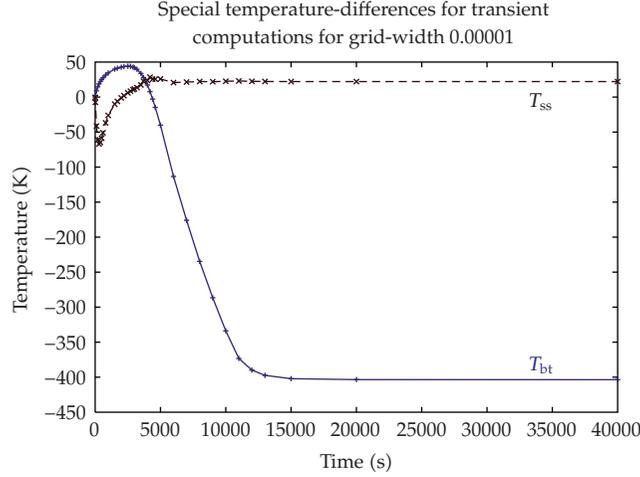
Figure 3: Transient results for the temperature differences T_{bt} and T_{ss}.

on the software package *pdelib* (cf. [25]) which uses the sparse matrix solver *PARDISO* (cf. [26]). We compute the coupled system consisting of the heat equations and Maxwell's equations. For the growth process, the temperature difference $T_{ss} = T(r_{source}, z_{source}) - T(r_{seed}, z_{seed})$ (with the coordinates $(r_{source}, z_{source}) = (0, 0.143)$ and $(r_{seed}, z_{seed}) = (0, 0.158)$, corresponding to the points T_{source} and T_{seed} in Figure 1) is crucial. On the other hand, in the physical growth experiments, usually only the temperatures $T(r_{bottom}, z_{bottom})$ and $T(r_{top}, z_{top})$ (with the coordinates $(r_{bottom}, z_{bottom}) = (0, 0.028)$ and $(r_{top}, z_{top}) = (0, 0.173)$, corresponding to the points T_{bottom} and T_{top} in Figure 1) are measurable and their difference $T_{bt} = T(r_{bottom}, z_{bottom}) - T(r_{top}, z_{top})$ is often used as an indicator for T_{ss}. In Figure 3, we present the temperature differences T_{ss} and T_{bt}. As a result of our computations, the temperature difference T_{bt} can only restrictively be used as an indicator for the temperature difference T_{ss} (cf. the discussions in [5, 9]).

The further computations are based on the stationary case, dealing with (2.1) by discarding the terms with a time derivative. For this case, the results are virtually equal to the one in the transient case with $t > 15000$ seconds. For the stationary results, we focus on the error analysis for the space dimension by applying the grid refinement. The solutions for the heat equation are computed at the points $T(r_{bottom}, z_{bottom})$ and $T(r_{top}, z_{top})$ for successive grids. For the error analysis, we apply the following error differences:

$$\epsilon_{abs} = \left| \tilde{T}_{j+1}(r, z) - \tilde{T}_j(r, z) \right|, \tag{4.1}$$

where $\tilde{T}_j(r, z)$ and $\tilde{T}_{j+1}(r, z)$ are solutions evaluated at the point (r, z) which has been computed using the grids j and $j + 1$, respectively. The elements of the grid $j + 1$ are approximately $1/4$ of the elements of the grid j. The results are presented in Table 1.

The result of the refinement indicates the reduction of the absolute difference as it is demanded for the convergence of the discretization method. The method is stabilized in the presented refinement by reducing the differences.

Table 1: Computations on different grids for the errors analysis with absolute differences (cf. (4.1)).

	Grid	Grid point $(0, 0.028)$ (T_{bottom})		Grid point $(0, 0.173)$ (T_{top})	
Level	Number of nodes	Solution T [K]	Absolute difference T [K]	Solution T [K]	Absolute difference T [K]
0	1532	2408.11		2813.29	
1	23017	2409.78	1.67	2812.78	1.01
2	91290	2410.35	0.57	2811.79	0.49
3	364225	2410.46	0.11	2811.60	0.19

Stationary temperature field

Height = 25 cm
T_{min} = 537.517 K
T_{max} = 3312.53 K
ΔT_{max} = 0 K

Heating power in crucible = 7811.89 W
heating power in coil = 2188.11 W

Prescribed power = 10000 W
Frequency = 10000 Hz

Coil:
5 rings
Top = 0.18 m
Bottom = 0.02 m

300 3300 T (K)   Radius = 8.4 cm

50   ΔT (K) between isolines
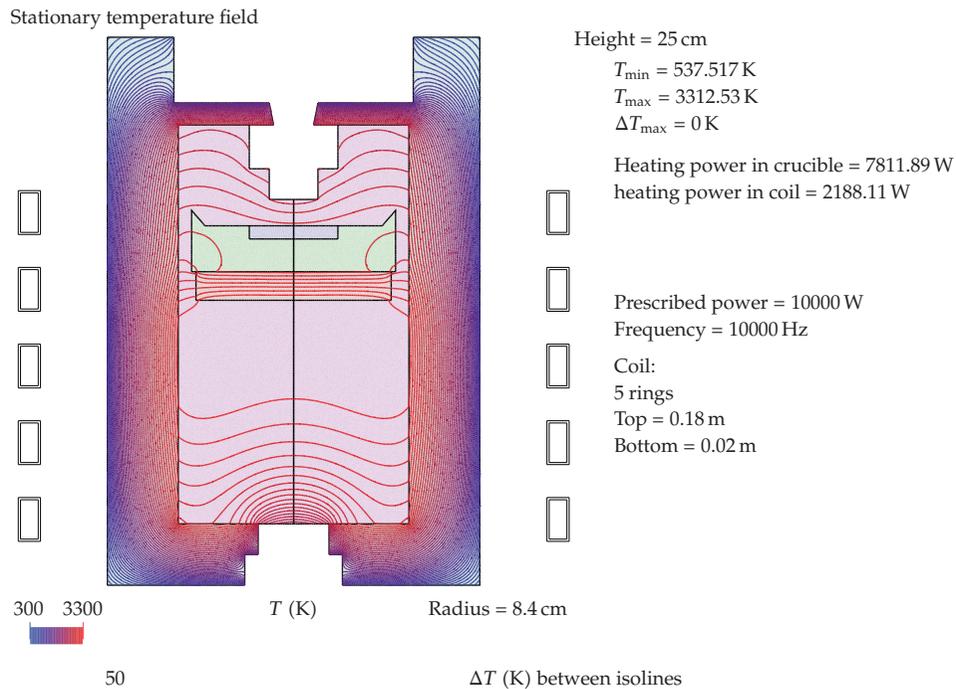
Figure 4: Temperature field for the apparatus simulated for the stationary case with 23017 nodes.

In Figure 4, the temperature field is presented for the stationary case. The temperature increases from the bottom to the middle of the graphite pot, and decreases from the middle to the top of the graphite pot.

4.2. Microscopic Model: Atomistic QM/MD Simulations of SiC Condensation on the C-face of Si(000-1)

The total time of the three condensation simulations was 24.02 picoseconds. This is admittedly a time too short for the study of crystal growth, which would ideally require annealing simulations on the order of several 100 nanoseconds, but this study is focusing on the initial stages of SiC aggregation and tries to identify key features in the condensation process. As such, this is at present rather a preliminary study exploring the applicability
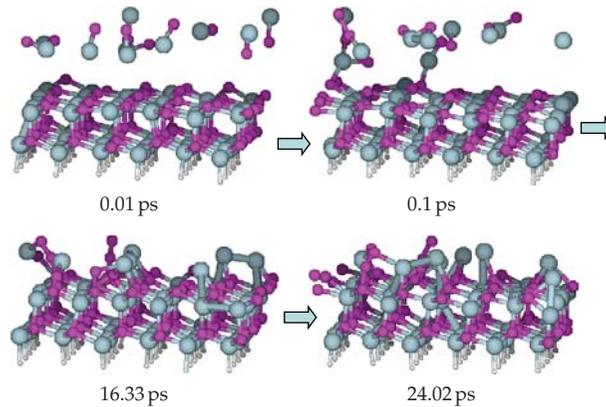
Figure 5: Simulation of the addition of 10 SiC atoms on the C-face of the (000-1) SiC surface from Figure 2. Blue spheres correspond to silicon atoms, purple spheres correspond to carbon atoms, and white spheres correspond to hydrogen atoms terminating the slab model in bulk direction. Times are given in picoseconds (ps), indicating that the moment the snapshots were taken during the dynamics simulations.

of QM/MD simulations for SiC crystal growth. We have first concentrated on the polar C-surface of SiC (0001) since it has a maximum of dangling bonds with highest reactivity. The Si-face and other nonpolar surfaces are much less reactive; see [27]. Since our seed crystal surface slab model contains only two SiC layers, we are also unable to address the issue of polymorphism at present, although it should be noted that our model system rather resembles the cubic 3C than the hexagonal polytypes.

T_f was chosen as 2000 K for all simulations, and representative snapshots from trajectory A50 are given in Figure 5. We find that under the present conditions with a relatively high density of SiC gas molecules, several of them attach very quickly to the surface (2 after 0.10 picosecond). Also, SiC molecules can react with each other to form dimers, preferably with C-C bonds. Eventually, an average of 5.3 SiC molecules become attached to the surface in the three simulations, with the other molecules being lost to the vacuum layer.

Once attached, the Si atoms on the surface prove to be highly mobile, as their bond radius is larger than the case of carbon, and the binding energies are lower [18]. The carbon atoms on the surface tend to form C_2 units, and behave similar to "wobbling C_2" entities that we had observed for high-temperature simulations of pure carbon; see [28]. It seems from our simulations at this stage that the system tries to reach an equilibrium with a constant number of C-C in the new layers, and that the Si atoms are more isolated, becoming occasionally attacked by a C2 dimer. In particular, C_2 units are oriented mainly perpendicular to the surface, while the more visible Si_2 dimers do not show such an alignment preference. The surface itself retains the structure of alternating Si–C units. A new layer of Si–C units is being deposited with a somewhat inhomogeneous structure containing C_2 and Si_2 units at first, and gradually becoming more homogeneous due to annealing.

5. Summary

We have presented a model for the heat transport inside a technical apparatus for crystal growth of SiC single crystals. We introduce the heat equation and the radiation of the apparatus and the coupled situation of the different materials. The equations are discretized

by the finite volume method and the complex material functions are embedded in this method. Transient and stationary results are presented leading to some information about the processes within the technical apparatus. We present numerical results for the stationary case to support the accuracy of our solutions. We also presented atomistic quantum chemical molecular dynamics (QM/MD) simulations based on the density-functional tight-binding (DFTB) method for initial reactions of gaseous SiC on the polar C-face of SiC(000-1). In our future work, we concentrate on further implementations and numerical methods for a crystal growth model and use kinetic data obtained from more accurate microscopic model simulations in the simulation of the heat transport. Once longer and a larger number of trajectories are obtained in our microsimulations, it will be possible to deduct an accurate QM/MD-based estimate for the bulk growth, in dependence on the temperature to our macrosimulations. This data will then enter the iterative solution of the heat and kinetics equations of the coupled macroscopic and microscopic models.

References

[1] St. Müller, R. C. Glass, H. M. Hobgood, et al., "Progress in the industrial production of SiC substrates for semiconductor devices," *Materials Science and Engineering B*, vol. 80, no. 1–3, pp. 327–331, 2001.

[2] X. Yao, B. He, H. Wang, and X. Zhou, "Numerical simulation of dendrite growth during solidification," *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 7, no. 2, pp. 171–176, 2006.

[3] O. Klein and P. Philip, "Transient numerical investigation of induction heating during sublimation growth of silicon carbide single crystals," *Journal of Crystal Growth*, vol. 247, no. 1-2, pp. 219–235, 2003.

[4] P. Philip, *Transient numerical simulation of sublimation growth of SiC bulk single crystals. Modeling, finite volume method, results*, Ph.D. thesis, Weierstrass-Institute for Applied Analysis and Stochastics, Berlin, Germany, 2003, Report No. 22.

[5] O. Klein, P. Philip, and J. Sprekels, "Modeling and simulation of sublimation growth of SiC bulk single crystals," *Interfaces and Free Boundaries*, vol. 6, no. 3, pp. 295–314, 2004.

[6] J. Rappaz and M. Swierkosz, "Modelling in numerical simulation of electromagnetic heating," in *Modelling and Optimization of Distributed Parameter Systems*, pp. 313–320, Chapman & Hall, New York, NY, USA, 1996.

[7] O. Klein and P. Philip, "Correct voltage distribution for axisymmetric sinusoidal modeling of induction heating with prescribed current, voltage, or power," *IEEE Transactions on Magnetics*, vol. 38, no. 3, pp. 1519–1523, 2002.

[8] J. Geiser, "R^3T: radioactive-retardation-reaction-transport-program for the simulation of radioactive waste disposals," Tech. Rep., Institute for Scientific Computation, Texas A&M University, College Station, Tex, USA, April 2004.

[9] O. Klein and P. Philip, "Transient temperature phenomena during sublimation growth of silicon carbide single crystals," *Journal of Crystal Growth*, vol. 249, no. 3-4, pp. 514–522, 2003.

[10] O. Klein and P. Philip, "Transient conductive-radiative heat transfer: discrete existence and uniqueness for a finite volume scheme," *Mathematical Models & Methods in Applied Sciences*, vol. 15, no. 2, pp. 227–258, 2005.

[11] M. Pons, M. Anikin, K. Chourou, et al., "State of the art in the modelling of SiC sublimation growth," *Materials Science and Engineering B*, vol. 61-62, pp. 18–28, 1999.

[12] O. Nilsson, H. Mehling, R. Horn, et al., "Determination of the thermal diffusivity and conductivity of monocrystalline silicon carbide (300–2300 K)," *High Temperatures-High Pressures*, vol. 29, no. 1, pp. 73–79, 1997.

[13] J. Geiser, "Iterative operator-splitting methods with higher-order time integration methods and applications for parabolic partial differential equations," *Journal of Computational and Applied Mathematics*, vol. 217, no. 1, pp. 227–242, 2008.

[14] J. F. Kanney, C. T. Miller, and C. T. Kelley, "Convergence of iterative split-operator approaches for approximating nonlinear reactive problems," *Advances in Water Resources*, vol. 26, no. 3, pp. 247–261, 2003.

[15] I. Faragó and J. Geiser, "Iterative operator-splitting methods for linear problems," *International Journal of Computational Science and Engineering*, vol. 3, no. 4, pp. 255–263, 2007.

[16] D. Porezag, Th. Frauenheim, Th. Köhler, G. Seifert, and R. Kaschner, "Construction of tight-binding-like potentials on the basis of density-functional theory: application to carbon," *Physical Review B*, vol. 51, no. 19, pp. 12947–12957, 1995.

[17] E. Rauls, Z. Hajnal, P. Deák, and Th. Frauenheim, "Theoretical study of the nonpolar surfaces and their oxygen passivation in 4*H*- and 6*H*-SiC," *Physical Review B*, vol. 64, no. 24, Article ID 245323, 7 pages, 2001.

[18] S. Irle, Z. Wang, G. Zheng, K. Morokuma, and M. Kusunoki, "Theory and experiment agree: single-walled carbon nanotube caps grow catalyst-free with chirality preference on a SiC surface," *The Journal of Chemical Physics*, vol. 125, no. 4, Article ID 044702, 5 pages, 2006.

[19] Z. Wang, S. Irle, G. Zheng, M. Kusunoki, and K. Morokuma, "Carbon nanotubes grow on the C face of SiC (0001) during sublimation decomposition: quantum chemical molecular dynamics simulations," *The Journal of Physical Chemistry C*, vol. 111, no. 35, pp. 12960–12972, 2007.

[20] C.-C. Chang and R.-J. Yang, "Hydrodynamic focusing effect on two-unmixed-fluid in microchannels," *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 9, no. 3, pp. 213–220, 2008.

[21] J. Fan, L. Wang, and L. Cheng, "Electrokinetic effects on flow and heat transfer in parallel-plate microchannels," *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 8, no. 3, pp. 335–345, 2007.

[22] J. Fan, L. Wang, and L. Cheng, "Forced convection in rectangular microchannels: electrokinetic effects," *International Journal of Nonlinear Sciences and Numerical Simulation*, vol. 8, no. 3, pp. 359–374, 2007.

[23] J.-H. He, Y.-Q. Wan, and L. Xu, "Nano-effects, quantum-like properties in electrospun nanofibers," *Chaos, Solitons & Fractals*, vol. 33, no. 1, pp. 26–37, 2007.

[24] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, New York, NY, USA, 1989.

[25] J. Fuhrmann, Th. Koprucki, and H. Langmach, "pdelib: an open modular tool box for the numerical solution of partial differential equations. Design patterns," in *Proceeding of the 14th GAMM Seminar on Concepts of Numerical Software*, W. Hackbusch and G. Wittum, Eds., Kiel, Germany, January 2001.

[26] O. Schenk and K. Gärtner, "Solving unsymmetric sparse systems of linear equations with PARDISO," *Future Generation Computer Systems*, vol. 20, no. 3, pp. 475–487, 2004.

[27] J. Pollmann, P. Krüger, and M. Sabisch, "Atomic and electronic structure of SiC surfaces from ab-initio calculations," *Physica Status Solidi (B)*, vol. 202, no. 1, pp. 421–445, 1997.

[28] S. Irle, G. Zheng, M. Elstner, and K. Morokuma, "Formation of fullerene molecules from carbon nanotubes: a quantum chemical molecular dynamics study," *Nano Letters*, vol. 3, no. 4, pp. 465–470, 2003.

Research Article

Free Vibration Analysis of Rectangular Orthotropic Membranes in Large Deflection

Zheng Zhou-Lian,[1,2] Liu Chang-Jiang,[1] He Xiao-Ting,[1] and Chen Shan-Lin[1]

[1] Department of Civil Engineering, Chongqing University, Chongqing 400045, China
[2] Key Laboratory of the Three Gorges Reservoir Region's Eco-Environment, Chongqing University, Ministry of Education, Chongqing 400045, China

Correspondence should be addressed to Zheng Zhou-Lian, zhengzhoulian@yahoo.com.cn

This paper reviewed the research on the vibration of orthotropic membrane, which commonly applied in the membrane structural engineering. We applied the large deflection theory of membrane to derive the governing vibration equations of orthotropic membrane, solved it, and obtained the power series formula of nonlinear vibration frequency of rectangular membrane with four edges fixed. The paper gave the computational example and compared the two results from the large deflection theory and the small one, respectively. Results obtained from this paper provide some theoretical foundation for the measurement of pretension by frequency method; meanwhile, the results provide some theoretical foundation for the research of nonlinear vibration of membrane structures and the response solving of membrane structures under dynamic loads.

1. Introduction

The membrane structure is a new structure system, which has been rising during the recent several dozens of years. Because of its economy, beauty, and less dead weight, it is widely applied to large span structures, such as large-scale stadium, exhibition center, works of decoration, and so on. The application of tensile membrane structure was the most extensive in each kind of membrane structure systems. The tensile membrane structure's stiffness is formed by the zonal and meridional pretensions. The tallying degree of the actual pretension value and the design pretension value directly influences the normal utilization and safety after the construction finished [1, 2]. Therefore, the nondestructive monitoring of the membrane structure's pretension is very important and necessary after the construction finished.

At present, most researches focused on design and construction of membrane structures; however, there is little research on the pretension measurement of membrane structures [1, 2]. The main methods of the pretension measurement include strain method, frequency method, deflection method, and "cable analogy" method [3]. If we study the application of frequency method, the vibration theory of membrane must be involved. Many scholars studied about the vibration theory of membrane. Their researches involve the problem of free vibration of a confocal composite elliptical membrane [4], the problem of fundamental frequency of rectangular membranes with an internal oblique support [5], the problem of free vibration of composite rectangular membranes with an oblique and a bent interface [6, 7], the problem of free in-plane vibration of an axially moving membrane [8]. However, these researches did not aim at this kind of orthotropic membrane that was used in construction field. Moreover, these researches have not obtained the power series formula of nonlinear vibration frequency of the orthotropic membrane.

In this paper, we studied the vibration of orthotropic membranes according to the large deflection theory of membrane [9, 10] and obtained the power series formula of nonlinear vibration frequency of rectangular membranes with four edges fixed. The paper gave the computational example and compared the two results from the large deflection theory and the small one, respectively.

2. Governing Equations and Boundary Conditions

The studied rectangular membrane is orthotropic. Its two orthogonal directions are the two principal fiber directions, and the material characteristics of the two principal fiber directions are different. Assume that the studied rectangular membrane is fixed on four edges. The two principal fiber directions are x and y, respectively. a and b denote the length of x and y direction, respectively; N_{0x} and N_{0y} denote initial tension in x and y, respectively, as shown in Figure 1.

According to the large deflection theory and D'Alembert's principle of membranes [8–10], the vibration partial differential equation and consistency equation of orthotropic membrane are

$$\rho \frac{\partial^2 w}{\partial t^2} - (N_x + N_{0x})\frac{\partial^2 w}{\partial x^2} - (N_y + N_{0y})\frac{\partial^2 w}{\partial y^2} = 0$$

$$\frac{1}{E_1 h}\frac{\partial^2 N_x}{\partial y^2} - \frac{\mu_2}{E_2 h}\frac{\partial^2 N_y}{\partial y^2} - \frac{\mu_1}{E_1 h}\frac{\partial^2 N_x}{\partial x^2} + \frac{1}{E_2 h}\frac{\partial^2 N_y}{\partial x^2} - \frac{1}{Gh}\frac{\partial^2 N_{xy}}{\partial x \partial y} = \left(\frac{\partial^2 w}{\partial x \partial y}\right)^2 - \frac{\partial^2 w}{\partial x^2}\frac{\partial^2 w}{\partial y^2},$$

$$(2.1)$$

where ρ denotes aerial density of membrane; N_x and N_y denote tension in x and y, respectively; N_{0x} and N_{0y} denote initial tension in x and y, respectively; N_{xy} denotes shear force; w denotes deflection: $w(x, y, t)$; h denotes membrane's thickness; E_1 and E_2 denote Young's modulus in x and y, respectively; G denotes shearing modulus; μ_1 and μ_2 denote Poisson's ratio in x and y, respectively.

While the membrane is in vibration, the effect of shearing stress is so small that we may take $N_{xy} = 0$. Introducing the stress function and letting $N_x = h(\partial^2 \varphi / \partial y^2)$, $N_y = h(\partial^2 \varphi / \partial x^2)$,
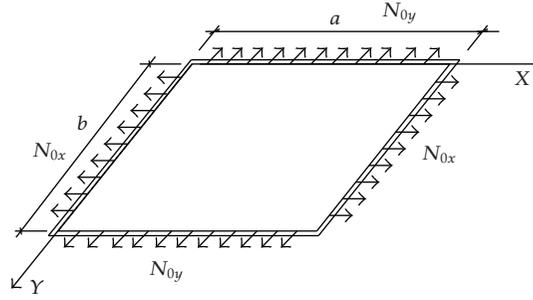
Figure 1: Rectangular membrane with four edges fixed.

$N_{0x} = h \cdot \sigma_{0x}$, and $N_{0y} = h \cdot \sigma_{0y}$, (2.1) can be simplified as follows:

$$\frac{\rho}{h} \frac{\partial^2 w}{\partial t^2} - \left(\sigma_{0x} + \frac{\partial^2 \varphi}{\partial y^2} \right) \frac{\partial^2 w}{\partial x^2} - \left(\sigma_{0y} + \frac{\partial^2 \varphi}{\partial x^2} \right) \frac{\partial^2 w}{\partial y^2} = 0, \tag{2.2}$$

$$\frac{1}{E_1} \frac{\partial^4 \varphi}{\partial y^4} + \frac{1}{E_2} \frac{\partial^2 \varphi}{\partial x^4} = \left(\frac{\partial^2 w}{\partial x \partial y} \right)^2 - \frac{\partial^2 w}{\partial x^2} \frac{\partial^2 w}{\partial y^2}, \tag{2.3}$$

where φ denotes stress function: $\varphi(x, y, t)$; σ_{0x} and σ_{0y} denote initial tensile stress in x and y, respectively.

The corresponding boundary conditions can be expressed as follows:

$$\begin{aligned} w(0, y, t) = 0, \qquad w(a, y, t) = 0, \\ w(x, 0, t) = 0, \qquad w(x, b, t) = 0. \end{aligned} \tag{2.4}$$

3. Solution of Free Vibration Frequency

Functions that satisfy the boundary conditions (2.4) are taken as follows:

$$\begin{aligned} w(x, y, t) = W(x, y) T(t), \\ \varphi(x, y, t) = \phi(x, y) T^2(t), \end{aligned} \tag{3.1}$$

where $W(x, y)$ is the given mode shape function, and $\phi(x, y)$ and $T(t)$ are the unknown functions.

Assume that the mode shape function is as follows:

$$W(x, y) = \sin \frac{m \pi x}{a} \sin \frac{n \pi y}{b}, \tag{3.2}$$

where m and n are integers and denote the sine half-wave number in x and y, respectively. Equation (3.2) satisfies the boundary conditions automatically.

Substituting (3.1) into (2.3) yields

$$\frac{1}{E_1}\frac{\partial^4 \phi}{\partial y^4} + \frac{1}{E_2}\frac{\partial^2 \phi}{\partial x^4} = \left(\frac{\partial^2 W}{\partial x \partial y}\right)^2 - \frac{\partial^2 W}{\partial x^2}\frac{\partial^2 W}{\partial y^2}. \tag{3.3}$$

Substituting (3.2) into (3.3) yields

$$\frac{1}{E_1}\frac{\partial^4 \phi}{\partial y^4} + \frac{1}{E_2}\frac{\partial^2 \phi}{\partial x^4} = \frac{m^2 n^2 \pi^4}{2a^2 b^2}\left(\cos\frac{2m\pi x}{a} + \cos\frac{2n\pi y}{b}\right). \tag{3.4}$$

Assume that the solution of (3.4) is

$$\phi(x,y) = \alpha\cdot\cos\frac{2m\pi x}{a} + \beta\cdot\cos\frac{2n\pi y}{b}. \tag{3.5}$$

Substituting (3.5) into (3.4) yields

$$\alpha = \frac{E_2 n^2 a^2}{32 m^2 b^2}, \qquad \beta = \frac{E_1 m^2 b^2}{32 n^2 a^2}. \tag{3.6}$$

Substituting (3.1) into (2.2), according to the Galerkin method [8, 10], yields

$$\iint_S \left(\frac{\rho}{h}\frac{\partial^2 w}{\partial t^2} - \frac{\partial^2 \varphi}{\partial y^2}\frac{\partial^2 w}{\partial x^2} - \frac{\partial^2 \varphi}{\partial x^2}\frac{\partial^2 w}{\partial y^2}\right) W(x,y)\,ds$$

$$= \iint_S \left[\frac{\rho}{h}W\frac{\partial^2 T(t)}{\partial t^2} - \left(\sigma_{0x}\frac{\partial^2 W}{\partial x^2} + \sigma_{0y}\frac{\partial^2 W}{\partial y^2}\right)T(t)\left(\frac{\partial^2 \phi}{\partial y^2}\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 \phi}{\partial x^2}\frac{\partial^2 W}{\partial y^2}\right)T^3(t)\right]W(x,y)\,ds$$

$$= 0. \tag{3.7}$$

Obviously, (3.7) is a nonlinear differential equation with respect to $T(t)$, and it can be expressed as follows:

$$A\cdot\frac{d^2 T(t)}{dt^2} - B\cdot T(t) - C\cdot T^3(t) = 0, \tag{3.8}$$

where

$$A = \iint_S \frac{\rho}{h}W^2\,ds = \iint_S \frac{\rho}{h}\sin^2\frac{m\pi x}{a}\sin^2\frac{n\pi y}{b}\,ds = \frac{\rho ab}{4h},$$

$$B = \iint_S \left(\sigma_{0x}\cdot\frac{\partial^2 W}{\partial x^2} + \sigma_{0y}\cdot\frac{\partial^2 W}{\partial y^2}\right)W\,ds = -\frac{\pi^2 ab}{4}\left(\frac{m^2}{a^2}\sigma_{0x} + \frac{n^2}{b^2}\sigma_{0y}\right), \tag{3.9}$$

$$C = \iint_S \left(\frac{\partial^2 \phi}{\partial y^2}\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 \phi}{\partial x^2}\frac{\partial^2 W}{\partial y^2}\right)W\,ds = -\frac{ab\pi^4}{64}\left(\frac{E_1 m^4}{a^4} + \frac{E_2 n^4}{b^4}\right).$$

Substituting the values of A, B, and C into (3.8) and dividing by $\rho ab/4h$ on two ends yields

$$\frac{d^2T(t)}{dt^2} + \frac{h\pi^2}{\rho}\left(\frac{m^2}{a^2}\sigma_{0x} + \frac{n^2}{b^2}\sigma_{0y}\right)T(t) + \frac{h\pi^4}{16\rho}\left(\frac{E_1m^4}{a^4} + \frac{E_2n^4}{b^4}\right)T^3(t) = 0. \qquad (3.10)$$

Letting $M = (h\pi^2/\rho)((m^2/a^2)\sigma_{0x} + (n^2/b^2)\sigma_{0y})$, $N = (\pi^4h/16\rho)((E_1m^4/a^4) + (E_2n^4/b^4))$ yields

$$\frac{d^2T(t)}{dt^2} + M{\cdot}T(t) + N{\cdot}T^3(t) = 0. \qquad (3.11)$$

Integrating (3.11) yields

$$\left(\frac{dT(t)}{dt}\right)^2 + M{\cdot}T^2(t) + \frac{N}{2}{\cdot}T^4(t) = H. \qquad (3.12)$$

In (3.12), the value of H is determined by the initial conditions. Assume that the initial displacement is $T|_{t=0} = T_0$. T_0 is the amplitude of the membrane, so the initial velocity is

$$\left.\frac{dT(t)}{dt}\right|_{t=0} = 0. \qquad (3.13)$$

Substituting $T|_{t=0} = T_0$ and (3.13) into (3.12) yields $H = MT_0^2 + (N/2)T_0^4$. Then substituting $H = MT_0^2 + (N/2)T_0^4$ into (3.12) yields

$$\frac{dT(t)}{dt} = T_0^2\sqrt{\frac{M}{T_0^2} + \frac{N}{2}}{\cdot}\sqrt{\left(1 - \frac{T^2}{T_0^2}\right){\cdot}\left(1 + \frac{NT_0^2}{2M + NT_0^2}{\cdot}\frac{T^2}{T_0^2}\right)}, \qquad (3.14)$$

letting $\lambda = T_0^2\sqrt{M/T_0^2 + N/2}$, $k^2 = NT_0^2/(2M + NT_0^2)$.

Integrating (3.11) by the separate variable method, we can obtain the period of the vibration of the membrane:

$$Z = \frac{4}{\lambda}\int_0^{T_0}\left[\left(1 - \frac{T^2}{T_0^2}\right)\left(1 + k^2{\cdot}\frac{T^2}{T_0^2}\right)\right]^{-1/2}dT. \qquad (3.15)$$

Letting $T/T_0 = \sin\theta$, (3.15) can be transformed into

$$Z = \frac{4T_0}{\lambda}\int_0^{\pi/2}(1 + k^2{\cdot}\sin^2\theta)^{-1/2}d\theta \quad (0 \le k{\cdot}\sin\theta \le 1), \qquad (3.16)$$

where $(1 + k^2 \cdot \sin^2\theta)^{-1/2}$ may be spread as a power series with respect to $k \cdot \sin\theta$:

$$(1 + k^2 \cdot \sin^2\theta)^{-1/2} = 1 - \frac{1}{2}k^2 \sin^2\theta + \frac{1 \cdot 3}{2 \cdot 4}k^4 \sin^4\theta + \cdots (-1)^n \frac{(2p-1)!!}{(2p)!!}(k \sin\theta)^{2p}. \quad (3.17)$$

Substituting (3.17) into (3.16) then solving (3.16) through integrating item by item [10, 11] yields

$$
\begin{aligned}
Z &= \frac{4T_0}{\lambda} \int_0^{\pi/2} \left[1 - \frac{1}{2}k^2 \sin^2\theta + \frac{1 \cdot 3}{2 \cdot 4}k^4 \sin^4\theta + \cdots (-1)^n \frac{(2p-1)!!}{(2p)!!}(k \sin\theta)^{2p} \right] d\theta \\
&= \frac{2\pi T_0}{\lambda} \left[1 - \left(\frac{1}{2}\right)^2 k^2 + \left(\frac{1 \cdot 3}{2 \cdot 4}\right)^2 k^2 + \cdots + (-1)^p \left(\frac{(2p-1)!!}{(2p)!!}\right)^2 k^{2p} \right] \\
&= \frac{2\pi T_0}{\lambda} \sum_{n=0}^{\infty} (-1)^p \left(\frac{(2p-1)!!}{(2p)!!}\right)^2 k^{2p},
\end{aligned}
\quad (3.18)
$$

where $p = 0, 1, 2, 3, \ldots$.

Therefore, the vibration frequency of the membrane is

$$\omega = \frac{2\pi}{Z} = \frac{\sqrt{M + (N/2)T_0^2}}{\sum_{n=0}^{\infty}(-1)^p((2p-1)!!/(2p)!!)^2\left(NT_0^2/(2M + NT_0^2)\right)^p}, \quad (3.19)$$

where $M = (h\pi^2/\rho)((m^2/a^2)\sigma_{0x} + (n^2/b^2)\sigma_{0y})$, $N = (\pi^4 h/16\rho)(E_1 m^4/a^4 + E_2 n^4/b^4)$, and T_0 is the amplitude of the membrane.

In (3.19), letting $T_0 \rightarrow 0$, (3.19) can be transformed into the formula obtained according to the small deflection:

$$\omega = \pi \sqrt{\frac{h}{\rho}} \cdot \sqrt{\frac{\sigma_{0x} \cdot m^2}{a^2} + \frac{\sigma_{0y} \cdot n^2}{b^2}}. \quad (3.20)$$

4. Computational Examples and Discussion

Take the membrane material commonly applied in project as an example. Young's moduli in x and y are $E_1 = 1.4 \times 10^6\,\mathrm{KN/m^2}$ and $E_2 = 0.9 \times 10^6\,\mathrm{KN/m^2}$, respectively; the aerial density of membranes is $\rho = 1.7\,\mathrm{kg/m^2}$; the membrane's thickness is $h = 1.0\,\mathrm{mm}$. Calculate the vibration frequency of the membrane according to (3.19).

We can draw the conclusion from the result of Table 1. If we exchange the two lengths of the two orthogonal directions and consider the orthotropic characteristic of the membrane, the result is dissimilar with the one before the exchange; however, if we only exchange the two lengths of the two orthogonal directions and do not consider the orthotropic characteristic, the result is similar with the one before the exchange. Therefore, we need to consider the orthotropic characteristic of membranes in practical engineering.

We can draw the conclusion from the result of Table 2. The initial displacement (the amplitude) has influenced the vibration frequency of the rectangle membrane when

Table 1: $T_0 = 0.1$ m, $\sigma_{0x} = \sigma_{0y} = 5.0 \times 10^3$ KN/m^2, and $m = n = 1$.

M	$a = 1$, $b = 1$	$a = 1$, $b = 2$	$a = 2$, $b = 1$	$a = 1$, $b = 3$	$a = 3$, $b = 1$	$a = 1$, $b = 4$	$a = 4$, $b = 1$	$a = 1$, $b = 5$	$a = 5$, $b = 1$
Rad/s	344.231	273.032	249.679	245.268	237.578	260.131	233.909	258.752	232.316
Hz	54.786	43.455	39.738	39.036	37.812	41.401	37.227	41.182	36.974

Table 2: $a = 1$ m, $b = 1$ m, $\sigma_{0x} = \sigma_{0y} = 5.0 \times 10^3$ KN/m^2, and $m = n = 1$.

M	$T_0 = 0.2$	$T_0 = 0.15$	$T_0 = 0.10$	$T_0 = 0.05$	$T_0 = 0.01$	$T_0 = 0.005$	$T_0 = 0.001$	$T_0 = 0.0001$	$T_0 \to 0$
Rad/s	544.651	439.21	344.231	270.86	242.227	286.325	240.962	240.949	240.948
Hz	86.684	69.902	54.786	43.109	38.552	45.570	38.350	38.348	38.348

Table 3: $a = 1$ m, $b = 1$ m, $\sigma_{0x} = \sigma_{0y} = 5.0 \times 10^3$ KN/m^2, and $T_0 = 0.10$ m.

Order	$m = 1, n = 1$	$m = 1, n = 2$	$m = 2, n = 1$	$m = 2, n = 2$	$m = 1, n = 3$	$m = 3, n = 1$	$m = 3, n = 3$
Rad/s	344.231	745.731	865.641	1089.30	1486.92	1799.04	2308.23
Hz	54.786	118.687	137.771	173.368	236.650	286.326	367.367

Table 4: $a = 1$ m, $b = 1$ m, $\sigma_{0x} = \sigma_{0y} = 5.0 \times 10^3$ KN/m^2, and $T_0 = 0.05$ m.

Order	$m = 1, n = 1$	$m = 1, n = 2$	$m = 2, n = 1$	$m = 2, n = 2$	$m = 1, n = 3$	$m = 3, n = 1$	$m = 3, n = 3$
Rad/s	270.86	499.358	546.065	688.462	880.846	1016.96	1317.63
Hz	43.109	79.475	86.909	109.572	140.191	161.854	209.707

Table 5: $a = 1$ m, $b = 1$ m, $\sigma_{0x} = \sigma_{0y} = 5.0 \times 10^3$ KN/m^2, and $T_0 = 0.01$ m.

Order	$m = 1, n = 1$	$m = 1, n = 2$	$m = 2, n = 1$	$m = 2, n = 2$	$m = 1, n = 3$	$m = 3, n = 1$	$m = 3, n = 3$
Rad/s	242.227	386.497	389.087	492.030	556.943	566.445	756.548
Hz	38.552	61.513	61.925	78.309	88.640	90.153	120.408

Table 6: $a = 1$ m, $b = 1$ m, and $\sigma_{0x} = \sigma_{0y} = 5.0 \times 10^3$ KN/m^2.

Order	$m = 1, n = 1$	$m = 1, n = 2$	$m = 2, n = 1$	$m = 2, n = 2$	$m = 1, n = 3$	$m = 3, n = 1$	$m = 3, n = 3$
Rad/s	240.949	380.974	380.974	481.898	538.779	538.779	722.847
Hz	38.348	60.634	60.634	76.697	85.749	85.749	115.045

calculated according to the large deflection theory. The frequency enlarged with the increase of the initial displacement, and the larger the initial displacement is, the larger the effect on the frequency is and vice versa. When the initial displacement approaches zero, the result is consistent with that obtained according to the small deflection theory.

The results from Table 3 to Table 6 are plotted as shown in Figure 2.

We can draw conclusions from the analysis of Figure 2. While order keeps unchanged, all frequency results based on the large deflection theory are larger than the corresponding ones based on the small deflection theory. The larger the initial displacement is, the larger the frequency is and vice versa. Specially, the smaller the initial displacement is, the closer the two results based on large and small deflection theories are.
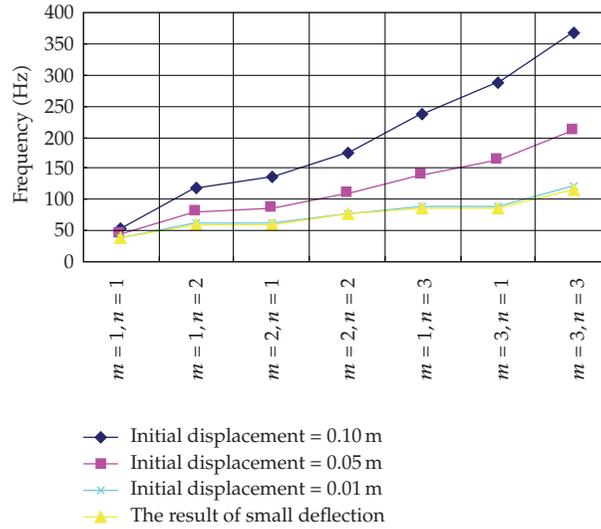
Figure 2: Comparative analysis of the results based on the large deflection theory and the small one.

The above conclusions are analyzed as follows.

(i) We considered the rigidity change caused by geometrical large deflection of membrane when calculating the frequency according to the large deflection theory. When the lateral displacement of membrane increased, the inner force increased, and the lateral rigidity also increased; then the vibration was quickened. Therefore, the membrane's vibration frequency will enlarge with the increase of initial displacement.

(ii) When initial displacement of membrane is very little, the change of lateral rigidity is also very little in the process of vibration, so it is negligible. In this case, the computational result based on the large deflection theory is very close to that based on the small deflection theory, which considers no changes of the lateral rigidity in the process of vibration.

5. Conclusions

(i) We applied the large deflection theory of membranes and D'Alembert's principle derived the governing vibration equations of rectangular membrane with four edges fixed, solved it, and obtained the power series formula of nonlinear vibration frequency of rectangular membranes with four edges fixed.

(ii) The frequency formula obtained in this paper is dependent on the initial conditions, as we considered the change of the lateral rigidity in the process of vibration. Therefore, the formula has reflected the geometric nonlinear characteristic of membrane structure's vibration. This is more tally with the actual situation and more reasonable than the result that is calculated according to the small deflection theory.

(iii) The nonlinear governing equation and the power series formula obtained in this paper provide some theoretical foundation for the measurement of pretension by frequency method; meanwhile, the results provide some theoretical foundation for the research of nonlinear vibration of membrane structures and the response solving of membrane structures under dynamic loads.

According to (3.20) (the frequency formula of the small deflection theory), by calculating the vibration frequency of the membrane, the results are listed in Table 6.

Acknowledgment

References

[1] B. Forster and M. Mollaert, *European Design Guide for Tensile Surface Structures*, China Machine Press, Beijing, China, 2006.

[2] Q.-L. Zhang, *Cable and Membrane Structures*, Tongji University Press, Shanghai, China, 2002.

[3] Z. J. Sun and Q. L. Zhang, "Simple and convenient membrane pretension measurement and testing experiments for measuring instrument," *Journal of Vibration, Measurement and Diagnosis*, vol. 25, no. 4, pp. 280–284, 2005.

[4] K. Sato, "Free vibration analysis of a composite elliptical membrane consisting of confocal elliptical parts," *Journal of Sound and Vibration*, vol. 34, no. 2, pp. 161–171, 1974.

[5] D. A. Vega, S. A. Vera, and P. A. A. Laura, "Fundamental frequency of vibration of rectangular membranes with an internal oblique support," *Journal of Sound and Vibration*, vol. 224, no. 4, pp. 780–783, 1999.

[6] S. W. Kang and J. M. Lee, "Free vibration analysis of composite rectangular membranes with an oblique interface," *Journal of Sound and Vibration*, vol. 251, no. 3, pp. 505–517, 2002.

[7] S. W. Kang, "Free vibration analysis of composite rectangular membranes with a bent interface," *Journal of Sound and Vibration*, vol. 272, no. 1-2, pp. 39–53, 2004.

[8] C. Shin, W. Kim, and J. Chung, "Free in-plane vibration of an axially moving membrane," *Journal of Sound and Vibration*, vol. 272, no. 1-2, pp. 137–154, 2004.

[9] S.-L. Chen and Z.-L. Zheng, "Large deformation of circular membrane under the concentrated force," *Applied Mathematics and Mechanics (English edition)*, vol. 24, no. 1, pp. 28–31, 2003.

[10] J. Wang, "Nonlinear free vibration of the circular plate with large deflection," *Journal of South China University of Technology*, vol. 29, no. 8, pp. 4–6, 2001.

[11] Y. Su and C.-H. Wang, "A new method for calculating ellipse integral," *Journal of Changchun University*, vol. 9, no. 3, pp. 17–22, 1999.

Research Article

Modified Jacobian Newton Iterative Method: Theory and Applications

Jürgen Geiser

Department of Mathematics, Humboldt-Universität zu Berlin, Unter den Linden 6,
D-10099 Berlin, Germany

Correspondence should be addressed to Jürgen Geiser, geiser@mathematik.hu-berlin.de

This article proposes a new approach to the construction of a linearization method based on the iterative operator-splitting method for nonlinear differential equations. The convergence properties of such a method are studied. The main features of the proposed idea are the linearization of nonlinear equations and the application of iterative splitting methods. We present an iterative operator-splitting method with embedded Newton methods to solve nonlinearity. We confirm with numerical applications the effectiveness of the proposed iterative operator-splitting method in comparison with the classical Newton methods. We provide improved results and convergence rates.

1. Introduction

In this paper we propose a modified Jacobian-Newton iterative method to solve nonlinear differential equations. In the first paper we concentrate on ordinary differential equations, but numerical results are also obtained for partial differential equations. Basic studies of the operator-splitting methods are found in [1, 2]. Further important research was carried out to obtain a higher order for the splitting methods (see [3]). For this reason, the iterative splitting methods became more important for linear and nonlinear differential equations, while simple increasing of iteration steps affects the order of the scheme (see [4]). An interesting topic is Newton's methods for nonlinear problems with specifications for numerical implementations, (see [5]). Efficient modifications of Newton's methods, with regard to the computation of the Jacobian matrices are discussed in [6]. In our paper we discuss the benefit of the combination of splitting and linearization methods (see theoretical frameworks [7, 8]).

The outline of the paper is as follows. For our mathematical model we describe the convection-diffusion-reaction equation in Section 2. The fractional splitting is introduced in Section 3. We present the iterative splitting methods in Section 4. Section 5 discusses the Newton methods and their modifications. In Section 6 we present the numerical results from the solution of selected model problems. We end the article in Section 7 with a conclusion and comments.

2. Mathematical Model

The motivation for the study presented below originates from a computational simulation of heat-transfer [9] and convection-diffusion-reaction-equations [10–13].

In the present paper we concentrate on ordinary differential equations, given as

$$\partial_t u(t) = A(u(t))\, u(t) + B(u(t))u(t), \quad t \in (0, T), \tag{2.1}$$

where the initial condition is $u(0) = u_0$. The operators $A(u)$ and $B(u)$ can be spatially discretized operators, that is, they can correspond to the discretized in space convection and diffusion operators (matrices). In the following, we deal with bounded nonlinear operators.

The aim of this paper is to present a new method based on Newton and iterative schemes.

In the next section we discuss the decoupling of the time-scale with a first-order fractional splitting method.

3. Fractional-Splitting Methods of First-Order for Linear Equations

First we describe the simplest operator-splitting, which is called *sequential operator-splitting*, for the following linear system of ordinary differential equations:

$$\partial_t u(t) = Au(t) + Bu(t), \quad t \in (0, T), \tag{3.1}$$

where the initial condition is $u(0) = u_0$. The operators A and B are linear and bounded operators in a Banach space (see also Section 2).

The sequential operator-splitting method is introduced as a method that solves two subproblems sequentially, where the different subproblems are connected via the initial conditions. This means that we replace the original problem (3.1) with the subproblems

$$\frac{\partial u^*(t)}{\partial t} = Au^*(t), \quad \text{with } u^*(t^n) = u^n,$$

$$\frac{\partial u^{**}(t)}{\partial t} = Bu^{**}(t), \quad \text{with } u^{**}(t^n) = u^*(t^{n+1}), \tag{3.2}$$

where the splitting time-step is defined as $\tau_n = t^{n+1} - t^n$. The approximated solution is $u^{n+1} = u^{**}(t^{n+1})$.

Clearly, the replacement of the original problem with the subproblems usually results in an error, called *splitting error*. The splitting error of the sequential operator-splitting method can be derived as (cf., e.g., [1, 2]).

$$\rho_n = \frac{1}{\tau_n} \left(\exp\left(\tau_n(A+B)\right) - \exp\left(\tau_n B\right) \exp\left(\tau_n A\right) \right) u(t^n)$$

$$= \begin{cases} 0, & \text{for } [A, B] = 0, \\ O(\tau_n), & \text{for } [A, B] \neq 0, \end{cases} \tag{3.3}$$

where $[A, B] := AB - BA$ is the commutator of A and B. Consequently, the splitting error is $O(\tau_n)$ when the operators A and B do not commute, otherwise the method is exact. Hence, by definition, the sequential operator-splitting is called the *first-order splitting method*.

4. The Iterative Splitting Method

The following algorithm is based on the iteration with fixed splitting discretization step-size τ. On the time interval $[t^n, t^{n+1}]$ we solve the following subproblems consecutively for $i = 0, 2, \ldots 2m$:

$$\frac{\partial u_i(x, t)}{\partial t} = A u_i(x, t) + B u_{i-1}(x, t), \quad \text{with } u_i(t^n) = u^n,$$

$$u_0(x, t^n) = u^n, \qquad u_{-1} = 0,$$

$$u_i(x, t) = u_{i-1}(x, t) = u_1, \quad \text{on } \partial\Omega \times (0, T), \tag{4.1}$$

$$\frac{\partial u_{i+1}(x, t)}{\partial t} = A u_i(x, t) + B u_{i+1}(x, t), \quad \text{with } u_{i+1}(x, t^n) = u^n,$$

$$u_i(x, t) = u_{i-1}(x, t) = u_1, \quad \text{on } \partial\Omega \times (0, T),$$

where u^n is the known split approximation at the time level $t = t^n$ (see [14]).

Remark 4.1. We can generalize the iterative splitting method to a multi-iterative splitting method by introducing new splitting operators, for example, spatial operators. Then we obtain multi-indices to control the splitting process; each iterative splitting method can be solved independently, while connecting with further steps to the multi-splitting methods. In the following we introduce the multi-iterative splitting method for a combined time-space splitting method.

5. The Modified Jacobian-Newton Methods and Fixpoint-Iteration Methods

In this section we describe the modified Jacobian-Newton methods and Fixpoint-iteration methods.

We propose for weak nonlinearities, for example, quadratic nonlinearity, the fixpoint iteration method, where our iterative operator splitting method is one, see [4]. For stronger nonlinearities, for example, cubic or higher order polynomial nonlinearities, the modified Jacobian method with embedded iterative-splitting methods is suggested.

The point of embedding the splitting methods into the Newton methods is to decouple the equation systems into simpler equations. Such simple equation systems can be solved with scalar Newton methods.

5.1. The Altered Jacobian-Newton Iterative Methods with Embedded Sequential Splitting Methods

We confine our attention to time-dependent partial differential equations of the form

$$\frac{dc}{dt} = A(c(t))c(t) + B(c(t))c(t), \quad \text{with } c(t^n) = c^n, \tag{5.1}$$

where $A(c), B(c) : \mathbf{X} \to \mathbf{X}$ are linear and densely defined in the real Banach space $\mathbf{X}$, involving only spatial derivatives of c, see [8]. We assume also that we have a weak nonlinear operator with $A(c)c = \lambda_1 c$ and $B(c)c = \lambda_2 c$, where λ_1 and λ_2 are constant factors.

In the following we discuss the embedding of a sequential splitting method into the Newton method.

The altered Jacobian-Newton iterative method with an embedded iterative splitting method is given as follows.

Newton's Method

$F(c) = dc/dt - A(c(t))c(t) - B(c(t))c(t)$ and we can compute $c^{(k+1)} = c^{(k)} - D(F(c^{(k)}))^{-1}F(c^{(k)})$, where $D(F(c))$ is the Jacobian matrix and $k = 0, 1, \ldots$.

We stop the iterations when we obtain: $|c^{(k+1)} - c^{(k)}| \leq$ err, where err is an error bound, for example, err $= 10^{-4}$.

We assume the spatial discretization, with spatial grid points, $i = 1, \ldots, m$ and obtain the differential equation system:

$$F(c) = \begin{pmatrix} F(c_1) \\ F(c_2) \\ \vdots \\ F(c_m) \end{pmatrix}, \tag{5.2}$$

where $c = (c_1, \ldots, c_m)T$ and m is the number of spatial grid points.

The Jacobian matrix for the equation system is given as:

$$DF(c) = \begin{pmatrix} \dfrac{\partial F(c_1)}{c_1} & \dfrac{\partial F(c_1)}{c_2} & \cdots & \dfrac{\partial F(c_1)}{c_m} \\ \dfrac{\partial F(c_2)}{c_1} & \dfrac{\partial F(c_2)}{c_2} & & \dfrac{\partial F(c_2)}{c_m} \\ \vdots & & & \\ \dfrac{\partial F(c_m)}{c_1} & \dfrac{\partial F(c_m)}{c_2} & \cdots & \dfrac{\partial F(c_m)}{c_m} \end{pmatrix}, \tag{5.3}$$

where $c = (c_1, \ldots, c_m)$.

The modified Jacobian is given as:

$$DF(c) = \begin{pmatrix} \dfrac{\partial F(c_1)}{c_1} + F(c_1) & \dfrac{\partial F(c_1)}{c_2} & \cdots & \dfrac{\partial F(c_1)}{c_m} \\ \dfrac{\partial F(c_2)}{c_1} & \dfrac{\partial F(c_2)}{c_2} F(c_2) & \cdots & \dfrac{\partial F(c_2)}{c_m} \\ \vdots & & & \\ \dfrac{\partial F(c_m)}{c_1} & \dfrac{\partial F(c_m)}{c_2} & \cdots & \dfrac{\partial F(c_m)}{c_m} + F(c_m) \end{pmatrix}, \tag{5.4}$$

where $c = (c_1, \ldots, c_n)$.

By embedding the sequential splitting method we obtain the following algorithm. We decouple into two equation systems:

$$F_1(u_1) = \partial_t u_1 - A(u_1)u_1 = 0 \quad \text{with } u_1(t^n) = c^n,$$
$$F_2(u_2) = \partial_t u_2 - B(u_2)u_2 = 0 \quad \text{with } u_2(t^n) = u_1(t^{n+1}), \tag{5.5}$$

where the results of the methods are $c(t^{n+1}) = u_2(t^{n+1})$, and $u_1 = (u_{11}, \ldots, u_{1n})$, $u_2 = (u_{21}, \ldots, u_{2n})$.

Thus we have to solve two Newton methods, each in one equations system. The contribution is to reduce the Jacobian matrix into a diagonal entry, for example, with a weighted Newton method, see [15]. The splitting method with embedded Newton method is given as follows.

Algorithm 5.1. We assume the spatial operators A and B are discretized, for example, finite difference or finite element methods; further all initial conditions and boundary conditions are discrete given. Then we can apply the Newton's method in its discrete form as:

$$u_1^{(k+1)} = u_1^{(k)} - D\left(F_1(u_1^{(k)})\right)^{-1}\left(\partial_t u_1^{(k)} - A(u_1^{(k)})u_1^{(k)}\right),$$

$$\text{with } D\left(F_1(u_1^{(k)})\right) = \frac{\partial}{\partial u_1^{(k)}}\left(\partial_t u_1^{(k)} - A(u_1^{(k)}) - \frac{\partial A(u_1^{(k)})}{\partial u_1^{(k)}} u_1^{(k)}\right),$$

$$u_1^{(k)}(t^n) = c^n, \quad k = 0, 1, 2, \ldots, K,$$

$$u_2^{(l+1)} = u_2^{(l)} - D\left(F_2(u_2^{(l)})\right)^{-1}\left(\partial_t u_2^{(l)} - B(u_2^{(l)})u_2^{(l)}\right),$$

$$\text{with } D\left(F_2(u_2^{(l)})\right) = \frac{\partial}{\partial u_1^{(k)}}\left(\partial_t u_2^{(k)} - B(u_2^{(l)}) - \frac{\partial B(u_2^{(l)})}{\partial u_2^{(l)}} u_2^{(l)}\right),$$

$$u_2^{(l)}(t^n) = u_1^K(t^{n+1}), \quad l = 0, 1, 2, \ldots, L. \tag{5.6}$$

where k and l are the iteration indices, K and L the maximal iterative steps for each part of the Newton's method. The maximal iterative steps allow us to have at least an error of:

$$
\begin{aligned}
\left| u_1^{(K)(t^{n+1})} - u_1^{(K-1)(t^{n+1})} \right| &\leq \text{ err}, \\
\left| u_2^{(L)(t^{n+1})} - u_2^{(L-1)(t^{n+1})} \right| &\leq \text{ err},
\end{aligned}
\tag{5.7}
$$

where err is the error bound, for example, err $= 10^{-6}$.

The approximated solution is given as:

$$
u\left(t^{n+1}\right) = u_2^{(L)(t^{n+1})}.
\tag{5.8}
$$

For the improvement method, we can apply the weighted Newton method. We try to skip the delicate outer diagonals in the Jacobian matrix and apply:

$$
u_1^{(k+1)} = u_1^{(k)} - \left(D\left(F_1\left(u_1^{(k)}\right)\right) + \delta_1\left(u_1^{(k)}\right) \right)^{-1} \left(F_1\left(u_1^{(k)}\right) + \epsilon\, u_1^{(k)} \right),
\tag{5.9}
$$

where the function δ can be applied as a scalar, for example, $\delta = 10^{-6}$, and the same with ϵ. It is important to ensure that δ is small enough to preserve the convergence.

Remark 5.2. If we assume that we discretize (5.5) with the backward-Euler method, for example,

$$
\begin{aligned}
F_1\left(u_1\left(t^{n+1}\right)\right) &= u_1\left(t^{n+1}\right) - u_1\left(t^n\right) - \Delta t A\left(u_1\left(t^{n+1}\right)\right)u_1\left(t^{n+1}\right) = 0 \quad \text{with } u_1\left(t^n\right) = c^n, \\
F_2\left(u_2\right) &= u_2\left(t^{n+1}\right) - u_2\left(t^n\right) - \Delta t B\left(u_2\left(t^{n+1}\right)\right)u_2\left(t^{n+1}\right) = 0 \quad \text{with } u_2\left(t^n\right) = u_1\left(t^{n+1}\right),
\end{aligned}
\tag{5.10}
$$

then we obtain the derivations $D(F_1(u_1(t^{n+1})))$ and $D(F_2(u_2(t^{n+1})))$

$$
\begin{aligned}
D\left(F_1\left(u_1\left(t^{n+1}\right)\right)\right) &= 1 - \Delta t\left(A\left(u_1\left(t^{n+1}\right)\right) + \frac{\partial A\left(u_1\left(t^{n+1}\right)\right)}{\partial u_1\left(t^{n+1}\right)} u_1\left(t^{n+1}\right) \right), \\
D\left(F_2\left(u_2\right)\right) &= 1 - \Delta t\left(B\left(u_2\left(t^{n+1}\right)\right) + \frac{\partial B\left(u_2\left(t^{n+1}\right)\right)}{\partial u_2\left(t^{n+1}\right)} u_2\left(t^{n+1}\right) \right).
\end{aligned}
\tag{5.11}
$$

We can apply the equation (5.9) analogously $u_2^{(l+1)}$.

5.2. Iterative Operator-Splitting Method as a Fixpoint Scheme

The iterative operator-splitting method is used as a fixpoint scheme to linearize the nonlinear operators, see [4, 16].

We confine our attention to time-dependent partial differential equations of the form:

$$
\frac{du}{dt} = A(u(t))u(t) + B(u(t))u(t), \quad \text{with } u(t^n) = c^n,
\tag{5.12}
$$

where $A(u), B(u) : \mathbf{X} \rightarrow \mathbf{X}$ are linear and densely defined in the real Banach space $\mathbf{X}$, involving only spatial derivatives of c, see [8]. In the following we discuss the standard iterative operator-splitting methods as a fixpoint iteration method to linearize the operators.

We split our nonlinear differential equation (5.12) by applying:

$$
\begin{aligned}
\frac{du_i(t)}{dt} &= A(u_{i-1}(t))u_i(t) + B(u_{i-1}(t))u_{i-1}(t), \quad \text{with } u_i(t^n) = c^n, \\
\frac{du_{i+1}(t)}{dt} &= A(u_{i-1}(t))u_i(t) + B(u_{i-1}(t))u_{i+1}, \quad \text{with } u_{i+1}(t^n) = c^n,
\end{aligned}
\tag{5.13}
$$

where the time-step is $\tau = t^{n+1} - t^n$. The iterations are $i = 1, 3, \ldots, 2m + 1$. $u_0(t) = c_n$ is the starting solution, where we assume the solution c^{n+1} is near c^n, or $u_0(t) = 0$. So we have to solve the local fixpoint problem. c^n is the known split approximation at the time level $t = t^n$.

The split approximation at time level $t = t^{n+1}$ is defined as $c^{n+1} = u_{2m+2}(t^{n+1})$. We assume the operators $A(u_{i-1}), B(u_{i-1}) : \mathbf{X} \rightarrow \mathbf{X}$ to be linear and densely defined on the real Banach space $\mathbf{X}$, for $i = 1, 3, \ldots, 2m + 1$.

Here the linearization is done with respect to the iterations, such that $A(u_{i-1}), B(u_{i-1})$ are at least non-dependent operators in the iterative equations, and we can apply the linear theory.

The linearization is at least in the first equation $A(u_{i-1}) \approx A(u_i)$, and in the second equation $B(u_{i-1}) \approx B(u_{i+1})$.

We have

$$
\| A(u_{i-1}(t^{n+1}))u_i(t^{n+1}) - A(u^{n+1})u(t^{n+1}) \| \leq \epsilon,
\tag{5.14}
$$

with sufficient iterations $i = \{1, 3, \ldots, 2m + 1\}$.

Remark 5.3. The linearization with the fixpoint scheme can be used for smooth or weak nonlinear operators, otherwise we lose the convergence behavior, while we did not converge to the local fixpoint, see [4].

The second idea is based on the Newton method.

5.3. Jacobian-Newton Iterative Method with Embedded Operator-Splitting Method

The Newton method is used to solve the nonlinear parts of the iterative operator-splitting method (see the linearization techniques in [4, 17]).

Newton Method

The function is given as:

$$
F(c) = \frac{\partial c}{\partial t} - A(c(t))c(t) - B(c(t))c(t) = 0,
\tag{5.15}
$$

The iteration can be computed as:

$$c^{(k+1)} = c^{(k)} - D(F(c^{(k)}))^{-1} F(c^{(k)}), \tag{5.16}$$

where $D(F(c))$ is the Jacobian matrix and $k = 0, 1, \ldots$ and $c = (c_1, \ldots, c_m)$ is the solution vector of the spatial discretized nonlinear equation.

We then have to apply the iterative operator-splitting method and obtain:

$$F_1(u_i) = \partial_t u_i - A(u_i)u_i - B(u_{i-1})u_{i-1} = 0, \quad \text{with } u_i(t^n) = c^n,$$
$$F_2(u_{i+2}) = \partial_t u_{i+1} - A(u_i)u_i - B(u_{i+1})u_{i+1} = 0, \quad \text{with } u_{i+1}(t^n) = c^n, \tag{5.17}$$

where the time-step is $\tau = t^{n+1} - t^n$. The iterations are $i = 1, 3, \ldots, 2m+1$. $c_0(t) = 0$ is the starting solution and c^n is the known split approximation at the time-level $t = t^n$. The results of the methods are $c(t^{n+1}) = u_{2m+2}(t^{n+1})$.

Thus we have to solve two Newton methods and the contribution will be to reduce the Jacobian matrix, for example, skip the diagonal entries. The splitting method with the embedded Newton method is given as:

$$u_i^{(k+1)} = u_i^{(k)} - D(F_1(u_i^{(k)}))^{-1} (\partial_t u_i^{(k)} - A(u_i^{(k)})u_i^{(k)} - B(u_{i-1}^{(k)})u_{i-1}^{(k)}),$$

$$\text{with } D(F_1(u_i^{(k)})) = -\left(A(u_i^{(k)}) + \frac{\partial A(u_i^{(k)})}{\partial u_i^{(k)}} u_i^{(k)}\right), \quad k = 0, 1, 2, \ldots, K, \quad \text{with } u_i(t^n) = c^n,$$

$$u_{i+1}^{(l+1)} = u_{i+1}^{(l)} - D(F_2(u_{i+1}^{(l)}))^{-1} (\partial_t u_{i+1}^{(l)} - A(u_i^{(k)})u_i^{(k)} - B(u_{i+1}^{(k)})u_{i+1}^{(k)} c_2^{(l)}),$$

$$\text{with } D(F_2(u_{i+1}^{(l)})) = -\left(B(u_{i+1}^{(l)}) + \frac{\partial B(u_{i+1}^{(l)})}{\partial u_{i+1}^{(l)}} u_{i+1}^{(l)}\right), \quad l = 0, 1, 2, \ldots, L, \quad \text{with } u_{i+1}(t^n) = c^n, \tag{5.18}$$

where the time-step is $\tau = t^{n+1} - t^n$. The iterations are: $i = 1, 3, \ldots, 2m + 1$. $c_0(t) = 0$ is the starting solution and c^n is the known split approximation at the time-level $t = t^n$. The results of the methods are $c(t^{n+1}) = u_{2m+2}(t^{n+1})$.

For the improvement by skipping the delicate outer diagonals in the Jacobian matrix, we apply $u_i^{(k+1)} = u_i^{(k)} - (D(F_1(u_i^{(k)})) + \delta_1(u_i^{(k)}))^{-1} (F_1(u_i^{(k)}) + \epsilon u_i^{(k)})$, and analogously $u_{i+1}^{(l+1)}$.

Remark 5.4. For the iterative operator-splitting method with the Newton iteration we have two iteration procedures. The first iteration is the Newton method to compute the solution of the nonlinear equations, and the second iteration is the iterative splitting method, which computes the resulting solution of the coupled equation systems. The embedded method is used for strong nonlinearities.

6. Numerical Results

In this section, we present the numerical results for nonlinear differential equation using several variations of the proposed Newton and iterative schemes as solvers.

6.1. First Numerical Example: Bernoulli Equation

As a nonlinear differential example, we choose the Bernoulli equation:

$$\frac{\partial u(t)}{\partial t} = (\lambda_1 + \lambda_3)u(t) + (\lambda_2 + \lambda_4)(u(t))^p, \quad t \in [0, T], \text{ with } u(0) = 1, \tag{6.1}$$

where the analytical solution can be derived as (see also [16]):

$$u(t) = \exp\left((\lambda_1 + \lambda_3)t\right)\left[-\frac{\lambda_2 + \lambda_4}{\lambda_1 + \lambda_3}\exp\left((\lambda_1 + \lambda_3)(p - 1)t\right) + c\right]^{1/(1-p)}. \tag{6.2}$$

Using $u(0) = 1$ we find that $c = 1 + (\lambda_2 + \lambda_4)/(\lambda_1 + \lambda_3)$, so

$$u(t) = \exp\left((\lambda_1 + \lambda_3)t\right)\left\{1 + \frac{\lambda_2 + \lambda_4}{\lambda_1 + \lambda_3}\left[1 - \exp\left((\lambda_1 + \lambda_3)(p - 1)t\right)\right]\right\}^{1/(1-p)}. \tag{6.3}$$

We choose $p = 2$, $\lambda_1 = -1$, $\lambda_2 = -0.5$, $\lambda_3 = -100$, $\lambda_4 = -20$ and, for example, $\Delta t = 10^{-2}$. The analytical solutions can be given as:

$$u(t)^{1-p} = u_0 \exp\left((1 - p)(\lambda_1 + \lambda_3)t\right) + \frac{\lambda_2 + \lambda_4}{\lambda_1 + \lambda_3}\left(\exp\left((1 - p)(\lambda_1 + \lambda_3)t\right) - 1\right). \tag{6.4}$$

We divide the time interval $[0, T]$, with $T = 1$, in n intervals with length $\tau_n = T/n$.

(1) The sequential operator-splitting method with analytical solutions is given as follows.

We apply the quasilinear iterative operator-splitting method:

$$\frac{du_1(t)}{dt} = A(u_1(t))u_1(t), \quad \text{with } u_1(t^n) = u^n,$$

$$\frac{du_2(t)}{dt} = B(u_2(t))u_2, \quad \text{with } u_2(t^n) = u_1^{n+1}, \tag{6.5}$$

with the nonlinear operators $A(u)u = \lambda_1 u(t) + \lambda_2(u(t))^{p-1}u$, $B(u)u = \lambda_3 u(t) + \lambda_4(u(t))^{p-1}u$. The result is given as $u_2(t^{n+1}) = u^{n+1}$.

We apply the Newton method and discretize the operators with time discretization methods such as backward-Euler or higher Runge-Kutta methods.

The analytical result for each equation part is given as:

$$u_1(t)^{1-p} = u(t^n) \exp\left((1-p)(\lambda_1)t\right) + \frac{\lambda_2}{\lambda_1}\left(\exp\left((1-p)(\lambda_1)t\right) - 1\right),$$

$$u_2(t^{n+1})^{1-p} = u_1(t^{n+1}) \exp\left((1-p)(\lambda_3)t\right) + \frac{\lambda_4}{\lambda_3}\left(\exp\left((1-p)(\lambda_3)t\right) - 1\right),$$

(6.6)

where the result is given as $u(t^{n+1}) = u_2(t^{n+1})$.

We can apply the simpler equations and solve the sequential operator-splitting method.

(2) The sequential operator-splitting method with embedded Newton method is given as follows.

We apply the quasilinear iterative operator-splitting method:

$$\frac{du_1(t)}{dt} = A(u_1(t))u_1(t), \quad \text{with } u_1(t^n) = u^n,$$

$$\frac{du_2(t)}{dt} = B(u_2(t))u_2, \quad \text{with } u_2(t^n) = u_1^{n+1},$$

(6.7)

with the nonlinear operators $A(u)u = \lambda_1 u(t) + \lambda_2(u(t))^{p-1}u$, $B(u)u = \lambda_3 u(t) + \lambda_4(u(t))^{p-1}u$. The result is given as $u_2(t^{n+1}) = u^{n+1}$.

We apply the Newton method and discretize the operators with time discretization methods such as backward-Euler or higher Runge-Kutta methods.

The splitting method with embedded Newton's method is given as

$$u_1^{(k+1)} = u_1^{(k)} - D\left(F_1(u_1^{(k)})\right)^{-1}\left(\partial_t u_1^{(k)} - A(u_1^{(k)})u_1^{(k)}\right),$$

$$\text{with } D\left(F_1(u_1^{(k)})\right) = -\left(A(u_1^{(k)}) + \frac{\partial A(u_1^{(k)})}{\partial u_1^{(k)}}u_1^{(k)}\right),$$

$$u_1^{(k)}(t^n) = c^n, \quad k = 0, 1, 2, \ldots, K,$$

$$u_2^{(l+1)} = u_2^{(l)} - D\left(F_2(u_2^{(l)})\right)^{-1}\left(\partial_t u_2^{(l)} - B(u_2^{(l)})u_2^{(l)}\right),$$

$$\text{with } D\left(F_2(u_2^{(l)})\right) = -\left(B(u_2^{(l)}) + \frac{\partial B(u_2^{(l)})}{\partial u_2^{(l)}}u_2^{(l)}\right),$$

$$u_2^{(l)}(t^n) = u_1^K(t^{n+1}), \quad l = 0, 1, 2, \ldots, L,$$

(6.8)

where we discretize the equations and obtain the discretized operators:

$$\partial_t u_1^{(k)} - A(u_1^{(k)})u_1^{(k)} = 0,$$

(6.9)

as

$$F_1\big(u_1\big(t^{n+1}\big)\big)u_1^{(k)}\big(t^{n+1}\big) - u_1\big(t^n\big) - \Delta t A\big(u_1^{(k)}\big(t^{n+1}\big)\big)u_1^{(k)}\big(t^{n+1}\big) = 0, \tag{6.10}$$

where we have the initialization of the Newton's method as $u_1^{(0)}\big(t^{n+1}\big) = 0$ or $u_1^{(0)}\big(t^{n+1}\big) = u_1\big(t^n\big)$.

For the second iteration equation we have

$$\partial_t u_2^{(l)} - B\big(u_2^{(l)}\big)u_2^{(l)}, \tag{6.11}$$

as

$$F_2\big(u_2\big(t^{n+1}\big)\big) = u_2^{(l)}\big(t^{n+1}\big) - u_2\big(t^n\big) - \Delta t B\big(u_2^{(l)}\big(t^{n+1}\big)\big)u_2^{(l)}\big(t^{n+1}\big) = 0, \tag{6.12}$$

where we have the initialization of the Newton's method as $u_2^{(0)}\big(t^{n+1}\big) = 0$ or $u_2^{(0)}\big(t^{n+1}\big) = u_1\big(t^n\big)$.

The derivations are given as:

$$D\big(F_1\big(u_1\big(t^{n+1}\big)\big)\big) = 1 - \Delta t\left(A\big(u_1\big(t^{n+1}\big)\big) + \frac{\partial A\big(u_1\big(t^{n+1}\big)\big)}{\partial u_1\big(t^{n+1}\big)} u_1\big(t^{n+1}\big)\right),$$
$$D\big(F_2\big(u_2\big)\big) = 1 - \Delta t\left(B\big(u_2\big(t^{n+1}\big)\big) + \frac{\partial B\big(u_2\big(t^{n+1}\big)\big)}{\partial u_2\big(t^{n+1}\big)} u_2\big(t^{n+1}\big)\right). \tag{6.13}$$

(3) The standard iterative operator-splitting method is given as follows.
We apply the quasilinear iterative operator-splitting method:

$$\frac{du_i(t)}{dt} = A\big(u_{i-1}(t)\big)u_i(t) + B\big(u_{i-1}(t)\big)u_{i-1}(t), \quad \text{with } u_i\big(t^n\big) = u^n,$$
$$\frac{du_{i+1}(t)}{dt} = A\big(u_{i-1}(t)\big)u_i(t) + B\big(u_{i-1}(t)\big)u_{i+1}, \quad \text{with } u_{i+1}\big(t^n\big) = u^n, \tag{6.14}$$

with the nonlinear operators $A(u)u = \lambda_1 u(t) + \lambda_2 (u(t))^{p-1} u$, $B(u)u = \lambda_3 u(t) + \lambda_4 (u(t))^{p-1} u$. The initialization of the fixpoint iteration is $u_0 = u^n$ or $u_0 = 0$ with $A(u_0) = \lambda_1$ and $B(u_0) = \lambda_3$.

For the iterations we can apply the analytical solution of each equation:

$$u_i(t) = u^n \exp\big(A\big(u_{i-1}(t)\big)t\big) + \big(A\big(u_{i-1}(t)\big)\big)^{-1}\big(B\big(u_{i-1}(t)\big)u_{i-1}(t)\big)\big(1 - \exp\big(A\big(u_{i-1}(t)\big)t\big)\big),$$
$$u_{i+1}(t) = u^n \exp\big(B\big(u_{i-1}(t)\big)t\big) + \big(B\big(u_{i-1}(t)\big)\big)^{-1}\big(A\big(u_{i-1}(t)\big)u_{i-1}(t)\big)\big(1 - \exp\big(B\big(u_{i-1}(t)\big)t\big)\big). \tag{6.15}$$

Further the iterative steps can be done.
(4) The Newton iterative method with embedded iterative operator-splitting method is given as follows.

We apply the quasilinear iterative operator-splitting method:

$$\frac{du_i(t)}{dt} = A(u_i(t))u_i(t) + B(u_{i-1}(t))u_{i-1}(t), \quad \text{with } u_i(t^n) = u^n,$$

$$\frac{du_{i+1}(t)}{dt} = A(u_i(t))u_i(t) + B(u_{i+1}(t))u_{i+1}, \quad \text{with } u_{i+1}(t^n) = u^n,$$

(6.16)

with the nonlinear operators $A(u)u = \lambda_1 u(t) + \lambda_2 (u(t))^{p-1}u$, $B(u)u = \lambda_3 u(t) + \lambda_4 (u(t))^{p-1}u$. The initialization of the fixpoint iteration is $u_0(t^{n+1}) = u^n$ or $u_0(t^{n+1}) = 0$.

The discretization of the nonlinear ordinary differential equation is performed with higher-order Runge-Kutta methods.

The Newton method is applied as:

$$u_i^{(k+1)} = u_i^{(k)} - D(F_1(u_i^{(k)}))^{-1} (\partial_t u_i^{(k)} - A(u_i^{(k)})u_i^{(k)} - B(u_{i-1})u_{i-1}),$$

$$D(F_1(u_i^{(k)})) = -\left(A(u_i^{(k)}) + \frac{\partial A(u_i^{(k)})}{\partial u_i^{(k)}} u_i^{(k)} \right), \quad k = 0, 1, 2, \ldots, K,$$

$$u_i(t^n) = c^n,$$

$$u_i(t^{n+1}) = u_i(t^{n+1})^{K+1}, \quad \text{where } |u_i(t^{n+1})^{K+1} - u_i(t^{n+1})^K| \le \text{err},$$

$$u_{i+1}^{(l+1)} = u_{i+1}^{(l)} - D(F_2(u_{i+1}^{(l)}))^{-1} (\partial_t u_{i+1}^{(l)} - A(u_i)u_i - B(u_{i+1}^{(l)})u_{i+1}^{(l)} c_2^{(l)}),$$

$$D(F_2(u_{i+1}^{(l)})) = -\left(B(u_{i+1}^{(l)}) + \frac{\partial B(u_{i+1}^{(l)})}{\partial u_{i+1}^{(l)}} u_{i+1}^{(l)} \right), \quad l = 0, 1, 2, \ldots, L,$$

$$u_{i+1}(t^n) = c^n,$$

$$u_{i+1}(t^{n+1}) = u_{i+1}(t^{n+1})^{L+1}, \quad \text{where } |u_{i+1}(t^{n+1})^{L+1} - u_{i+1}(t^{n+1})^L| \le \text{err}.$$

(6.17)

Here the time-step is $\tau = t^{n+1} - t^n$. The iterations are $i = 1, 3, \ldots, 2m + 1$. $u_0(t) = 0$ is the starting solution and c^n is the known split approximation at the time-level $t = t^n$. The results of the methods are $u(t^{n+1}) = u_{2m+2}(t^{n+1})$.

We apply the discretization methods for the iteration steps.

We discretize the equations

$$\partial_t u_i^{(k)} - A(u_i^{(k)})u_i^{(k)} - B(u_{i-1})u_{i-1} = 0,$$

(6.18)

as

$$F_1(u_i^{(k)}(t^{n+1})) = u_i^{(k)}(t^{n+1}) - u_i(t^n) - \Delta t (A(u_i^{(k)}(t^{n+1}))u_i^{(k)}(t^{n+1}) + B(u_{i-1}(t^{n+1}))u_{i-1}(t^{n+1})),$$

(6.19)

where we have the initialization of the Newton's method as $u_i^{(0)}(t^{n+1}) = 0$ or $u_i^{(0)}(t^{n+1}) = u(t^n)$.

Table 1: Numerical results for the Bernoulli equation with sequential operator-splitting method.

Time partitions	Approximated solution	Error
1	0.0000000000	6.620107e-044
2	0.0000000000	5.874983e-023
3	0.0000000000	6.351951e-016
4	0.0000000000	1.917794e-012
5	0.0000000002	2.232302e-010
10	0.0000023626	2.362646e-006
50	0.0015822287	1.582229e-003

Table 2: Numerical results for the Bernoulli equation with sequential operator-splitting method with embedded Newton's method.

Time partitions	Approximated solution	Error
1	0.4705129443	4.705129e-001
4	0.0546922483	5.469225e-002
5	0.0269954419	2.699544e-002
10	0.0008034713	8.034713e-004
15	0.0000000000	1.137634e-044
100	0.0000000000	1.137634e-044

For the second iteration equation we have:

$$\partial_t u_2^{(l)} - A(u_i)u_i - B(u_2^{(l)})u_2^{(l)} = 0, \tag{6.20}$$

as

$$F_2\big(u_{i+1}^{(l)}(t^{n+1})\big) = u_{i+1}^{(l)}(t^{n+1}) - u_{i+1}(t^n) - \Delta t\big(A(u_i(t^{n+1}))u_i(t^{n+1}) + B(u_{i+1}^{(l)}(t^{n+1}))u_{i+1}^{(l)}(t^{n+1})\big), \tag{6.21}$$

where we have the initialization of the Newton's method as $u_{i+1}^{(0)}(t^{n+1}) = 0$ or $u_{i+1}^{(0)}(t^{n+1}) = u(t^n)$.

The derivations are given as:

$$\begin{aligned} D\big(F_1\big(u_i^{(k)}(t^{n+1})\big)\big) &= 1 - \Delta t\left(A(u_{i+1}^{(k)}(t^{n+1})) + \frac{\partial A(u_{i+1}^{(k)}(t^{n+1}))}{\partial u_{i+1}^{(k)}(t^{n+1})}u_{i+1}^{(k)}(t^{n+1})\right), \\ D\big(F_2\big(u_{i+1}^{(l)}(t^{n+1})\big)\big) &= 1 - \Delta t\left(B(u_{i+1}^{(l)}(t^{n+1})) + \frac{\partial B(u_{i+1}^{(l)}(t^{n+1}))}{\partial u_{i+1}^{(l)}(t^{n+1})}u_{i+1}^{(l)}(t^{n+1})\right). \end{aligned} \tag{6.22}$$

Our numerical results for the different methods are presented in Tables 1, 2, 3, and 4. The errors of the methods are shown in Figures 1, 2, and 3. We chose different iteration steps and time partitions. The error between the analytical and numerical solution is shown with the supremum norm at time $T = 1.0$.

Table 3: Numerical results for the Bernoulli equation with iterative operator-splitting method.

Time partitions	Number of iter.	Approximated solution	Error
1	2	0.0125000000	1.250000e-002
1	4	0.2927814810	2.927815e-001
1	10	0.0109667158	1.096672e-002
1	50	0.0109556732	1.095567e-002
5	2	0.0109913109	1.099131e-002
5	4	0.3152826900	3.152827e-001
5	10	0.0108511723	1.085117e-002
5	50	0.0108509643	1.085096e-002
10	2	0.0108995483	1.089955e-002
10	4	0.2437741856	2.437742e-001
10	10	0.0108426328	1.084263e-002
10	50	0.0108426158	1.084262e-002
50	2	0.0149667882	1.496679e-002
50	4	0.0166913971	1.669140e-002
50	10	0.0157464111	1.574641e-002
50	50	0.0159933864	1.599339e-002
100	2	0.0154572223	1.545722e-002
100	4	0.0160048071	1.600481e-002
100	10	0.0158481781	1.584818e-002
100	50	0.0158673179	1.586732e-002
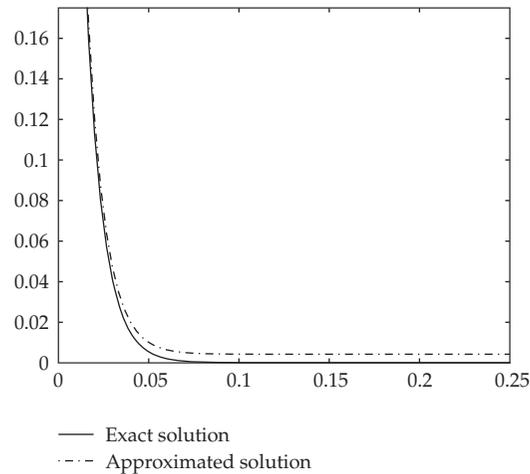
Figure 1: Analytical and approximated solution with sequential operator-splitting method.

The experiments show the reduced errors for more iteration steps and more time partitions. Because of the time-discretization method for ODEs, we restrict the number of iteration steps to a maximum of five. If we restrict the error bound to 10^{-3}, two iteration steps and five time partitions give the most effective combination.

Table 4: Numerical results for the Bernoulli equation with iterative operator-splitting method with embedded Newton's method.

Time partitions	Number of iter.	Approximated solution	Error
1	2	0.0000000000	1.137634e-044
1	4	0.0000000000	1.137634e-044
1	10	0.0000000000	1.137634e-044
1	20	0.0000000000	1.137634e-044
2	2	0.0000000000	1.137634e-044
2	4	0.0000000000	1.137634e-044
2	10	0.0000000000	1.137634e-044
2	20	0.0000000000	1.137634e-044

— Exact solution
·-·- Approximated solution
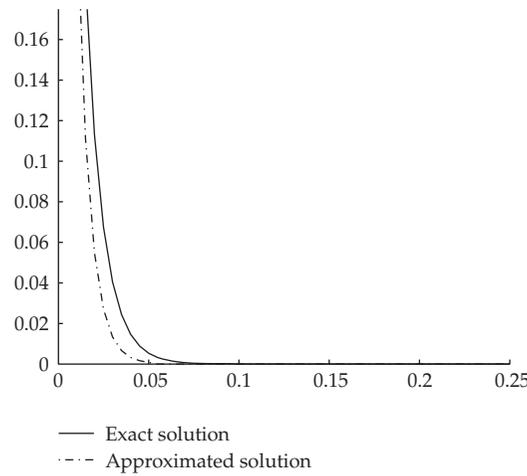
Figure 2: Analytical and approximated solution with sequential operator-splitting method with embedded Newton's method.

6.2. Second Numerical Example: Mixed Convection-Diffusion and Burgers Equation

We deal with a 2D example which is a mixture of a convection-diffusion and Burgers equation. We can derive an analytical solution:

$$
\begin{aligned}
\partial_t u = {}& -\frac{1}{2}u\partial_x u - \frac{1}{2}u\partial_y u - \frac{1}{2}\partial_x u - \frac{1}{2}\partial_y u \\
& + \mu(\partial_{xx} u + \partial_{yy} u) + f(x,y,t), \quad (x,y,t) \in \Omega \times [0,T], \\
u(x,y,0) = {}& u_{\text{ana}}(x,y,0), \quad (x,y) \in \Omega, \\
u(x,y,t) = {}& u_{\text{ana}}(x,y,t) \quad \text{on } \partial\Omega \times [0,T],
\end{aligned}
\tag{6.23}
$$

where $\Omega = [0,1] \times [0,1]$, $T = 1.25$, and μ is the viscosity.
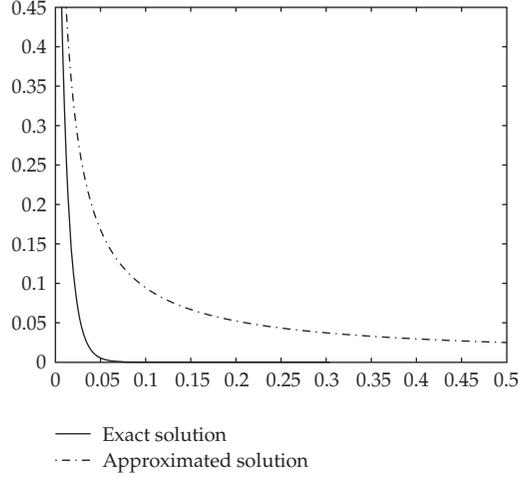
Figure 3: Analytical and approximated solution with iterative operator-splitting method.

The analytical solution is given as

$$u_{\text{ana}}(x, y, t) = \left(1 + \exp\left(\frac{x + y - t}{2\mu}\right)\right)^{-1} + \exp\left(\frac{x + y - t}{2\mu}\right), \qquad (6.24)$$

where we compute $f(x, y, t)$ accordingly.

We split the convection-diffusion and the Burgers equation. The operators are given as:

$$A(u)u = -\frac{1}{2}u\partial_x u - \frac{1}{2}u\partial_y u + \frac{1}{2}\mu(\partial_{xx}u + \partial_{yy}u), \qquad (6.25)$$

hence

$$A(u) = \frac{1}{2}\left(-u\partial_x - u\partial_y + \mu(\partial_{xx} + \partial_{yy})\right) \quad \text{(the Burgers term)},$$

$$Bu = -\frac{1}{2}\partial_x u - \frac{1}{2}\partial_y u + \frac{1}{2}\mu(\partial_{xx}u + \partial_{yy}u) + f(x, y, t) \quad \text{(the convection-diffusion term)}.$$

$$(6.26)$$

For the first equation we apply the nonlinear Algorithm 5.1 and obtain

$$A(u_{i-1})u_i = -\frac{1}{2}u_{i-1}\partial_x u_i - \frac{1}{2}u_{i-1}\partial_y u_i + \frac{1}{2}\mu(\partial_{xx}u_i + \partial_{yy}u_i),$$

$$Bu_{i-1} = \frac{1}{2}\left(-\partial_x - \partial_y + \mu(\partial_{xx} + \partial_{yy})\right)u_{i-1}, \qquad (6.27)$$

and we obtain linear operators, because u_{i-1} is known from the previous time-step.

Table 5: Numerical results for the mixed convection-diffusion and Burgers equation with viscosity $\mu = 0.5$, initial condition $u_0(t) = c_n$, and four iterations per time-step.

$\Delta x = \Delta y$	Δt	err_{L_1}	$\text{err}_{\max}$	ρ_{L_1}	$\rho_{\max}$
1/5	1/20	0.0137	0.0354		
1/10	1/20	0.0055	0.0139	1.3264	1.3499
1/20	1/20	0.0017	0.0043	1.6868	1.6900
1/40	1/20	$8.8839 \cdot 10^{-5}$	$3.8893 \cdot 10^{-4}$	4.2588	3.4663
1/5	1/40	0.0146	0.0377		
1/10	1/40	0.0064	0.0160	1.1984	1.2315
1/20	1/40	0.0026	0.0063	1.3004	1.3375
1/40	1/40	$8.2653 \cdot 10^{-4}$	0.0021	1.6478	1.6236

Table 6: Numerical results for the mixed convection-diffusion and Burgers equation with viscosity $\mu = 5$, initial condition $u_0(t) = c_n$, and two iterations per time-step.

$\Delta x = \Delta y$	Δt	err_{L_1}	$\text{err}_{\max}$	ρ_{L_1}	$\rho_{\max}$
1/5	1/20	$1.3166 \cdot 10^{-5}$	$2.9819 \cdot 10^{-5}$		
1/10	1/20	$5.6944 \cdot 10^{-6}$	$1.3541 \cdot 10^{-5}$	1.2092	1.1389
1/20	1/20	$1.6986 \cdot 10^{-6}$	$4.5816 \cdot 10^{-6}$	1.7452	1.5634
1/40	1/20	$7.8145 \cdot 10^{-7}$	$2.0413 \cdot 10^{-6}$	1.1201	1.1663
1/5	1/40	$1.4425 \cdot 10^{-5}$	$3.2036 \cdot 10^{-5}$		
1/10	1/40	$7.2343 \cdot 10^{-6}$	$1.5762 \cdot 10^{-5}$	0.9957	1.0233
1/20	1/40	$3.0776 \cdot 10^{-6}$	$6.7999 \cdot 10^{-6}$	1.2330	1.2129
1/40	1/40	$9.8650 \cdot 10^{-7}$	$2.3352 \cdot 10^{-6}$	1.6414	1.5420

In the second equation we obtain by using Algorithm 5.1:

$$
\begin{aligned}
A(u_{i-1})u_i &= -\frac{1}{2}u_{i-1}\partial_x u_i - \frac{1}{2}u_{i-1}\partial_y u_i + \frac{1}{2}\mu(\partial_{xx}u_i + \partial_{yy}u_i), \\
Bu_{i+1} &= \frac{1}{2}\left(-\partial_x - \partial_y + \mu(\partial_{xx} + \partial_{yy})\right)u_{i+1},
\end{aligned}
\tag{6.28}
$$

and we have linear operators.

We deal with different viscosities μ as well as different step-sizes in time and space. We have the following results (see Tables 5 and 6).

Figure 4 presents the profile of the 2D linear and nonlinear convection-diffusion equation.

Remark 6.1. In the examples, we deal with more iteration steps to obtain higher-order convergence results. In the first test we have four iterative steps but a smaller viscosity ($\mu = 0.5$) such that we can reach at least a second-order method. In the second test we use a high viscosity about $\mu = 5$ and get the second-order result with two iteration steps. Here we see the loss of differentiability, that becomes stiff equation parts. To obtain the same results, we have to increase the number of iteration steps. Therefore we can show an improvement of the convergence order with respect to the iteration steps.

Numeric solution wave equation

Numeric solution wave equation

(a)

(b)
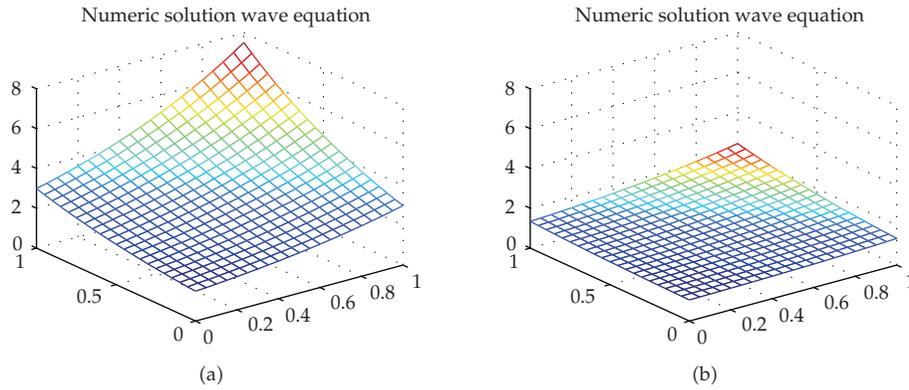
Figure 4: Mixed convection-diffusion and Burgers equation at initial time $t = 0.0$ (a) and end time $t = 1.25$ (b) for viscosity $\mu = 0.5$.

Numeric solution wave equation

Numeric solution wave equation

(a)

(b)
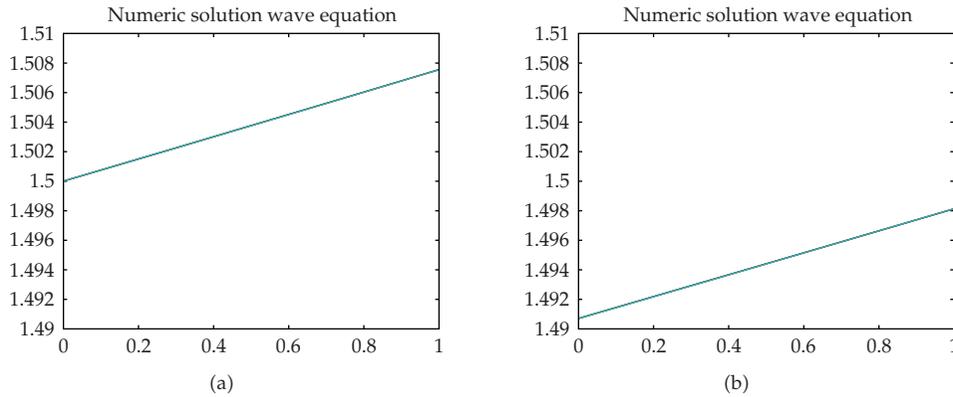
Figure 5: 1D momentum equation at initial time $t = 0.0$ (a) and end time $t = 1.25$ (b) for $\mu = 5$ and $v = 0.001$.

6.3. Third Numerical Example: Momentum Equation (Molecular Flow)

We deal with an example of a momentum equation, that is used to model the viscous flow of a fluid.

$$\partial_t \mathbf{u} = -\mathbf{u} \cdot \nabla \mathbf{u} + 2\mu \nabla \left(D(\mathbf{u}) + \frac{1}{3} \nabla \mathbf{u} \right) + \mathbf{f}(x, y, t), \quad (x, y, t) \in \Omega \times [0, T],$$

$$\mathbf{u}(x, y, 0) = \mathbf{g}_1(x, y), \quad (x, y) \in \Omega, \tag{6.29}$$

$$\mathbf{u}(x, y, t) = \mathbf{g}_2(x, y, t) \quad \text{on } \partial\Omega \times [0, T], \text{ (enclosed flow)},$$

where $\mathbf{u} = (u_1, u_2)^t$ is the solution and $\Omega = [0, 1] \times [0, 1]$, $T = 1.25$, $\mu = 5$, and $\mathbf{v} = (0.001, 0.001)^t$ are the parameters and I is the unit matrix.

(a)

(b)

(c)

(d)
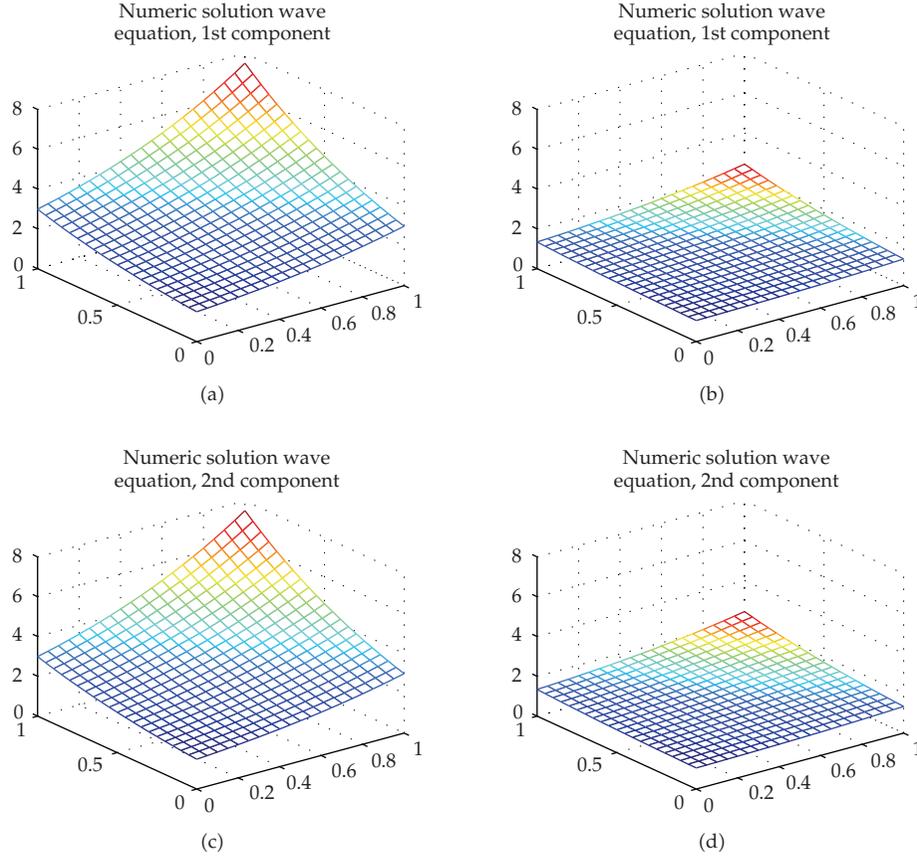
Figure 6: 2D momentum equation at initial time $t = 0.0$ ((a), (c)) and end time $t = 1.25$ ((b), (d)) for $\mu = 0.5$ and $v = (1, 1)^t$ for the first and second component of the numerical solution.

The nonlinear function $D(\mathbf{u}) = \mathbf{u} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{u}$ is the viscosity flow, and $\mathbf{v}$ is a constant velocity.

We can derive the analytical solution with respect to the first two test examples with the functions:

$$
\begin{aligned}
u_1(x, y, t) &= \left(1 + \exp\left(\frac{x + y - t}{2\mu}\right)\right)^{-1} + \exp\left(\frac{x + y - t}{2\mu}\right), \\
u_2(x, y, t) &= \left(1 + \exp\left(\frac{x + y - t}{2\mu}\right)\right)^{-1} + \exp\left(\frac{x + y - t}{2\mu}\right).
\end{aligned}
\tag{6.30}
$$

For the splitting method our operators are given as:

$$
\begin{aligned}
A(\mathbf{u})\mathbf{u} &= -\mathbf{u}\nabla\mathbf{u} + 2\mu\nabla D(\mathbf{u}) \quad \text{(the nonlinear operator)}, \\
B\mathbf{u} &= \frac{2}{3}\mu\Delta\mathbf{u} \quad \text{(the linear operator)}.
\end{aligned}
\tag{6.31}
$$

We deal first with the one-dimensional case,

$$\partial_t u = -u \cdot \partial_x u + 2\mu\partial_x\left(D(u) + \frac{1}{3}\partial_x u\right) + f(x,t), \quad (x,t) \in \Omega \times [0,T],$$

$$u(x,0) = g_1(x), \quad (x) \in \Omega, \tag{6.32}$$

$$u(x,t) = g_2(x,t) \quad \text{on } \partial\Omega \times [0,T], \text{ (enclosed flow)},$$

where u is the solution and $\Omega = [0,1]$, $T = 1.25$, $\mu = 5$, and $v = 0.001$ are the parameters.

Then the operators are given as:

$$A(u)u = -u\partial_x u + 2\mu\partial_x D(u) \quad \text{(the nonlinear operator)},$$

$$Bu = \frac{2}{3}\mu\partial_{xx}u \quad \text{(the linear operator)}. \tag{6.33}$$

For the iterative operator-splitting as fixed point scheme, we have the following results (see Tables 7 and 9).

Figure 5 presents the profile of the 1D momentum equation.

We have the following results for the 2D case (see Tables 10, 11, and 12).

Figure 6 presents the profile of the 2D momentum equation.

For the Newton operator-splitting method we obtain the following functional matrices for the one-dimensional case:

$$DF(u) = (4\mu - 1)\partial_x u,$$

$$D(F(u)) = -\begin{pmatrix} \partial_{u_1} F_1(u) & \partial_{u_2} F_1(u) \\ \partial_{u_1} F_2(u) & \partial_{u_2} F_2(u) \end{pmatrix}$$

$$= -\begin{pmatrix} -\partial_x u_1 + 4\mu\partial_x u_1 & -\partial_x u_2 + 4\mu\partial_x u_2 \\ -\partial_y u_1 + 4\mu\partial_y u_1 & -\partial_y u_2 + 4\mu\partial_y u_2 \end{pmatrix} \tag{6.34}$$

$$= (4\mu - 1)\nabla u.$$

For the two-dimensional case, we use:

$$A(\mathbf{u})\mathbf{u} = -\mathbf{u}\nabla\mathbf{u} + 2\mu\nabla D(\mathbf{u})$$

$$= -\begin{pmatrix} u_1\partial_x u_1 + u_2\partial_x u_2 \\ u_1\partial_y u_1 + u_2\partial_y u_2 \end{pmatrix} + 2\mu\begin{pmatrix} 2u_1\partial_x u_1 + 2u_2\partial_x u_2 + v_1\partial_x u_1 + v_2\partial_x u_2 \\ 2u_1\partial_y u_1 + 2u_2\partial_y u_2 + v_1\partial_y u_1 + v_2\partial_y u_2 \end{pmatrix}. \tag{6.35}$$

Here, we do not need the linearization and apply the standard iterative splitting method.

We only linearize the first split step and therefore we can relax this step with the second linear split step. Therefore we obtain stable methods, see [15]. For the Newton iterative method, we have the following results, see Table 8.

Table 7: Numerical results for the 1D momentum equation with $\mu = 5$, $v = 0.001$, initial condition $u_0(t) = c_n$, and two iterations per time-step.

Δx	Δt	err_{L_1}	$\mathrm{err}_{\max}$	ρ_{L_1}	$\rho_{\max}$
1/10	1/20	0.0213	0.0495		
1/20	1/20	0.0203	0.0470	0.0689	0.0746
1/40	1/20	0.0198	0.0457	0.0401	0.0402
1/80	1/20	0.0195	0.0450	0.0216	0.0209
1/10	1/40	0.0134	0.0312		
1/20	1/40	0.0117	0.0271	0.1957	0.2009
1/40	1/40	0.0108	0.0249	0.1213	0.1211
1/80	1/40	0.0103	0.0238	0.0682	0.0674
1/10	1/80	0.0094	0.0217		
1/20	1/80	0.0073	0.0169	0.3591	0.3641
1/40	1/80	0.0062	0.0143	0.2451	0.2448
1/80	1/80	0.0056	0.0129	0.1478	0.1469

Table 8: Numerical results for the 1D momentum equation with $\mu = 5$, $v = 0.001$, initial condition $u_0(t) = c_n$, two iterations per time-step and $K = 1$ using Newton iterative method.

Δx	Δt	err_{L_1}	$\mathrm{err}_{\max}$	ρ_{L_1}	$\rho_{\max}$
1/10	1/20	0.0180	0.0435		
1/20	1/20	0.0120	0.0276	0.5867	0.6550
1/40	1/20	0.0095	0.0227	0.3311	0.2870
1/80	1/20	0.0085	0.0208	0.1706	0.1231
1/10	1/40	0.0172	0.0459		
1/20	1/40	0.0125	0.0305	0.4652	0.5884
1/40	1/40	0.0108	0.0253	0.2366	0.2698
1/80	1/40	0.0097	0.0235	0.1191	0.1111
1/10	1/80	0.0166	0.0475		
1/20	1/80	0.0132	0.0338	0.3327	0.4917
1/40	1/80	0.0119	0.0280	0.1640	0.2734
1/80	1/80	0.0112	0.0265	0.0802	0.0779

Remark 6.2. In the more realistic examples of 1D and 2D momentum equations, we can also observe the stiffness problem, which we obtain with a more hyperbolic behavior. In the 1D experiments we deal with a more hyperbolic behavior and can obtain at least first-order convergence with two iterative steps. In the 2D experiments we obtain nearly second-order convergence results with two iterative steps, if we increase the parabolic behavior, for example, larger μ and **v** values. For such methods, we have to balance the usage of the iterative steps with refinement in time and space with respect to the hyperbolicity of the equations. At least we can obtain a second-order method with more than two iterative steps. Therefore the stiffness influences the number of iterative steps.

Table 9: Numerical results for the 1D momentum equation with $\mu = 50$, $v = 0.1$, initial condition $u_0(t) = c_n$, and two iterations per time-step.

Δx	Δt	err_{L_1}	$\text{err}_{\max}$	ρ_{L_1}	$\rho_{\max}$
1/10	1/20	$2.7352 \cdot 10^{-6}$	$6.4129 \cdot 10^{-6}$		
1/20	1/20	$2.3320 \cdot 10^{-6}$	$5.4284 \cdot 10^{-6}$	0.2301	0.2404
1/40	1/20	$2.1144 \cdot 10^{-6}$	$4.9247 \cdot 10^{-6}$	0.1413	0.1405
1/80	1/20	$2.0021 \cdot 10^{-6}$	$4.6614 \cdot 10^{-6}$	0.0787	0.0793
1/10	1/40	$2.1711 \cdot 10^{-6}$	$5.2875 \cdot 10^{-6}$		
1/20	1/40	$1.7001 \cdot 10^{-6}$	$4.1292 \cdot 10^{-6}$	0.3528	0.3567
1/40	1/40	$1.4388 \cdot 10^{-6}$	$3.4979 \cdot 10^{-6}$	0.2408	0.2394
1/80	1/40	$1.3023 \cdot 10^{-6}$	$3.1694 \cdot 10^{-6}$	0.1438	0.1423
1/10	1/80	$1.6788 \cdot 10^{-6}$	$4.1163 \cdot 10^{-6}$		
1/20	1/80	$1.1870 \cdot 10^{-6}$	$2.9138 \cdot 10^{-6}$	0.5001	0.4984
1/40	1/80	$9.1123 \cdot 10^{-7}$	$2.2535 \cdot 10^{-6}$	0.3814	0.3707
1/80	1/80	$7.6585 \cdot 10^{-7}$	$1.9025 \cdot 10^{-6}$	0.2507	0.2443

Table 10: Numerical results for the 2D momentum equation with $\mu = 2$, $v = (1,1)^t$, initial condition $u_0(t) = c_n$, and two iterations per time-step.

$\Delta x = \Delta y$	Δt	err_{L_1} 1st c.	$\text{err}_{\max}$ 1st c.	ρ_{L_1} 1st c.	$\rho_{\max}$ 1st c.	err_{L_1} 2nd c.	$\text{err}_{\max}$ 2nd c.	ρ_{L_1} 2nd c.	$\rho_{\max}$ 2nd c.
1/5	1/20	0.0027	0.0112			0.0145	0.0321		
1/10	1/20	0.0016	0.0039	0.7425	1.5230	0.0033	0.0072	2.1526	2.1519
1/20	1/20	0.0007	0.0022	1.2712	0.8597	0.0021	0.0042	0.6391	0.7967
1/5	1/40	0.0045	0.0148			0.0288	0.0601		
1/10	1/40	0.0032	0.0088	0.5124	0.7497	0.0125	0.0239	1.2012	1.3341
1/20	1/40	0.0014	0.0034	1.1693	1.3764	0.0029	0.0054	2.1263	2.1325
1/5	1/80	0.0136	0.0425			0.0493	0.1111		
1/10	1/80	0.0080	0.0241	0.7679	0.8197	0.0278	0.0572	0.8285	0.9579
1/20	1/80	0.0039	0.0113	1.0166	1.0872	0.0115	0.0231	1.2746	1.3058

Table 11: Numerical results for the 2D momentum equation for the first component with $\mu = 50$, $v = (100, 0.01)^t$, initial condition $u_0(t) = c_n$, and two iterations per time-step.

$\Delta x = \Delta y$	Δt	err_{L_1} 1st c.	$\text{err}_{\max}$ 1st c.	ρ_{L_1} 1st c.	$\rho_{\max}$ 1st c.
1/5	1/20	$1.5438 \cdot 10^{-5}$	$3.4309 \cdot 10^{-5}$		
1/10	1/20	$4.9141 \cdot 10^{-6}$	$1.0522 \cdot 10^{-5}$	1.6515	1.7052
1/20	1/20	$1.5506 \cdot 10^{-6}$	$2.9160 \cdot 10^{-6}$	1.6641	1.8513
1/5	1/40	$2.8839 \cdot 10^{-5}$	$5.5444 \cdot 10^{-5}$		
1/10	1/40	$1.3790 \cdot 10^{-5}$	$2.3806 \cdot 10^{-5}$	1.0645	1.2197
1/20	1/40	$3.8495 \cdot 10^{-6}$	$6.8075 \cdot 10^{-6}$	1.8408	1.8061
1/5	1/80	$3.1295 \cdot 10^{-5}$	$5.5073 \cdot 10^{-5}$		
1/10	1/80	$1.7722 \cdot 10^{-5}$	$2.6822 \cdot 10^{-5}$	0.8204	1.0379
1/20	1/80	$7.6640 \cdot 10^{-6}$	$1.1356 \cdot 10^{-5}$	1.2094	1.2400

Table 12: Numerical results for the 2D momentum equation for the second component with $\mu = 50$, $v = (100, 0.01)^t$, initial condition $u_0(t) = c_n$, and two iterations per time-step.

$\Delta x = \Delta y$	Δt	err_{L_1} 2nd c.	err_{max} 2nd c.	ρ_{L_1} 2nd c.	ρ_{max} 2nd c.
1/5	1/20	$4.3543 \cdot 10^{-5}$	$1.4944 \cdot 10^{-4}$		
1/10	1/20	$3.3673 \cdot 10^{-5}$	$7.9483 \cdot 10^{-5}$	0.3708	0.9109
1/20	1/20	$2.6026 \cdot 10^{-5}$	$5.8697 \cdot 10^{-5}$	0.3717	0.4374
1/5	1/40	$3.4961 \cdot 10^{-5}$	$2.2384 \cdot 10^{-4}$		
1/10	1/40	$1.7944 \cdot 10^{-5}$	$8.9509 \cdot 10^{-5}$	0.9622	1.3224
1/20	1/40	$1.5956 \cdot 10^{-5}$	$3.6902 \cdot 10^{-5}$	0.1695	1.2783
1/5	1/80	$9.9887 \cdot 10^{-5}$	$3.3905 \cdot 10^{-4}$		
1/10	1/80	$3.5572 \cdot 10^{-5}$	$1.3625 \cdot 10^{-4}$	1.4896	1.3153
1/20	1/80	$1.0557 \cdot 10^{-5}$	$4.4096 \cdot 10^{-5}$	1.7525	1.6275

7. Conclusion and Discussion

We present decomposition methods for differential equations based on iterative and non-iterative methods. The nonlinear equations are solved with embedded Newton's methods. We present new ideas on linearization to obtain more accurate results. The superiority of the new embedded Newton's methods over the traditional sequential methods is demonstrated in examples, especially through their simple implementation. Further, we have the smoothing properties of the iterative scheme that allow a balance between the nonlinear and the linear terms. The results show more accurate solutions with respect to time decomposition. In the future the iterative operator-splitting method can be generalized for multi-dimensional problems and also for non-smooth and nonlinear problems in time and space. In next paper we discuss error analysis of nonlinear methods.

References

[1] G. I. Marchuk, "Some applicatons of splitting-up methods to the solution of problems in mathematical physics," *Aplikace Matematiky*, vol. 1, pp. 103–132, 1968.

[2] G. Strang, "On the construction and comparison of difference schemes," *SIAM Journal on Numerical Analysis*, vol. 5, no. 3, pp. 506–517, 1968.

[3] H. Yoshida, "Construction of higher order symplectic integrators," *Physics Letters A*, vol. 150, no. 5–7, pp. 262–268, 1990.

[4] J. F. Kanney, C. T. Miller, and C. T. Kelley, "Convergence of iterative split-operator approaches for approximating nonlinear reactive problems," *Advances in Water Resources*, vol. 26, no. 3, pp. 247–261, 2003.

[5] P. Deuflhard, *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithm*, vol. 35 of *Springer Series in Computational Mathematics*, Springer, Berlin, Germany, 2004.

[6] P. Deuflhard, "A modified Newton method for the solution of ill-conditioned systems of nonlinear equations with application to multiple shooting," *Numerische Mathematik*, vol. 22, no. 4, pp. 289–315, 1974.

[7] M. Ruzicka, *Nichtlineare Funktionalanalysis*, Springer, Berlin, Germany, 2004.

[8] E. Zeidler, *Nonlinear Functional Analysis and Its Applications. II/B: Nonlinear Monotone Operators*, Springer, New York, NY, USA, 1990.

[9] J. Geiser, O. Klein, and P. Philip, "Anisotropic thermal conductivity in apparatus insulation: numerical study of effects on the temperature field during sublimation growth of silicon carbide single crystals," preprint no. 1034 of Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany, 2005.

[10] J. Geiser, "Discretization methods with embedded analytical solutions for convection dominated transport in porous media," in *Proceedings of the 3rd International Conference on Numerical Analysis and Its Applications (NAA '04)*, vol. 3401 of *Lecture Notes in Computer Science*, pp. 288–295, Springer, Rousse, Bulgaria, June-July 2004.

[11] W. Hundsdorfer and J. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, vol. 33 of *Springer Series in Computational Mathematics*, Springer, Berlin, Germany, 2003.

[12] K. H. Karlsen, K.-A. Lie, J. R. Natvig, H. F. Nordhaug, and H. K. Dahle, "Operator splitting methods for systems of convection-diffusion equations: nonlinear error mechanisms and correction strategies," *Journal of Computational Physics*, vol. 173, no. 2, pp. 636–663, 2001.

[13] K. H. Karlsen and N. H. Risebro, "Corrected operator splitting for nonlinear parabolic equations," *SIAM Journal on Numerical Analysis*, vol. 37, no. 3, pp. 980–1003, 2000.

[14] I. Faragó and J. Geiser, "Iterative operator-splitting methods for linear problems," *International Journal of Computational Science and Engineering*, vol. 3, no. 4, pp. 255–263, 2007.

[15] C. T. Kelly, *Iterative Methods for Linear and Nonlinear Equations*, Frontiers in Applied Mathematics, SIAM, Philadelphia, Pa, USA, 1995.

[16] J. Geiser, "Iterative operator-splitting methods with higher-order time integration methods and applications for parabolic partial differential equations," *Journal of Computational and Applied Mathematics*, vol. 217, no. 1, pp. 227–242, 2008.

[17] K. H. Karlsen and N. H. Risebro, "An operator splitting method for nonlinear convection-diffusion equations," *Numerische Mathematik*, vol. 77, no. 3, pp. 365–382, 1997.

Research Article

The Staircase Structure of the Southern Brazilian Continental Shelf

M. S. Baptista[1, 2] and L. A. Conti[3]

[1] *Centro de Matemática da Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*

[2] *Institute for Complex Systems and Mathematical Biology, King's College, University of Aberdeen, AB24 3UE Aberdeen, UK*

[3] *Escola de Artes Ciências e Humanidades, Universidade de São Paulo, Rua Arlindo Bettio 1000, 03828-000 São Paulo, SP, Brazil*

Correspondence should be addressed to M. S. Baptista, murilo.baptista@fc.up.pt

We show some evidences that the Southeastern Brazilian Continental Shelf (SBCS) has a devil's staircase structure, with a sequence of scarps and terraces with widths that obey fractal formation rules. Since the formation of these features is linked with the sea-level variations, we say that the sea level changes in an organized pulsating way. Although the proposed approach was applied in a particular region of the Earth, it is suitable to be applied in an integrated way to other shelves around the world, since the analyses favor the revelation of the global sea-level variations.

1. Introduction

During the late quaternary period, after the last glacial maximum (LGM), from 18000 years ago till present, a global warming was responsible for the melting of the glaciers leading to a fast increase in the sea level. In approximately 13000 years, the sea level rised up to about 120 meters, reaching the actual level. However, the sea level did not go up in a continuous fashion, but rather, it has evolved in a pulsatile way, leaving behind a signature of what actually happened, the continental shelf, that is, the seafloor.

Continental shelves are located at the boundary with the land so that they are shaped by both marine and terrestrial processes. Sea-level oscillations incessantly transform terrestrial areas in marine environments and vice versa, thus increasing the landscape complexity [1]. The presence of regions with abnormal slope as well as the presence of terraces on a continental shelf are indicators of sea-level positions after the last glacial maximum (LGM), when large ice sheets covered high latitudes of Europe and North

America, and sea levels stood about 120–130 m lower than today [2]. Geomorphic processes responsible for the formation of these terraces and discontinuities on the bottom of the sea topography are linked to the coastal dynamics during eustatic processes associated with both erosional or depositional forcing (wave-cut and wave-built terraces, resp., [3]).

The irregular distribution of such terraces and shoreface sediments is mainly controlled by the relationship between shelf paleo-physiography and changes on the sea level and sediment supply which reflect both global and local processes. Several works have dealt with mapping and modeling the distribution of shelf terraces in order to understand the environmental consequences of climate change and sea-level variations after the LGM [4–6].

In this period of time the sea-level transgression was punctuated by at least six relatively short flooding events that collectively accounted for more than 90 m of the 120 m rise. Most (but not all) of the floodings appear to correspond with paleoclimatic events recorded in Greenland and Antarctic ice cores, indicative of the close coupling between rapid climate change, glacial melt, and corresponding sea-level rise [7].

In this work, we analyze data from the Southeastern Brazilian Continental Shelf (SBCS) located in a typical sandy passive margin with the predominance of palimpsests sediments. The mean length is approximately 250 km, and the shelf break is located at 150 m depth. It is a portion of a greater geomorphologic region of the southeastern Brazilian coast called São Paulo Bight, an arc-shaped part of the southeastern Brazilian margin. The geology and topography of the immersed area are very peculiar, represented by the Mesozoic/Cenozoic tectonic processes that generated the mountainous landscapes known as "Serra do Mar." These landscapes (with mean altitudes of 800 m) have a complex pattern that characterizes the coastal morphology and leads to several scarps intercalated with small coastal plains and pocket beaches.

This particular characteristic determines the development of several small-size fluvial basins and absence of major rivers conditioning low sediment input, what tends to preserve topographic signatures of the sea-level variations.

For the purpose of the present study, we select three parallel profiles acquired from echo-sounding surveys, since for all the considered profiles, the same similar series of sequences of terraces were found. These profiles [8–10] are transversal to the coastline and the isobaths trend, and they extend from a 20 m to a 120 m depth.

The importance of understanding the formation of these ridges is that it can tell us about the coastal morphodynamic conditions, inner shelf processes, and the characteristics of periods of the sea-level regimes standstills (paleoshores). In particular, the widths of the terraces are related to the time the sea level "stabilized." All this information is vital for the better understanding of the late quaternary climate changes dynamic.

We find relations between the widths of the terraces that follow a self-affine pattern description. These relations are given by a mathematical model, which describes an order of appearance for the terraces. Our results suggest that this geomorphological structure for the terraces can be described by a devil's staircase [11]; a staircase with infinitely many steps in between two steps. This property gives the name "devil" to the staircase, once an idealized being would take an infinite time to go from one step to another. So, the seafloor morphology is self-affine (fractal structure) as reported in [12, 13], but according to our findings, it has a special kind of self-affine structure, the devil's staircase structure.

A devil's staircase as well as other self-affine structure are the response of an oscillatory system when excited by some external force. The presence of a step means that while varying some internal parameter, the system preserves some averaged regular behavior; a consequence of the stable frequency-locking regime between a natural frequency of the

system and the frequency of the excitation. This staircase as well as other self-affine structures are characterized by the presence of steps whose widths are directly related to the rational ratio between the natural frequency of the system and the frequency of the excitation.

In a similar fashion, we associate the widths of the terraces with rational numbers that represent two hypothetical frequencies of oscillation which are assumed to exist in the system that creates the structure of the SBCS, here regarded as the sea-level dynamics (SLDs), also known as the sea-level variations. Then, once these rational numbers are found, we show that the relative distances between triples of terraces (associated with some hypothetical frequencies) follow similar scalings found in the relative distance between triples of plateaus (associated with these same frequencies) observed in the devil's staircase.

The seafloor true structure, apart from the dynamics that originated it, is also a very relevant issue, specially for practical applications. For example, one can measure the seafloor with one resolution and then reconstruct the rest based on some modeling [14]. As we show in this work (Section 5), a devil's staircase structure fits remarkably well the experimental data.

Our paper is organized as follows. In Section 2, we describe the data to be analyzed. In Section 3, we describe which kind of dynamical systems can create a devil's staircase and how one can detect its presence in experimental data based on only a few observations. In Section 4, we show the evidences that led us to characterize the SBCS as a devil's staircase, and in Section 5 we show how to construct seafloor profiles based on the devil's staircase geometry. Finally, in Section 6, we present our conclusions, discussing also possible scenarios for the future of the sea-level dynamics under the perspective of our findings.

2. Data

The data consists of the tree profiles given in Figures 1(a)-1(b). The profile considered for our analyses is shown in Figure 1(b), where we show the continental shelf of the State of São Paulo, in a transversal cut in the direction: inner shelf (coast) → shelfbreak (open sea). The horizontal axis represents the distance to the coast and the vertical axis, the sea level (depth), d. We are interested in the terraces widths and their respective depths.

The profiles shown in Figure 1 were the result of a smoothing (filtering) process from the original data collected by Sonar (the profiles used in this work were obtained by the use of polinomial spilines in a two dimensional grip of points 10 km appart. This also imposes limits in the indentification of terraces that have small widths). The smoothing process is needed to eliminate from the measured data the influence of the oscillations of the ship where the sonar is located and local oscillations on the sea floor probably due to the stream flows.

Smaller topographic terraces could be smoothed or masked due to several processes such as coastal dynamic erosional during sea-level rising, Holocene sediment cover, and erosional processes associated with modern hidrodynamic pattern (geostrophic currents). For that reason we only consider the largest ones, as the ones shown in Figure 2 (located at $d = -30.01$ m with the width of $l = 6.06$ km). As one can see, the edges of the terraces are not so sharp as one would expect from a staircase plateau. Again, this is due to the action of the sea waves and stream flows throughout the time. To reconstruct what we believe to be the original terrace, we consider that its depth is given by the depth of the middle point, and its width is given by the minimal distance between two parallel lines placed along the scarps of the terrace edges. Using this procedure, we construct Table 1 with the largest and more relevant terraces found.

(a)   (b)
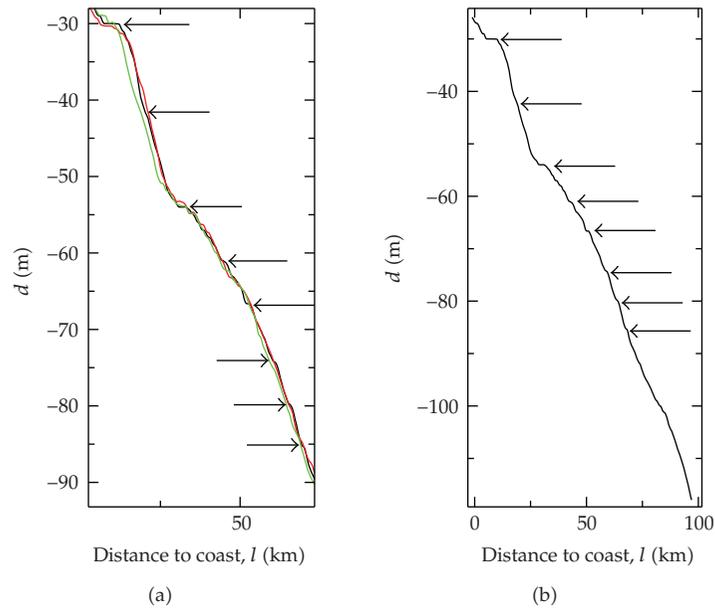
Figure 1: Profiles (depth versus the distance to the coast) of the Southeastern Brazilian Continental Shelf. The arrows indicate the terraces considered in our analyses. The profile shown with a thick black line is the profile chosen for our derivations, reproduced also in (b). The other two profiles had their original position of the two axes shifted by a constant value such that one can also identify the terraces observed in the chosen profile in these other two. Note that a translation of the profiles by a constant value has no effect on any of the scalings observed. The reason of this mismatch between the profiles is due to the local geometry of the coast at the time the sea reached that level.
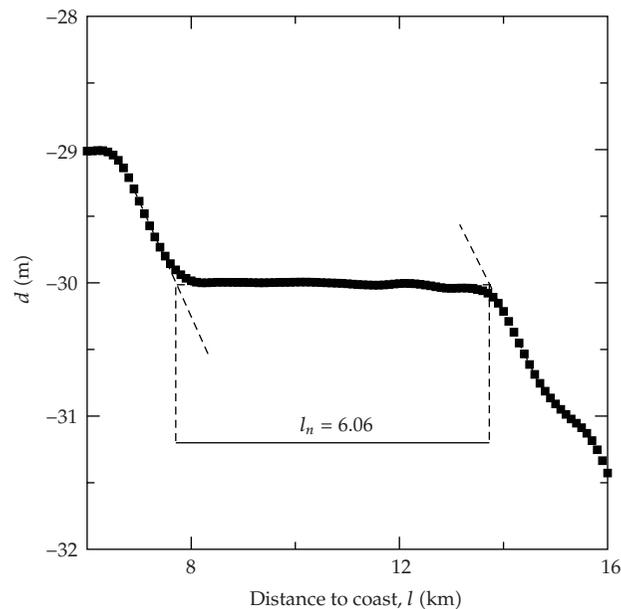
Figure 2: Reconstruction of the terraces. The width of the terrace is given by the minimal distance between the two parallel dashed lines.

Table 1: Terrace widths and depths. While the depths present no representative deviation, the deviation in the widths becomes larger for deeper terraces. The deviation in the widths is estimated by calculating the widths assuming many possible configurations between the placement of the two parallel lines used to calculate the widths.

n	d_n (m)	l_n (km)
1	−30.01	6.06 ± 0.05
2	−41.86	1.59 ± 0.05
3	−54.01	2.93 ± 0.05
4	−61.14	1.73 ± 0.05
5	−66.69	2.21 ± 0.05
6	−74.33	0.80 ± 0.1
7	−79.75	0.80 ± 0.1
8	−85.30	0.80 ± 0.1

We identify a certain terrace introducing a lower index n in l and d, according to their chronological order of appearance. More recent appearance (closer to the coast, less deep) smaller is the index n. We consider the more recent data to have a zero distance from the coast, but in fact, this data is positioned at about 15 km away from the shore, where the bottom of the sea is not affected by the turbulent zone caused by the breakout of the waves. The profile of Figure 1(b) was the one chosen among the other tree profiles because from it we could more clearly identify the largest number of relevant terraces by Sonar (the profiles used in this work were obtained by the use of polinomial spilines in a two dimensional grip of points 10 km appart. This also imposes limits in the indentification of terraces that have small widths).

3. The Devil's Staircase

Frequency locking is a resonant response occurring in systems of coupled oscillators or oscillators coupled to external forces. The first relevant report about this phenomenon was given by Christian Huygens in the 17th century. He observed that two clocks back to back in the wall, set initially with slightly different frequencies, would have their oscillations coupled by the energy transfer throughout the wall, and then they would eventually have their frequencies synchronized. Usually, we expect that a harmonic P_{w1} of one oscillatory system locks with a harmonic Q_{w2} of the other oscillatory system, leading to a locked system working in the rational ratio P/Q [15].

To understand what are the dynamics responsible for the onset of a frequency-locked oscillation, that is, the reasons for which a system either locks or unlocks, we present the simplest model one can come up with to describe a more general oscillator. This model is described by an angle θ, which is changed (after one period) to the angle $f(\theta)$. So, $f(\theta) = \theta + \Omega$. Then, this model (a rotation) is perturbed by an external force g, resulting in the following model:

$$f(\theta) = \theta + \Omega - g(\theta, K) \pmod 1, \tag{3.1}$$

where

$$g(\theta, K) = \frac{K}{2\pi} \sin 2\pi\theta. \tag{3.2}$$

Despite this map simplicity, the same cannot be said about its complexity [16]. Arnol'd (see [17]) studied this map in detail aiming to understand how an oscillatory system would undergo into periodic stable state when perturbed by an external perturbation.

For $K = 0$, (3.1) represents a pure rotation, which is topologically equivalent to a twice continuously differentiable orienting preserving mapping of the circle onto itself (Theorem of Denjoy; see [18]). Therefore, the simple (3.1) can be considered as a model for many types of oscillatory systems. In fact, (3.1) represents a more complicated system, a three-dimensional torus with frequencies w_1 and w_2, when viewed by a Poincaré map. Thus, Ω in (3.1) represents the ratio w_1/w_2. When $w_1/w_2 = p/q$ (with $p \leq q$) is rational, this map has a period p motion, and its trajectory, that is, the value of θ, assumes the same value after q iterations. For $K = 0$, the so-called winding number W is exactly equal to Ω, that is, $W = p/q$.

For $K \neq 0$ (nonlinear case), W is defined by

$$W(\Omega, K) = \lim_{n \to \infty} \frac{h(\theta_0, K) + h(\theta_1, K) + \cdots + h(\theta_{n-1}, K)}{n}, \tag{3.3}$$

where

$$h(\theta, K) = \Omega + g(\theta, K). \tag{3.4}$$

For $K < 1$, (3.1) is monotonic and invertible. For $K = 1$, it develops a cubic inflection point at $\theta = 0$. The map is still invertible but the inverse has a singularity. For $K > 1$, the map is noninvertible.

Arnol'd wanted to understand how periodic oscillations would appear as one increases K from zero to positive values. He observed that a quasiperiodic oscillation, for an irrational Ω and $K = 0$, would turn into a periodic oscillation as one varies K, from zero to positive values. He demonstrated that a periodic oscillation has probability zero of being found for $K = 0$ (rational numbers set is countable while the irrational numbers set is uncountable) and positive probability of being found for $K > 0$. He also observed that fixing a positive value K, the winding number W in (3.3) is a continuous but not differentiable function of Ω, as one can see in Figure 3, forming a stair-like structure.

If $W(\Omega, K)$ is rational, it can be represented by the ratio between two integer numbers as $W = P/Q$. At this point, the frequency Ω and the frequency of the function g are locked, producing the phenomenon of frequency locking, when $W(K, \Omega)$ does not change its value within an interval $\Delta\Omega$ (a plateau) of values of Ω. In fact, the smaller the denominator Q is, the larger the interval $\Delta\Omega$ is.

As one changes Ω, plateaus for W rational appear following a natural order described by the Farey mediant. Given two plateaus that represent P_1/Q_1 and P_3/Q_3 winding numbers,
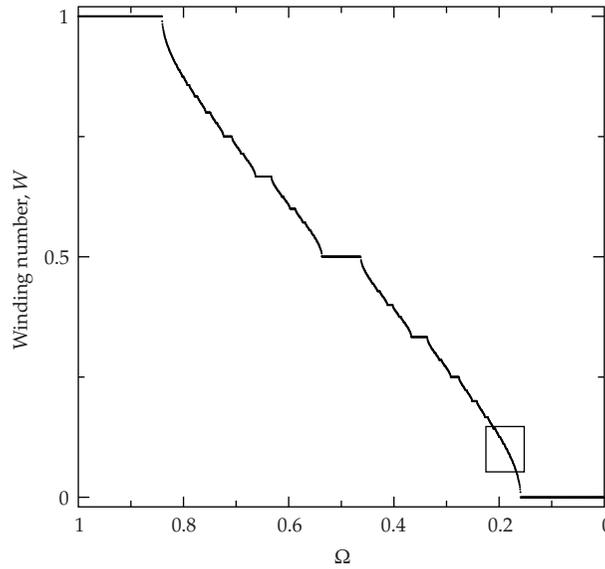
Figure 3: A complete devil's staircase, obtained from (3.1), for $K = 1$.

with plateau widths of $\Delta\Omega_1$ and $\Delta\Omega_3$, respectively, there exists another plateau positioned at a winding number W within the interval $[P_1/Q_1, P_3/Q_3]$ given by

$$\frac{P_2}{Q_2} = \frac{P_1 + P_3}{Q_1 + Q_3}. \tag{3.5}$$

The Farey mediant gives the rational with the smallest integer denominator that is within the interval $[P_1/Q_1, P_3/Q_3]$. Therefore, the $\Delta\Omega(P_2/Q_2)$ plateau is smaller than $\Delta\Omega(P_1/Q_1)$ and $\Delta\Omega(P_3/Q_3)$, but is bigger than any other possible plateau. Organizing the rationals according to the Farey mediant creates a hierarchical level of rationals, which are called Farey Tree. The plateaus $\Delta\Omega_1$ and $\Delta\Omega_3$ are regarded as the parents, and $\Delta\Omega_2$ as the daughter plateau.

The interesting case of (3.1) for our purpose is exactly when $K = 1$. For that case, one can find periodic orbits with any possible rational winding number, as one varies Ω. What means that the probability of finding quasiperiodic oscillation in (3.1) is zero by a random choice of the Ω value. Also, for $K > 1$, due to the overlap of resonances (periodic oscillation), chaos is possible.

The devil's staircase can be fully characterized by the relations between the plateaus widths and the relations between the gaps between two of them. While the plateau widths are linked to the probability that one has to find periodic oscillations, the gaps widths between plateaus are linked to the probability that one has to find quasiperiodic oscillations, in (3.1).

There are many scaling laws relating the plateau widths [15, 19, 20]. There are local scalings, which relate the widths of plateaus that appear close to a specific winding number, for example, the famous golden mean $W_G = (\sqrt{5}-1)/2$. However, we will focus our attention on the global scalings, which can be experimentally observed, and only for the case where $K = 1$. For this case, we are interested in two scalings; the one that relates the plateau widths with the respective winding numbers in the form $1/Q$ (the largest plateaus), and the one that describes the structure of the complementary set to the plateaus $\Delta\Omega$, that is, the structure of
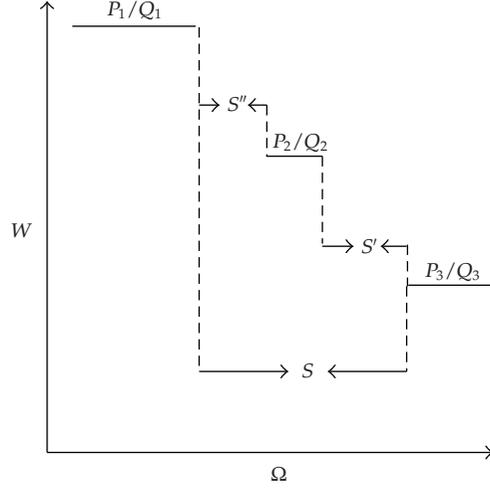
$S = 22.7\,\text{km}$
$S' = 2.5\,\text{km}$
$S'' = 9.6\,\text{km}$

Figure 4: Representation of the gaps S'', S', and S, used to estimate the fractal dimension D_0 of the complementary set, using (3.7).

the gaps between plateaus. The structure of the plateaus is a Cantor set as well as the structure of the complementary set. Therefore, a characterization of these sets can be done in terms of the fractal dimension D_0 [21] of the complementary set.

The first scaling is [15]

$$\Delta\Omega\left(\frac{1}{Q}\right) \propto \frac{1}{Q^\gamma} \quad (\gamma > 3),$$ (3.6)

The second scaling relates the widths of the complementary set as one goes to smaller and smaller scales. These widths are related to a power-scaling law whose coefficient D_0 is the fractal dimension of the complementary set. For $K = 1$, we have that the fractal dimension of the complementary set is $D_0 \cong 0.87$. This is a universal scaling. Since the complementary set of the plateaus represents the irrational rotations, the smaller is its fractal dimension, the smaller is the probability of finding quasiperiodic oscillation.

For experimental data, the determination of D_0 is difficult to obtain because the dimension measures a microscopic quantity of the plateaus widths, and in experimental data one can only observe the largest plateaus. Fortunately, an approximation D' for D_0 can be obtained from the largest plateaus by using the idea proposed in [22],

$$\left(\frac{S'}{S}\right)^{D'} + \left(\frac{S''}{S}\right)^{D'} = 1.$$ (3.7)

where S', S, and S'' are represented in Figure 4.

In case one has $K \cong 1$ ($K < 1$), we do not have a complete devil's staircase. In other words, winding numbers with denominators larger than a given $\tilde{Q}$ are cut off from the Farey Tree. Using this information we can estimate the value of K through the largest denominator observed [15],

$$\tilde{Q} \geq \frac{1}{1 - K}. \tag{3.8}$$

Finally, we would like to stress that while in (3.1) the plateaus of the devil's staircase are positioned at winding numbers defined by (3.3), in nature, devil's staircases have plateaus positioned at some accessible measurement.

In the driven Rayleigh-Bénard experiment [23, 24], convection rolls appear in a small brick-shaped cell filled with mercury, for a critical temperature difference between the upper and lower plates. As one perturbs the cell by a constant external magnetic field parallel to the axes of the rolls and by the introduction of an AC electrical current sheet pulsating with a frequency f_e and amplitude B, a devil's staircase is found in the variable f_i, the main frequency of the power spectra of the fluid velocity. As one varies the external frequency f_e, stable oscillations take place at a frequency ratio f_i/f_e, for a given value of f_e. In analogy to the devil's staircase of (3.1), f_e should be thought of as playing the same role of Ω in (3.1), and the ratio f_i/f_e as playing the same role of the winding number W.

A devil's staircase can also be observed (see [25]) in the amount of information H (topological entropy) that an unstable chaotic set has in terms of an interval of size ϵ, used to create the set. To generate the unstable chaotic set, we eliminate all possible trajectories of a stable chaotic set that visits this interval ϵ. In analogy with the devil's staircase of the circle map, ϵ should be related to $\Delta\Omega$, while H to W.

The first proof of a complete devil's staircase in a physical model was given in [26], in the one-dimensional ising model with convex long-range anti-ferromagnetic interactions.

In [27], it was found that a model for the El Niňo, a phenomenon that is the result of a tropical ocean-atmosphere interaction when coupled nonlinearly with the Earth's annual cycle, could undergo a transition to chaos through a series of frequency-locked steps. The overlapping of these resonances, which are the steps of the devil's staircase, leads to the chaotic behavior.

4. A Devil's Staircase in the Southern Brazilian Continental Shelf

Understanding the sea floor is a way for understanding how the sea level changed. Models associating local features in continental shelves and glacio global sea-level variations have been developed for the formation of wave-cut and wave-built terraces at various sea levels. In [28] a process-response model for the development of barrier beaches during sea level rise is described, and in [29] a model was developed for the formation of rock platforms during changing sea level [30]. These models, though, consider the formation of local features, as a particular terrace, and do not intend to make a large picture of a continental shelf.

On the other hand, fractal geometry approach has been applied to describe large-scale topographic features associated with geomorphic processes [31–34]. These models are considered to be of potential methodological interest in sea floor since "it is heuristically clear that seafloor or landscape topography is best described by fractal geometry, because of the different and increasingly detailed features appearing at increasing resolution" [35].

The purpose of this work is to show how to find a fractal geometry that fits reasonably good the SBCS.

The main premise that guides the application of the devil's staircase model to the shelf is that the rules found in quantities related to the widths and depths of the terraces obey the same rules found in a complete devil's staircase for the frequency-locked intervals $\Delta\Omega$ and their rational winding number, W. Thus, we assume that the terrace widths l_n play the same role of the frequency-locked intervals $\Delta\Omega$, and the terrace depths play the same role of the rational winding number W.

In order to interpret the shelf as a devil's staircase, we have to show that the terraces appear in positions which respect the Farey mediant, the rule that describes the winding number "positions" of the many plateaus. For that, we verify whether the position of the terraces at d_n can be associated with hypothetical frequency ratios, denoted by $w_n = p_n/q_n$, which respects the Farey mediant. In doing so, we want the metric of three adjacent terraces to respect the Farey mediant. In addition, we also assume that the larger terraces are the parents of the Farey Tree, while the smaller terraces between two larger ones are the daughters. Thus, for each triple of terrace, we want

$$\frac{d_{n+2} + d_n}{d_{n+1}} = \frac{w_{n+2} + w_n}{w_{n+1}}.$$
(4.1)

One could have considered other ways to relate the depths and the frequency ratios. The one chosen in (4.1) is used in order to account for the fact that while d_n is negative, w_n is not.

From (4.1) it becomes clear that for the further analysis the depth of a particular terrace does not play a very important role as the ratio between the depths of triple of terraces that contains this particular terrace. These ratios may eliminate the influence of the local morphology and the influence of the local sea-level dynamics into the formation of the shelf. Therefore, the proposed quantity might be suitable for an integrated analysis of the different shelves all over the world, specially the ones affected by local geomorphological characteristics.

From the Farey mediant, we have a way to obtain the frequency ratios associated to each terrace,

$$w_{n+1} = \frac{p_n + p_{n+2}}{q_n + q_{n+2}}.$$
(4.2)

Therefore, combining (4.1) and (4.2), we obtain

$$\frac{p_n + p_{n+2}}{q_n + q_{n+2}} = \frac{p_{n+2}/q_{n+2} + p_n/q_n}{E},$$
(4.3)

which results in

$$(p_n + p_{n+2})(E - 1)q_n q_{n+2} p_n = p_n q_{n+2}^2 + q_n^2 p_{n+2},$$
(4.4)
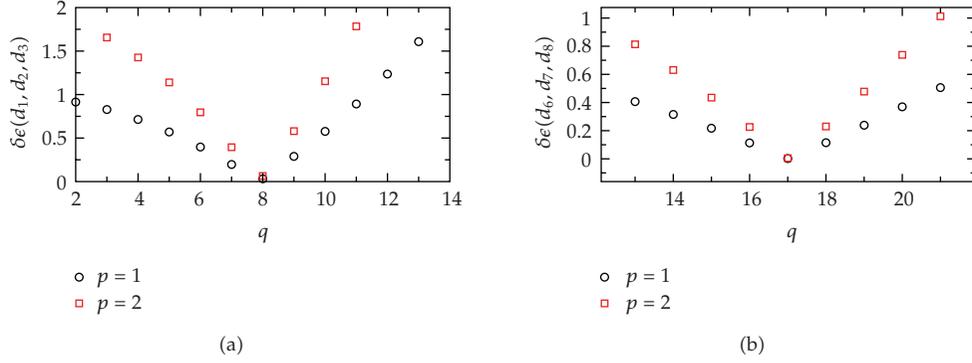
Figure 5: (a) Values of $\delta\epsilon(d_1, d_2, d_3)$ and (b) $\delta\epsilon(d_6, d_7, d_8)$, for different values of p and q. $\delta\epsilon$ is the difference between the left- and the right-hand sides of (4.4).

where E is defined by

$$\frac{d_{n+2} + d_n}{d_{n+1}} = E. \tag{4.5}$$

We do not expect to have (4.4) satisfied. We only require that the difference between the left- and right-hand sides of this equation, regarded as $\delta\epsilon$, is the lowest possible, among all possible values for p_m and q_m (with $m = n, n + 2$), for a given E, with the restriction that the considered largest terraces are related to largest plateaus of (3.1), and thus $p_{m+2} = p_m$ and $q_{m+2} = q_m + 1$, and $\delta\epsilon \ll 1$. Doing so, we find the rationals associated with the terraces, which are shown in Table 2. The minimal value of $\delta\epsilon$, denoted by $\min[\delta\epsilon]$, is $\min[\delta\epsilon(d_1, d_2, d_3)] = 0.032002$, with $p_1/q_1 = 1/8$ for the terrace 1, and $p_3/q_3 = 1/9$, for the terrace 3. We also find that $\min[\delta\epsilon(d_6, d_7, d_8)] = 0.002344$, with $p_6/q_6 = 1/17$, for the terrace 6, and $p_8/q_8 = 1/18$, for the terrace 8. These minimal values can be seen in Figures 5(a) and 5(b), where we show the values of $\delta\epsilon(d_1, d_2, d_3)$, in Figure 5(a), and the values of $\delta\epsilon(d_6, d_7, d_8)$, in Figure 5(b), for different values of p and q. Using bigger values for p has the only effect to increase the value of $\delta\epsilon$.

We have not identified rationals that can be associated with the terraces 4 and 5, which means that for p and q within $p = [1, 50]$ and $q = [1, 400]$, we find that $\delta\epsilon > 1$. We have assumed that they could be either a daughter or a parent.

From now on, when convenient, we will drop the index n and represent each terrace by the associated frequency ratio. So, the terrace 1, for $n = 1$, is represented as the terrace with $w = 1/8$.

Table 2 can be represented in the form of the Farey Tree as shown in Figure 6. The branch of rationals in the Farey Tree in the form $1/q$ belongs to the most stable branch, which means that the observed terraces should have the largest widths. We believe that the other less important branches of the complete devil's staircase present in the data were smoothed out by the action of the waves and the flow streams throughout the time, and at the present time cannot be observed.

Notice that as the time goes by, the frequency ratios are increasing their absolute value, which means that if this tendency is preserved in the future, we should expect to see larger terraces.

Table 2: Integers associated with the n considered terraces, with $n = 1, \ldots, 8$.

n	p_n	q_n
1	1	8
2	2	17
3	1	9
4	—	—
5	—	—
6	1	17
7	2	35
8	1	18
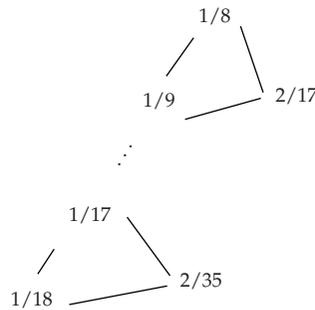
Figure 6: Farey Tree representing the frequency ratios associated with the major terraces.

In the following, we will try to recover in the experimental profile, the universal scaling laws of (3.6) and (3.7). Regarding (3.6), we find that l scales as $1/q^{-3.60}$, as shown in Figure 7, which is the expected global universal scaling for a complete devil's staircase. Regarding (3.7), and calculating S', S'', and S using the triple of terraces with widths $l(w = 1/8)$, $l(w = 2/17)$, and $l(w = 1/9)$, as represented in Figure 4, we find $D' = 0.89$. Using the triple of terraces ($n = 3$, $n = 4$, $n = 5$), we find that $D' = 0.87$. Both results are very close from the universal fractal dimension $D_0 \cong 0.87$, found for a complete devil's staircase.

5. Fitting the SBCS

Motivated by our previous results, we fit the observed shelf as a complete devil's staircase, using (3.1). Notice that the only requirement for (3.1) to generate a complete devil's staircase is that the function g has a cubic inflection point at the critical parameter $K = 1$. Whether (3.1) is indeed an optimal modeling for the shelf is beyond the scope of the present study. We only chose this map because it is a well-known system, and it captures most of the relevant characteristic a dynamical system needs to fulfill in order to create a devil's staircase.

We model the SBCS as a complete devil's staircase, but we rescale the winding number W into the observed terrace depth. So, we transform the complete devil's staircase of Figure 8 as good as possible into the profile of Figure 1(b), by rescaling the vertical axis of the staircase in Figure 8.

We do that by first obtaining the function F (see Figure 9) whose application into the terrace depth $d(w)$ gives the frequency ratio $w_n = p_n/q_n$ associated with the terrace.
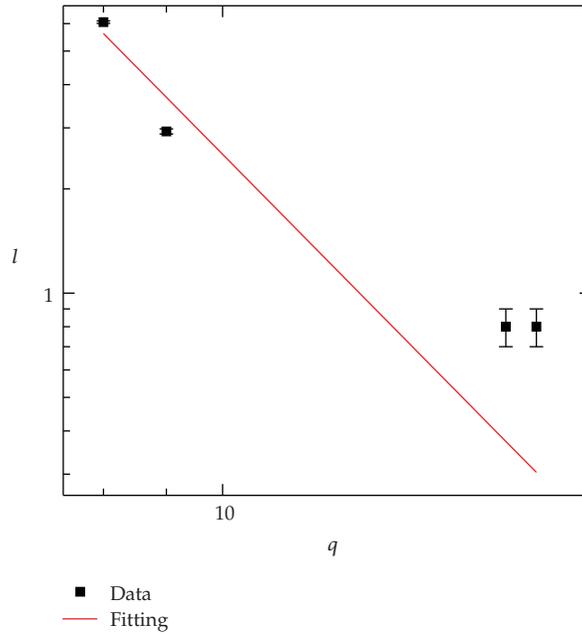
Figure 7: Scaling between the $1/q$-terrace widths and the value of q.
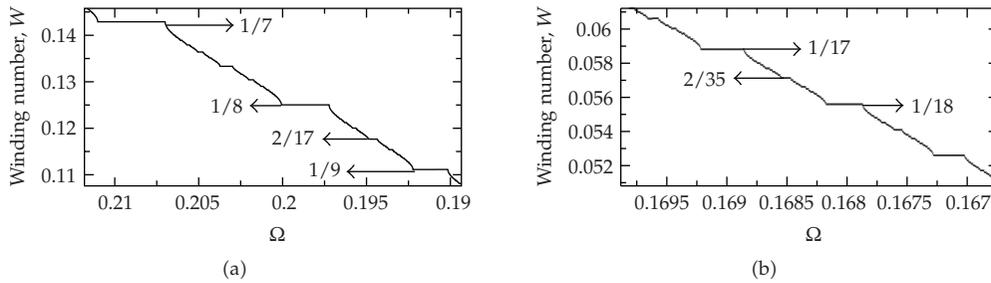
Figure 8: Magnifications of the small box of Figure 3, showing the plateaus of the devil's staircase of (3.1) that appear for the same frequency ratios associated with the triple of terraces $w = (1/8, 2/17, 1/9)$, in (a), and $w = (1/17, 2/35, 1/18)$, in (b).

For the triple of terraces $w = (1/8, 2/17, 1/9)$, we obtain

$$F(d[\text{km}]) = 0.14219 + 0.00057853d \, [\text{km}], \tag{5.1}$$

and for the triple of terraces $w = (1/17, 2/35, 1/18)$, we obtain

$$F(d[\text{km}]) = 0.080941 + 0.00029786d \, [\text{km}]. \tag{5.2}$$

Therefore, we assume that, locally, the frequency ratios are linearly related to the depth of the terraces.
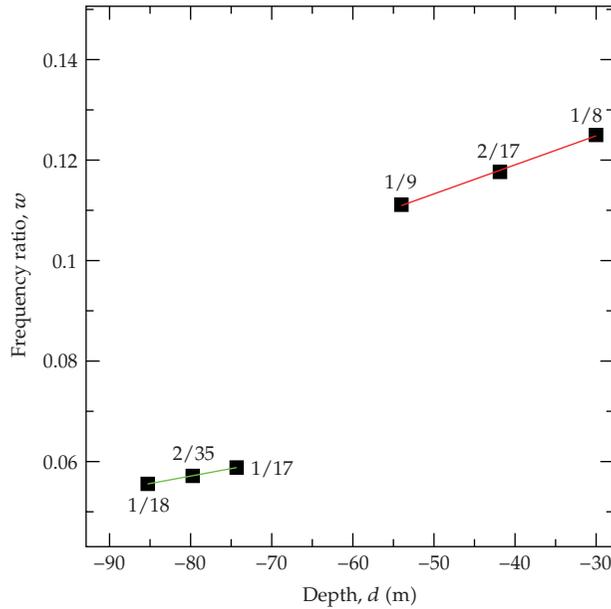
Figure 9: The function F, which is a linear relation between the frequency ratios associated with the terraces and their depths for the triple of terraces $w = (1/8, 2/17, 1/9)$ and $w = (1/17, 2/35, 1/18)$.
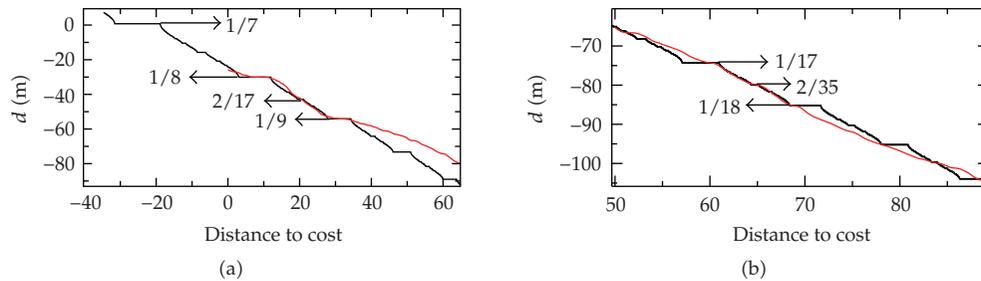
Figure 10: (a) Rescaling of Figure 8(a) in black, showing that the devil's staircase fits well the terraces with $w = (1/8, 2/17, 1/9)$ of the profile of Figure 1(b), in gray. (b) Rescaling of Figure 8(b) in black, showing that the devil's staircase fits well the terraces with $w = (1/17, 2/35, 1/18)$ of the profile of Figure 1(b), in gray.

Then, we rescale the vertical axis of the staircases in Figures 8(a)-8(b) and calculate an equivalent depth, d, for the winding number W by using

$$d = F^{-1}(W). \tag{5.3}$$

We also allow tiny adjustments in the axes for a best fitting. The result is shown in Figure 10(a) for the triple of terraces $w = (1/8, 2/17, 1/9)$ and in Figure 10(b) for the triple of terraces $w = (1/17, 2/35, 1/18)$. We see that locally, for a short time interval, we can have a good agreement of the terrace widths and positions, with the rescaled devil's staircase. However, globally, the fitting in Figure 10(a) does not do well, as it is to be expected since

the function F is only locally well defined, and it changes depending on the depths of the terraces. Notice however that this short time interval is not so short since the time interval correspondent to a triple of terraces is of the order of a few hundred years.

The assumption made that $K = 1$ is also supported from (3.8). Using this equation, we can obtain an estimation of the maximum value of K from a terrace with a frequency ratio that has the largest denominator. In our case, we observed $w = 2/35$. Using $\widetilde{Q} = 35$ in (3.8), we obtain $K \leq 0.97$.

In Figure 10(a), we see a $1/7$ plateau positioned in the zero sea level. That is the current level. Thus, the model predicts that nowadays we should have a large terrace, which might imply in an average stabilization of the sea level for a large period of time. However, this prediction might not correspond to reality if the sea dynamics responsible for the creation of the observed continental shelf suffered structurally modifications.

6. Conclusions

We have shown some experimental evidences that the Southern Brazilian Continental Shelf (SBCS) has a structure similar to the devil's staircase. That means that the terraces found in the bottom of the sea are not randomly distributed but they occur following a dynamical rule. This finding lead us to model the SBCS as a complete devil's staircase, in which, between two real terraces, we suppose an infinite number of virtual (smaller) ones. We do not find these later ones, either because they have been washed out by the stream flow or simply due to the fact that the time period in which the sea-level dynamics (SLDs) stayed locked was not sufficient to create a terrace. By our hypothesis, the SLDs creates a terrace if it is a dynamics in which two relevant frequencies are locked in a rational ratio.

This special phase-locked dynamics possesses a critical characteristic; large changes in some parameter responsible for a relevant natural frequency of the SLDs might not destroy the phase-locked regime, which might imply that the averaged sea level would remain still. On the other hand, small changes in the parameter associated with an external forcing of the SLDs could be catastrophic, inducing a chaotic SLDs, what would mean a turbulent averaged sea-level rising/regression.

In order to interpret the shelf as a devil's staircase, we have shown that the terraces appear in an organized way according to the Farey mediant, the rule that describes the way plateaus appear in the devil's staircase. That allows us to "name" each terrace depth, d_n, by a rational number, w_n, regarded as the hypothetical frequency ratio. Arguably, these ratios represent the ratio between real frequencies that are present in the SLDs. It is not the scope of the present work to verify this hypothesis; however, one way to check if the hypothetical frequency ratios are more than just a mathematical artifact would be to check if the SLDs has, nowadays, two relevant frequencies in a ratio $1/7$, as predicted.

The newly proposed approach to characterize the SBCS relies mainly on the ratios between terraces widths and between terraces depths. While single terrace widths and depths are strongly influenced by local properties of the costal morphology and the local sea-level variations, the ratios between terrace widths and depths should be a strong indication of the global sea-level variations. Therefore, the newly proposed approach has a general character, and it seems to be appropriated as a tool of analysis to other continental shelves around the world.

Reminding that the local morphology of the studied area, the "Serra do Mar," does not have a strong impact in the formation of the shelf and assuming that the local SLDs is not

directly involved in the formation of the large terraces considered in our analyses, thus, our results should reflect mainly the action of the global SLDs.

In fact, analytical models [36–38] that study the relation between the sea-level variation and the climate change during the last three decades have shown that there are preferred oceanic locations at which the global sea-level variation can be well described by the local variation.

If the characteristics observed locally in the São Paulo Bight indeed reflect the effect of the global SLDs, then the global SLDs might be a critical system. Hopefully, the environmental changes caused by the modern men have not yet made any significant change in a relevant parameter of this global system.

Acknowledgments

References

[1] K. Lambeck and J. Chappell, "Sea level change through the last glacial cycle," *Science*, vol. 292, no. 5517, pp. 679–686, 2001.

[2] H. Heinrich, "Origin and consequences of cyclic ice rafting in the Northeast Atlantic Ocean during the past 130,000 years," *Quaternary Research*, vol. 29, no. 2, pp. 142–152, 1988.

[3] T. Goslar, M. Arnold, N. Tisnerat-Laborde, J. Czernik, and K. Więckowski, "Variations of Younger Dryas atmospheric radiocarbon explicable without ocean circulation changes," *Nature*, vol. 403, no. 6772, pp. 877–880, 2000.

[4] J. Adams, M. Maslin, and E. Thomas, "Sudden climate transitions during the Quaternary," *Progress in Physical Geography*, vol. 23, no. 1, pp. 1–36, 1999.

[5] J. T. Andrews, "Abrupt changes (Heinrich events) in late Quaternary North Atlantic marine environments: a history and review of data and concepts," *Journal of Quaternary Science*, vol. 13, no. 1, pp. 3–16, 1998.

[6] W. S. Broecker, "Paleocean circulation during the last deglaciation: a bipolar seesaw?" *Paleoceanography*, vol. 13, no. 2, pp. 119–121, 1998.

[7] K. C. Taylor, P. A. Mayewski, R. B. Alley, et al., "The Holocene-Younger Dryas transition recorded at Summit, Greenland," *Science*, vol. 278, no. 5339, pp. 825–827, 1997.

[8] V. V. Furtado, M. M. Mahiques, and M. G. Tessler, "Utilização de feições topográficas submersas na correlação de paleoníveis marinhos: uma avaliação," in *Proceedings of Anais do III Congresso da Associação Brasileira de Estudos do Quaternário (ABEQUA '92)*, pp. 175–186, Belo Horizonte, Brazil, October 1992.

[9] L. A. Conti and V. V. Furtado, "Geomorphological indicators of Quaternary sea levels on the continental shelf of southeastern Brazil," in *Proceedings of the 1st Annual Conference of International Geological Correlation Program (IGCP '01)*, Hong Kong, October 2001.

[10] I. C. S. Corrêa, "Les variations du niveau de la mer durant les derniers 17.500 ans BP: l'exemple de la plate-forme continentale du Rio Grande do Sul-Brésil," *Marine Geology*, vol. 130, no. 1-2, pp. 163–178, 1996.

[11] B. B. Mandelbrot, *Fractals: Form, Chance, and Dimension*, W.H. Freeman, San Francisco, Calif, USA, 1977.

[12] U. C. Herzfeld, I. I. Kim, and J. A. Orcutt, "Is the ocean floor a fractal?" *Mathematical Geology*, vol. 27, no. 3, pp. 421–462, 1995.

[13] J. A. Goff, D. L. Orange, L. A. Mayer, and J. E. Hughes Clarke, "Detailed investigation of continental shelf morphology using a high-resolution swath sonar survey: the Eel margin, northern California," *Marine Geology*, vol. 154, no. 1–4, pp. 255–269, 1999.

[14] J.-C. Mareschal, "Fractal reconstruction of sea-floor topography," *Pure and Applied Geophysics*, vol. 131, no. 1-2, pp. 197–210, 1989.

[15] M. H. Jensen, P. Bak, and T. Bohr, "Transition to chaos by interaction of resonances in dissipative systems—I: circle maps," *Physical Review A*, vol. 30, no. 4, pp. 1960–1969, 1984.

[16] J. Argyris, G. Faust, and M. Haase, *An Exploration of Chaos: An Introduction for Natural Scientists and Engineers*, Texts on Computational Mechanics, 7, North-Holland, Amsterdam, The Netherlands, 1994.

[17] V. I. Arnol'd, "Small divisors—I: on mappings of the circle onto itself," *American Mathematical Society Translations: Series A*, vol. 46, pp. 213–284, 1965.

[18] V. I. Arnol'd, *Geometrical Methods in the Theory of Ordinary Differential Equations*, vol. 250 of *Grundlehren der Mathematischen Wissenschaften*, Springer, New York, NY, USA, 2nd edition, 1988.

[19] M. H. Jensen, P. Bak, and T. Bohr, "Complete devil's staircase, fractal dimension, and universality of mode- locking structure in the circle map," *Physical Review Letters*, vol. 50, no. 21, pp. 1637–1639, 1983.

[20] P. Cvitanovic, B. Shraiman, and B. Söderberg, "Scaling laws for mode lockings in circle maps," *Physica Scripta*, vol. 32, no. 4, pp. 263–270, 1985.

[21] J. D. Farmer, E. Ott, and J. A. Yorke, "The dimension of chaotic attractors," *Physica D*, vol. 7, no. 1–3, pp. 153–180, 1983.

[22] H. G. E. Hentschel and I. Procaccia, "The infinite number of generalized dimensions of fractals and strange attractors," *Physica D*, vol. 8, no. 3, pp. 435–444, 1983.

[23] J. Stavans, F. Heslot, and A. Libchaber, "Fixed winding number and the quasiperiodic route to chaos in a convective fluid," *Physical Review Letters*, vol. 55, no. 6, pp. 596–599, 1985.

[24] M. H. Jensen, L. P. Kadanoff, A. Libchaber, I. Procaccia, and J. Stavans, "Global universality at the onset of chaos: results of a forced Rayleigh-Bénard experiment," *Physical Review Letters*, vol. 55, no. 25, pp. 2798–2801, 1985.

[25] M. S. Baptista, E. E. Macau, and C. Grebogi, "Conditions for efficient chaos-based communication," *Chaos*, vol. 13, no. 1, pp. 145–150, 2003.

[26] P. Bak and R. Bruinsma, "One-dimensional Ising model and the complete devil's staircase," *Physical Review Letters*, vol. 49, no. 4, pp. 249–251, 1982.

[27] F.-F. Jin, J. D. Neelin, and M. Ghil, "El Niño on the devil's staircase: annual subharmonic steps to chaos," *Science*, vol. 264, no. 5155, pp. 70–72, 1994.

[28] D. L. Inman, S. A. Jenkins, and P. M. Masters, *Modeling Platforms, Terraces and Coastal Evolution*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[29] J. E. A. Storms, G. J. Weltje, J. J. van Duke, C. R. Geel, and S. B. Kroonenberg, "Process-response modeling of wave-dominated coastal systems: simulating evolution and stratigraphy on geological timescales," *Journal of Sedimentary Research*, vol. 72, no. 2, pp. 226–239, 2002.

[30] A. S. Trenhaile, "Modeling the development of marine terraces on tectonically mobile rock coasts," *Marine Geology*, vol. 185, no. 3-4, pp. 341–361, 2002.

[31] A. Baldassarri, M. Montuori, O. Prieto-Ballesteros, and S. C. Manrubia, "Fractal properties of isolines at varying altitude revealing different dominant geological processes on Earth," *Journal of Geophysical Research E*, vol. 113, no. 9, Article ID E09002, 2008.

[32] Q.-C. Sung and Y.-C. Chen, "Self-affinity dimensions of topography and its implications in morphotectonics: an example from Taiwan," *Geomorphology*, vol. 62, no. 3-4, pp. 181–198, 2004.

[33] T. H. Wilson and J. Dominic, "Fractal interrelationships between topography and structure," *Earth Surface Processes and Landforms*, vol. 23, no. 6, pp. 509–525, 1998.

[34] P. S. Dodds and D. H. Rothman, "Scaling, universality, and geomorphology," *Annual Review of Earth and Planetary Sciences*, vol. 28, pp. 571–610, 2000.

[35] U. C. Herzfeld and C. Overbeck, "Analysis and simulation of scale-dependent fractal surfaces with application to seafloor morphology," *Computers & Geosciences*, vol. 25, no. 9, pp. 979–1007, 1999.

[36] J. A. Clark, W. E. Farrell, and W. R. Peltier, "Global changes in postglacial sea level: a numerical calculation," *Quaternary Research*, vol. 9, no. 3, pp. 265–287, 1978.

[37] K. Fleming, P. Johnston, D. Zwartz, Y. Yokoyama, K. Lambeck, and J. Chappell, "Refining the eustatic sea-level curve since the Last Glacial Maximum using far- and intermediate-field sites," *Earth and Planetary Science Letters*, vol. 163, no. 1–4, pp. 327–342, 1998.

[38] R. B. Alley, P. U. Clark, P. Huybrechts, and I. Joughin, "Ice-sheet and sea-level changes," *Science*, vol. 310, no. 5747, pp. 456–460, 2005.

Research Article

Solitons, Peakons, and Periodic Cuspons of a Generalized Degasperis-Procesi Equation

Jiangbo Zhou and Lixin Tian

Nonlinear Scientific Research Center, Faculty of Science, Jiangsu University, Zhenjiang, Jiangsu 212013, China

Correspondence should be addressed to Jiangbo Zhou, zhoujiangbo@yahoo.cn

We employ the bifurcation theory of planar dynamical systems to investigate the exact travelling wave solutions of a generalized Degasperis-Procesi equation $u_t - u_{xxt} + 4uu_x + \gamma(u - u_{xx})_x = 3u_x u_{xx} + uu_{xxx}$. The implicit expression of smooth soliton solutions is given. The explicit expressions of peaked soliton solutions and periodic cuspon solutions are also obtained. Further, we show the relationship among the smooth soliton solutions, the peaked soliton solutions, and the periodic cuspon solutions. The physical relevance of the found solutions and the reason why these solutions can exist in this equation are also given.

1. Introduction

Recently, Degasperis and Procesi [1] derived a nonlinear dispersive equation

$$u_t - u_{xxt} + 4uu_x = 3u_x u_{xx} + uu_{xxx} \tag{1.1}$$

which is called the Degasperis-Procesi equation. Here $u(t, x)$ represents the fluid velocity at time t in the x direction in appropriate nondimensional units (or, equivalently the height of the water's free surface above a flat bottom). The nonlinear convection term uu_x in (1.1) causes the steepening of wave form, whereas the nonlinear dispersion effect term $3u_x u_{xx} + uu_{xxx} = ((1/2)u^2)_{xxx}$ in (1.1) makes the wave form spread. Equation (1.1) can be regarded as a model for nonlinear shallow water dynamics. Degasperis et al. [2] showed that the (1.1) is integrable by deriving a Lax pair and a bi-Hamiltonian structure for it. Yin proved local well posedness to (1.1) with initial data $u_0 \in H^s(\mathbb{R})$, $s > 3$ on the line [3] and on the circle [4]. The global existence of strong solutions and weak solutions to (1.1) is investigated in [4–10]. The solution to Cauchy problem of (1.1) can also blow up in finite time when the initial data satisfies certain sign condition[7–10]. Vakhnenko and Parkes [11] obtained

periodic and solitary-wave solutions of (1.1). Matsuno [12, 13] obtained multisoliton, cusp and loop soliton solutions of (1.1). Lundmark and Szmigielski [14] investigated multipeakon solutions of (1.1). Lenells [15] classified all weak travelling wave solutions. The shock wave solutions of (1.1) are investigated in [16, 17].

Yu and Tian [18] investigated the following generalized Degasperis-Procesi equation:

$$u_t - u_{xxt} + 4uu_x = 3u_x u_{xx} + uu_{xxx} - \gamma u_{xxx}, \tag{1.2}$$

where γ is a real constant, and the term u_{xxx} denotes the linear dispersive effect. They obtained peaked soliton solutions and period cuspon solutions of (1.2). Unfortunately, they did not obtain smooth soliton solutions of (1.2).

In this paper, we are interesting in the following generalized Degasperis-Procesi equation:

$$u_t - u_{xxt} + 4uu_x + \gamma(u - u_{xx})_x = 3u_x u_{xx} + uu_{xxx}, \tag{1.3}$$

where γ is a real constant, the term u_x denotes the dissipative effect and the term u_{xxx} represents the linear dispersive effect. Employing the bifurcation theory of planar dynamical systems, we obtain the analytic expressions of smooth solitons, peaked solitons, and period cuspons of (1.3). Our work covers and supplements the results obtained in [18].

The remainder of the paper is organized as follows. In Section 2, using the travelling wave transformation, we transform (1.3) into the planar dynamical system (2.3) and then discuss bifurcations of phase portraits of system (2.3). In Section 3, we obtain the implicit expression of smooth solitons, the explicit expressions of peaked solitons and periodic cuspon solutions. At the same time, we show that the limits of smooth solitons and periodic cusp waves are peaked solitons. In Section 4, we discuss the physical relevance of the found solutions and give the reason why these solutions can exist in (1.3).

2. Bifurcations of Phase Portraits of System (2.3)

We look for travelling wave solutions of (1.3) in the form of $u(x, t) = \varphi(x - ct) = \varphi(\xi)$, where c is the wave speed and $\xi = x - ct$. Substituting $u = \varphi(\xi)$ into (1.3), we obtain

$$-c\varphi' + c\varphi''' + 4\varphi\varphi' - 3\varphi'\varphi'' - \varphi\varphi''' + \gamma\varphi' - \gamma\varphi''' = 0. \tag{2.1}$$

By integrating (2.1) once we have

$$\varphi''(\varphi - c + \gamma) = g - (c - \gamma)\varphi + 2\varphi^2 - (\varphi')^2, \tag{2.2}$$

where g is the integral constant.

Let $y = \varphi'$, then we get the following planar dynamical system:

$$\frac{d\varphi}{d\xi} = y,$$

$$\frac{dy}{d\xi} = \frac{g - (c - \gamma)\varphi + 2\varphi^2 - y^2}{\varphi - c + \gamma},$$

(2.3)

with a first integral

$$H(\varphi, y) = (\varphi - c + \gamma)^2 \left(y^2 - \varphi^2 - g\right) = h,$$

(2.4)

where h is a constant.

Note that (2.3) has a singular line $\varphi = c - \gamma$. To avoid the line temporarily we make transformation $d\xi = (\varphi - c + \gamma)d\zeta$. Under this transformation, (2.3) becomes

$$\frac{d\varphi}{d\zeta} = (\varphi - c + \gamma)y,$$

$$\frac{dy}{d\zeta} = g - (c - \gamma)\varphi + 2\varphi^2 - y^2.$$

(2.5)

System (2.3) and system (2.5) have the same first integral as (2.4). Consequently, system (2.5) has the same topological phase portraits as system (2.3) except for the straight line $\varphi = c - \gamma$. Obviously, $\varphi = c - \gamma$ is an invariant straight-line solution for system (2.5).

For a fixed h, (2.4) determines a set of invariant curves of system (2.5). As h is varied, (2.4) determines different families of orbits of system (2.5) having different dynamical behaviors. Let $M(\varphi_e, y_e)$ be the coefficient matrix of the linearized system of (2.5) at the equilibrium point (φ_e, y_e), then

$$M(\varphi_e, y_e) = \begin{pmatrix} y_e & \varphi_e - c + \gamma \\ 4\varphi_e - c + \gamma & -2y_e \end{pmatrix},$$

(2.6)

and at this equilibrium point, we have

$$J(\varphi_e, y_e) = \det M(\varphi_e, y_e) = -2y_e^2 - (\varphi_e - c + \gamma)(4\varphi_e - c + \gamma),$$

$$p(\varphi_e, y_e) = \text{trace}(M(\varphi_e, y_e)) = -y_e.$$

(2.7)

By the qualitative theory of differential equations (see [19]), for an equilibrium point of a planar dynamical system, if $J < 0$, then this equilibrium point is a saddle point; it is a center point if $J > 0$ and $p = 0$; if $J = 0$ and the Poincaré index of the equilibrium point is 0, then it is a cusp.

By using the first integral value and properties of equilibrium points, we obtain the bifurcation curves as follows:

$$g_1(c) = \frac{(c-\gamma)^2}{8},$$

$$g_2(c) = -(c-\gamma)^2. \tag{2.8}$$

Obviously, the two curves have no intersection point and $g_2(c) < 0 < g_1(c)$ for arbitrary constants $c \neq \gamma$.

Using bifurcation method of vector fields (e.g., [19]), we have the following result which describes the locations and properties of the singular points of system (2.5).

Theorem 2.1. *For a given constant wave speed $c \neq 0$, let*

$$\varphi_0^{\pm} = \frac{c-\gamma \pm \sqrt{(c-\gamma)^2 - 8g}}{4} \quad \text{for } g \leq g_1(c), \tag{2.9}$$

$$y_0^{\pm} = \pm\sqrt{(c-\gamma)^2 + g} \quad \text{for } g \geq g_2(c). \tag{2.10}$$

When $c = \gamma$,

(1) *if $g < 0$, then system (2.5) has two equilibrium points $(-\sqrt{-g/2}, 0)$ and $(\sqrt{-g/2}, 0)$, which are saddle points;*

(2) *if $g = 0$, then system (2.5) has only one equilibrium point $(0,0)$, which is a cusp;*

(3) *if $g > 0$, then system (2.5) has two equilibrium points $(0, -\sqrt{g})$ and $(0, \sqrt{g})$, which are saddle points.*

When $c \neq \gamma$,

(1) *if $g < g_2(c)$, then system (2.5) has two equilibrium points $(\varphi_0^-, 0)$ and $(\varphi_0^+, 0)$. They are saddle points:*

   (i) *if $c > \gamma$, then $\varphi_0^- < -1/2(c-\gamma) < 1/4(c-\gamma) < c - \gamma < \varphi_0^+$,*

   (ii) *if $c < \gamma$, then $\varphi_0^- < c - \gamma < 1/4(c-\gamma) < -1/2(c-\gamma) < \varphi_0^+$;*

(2) *if $g = g_2(c)$, then system (2.5) has three equilibrium points $(\varphi_0^-, 0)$, $(\varphi_0^+, 0)$, and $(c-\gamma, 0)$. $(c-\gamma, 0)$ is a cusp:*

   (i) *if $c > \gamma$, then $\varphi_0^- = -(1/2)(c-\gamma) < (1/4)(c-\gamma) < c - \gamma = \varphi_0^+$. $(\varphi_0^-, 0)$ is a saddle point, while $(\varphi_0^+, 0)$ is a degenerate center point,*

   (ii) *if $c < \gamma$, then $\varphi_0^- = c - \gamma < 1/4(c-\gamma) < -1/2(c-\gamma) = \varphi_0^+$. $(\varphi_0^-, 0)$ is a degenerate center point, while $(\varphi_0^+, 0)$ is a saddle point;*
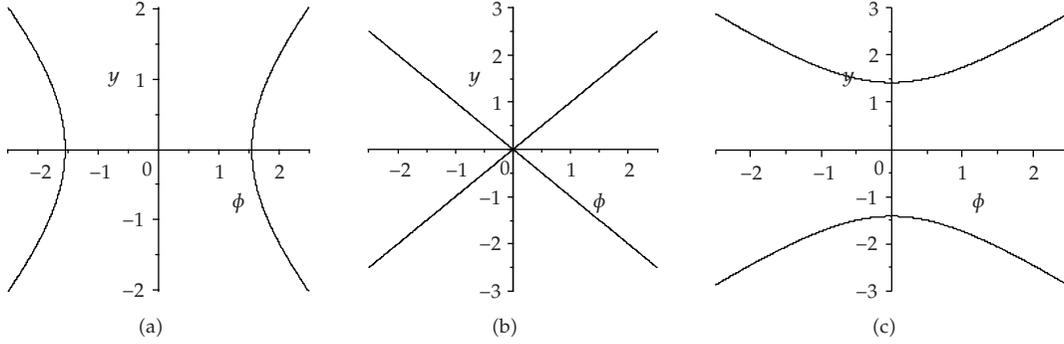
Figure 1: The phase portraits of system $(2.5)(c = \gamma)$. (a) $g < 0$; (b) $g = 0$; (c) $g > 0$.

(3) if $g_2(c) < g < g_1(c)$, then system (2.5) has four equilibrium points $(\varphi_0^-, 0)$, $(\varphi_0^+, 0)$, $(c - \gamma, y_0^-)$, and $(c - \gamma, y_0^+)$. $(c - \gamma, y_0^-)$ and $(c - \gamma, y_0^+)$ are two saddle points:

   (i) if $c > \gamma$, then $\varphi_0^- < 1/4(c - \gamma) < \varphi_0^+ \leq c - \gamma$. $(\varphi_0^-, 0)$ is a saddle point, while $(\varphi_0^+, 0)$ is a center point,

   (ii) if $c < \gamma$, then $c - \gamma \leq \varphi_0^- < 1/4(c - \gamma) < \varphi_0^+$. $(\varphi_0^-, 0)$ is a center point, while $(\varphi_0^+, 0)$ is a saddle point.

   Specially, when $g = 0$,

(i) if $c > \gamma$, then the three saddle points $(\varphi_0^-, 0)$, $(c - \gamma, y_0^-)$, and $(c - \gamma, y_0^+)$ form a triangular orbit which encloses the center point $(\varphi_0^+, 0)$,

(ii) if $c < \gamma$, then the three saddle points $(\varphi_0^+, 0)$, $(c - \gamma, y_0^-)$, and $(c - \gamma, y_0^+)$ form a triangular orbit which encloses the center point $(\varphi_0^-, 0)$;

(4) if $g = g_1(c)$, then system (2.5) has three equilibrium points $((c - \gamma)/4, 0)$, $(c - \gamma, y_0^-)$, and $(c - \gamma, y_0^+)$. $((c - \gamma)/4, 0)$ is a degenerate center point, while $(c - \gamma, y_0^-)$ and $(c - \gamma, y_0^+)$ are two saddle points;

(5) if $g > g_1(c)$, then system (2.5) has two equilibrium points $(c - \gamma, y_0^-)$ and $(c - \gamma, y_0^+)$, which are saddle points.

Corresponding to the case $c = \gamma$ and the case $c \neq \gamma$, we show the phase portraits of system (2.5) in Figures 1 and 2, respectively.

3. Solitons, Peakons, and Periodic Cusp Wave Solutions

Theorem 3.1. *Given arbitrary constant $c \neq \gamma$, let $\xi = x - ct$, then*

   (1) *when $0 < g < g_1(c)$,*

   (i) *if $c > \gamma$, then (1.3) has the following smooth hump-like soliton solutions:*

$$\beta_1(\varphi_1^+) = \beta_1(\varphi)e^{-|\xi|} \quad for \ \varphi_0^- < \varphi < \varphi_1^+, \tag{3.1}$$

   (ii) *if $c < \gamma$, then (1.3) has the following smooth valley-like soliton solutions:*

$$\beta_2(\varphi) = \beta_2(\varphi_1^-)e^{-|\xi|} \quad for \ \varphi_1^- < \varphi < \varphi_0^+, \tag{3.2}$$
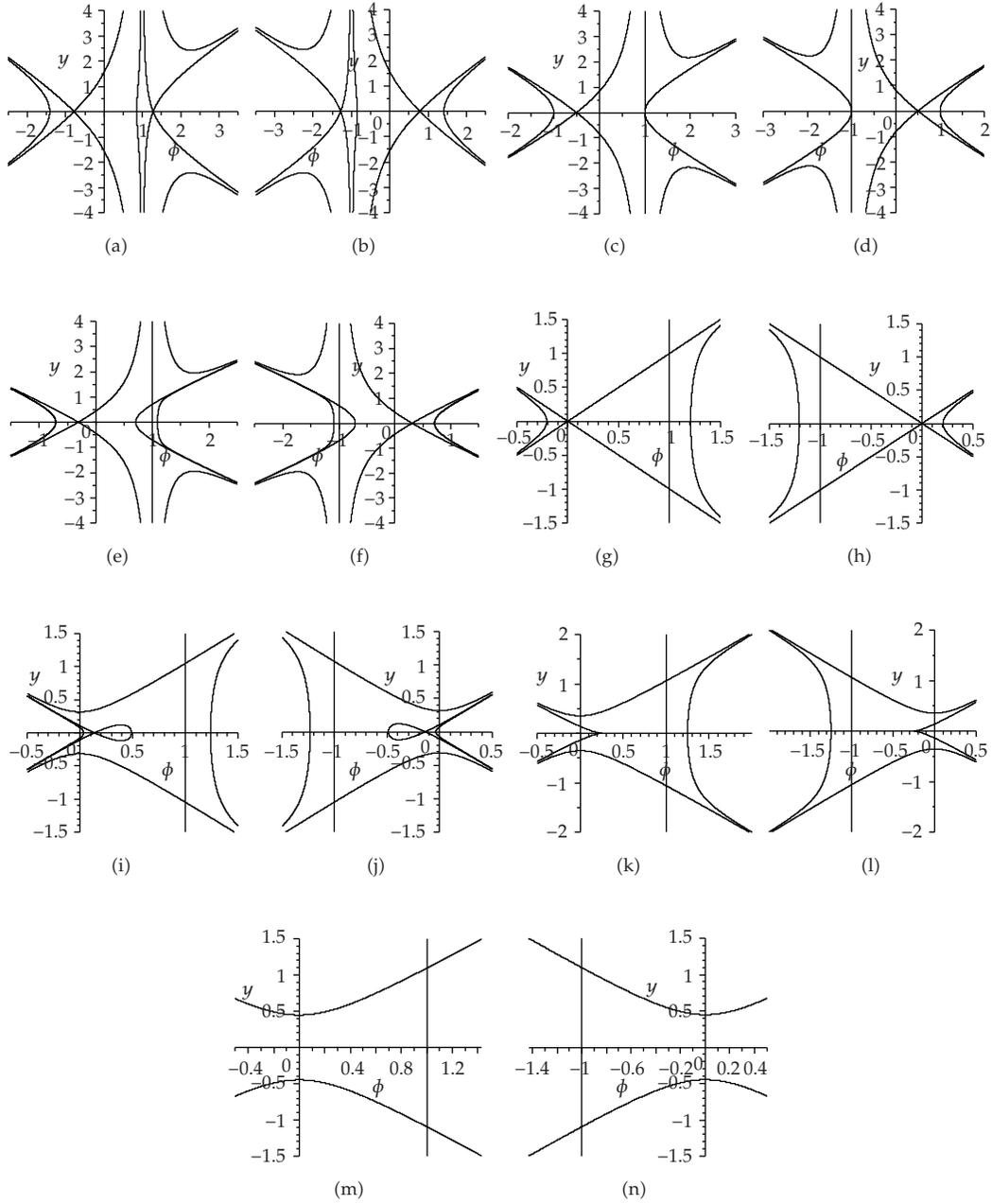
Figure 2: The phase portraits of system (2.5) ($c \neq \gamma$). (a) $g < g_2(c), c > \gamma$; (b) $g < g_2(c), c < \gamma$; (c) $g = g_2(c), c > \gamma$; (d)$g = g_2(c), c < \gamma$; (e) $g_2(c) < g < 0, c > \gamma$; (f) $g_2(c) < g < 0, c < \gamma$; (g) $g = 0, c > \gamma$; (h) $g = 0, c < \gamma$; (i) $0 < g < g_1(c), c > \gamma$; (j) $0 < g < g_1(c), c < \gamma$; (k) $g = g_1(c), c > \gamma$; (l) $g = g_1(c), c < \gamma$; (m)$g > g_1(c), c > \gamma$; (n)$g > g_1(c), c < \gamma$.

(2) *when $g = 0$, (1.3) has the following peaked soliton solutions:*

$$\varphi = (c - \gamma)e^{-|\xi|}, \tag{3.3}$$

(3) *when $g_2(c) < g < 0$,*

(i) *if $c > \gamma$, then (1.3) has the following periodic cusp wave solutions:*

$$u(x,t) = \varphi_3(x - ct - 2nT) \quad \text{for } (2n-1)T < x - ct < (2n+1)T, \tag{3.4}$$

(ii) *if $c < \gamma$, then (1.3) has the following periodic cusp wave solutions:*

$$u(x,t) = \varphi_4(x - ct - 2nT) \quad \text{for } (2n-1)T < x - ct < (2n+1)T, \tag{3.5}$$

where

$$\beta_1(\varphi) = \frac{\left(2\sqrt{\varphi^2 + l_1\varphi + l_2} + 2\varphi + l_1\right)(\varphi - \varphi_0^-)^{\alpha_1}}{\left(2\sqrt{a_1}\sqrt{\varphi^2 + l_1\varphi + l_2} + b_1\varphi + l_3\right)^{\alpha_1}},$$

$$\beta_2(\varphi) = \frac{\left(2\sqrt{\varphi^2 + m_1\varphi + m_2} + 2\varphi + m_1\right)(\varphi - \varphi_0^+)^{\alpha_2}}{\left(2\sqrt{a_2}\sqrt{\varphi^2 + m_1\varphi + m_2} + b_2\varphi + m_3\right)^{\alpha_2}},$$

$$l_1 = -\frac{3(c - \gamma) + \sqrt{(c - \gamma)^2 - 8g}}{2},$$

$$l_2 = \frac{3(c - \gamma)^2 - 4g + 5(c - \gamma)\sqrt{(c - \gamma)^2 - 8g}}{8},$$

$$l_3 = \frac{(c - \gamma)^2 - 4g + 3(c - \gamma)\sqrt{(c - \gamma)^2 - 8g}}{2},$$

$$m_1 = -\frac{3(c - \gamma) - \sqrt{(c - \gamma)^2 - 8g}}{2},$$

$$m_2 = \frac{3(c - \gamma)^2 - 4g - 5(c - \gamma)\sqrt{(c - \gamma)^2 - 8g}}{8},$$

$$m_3 = \frac{(c - \gamma)^2 - 4g - 3(c - \gamma)\sqrt{(c - \gamma)^2 - 8g}}{2},$$

$$a_1 = \frac{(c - \gamma)^2 - 8g + 3(c - \gamma)\sqrt{(c - \gamma)^2 - 8g}}{4},$$

$$a_2 = \frac{(c-\gamma)^2 - 8g - 3(c-\gamma)\sqrt{(c-\gamma)^2 - 8g}}{4},$$

$$b_1 = -(c-\gamma) - \sqrt{(c-\gamma)^2 - 8g},$$

$$b_2 = -(c-\gamma) + \sqrt{(c-\gamma)^2 - 8g},$$

$$\alpha_1 = -\frac{3(c-\gamma) + \sqrt{(c-\gamma)^2 - 8g}}{2\sqrt{(c-\gamma)^2 - 8g + 3(c-\gamma)\sqrt{(c-\gamma)^2 - 8g}}},$$

$$\alpha_2 = \frac{-3(c-\gamma) + \sqrt{(c-\gamma)^2 - 8g}}{2\sqrt{(c-\gamma)^2 - 8g - 3(c-\gamma)\sqrt{(c-\gamma)^2 - 8g}}},$$

$$\varphi_3(\xi) = l_+ e^{-|\xi|} + l_- e^{|\xi|} \quad for \ \sqrt{-g} \leq \varphi_3 \leq c - \gamma,$$

$$\varphi_4(\xi) = l_+ e^{|\xi|} + l_- e^{-|\xi|} \quad for \ c - \gamma \leq \varphi_4 \leq -\sqrt{-g},$$

$$l_\pm = \frac{c - \gamma \pm \sqrt{(c-\gamma)^2 + g}}{2},$$

$$T = \left| \ln\left(\sqrt{-g} + \sqrt{-2g}\right) - \ln(2l_-) \right|.$$

$$\varphi_1^\pm = \frac{3}{4}(c-\gamma) \pm \frac{1}{4}\sqrt{(c-\gamma)^2 - 8g} \mp \frac{1}{2}\sqrt{(c-\gamma)^2 \mp (c-\gamma)\sqrt{(c-\gamma)^2 - 8g}},$$

$$(3.6)$$

φ_0^+ and φ_0^- are as in (2.9).

Before proving this theorem, we take a set of data and employ Maple to display the graphs of smooth solion, peaked soliton and periodic cuspon solutions of (1.3), see Figures 3, 4, 5, 6 and 7.

Proof. Usually, a solion solution of (1.3) corresponds to a homoclinic orbit of system (2.5), and a periodic travelling wave solution of (1.3) corresponds to a periodic orbit of system (2.5). The graphs of homoclinic orbit, periodic orbit of system (2.5), and their limit curves are shown in Figure 8.

(1) When $0 < g < g_1(c), c > \gamma$, system (2.5) has a homoclinic orbit (see Figure 8(a)). This homoclinic orbit can be expressed as

$$y = \pm\frac{(\varphi - \varphi_0^-)\sqrt{\varphi^2 + l_1\varphi + l_2}}{c - \gamma - \varphi} \quad for \ \varphi_0^- < \varphi < \varphi_1^+. \tag{3.7}$$

Substituting (3.7) into the first equation of system (2.3) and integrating along this homoclinic orbit, we obtain (3.1).
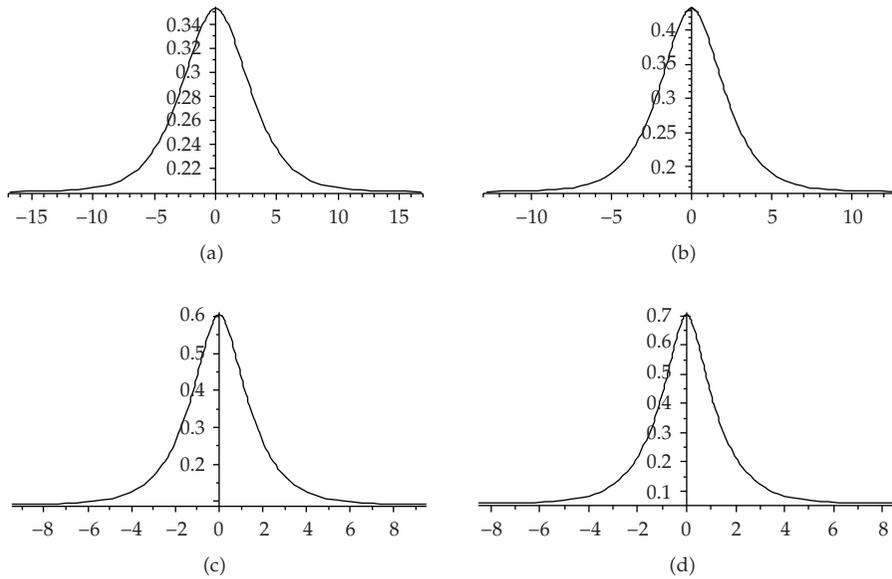
Figure 3: Smooth hump-like soliton solutions of (1.3) $(c = 2, \gamma = 1)$. (a) $g = 0.12$; (b) $g = 0.1$; (c) $g = 0.075$; (d) g=0.05.
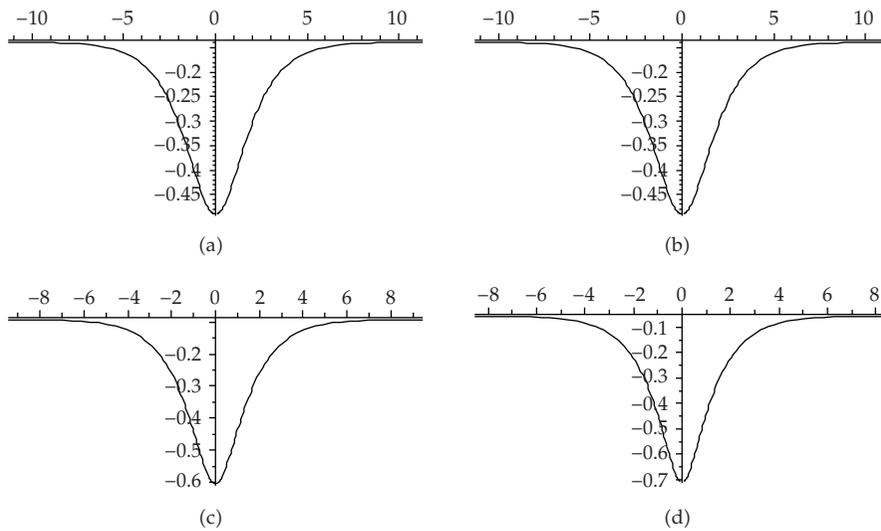
Figure 4: Smooth valley-like soliton solutions of (1.3) $(c = 1, \gamma = 2)$. (a) $g = 0.12$; (b) $g = 0.1$; (c) $g = 0.075$; (d) $g = 0.05$.

When $0 < g < g_1(c), c < \gamma$, we can obtain (3.2) in similar way.

(2) When $g = 0, c > \gamma$, system (2.5) has a homoclinic orbit that consists of the following three line segments (see Figure 8(c)):

$$y = \pm\varphi \quad \text{for } \varphi_0^- \leq \varphi \leq \varphi_1^+, \tag{3.8}$$

$$\varphi = c - \gamma \quad \text{for } -\sqrt{(c-\gamma)^2 + g} \leq y \leq \sqrt{(c-\gamma)^2 + g}. \tag{3.9}$$

(a)

(b)
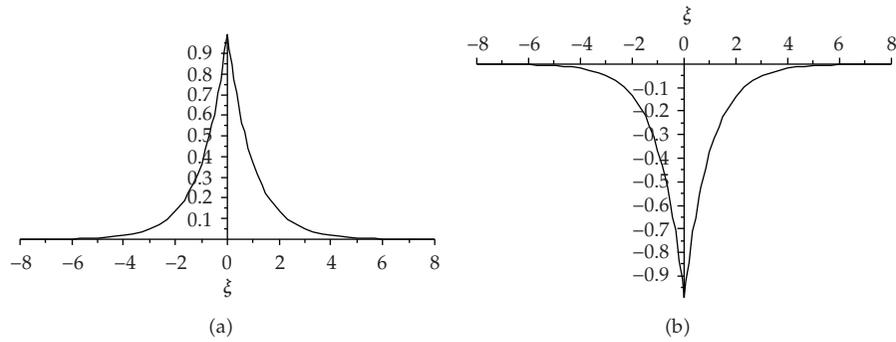
Figure 5: Peaked soliton solutions of (1.3). (a) $c = 2$, $\gamma = 1$; (b) $c = 1$, $\gamma = 2$.

(a)

(b)

(c)

(d)
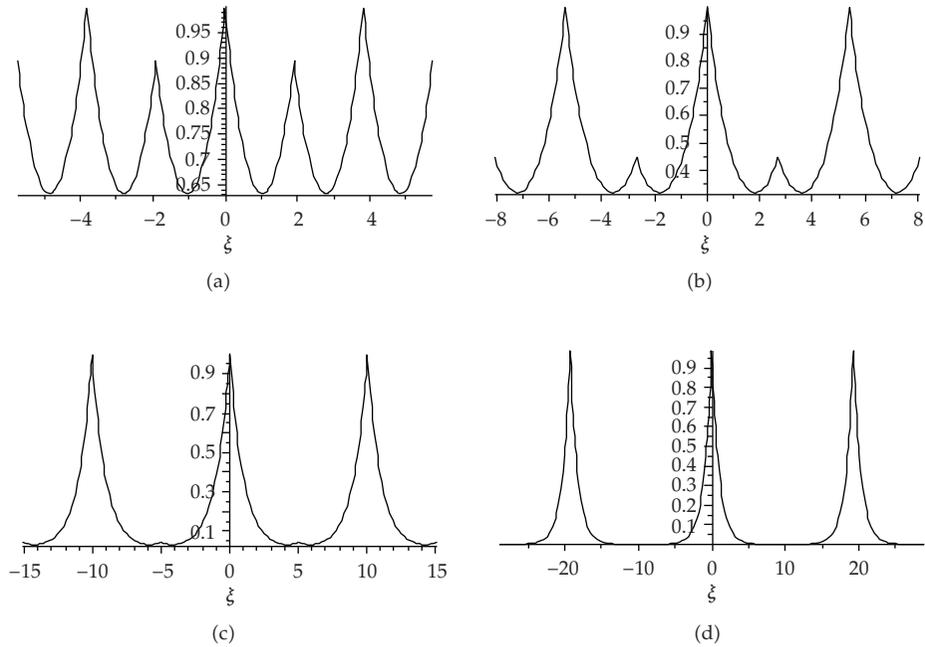
Figure 6: Period cuspon solutions of (1.3) ($c = 2$, $\gamma = 1$). (a) $g = -0.4$; (b) $g = -0.1$; (c) $g = -0.01$; (d) $g = -0.0000001$.

Substituting (3.8) into the first equation of system (2.3) and integrating along this orbit, we obtain (3.3).

When $g = 0, c < \gamma$, we can also obtain (3.3).

(3) When $g_2(c) < g < 0, c > \gamma$, system (2.5) has a periodic orbit (see Figure 8(e)). This periodic orbit can be expressed as

$$y = \pm\sqrt{\varphi^2 + g} \quad \text{for } \sqrt{-g} \leq \varphi \leq c - \gamma, \tag{3.10}$$

$$\varphi = c - \gamma \quad \text{for } -\sqrt{(c-\gamma)^2 + g} \leq y \leq \sqrt{(c-\gamma)^2 + g}. \tag{3.11}$$

(a)

(b)

(c)

(d)
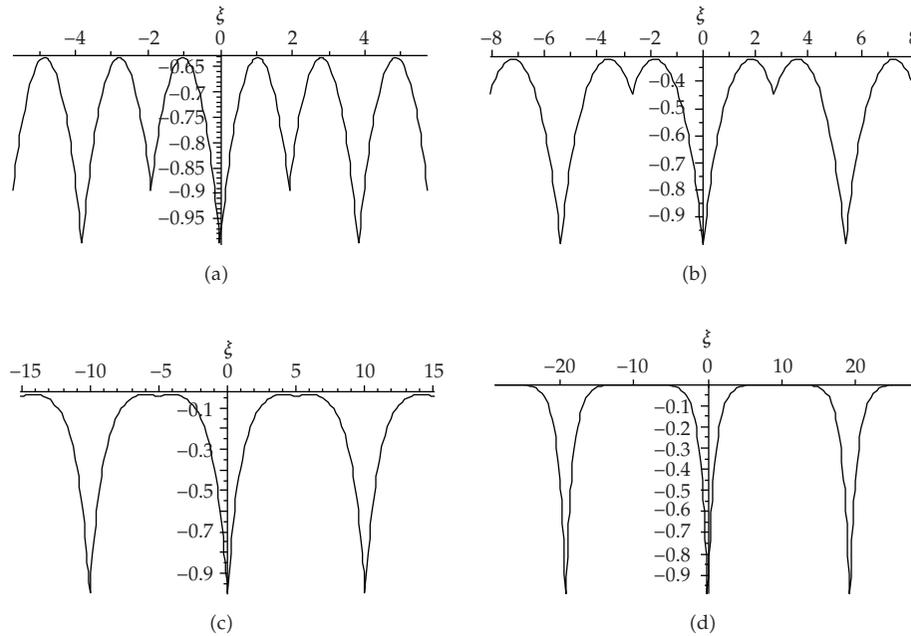
Figure 7: Period cuspon solutions of (1.3) ($c = 1$, $\gamma = 2$). (a) g =-0.4; (b) $g = -0.1$; (c) $g = -0.01$; (d) $g = -0.0000001$.

Substituting (3.10) into the first equation of system (2.3) and integrating along this periodic orbit, we obtain (3.4).

When $g_2(c) < g < 0, c < \gamma$, we can obtain (3.5). □

Remark 3.2. From the above discussion, we can see that when $g < 0, g \rightarrow 0$, the period of the periodic cusp wave solution becomes bigger and bigger, and the periodic cuspon solutions (3.4) and (3.5) tend to the peaked soliton solutions (3.3). When $g > 0, g \rightarrow 0$, the smooth hump-like soliton solutions (3.1) and the smooth valley-like soliton solutions (3.2) lose their smoothness and tend to the peaked soliton solutions (3.3).

4. Discussion

In this paper, we obtain the solitons, peakons, and periodic cuspons of a generalized Degasperis-Procesi equation (1.3). These solitons denote the nonlinear localized waves on the shallow water's free surface that retain their individuality under interaction and eventually travel with their original shapes and speeds. The balance between the nonlinear steepening and dispersion effect under (1.3) gives rise to these solitons.

The peakon travels with speed equal to its peak amplitude. This solution is nonanalytic, having a jump in derivative at its peak. Peakons are true solitons that interact via elastic collisions under (1.3). We claim that the existence of a singular straight line for the planar dynamical system (2.3) is the original reason why the travelling waves lose their smoothness.

Also, the periodic cuspon solution is nonanalytic, having a jump in derivative at its each cusp.
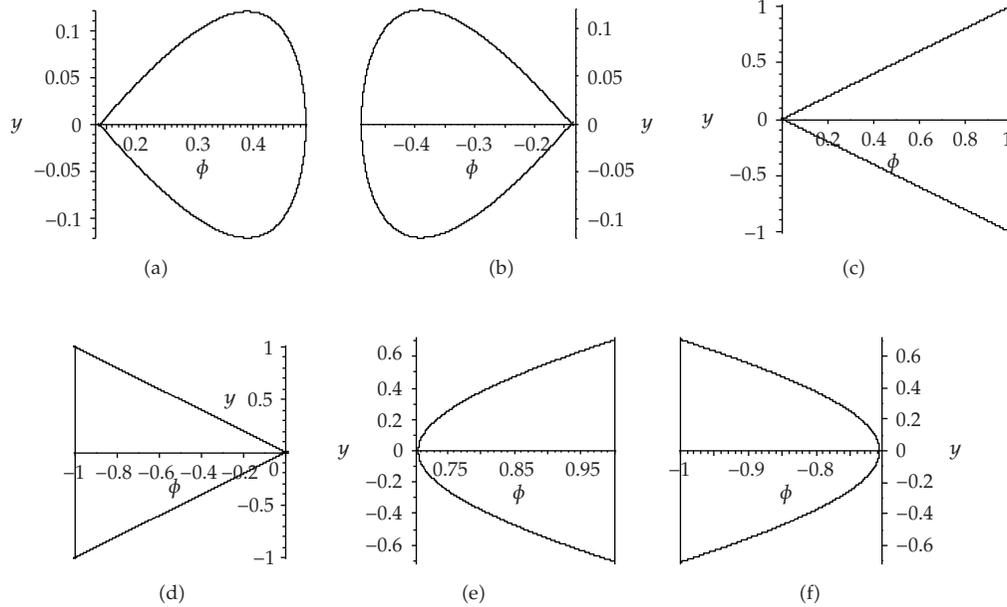
Figure 8: The orbits of system (2.5) connecting with the saddle points. (a) $0 < g < g_1(c), c > \gamma$; (b) $0 < g < g_1(c), c < \gamma$; (c) $g = 0, c > \gamma$; (d) $g = 0, c < \gamma$; (e) $g_2(c) < g < 0, c > \gamma$; (f) $g_2(c) < g < 0, c < \gamma$.

Acknowledgments

References

[1] A. Degasperis and M. Procesi, "Asymptotic integrability," in *Symmetry and Perturbation Theory (Rome, 1998)*, A. Degasperis and G. Gaeta, Eds., pp. 23–37, World Scientific, River Edge, NJ, USA, 1999.

[2] A. Degasperis, D. D. Kholm, and A. N. I. Khon, "A new integrable equation with peakon solutions," *Theoretical and Mathematical Physics*, vol. 133, no. 2, pp. 1463–1474, 2002.

[3] Z. Yin, "On the Cauchy problem for an integrable equation with peakon solutions," *Illinois Journal of Mathematics*, vol. 47, no. 3, pp. 649–666, 2003.

[4] Z. Yin, "Global weak solutions for a new periodic integrable equation with peakon solutions," *Journal of Functional Analysis*, vol. 212, no. 1, pp. 182–194, 2004.

[5] Z. Yin, "Global solutions to a new integrable equation with peakons," *Indiana University Mathematics Journal*, vol. 53, no. 4, pp. 1189–1209, 2004.

[6] Z. Yin, "Global existence for a new periodic integrable equation," *Journal of Mathematical Analysis and Applications*, vol. 283, no. 1, pp. 129–139, 2003.

[7] Y. Zhou, "Blow-up phenomenon for the integrable Degasperis-Procesi equation," *Physics Letters A*, vol. 328, no. 2-3, pp. 157–162, 2004.

[8] J. Escher, Y. Liu, and Z. Yin, "Global weak solutions and blow-up structure for the Degasperis-Procesi equation," *Journal of Functional Analysis*, vol. 241, no. 2, pp. 457–485, 2006.

[9] Z. Yin, "Well-posedness, blowup, and global existence for an integrable shallow water equation," *Discrete and Continuous Dynamical Systems. Series A*, vol. 11, no. 2-3, pp. 393–411, 2004.

[10] Y. Liu and Z. Yin, "Global existence and blow-up phenomena for the Degasperis-Procesi equation," *Communications in Mathematical Physics*, vol. 267, no. 3, pp. 801–820, 2006.

[11] V. O. Vakhnenko and E. J. Parkes, "Periodic and solitary-wave solutions of the Degasperis-Procesi equation," *Chaos, Solitons & Fractals*, vol. 20, no. 5, pp. 1059–1073, 2004.

[12] Y. Matsuno, "Multisoliton solutions of the Degasperis-Procesi equation and their peakon limit," *Inverse Problems*, vol. 21, no. 5, pp. 1553–1570, 2005.

[13] Y. Matsuno, "Cusp and loop soliton solutions of short-wave models for the Camassa-Holm and Degasperis-Procesi equations," *Physics Letters A*, vol. 359, no. 5, pp. 451–457, 2006.

[14] H. Lundmark and J. Szmigielski, "Multi-peakon solutions of the Degasperis-Procesi equation," *Inverse Problems*, vol. 19, no. 6, pp. 1241–1245, 2003.

[15] J. Lenells, "Traveling wave solutions of the Degasperis-Procesi equation," *Journal of Mathematical Analysis and Applications*, vol. 306, no. 1, pp. 72–82, 2005.

[16] H. Lundmark, "Formation and dynamics of shock waves in the Degasperis-Procesi equation," *Journal of Nonlinear Science*, vol. 17, no. 3, pp. 169–198, 2007.

[17] J. Escher, Y. Liu, and Z. Y. Yin, "Shock waves and blow-up phenomena for the periodic Degasperis-Procesi equation," *Indiana University Mathematics Journal*, vol. 56, no. 1, pp. 87–117, 2007.

[18] L. Yu, L. Tian, and X. Wang, "The bifurcation and peakon for Degasperis-Procesi equation," *Chaos, Solitons & Fractals*, vol. 30, no. 4, pp. 956–966, 2006.

[19] D. Luo, X. Wang, D. Zhu, and M. Han, *Bifurcation Theory and Methods of Dynamical Systems*, vol. 15 of *Advanced Series in Dynamical Systems*, World Scientific, River Edge, NJ, USA, 1997.

Research Article

Chaotic Patterns in Aeroelastic Signals

F. D. Marques and R. M. G. Vasconcellos

Laboratory of Aeroelasticity, School of Engineering of São Carlos, University of São Paulo,
Avenue Trabalhador Sancarlense, 400, 13566-590 São Carlos, SP, Brazil

Correspondence should be addressed to F. D. Marques, fmarques@sc.usp.br

This work presents the analysis of nonlinear aeroelastic time series from wing vibrations due to airflow separation during wind tunnel experiments. Surrogate data method is used to justify the application of nonlinear time series analysis to the aeroelastic system, after rejecting the chance for nonstationarity. The singular value decomposition (SVD) approach is used to reconstruct the state space, reducing noise from the aeroelastic time series. Direct analysis of reconstructed trajectories in the state space and the determination of Poincaré sections have been employed to investigate complex dynamics and chaotic patterns. With the reconstructed state spaces, qualitative analyses may be done, and the attractors evolutions with parametric variation are presented. Overall results reveal complex system dynamics associated with highly separated flow effects together with nonlinear coupling between aeroelastic modes. Bifurcations to the nonlinear aeroelastic system are observed for two investigations, that is, considering oscillations-induced aeroelastic evolutions with varying freestream speed, and aeroelastic evolutions at constant freestream speed and varying oscillations. Finally, Lyapunov exponent calculation is proceeded in order to infer on chaotic behavior. Poincaré mappings also suggest bifurcations and chaos, reinforced by the attainment of maximum positive Lyapunov exponents.

1. Introduction

The assessment of aeroelastic phenomena with linear models has provided a reasonable amount of tools for the analysis of most of adverse instability behavior [1, 2]. Nonetheless, aeronautical engineering has shown advances that lead to faster and lighter aircraft, thereby increasing the risk of moderate or severe nonlinear aeroelastic problems.

Nonlinear behavior is inherent to aeroelastic systems and can be associated with aerodynamic sources (compressibility, separated flows, aerodynamic heating, and turbulence effects) and structural sources (effects of aging, loose attachments, material features, and large deformations) [3–5]. Aeroelastic systems can face those effects, for instance, in transonic flight, high angle of attack manoeuvres, and in all cases leading to complex models beyond linearity suppositions. Nonlinear systems typically present features like,

multiple equilibrium points, bifurcations, limit-cycle oscillations, and chaos [6, 7]. The presence of such effects results in modifications to the aeroelastic dynamics, leading to more laborious prediction of instabilities. For instance, the flutter phenomenon in the presence of nonlinearities happens in a different way to that foreseen in linear models.

Recently, nonlinear aeroelasticity research has been performed for a greater number of groups using advanced CFD methods, reduced-order models, and other methodologies [8–10]. However, these methodologies present deficiencies such as losses in the analysis of the physical phenomenon and little flexibility to evaluate different flight regimes using the same model. To validate and verify the mathematical models, experimental analysis makes it possible to observe the system dynamics without neglecting important effects. Experimental data furnishes sequences of measurements that correspond to time series with embedded system dynamics. Time series analysis techniques, such as state space reconstruction and Lyapunov exponents, can be used in these time series to access important information in the system dynamics. Therefore, it seems reasonable that by examining raw experimental data with techniques from time series analysis, one can better assess the effects of aerodynamic and/or structural nonlinearities on aeroelastic systems. Moreover, such insight may provide important tools to support and improve mathematical modeling for aeroelastic analysis. Nonlinear time series analysis techniques can also have an important impact in flutter flight tests, helping in the extraction of instability parameters which are typically surrounded by uncertainties.

While many processes in nature seem a priori very unlikely to be linear, their possible nonlinear nature might not be evident in specific aspects of their dynamics. Moreover, there is always the danger that one is dealing with nonstationary time series, particularly in the case of experiments. Testing the properties of a time series is the most prudent action before starting to draw any conclusion on a system behavior. A variety of techniques to check time series stationarity are available [11], and the method of surrogate data can also be used to justify the application of nonlinear time series analysis techniques excluding the linear hypothesis [12, 13].

Typical dynamic system responses can be assessed by means of reconstructing the state space from time series using the so-called method of *singular value decomposition* (SVD) [14]. The SVD method uses the properties of the covariance matrix to produce uncorrelated coordinates; as a result of the process the data is filtered, diminishing complications caused by the noise present in experimental data.

With reconstructed state spaces and Poincaré sections, it is possible to identify structures associated with limit-cycle oscillations (LCOs) and chaotic patterns [15, 16]. Chaotic behavior may be characterized by the divergence between neighbor trajectories in state space [17]. The assessment of Lyapunov exponents can be used to quantify this divergence [15].

The purpose of this work is to apply techniques from time series analysis for the investigation of nonlinear aeroelastic response behavior present in experimental data obtained from a wind tunnel tests. The experimental apparatus comprises a flexible wing model and by exposing it to the wind tunnel airflow, motion-induced aeroelastic responses occur. By inducing motions at higher angles of attack, flow separation introduces severe unsteady aerodynamic nonlinearity into the system. Incidence oscillatory variations are achieved using a turntable that supports the wing model, and structural deformations are captured by strain gages, thereby providing information on the aeroelastic responses. In this way, investigations may be made for various wind tunnel freestream speed and turntable oscillation frequencies.

The resulting aeroelastic signals are firstly checked for stationarity and nonlinearity properties. Runtest and reverse arrangements followed by surrogate data tests are used to the aforementioned time series properties checking.

The SVD method is used to reconstruct the state space from aeroelastic time series. Evolutions of reconstructed state space and respective Poincaré mapping with parameter variation are presented and discussed. Finally, the method for estimating the largest Lyapunov exponents is applied to identify and reinforce the suspect for the presence of deterministic chaotic behavior in the aeroelastic response time series.

2. State Space Reconstruction

State space reconstruction approaches use time histories or time series ($\mathbf{s}(t)$) to extract the dynamics of a system. Reconstruction techniques are based on Taken's embedded theorem [18], which establishes that a time series $\mathbf{s}(t)$ has information on nonobservable states. With $\mathbf{s}(t)$ it is possible to reconstruct the state space of the system comparable to the real case preserving the invariants of the system, for example, attractor dimension and Lyapunov exponents. This statement has been proven numerically by Packard et al. [19] and Takens [18].

There are different methods to reconstruct the state space, like the *method of delays* (MODs) and the *singular value decomposition method* (SVD). The MOD is the most explored reconstruction method in literature. In this technique a time delay (τ) and an embedding dimension (d) are required to generate delayed coordinates from time series [20].

The reconstruction method based on singular value decomposition (SVD) has been proposed by Broomhead and King [14]. The methodology eliminates the need for a time delay parameter by using the properties of the covariance matrix of the data to generate uncorrelated coordinates. One of the advantages of SVD is the capacity of filtering the time series as a result of the reconstruction process. Kugiumtzis and Christophersen [21] and Vasconcellos [16] have compared MOD and SVD, and they concluded that SVD is more reliable for noisy data, since MOD may lead to arbitrary conclusions because the approaches to obtain τ and d are sensitive to noise [16, 22]. Therefore in principle, the SVD approach is more suitable for experimentally acquired time series.

The SVD approach for state space reconstruction needs the covariance matrix constructed from data contained in the time series $\mathbf{s}(t)$. In this case, each state of the system can be considered a statistical variable, and the diagonalization of covariance matrix separates the states by their variance, allowing the assessment of the system dynamics from those states that have higher variance. As small variance states are dominated by noise, reconstruction is basically attained with filtering. The application of an n-size window to a time series of N_T data points results in a sequence of $N = N_T - (n-1)$ vectors in the embedding space, that is, $\{\mathbf{x}_i \in \mathfrak{R}^n \mid i = 1, 2, \ldots, N\}$. Such a sequence can be used to construct a so-called trajectory matrix ($\mathbf{X}$), which contains the complete record of patterns that have occurred within the window, that is:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^{\mathrm{T}} & \mathbf{x}_2^{\mathrm{T}} & \cdots & \mathbf{x}_N^{\mathrm{T}} \end{bmatrix}. \tag{2.1}$$

The columns of trajectory matrix constitute the state vectors $\mathbf{x}_i$ on the reconstructed trajectory in embedding space. The N state vectors in embedding space are used, in order to find a set of linearly independent vectors in $\mathfrak{R}^n$, which describe efficiently the attracting

manifold in state space [23]. The $\mathbf{X}$ matrix can be decomposed according to the following relation:

$$\mathbf{X} = \mathbf{SDC}^{\mathrm{T}}, \tag{2.2}$$

where $\mathbf{S} = [\mathbf{s}_1 \ \ \mathbf{s}_2 \ \ \cdots \ \ \mathbf{s}_n]$ and $\mathbf{C} = [\mathbf{c}_1 \ \ \mathbf{c}_2 \ \ \cdots \ \ \mathbf{c}_n]$ are matrices of the respective singular vectors, and $\mathbf{D} = \mathrm{diag}[\sigma_1, \sigma_2, \ldots, \sigma_n]$ is a diagonal matrix of the singular values [14].

The number of independent eigenvectors $\mathbf{c}_i$, which are relevant for the description of the system dynamics, is equal to the number of nonzero eigenvalues σ_i, and they also are the new basis for the trajectories projection. The trajectory can be described in the new basis by projecting the trajectory matrix on the basis by the $\mathbf{XC}$ product, where $\mathbf{C} = \{\mathbf{c}_i, i = 1, \ldots, \mathrm{rank}(\mathbf{XX}^{\mathrm{T}})\}$. The new trajectory matrix $\mathbf{XC}$ is described by the relation:

$$(\mathbf{XC})^{\mathrm{T}}(\mathbf{XC}) = \mathbf{D}^2. \tag{2.3}$$

The relationship given by (2.3) corresponds to the diagonalization of the new covariance matrix, so that in the basis $\mathbf{c}_i$ the components of trajectory are uncorrelated.

When the system is perturbed by external noise the trajectory begins to be diffuse in directions corresponding to zero eigenvalues, where the external perturbation dominates. The projection of trajectory matrix in the basis $\mathbf{C}$ works as a lowpass filter for the entire trajectory. Moreover, the SVD method permits the reconstruction of the original trajectory excluding all dimensions dominated by noise and the retrieval of a filtered time series. The presence of a nonzero constant background, or noise floor, in the spectrum σ_i is sufficient to distinguish the deterministic components [14]. The original trajectory may be reconstructed by

$$\mathbf{X}_d = \sum_{\sigma > \mathrm{noise}} \{\mathbf{Xc}_i\} \mathbf{c}_i^{\mathrm{T}}, \tag{2.4}$$

where σ_i corresponds to a singular value above the noise floor.

3. Tools to Characterize Chaotic Patterns

This section presents some techniques that can be employed for the characterization of complex nonlinear dynamics. There is suggested, as first steps toward chaotic patterns assessment, tests for stationarity and the surrogate data test to the experimental time series. If a time series originates from a unknown process, it is important to investigate if the system parameters remain constant or not during the experiment and whether the data does or not some nonlinear deterministic dependencies [24].

Violations of the fact that the dynamical properties of the system underlying a signal must not change during the observation period can be checked simply by measuring them for several segments of the data set [13]. To investigate if the signal contains some nonlinear deterministic dependencies, a surrogate data test can be useful. The basic idea with respect to the surrogate data technique is to make some hypotheses about the data and then try to contradict this hypothesis. A widely used hypothesis is that colored noise data is generated by a linear stochastic process. Therefore, the data is modified in such a way that the

complete structure, except for the assumed properties, will be destroyed. This may be done by Fourier transforming the data, and by randomly shuffling the phases, the power spectrum or equivalently the autocorrelation function is not affected. A new time series with the same power spectrum is obtained by transforming back into the time domain.

If the original data is just colored noise, estimators of dimension, average mutual information, Lyapunov exponents, prediction errors, and so forth should give the same results for the original time series and the surrogates. If however, the analysis yields significant differences, the original data is more than "just noise" [24].

To improve the statistical robustness, several surrogates are generated. Furthermore, it is necessary to take into account a possibly static nonlinear transformation of the data that would distort the Gaussian distribution of the assumed colored noise as done by Theiler et al. [25].

A powerful tool for the verification of complex dynamics, in particular, to identify chaotic patterns is Poincaré mapping. The Poincaré section of the state space dynamics simplifies the geometric description of the dynamics by removing one of the state space dimensions. For instance, a three-dimensional state space presents the Poincaré section as a two-dimensional plane chosen in such way that the trajectories intersect it transversely.

The key point is that this simplified geometry contains the essential information about the periodicity, quasiperiodicity, chaosity, and bifurcations of the system dynamics [17]. Bifurcation, in this case, is the term used to describe any sudden change in the dynamics of the system due to the respective parametric change. Therefore, for any change on the attractor geometry with a parameter variation, bifurcations can be visualized by plotting one Poincaré section for each parameter value. The Poincaré section computation has been based on Merkwirth et al. [26] and Kantz and Shreiber [13], which proposes the section extracted directly from an embedded time series. The result is a set of $(n-1)$-dimensional vector points, used to perform an orthogonal projection.

Lyapunov exponents determination furnishes important indications with respect to chaotic patterns of dynamic systems. Lyapunov exponents describe the mean exponential increase or decrease of small perturbations on an attractor and are invariant with respect to diffeomorphic changes of the coordinate system [24]. When the largest Lyapunov exponent is positive, the system is said to be chaotic. Direct methods to quantify the largest Lyapunov exponent estimate the divergent motion from the reconstructed space state, without fitting a model to the data.

The method proposed by Sato et al. [27] considers the average exponential growth of the distance of neighboring orbits on a logarithmic scale via prediction error on the number of time steps k, that is:

$$p(k) = \frac{1}{Nt_s} \sum_{n=1}^{N} \log_2 \left(\frac{\|\mathbf{y}^{n+k} - \mathbf{y}^{nn+k}\|}{\|\mathbf{y}^n - \mathbf{y}^{nn}\|} \right), \tag{3.1}$$

where N is the number of data points, t_s is the sampling period, and $\mathbf{y}^{nn}$ is the nearest neighbor of $\mathbf{y}^n$.

The dependence of the prediction error $p(k)$ on the number of time steps k may be divided into three phases: the transient, corresponding to the first phase, where the neighboring orbit converges to the direction corresponding to the largest Lyapunov exponent; the second phase, where the distance grows exponentially with $e^{(\lambda_1 t_s k)}$ until it exceeds the range of validity of the linear approximation of the flow around the reference orbit $\mathbf{y}^{n+k}$;
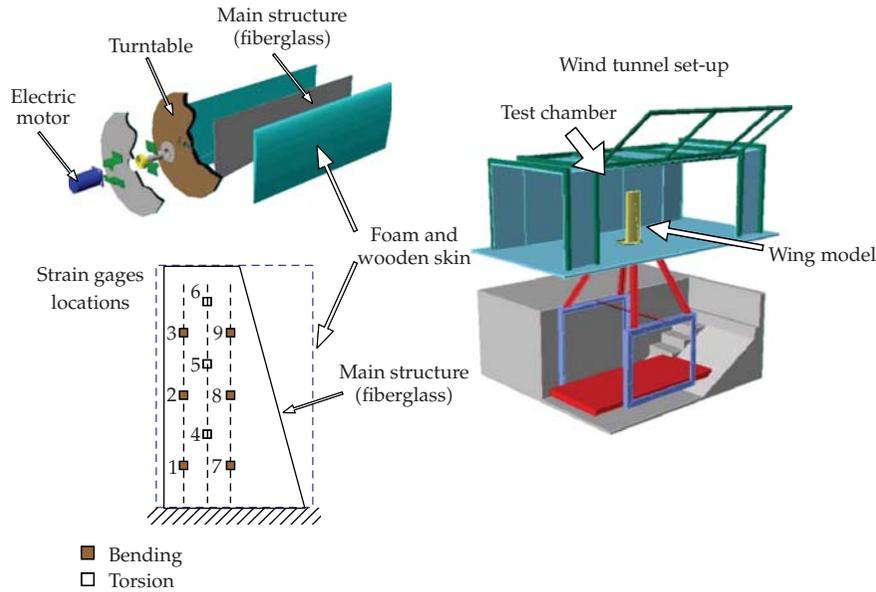
Figure 1: Experimental set-up and strain gages locations.

then, the last phase begins, where the distance increases slower than exponentially until it decreases again due to foldings in the state space [24]. If the second phase is sufficient long, a linear segment with slope λ appears in $p(k)$ versus k plot. This slope value (λ) is associated with the Lyapunov exponent value. This also provides a direct verification of the exponential growth of distances to distinguish deterministic chaos from stochastic processes, where a *non*exponential separation of trajectories occurs [24].

4. Experimental Apparatus and Database

The experimental apparatus comprises an aeroelastic wing model mounted over a turntable device driven by a brushless electrical motor and an acquisition system. The wing model is tested in a wind tunnel test section of approximately $2\,m^2$ cross-section area and a maximum flow speed of $50\,m/s$. The wing model was fixed to a turntable that allowed various angles of incidence of the model, thereby providing exploration of a variety of motion-induced aeroelastic responses over a range of airflow velocities.

The wing model main structure was constructed with fiberglass and epoxy resin with a taper ratio of 1 : 1.67, where the width at the root is $250\,mm$ and the semispan is $800\,mm$. To provide aerodynamic shape of NACA0012 airfoil, high-density foam and a thick wooden skin were used, and the chord was fixed in $290\,mm$ from root to the tip. In order to reduce the effect of aerodynamic cover to the wing structure stiffness, both foam and wooden shell have been segmented at each $100\,mm$ spanwise. Strain gages were fixed to the plate surface to register the dynamic response of the wing main structure. The strain gages were distributed along three spanwise lines. The first and the last lines received three strain gages each, to register bending motions. The intermediate line also received three strain gages, arranged in this case to register torsional motion. Figure 1 illustrates the experimental apparatus with indications of the strain gage locations on the wing model structure.

Table 1: Experimental test cases.

| Fixed oscillatory frequency (at 10.0 rad/s) | Fixed freestream speed (at 15.0 m/s) |
Freestream speeds (m/s)	Turntable oscillatory frequency (rad/s)
8.28	2.0
9.97	4.0
11.64	6.0
13.30	8.0
14.97	10.0

Data acquisition and motion control of the brushless electrical motor were achieved using a dSPACE DS1103 PPC controller board and real-time interface for SIMULINK. An HBM KWS 3073 amplifier was used to acquire and amplify the strain gages signal. The resulting signals are acquired by the dSPACE controller board, for subsequent storage into a PC compatible computer.

5. Results and Discussion

During experiments, oscillatory motions of the turntable were executed at relatively low amplitude, that is, 5.5° incidence angle, but such oscillations have been considered around an average incidence angle of 9.5°. For these cases, highly unsteady separated flow occurs, inducing complex aeroelastic responses to the wing model. These cases furnish an adequate database for nonlinear aeroelastic response phenomena analysis.

The cases under consideration are summarized in Table 1 and were collected from strain gage at position 1 for bending measurement (cf. Figure 1). Table 1 indicates that aeroelastic time series has been acquired for a range of freestream speed (U), at a fixed turntable oscillatory frequency (ω), and for a range of turntable oscillatory frequencies at a fixed airflow velocity. Both cases provide essential information on motion-induced aeroelastic responses. Each aeroelastic time series has been filtered by SVD method and checked for stationarity with runtest and reverse arrangements test [11], prior to any analysis. All time series considered in this work have passed at the significance level of 0.05. Figure 2 presents a typical aeroelastic response, in this case for $U = 14.97$ m/s and $\omega = 10.0$ rad/s, where the existence of complex aeroelastic response can be observed.

In order to justify the use of nonlinear analysis techniques, a surrogate data test was performed. Using Algorithm II of Theiler et al. [25], 99 surrogates were generated and the correlation sum was computed for each by using the algorithm proposed by Grassberger and Procaccia [28]. The correlation sum assesses the relative number of neighboring points closer than r [24] and is given by

$$C_d(r) = \frac{2}{(N-c)(N-1-c)} \sum_{i=1}^{N} \sum_{j=1}^{i-c} H\left(r - \left\| \mathbf{y}^i - \mathbf{y}^j \right\|\right), \qquad (5.1)$$

where N is the number of data points, H is the Heaviside function with $H(x) = 1$ for $x > 1$ and zero elsewhere, c is a constant accounting for some correlation length (used to omit points that are close neighbors in time), $\left\| \mathbf{y}^i - \mathbf{y}^j \right\|$ is the mutual distance between the points in question, and d is the embedding dimension.
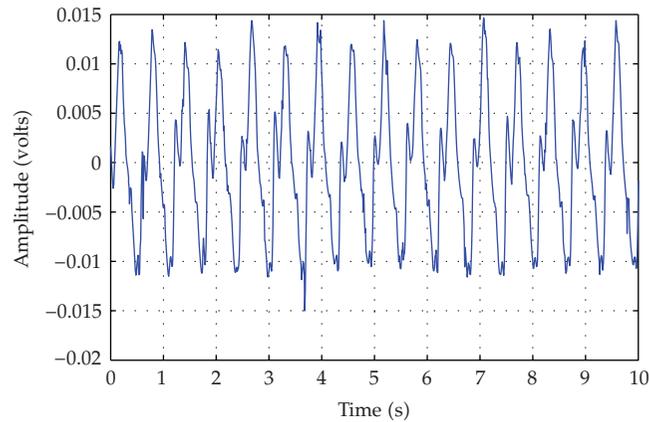
Figure 2: Aeroelastic time series—strain gage measurement at position 1 ($U = 14.97\,\text{m/s}$, $\omega = 10.0\,\text{rad/s}$).

Figures 3 and 4 present the correlation sum of these 99 surrogates in the mean line with the error bars; the correlation sum computed for the SVD-filtered acquired data is present in the continuous line. Clearly, the data is out of correlation sum distribution generated for purely linear stochastic surrogate signal. This evidence reinforces that the signal may be representative of a deterministic nonlinear process. These results provide enough information to qualify further investigation using nonlinear analysis tools to characterize possible aeroelastic chaotic patterns.

Here, state spaces have been reconstructed by the SVD approach. As an example, Figure 5 shows the singular spectrum and the accumulated variance of the considered singular values for one case ($U = 14.97\,\text{m/s}$, $\omega = 10.0\,\text{rad/s}$, and strain gage at position 1—cf. Figure 1), clearly revealing three eigenvalues above the noise floor. This indicates embedding dimension 3, which was confirmed as the same for all the other cases. The state space was reconstructed considering only these tree singular values, which represents more than 99% of the total variance. Figure 6 presents an example of reconstructed state space in terms of the projections onto the three mutually orthogonal planes spanned by the singular vectors ($\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$) and the three-dimensional view. The reconstructed trajectories present complex shapes, with the presence of more than one center of rotation, which is an indication of a chaotic pattern [29].

For the cases presented in Table 1 where a fixed oscillation frequency is considered (in this case, $\omega = 10.0\,\text{rad/s}$) for a range of airflow velocities, respective aeroelastic responses have been used to reconstruct the spaces via the SVD technique. The results furnish an evolution of the reconstructions with respect to freestream speed, allowing investigations of bifurcations and chaotic patterns. Figure 7 shows the evolution of trajectories in reconstructed state space due to freestream speed variation. The occurrence of bifurcation is clear, mainly due to the transition between trajectories at 11.64 and 13.30 m/s as well as between 13.30 and 14.97 m/s. In previous works, Vasconcellos [16] and Marques et al. [30] encountered evidence that bifurcations occur in a similar system. For increasing freestream speed, separated flow intensity also increases and different nonlinear mechanisms for this effect occur.

Poincaré sections of reconstructed state spaces have been determined, in order to supply an easier visualization of the aforementioned transitions or bifurcations. In Figure 8 one may observe considerable changes in the Poincaré sections, as the speed increases, with
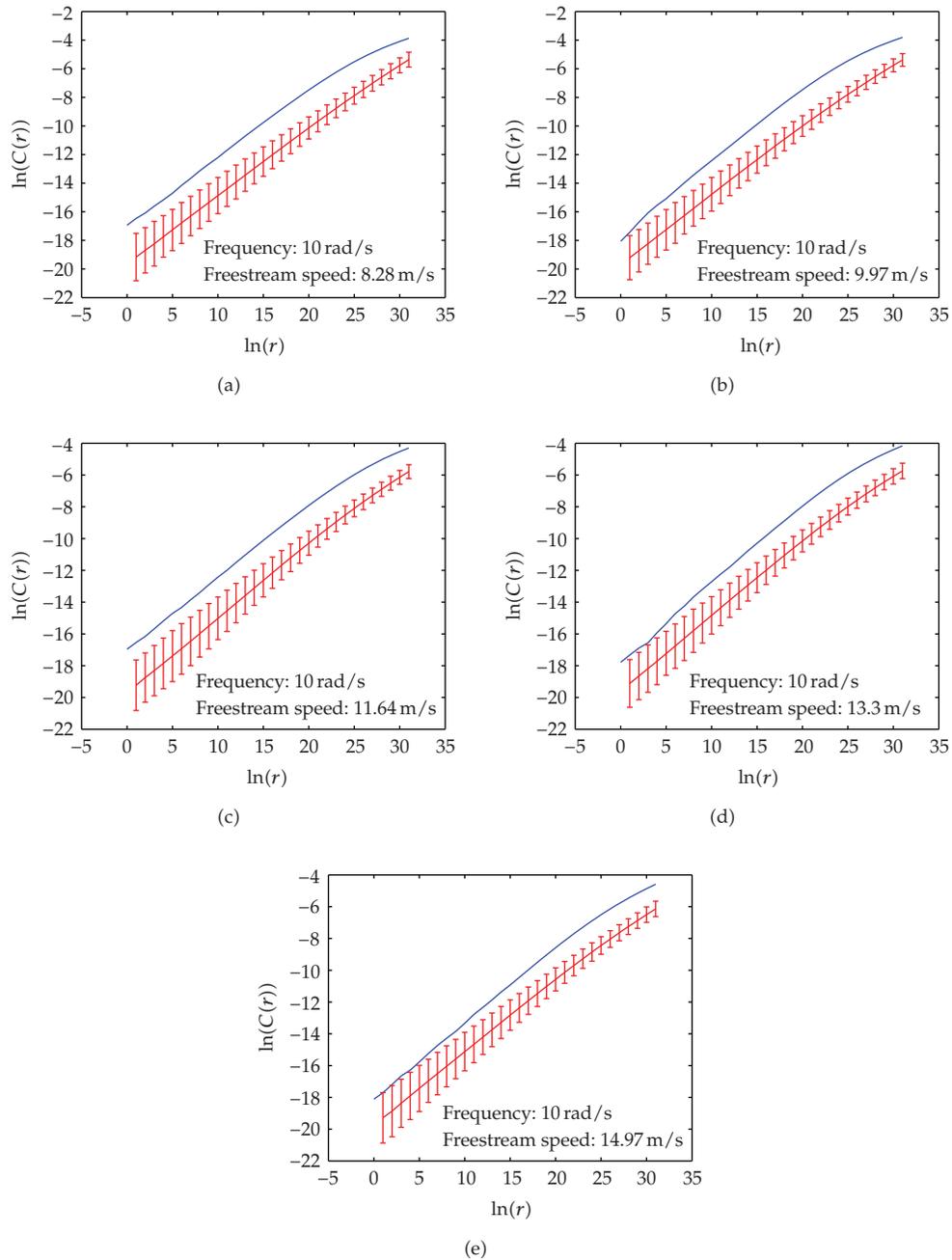
Figure 3: Correlation sum of 99 surrogates (line with error bars) and the tested data for the fixed oscillation frequency conditions.

the amplitude of the motions enlarging. Between third and fourth sections (corresponding to freestream speeds of 11.64 and 13.30 m/s), considerable change in the Poincaré section shape can be observed. The same complex behavior occurs between the fourth and fifth sections (corresponding to freestream speeds of 13.30 and 14.97 m/s). In all these cases, one can infer

(a)

(b)

(c)

(d)

(e)
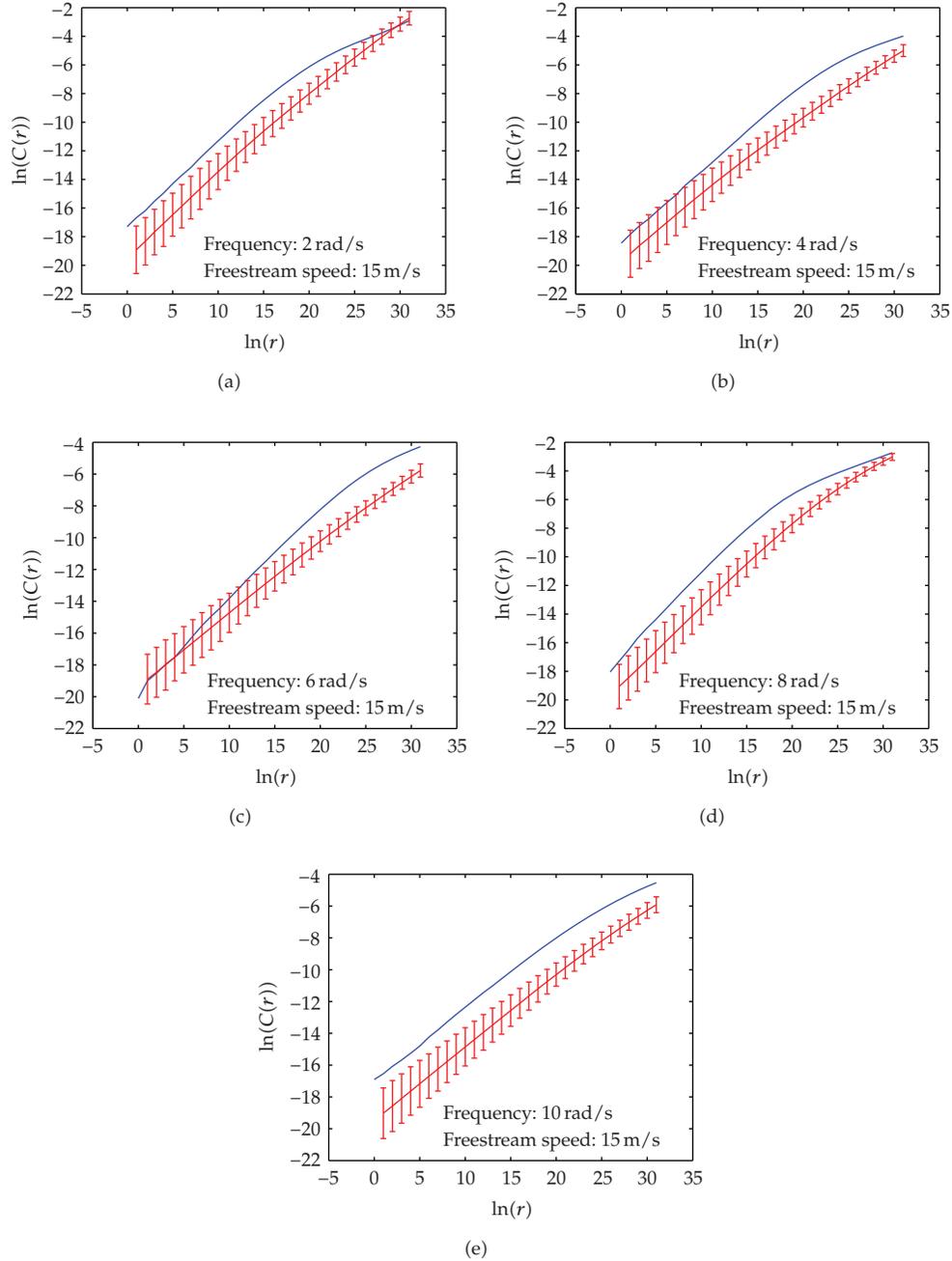
Figure 4: Correlation sum of 99 surrogates (line with error bars) and the tested data for the fixed freestream speed conditions.

that the aeroelastic system response is complex, revealing bifurcations associated separated flowfield effects as well as with distributed structural nonlinearities. Projected Poincaré sections, as illustrated in Figure 9, show an alternative way to visualize the bifurcations with respect to airflow velocity evolution.
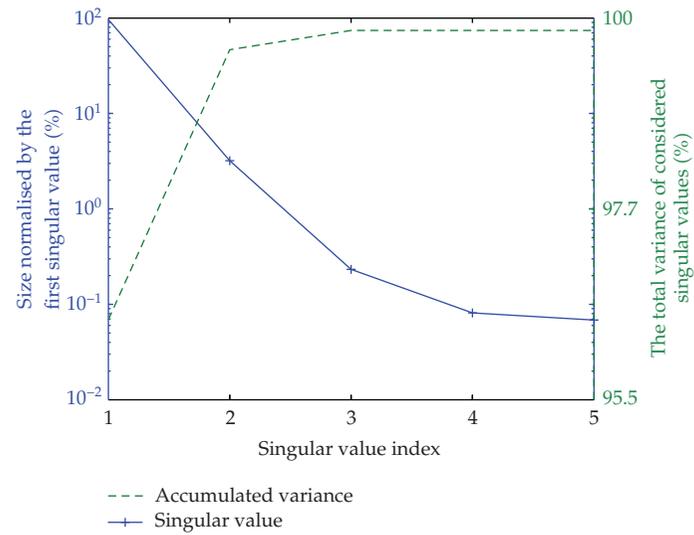
Figure 5: Singular spectrum and accumulated variance for aeroelastic time series (case: U = 14.97 m/s, ω = 10.0 rad/s, and strain gage at position 1—cf. Figure 1).
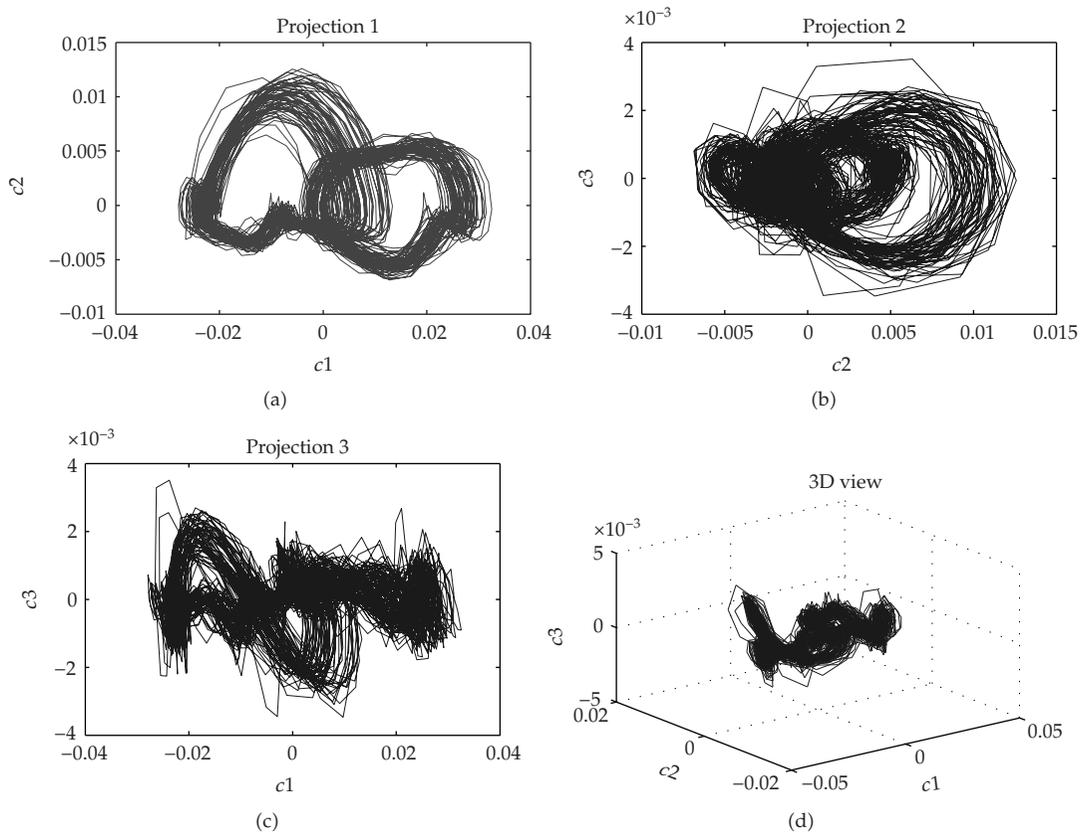
Figure 6: Reconstructed state space for aeroelastic response (case: U = 14.97 rad/s, ω = 10.0 rad/s, and strain gage at position 1—cf. Figure 1), including projections in three orthogonal planes and 3D view.
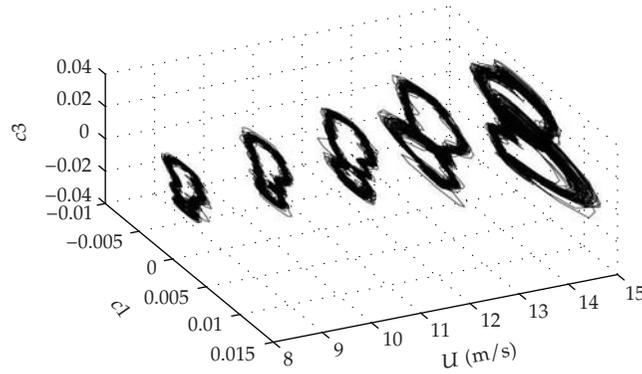
Figure 7: The evolution of state space reconstructions for the case of freestream speed variation: 8.28, 9.97, 11.64, 13.30, and 14.97 m/s, respectively, at fixed oscillatory turntable frequency of $\omega = 10.0$ rad/s.
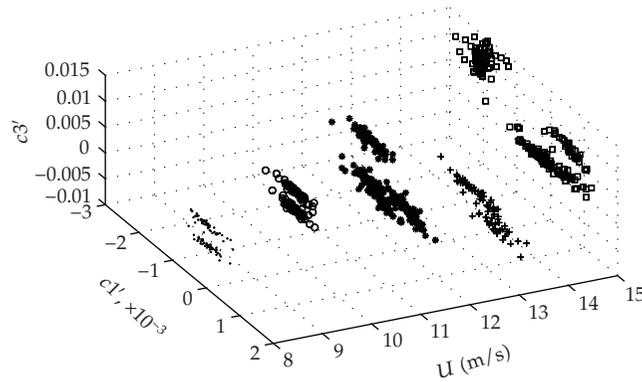
Figure 8: The evolution of Poincaré sections for the case of freestream speed variation: 8.28, 9.97, 11.64, 13.30, and 14.97 m/s, respectively, at fixed oscillatory turntable frequency of $\omega = 10.0$ rad/s.
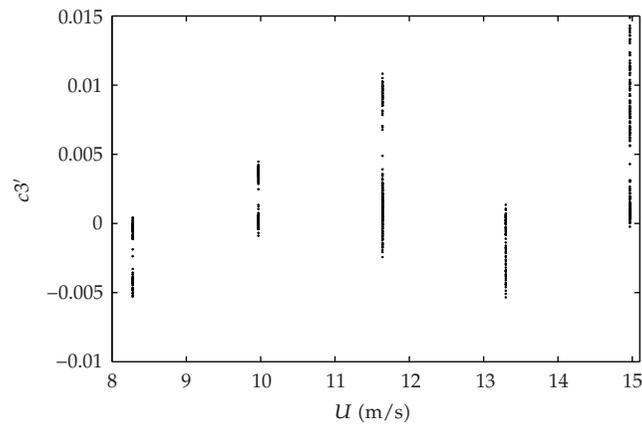
Figure 9: Projection of the Poincaré sections evolution with the freestream speeds: 8.28, 9.97, 11.64, 13.30, and 14.97 m/s, respectively, at fixed oscillatory turntable frequency of $\omega = 10.0$ rad/s.

Table 2: Lyapunov exponents by prediction error technique [27] for fixed turntable oscillatory frequency ($\omega = 10.0\,\text{rad/s}$) and a range of freestream speeds (cf. Table 1).

Freestream speed (m/s)	8.28	9.97	11.64	13.30	14.97
Largest exponent	0.50	0.54	0.56	0.56	0.57

The final step in the investigation of the complex nonlinear behavior of the aeroelastic signals is the determination of the largest Lyapunov exponent. Here, the exponent for each of the aeroelastic responses, for fixed turntable oscillatory frequency in a range of airflow velocities (cf. Table 1), is summarized in Table 2. The calculations were executed using the prediction error technique as proposed by Sato et al. [27]. It may be observed that the largest Lyapunov exponent increases with freestream speed. In all conditions, the largest Lyapunov exponents are positive, indicating chaotic behavior, what implies that the encountered bifurcations from inspecting state space reconstructions and Poincaré mappings are chaos-chaos bifurcations. The occurrence of chaotic motions may cause degradation of aircraft flight performance, leading to future structural problems due to material fatigue. Moreover, abrupt dynamical behavior changes due to bifurcations may yield severe structural damage or total failure.

Figure 10 can be seen as a complementary result, because it shows plottings of solutions for (3.1). The presence of the linear segment slope ensures deterministic chaos occurence, thereby validating surrogate data tests. In Marques et al. [20] and Simoni [31], the method developed by Wolf et al. [32] has been used to estimate the Lyapunov exponents, considering the analysis for similar motion-induced aeroelastic time series. Positive Lyapunov exponents have also been encountered, with values of approximately 0.3. Comparative results between techniques for the largest Lyapunov exponents for nonlinear aeroelastic responses can be found in Marques et al. [33].

The following results are related to investigations of chaotic patterns of aeroelastic responses for a range of turntable oscillatory frequencies, while the wind tunnel freestream speed is kept fixed (in this case, $U = 15.0\,\text{m/s}$). The respective oscillatory frequency range can be seen in Table 1. The state spaces are reconstructed for all these conditions, and Figure 11 shows the evolution of trajectories in state space within the range of turntable oscillatory frequencies. Here, a considerable change in trajectory patterns may be observed, with a clear increase in the amplitude of motion until $8.0\,\text{rad/s}$, followed by a sudden change in shape and amplitude at around $10.0\,\text{rad/s}$. Such behavior may be associated with the so-called bifurcation crises, in which chaotic attractors and their basin of attraction suddenly disappear or expand; the sudden expansion or contraction of a chaotic attractor is called an *interior crises* [17].

The physical events related to these results indicate that separated flow effects and aeroelastic modes interaction play an important role in the nonlinear behavior. Bifurcation crises phenomenon manifests itself due to highly separated flow nonlinearities together with oscillatory evolution leading to nonlinear couplings between different aeroelastic modes.

Poincaré sections obtained from reconstructed state spaces may also be used to verify the peculiar changes in trajectory shape and amplitude. In Figure 12, one may observe considerable changes in Poincaré sections, as the turntable oscillatory frequencies increase, thereby indicating the existence of bifurcations. Again, in all these cases the Poincaré sections suggest complexity of the aeroelastic system and several changes in geometry of state space occur. Again, projected Poincaré sections as shown in Figure 13 can be used to infer the presence of bifurcations.

(a)

(b)

(c)

(d)

(e)
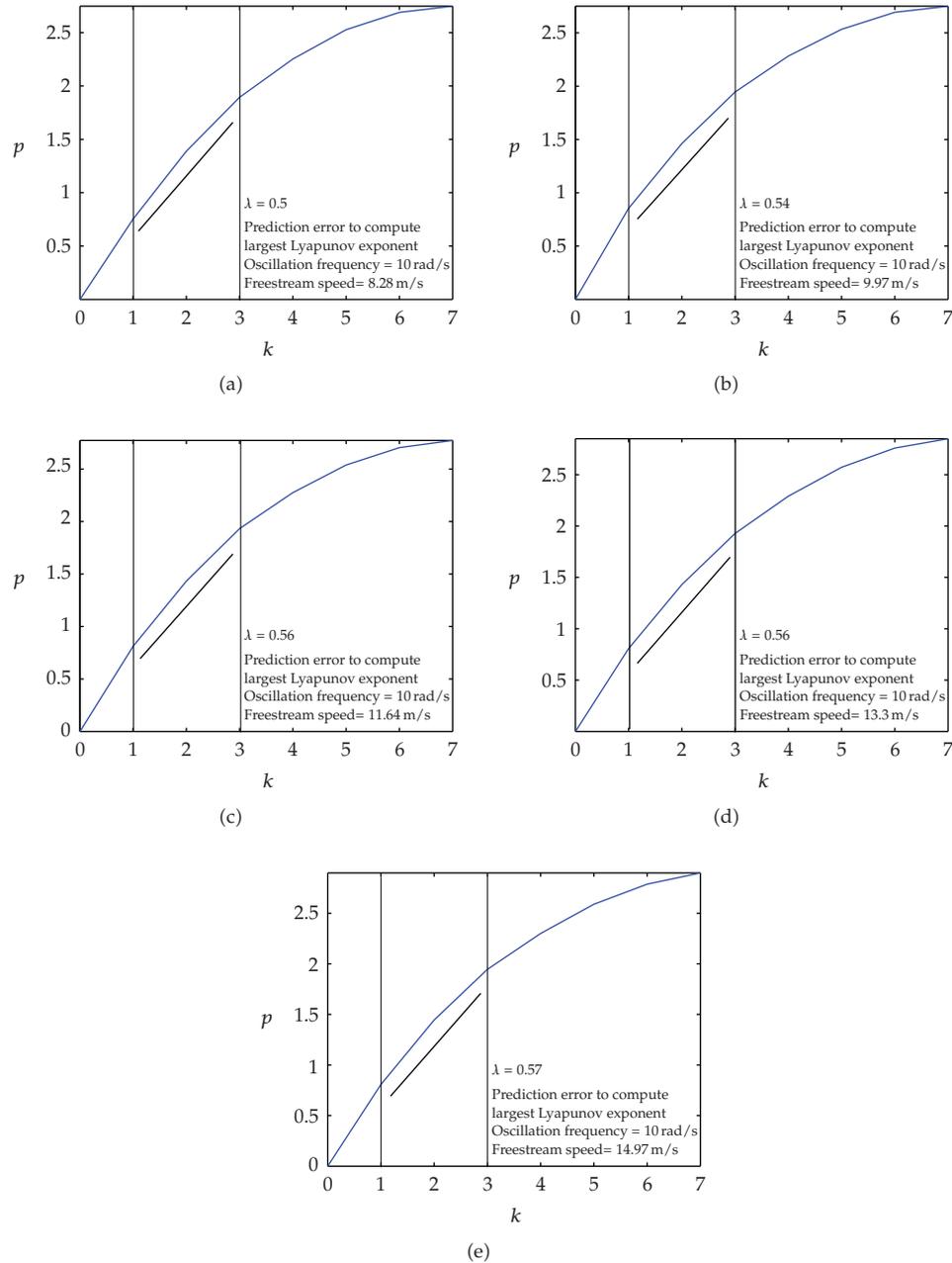
Figure 10: Largest Lyapunov exponents with the freestream speeds: 8.28, 9.97, 11.64, 13.30, and 14.97 m/s, respectively, at fixed oscillatory turntable frequency of $\omega = 10.0$ rad/s.

The largest Lyapunov exponent was also computed via prediction error technique [27], for turntable frequency variation cases (cf. Table 1). In all conditions, the largest Lyapunov exponents are positive as presented in Table 3, indicating chaotic patterns for the aeroelastic wing responses and chaos-chaos bifurcations.
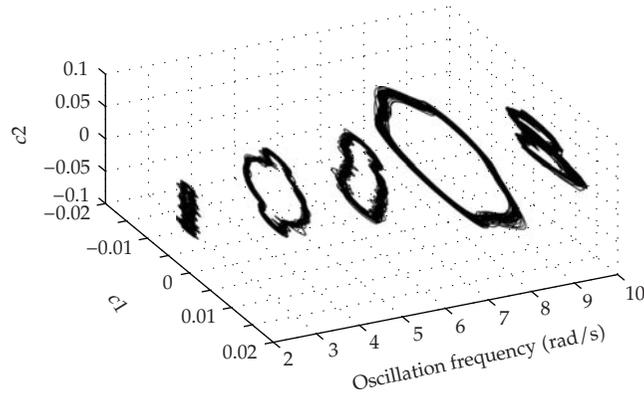
Figure 11: Reconstructed state space evolution with turntable oscillatory frequencies: 2.0, 4.0, 6.0, 8.0, and 10.0 rad/s, respectively, at fixed wind tunnel freestream speed of $U = 15.0$ m/s.
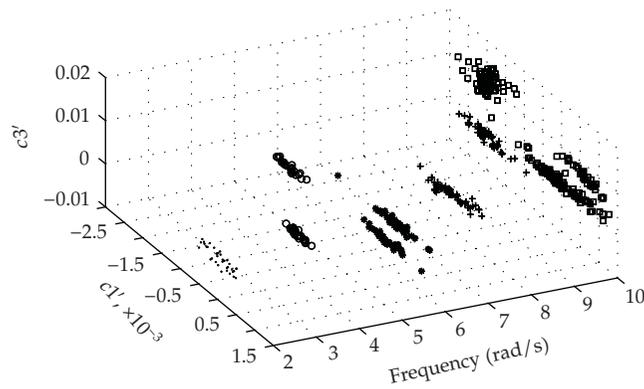

Figure 12: Poincaré sections evolution with turntable oscillatory frequencies: 2.0, 4.0, 6.0, 8.0, and 10.0 rad/s, respectively, at fixed wind tunnel freestream speed of $U = 15.0$ m/s.
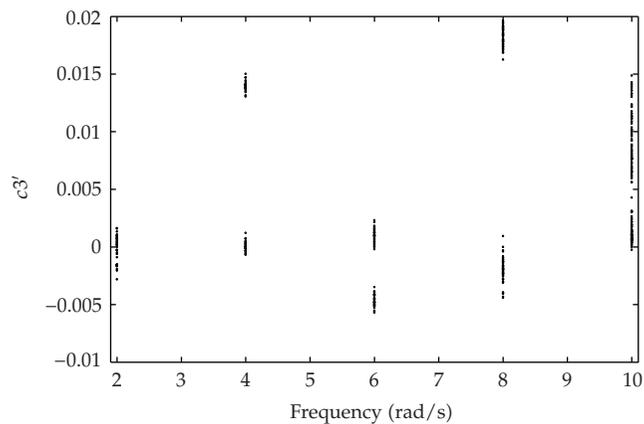
Figure 13: Projection of the Poincaré sections evolution with turntable oscillatory frequencies: 2.0, 4.0, 6.0, 8.0, and 10.0 rad/s, respectively, at fixed wind tunnel freestream speed of $U = 15.0$ m/s.

(a)

(b)

(c)

(d)

(e)
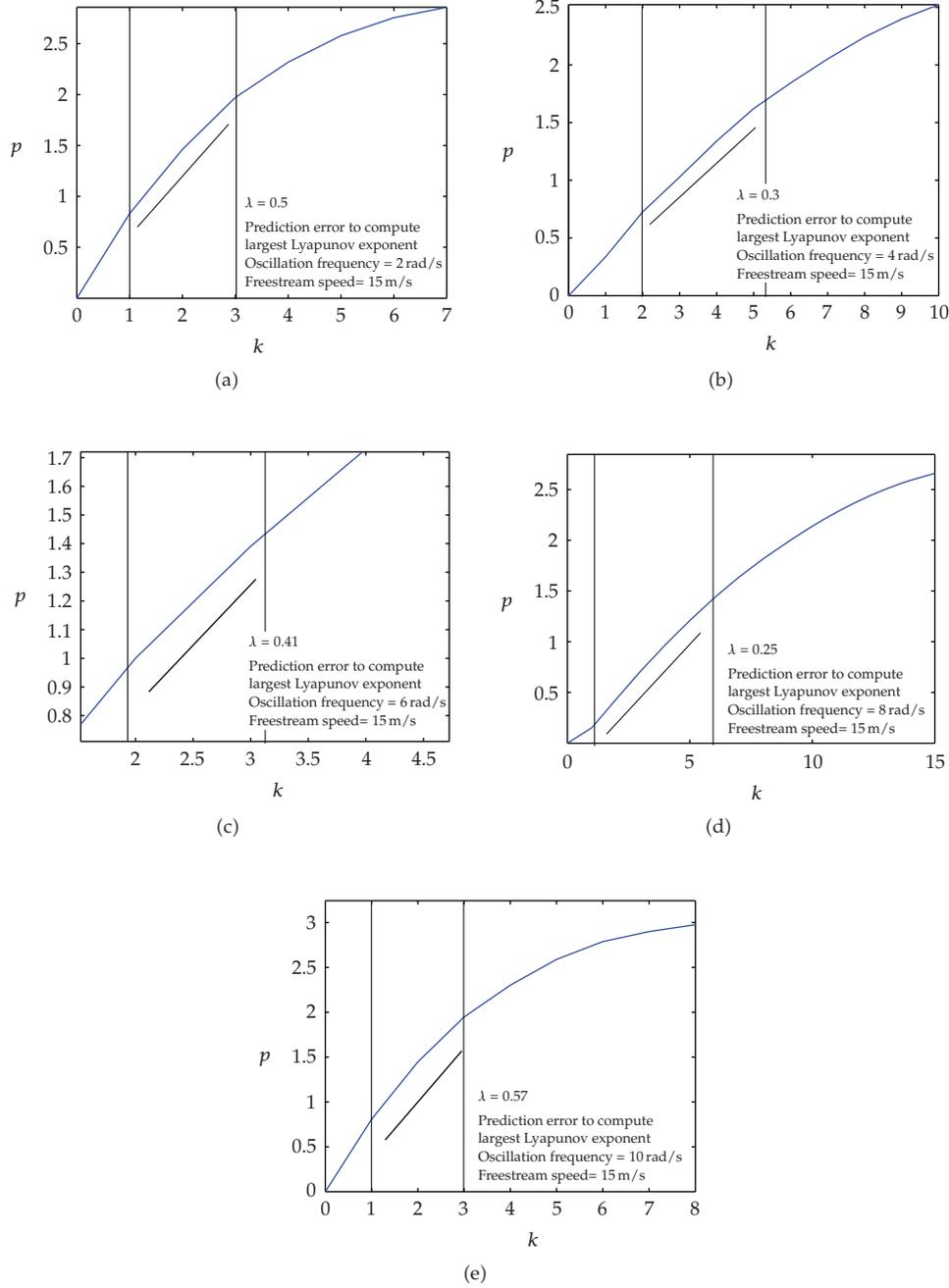
Figure 14: Largest Lyapunov exponents via prediction error computed for turntable oscillatory frequencies: 2.0, 4.0, 6.0, 8.0, and 10.0 rad/s, respectively, at fixed wind tunnel freestream speed of $U = 15.0$ m/s.

Similarly to the previous cases (freestream speed range), Lyapunov exponents for that analysis have been obtained from plottings as presented in Figure 14 and summarized in Table 3.

Table 3: Lyapunov exponents by prediction error technique [27] for fixed wind tunnel freestream speed ($U = 15.0$ m/s) and a range of turntable oscillatory frequencies (cf. Table 1).

Turntable oscillatory frequency (rad/s)	2.0	4.0	6.0	8.0	10.0
Largest exponent	0.50	0.30	0.41	0.25	0.57

6. Concluding Remarks

Techniques from nonlinear time series analysis theory have been presented in this work to investigate chaotic patterns of nonlinear motion-induced aeroelastic responses. Experimental tests with a wind tunnel aeroelastic wing model mounted on an oscillatory turntable have been executed with highly separated flow field conditions. Aeroelastic time series have been obtained from strain gages measurements, which were used directly with a variety of time series analysis tools. The time series have been tested using the surrogate data method, in order to investigate whether or not the data was representative of a nonlinear process. The results justify the application of techniques in order to search for bifurcations and chaotic patterns, since the linear hypothesis could be rejected.

The SVD method has been used to reconstruct the state spaces from the experimentally acquired aeroelastic time series, and the trajectories and subsequent assessment of the Poincaré sections have indicated complex behavior, such as bifurcations and chaos.

The evolution with freestream speed for a fixed turntable oscillatory frequency suggests the occurrence of chaos-chaos bifurcations, since changes in the shape of the attractor and Poincaré sections have been observed and all largest Lyapunov exponents are positive. Moreover, evolutions in terms of turntable oscillation frequency at a fixed wind tunnel freestream velocity also show the occurrence of bifurcations. Reconstructed spaces have also revealed complex motion amplitude changes with respect to parametric variation (freestream speed or turntable oscillatory frequency).

The occurrence of bifurcations, mainly in the cases where a sudden increase in amplitude of motion happens, reinforces the importance of nonlinear behavior study in aeroelastic systems. Further investigations to check experimental nonlinear aeroelastic response features with other time series analysis tools are planned.

Acknowledgments

References

[1] P. P. Friedmann, "The renaissance of aeroelasticity and its future," in *Proceedings of the International Forum on Aeroelasticity and Structural Dynamics (CEAS '97)*, pp. 19–49, Rome, Italy, June 1997.

[2] I. E. Garrick, "Aeroelasticity—frontiers and beyond," *AIAA Journal of Aircraft*, vol. 13, no. 9, pp. 641–657, 1976.

[3] L. E. Ericsson and J. P. Reding, "Fluid dynamics of unsteady separated flow. Part II. Lifting surfaces," *Progress in Aerospace Sciences*, vol. 24, no. 4, pp. 249–356, 1987.

[4] B. H. K. Lee, S. J. Price, and Y. S. Wong, "Nonlinear aeroelastic analysis of airfoils: bifurcation and chaos," *Progress in Aerospace Sciences*, vol. 35, no. 3, pp. 205–334, 1999.

[5] E. H. Dowell and D. Tang, "Nonlinear aeroelasticity and unsteady aerodynamics," *AIAA Journal of Aircraft*, vol. 40, no. 9, pp. 1697–1707, 2002.

[6] H. Alighanbari and B. H. K. Lee, "Analysis of nonlinear aeroelastic signals," *AIAA Journal of Aircraft*, vol. 40, no. 3, pp. 552–558, 2003.

[7] E. F. Sheta, V. J. Harrand, D. E. Thompson, and T. W. Strganac, "Computational and experimental investigation of limit cycle oscillations of nonlinear aeroelastic systems," *AIAA Journal of Aircraft*, vol. 39, no. 1, pp. 133–141, 2002.

[8] J. W. Edwards, "Computational aeroelasticity," in *Structural Dynamics and Aeroelasticity*, A. K. Noor and S. L. Venner, Eds., vol. 5 of *Flight Vehicle Materials, Structures and Dynamics—Assessment and Future Directions*, pp. 393–436, ASME, New York, NY, USA, 1993.

[9] J. G. Leishman and T. S. Beddoes, "A semi-empirical model for dynamic stall," *Jornal of the American Helicopter Society*, vol. 34, no. 3, pp. 3–17, 1989.

[10] F. D. Marques, *Multi-layer functional approximation of non-linear unsteady aerodynamic response*, Ph.D. thesis, University of Glasgow, Glasgow, UK, 1997.

[11] J. S. Bendat and A. G. Piersol, *Random Data: Analysis & Measurement Procedures*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1986.

[12] T. Schreiber and A. Schmitz, "Surrogate time series," *Physica D*, vol. 142, no. 3-4, pp. 346–382, 2000.

[13] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge, UK, 2nd edition, 2004.

[14] D. S. Broomhead and G. P. King, "Extracting qualitative dynamics from experimental data," *Physica D*, vol. 20, no. 2-3, pp. 217–236, 1986.

[15] A. H. Nayfeh and B. Balachandran, *Applied Nonlinear Dynamics*, John Wiley & Sons, New York, NY, USA, 1995.

[16] R. M. G. Vasconcellos, *Reconstrução de espa ços de estados aeroelásticos por decomposiçãao em valores singulares*, M.S. thesis, Universidade de São Paulo—EESC-USP, São Paulo, Brazil, 2007.

[17] R. C. Hilborn, *Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineers*, The Clarendon Press, Oxford University Press, New York, NY, USA, 2nd edition, 2000.

[18] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, vol. 898, pp. 366–381, Springer, Berlin, Germany, 1981.

[19] N. J. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a time series," *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.

[20] F. D. Marques, E. M. Belo, V. A. Oliveira, J. R. Rosolen, and A. R. Simoni, "On the investigation of state space reconstruction of nonlinear aeroelastic response time series," *Shock and Vibration*, vol. 13, no. 4-5, pp. 393–407, 2006.

[21] D. Kugiumtzis and N. Christophersen, "State space reconstruction: method of delays vs singular spectrum approach," Research Report 236, Department of Informatics, University of Oslo, Oslo, Norway, 1997.

[22] M. Casdagli, S. Eubank, D. Farmer, and J. Gibson, "State space reconstruction in the presence of noise," *Physica D*, vol. 51, no. 1–3, pp. 52–98, 1991.

[23] M. A. Athanasiu and G. P. Pavlos, "SVD analysis of the magnetospheric AE index time series and comparison with low-dimensional chaotic dynamics," *Nonlinear Processes in Geophysics*, vol. 8, no. 1-2, pp. 95–125, 2001.

[24] U. ParlitzJ. A. K. Suykens and J. Vandewalle, "Nonlinear time-series analysis," in *Nonlinear Modeling-Advanced Black-Box Techniques*, pp. 209–239, Kluwer Academic Publishers, Boston, Mass, USA, 1998.

[25] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank, and J. D. Farmer, "Using surrogate data to detect nonlinearity in time series," in *Nonlinear Modeling and Forecasting*, vol. 12 of *SFI Studies in the Sciences of Complexity*, pp. 163–188, Addison-Wesley, Reading, Mass, USA, 1992.

[26] C. Merkwirth, U. Parlitz, and W. Lauterborn, "TSTOOL—a software package for nonlinear time series analysis," in *Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling*, J. A. Suykens and J. Vandewalle, Eds., Katholieke Universiteit Leuven, Leuven, Belgium, July 1998.

[27] S. Sato, M. Sano, and Y. Sawada, "Practical methods of measuring the generalized dimension and the largest Lyapunov exponent in high-dimensional chaotic systems," *Progress of Theoretical Physics*, vol. 77, no. 1, pp. 1–5, 1987.

[28] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D*, vol. 9, no. 1-2, pp. 189–208, 1983.

[29] T. Yalçinkaya and Y.-C. Lai, "Phase characterization of chaos," *Physical Review Letters*, vol. 79, no. 20, pp. 3885–3888, 1997.

[30] F. D. Marques, E. M. Belo, V. A. Oliveira, J. R. Rosolen, and A. R. Simoni, "Non-linear phenomena analysis of stall-induced aeroelastic oscillations," in *Proceedings of the 45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, vol. 6, pp. 4507–4513, Palm Springs, Calif, USA, April 2004.

[31] A. R. Simoni, *Análise de séries temporais esperimentais não lineares*, Ph.D. thesis, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Paulo, Brazil, 2007.

[32] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D*, vol. 16, no. 3, pp. 285–317, 1985.

[33] F. D. Marques, R. M. G. Vasconcellos, and A. R. Simoni, "Analysis of an experimental aeroelastic system through nonlinear time series," in *Proceedings of the International Symposium on Dynamic Problems of Mechanics (DINAME '09)*, Angra dos Reis, Brazil, March 2009.

Research Article

Limit Cycle Prediction Based on Evolutionary Multiobjective Formulation

M. Katebi,[1] H. Tawfik,[1] and S. D. Katebi[2]

[1] *School of Computing Research, Liverpool Hope University, L16 9JD Liverpool, UK*
[2] *Department of Computer Science and Engineering, School of Engineering, Shiraz University, 71348-51154 Shiraz, Iran*

Correspondence should be addressed to S. D. Katebi, katebi@shirazu.ac.ir

This paper is concerned with an evolutionary search for limit cycle operation in a class of nonlinear systems. In the first part, single input single output (SISO) systems are investigated and sinusoidal input describing function (SIDF) is extended to those cases where the key assumption in its derivation is violated. Describing function matrix (DMF) is employed to take into account the effects of higher harmonic signals and enhance the accuracy of predicting limit cycle operation. In the second part, SIDF is extended to the class of nonlinear multiinput multioutput (MIMO) systems containing separable nonlinear elements of any general form. In both cases linearized harmonic balance equations are derived and the search for a limit cycle is formulated as a multiobjective problem. Multiobjective genetic algorithm (MOGA) is utilized to search the space of parameters of theoretically possible limit cycle operations. Case studies are presented to demonstrate the effectiveness of the proposed approach.

1. Introduction

The theory of linear dynamic systems is now well understood and is widely applied to many fields of engineering such as robotics, processes control, ship stirring to name a few. However, nonlinear systems have received less attention, the reason being the diversity and complexity of these systems. With the advent of fast and powerful digital computers, research for a more precise and accurate analysis of nonlinear systems has grown considerably [1–3]. One such method which traditionally has been applied is the replacement of nonlinear behavior with a quasi-linear gain called describing function (DF) [4].

Describing function theory and techniques represent a powerful mathematical approach for analyzing and designing nonlinear systems. The main motivation for DF
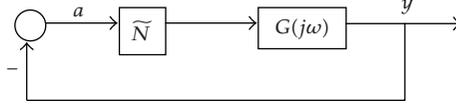
Figure 1: Feed back configuration for nonlinear systems.

techniques is the need to understand the behavior of nonlinear systems, which in turn is based on the simple fact that every system is nonlinear, except in very limited operating regimes. The basic philosophy of DF is to replace each nonlinear element with a quasi-linear descriptor or describing function. The functional form of such a descriptor is governed by several factors, the type of input signal, (which is assumed in advance), and the approximation criterion, for example, minimization of mean squared error. This technique is dealt with very thoroughly in a number of texts for the case of nonlinear systems with a single nonlinearity [4]. One category of DFs that has been particularly successful is the sinusoidal input describing function (SIDF).

The fundamental ideas and use of the SIDF approach can best be introduced by overviewing the most common application, limit cycle analysis for a system with a single nonlinearity. A limit cycle (LC) is a periodic signal, $x_{LC}(t + T) = x(t)$ for all t and for some T (the period), such that perturbed solutions either approach x_{LC} (a stable limit cycle) or diverge from it (an unstable one). Recently several analytical techniques have been proposed in the framework of the Bifurcation theory for investigating different aspects of limit cycle operation [5, 6]. Nonlinear techniques for modeling the periodic signal in general and limit cycles in particular have also been suggested [7]. However, an approach to LC analysis that has gained widespread acceptance is the graphical frequency-domain SIDF method [8, 9].

In this paper the describing function method is first extended to the case of single input single output (SISO) systems, where the condition for elementary SIDF application is not strictly satisfied [10]. When the linear system element does not attenuate the super-harmonic components around the feedback loop, the input to the nonlinear element will not be a pure sinusoid and will be a distorted waveform containing higher harmonics. The first part of this paper uses the (DFM), to account for higher harmonics and widen the scope of applications of the SIDF method [11].

The second part of the paper further extends the SIDF techniques to a class of multiloop nonlinear systems in which the nonlinear elements are separable from the linear part. In both cases, emphasis is placed on the multiobjective formulation of predicting the limit cycle operation and the subsequent solution of the harmonically linearised system equations by the multiobjective genetic algorithms (MOGA).

2. Harmonic Analysis

The harmonic balance equation for the autonomous feedback configuration of Figure 1, in which the nonlinearity exists as a separable element in an otherwise linear system, is given as

$$1 + N(a, \omega)G(j\omega) = 0, \tag{2.1}$$

where $N(a, \omega)$ is the SIDF representing the nonlinear element and $G(j\omega)$ is the frequency transfer function of the linear part. In a simple harmonic analysis, some form of the solution

of (2.1) is sought. However, the valid application of the SIDF requires that the input signal to the nonlinear element be essentially sinusoidal in form. This is a condition which imposes an overall low-pass frequency characteristic on the linear system elements such that the super-harmonic signals are attenuated around the feedback loop. Based on the assumption that the input to the nonlinear element is a pure sinusoidal, the SIDF is derived as

$$N(a, \omega) = \frac{(b_1 + jc_1)e^{j\omega t}}{ae^{j\omega t}} = \frac{1}{a}(b_1 + jc_1), \tag{2.2}$$

where b_1 and c_1 are the first harmonics in the Fourier series of N, and a is the amplitude of the input sinusoidal. In general the SIDF is a function of both input amplitude and input frequency, however, for single valued, sector bounded nonlinearities which are the subject of this work, the describing function is real and only a function of input amplitude a, therefore for brevity $N(a)$ is used to denote the SIDF.

3. Extension of Higher Harmonic Analysis to SISO Systems

In cases where the requirements of the overall low-pass characteristics of the linear system elements are met, the conventional graphical technique in the frequency domain has been shown to give acceptable results [9]. This elementary SIDF analysis and the associate graphical solution are well documented [4]. However, in those cases where the amplitude of higher harmonics of the Fourier series at the output of the nonlinear element are significant and are not suppressed around the feedback loop, the key assumption in derivation of the SIDF is violated [10]. The reason being that the input to the nonlinear element will be a distorted waveform containing the fundamental as well as other harmonics. One method to remedy this circumstance is to find error bounds for the elementary SIDF. Based on Tsypkin's method, a strategy with which to find systems with a low-pass linear part for which the describing function technique erroneously predicts limit cycles is outlined in [10]. A more attractive method that widens the scope of the SIDF applications, as well as achieving a more accurate limit cycle prediction is to account for additional harmonics. An appropriate technique is the so called "describing function matrix" (DFM), which is a mathematical formulation that allows an arbitrary number of harmonics to be taken into account in the SIDF analysis [11, 12].

Consider a nonlinear element N which gives output $y = N(x)$ for an input x. It is required to examine the behavior of N under an input which is exactly periodic. Let this be

$$x = \text{Re}\left[\sum_{r=0}^{\infty} a_r e^{jr\omega t}\right], \tag{3.1}$$

where $\overline{a} = (a_0, a_1, a_2 \ldots)^T$ is the vector of amplitudes of the harmonic components at the input to the nonlinear element and for a finite number of harmonics ($r = m$), $\overline{a}_m = (a_0, a_1, a_2, \ldots, a_m)^T$, T denotes transposition of $\overline{a}$. Let C be an analogous vector of complex Fourier coefficients of the nonlinearity output y, then an infinite matrix N (the describing function matrix) may be defined as follows:

The 0th column is given by $N_{k_0} = C(a_0)/a_0$,

The 1st column is given by $N_{k_1} = [(C_k(a_1) - C_k(a_0))/a_1]$,

where, for example, the (3.1), (2.1) element represents the amplitude of the 3rd output harmonic divided by that of the 1st input harmonic. In general, the jth column is given by,

$$N_{k_j} = \left[\frac{C_k(a_j) - C_k(a_{j-1})}{a_j} \right], \tag{3.2}$$

where k signifies the row number and is the index of harmonic components. If the nonlinear element is not frequency dependent and no bias level is present, then the 0th column can be ignored, and in this case the first element of the first column is the normal SIDF. Also note that if the nonlinear element is symmetrical and odd valued, then even rows ($k = 2, 4, 6, \ldots$) can be ignored.

Now define the linear part $G^m = \mathrm{Diag}(g(0j\omega), g(j\omega), g(j2\omega), \ldots, g(jm\omega))$ and representing the nonlinear element with DFM for m harmonic components, the condition for existence of oscillation is that the following simultaneous harmonic balance equation (3.3)

$$(I + N^m G^m)\bar{a}_m = 0 \tag{3.3}$$

has a solution $\psi_m = \left[\begin{smallmatrix} \omega \\ a^m \end{smallmatrix} \right]$. The reason for G being a diagonal matrix is that different harmonics do not interact in passing through the linear element.

Consider the nonlinear SISO system shown in Figure 1 in which $G(j\omega)$ and $N(a)$ are scalars. Assume a fundamental plus a third harmonic to be present at the input to the nonlinear element. Further, assume that the nonlinear element is not frequency dependent, but is symmetrical and odd valued function. The describing function matrix for an input of the following form

$$x(t) = a_1 \sin \omega t + a_3 \sin(3\omega t + \phi) \tag{3.4}$$

can be constructed as

$$N \triangleq \begin{bmatrix} \dfrac{c_1(a_1)}{a_1} & \dfrac{c_1(a_1, a_3) - c_1(a_1)}{a_3} \\[3mm] \dfrac{c_3(a_1)}{a_1} & \dfrac{c_3(a_1, a_3) - c_3(a_3)}{a_3} \end{bmatrix}, \tag{3.5}$$

where ϕ is the phase shift between the two harmonics and $c_1(a_1)$, $c_3(a_3)$ are respectively, the first and the third coefficients of Fourier components at the nonlinear output, due to the component $a_1 \sin(\omega t)$ only, and $c_1(a_1, a_3)$, $c_3(a_1, a_3)$ are the corresponding coefficients due to the sum of the two signals at the input to the nonlinear element.

In the case of only two harmonics such as $x(t)$ given by (3.4), the harmonic balance equation (3.3) is reduced to the following two simultaneous equations

$$(I + N^2 G^2), \tag{3.6}$$

where N^2 is the second order (first and the third harmonic) DFM and G^2 is a 2×2 diagonal matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} n_{11}(a_1) & n_{12}(a_1, a_3) \\ n_{21}(a_1) & n_{22}(a_1, a_3) \end{pmatrix} \begin{pmatrix} g(j\omega) & 0 \\ 0 & g(j3\omega) \end{pmatrix} \begin{pmatrix} a_1 \\ a_3 \angle \phi \end{pmatrix} = 0, \tag{3.7}$$

equation (3.7) may be written as

$$\begin{aligned} (1 + n_{11}(a_1)g(j\omega))a_1 + n_{12}(a_1, a_3)g(j3\omega)a_3 \angle \phi &= 0, \\ n_{21}(a_1)g(j\omega)a_1 + (1 + n_{22}(a_1, a_3)g(j3\omega))a_3 \angle \phi &= 0, \end{aligned} \tag{3.8}$$

where $n_{11}, n_{12}, n_{21}, n_{22}$ represent the elements of matrix N as defined by (3.5).

Rearranging the second equation in the equation set (3.8)

$$a_3 \angle \phi = - \left| \frac{n_{21}(a_1)g(j\omega)}{1 + n_{22}(a_1, a_3)g(j3\omega)} \right| a_1. \tag{3.9}$$

A search has to be carried out in the space of $(a_1, \phi, \omega, a_3$ in that order) to find those values of a_3 and ϕ which will satisfy (3.9). If such values are found, the result will be substituted into the first equation of equation set (3.8), to test for this equation which governs the fundamental oscillation.

Substituting (3.9) into the first equation of equation set (3.8) and rearranging yields

$$n_{11}(a_1)g(j\omega)a_1 + G_1 = -1, \tag{3.10}$$

where $G_1 = -n_{12}(a_1, a_3)g(j3\omega)n_{21}(a_1)g(j\omega)/(1 + n_{22}(a_1, a_3)g(j3\omega))$, (3.10) represents the overall harmonic balance equation for both the fundamental and the third harmonic components. Hence, a limit cycle with parameters (a_1, a_3, ω, ϕ) exists when the following two equations are simultaneously satisfied;

$$\begin{aligned} |n_{11}(a_1)g(j\omega)a_1 + G_1| &= 1, \\ \angle(n_{11}(a_1)g(j\omega)a_1 + G_1) &= 180 \pm 2k\pi, \quad k = 0, 1, \dots. \end{aligned} \tag{3.11}$$

Alternatively, (3.8) may be rearranged as follows and be directly used as the objective (fitness function) in the subsequent MOGA search as the satisfaction of these objectives implies the solution of the two simultaneous harmonic balance equations (3.8)

$$\begin{aligned} \text{Obj}_1 &= |(1 + n_{11}(a_1)g(j\omega))a_1 + n_{12}(a_1, a_3)g(j3\omega)a_3 \exp(j\phi)| \le \varepsilon, \\ \text{Obj}_2 &= |n_{21}(a_1)g(j\omega)a_1 + (1 + n_{22}(a_1, a_3)g(j3\omega))a_3 \exp(j\phi)| \le \varepsilon. \end{aligned} \tag{3.12}$$

For the exact solution of the simultaneous equations (3.12), values of Obj_1 and Obj_2 should reach zero but ε is a small positive number (as near to zero as possible). An intelligent search based on MOGA with the above defined objective (fitness) function is carried out over the

space of (a_1, a_3, ω, ϕ) to facilitate an efficient and an accurate prediction of limit cycle in the presence of higher harmonics.

4. Extension of Harmonic Analysis to MIMO System

Extension of describing function techniques to multiloop nonlinear systems is not new and follows the development of the frequency domain multivariable linear theory [13, 14]. A numerical technique with a graphical interpretation for quantifying the limit cycle parameters in multivariable systems in the frequency domain is reported in [15]. The extension of the graphical techniques to multiloop systems with coupled multivalued nonlinear elements is reported in [16]. Another graphical method based on the phasor diagram which particularly gives accurate limit cycle prediction for relay systems is developed in [8]. A computer aided design (CAD) tool for limit cycle prediction aimed at educational purposes is reported in [9]. Another novel numerical technique based on deriving the least damped eigenvalue to the imaginary axis for nonlinear systems with multiple nonlinearities is suggested in [17]. Based on defining an appropriate error function, the authors use both eigenvalue and eigenvectors to formulate a generalized Newton-Raphson method to solve for the state variable amplitude in a minimum norm sense [17]. Most of the above mentioned techniques are essentially dependent on the graphical displays in the frequency domain. While this is useful for getting insight into the subsequent compensator design, an efficient technique for accurately quantifying the limit cycle parameters is still highly desirable.

In this paper, a multiobjective formulation is presented to search numerically for limit cycles in a class of multiloop nonlinear systems. The approach is computationally efficient and is based on multiobjective genetic algorithms (MOGA).

5. Limit Cycle Prediction in Nonlinear Multivariable Systems

Provided that certain limitations are placed on the form of the linear system elements, the extension of the of harmonic linearization to multivariable systems is conceptually straightforward. The equation governing limit cycle operation in the autonomous multivariable nonlinear feedback system of Figure 1 can be expressed as

$$(T(a, j\omega) + I)a = 0, \tag{5.1}$$

where $T(a, j\omega) = N(a, j\omega)G(j\omega)$ and $N(a, \omega)$ is the matrix of sinusoidal input describing function corresponding to the nonlinear elements of N, and a is the column vector of the sinusoid at the inputs to these elements. The use of a single sinusoidal describing function analysis implies the following assumptions:

(a) Each element of $G(j\omega)$ acts as a low pass system so that higher harmonic signal components are effectively suppressed.

(b) If a limit-cycle is present, then all loops will oscillate at the same frequency. Experience indicates that this is a reasonable assumption, particularly if the nonlinear elements are similar and if the dominant linear elements have approximately the same frequency characteristics.

Equation (5.1) will have a nontrivial solution only if

$$\det[N(a,\omega)G(j\omega) + I] = 0. \tag{5.2}$$

Thus for no limit cycle to exist, no eigenvalue of $N(a,\omega)G(j\omega)$ can equal $(-1, j0)$.

The conventional graphical frequency domain method for the solution of (5.2) for single input single output systems with a single nonlinear element may be extended to MIMO systems employing Nyquist or the Inverse Nyquist Array. Invoking the Inverse Nyquist Array method [13], the possibility of limit cycle existence may be examined by studying the following inequalities:

$$\left|n_{kk}(a_k) + \widehat{g}_{kk}(j\omega)\right| > \sum_{j \neq k}\left|n_{jk}(a_k) + \widehat{g}_{jk}(j\omega)\right|, \tag{5.3}$$

$$\left|n_{kk}(a_k) + \widehat{g}_{kk}(j\omega)\right| > \sum_{j \neq k}\left|n_{jk}(a_k)\right| + \sum_{j \neq k}\left|\widehat{g}_{jk}(j\omega)\right|, \tag{5.4}$$

where $n_{jk}(a)$ and $\widehat{g}_{jk}(j\omega)$ are the jkth elements of the matrices, $N(a)$ for single valued nonlinear elements and $G^{-1}(j\omega)$, respectively. Inequality (5.3) implies that no limit cycle operation is theoretically possible if Gershgorin bands associated with each diagonal element of $(N(a)+G^{-1}(j\omega))$ do not encompass the origin in the complex frequency domain. Inequality (5.4) implies the same result, provided that the bands traced out by the Gershgorin discs on the loci of $\widehat{g}_{ii}(j\omega)$ and $n_{ii}(a)$ do not intercept for every i. The former representation is computationally more demanding, but gives less conservative results because of the weakening step between inequalities (5.3) and (5.4). The Gershgorin discs set bounds to the eigenvalue locations and any method of limit cycle prediction based on band intersection should have a tendency to be conservative.

An alternative approach lies in calculating eigenvalue of the harmonically linearised return ratio system equation. For a general system which contains both on and off diagonal nonlinear elements, the computational effort involved in determining the eigenvalue becomes almost formidable as, at any frequency, the return ratio matrix is a function of the signal amplitude at the input to the nonlinear elements [16].

A numerically based technique called the sequential loop balance method has also been devised [15]. This method is based on (5.1) for which $T(a, j\omega)$ has elements of the form $t_{ij}(a_j, j\omega)$, and a is a column vector at the input to the nonlinear elements such that $a_j = A_j \exp(j\phi_i)$, $j = 1, 2, \dots, n$, for all $j, i = 1, 2, \dots m$ over arbitrary ranges A, ω, and ϕ, a possible infinite number of solutions may exist for (5.1). For a specified value of frequency and a specified range of discrete values of the reference signal $a_1 = A_1 e^{j0}$, $A_1 > A_2, A_3, \dots, A_n$ a finite number of p sets of sinusoids $(a_{1k}, a_{2k}, \dots, a_{nk})$, $k = 1, 2, \dots, p$ which will satisfy the condition for harmonic signal balance in the nth system loop as given by (5.5) are found;

$$(t_{nn} + 1)a_n + \sum_{j=1}^{n-1} t_{nj} a_j = 0. \tag{5.5}$$

Next, the $(n-1)$th loop is considered but using only those finite sets of solution sinusoids derived from (5.5). If there are q solution sets of sinusoids $(q \leq p)$ which satisfy both equations, the $(n-2)$th loop equation is next examined using only the q sets of solution
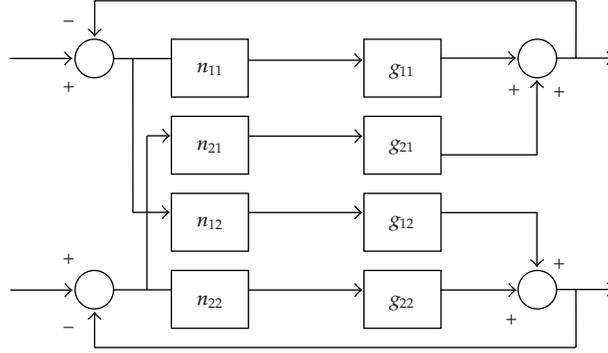
Figure 2: A two-input two-output nonlinear system.

sinusoids and the process is repeated in a sequential manner until loop 1 is reached. Clearly this method is also computationally demanding due to an indirected search over the specified ranges of parameters. Further, it does not search for the phase difference between the oscillating loops, and as discrete data are used in the computation, there is the possibility of solution sets which exist over the data intervals.

Considering the 2×2 autonomous system shown in Figure 2, the set of equations governing this model is given by (5.6)

$$
\begin{aligned}
\left(1 + n_{11}(a_1) g_{11}(j\omega)\right) a_1 + n_{21}(a_2) g_{21}(j\omega) a_2 e^{j\phi} = 0, \\
\left(n_{12}(a_1) g_{12}(j\omega) a_1 + 1 + n_{22}(a_2) g_{22}(j\omega)\right) a_2 e^{j\phi} = 0.
\end{aligned}
\tag{5.6}
$$

In this case there are two equations and four variables (unknown). The solution of these equations is sought over a range of specific values of a_1, a_2, ω and φ, where a_1 and a_2 are amplitudes of limit cycles in loop 1 and 2, respectively, ω is frequency of oscillation for both loops and ϕ is the phase-shift between the loops.

6. Multiobjective Genetic Algorithms

Genetic algorithms (GAs) are search procedures based on the evolutionary process in nature. They differ from other approaches in that they use probabilistic and not deterministic criteria for progressing the search. The idea is that GA operates on a population of individuals, each individual representing a potential solution of the problem, and applies the principle of survival of the fittest to the population, so that the individuals evolve towards better solutions of the problem. Each individual is given a chromosoidal representation, which corresponds to the genotype of an individual in nature. Three operations can be performed on individuals in the population, *selection*, *cross-over* and *mutation*. These correspond to the selection of individuals in nature for breeding, where the fitter members of a population breed and pass-on their genetic material. The cross-over corresponds to the combination of genes by mating, and mutation to genetic mutation in nature. The selection is biased so that the "fittest" individuals are more likely to be selected for cross-over, the fitness being a function of the criteria which is being minimized. By means of these operations, the population will evolve towards a solution. Most GAs have been used for single objective optimization problems

[18], although several multiobjective schemes have been proposed [19, 20]. Applications of multiobjective evolutionary schemes to control systems analysis and design have been widespread [21, 22]. A formulation called the multiobjective genetic algorithm (MOGA) maintains the genuine multiobjective nature of the problem, and is essentially the scheme adapted here [23]. Further details for the MOGA can be found in [20]. The design philosophy of the MOGA differs from other methods in that a set of simultaneous solutions are sought and the designer then selects the best solution from the set. The idea behind the MOGA is to develop a population of Pareto-optimal or near Pareto-optimal solutions. The aim is to find a set of solutions which are nondominated and satisfy a set of inequalities. An individual j with a set of objective functions $\varphi^j = (\varphi_1^j, \ldots, \varphi_n^j)$ is said to be *nondominated* if for a population of N individuals there are no other individuals $k = 1, \ N; k \neq j$ such that

$$
\begin{aligned}
&\text{(a)} \quad \varphi_i^k \leq \varphi_i^j \quad \forall\, i = 1, 2, \ldots, n, \\
&\text{(b)} \quad \varphi_i^k < \varphi_i^j \quad \text{for at least one } i.
\end{aligned}
\tag{6.1}
$$

The MOGA is set into a multiobjective context by means of the fitness function. Individuals are ranked on the basis of the number of other individuals they are dominated by for the unsatisfied inequalities. Each individual is then assigned fitness according to their rank. The mechanism is described in detail in [20]. To summarize, the MOGA problem could be stated as:

Find a set of M admissible points P_j, $j = 1, \ldots, M$ such that

$$
\varphi_i^j \leq \varepsilon_i \quad (j = 1, 2, \ldots, m), \ (i = 1, 2, \ldots, n).
\tag{6.2}
$$

And such that φ^j $(j = 1, \ldots, M)$ are nondominated.

Genetic algorithms are naturally parallel and hence lend themselves well to multiobjective settings. They also work well on nonsmooth objective functions. Thus MOGA can be used to search the existence of any possible limit cycle operation in nonlinear MIMO systems.

7. Applications

7.1. SISO System

The feedback configuration of the single input single output system considered in this example is shown in Figure 3.

A Fourier analysis of the nonlinearity output shows that the amplitude of the third harmonic component is significant. Therefore, the elementary describing function analysis does not give accurate limit cycle prediction. Based on the calculation of the required Fourier coefficients at the output of the nonlinear element for the input of the form given by (3.4), the describing function matrix as defined by (3.5) is derived as,

$$
\text{DFM} = \begin{bmatrix} \dfrac{1.91}{a_1} & \dfrac{-1.12}{a_3} \\[2mm] \dfrac{0.64}{a_1} & \dfrac{0.89}{a_3} \end{bmatrix}
\tag{7.1}
$$
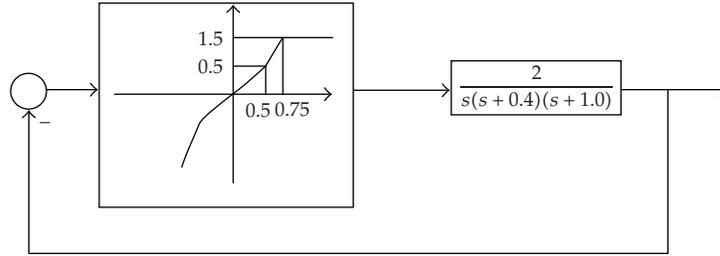
Figure 3: SISO nonlinear feedback system.

Table 1: MOGA parameters for the SISO system.

Number of generation	Population size	Selection method	Representation method	Multiobjective method	Mutation probability	Cross over probability
50	100	Tournament	Binary 8 bit for each parameter	Pareto	0.02	0.7

Table 2: Results of limit cycle prediction with MOGA for SISO system.

Best ranked	ω rad./sec.	a_1	a_3	φ radians	Objective 1	Objective 2
1	0.68	6.853	0.103	5.41	0.01	0.015
2	0.68	6.854	0.103	5.41	0.012	0.006
3	0.68	6.858	0.103	5.41	0.009	0.08
Simulation (actual)	**0.69**	**6.9**	**0.1**	**5.21**		

a_1 and a_3 denote the amplitude of the first and third harmonics at the input to the nonlinear element. The harmonic balance equation is formulated in the presence of the third harmonic using (3.12) with $\varepsilon = 0.01$. The range of limit cycle parameters are specified as

$$0.1 \leq \omega \leq 1.5, \qquad 3 \leq a_1 \leq 8.0, \qquad 0.05 \leq a_3 \leq 0.5, \qquad 0.1 \leq \phi \leq 2\pi. \tag{7.2}$$

The population is initialized randomly and the MOGA program is run to search the space of these parameters for the simultaneous satisfaction of objectives (3.12). The parameters of MOGA are shown in Table 1 and after 18 generations the algorithm converged with the value of $\varepsilon = 0.012$. The first three solutions are ranked based on the concept of non dominance and are listed in Table 2. The results obtained from simulation are also shown in bold in Table 2 and are comparable with those predicted by MOGA. The phase space of the predicted limit cycle is shown in Figure 4.

The computation time required depends largely on the number of generations, population size and on the number of bits specified for each parameter. For this example with the above specifications, 50 generations took 6.2 seconds on a Toshiba Centrino 1.6 with 512 M Ram and 2 M cashe.
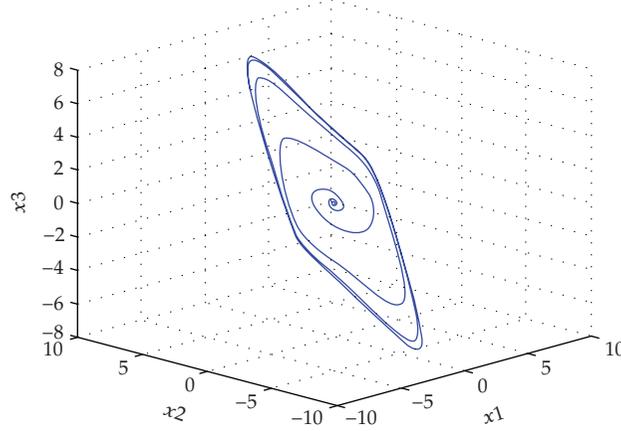
Figure 4: Phase space of limit cycle for SISO example.

7.2. MIMO Systems

In this section, two 2×2 nonlinear systems are presented. In the first example the elements of the nonlinear matrix are similar and consist of four ideal relays.

In both examples the numerical solution of (5.6) is formulated as a 2-objective problem, and the satisfaction of these objectives implies the solution of the simultaneous harmonic balance (5.6)

$$\text{Obj}_1 = \left| (1 + n_{11}(a_1)g_{11}(j\omega))a_1 + n_{21}(a_2)g_{21}(j\omega)a_2e^{\phi} \right| \leq \varepsilon,$$
$$\text{Obj}_2 = \left| n_{12}(a_1)g_{12}(j\omega)a_1 + (1 + n_{22}(a_2)g_{22}(j\omega))a_2e^{j\phi} \right| \leq \varepsilon, \tag{7.3}$$

where Obj_1 and Obj_2 are the values of the right-hand side of (5.6) which ideally should tend towards zero for specific values of a_1, a_2, ω, and ϕ. Upper and lower bounds are specified for ω, a_1, a_2, and ϕ, then the real generational MOGA with specified selection method, cross over, mutation rate and population size is called to search over the parameter space of ω, a_1, a_2, and ϕ. The required numerical accuracy may be achieved by specifying the number of genes (binary bits) for each individual in accordance with the value of ε. If conditions for inequalities (7.3) exist, then MOGA converges to the correct values of limit cycle parameters after a number of generations. If finer and more accurate parameter values are required, the bounds on the parameters ω, a_1, a_2, and ϕ may be tightened and the number of genes increased on the subsequent run of the MOGA program. In order to avoid the local minima and premature convergence the mutation probability rate may be taken higher at the beginning and decreased exponentially toward the end of the run. One advantage of limit cycle prediction based on MOGA is that multiple solutions may be distinguished by using the Niching mechanism [19].

Due to the inherent approximation in using the SIDF, and also the nature of multiobjective formulation, it may not be possible to reach the exact minimum which is zero, as required by the equation set (7.3). Therefore MOGA may converge to a set of Pareto optimal solutions and in the following examples the three best ranked are given as the final solutions.
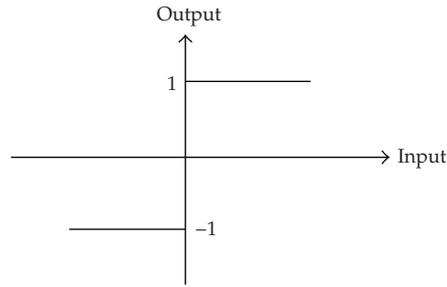
**Figure 5:** Ideal relay-nonlinear elements for MIMO Example 7.1.
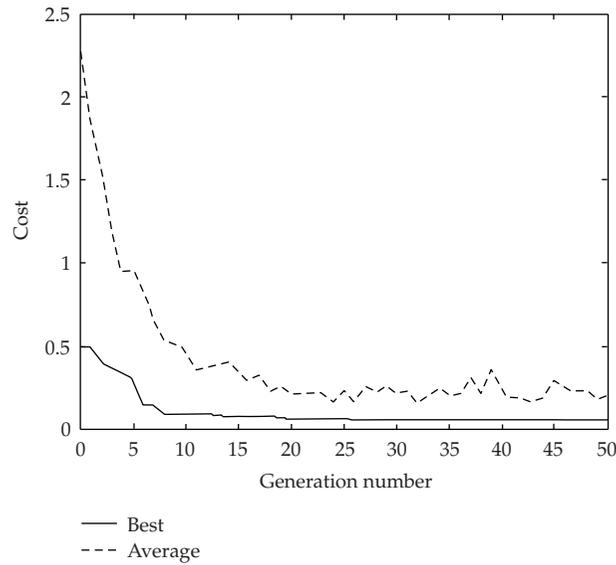
Figure 6: Convergence curve for MIMO Example 7.1.

Example 7.1. The nonlinear control system with two inputs and two outputs has the configuration of Figure 2 with four similar nonlinear elements ($n_{11} = n_{12} = n_{21} = n_{22}$) of the form shown in Figure 5.

The linear system matrix is

$$G = \begin{bmatrix} \dfrac{0.5}{s\left(s^2 + 1.5s + 0.5\right)} & \dfrac{-0.15}{s^2 + s + 1} \\[2ex] \dfrac{0.8}{s^2 + 3s + 2} & \dfrac{1}{s\left(s^2 + 1.6s + 0.8\right)} \end{bmatrix}. \tag{7.4}$$

In order to make the MOGA search more realistic, upper and lower bounds are specified for all parameters of possible limit cycle operation, namely, the frequency and amplitude for each loop. For this example, lower and upper bounds for w, a_1 and a_2 are specified as 0.1 and 1.5 for all three parameters and 0.1 to 2π radians for ϕ and $\varepsilon = 0.1$. Parameters of MOGA were also specified as in Table 3. The search is monitored interactively and if within the specified
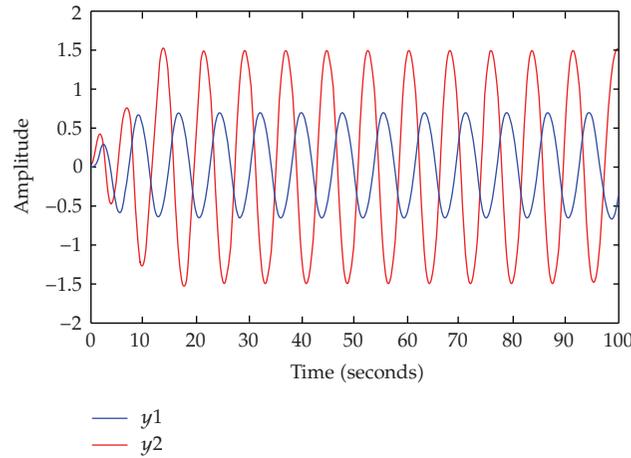
Figure 7: Simulation results for MIMO Example 7.1.

Table 3: MOGA parameters for MIMO Example 7.1.

Number of generation	Population size	Selection method	Representation method	Multiobjective method	Mutation probability	Cross over probability
50	50	Tournament	Binary 8 bit for each parameter	Pareto	0.02	0.7

Table 4: Results of limit cycle prediction with MOGA for Example 7.1.

Best ranked	ω	a_1	a_2	φ	Obj 1	Obj 2
1	0.86396	0.617785	1.42203	2.17477	0.0317952	0.0150623
2	0.863865	0.617787	1.41485	2.15623	0.0336897	0.00619064
3	0.845835	0.619658	1.44005	2.48668	0.00917439	0.0861531
Simulation (actual)	**0.863**	**0.620**	**1.406**	**2.168**		

range of the parameters feasible solutions exist, MOGA converges to solutions which are ranked according to their fitness values measured by the concept of non dominance, that is, Pareto optimal. For a population size of 50, it was observed that initially a large number of solutions exist. However, as the search progressed most members of the population tend to be concentrated around a narrow range of parameters. After 25 generations the algorithm converged and the results for the first three solutions are given in Table 4 and the convergence curve is shown in Figure 6. In Figure 6 the cost axis is a measure of fitness and the doted curve represent the sum of the average of the two objectives (average (Obj_1) + Average (Obj_2)) and the solid curve represent the sum of the best values of the two objectives (best (Obj_1) + best (Obj_2)). With the parameters given in Table 3 and the above specified ranges for parameters, the MOGA algorithm converged within 25 generations, consuming 5.2 seconds of CUP time on a Toshiba Centrino 1.6 with 512 M Ram and 2 M cashe. For the purpose of having an appropriate axes for the convergence curve the program was allowed to execute all 50 generations which took 11 seconds on the same machine.
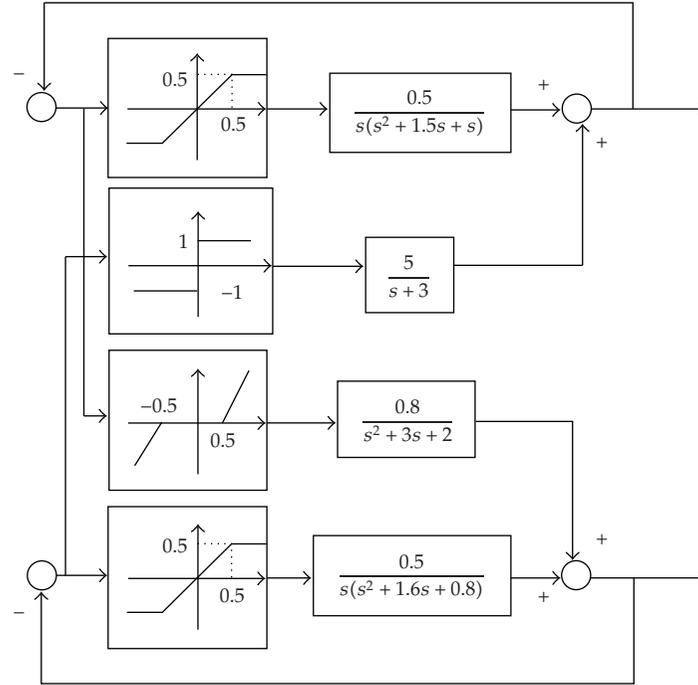
Figure 8: Two dimensional nonlinear systems for Example 7.2.

Table 5: Results of limit-cycle prediction with MOGA for MIMO Example 7.2.

	ω	a_1	a_2	φ	Obj 1	Obj 2
1	0.517728	1.18528	1.73296	3.6772	0.0229142	0.0611446
2	0.517638	1.18025	1.73396	3.67718	0.0181093	0.0655828
3	0.530139	1.16228	1.68292	3.67689	0.0410611	0.0383831
Simulation (actual)	**0.5400**	**1.073**	**1.543**	**3.500**		

In order to validate the results given by Table 4, the nonlinear system was simulated. The results of simulation (actual LC) compare well with the results obtained by MOGA. This is also shown in bold in Table 4 and in Figure 7. The reason for the very accurate limit cycle prediction in this example is two folds, one is that all nonlinear elements have the same characteristics and the other is that the linear system elements have a low pass frequency characteristics.

Example 7.2. The nonlinear system for this example is of the same configuration as in Figure 2, and is shown in Figure 8. In this case the nonlinear matrix consists of different nonlinear behavior such as saturation, ideal relay, and dead zone.

The same MOGA parameters were chosen as in Table 3 and the following ranges were specified for the parameters

$$0.1 \leq \omega \leq 1.0, \qquad 0.5 \leq a_1 \leq 1.5, \qquad 1.0 \leq a_2 \leq 2.0, \qquad 0.1 \leq \phi \leq 2\pi. \tag{7.5}$$
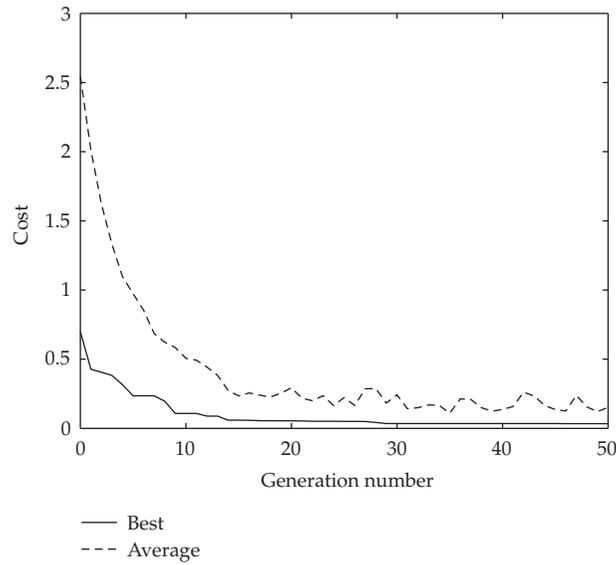
**Figure 9:** Convergence curve for MIMO Example 7.2.
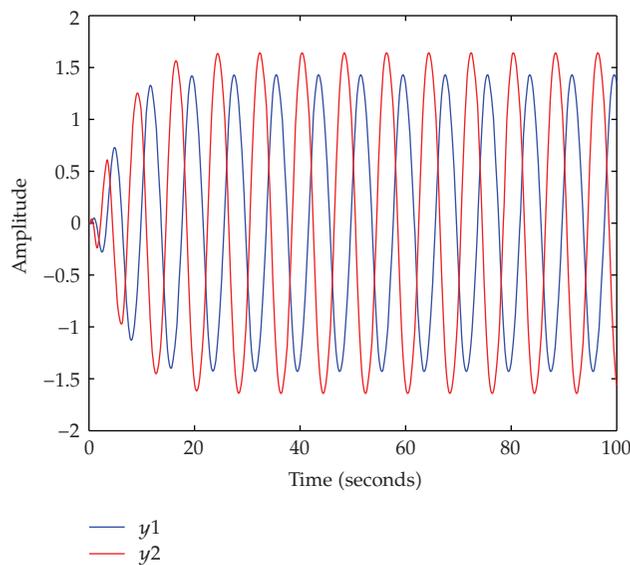
Figure 10: Simulation results for MIMO Example 7.2.

For this example, the algorithm converged within 30 generations, consuming 31.2 seconds of CUP time on the same machine. For all 50 generations 52 seconds of the CPU time is consumed on the same machine. The convergence curve is shown in Figure 9.

Contrary to Example 7.1 the results of simulation for this example, do not agree well with the results predicted by MOGA. This is shown (bold) in Table 5 and in Figure 10. As seen in Figure 10, the oscillation in loop 2 is distorted due to the higher harmonic components. This is caused by the frequency characteristics of element g_{21} of the linear part, which do

not strictly conform to the filter hypothesis, thus reducing the describing function accuracy and hence the reason for the small discrepancy between simulation and MOGA results. The development of similar analysis as for SISO systems will improve the SIDF accuracy significantly.

8. Conclusion

The single sinusoidal input describing function is extended by means of describing function matrix to account for higher harmonic signal components. In the analysis of single loop nonlinear systems, the describing function matrix provides a rigorous method, enhances limit cycle prediction and widens the scope of the valid application of the SIDF to those cases for which the assumption of the filter hypothesis is not strictly met. It is shown how, by formulating limit cycle prediction in the presence of higher harmonics as a multiobjective problem and using the MOGA procedure, a solution to the DF matrix equation can be obtained efficiently in numerical form. This is useful in illustrating both how a multiplicity of possible solutions may arise, and for emphasizing the effects of higher harmonic signal content.

Next, the single sinusoidal input describing function technique is extended to multiloop nonlinear systems. For the class of separable nonlinear elements of any general form, the linearized harmonic balance equations are derived. The search for limit cycle is formulated as a multiobjective problem. The multiobjective genetic algorithm is employed to obtain quantitative values for parameters of possible limit cycle operation. The MOGA search space is the space of the limit cycle parameters such as amplitudes, frequency and phase difference between the interacting loops. The algorithm is computationally efficient and provides very accurate results when the conditions for the SIDF application are satisfied.

In both cases of SISO and MIMO systems, emphasis is placed on the formulation of limit cycle parameters as a multiobjective problem and subsequent solution by the evolutionary MOGA method. Finally, examples of use are given to demonstrate the effectiveness of the proposed approach.

References

[1] R. Hilborn, *Chaos and Nonlinear Dynamics*, Oxford University Press, Oxford, UK, 2001.

[2] M. Vidyasagar, *Nonlinear Systems Analysis*, vol. 42 of *Classics in Applied Mathematics*, SIAM, Philadelphia, Pa, USA, 2nd edition, 2002.

[3] H. Khalil, *Nonlinear Systems*, Prentice-Hall, Upper Saddle River, NJ, USA, 3rd edition, 2002.

[4] J. H. Taylor, "Describing functions," in *Electrical Engineering Encyclopedia*, John Wiley & Sons, New York, NY, USA, 1999.

[5] M. Basso, R. Genesio, and A. Tesi, "A frequency method for predicting limit cycle bifurcations," *Nonlinear Dynamics*, vol. 13, no. 4, pp. 339–360, 1997.

[6] F. Angulo, M. di Bernardo, E. Fossas, and G. Olivar, "Feedback control of limit cycles: a switching control strategy based on nonsmooth bifurcation theory," *IEEE Transactions on Circuits and Systems I*, vol. 52, no. 2, pp. 366–378, 2005.

[7] E. Abd-Elrady, *Nonlinear approaches to periodic signal modeling*, Ph.D. thesis, Division of Systems and Control, Department of Information Technology, Uppsala University, Uppsala, Sweden, January 2005.

[8] K. C. Patra and Y. P. Singh, "Graphical method of prediction of limit cycle for multivariable nonlinear systems," *IEE Proceedings: Control Theory and Applications*, vol. 143, no. 5, pp. 423–428, 1996.

[9] F. Paoletti, A. Landi, and M. A. Innocentti, "A CAD tool for limit cycle prediction in nonlinear systems," *IEEE Transactions on Education*, vol. 39, no. 4, pp. 505–511, 1996.

[10] S. Engelberg, "Limitations of the describing function for limit cycle prediction," *IEEE Transactions on Automatic Control*, vol. 47, no. 11, pp. 1887–1890, 2002.

[11] A. I. Mees, "The describing function matrix," *IMA Journal of Applied Mathematics*, vol. 10, no. 1, pp. 46–67, 1972.

[12] S. D. Katebi and M. R. Katebi, "Combined frequency and time domain technique for the design of compensators for non-linear feedback control systems," *International Journal of Systems Science*, vol. 18, no. 11, pp. 2001–2017, 1987.

[13] H. H. Rosenbrock, *Computer-Aided Control System Design*, Academic Press, New York, NY, USA, 1974.

[14] A. G. J. MacFarlane and J. J. Belletrutti, "The characteristic locus design method," *Automatica*, vol. 9, pp. 575–588, 1973.

[15] J. O. Gray and P. M. Taylor, "Computer aided design of multivariable nonlinear control systems using frequency domain techniques," *Automatica*, vol. 15, no. 3, pp. 281–297, 1979.

[16] S. D. Katebi and M. R. Katebi, "Control design for multivariable multivalued nonlinear systems," *Systems Analysis Modelling Simulation*, vol. 15, no. 1, pp. 13–37, 1994.

[17] V. K. Pillai and H. D. Nelson, "A new algorithm for limit cycle analysis of nonlinear control systems," *Journal of Dynamic Systems, Measurement, and Control*, vol. 110, no. 3, pp. 272–277, 1988.

[18] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, Mass, USA, 1989.

[19] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, Wiley-Interscience Series in Systems and Optimization, John Wiley & Sons, Chichester, UK, 2001.

[20] C. A. Coello Coello, D. A. Van Veldhuizen, and G. B. Lamont, *Evolutionary Algorithms for Solving Multi-Objective Problems*, vol. 5 of *Genetic Algorithms and Evolutionary Computation*, Kluwer Academic Publishers, New York, NY, USA, 2002.

[21] G. P. Liu, J. B. Yang, and J. F. Whidborne, *Multi-Objective Optimisation and Control*, Research Studies Press, Baldock, UK, 2003.

[22] Mo. Jamshidi, L. D. Santos Coelho, R. A. Krohling, and P. J. Fleming, *Robust Control Design Using Genetic Algorithms*, CRC Press, Boca Raton, Fla, USA, 2002.

[23] C. M. Fonseca and P. J. Fleming, "Multi-objective genetic algorithms," in *Proceedings of the IEE Colloquium on Genetic Algorithms for Control Systems Engineering*, pp. 1–5, London, UK, May 1993.

Research Article

Intermittent Behavior and Synchronization of Two Coupled Noisy Driven Oscillators

Ângela Maria dos Santos,[1, 2] Sérgio Roberto Lopes,[1] and Ricardo Luiz Viana[1]

[1] *Departamento de Física, Universidade Federal do Paraná, Caixa Postal 19044, 81531-990 Curitiba, Paraná, Brazil*

[2] *Setor Escola Técnica, Universidade Federal do Paraná, UNED Paranaguá, 83215-750 Paranaguá, Paraná, Brazil*

Correspondence should be addressed to Ricardo Luiz Viana, viana@fisica.ufpr.br

The coupled system of two forced Liénard-type oscillators has applications in diode-based electric circuits and phenomenological models for the heartbeat. These systems typically exhibit intermittent transitions between laminar and chaotic states; what affects their performance and, since noise is always present in such systems, dynamical models should include these effects. Accordingly, we investigated numerically the effect of noise in two intermittent phenomena: the intermittent transition to synchronized behavior for identical and unidirectionally coupled oscillators, and the intermittent transition to chaos near a periodic window of bidirectionally coupled oscillators. We found that the transition from a nonsynchronized to a synchronized state exhibits a power-law scaling with exponent 3/2 characterizing on-off intermittency. The inclusion of noise adds an exponential tail to this scaling.

1. Introduction

Intermittency is a ubiquitous phenomenon in nonlinear dynamics. It consists of the intermittent switching between a laminar phase of regular behavior and irregular bursts. In one-dimensional quadratic-type maps it was first associated with a saddle-node bifurcation by Pomeau and Manneville, who also described its scaling characteristics [1]. A comprehensive theory of intermittency for such systems is now available [2]. Another context in which intermittency appears is related to the synchronization of coupled nonlinear oscillators.

Synchronization of nonlinear oscillators is a subject with a venerable history dating back from the early observation by Huygens that two pendula suspended from the same

frame—which provides the mechanical coupling—can synchronize their librations so as to become antiphase [3]. Most recently, the probability of synchronizing chaotic oscillations has opened a wide horizon of applications ranging from electronic circuits [4] and lasers [5] to biological rhythms like heartbeat [6].

A paradigmatic example of a nonlinear oscillator is the van der Pol equation [7] $\ddot{x} + \mu(x^2 - 1)\dot{x} + \omega_0^2 x = \rho \sin(\omega t)$, which was originally introduced to model the behavior of a AC-driven circuit with a nonlinear resistance, such as that yielded by a triode vacuum tube [8] and by Zener diodes [9] but which has also been used in other contexts, like in the phenomenological description of the heartbeat [10]. Actually the van der Pol equation belongs to a more general class of Liénard-type oscillators, whose dynamics in the presence of external forcing has been investigated, showing a rich behavior including crises, intermittency, and chaos [11].

The question of how two or more Liénard-type oscillators can synchronize their motions arises in the study of coupled vacuum-tube circuits [12] but is also relevant to the understanding of the mechanisms coupling the heartbeat with the nerve conduction [11]. The synchronization of two coupled Liénard-type oscillators, one of them being forced, has been considered in an earlier work of the authors, where we show the existence of different types of synchronization, according to the forcing and coupling parameters used [13].

In this work we focus on the influence of parametric noise in intermittency phenomena numerically observed in two Liénard-type forced oscillators. The first case is related to the intermittent transition to synchronized behavior in such systems in the presence of noise. Bearing in mind the usefulness of Liénard-type oscillators to model vacuum-tube circuits, we can regard the presence of parametric noise as unavoidable, since virtually each circuit component has a fluctuating magnitude (like resistances, capacitances, or inductances) within a given noise level [14]. The second case to be treated here is the influence of noise in the intermittent behavior related to a periodic window existing for a parameter range where chaotic behavior is dominant.

The rest of the paper is organized as follows. Section 2 introduces the theoretical model, as well as the basic concepts to be used in the discussion of synchronization. Section 3 considers the case of intermittent transition to synchronization in two identical oscillators with unidirectional coupling and the presence of noise. Section 4 is devoted to the case of intermittent transition to chaos in the vicinity of a periodic window, for two nonidentical oscillators with a bidirectional coupling. Our conclusions are left to the last section.

2. Coupled Oscillators and Their Synchronization

The triode circuit is a standard textbook example of the Van der Pol equation [12, 15]. However, a more realistic description of such a circuit does not lead us to the Van der Pol but instead to a more general class of Liénard-type equations:

$$\ddot{x} + \mu\left(x^2 + \sigma x - 1\right)\dot{x} + \omega_0^2 x = \rho \sin(\omega t), \tag{2.1}$$

where ω_0 is the frequency of the unforced oscillations; ρ, ω, are, respectively, the amplitude and frequency of an external AC-voltage, and μ, σ are coefficients whose values are drawn from the triode characteristic curve, supposed a third-order polynomial in the grid voltage,

whose normalized form is represented by the variable x [15]. The particular case $\sigma = 0$ yields the usual AC-driven Van der Pol equation.

Moreover, we can rewrite (2.1) in the form

$$\ddot{x} + \mu(x - w_1)(x - w_2)\dot{x} + \omega_0^2 x = \rho \sin(\omega t), \tag{2.2}$$

where

$$w_{1,2} \equiv -\frac{\sigma}{2} \pm \left[\left(\frac{\sigma}{2}\right)^2 + 1\right]^{1/2} \tag{2.3}$$

are the roots of the quadratic resistance, usually of opposite signs.

We consider two such circuits, of which only one is driven by an AC-voltage, and they are supposed to be almost identical, except for their natural frequencies $b_1 = \omega_{01}^2$ and $b_2 = \omega_{02}^2$, which will be considered as variable parameters of the coupled system, whose equations are

$$\dot{x}_1 = x_2, \tag{2.4}$$

$$\dot{x}_2 = -\mu(x_1 - w_1)(x_1 - w_2)x_2 - b_1 x_1 + \rho \sin(\omega t) + c_1(x_3 - x_1), \tag{2.5}$$

$$\dot{x}_3 = x_4, \tag{2.6}$$

$$\dot{x}_4 = -\mu(x_3 - w_1)(x_3 - w_2)x_4 - b_2 x_3 + c_2(x_1 - x_3), \tag{2.7}$$

where the pairs $x_{1,2}$ and $x_{3,4}$ stand for each coupled circuit. We have assumed an asymmetric diffusive coupling whose strengths, c_1 and c_2, take on different values. If $c_1 = 0$, we have a unidimensional coupling, or master-slave configuration, whereas for $c_1 > 0$ (but small) and $c_2 > 0$, the coupling is bidirectional, although it is strongly asymmetric.

The rationale for using such coupling schemes lies in the modeling of the interaction between the heart pacemakers, the sino-atrial (SA), and atrio-ventricular (AV) nodes [13]. The SA node is the primary pacemaker of the heart, and the electrical impulse it generates spreads out through the myocardium, reaching the AV node. Hence, the coupling should be taken either unidirectional or bidirectional but strongly asymmetric. In both cases we use to call the (x_1, x_2) oscillator the driving one, whereas (x_3, x_4) the response oscillator, corresponding to the SA and AV nodes, respectively.

We will use throughout this work the following values for the system parameters: $\mu = -1.45$, $w_1 = -0.2$, $w_2 = 1.9$, $\rho = 0.95$, and $\omega = 1.0$, letting the normal mode frequencies ($b_{1,2}$) and the coupling constants ($c_{1,2}$) to be the parameters to be varied. We have integrated numerically the coupled system of first-order differential equations (2.4)–(2.7) by using a predictor-corrector routine based on the Adams method [16].

For all the cases studied in this paper the system asymptotic behavior will consist of a periodic or chaotic orbit which encircles the points $(x_1, x_2) = (0, 0)$ and $(x_3, x_4) = (0, 0)$ for the driving and response oscillator, respectively. In such cases we can define geometrical phases as

$$\phi_1(t) = \arctan\left[\frac{x_2(t)}{x_1(t)}\right],$$

$$\phi_2(t) = \arctan\left[\frac{x_4(t)}{x_3(t)}\right]. \tag{2.8}$$

In cases, however, for which we do not have an orbit encircling a point, as in funnel attractor, for example, the phase should be defined using other methods, like Poincaré sections or Hilbert transforms [17].

The winding number, defined as

$$\Omega_i = \langle \dot{\phi}_i(t) \rangle_T, \quad (i = 1, 2), \tag{2.9}$$

is the average time rate of the phases of both oscillators and stands for their endogenous frequencies. A weak type of synchronization between the driving and response oscillator consists in the equality of their winding numbers, $\Omega_1(t) = \Omega_2(t)$, regardless of the actual value each phase takes on, and called frequency synchronization (FS) [18]. When the phases themselves are equal, we speak of phase synchronization (PS): $\phi_1(t) = \phi_2(t)$ [19].

By way of contrast, the strongest type of synchronization is complete synchronization (CS), for which the positions and velocities themselves (and not only the phases) are equal: $x_1(t) = x_3(t), x_2(t) = x_4(t)$ [4]. However, CS occurs only if the coupled oscillators are identical [17]. If the coupled oscillators have slightly distinct parameters and the coupling is not too strong, it is possible to find a weaker effect called lag synchronization (LS), defined as the approximate equality of the state variables, delayed by a given time lag τ: $x_1(t) \approx x_3(t - \tau)$, $x_2(t) \approx x_4(t - \tau)$ [19].

Finally, if the oscillator parameters are widely different, as in the case we are investigating here, there is no longer LS because the oscillator positions and velocities differ by a large amount [13]. Even in this case, however, it is still possible to find generalized synchronization (GS), which is observed when there exists a functional relationship between the amplitudes of the two coupled oscillators: $x_2(t) = F(x_1(t))$ and can occur for nonidentical systems [20]. An even weaker effect is generalized lag synchronization (GLS), when the functional relationship between the variables holds up to a time delay τ: $x_2(t) = F(x_1(t - \tau))$.

We will introduce extrinsic noise on the driving oscillator, by adding to (2.5) a term σR_n, where σ is the noise level, and R_n is a pseudorandom variable, with values uniformly distributed in the interval $[-1/2, +1/2]$ and applied at each period of external force. The noisy term is applied at each integration step, so that it plays the role of a stochastic perturbation of a deterministic system rather than a stochastic differential equation, which would need specific integration techniques to be numerically solved [21].

We have considered also this type of noisy term applied on the response oscillator, when both systems are identical, but the results do not differ appreciably from the case we consider here. On the other hand, the inclusion parametric noise (i.e., noise terms applied to system parameters like normal mode frequencies or coupling constants) leads to qualitatively different phenomena, like GLS states, and which we have considered in a recent paper [22].

3. Intermittent Synchronization in Unidirectional Coupling

We initiate our analysis by the case of unidirectional coupling ($c_1 = 0$), for which the response oscillator ($x_3 - x_4$) is slaved under the driving of the master oscillator ($x_1 - x_2$). Moreover, we will consider both oscillators as being identical, in the sense that $b_1 = b_2 = 1.0$, that is, their normal mode frequencies take on exactly the same value. As a consequence, a CS state $x_1 = x_3, x_2 = x_4$ is possible and defines a synchronization manifold S in the phase space. If S is transversely asymptotically stable, there follows that small displacements along directions transversal to S will shrink down to zero as time tends to infinity. The conditional
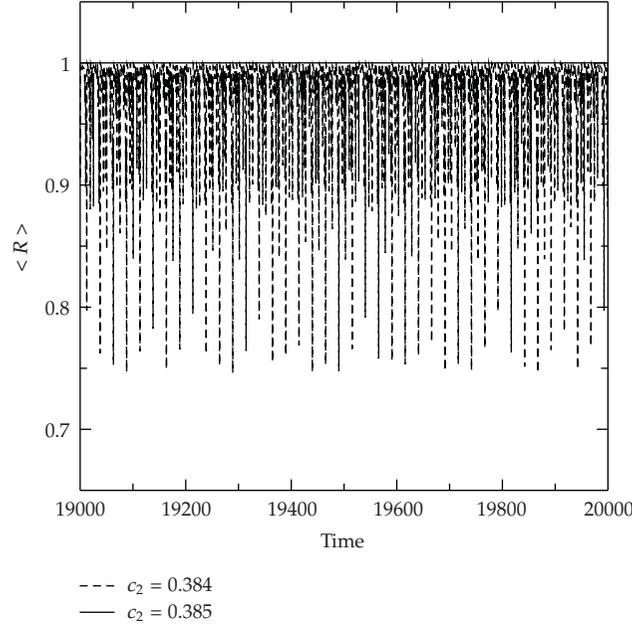
Figure 1: Order parameter for two values of c_2 at the neighborhood of the transition to a synchronized state. The time in the horizontal axis is measured in arbitrary units.

Lyapunov exponents λ_{cond} for the response oscillator are computed under the constraint that the trajectory lies entirely in S for all times. The negativity of all conditional exponents is a necessary condition for the CS state to be transversely asymptotically stable [17].

A useful numerical diagnostic of CS is the order parameter

$$R(t) \equiv \frac{1}{2}\left|e^{2\pi i x_2(t)} + e^{2\pi i x_4(t)}\right|, \tag{3.1}$$

which measures the degree of coherence between oscillators. For completely incoherent evolution R approaches zero, whereas for CS states R goes to the unity. Oscillations of $R(t)$ around values less than unity indicate partial coherence and no synchronization at all. Figure 1 presents the time evolution of the order parameter for two values of c_2, before and after a transition to CS occurring at $c_2 = c_2^{CR} \approx 0.3845$. For $c_2 \lesssim c_2^{CR}$ the order parameter oscillates with an average below unity, whereas for $c_2 \gtrsim c_2^{CR}$ it settles down at unity without noticeable fluctuations.

The approach to the CS state is characterized by the appearance of weaker forms of synchronization, like PS, and their breakdown. This is a particularly interesting point to investigate the role of noise on synchronization properties of our model. The temporal evolution of the phase difference $\phi_1(t) - \phi_2(t)$ between the coupled oscillators is plotted in Figure 2, starting from an arbitrarily chosen time. For $c_2 \gtrsim c_2^{CR}$ (thick dashed line) the phase difference is zero, which is an obvious consequence of the CS state. On the other hand, for $c_2 \lesssim c_2^{CR}$ (thin dashed line), the phases increase monotonically with time as well as their
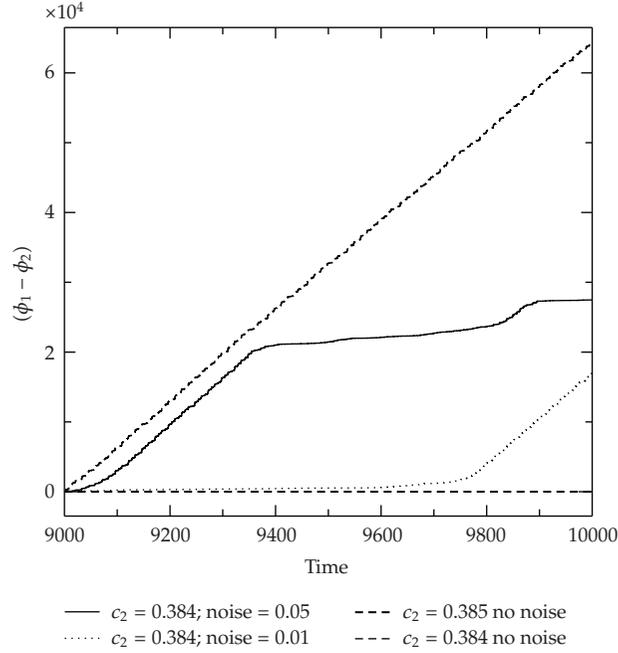
Figure 2: Time evolution of the phase difference between the coupled oscillator with and without noise, for two values of c_2 at the neighborhood of the transition to a synchronized state.

difference, due to absence of synchronization. This phase diffusion can be described by a Gaussian dependence with time

$$\langle \phi_1(t) - \phi_2(t) \rangle \approx 2Dt, \tag{3.2}$$

where D is the diffusion coefficient, and $\langle \cdots \rangle$ stands for an average over many randomly chosen initial condition. From Figure 2 we estimate that $D \sim 30$.

If we add a noise level $\sigma = 0.05$ to the precritical case (solid thin line), the phases initially diffuse but suddenly the phase difference stays for some time at a nearly constant value, indicating a transient PS state which breaks down and yields phase diffusion again, and so on. These phase slips, as they are usually called, are characteristic of the onset of desynchronization [23]. The average duration of laminar PS states increases as the noise level is diminished (see the dashed line for $\sigma = 0.01$ in Figure 2).

In the absence of noise, we observed that the oscillators present FS, characterized by $\Omega_1 - \Omega_2 = 0$, irrespective of exhibiting a CS state (Figure 3(a)). In such case, even though the dynamics for each oscillator is chaotic, their phases are correlated enough to warrant the equality of their time rates. Hence FS can be considered more robust than CS or even PS in our system. This is even more evident when noise is added to the system (Figure 3(b)). The FS state survives with small spikes for the interval of c_2-values studied, except for a narrow interval centered at 0.5 for which there is a (still low) peak of winding number mismatch of $\sim 10^{-3}$.

No noise

(a)

Noise = 0.1

(b)
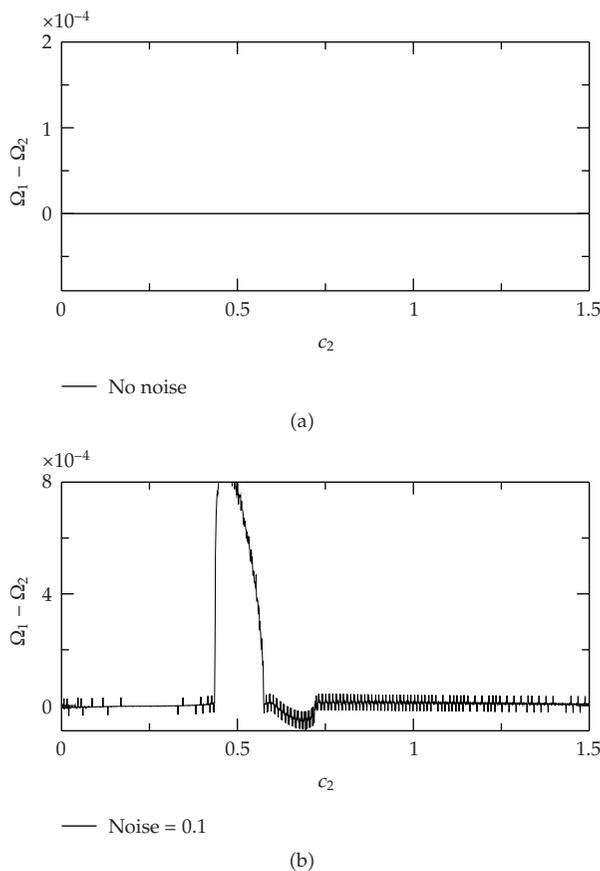
Figure 3: Winding number difference between the coupled oscillator with and without noise, as a function of the coupling strength.

Another dynamical feature observed in the neighborhood of the synchronization transition at $c_2 = c_2^{CR} \approx 0.3845$ is the presence of intermittent bursts of nonsynchronized behavior intercalated with laminar intervals of synchronized behavior. These features are characteristic of the so-called *on-off intermittency*, where there is an invariant manifold on which the system would lie during laminar intervals of duration τ_i, the burst representing excursions off this manifold [24]. On-off intermittency has been observed in coupled systems, like coupled map lattices, whose invariant manifold is the synchronization manifold [25, 26].

On-off intermittency has a numerical signature, which is the scaling obeyed by the statistical distribution $P(\tau)$ of the durations of laminar (or interburst) intervals of synchronization: $P(\tau) \sim \tau^{-3/2}$. The presence of noise in on-off intermittency scenarios introduces an additional tail of exponential dependence [27]. We actually observed this universal scaling for values slightly after and before the transition at $c_2 = c_2^{CR}$, with nonzero noise levels. Our results are shown in Figures 4(a) and 4(b) for $c_2 \lesssim c_2^{CR}$ and $c_2 \gtrsim c_2^{CR}$, respectively, with a 1% noise level, where the diamonds stand for numerical data, with two different scaling regions: one power-law scaling for small interburst intervals, with exponent $-3/2$ within the numerical accuracy, and an exponential tail for large interburst intervals.

(a)

(b)

(c)

(d)
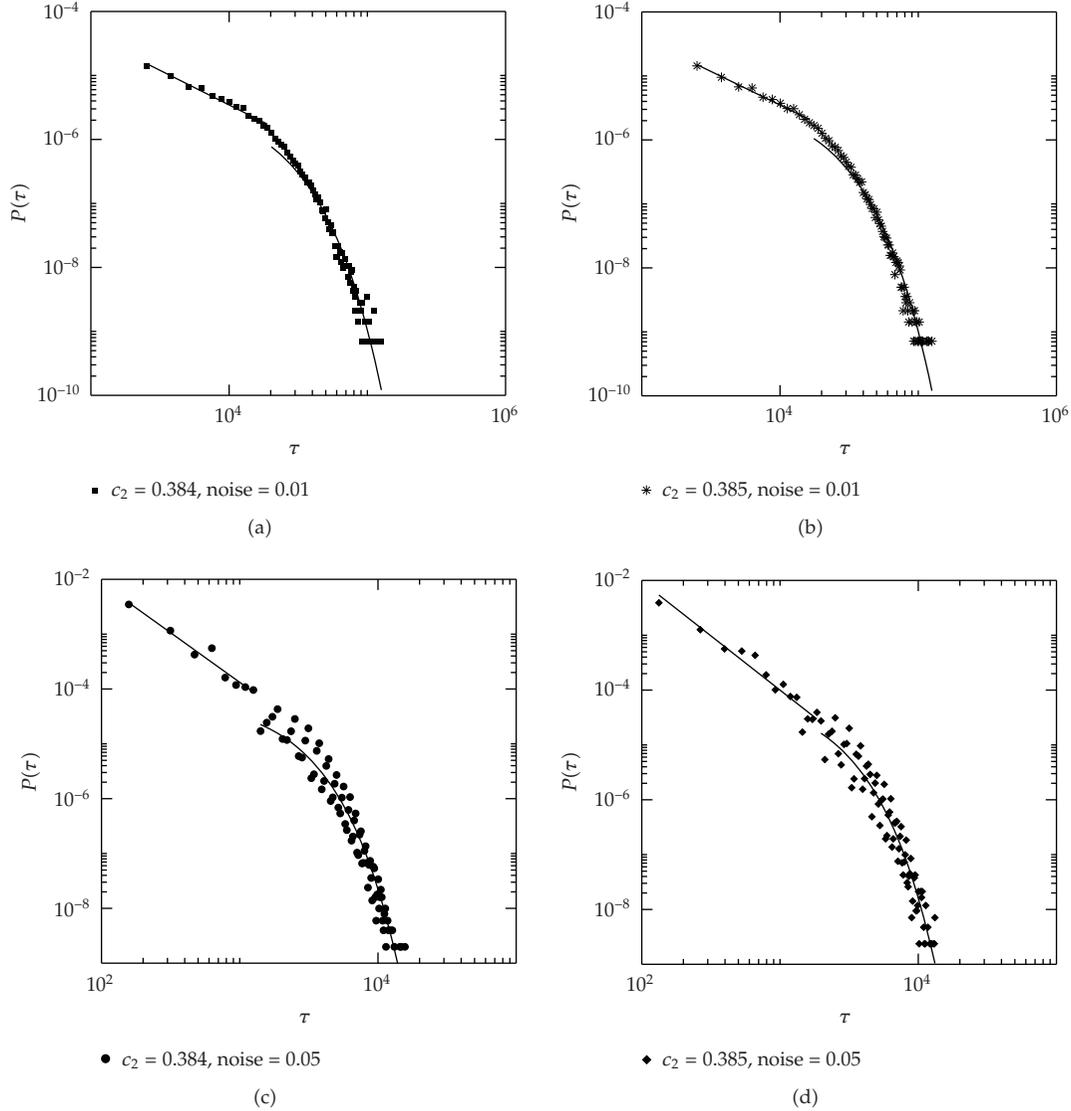
Figure 4: Probability distribution for the duration of laminar synchronized intervals between consecutive bursts of nonsynchronized behavior, for (a) $c_2 = 0.384 \gtrsim c_2^{CR} c_2 = 0.384 \gtrsim c_2^{CR}$, and (b) $c_2 = 0.385 \gtrsim c_2^{CR} c_2 = 0.385 \gtrsim c_2^{CR}$ with a 1%1% noise level. (c) and (d) correspond to (a) and (b), respectively, with a 5%5% noise level. The lines shown are least-squares fits evidencing two different scalings for numerical data, and the symbols stand for the numerical results.

The two scalings are roughly separated by a shoulder which, according to the general theory of noisy on-off intermittency, defines a crossover time whose value depends on the noise level [27]. We verified this point by considering, in Figures 4(c) and 4(d), the cases before and after the synchronization transition, respectively, but with a higher (5%) noise level. We observe that the crossover time, which is *circa* 10^4 for weak noise, decreases to 10^3 for stronger noise, diminishing the noiseless power-law scaling, as expected.
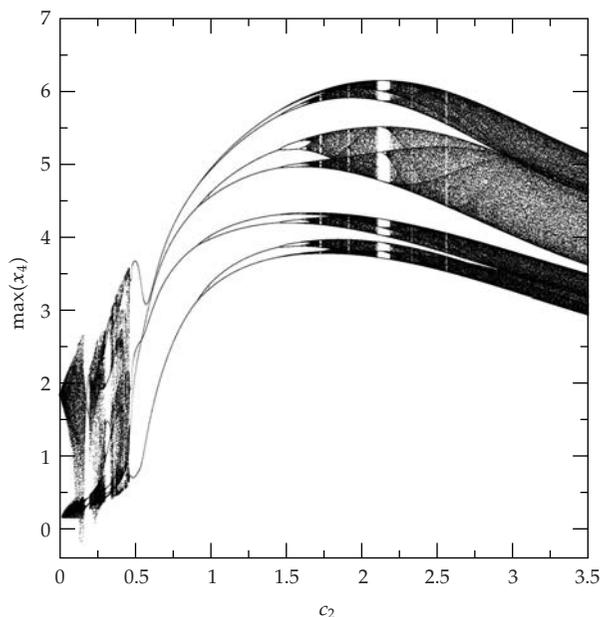
Figure 5: Bifurcation diagram for $x_n = \max(x_4)$, the local maxima of $x_3(t)$ versus c_2.

4. Intermittent Chaotic Bursts in Bidirectional Coupling

Now we present results considering the case where $c_1 \neq 0$ and $c_2 \neq 0$, what characterizes a bidirectional form of coupling. We choose low values of $c_1 \gtrsim 0.01$, because of the strong asymmetrical character of the bidirectional coupling and kept c_2 in the $[2.0, 2.2]$ interval. In this section, also, we relax the hypothesis we previously made of identical oscillators and allow for a mismatch of the normal mode frequencies: $b_1 = 1.0$, and $b_2 = 0.66$, the other parameter values being the same as those used in the previous section.

The dependence of $x_n = \max(x_4)$, which are the local maxima of $x_4(t)$, is plotted in Figure 5 as a function of c_2. We can distinguish two different parts in this diagram: (i) a quasiperiodic region for $0 < c_2 < c_{II} = 0.45507$; (ii) a period-doubling cascade for $c_2 > c_{II}$. In the latter, we focus on the narrow period-12 window starting at $c_2 = c_I = 2.101503$. We will consider the effects of noise in the vicinity of the critical points c_I and c_{II}, since they are very sensitive with respect to variations in the coupling strength and are related with sudden changes in the dynamical behavior of the system.

In the neighborhood of the period-12 window, that is, at $c_2 = c_I$, the four-band chaotic attractor suddenly disappears, and a saddle-node bifurcation occurs forming a stable and an unstable orbit of period-12. This can be explicitly verified in Figures 6(a) and 6(b), where we plot the 12th return map for the variable $x_n = \max(x_4)$, respectively, just before and just after the saddle-node bifurcation. For $c_2 = 2.101504 \gtrsim c_I$ (Figure 6(a)) the periodic points are the intersections of the return map with the 45° line, corresponding to the regular time series in Figure 6(c). When $c_2 = 2.101470 \lesssim c_I$ (Figure 6(b)) there are no intersections with the 45° line, but since we are too close to the bifurcation point, there exists a narrow bottleneck between the curves. When the trajectory enters such bottlenecks, the resulting behavior is the laminar interval, followed by chaotic bursts when the trajectory eventually exits the nozzle and is randomly reinjected to its vicinity (Figure 6(d)).

(a)

(b)

(c)

(d)
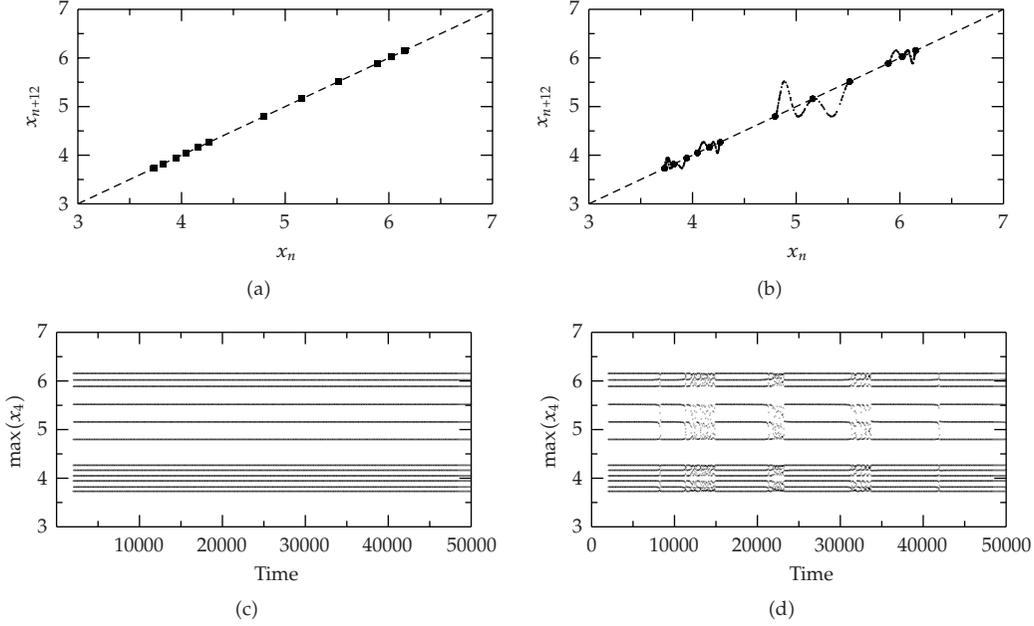
Figure 6: Return maps for $x_{n+12}x_{n+12}$ versus $x_n x_n$ for (a) $c_2 = 2.101504 \gtrsim c_I c_2 = 2.101504 \gtrsim c_I$ and (b) $c_2 = 2.101470 \lesssim c_I c_2 = 2.101470 \lesssim c_I$. (c) and (d) are the corresponding time series for the cases (a) and (b), respectively.

Let $\langle \tau_{\text{int}} \rangle$ be the average duration of the laminar intervals between consecutive bursts. If we approach the critical value c_I from below, this average duration increases according to a power-law: $\langle \tau_{\text{int}} \rangle \sim (c_2 - c_I)^{-\varpi}$ (Figure 7(a)), where the scaling exponent $\varpi = 1/2$ within the numerical precision. This agrees with the Pomeau-Manneville type-I intermittency scenario and shows that the essential dynamics leading to intermittent behavior near c_I is one-dimensional [1].

We can now investigate the role of a noise level on this average duration of laminar intervals. Figure 7(b) plots $\langle \tau \rangle$ versus $c_2 - c_I$ for various noise levels. If we are far enough from the critical value c_I, the scaling is essentially the same as in the noiseless case. If the noise level is too small—for example, 2×10^{-10}, represented by circles in Figure 7(b)—the scaling is practically unchanged even very close to the critical point. However, for higher noise levels, the scaling holds only to a certain minimum distance, between 10^{-3} and 10^{-4}, below which the value of $\langle \tau_{\text{int}} \rangle$ is barely affected by the distance $c_2 - c_I$. This maximum distance increases with the noise level [2].

In the context of the heartbeat model described in [13], intermittency can be a highly undesirable feature, since the irregular alternations between laminar and chaotic states make it difficult to either control or suppress pathological rhythms, like those characterizing cardiac arrhythmias [6]. If one tries to control such rhythms, for example, using OGY chaos control, it is required that the system attractor be chaotic enough so as to present recurrence of trajectories, a necessary condition to control the system by applying only small perturbation [28]. As an example of the applicability of our results, let us imagine that an experimentally obtained data series shows an intermittent alternation between regular and arrhythmic heartbeats (this can be done by usual electrocardiographic recording).

(a)

(b)
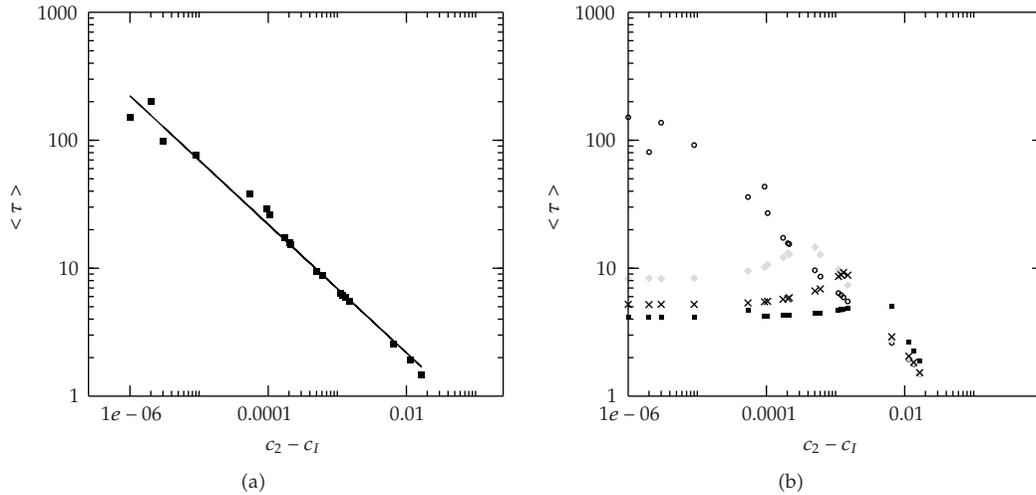
Figure 7: Average duration of interburst laminar intervals versus the difference $c_2 - c_I$, where $c_I = 2.101503$ marks the onset of a period-12 window. (a) Noiseless case. The solid line is a least squares fit with slope 0.5. (b) The effect of various noise levels: 2.0×10^{-10} (empty circles); 5.0×10^{-10} (vertical crosses); 8.0×10^{-10} (inclined crosses); 2.0×10^{-9} (filled squares).

It is possible, at least in principle, to record the durations of the laminar (interburst) intervals and make a histogram of the laminar times $\overline{P}(\tau_{\text{int}})$, which can be then used to compute an average duration $\langle \overline{\tau}_{\text{int}} \rangle$. Let us also suppose that one has a control parameter p that can be varied over an interval. In an experimental setting this control parameter could be, for example, an external factor like the amount of some chemical influencing arrhythmic behavior. For example, cardiac arrhythmias can be induced by the drug ouabain in rabbit ventricle [29]. Finally, if we were able to record the critical dose p_C for which laminar regions disappear at all and intermittent behavior becomes stable chaos, the characterization of $\langle \overline{\tau}_{\text{int}} \rangle$ as a function of the parameter difference $|p - p_c|$ could provide an empirical verification of the results we obtained from an analytical model. Moreover, if the scaling law for the numerically determined probability distribution $\overline{P}(\tau_{\text{int}})$ was found to be a power-law with exponent $3/2$, this would be a strong evidence in favor of a universal on-off intermittency mechanism that, since an invariant manifold is supposed for this to occur, what could shed some light on the dynamical mechanisms underlying intermittent behavior in this system [24].

5. Conclusions

We have studied the effect of parametric noise in the coupled system of two Liénard-type oscillators with external periodic forcing, focusing on two different intermittent phenomena exhibited by the system under distinct types of coupling. In the unidirectional coupling or master-slave configuration, we have analyzed the occurrence of complete synchronization of identical oscillators and have determined the necessary coupling strength for a transition from a nonsynchronized to a synchronized state. Near this transition there is an intermittent switching between laminar phases of synchronized (albeit chaotic) behavior and bursts of

nonsynchronized dynamics. We verified the universal 3/2 power-law scaling, obeyed by on-off intermittency, for the statistical distribution of the duration of the synchronized laminar phases. The inclusion of noise modifies this scaling according to the general description by adding an exponential tail (for large times) to the power-law scaling (for short times).

We also verified the presence of other types of synchronization, like phase and frequency synchronization, and observed that the latter is robust in the sense that it is not likely to cease with addition of white noise. In the bidirectional coupling of nonidentical oscillators (because of a mismatch of their natural frequencies), we no longer have synchronization, and the intermittent phenomenon of interest is the transition to chaos in the beginning of a periodic window for a parameter range where chaos is the dominant feature. We verified that this transition obeys the Pomeau-Manneville type-I intermittency scenario, by considering the statistical properties of the average laminar durations as well as evidencing the saddle-node bifurcation which is the mechanism underlying the phenomenon. The addition of noise affects these properties in the way predicted for one-dimensional maps. Finally, the results of this paper can be applied to a number of physical systems described by Liénard-type oscillators. Two representative examples are electronic circuits using tunnel diodes, like Zener diode, and models of the heartbeat. The statistical nature of our numerical results makes them amenable to further comparisons with experimental investigations of intermittent behavior.

Acknowledgments

References

[1] Y. Pomeau and P. Manneville, "Intermittent transition to turbulence in dissipative dynamical systems," *Communications in Mathematical Physics*, vol. 74, no. 2, pp. 189–197, 1980.

[2] J. E. Hirsch, B. A. Huberman, and D. J. Scalapino, "Theory of intermittency," *Physical Review A*, vol. 25, no. 1, pp. 519–532, 1982.

[3] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Science*, vol. 12 of *Cambridge Nonlinear Science Series*, Cambridge University Press, Cambridge, UK, 2001.

[4] J. F. Heagy, T. L. Carroll, and L. M. Pecora, "Synchronous chaos in coupled oscillator systems," *Physical Review E*, vol. 50, no. 3, pp. 1874–1885, 1994.

[5] R. Roy and K. S. Thornburg Jr., "Experimental synchronization of chaotic lasers," *Physical Review Letters*, vol. 72, no. 13, pp. 2009–2012, 1994.

[6] L. Glass, "Synchronization and rhythmic processes in physiology," *Nature*, vol. 410, no. 6825, pp. 277–284, 2001.

[7] J. Guckenheimer, "Dynamics of the van der Pol equation," *IEEE Transactions on Circuits and Systems*, vol. 27, no. 11, pp. 983–989, 1980.

[8] B. van der Pol, "Forced oscillationsin a circuit with non-linear resistance," *Philosophical Magazine*, vol. 3, pp. 65–80, 1927.

[9] J. C. Biswas and V. Mitra, "Negative resistance in Zener diodes," *Solid-State Electronics*, vol. 21, no. 8, pp. 1053–1055, 1978.

[10] B. van der Pol and J. van der Mark, "The heartbeat considered as a relaxation oscillation, and an electrical model of the heart," *Philosophical Magazine*, vol. 6, pp. 763–775, 1928.

[11] O. Kongas, R. von Hertzen, and J. Engelbrecht, "Bifurcation structure of a periodically driven nerve pulse equation modelling cardiac conduction," *Chaos, Solitons & Fractals*, vol. 10, no. 1, pp. 119–136, 1999.

[12] M. Kontorovich, *Nonlinear Oscillations in Radio Engineering*, Mir, Moscow, Russia, 1976.

[13] A. M. dos Santos, S. R. Lopes, and R. L. Viana, "Rhythm synchronization and chaotic modulation of coupled van der Pol oscillators in a model for the heartbeat," *Physica A*, vol. 338, no. 3-4, pp. 335–355, 2004.

[14] T. Kapitaniak, *Chaos in Systems with Noise*, World Scientific, Singapore, 2nd edition, 1990.

[15] P. Hagedorn, *Nichtlineare Schwingungen*, Akademiscbe Verlagsellschaft, Wiesbaden, Germany, 1978.

[16] A. C. Hindmarch, "ODEPACK: a systematized collection of ODE solvers," in *Scientific Computing*, R. S. Stepleman, M. Carver, R. Peskin, W. F. Ames, and R. Vichnevetsky, Eds., vol. 1 of *IMACS Transactions on Scientific Computation*, pp. 55–64, North-Holland, Amsterdam, The Netherlands, 1983.

[17] S. Boccaletti, J. Kurths, G. Osipov, D. L. Valladares, and C. S. Zhou, "The synchronization of chaotic systems," *Physics Report*, vol. 366, no. 1-2, pp. 1–101, 2002.

[18] P. S. Landa, *Regular and Chaotic Oscillations*, Foundations of Engineering Mechanics, Springer, Berlin, Germany, 2001.

[19] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "From phase to lag synchronization in coupled chaotic oscillators," *Physical Review Letters*, vol. 78, no. 22, pp. 4193–4196, 1997.

[20] S. Taherion and Y.-C. Lai, "Observability of lag synchronization of coupled chaotic oscillators," *Physical Review E*, vol. 59, no. 6, pp. R6247–R6250, 1999.

[21] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin, Germany, 2000.

[22] A. M. dos Santos, S. R. Lopes, and R. L. Viana, "Synchronization regimes for two coupled noisy Liénard-type driven oscillators," *Chaos, Solitons & Fractals*, vol. 36, no. 4, pp. 901–910, 2008.

[23] Z. Zheng, G. Hu, and B. Hu, "Phase slips and phase synchronization of coupled oscillators," *Physical Review Letters*, vol. 81, no. 24, pp. 5318–5321, 1998.

[24] N. Platt, E. A. Spiegel, and C. Tresser, "On-off intermittency: a mechanism for bursting," *Physical Review Letters*, vol. 70, no. 3, pp. 279–282, 1993.

[25] R. L. Viana, C. Grebogi, S. E. de S. Pinto, S. R. Lopes, A. M. Batista, and J. Kurths, "Validity of numerical trajectories in the synchronization transition of complex systems," *Physical Review E*, vol. 68, no. 6, Article ID 067204, 4 pages, 2003.

[26] R. L. Viana, C. Grebogi, S. E. de S.pinto, S. R. Lopes, A. M. Batista, and J. Kurths, "Bubbling bifurcation: loss of synchronization and shadowing breakdown in complex systems," *Physica D*, vol. 206, no. 1-2, pp. 94–108, 2005.

[27] P. W. Hammer, N. Platt, S. M. Hammel, J. F. Heagy, and B. D. Lee, "Experimental observation of on-off intermittency," *Physical Review Letters*, vol. 73, no. 8, pp. 1095–1098, 1994.

[28] E. Ott, C. Grebogi, and J. A. Yorke, "Controlling chaos," *Physical Review Letters*, vol. 64, no. 11, pp. 1196–1199, 1990.

[29] A. Garfinkel, M. L. Spano, W. L. Ditto, and J. N. Weiss, "Controlling cardiac chaos," *Science*, vol. 257, no. 5074, pp. 1230–1235, 1992.

Research Article

Chaos Synchronization between Two Different Fractional Systems of Lorenz Family

A. E. Matouk[1, 2]

[1] *Mathematics Department, Faculty of Science, Mansoura University, Mansoura 35516, Egypt*
[2] *Mathematics Department, Faculty of Science, Hail University, Hail 2440, Saudi Arabia*

Correspondence should be addressed to A. E. Matouk, aematouk@hotmail.com

This work investigates chaos synchronization between two different fractional order chaotic systems of Lorenz family. The fractional order Lü system is controlled to be the fractional order Chen system, and the fractional order Chen system is controlled to be the fractional order Lorenz-like system. The analytical conditions for the synchronization of these pairs of different fractional order chaotic systems are derived by utilizing Laplace transform. Numerical simulations are used to verify the theoretical analysis using different values of the fractional order parameter.

1. Introduction

Fractional calculus has been known since the early 17th century [1, 2]. It has useful applications in many fields of science like physics [3], engineering [4], mathematical biology [5, 6], and finance [7, 8].

The fractional order derivatives have many definitions; one of them is the Riemann-Liouville definition [9] which is given by

$$D^\alpha f(t) = \frac{d^l}{dt^l} J^{l-\alpha} f(t), \quad \alpha > 0, \tag{1.1}$$

where J^θ is the θ-order Riemann-Liouville integral operator which is given as

$$J^\theta u(t) = \frac{1}{\Gamma(\theta)} \int_0^t (t - \tau)^{\theta-1} u(\tau) d\tau, \qquad \theta > 0. \tag{1.2}$$

However, the most common definition is the Caputo definition [10], since it is widely used in real applications:

$$D_*^\alpha f(t) = J^{l-\alpha} f^{(l)}(t), \tag{1.3}$$

where $f^{(l)}$ represents the l-order derivative of $f(t)$ and $l = [\alpha]$; this means that l is the first integer which is not less than α. The operator D_*^α is called "the Caputo differential operator of order α." Hence, I choose the Caputo type throughout this paper.

On the other hand, chaos has been studied and developed with much interest by scientists since the birth of Lorenz chaotic attractor in 1963 [11]. Chen attractor is similar to Lorenz attractor but not topologically equivalent [12]. Recently, Lü et al. found a new chaotic system which connects the Lorenz and Chen attractors, according to the conditions formulated by Vaněček and Čelikovský, and it is called Lü system [13]. Afterwards, chaos in fractional order dynamical systems has become an interesting topic. In [14] chaotic behaviors of the fractional order Lorenz system are studied. Moreover, chaotic behaviors have also been found in the fractional order Chen system [15] and the fractional order Lü system [16]. Furthermore, Chaos synchronization in fractional order chaotic systems starts to attract increasing attention [16–20]. However, it has been studied very well in the case of integer order chaotic systems, due to its potential applications in physical, chemical, and biological systems [21–24] and secure communications [25].

The generalized synchronization between two different fractional order systems is investigated in [26]. However, in this paper, I investigate the conditions of chaos synchronization between two different fractional order chaotic systems of Lorenz family by designing suitable linear controllers. I give examples to achieve chaos synchronization of two pairs of different fractional order chaotic systems (fractional Chen & fractional Lü, fractional Lorenz-like, and fractional Chen) in drive-response structure. Conditions for achieving chaos synchronization using linear control method are further discussed using Laplace transform theory.

2. Systems Description

The fractional order Chen system isgiven as follows:

$$\frac{d^\alpha x}{dt^\alpha} = a(y - x), \qquad \frac{d^\alpha y}{dt^\alpha} = (c - a)x - xz + cy, \qquad \frac{d^\alpha z}{dt^\alpha} = xy - bz. \tag{2.1}$$

Here and throughout, $(a, b, c) = (35, 3, 28)$ where α is the fractional order. In the following I choose $\alpha = 0.9$ at which system (2.1) exhibits chaotic attractor (see Figure 1).

The fractional order Lü system is given as follows

$$\frac{d^\alpha x}{dt^\alpha} = r(y - x), \qquad \frac{d^\alpha y}{dt^\alpha} = -xz + py, \qquad \frac{d^\alpha z}{dt^\alpha} = xy - qz. \tag{2.2}$$

Here and throughout, $(r, p, q) = (35, 28, 3)$. By choosing $\alpha = 0.9$, system (2.2) has chaotic attractor (see Figure 2).
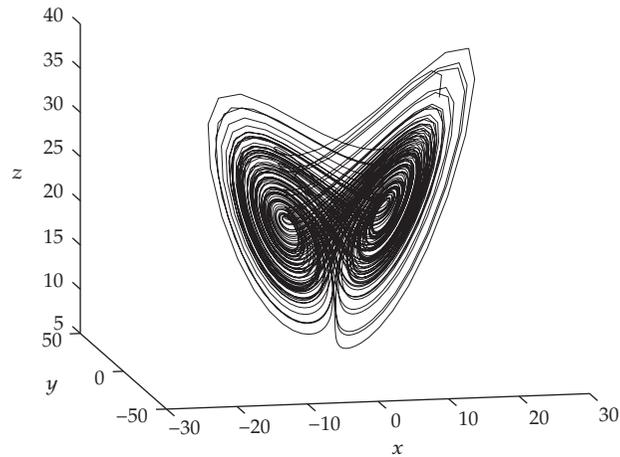
Figure 1: Chaotic attractor of the fractional order Chen system (2.1) with $\alpha = 0.9$ and $(a, b, c) = (35, 3, 28)$.
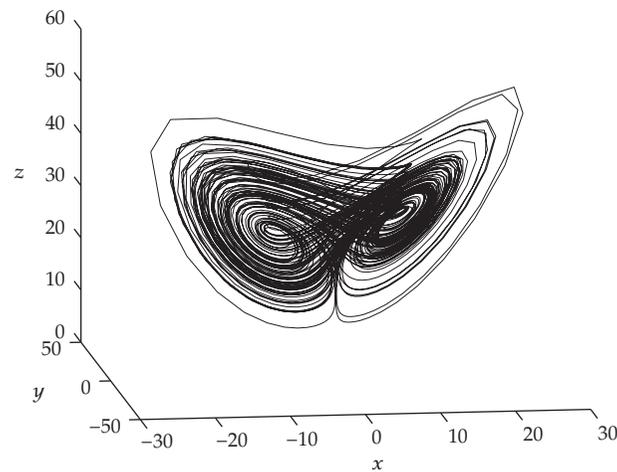

Figure 2: Chaotic attractor of the fractional order Lü system (2.2) with $\alpha = 0.9$ and $(r, p, q) = (35, 28, 3)$.

The fractional order Lorenz-like system [27] is described by

$$\frac{d^\alpha x}{dt^\alpha} = \sigma(y - x), \qquad \frac{d^\alpha y}{dt^\alpha} = \rho x - xz + \gamma y, \qquad \frac{d^\alpha z}{dt^\alpha} = xy - \beta z, \tag{2.3}$$

which has a chaotic attractor as shown in Figure 3 when $\beta = 2.8$, $\gamma = 10.6$, $\rho = 14$, $\sigma = 20$, and $\alpha = 0.9$.

It should be also noted that, the systems (2.1), (2.2), and (2.3) are still chaotic at the fractional order values $\alpha = 0.95$ and $\alpha = 0.99$.
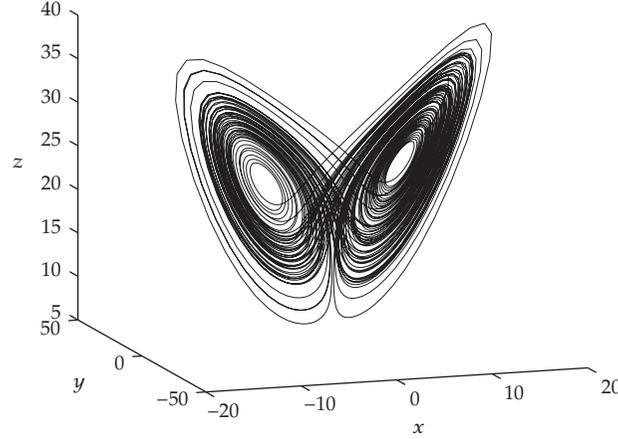
Figure 3: Chaotic attractor of the fractional order Lorenz-like system (2.3) with $\alpha = 0.9$ and $(\beta, \gamma, \rho, \sigma) = (2.8, 10.6, 14, 20)$.

3. Synchronization between Two Different Fractional Order Systems

Consider the master-slave (or drive-response) synchronization scheme of two autonomous different fractional order chaotic systems:

$$\frac{d^\alpha X}{dt^\alpha} = f(X), \qquad \frac{d^\alpha Y}{dt^\alpha} = g(Y) + U(t), \tag{3.1}$$

where α is the fractional order, $X \in R^n$, $Y \in R^n$ represent the states of the drive and response systems, respectively, $f : R^n \rightarrow R^n$, $g : R^n \rightarrow R^n$ are the vector fields of the drive and response systems, respectively. The aim is to choose a suitable linear control function $U(t) = (u_1, \ldots, u_n)^T$ such that the states of the drive and response systems are synchronized (i.e., $\lim_{t \rightarrow \infty} \|X - Y\| = 0$, where $\| \cdot \|$ is the Euclidean norm).

3.1. Synchronization between Chen and Lü Fractional Order Systems

In this subsection, the goal is to achieve chaos synchronization between the fractional order Chen system and the fractional order Lü system by using the fractional order Chen system to drive the fractional order Lü system. The drive and response systems are given as follows:

$$\frac{d^\alpha x_m}{dt^\alpha} = a(y_m - x_m), \qquad \frac{d^\alpha y_m}{dt^\alpha} = (c - a)x_m - x_m z_m + c y_m, \qquad \frac{d^\alpha z_m}{dt^\alpha} = x_m y_m - b z_m, \tag{3.2}$$

$$\frac{d^\alpha x_s}{dt^\alpha} = r(y_s - x_s) + u_1, \qquad \frac{d^\alpha y_s}{dt^\alpha} = -x_s z_s + p y_s + u_2, \qquad \frac{d^\alpha z_s}{dt^\alpha} = x_s y_s - q z_s + u_3, \tag{3.3}$$

where u_1, u_2, and u_3 are the linear control functions. Define the error variables as follows:

$$e_1 = x_s - x_m, \qquad e_2 = y_s - y_m, \qquad e_3 = z_s - z_m. \tag{3.4}$$

By subtracting (3.2) from (3.3) and using (3.4), we obtain

$$
\begin{aligned}
\frac{d^\alpha e_1}{dt^\alpha} &= r(e_2 - e_1) + (r - a)(y_m - x_m) + u_1, \\
\frac{d^\alpha e_2}{dt^\alpha} &= pe_2 - z_m e_1 - x_m e_3 - e_1 e_3 - (c - a)x_m + (p - c)y_m + u_2, \\
\frac{d^\alpha e_3}{dt^\alpha} &= -qe_3 + y_m e_1 + x_m e_2 + e_1 e_2 - (q - b)z_m + u_3.
\end{aligned}
\tag{3.5}
$$

Now, by letting

$$
\begin{aligned}
u_1 &= (a - r)(y_m - x_m), \\
u_2 &= (c - a)x_m + (c - p)y_m - k_1(y_s - y_m), \\
u_3 &= (q - b)z_m - k_2(z_s - z_m),
\end{aligned}
\tag{3.6}
$$

where $k_1, k_2 \geq 0$, then the error system (3.5) is reduced to

$$
\begin{aligned}
\frac{d^\alpha e_1}{dt^\alpha} &= r(e_2 - e_1), \\
\frac{d^\alpha e_2}{dt^\alpha} &= (p - k_1)e_2 - z_m e_1 - x_m e_3 - e_1 e_3, \\
\frac{d^\alpha e_3}{dt^\alpha} &= -(q + k_2)e_3 + y_m e_1 + x_m e_2 + e_1 e_2.
\end{aligned}
\tag{3.7}
$$

By taking the Laplace transform in both sides of (3.7), letting $E_i(s) = L\{e_i(t)\}$ where $(i = 1, 2, 3)$, and applying $L\{d^\alpha e_i/dt^\alpha\} = s^\alpha E_i(s) - s^{\alpha-1}e_i(0)$, we obtain

$$
\begin{aligned}
s^\alpha E_1(s) - s^{\alpha-1}e_1(0) &= r(E_2(s) - E_1(s)), \\
s^\alpha E_2(s) - s^{\alpha-1}e_2(0) &= (p - k_1)E_2(s) - L\{x_m e_3\} - L\{z_m e_1\} - E_1(s)E_3(s), \\
s^\alpha E_3(s) - s^{\alpha-1}e_3(0) &= -(q + k_2)E_3(s) + L\{y_m e_1\} + L\{x_m e_2\} + E_1(s)E_2(s).
\end{aligned}
\tag{3.8}
$$

Proposition 3.1. *If $E_1(s)$, $E_2(s)$ are bounded and $p - k_1 \neq 0$, then the drive and response systems (3.2) and (3.3) will be synchronized under a suitable choice of k_1 and k_2.*

Proof. Rewrite (3.8) as follows:

$$
\begin{aligned}
E_1(s) &= \frac{rE_2(s)}{s^\alpha + r} + \frac{s^{\alpha-1}e_1(0)}{s^\alpha + r}, \\
E_2(s) &= -\frac{L\{z_m e_1\}}{s^\alpha - p + k_1} - \frac{L\{x_m e_3\}}{s^\alpha - p + k_1} - \frac{E_1(s)E_3(s)}{s^\alpha - p + k_1} + \frac{s^{\alpha-1}e_2(0)}{s^\alpha - p + k_1}, \\
E_3(s) &= \frac{L\{y_m e_1\}}{s^\alpha + q + k_2} + \frac{L\{x_m e_2\}}{s^\alpha + q + k_2} + \frac{E_1(s)E_2(s)}{s^\alpha + q + k_2} + \frac{s^{\alpha-1}e_3(0)}{s^\alpha + q + k_2}.
\end{aligned}
\tag{3.9}
$$

Using the final value theorem of the Laplace transform, it follows that

$$\lim_{t\to\infty} e_1(t) = \lim_{s\to 0^+} sE_1(s) = \lim_{s\to 0^+} sE_2(s) = \lim_{t\to\infty} e_2(t),$$

$$\lim_{t\to\infty} e_2(t) = \lim_{s\to 0^+} sE_2(s) = \frac{1}{p-k_1}\lim_{s\to 0^+} sL\{x_m e_3\} + \frac{1}{p-k_1}\lim_{s\to 0^+} sL\{z_m e_1\} + \frac{1}{p-k_1}\lim_{t\to\infty} e_1(t)\cdot\lim_{t\to\infty} e_3(t),$$

$$\lim_{t\to\infty} e_3(t) = \lim_{s\to 0^+} sE_3(s) = \frac{1}{q+k_2}\lim_{s\to 0^+} sL\{y_m e_1\} + \frac{1}{q+k_2}\lim_{s\to 0^+} sL\{x_m e_2\} + \frac{1}{q+k_2}\lim_{t\to\infty} e_1(t)\cdot\lim_{t\to\infty} e_2(t).$$

$$(3.10)$$

Since $E_1(s)$, $E_2(s)$ are bounded and $p - k_1 \neq 0$ then $\lim_{t\to\infty} e_1(t) = \lim_{t\to\infty} e_2(t) = 0$. Now, owing to the attractiveness of the attractors of systems (2.1) and (2.2), there exists $\eta > 0$ such that $|x_i(t)| \leq \eta < \infty$, $|y_i(t)| \leq \eta < \infty$, and $|z_i(t)| \leq \eta < \infty$ where i refers to the index of the drive or response variables. Therefore, $\lim_{t\to\infty} e_3(t) = 0$. This implies that

$$\lim_{t\to\infty} e_i(t) = 0, \quad i = 1, 2, 3. \tag{3.11}$$

Consequently, the synchronization between the drive and response systems (3.2) and (3.3) is achieved. □

3.1.1. Numerical Results

An efficient method for solving fractional order differential equations is the predictor-correctors scheme or more precisely, PECE (Predict, Evaluate, Correct, Evaluate) technique which has been investigated in [28, 29], and represents a generalization of the Adams-Bashforth-Moulton algorithm. It is used throughout this paper.

Based on the above mentioned discretization scheme, the drive and response systems (3.2) and (3.3) are integrated numerically with the fractional orders $\alpha = 0.9, 0.95, 0.99$ and using the initial values $x_m(0) = 15$, $y_m(0) = 20$, $z_m(0) = 29$ and $x_s(0) = 10$, $y_s(0) = 15$, $z_s(0) = 25$. From Figure 4, it is clear that the synchronization is achieved for all these values of fractional order when $k_1 = 20$ and $k_2 = 10$.

3.2. Synchronization between Lorenz-Like and Chen Fractional Order Systems

In this case it is assumed that, the fractional order Lorenz-like system drives the fractional order Chen system. The drive and response systems are defined as follows:

$$\frac{d^\alpha x_m}{dt^\alpha} = \sigma(y_m - x_m), \qquad \frac{d^\alpha y_m}{dt^\alpha} = \rho x_m - x_m z_m + \gamma y_m, \qquad \frac{d^\alpha z_m}{dt^\alpha} = x_m y_m - \beta z_m, \tag{3.12}$$

$$\frac{d^\alpha x_s}{dt^\alpha} = a(y_s - x_s) + v_1, \qquad \frac{d^\alpha y_s}{dt^\alpha} = (c - a)x_s - x_s z_s + c y_s + v_2, \qquad \frac{d^\alpha z_s}{dt^\alpha} = x_s y_s - b z_s + v_3,$$

$$(3.13)$$
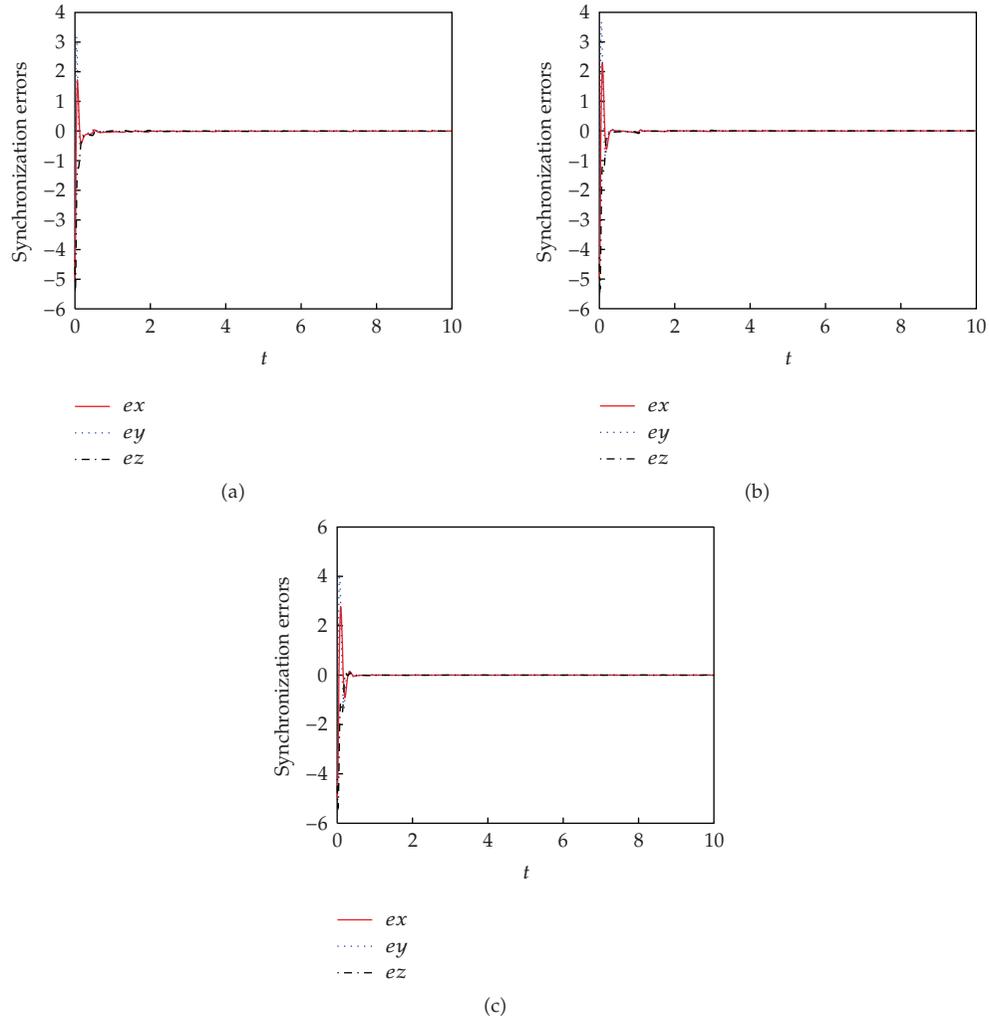
Figure 4: Synchronization errors of the drive system (3.2) and response system (3.3) using $k_1 = 20$, $k_2 = 10$ and fractional orders: (a) $\alpha = 0.9$, (b) $\alpha = 0.95$, and (c) $\alpha = 0.99$.

where v_1, v_2, and v_3 are the linear control functions. The error variables are given by

$$e_1 = x_s - x_m, \qquad e_2 = y_s - y_m, \qquad e_3 = z_s - z_m. \tag{3.14}$$

By subtracting (3.12) from (3.13) and using (3.14), we get

$$\frac{d^\alpha e_1}{dt^\alpha} = a(e_2 - e_1) + (a - \sigma)(y_m - x_m) + v_1,$$

$$\frac{d^\alpha e_2}{dt^\alpha} = c e_2 - z_m e_1 - x_m e_3 - e_1 e_3 + (c - a)x_s - \rho x_m + (c - \gamma)y_m + v_2, \tag{3.15}$$

$$\frac{d^\alpha e_3}{dt^\alpha} = -b e_3 + y_m e_1 + x_m e_2 + e_1 e_2 + (\beta - b)z_m + v_3.$$

Now, by choosing

$$v_1 = (\sigma - a)(y_m - x_m), \qquad v_2 = \rho x_m + (a - c)x_s + (\gamma - c)y_m - k_1 e_2, \qquad v_3 = (b - \beta)z_m - k_2 e_3,$$

(3.16)

where k_1, $k_2 \geq 0$, then the error system (3.15) is rewritten as

$$\frac{d^\alpha e_1}{dt^\alpha} = a(e_2 - e_1),$$

$$\frac{d^\alpha e_2}{dt^\alpha} = (c - k_1)e_2 - z_m e_1 - x_m e_3 - e_1 e_3,$$

(3.17)

$$\frac{d^\alpha e_3}{dt^\alpha} = -(b + k_2)e_3 + y_m e_1 + x_m e_2 + e_1 e_2.$$

Take Laplace transform in both sides of (3.17), let $E_i(s) = L\{e_i(t)\}$, where $(i = 1, 2, 3)$, and apply $L\{d^\alpha e_i / dt^\alpha\} = s^\alpha E_i(s) - s^{\alpha-1} e_i(0)$. After that, by doing similar analysis like the previous subsection, we obtain

$$\lim_{t \to \infty} e_1(t) = \lim_{s \to 0^+} s E_1(s) = \lim_{s \to 0^+} s E_2(s) = \lim_{t \to \infty} e_2(t),$$

$$\lim_{t \to \infty} e_2(t) = \lim_{s \to 0^+} s E_2(s) = \frac{1}{c - k_1} \lim_{s \to 0^+} s L\{x_m e_3\} + \frac{1}{c - k_1} \lim_{s \to 0^+} s L\{z_m e_1\} + \frac{1}{c - k_1} \lim_{t \to \infty} e_1(t) \cdot \lim_{t \to \infty} e_3(t),$$

$$\lim_{t \to \infty} e_3(t) = \lim_{s \to 0^+} s E_3(s) = \frac{1}{b + k_2} \lim_{s \to 0^+} s L\{y_m e_1\} + \frac{1}{b + k_2} \lim_{s \to 0^+} s L\{x_m e_2\} + \frac{1}{b + k_2} \lim_{t \to \infty} e_1(t) \cdot \lim_{t \to \infty} e_2(t).$$

(3.18)

If we assume that $c - k_1 \neq 0$ and $E_1(s)$, $E_2(s)$ are bounded, then it follows that $\lim_{t \to \infty} e_1(t) = \lim_{t \to \infty} e_2(t) = 0$. Now, owing to the attractiveness of the attractors of systems (2.1) and (2.3), there exists $\xi > 0$ such that $|x_i(t)| \leq \xi < \infty$, $|y_i(t)| \leq \xi < \infty$, and $|z_i(t)| \leq \xi < \infty$ where i refers to the index of the drive or response variables. Therefore, $\lim_{t \to \infty} e_3(t) = 0$. Consequently,

$$\lim_{t \to \infty} e_i(t) = 0, \quad i = 1, 2, 3.$$

(3.19)

Thus, the states of the drive system (3.12) are synchronized with the states of the response system (3.13), as the controllers (3.16) are activated.

3.2.1. Numerical Results

Numerical simulations are carried out to integrate the drive and response systems (3.12) and (3.13) using the predictor-correctors scheme, with the fractional orders $\alpha = 0.9$, 0.95, 0.99 and the initial values $x_m(0) = 10$, $y_m(0) = 16$, $z_m(0) = 25$ and $x_s(0) = 15$, $y_s(0) = 20$, $z_s(0) = 29$. Thus, the drive and response systems (3.12) and (3.13) are synchronized in such a successful way for all at the above-mentioned fractional orders values, using the linear controllers (3.16) with $k_1 = 20$ and $k_2 = 10$ (see Figure 5).

(a)

(b)

(c)
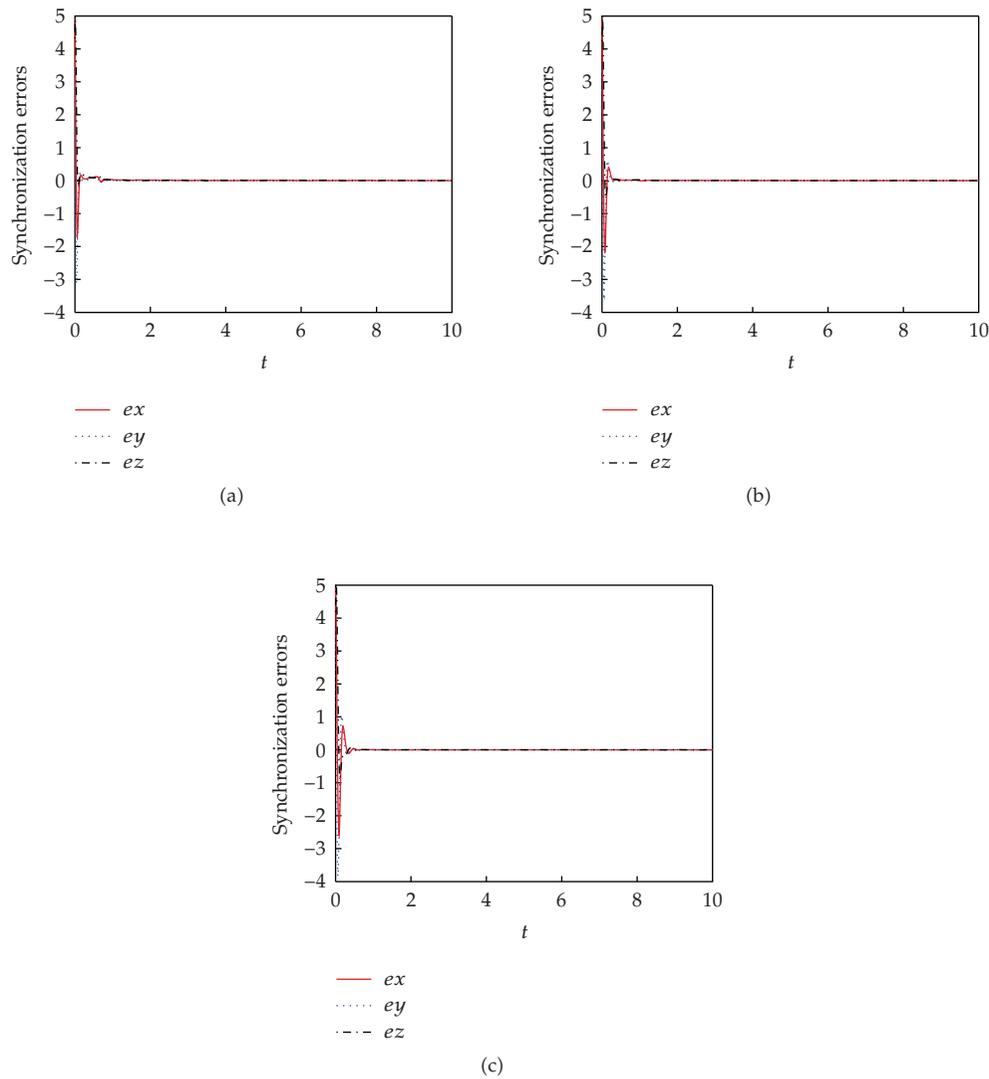
Figure 5: Synchronization errors of the drive system (3.12) and response system (3.13) using $k_1 = 20$, $k_2 = 10$ and fractional orders: (a) $\alpha = 0.9$, (b) $\alpha = 0.95$, and (c) $\alpha = 0.99$.

4. Conclusion

Chaos synchronization between two different fractional order chaotic systems has been studied using linear control technique. Fractional order Chen system has been used to drive fractional order Lü system, and fractional order Lorenz-like system has been used to drive fractional order Chen system. Conditions for chaos synchronization have been investigated theoretically by using Laplace transform. Numerical simulations have been carried out using different fractional order values to show the effectiveness of the proposed synchronization techniques.

Acknowledgments

The author wishes to thank the reviewers and the associate editor for their careful reading and efforts and for providing some helpful suggestions. Also I wish to thank Professor E. Ahmed and Dr. Faycal Ben Adda for discussion and help.

References

[1] P. L. Butzer and U. Westphal, *An Introduction to Fractional Calculus*, World Scientific, Singapore, 2000.

[2] K. S. Miller and B. Ross, *An Introduction to the Fractional Calculus and Fractional Differential Equations*, A Wiley-Interscience Publication, John Wiley & Sons, New York, NY, USA, 1993.

[3] R. Hilfer, Ed., *Applications of FractionalCalculus in Physics*, World Scientific, River Edge, NJ, USA, 2000.

[4] H. H. Sun, A. A. Abdelwahab, and B. Onaral, "Linear approximation of transfer function with a pole of fractional power," *IEEE Transactions on Automatic Control*, vol. 29, no. 5, pp. 441–444, 1984.

[5] E. Ahmed and A. S. Elgazzar, "On fractional order differential equations model for nonlocal epidemics," *Physica A*, vol. 379, no. 2, pp. 607–614, 2007.

[6] A. M. A. El-Sayed, A. E. M. El-Mesiry, and H. A. A. El-Saka, "On the fractional-order logistic equation," *Applied Mathematics Letters*, vol. 20, no. 7, pp. 817–823, 2007.

[7] N. Laskin, "Fractional market dynamics," *Physica A*, vol. 287, no. 3-4, pp. 482–492, 2000.

[8] W.-C. Chen, "Nonlinear dynamics and chaos in a fractional-order financial system," *Chaos, Solitons & Fractals*, vol. 36, no. 5, pp. 1305–1314, 2008.

[9] I. Podlubny, *Fractional Differential Equations*, vol. 198 of *Mathematics in Science and Engineering*, Academic Press, San Diego, Calif, USA, 1999.

[10] M. Caputo, "Linear models of dissipation whose Q is almost frequency independent-II," *Geophysical Journal of the Royal Astronomical Society*, vol. 13, pp. 529–539, 1967.

[11] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of the Atmospheric Sciences*, vol. 20, pp. 130–141, 1963.

[12] G. Chen and T. Ueta, "Yet another chaotic attractor," *International Journal of Bifurcation and Chaos*, vol. 9, no. 7, pp. 1465–1466, 1999.

[13] J. Lü, G. Chen, and S. Zhang, "The compound structure of a new chaotic attractor," *Chaos, Solitons & Fractals*, vol. 14, no. 5, pp. 669–672, 2002.

[14] I. Grigorenko and E. Grigorenko, "Chaotic dynamics of the fractional Lorenz system," *Physical Review Letters*, vol. 91, no. 3, Article ID 034101, 4 pages, 2003.

[15] C. Li and G. Peng, "Chaos in Chen's system with a fractional order," *Chaos, Solitons & Fractals*, vol. 22, no. 2, pp. 443–450, 2004.

[16] J. G. Lu, "Chaotic dynamics of the fractional-order Lü system and its synchronization," *Physics Letters A*, vol. 354, no. 4, pp. 305–311, 2006.

[17] T. Zhou and C. Li, "Synchronization in fractional-order differential systems," *Physica D*, vol. 212, no. 1-2, pp. 111–125, 2005.

[18] G. Peng, "Synchronization of fractional order chaotic systems," *Physics Letters A*, vol. 363, no. 5-6, pp. 426–432, 2007.

[19] Z.-M. Ge and W.-R. Jhuang, "Chaos, control and synchronization of a fractional order rotational mechanical system with a centrifugal governor," *Chaos, Solitons & Fractals*, vol. 33, no. 1, pp. 270–289, 2007.

[20] C. Li and J. Yan, "The synchronization of three fractional differential systems," *Chaos, Solitons & Fractals*, vol. 32, no. 2, pp. 751–757, 2007.

[21] A. E. Matouk and H. N. Agiza, "Bifurcations, chaos and synchronization in ADVP circuit with parallel resistor," *Journal of Mathematical Analysis and Applications*, vol. 341, no. 1, pp. 259–269, 2008.

[22] A. Uchida, S. Kinugawa, and S. Yoshimori, "Synchronization of chaos in two microchip lasers by using incoherent feedback method," *Chaos, Solitons & Fractals*, vol. 17, no. 2-3, pp. 363–368, 2003.

[23] Y.-N. Li, L. Chen, Z.-S. Cai, and X.-Z. Zhao, "Study on chaos synchronization in the Belousov-Zhabotinsky chemical system," *Chaos, Solitons & Fractals*, vol. 17, no. 4, pp. 699–707, 2003.

[24] B. Blasius, A. Huppert, and L. Stone, "Complex dynamics and phase synchronization in spatially extended ecological systems," *Nature*, vol. 399, no. 6734, pp. 354–359, 1999.

[25] K. M. Cuomo and A. V. Oppenheim, "Circuit implementation of synchronized chaos with applications to communications," *Physical Review Letters*, vol. 71, no. 1, pp. 65–68, 1993.

[26] W. Deng, "Generalized synchronization in fractional order systems," *Physical Review E*, vol. 75, no. 5, Article ID 056201, 7 pages, 2007.

[27] X. Wang and Y. He, "Projective synchronization of fractional order chaotic system based on linear separation," *Physics Letters A*, vol. 372, no. 4, pp. 435–441, 2008.

[28] K. Diethelm and N. J. Ford, "Analysis of fractional differential equations," *Journal of Mathematical Analysis and Applications*, vol. 265, no. 2, pp. 229–248, 2002.

[29] K. Diethelm, N. J. Ford, and A. D. Freed, "A predictor-corrector approach for the numerical solution of fractional differential equations," *Nonlinear Dynamics*, vol. 29, no. 1–4, pp. 3–22, 2002.

Research Article

Diffusive Synchronization of Hyperchaotic Lorenz Systems

Ruy Barboza

Department of Electrical Engineering, School of Engineering at São Carlos, University of São Paulo, 13566-590 São Carlos, SP, Brazil

Correspondence should be addressed to Ruy Barboza, rbarboza@sel.eesc.usp.br

The synchronizing properties of two diffusively coupled hyperchaotic Lorenz 4D systems are investigated by calculating the transverse Lyapunov exponents and by observing the phase space trajectories near the synchronization hyperplane. The effect of parameter mismatch is also observed. A simple electrical circuit described by the Lorenz 4D equations is proposed. Some results from laboratory experiments with two coupled circuits are presented.

1. Introduction

Coupled oscillators are currently studied in physics, chemistry, biology, neural networks, and other fields. A large number of coupled oscillators form a complex system for which the investigation of coherence and synchronization is important and in the last decade the discovery that chaotic systems can synchronize added interest to this topic [1–4]. An approach for contributing to the investigation of large networks is studying the properties of a small number of coupled oscillators—this is the approach used in the present work. Chaos synchronization is also of interest in secure communication systems. For such application, hyperchaos has drawn attention for providing more complex waveforms than simply chaotic systems, thus improving the masking process. This is because hyperchaos is characterized by at least two positive Lyapunov exponents, while simple chaos shows a single one. Related to this subject is the question on how many variables are necessary to be coupled in order to obtain synchronization. Although chaotic systems can synchronize by a single variable coupling, it was for some time believed that in the case of hyperchaos the minimum number of coupling variables had to be equal to the number of positive Lyapunov exponents [5]. It was later demonstrated in [6] that it is not true, and some

hyperchaotic systems can achieve synchronization by a single variable. Other systems, however, for example the Rössler equations for hyperchaos, are unable to synchronize by only one of its variables [6, 7]. Another problem refers to synchronization under parameter mismatch, since usually perfect synchronization is achieved only if the coupling systems are identical. In this work we are concerned with the question of whether a hyperchaotic Lorenz system can synchronize and with which variables. The effect of mismatch is also observed. We present some results from numerical and laboratory experiments on an eight-dimensional dynamical system obtained by diffusively coupling two hyperchaotic Lorenz systems.

2. Equations and Numerical Simulations

2.1. Diffusive Coupling

Consider the four-dimensional system $d\mathbf{x}/dt = \mathbf{f}(\mathbf{x})$, where $\mathbf{x} = [x, y, z, w]^t$. *Unidirectional* diffusive coupling involving two such systems is obtained by using the system $d\mathbf{x}_1/dt = \mathbf{f}_1(\mathbf{x}_1)$ to drive the response system $d\mathbf{x}_2/dt = \mathbf{f}_2(\mathbf{x}_2) + \mathbf{K}(\mathbf{x}_1 - \mathbf{x}_2)$. In this work we consider $\mathbf{K}$ a diagonal matrix, where each element k_{ii} is the strength of the coupling related to the corresponding variable. We also consider $\mathbf{f}_1(\mathbf{x}) = \mathbf{f}_2(\mathbf{x}) = \mathbf{f}(\mathbf{x})$, and each entry k_{11}, k_{22}, k_{33}, or k_{44} is either zero or equal to a positive constant k. For example, the case $k_{11} = 0$, $k_{22} = k$, $k_{33} = 0$, and $k_{44} = k$ means simultaneous y-y and w-w coupling. The two coupled systems form a compound eight-dimensional system. If the systems synchronize, the motion must remain on the hyperplane $\mathbf{x}_1 = \mathbf{x}_2$. For small $(\mathbf{x}_1 - \mathbf{x}_2)$ we have $\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2) \approx (\partial \mathbf{f}(\mathbf{x}_1)/\partial \mathbf{x}_1) \cdot (\mathbf{x}_1 - \mathbf{x}_2)$, so we can write the variational equation

$$\frac{d\mathbf{x}_\perp}{dt} = \mathbf{A}(\mathbf{x}_1)\, \mathbf{x}_\perp, \tag{2.1}$$

where $\mathbf{A} = \partial \mathbf{f}(\mathbf{x}_1)/\partial \mathbf{x}_1 - \mathbf{K}$ and $\mathbf{x}_\perp = (\mathbf{x}_1 - \mathbf{x}_2)$. In the case of *bidirectional* diffusive coupling we have $d\mathbf{x}_1/dt = \mathbf{f}(\mathbf{x}_1) + \mathbf{K}(\mathbf{x}_2 - \mathbf{x}_1)$ and $d\mathbf{x}_2/dt = \mathbf{f}(\mathbf{x}_2) + \mathbf{K}(\mathbf{x}_1 - \mathbf{x}_2)$, therefore (2.1) also applies, now with $\mathbf{A} = \partial \mathbf{f}(\mathbf{x}_1)/\partial \mathbf{x}_1 - 2\mathbf{K}$. We use (2.1) as the locally linear dynamical system associated to a fiducial trajectory [8] of the coupled system to calculate the transverse Lyapunov exponents [2, 3, 7, 9], so called because the perturbation $\mathbf{x}_\perp$ is transverse to the synchronization hyperplane. The coupled system will remain stably synchronized if all transverse Lyapunov exponents (TLEs) are negative.

2.2. Complete Replacement

If the drive system 1 transmits the scalar component x_1 and the corresponding variable x_2 of the response system 2 is replaced by the transmitted one x_1, this is called complete replacement [4]. As explained in [4], unidirectional diffusive coupling and complete replacements are related, since at very high values of k the variable x_1 slaves x_2. Therefore, in our numerical and experimental investigations, complete replacement of one or more variables corresponds to $k \to \infty$.

2.3. Coupled Hyperchaotic Lorenz Systems

The preceding coupling scheme will now be applied to systems described by the following Lorenz equations linearly extended to four dimensions:

$$
\begin{aligned}
\frac{dx}{dt} &= \sigma(y - x), \\
\frac{dy}{dt} &= x(r - z) - y + w, \\
\frac{dz}{dt} &= xy - bz, \\
\frac{dw}{dt} &= -\gamma x,
\end{aligned}
\tag{2.2}
$$

which shows hyperchaos and was theoretically analyzed in [10]. In the case of unidirectional coupling the eight-dimensional system is given by

$$
\begin{aligned}
\frac{dx_1}{dt} &= \sigma(y_1 - x_1), & \frac{dx_2}{dt} &= \sigma(y_2 - x_2) + k_{11}(x_1 - x_2), \\
\frac{dy_1}{dt} &= x_1(r - z_1) - y_1 + w_1, & \frac{dy_2}{dt} &= x_2(r - z_2) - y_2 + w_2 + k_{22}(y_1 - y_2), \\
\frac{dz_1}{dt} &= x_1 y_1 - bz_1, & \frac{dz_2}{dt} &= x_2 y_2 - bz_2 + k_{33}(z_1 - z_2), \\
\frac{dw_1}{dt} &= -\gamma x_1, & \frac{dw_2}{dt} &= -\gamma x_2 + k_{44}(w_1 - w_2).
\end{aligned}
\tag{2.3}
$$

For this system the matrix $\mathbf{A}$ is

$$
\mathbf{A} = \begin{bmatrix}
-\sigma - k_{11} & \sigma & 0 & 0 \\
z_1 - r & -1 - k_{22} & -x_1 & 1 \\
y_1 & x_1 & -b - k_{33} & 0 \\
-\gamma & 0 & 0 & -k_{44}
\end{bmatrix}
\tag{2.4}
$$

for unidirectional coupling. In the case of bidirectional coupling, the only modification is replacing k_{ii} by $2k_{ii}$ along the diagonal. The synchronizing properties of system (2.2) will now be numerically investigated by calculating the TLE of (2.3) as a function of the coupling strength k. Parameter mismatch is also examined. In the following, only unidirectional coupling is considered.

2.3.1. Parameters $\sigma = 10$, $b = 8/3$, $r = 30$, $\gamma = 10$

We first illustrate the synchronizing properties of (2.2) for the above parameter values. The values $\sigma = 10$ and $b = 8/3$ are the classical, or most popular, used in studies of the original

(a)

(b)

(c)

(d)
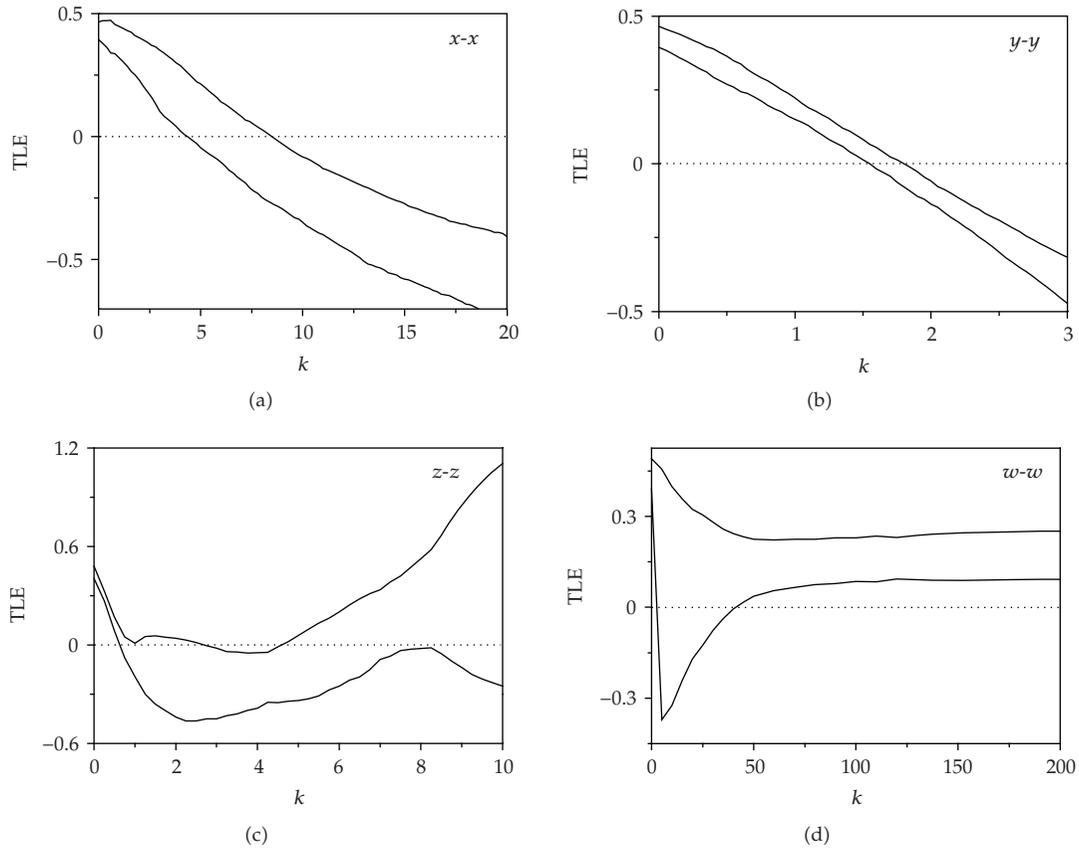
Figure 1: Transversal Lyapunov exponents (TLEs) for $\sigma = 10$, $b = 8/3$, $r = 30$, $\gamma = 10$.

Lorenz 3D system. For the hyperchaotic Lorenz 4D system the extra parameter $\gamma = 10$ is included. In Figure 1, the two largest TLE for single-variable coupling are plotted as a function of k, for $r = 30$. The synchronization thresholds are $k = 8.1$ and $k = 1.8$ for x-x and y-y couplings, respectively. The systems will never get synchronized if w alone is the coupling variable. On the other hand, z-z coupling provides a small window of stable synchronization.

2.3.2. Parameters $\sigma = 4$, $b = 0.3$, $r = 30$, $\gamma = 1.6$

The above values of the parameters are of interest in this work because small values of σ and b are easier to realize with practical component values in the circuit model presented in Section 3. (At this point it is worth remembering the observation by Sparrow [11] that small b leads to very complex behavior of the Lorenz equations.) Therefore, the system properties for such small parameter values will be examined with more detail in the following. In Figure 2, the Lyapunov spectrum for these parameter values is plotted as a function of r, showing a broad range of hyperchaotic behavior. Also shown is the one-dimensional bifurcation diagram along the same r range.

(a)

(b)
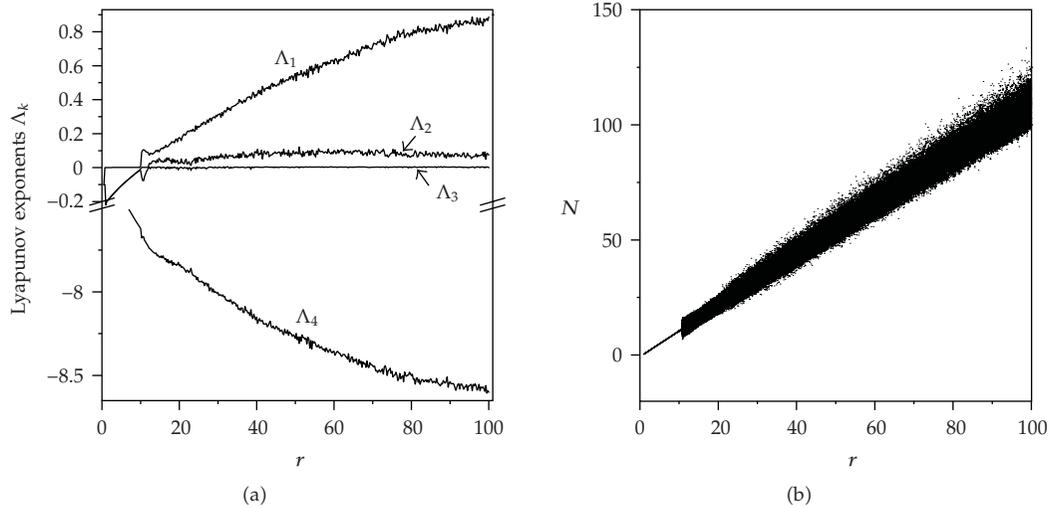
Figure 2: (a) Lyapunov spectrum $\Lambda_k(r)$; (b) bifurcation diagram obtained from the crossings through the Poincaré surface $dz/dt = 0$, corresponding to maxima of $z(t)$. Parameter values: $\sigma = 4$, $b = 0.3$, $r = 30$, $\gamma = 1.6$.

The two largest TLEs for single-variable coupling, plotted as a function of k, are shown in Figures 3(a)–3(c). The synchronization thresholds are $k = 3.2$ and $k = 1.3$ for x-x and y-y couplings, respectively. For z-z coupling, the system shows a window of stable synchronization ranging from $k = 0.7$ to $k = 9.0$. The system does not synchronize in the case of w-w coupling. However, the nonsynchronizing variables z and w, when working together in the z-z plus w-w double-coupling scheme, provide stable synchronization above $k = 0.34$. In the case of all-variable coupling, stable synchronization is achieved above $k = 0.21$.

2.3.3. Parameter Mismatch

A qualitative method of investigating the hyperchaos synchronization phenomena is by observing the projections of the eight-dimensional attractor onto the planes (x_1, x_2), (y_1, y_2), (z_1, z_2), and (w_1, w_2). In these planes the straight lines $x_1 = x_2$, $y_1 = y_2$, $z_1 = z_2$, and $w_1 = w_2$ correspond to the synchronization hyperplane. In the following we show only the (z_1, z_2) plane, since the components z_1 and z_2 seem to be the most difficult to synchronize. For all plots we used x-x coupling. Figure 4 refers to identical parameters (i.e., without mismatch), illustrating the inability of the systems to synchronize if k is less than the threshold value obtained from Figure 3(a), while perfect synchronization is achieved above the threshold: the same alignment along the diagonal is observed in all the four projection planes.

In Figures 5 and 6, we observe the effect of mismatch on synchronization (for x-x coupling). In these examples we applied the same mismatch to all parameters, that is, $\Delta\sigma/\sigma = \Delta b/b = \Delta r/r = \Delta\gamma/\gamma$. In Figure 5, where $k = 4.0$, we see that for k values just above the threshold, some good degree of synchronization is obtained for 1% mismatch; however, for 5% large deviations from the diagonal are observed. Figure 6 shows the effect of mismatch for 1% and 5% in the case of $k = 10$.
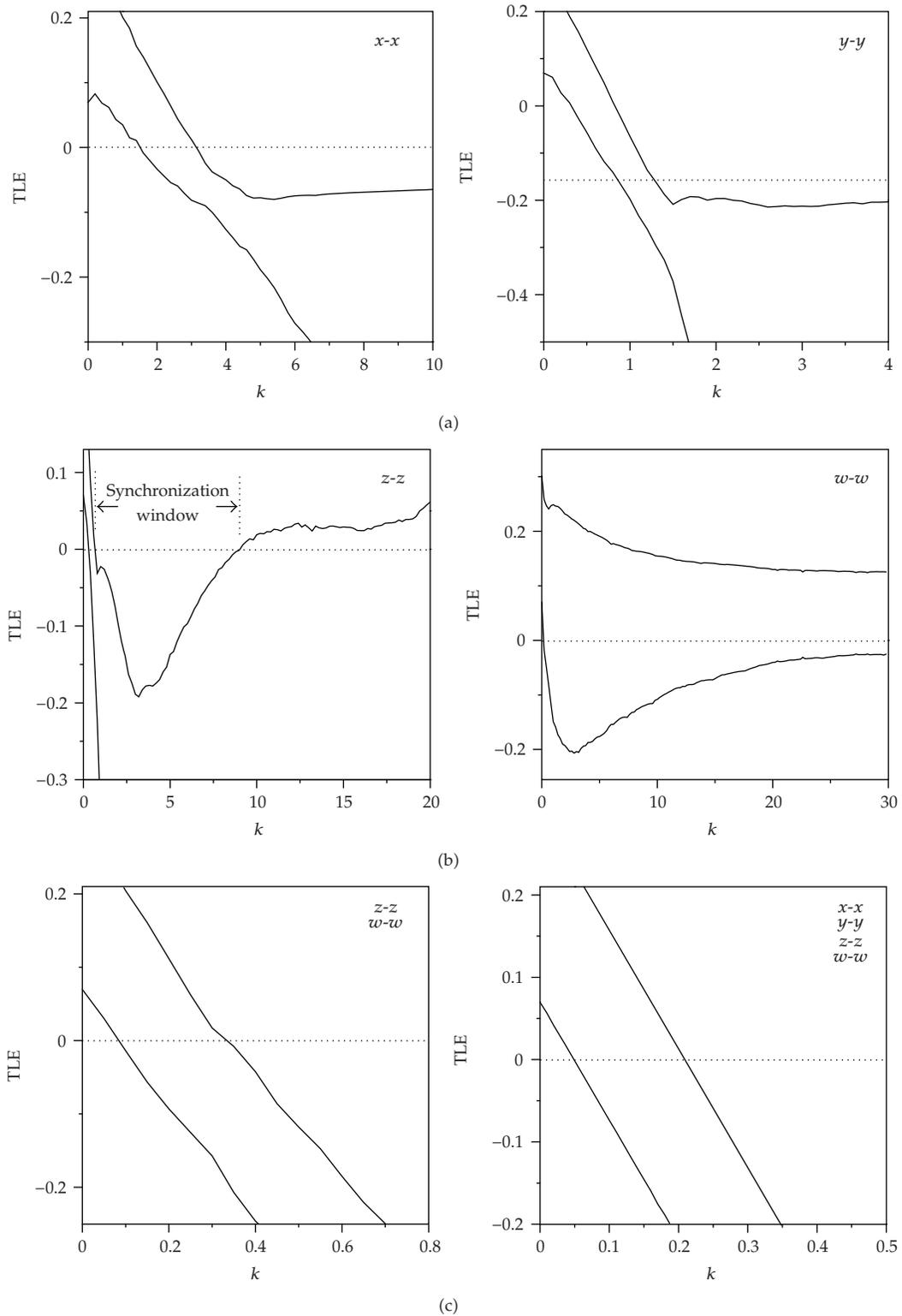
Figure 3: Transversal Lyapunov exponents (TLEs) for $\sigma = 4$, $b = 0.3$, $r = 30$, $\gamma = 1.6$. (a) and (b) One-variable coupling; (c) two- and four-variable couplings.

(a) (b)
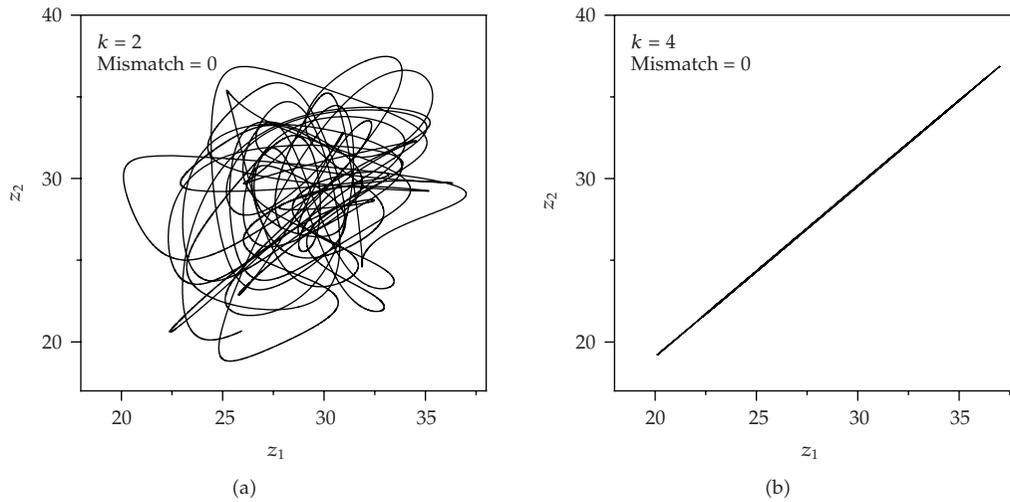
Figure 4: Synchronization for identical parameters: k below and above the threshold $k = 3.2$ ($\sigma = 4$, $b = 0.3$, $r = 30$, $\gamma = 1.6$).


(a) (b)
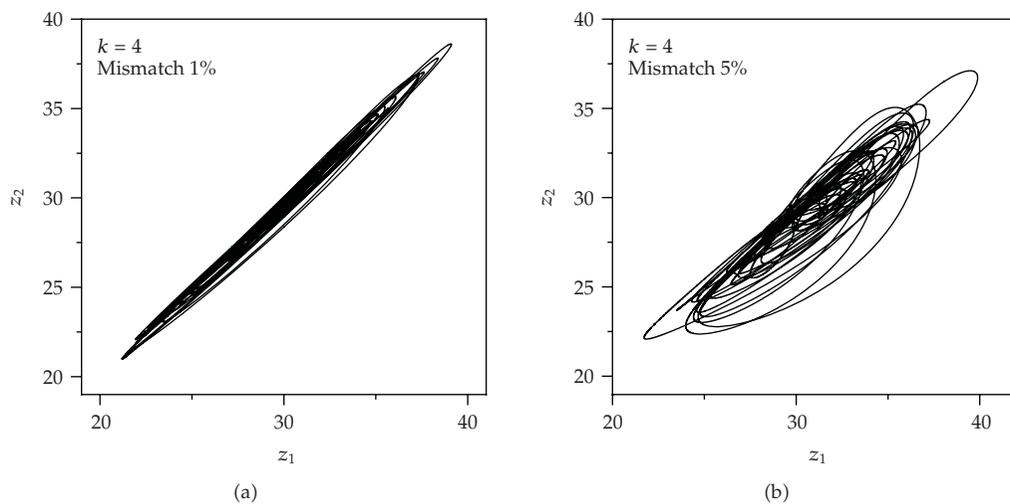
Figure 5: Effect of mismatch on synchronization for $k = 4$ ($\sigma = 4$, $b = 0.3$, $r = 30$, $\gamma = 1.6$).

3. Experiments with a Simple Electrical Circuit

3.1. Circuit Description and Equations

The Lorenz system, being one of the most important paradigms of chaos, has inspired many attempts to make a physical system representing its equations, mainly in the form of an electrical circuit. Some authors have proposed replacing the cross-products of variables by discontinuities (switching circuits) as in [12, 13], or by continuous piecewise linear resistors [14], thus resulting in very simple and practical circuits, although not truly described by the Lorenz equations. More accurate realization, though more complex, is by the analog computer approach using smooth cross-product functions, as in [15], which employs 10
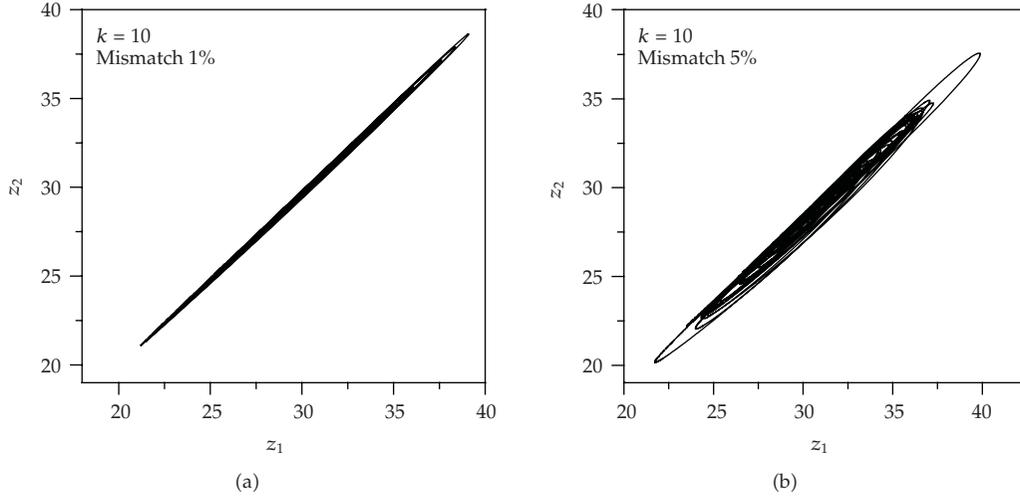
Figure 6: Effect of mismatch on synchronization for $k = 10$ ($\sigma = 4$, $b = 0.3$, $r = 30$, $\gamma = 1.6$).

integrated circuits (2 multipliers and 8 op amps) and 23 passive components, therefore a total of 33 circuit components for the Lorenz 3D circuit. In the present work we are proposing a simpler easy-to-build circuit with smooth functions, aiming at encouraging more experimental approaches on hyperchaos investigation, even by those researchers not trained in electronics. As stated in [3], to facilitate experiments with coupled chaotic oscillators the circuit is required to exhibit chaos in a large range of parameters in order that the coupling will not destroy the attractor, and it is needed to be simple enough so that several practically identical oscillators can be easily constructed. The simplest possible smooth Lorenz 3D circuit appeared in [16], using only 2 integrated circuits (2 multipliers) and 7 passive components (a total of 9 components, thus about 70% smaller than the circuit by Cuomo et al. in [15]), as shown in Figure 7(a). The good performance of this circuit is illustrated in Figure 7(b), which shows the experimental attractor and examples of single-cusp and double-cusp Lorenz maps (displayed in-line by the circuit, via a Poincaré-section circuitry). In the present work we extended to 4D that simplest circuit by adding 2 op amps and 6 passive components, obtaining the hyperchaotic Lorenz circuit shown in Figure 8(a), redrawn in Figure 8(b) using circuit theoretic symbols. The following equations describe the circuit:

$$
\begin{aligned}
C_1 \frac{dv_1}{dt} &= i - \frac{v_1}{R_1}, \\
L \frac{di}{dt} &= -\frac{v_1 v_2}{10} - R_2 i + v_3, \\
C_2 \frac{dv_2}{dt} &= \frac{R_2 i v_1}{10 R_4} - \frac{v_2 + E}{R_3}, \\
C_3 \frac{dv_3}{dt} &= -\frac{v_1}{R_1},
\end{aligned}
\tag{3.1}
$$

where we used the multipliers transfer function $W = 0.1(X_1 - X_2)(Y_1 - Y_2) + Z$, where $Z = v_1 + v_3$ on the input of the first multiplier (on the left side in Figure 8(a)), and $Z = v_2$ on

(a)

(b)
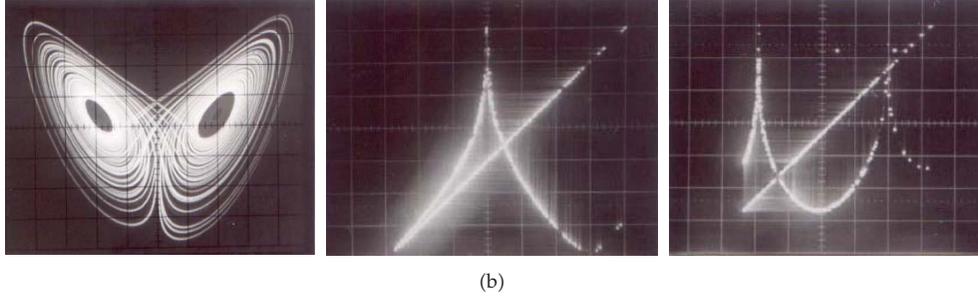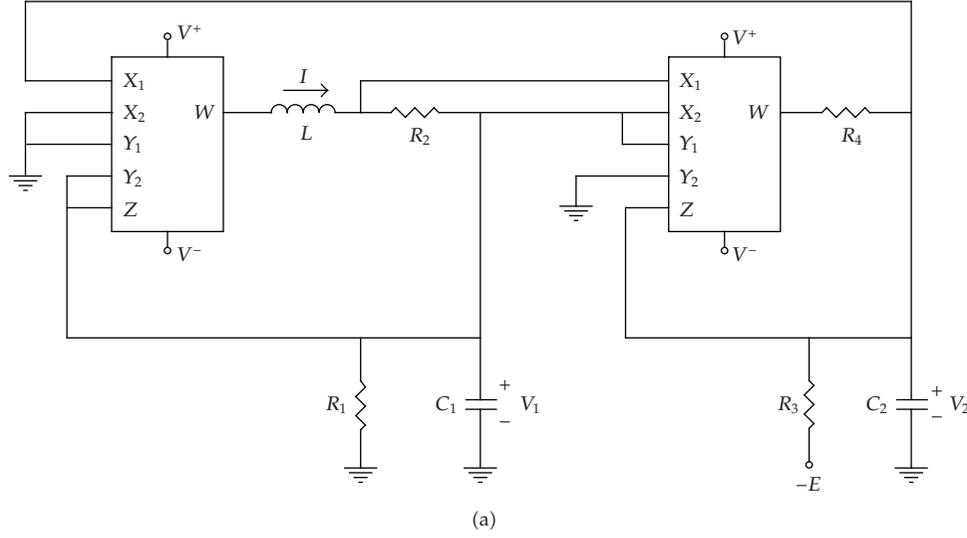
Figure 7: (a) Simplest circuit for the 3D standard Lorenz equations. (b) Experimental results from the circuit of Figure 7(a): butterfly attractor on the plane $v_1 \times v_2$; Lorenz maps showing single and double cusps $v_2(t_n) \times v_2(t_{n+1})$. The straight line in each picture is given by $v_2(t_{n+1}) = v_2(t_n)$. $L = 10\,\text{mH}$, $C_1 = 22\,\text{nF}$, $C_2 = 1\,\text{nF}$, $R_1 = 1.5\,\text{k}\Omega$, $R_2 = 100\,\Omega$, $R_3 = 160\,\text{k}\Omega$, $R_4 = 1\,\text{k}\Omega$.

the input of the second one. In deriving (3.1) we assumed $R_8 = 2R_9$ and $R_5 = R_6 = R_7 \gg R_1$. Now, by defining the new variables and parameters

$$x = \frac{v_1}{10}\sqrt{\frac{L}{R_1 R_4 C_2}}, \qquad y = \frac{R_1 i}{10}\sqrt{\frac{L}{R_1 R_4 C_2}}, \qquad z = \frac{v_2}{10}\frac{R_1}{R_2} + r,$$

$$w = \frac{v_3}{10}\frac{R_1}{R_2}\sqrt{\frac{L}{R_1 R_4 C_2}}, \qquad \sigma = \frac{L}{R_1 R_2 C_1}, \qquad b = \frac{L}{R_2 R_3 C_2}, \tag{3.2}$$

$$r = \frac{R_1 E}{10 R_2}, \qquad \gamma = \frac{L}{R_2^2 C_3},$$

we obtain (2.2). Therefore, the proposed circuit realizes, exactly, the Lorenz hyperchaotic system given by (2.2), obviously with some usual practical restrictions imposed by parasitic effects and finite bandwidth, slew-rate, excursion range, and so forth.

(a)

(b)
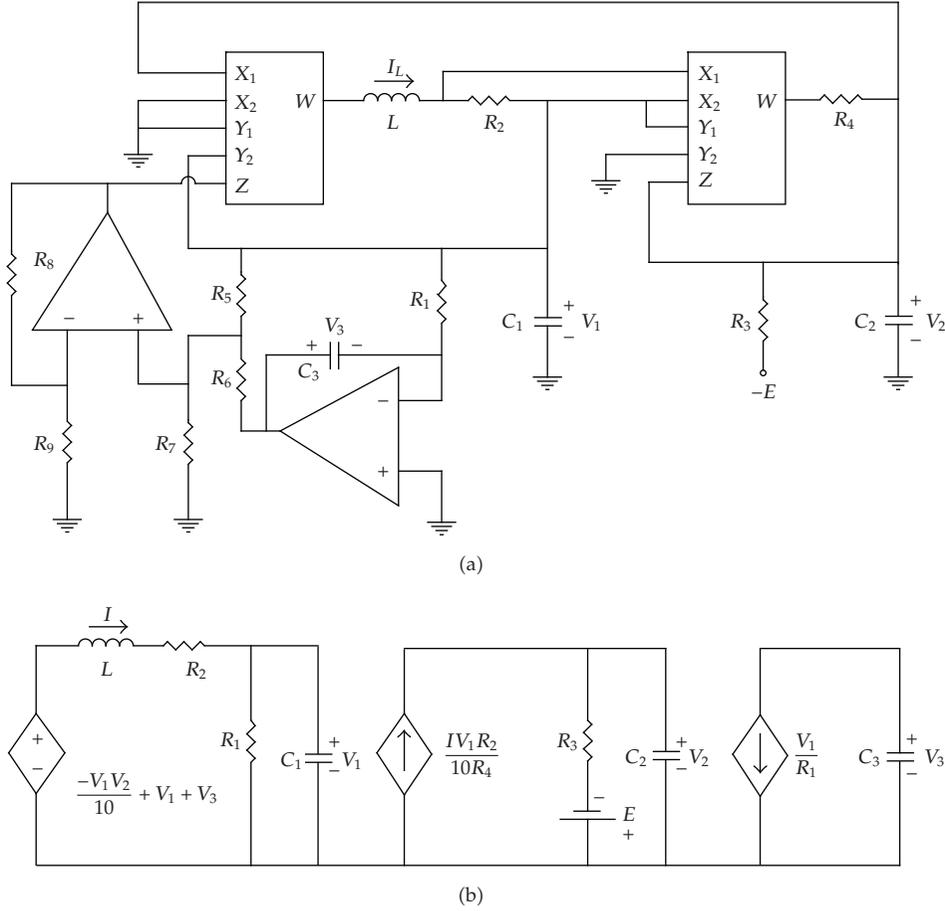
Figure 8: (a) Simplest circuit for the 4D hyperchaotic Lorenz system of (2.2). (b) Equivalent Lorenz hyperchaotic circuit.

3.2. Some Experimental Results

For our experiments, two circuits following the schematic diagram of Figure 8(a) were constructed, each one using two AD633 multipliers and two LM351 op amps, all powered with ±15 V, and the following passive components: L = 11 mH, C_1 = 22 nF, C_2 = 1 nF, C_3 = 1 μF, R_1 = 1.5 kΩ, R_2 = 80 Ω, R_3 = 470 kΩ, R_4 = 1 kΩ, $R_5 = R_6 = R_7$ = 100 kΩ, R_8 = 2 kΩ, R_9 = 1 kΩ. Using (3.2), the corresponding parameters of (2.2) are σ = 4.2, b = 0.3, γ = 1.7. We worked with E = 15 V, giving r = 28. (Note: we verified that with an independent DC power source it is possible to use E values up to 30 V, or r = 56, without waveform clipping.) The six projections of the experimental hyperchaotic attractor are shown in Figure 9(b); the calculated attractor is shown in Figure 9(a). For the experiments on synchronization, bidirectional diffusive coupling can be obtained simply by connecting a resistor R linking the capacitor C_1 of the first circuit with the capacitor C_1' of the second circuit, since in this work we have tested only x-x, or v_1-v_1, coupling. For unidirectional coupling a voltage follower is added in series with the resistor R, as sketched in Figure 10. The coupling strength is given by the relation $k = \sigma R_1/R$, as can be easily verified by adding the term $(v_1' - v_1)/R$ to the first

(a)

(b)
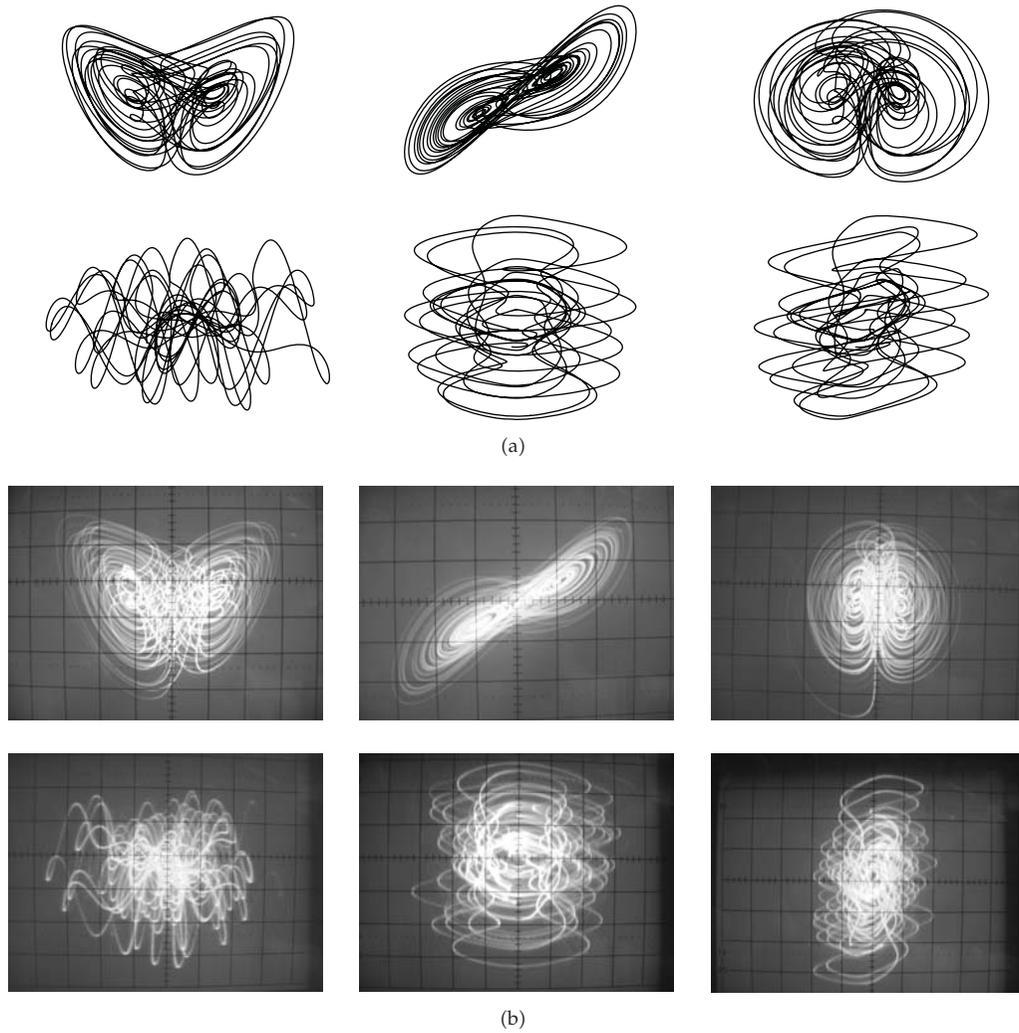
Figure 9: (a) Projections of a calculated hyperchaotic attractor described by (2.2) for $\sigma = 4.2$, $b = 0.3$, $\gamma = 1.7$, and $r = 28$. From left to right, top: $x \times z$; $x \times y$; $y \times z$; bottom: $w \times z$; $x \times w$; $y \times w$. (b) Experimental hyperchaotic Lorenz attractor generated by the circuit of Figure 8, as projected on the oscilloscope screen. From left to right, top: $v_1 \times v_2$; $v_1 \times i$; $i \times v_2$; bottom: $v_3 \times v_2$; $v_1 \times v_3$; $i \times v_3$. The same parameter values as those of Figure 9(a)—see text.


(a)   (b)
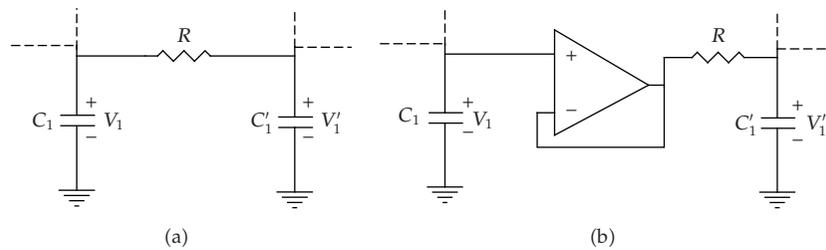
Figure 10: Coupling schemes: (a) bidirectional; (b) unidirectional.

(v_2, v_2') (v_1, v_1') (i, i') (v_3, v_3')

A

B

C

(a)

(v_2, v_2') (v_1, v_1') (i, i') (v_3, v_3')

D

E

F

(b)
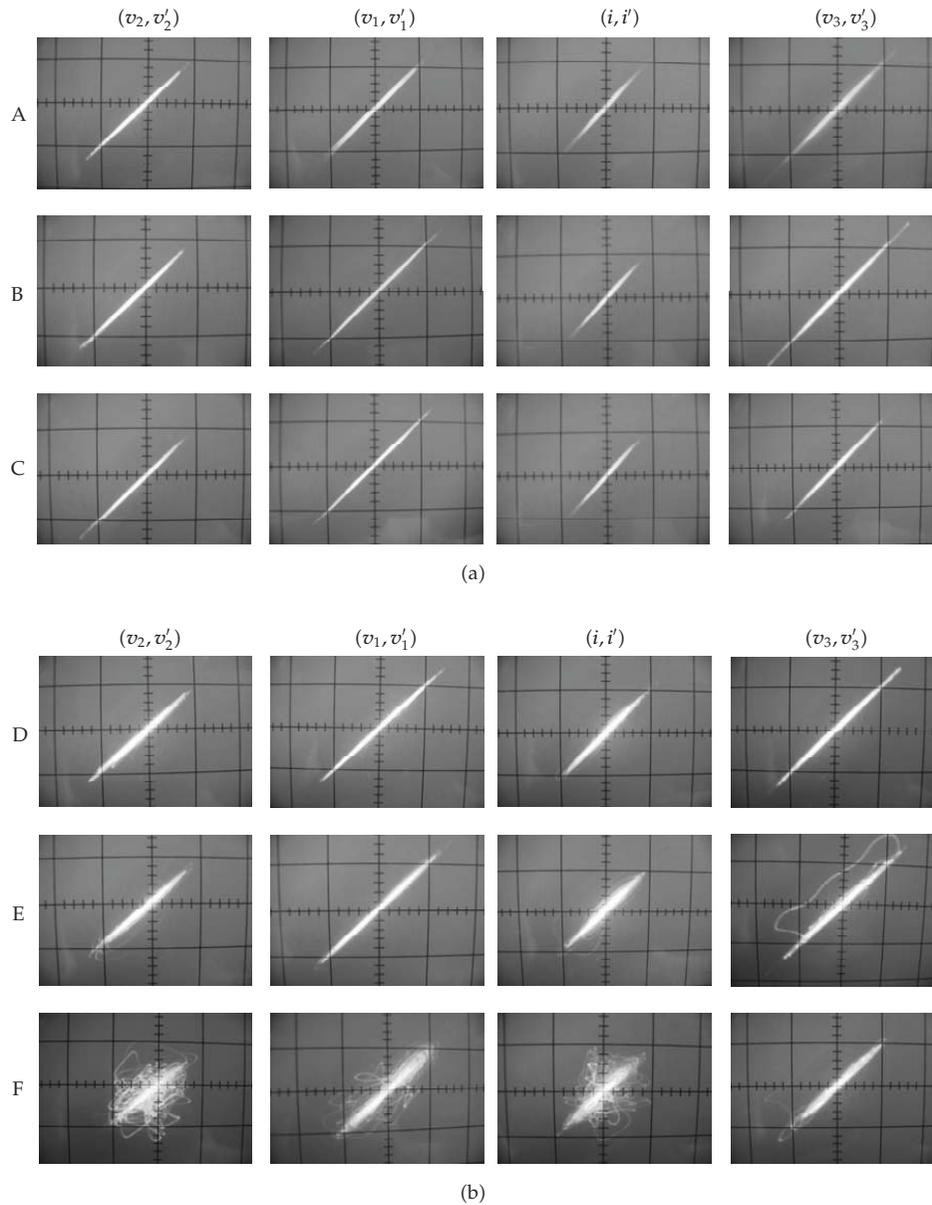
Figure 11: (a) Experimental trajectories near the synchronization hyperplane for several values of the unidirectional coupling resistor. A: $0\,\Omega$; B: $100\,\Omega$; C: $220\,\Omega$. (b) Continuation of Figure 11(a). D: $470\,\Omega$; E: $1.0\,\mathrm{k}\Omega$; F: $2.2\,\mathrm{k}\Omega$.

of (3.1). Note that each variable of the second circuit is represented by the same symbol as the corresponding one of the first circuit, but with an uppercase prime. The four projections of the trajectories near the synchronization hyperplane are shown in Figures 11(a) and 11(b) for several values of the coupling resistor R. The waveforms $v_2(t)$ and $v_2'(t)$ are shown in Figure 12. All these results refer to unidirectional coupling through the variables v_1 and v_1'.
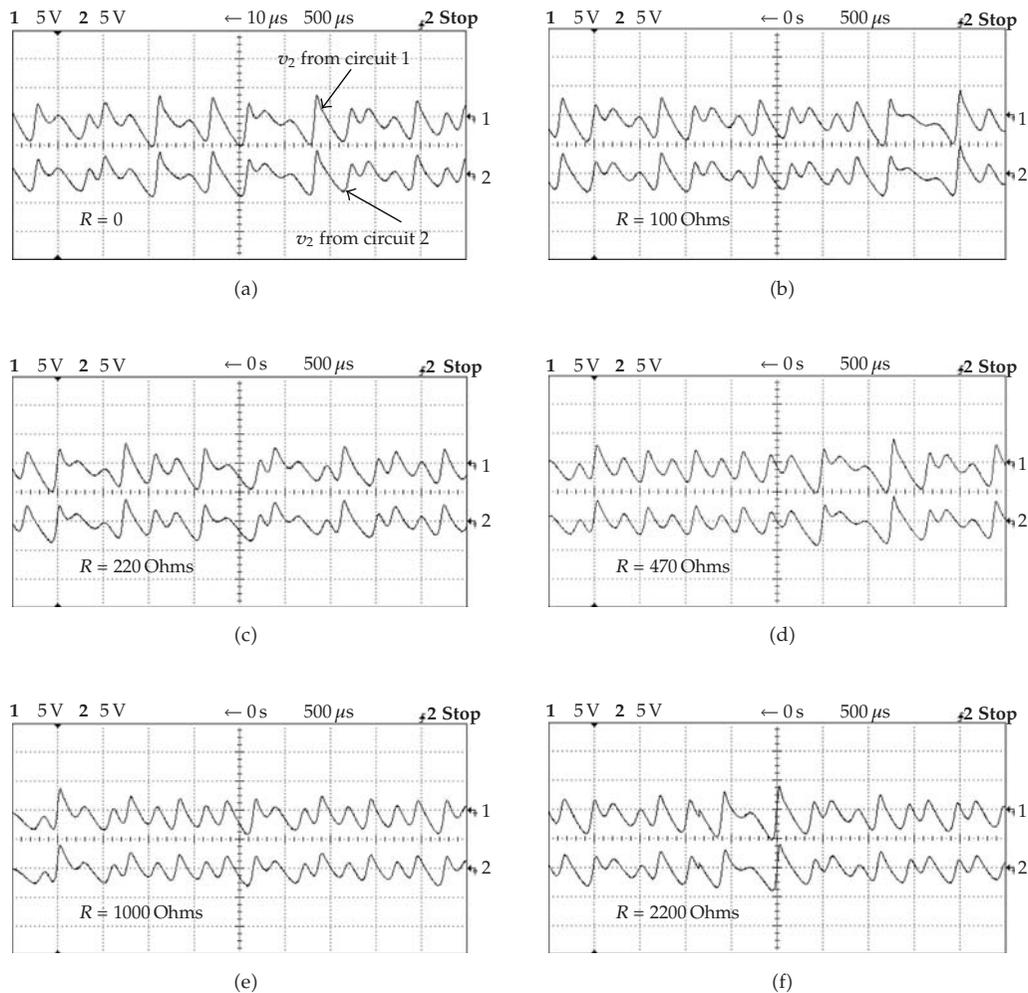
Figure 12: Experimental waveforms v_2 and v_2' corresponding to the first column of Figure 11.

4. Conclusion

In this work the synchronizing properties of diffusively coupled hyperchaotic Lorenz 4D systems, described by (2.2), have been studied both numerically and experimentally. The numerical investigation was realized by calculating the transverse Lyapunov exponents as a function of the coupling strength k, and also by visually inspecting the phase space trajectories near the synchronization hyperplane. We concluded that using a single coupling variable, either x or y (but neither z nor w), guarantees stable synchronization. Although z and w are not good choices for the single-variable scheme, double coupling with both z and w easily provides synchronization. We also verified that a small degree of parameter mismatch seems tolerable. For the laboratory work a very simple electrical circuit described by the Lorenz 4D system was proposed and described here for the first time. The experiments confirmed the qualitative behavior predicted by the numerical approach.

Acknowledgments

References

[1] L. M. Pecora and T. L. Carroll, "Synchronization in chaotic systems," *Physical Review Letters*, vol. 64, no. 8, pp. 821–824, 1990.

[2] L. M. Pecora and T. L. Carroll, "Driving systems with chaotic signals," *Physical Review A*, vol. 44, no. 4, pp. 2374–2383, 1991.

[3] J. F. Heagy, T. L. Carroll, and L. M. Pecora, "Synchronous chaos in coupled oscillator systems," *Physical Review E*, vol. 50, no. 3, pp. 1874–1885, 1994.

[4] L. M. Pecora, T. L. Carroll, G. A. Johnson, D. J. Mar, and J. F. Heagy, "Fundamentals of synchronization in chaotic systems, concepts, and applications," *Chaos*, vol. 7, no. 4, pp. 520–543, 1997.

[5] K. Pyragas, "Predictable chaos in slightly perturbed unpredictable chaotic systems," *Physics Letters A*, vol. 181, no. 3, pp. 203–210, 1993.

[6] J. H. Peng, E. J. Ding, M. Ding, and W. Yang, "Synchronizing hyperchaos with a scalar transmitted signal," *Physical Review Letters*, vol. 76, no. 6, pp. 904–907, 1996.

[7] A. Tamasevicious and A. Cenis, "Synchronizing hyperchaos with a single variable," *Physical Review E*, vol. 55, no. 1, pp. 297–299, 1997.

[8] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D*, vol. 16, no. 3, pp. 285–317, 1985.

[9] J. N. Blakely and D. J. Gauthier, "Attractor bubbling in coupled hyperchaotic oscillators," *International Journal of Bifurcation and Chaos*, vol. 10, no. 4, pp. 835–847, 2000.

[10] R. Barboza, "Dynamics of a hyperchaotic Lorenz system," *International Journal of Bifurcation and Chaos*, vol. 17, no. 12, pp. 4285–4294, 2007.

[11] C. Sparrow, *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*, vol. 41 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1982.

[12] S. Özoğuz, A. S. Elwakil, and M. P. Kennedy, "Experimental verification of the butterfly attractor in a modified Lorenz system," *International Journal of Bifurcation and Chaos*, vol. 12, no. 7, pp. 1627–1632, 2002.

[13] E. H. Baghious and P. Jarry, ""Lorenz attractor" from differential equations with piecewise-linear terms," *International Journal of Bifurcation and Chaos*, vol. 3, no. 1, pp. 201–210, 1993.

[14] R. Tokunaga, T. Matsumoto, L. O. Chua, and S. Miyama, "The piecewise-linear Lorenz circuit is chaotic in the sense of Shilnikov," *IEEE Transactions on Circuits and Systems*, vol. 37, no. 6, pp. 766–786, 1990.

[15] K. M. Cuomo, A. V. Oppenheim, and S. H. Strogatz, "Synchronization of Lorenz-based chaotic circuits with applications to communications," *IEEE Transactions on Circuits and Systems II*, vol. 40, no. 10, pp. 626–633, 1993.

[16] R. Barboza, "Experiments on Lorenz system," in *Proceedings of the International Symposium on Nonlinear Theory and Its Applications (NOLTA '04)*, vol. 1, pp. 529–532, Fukuoka, Japan, November 2004.

Research Article

Synchronization of Discrete-Time Chaotic Systems in Bandlimited Channels

Marcio Eisencraft,[1, 2] Renato D. Fanganiello,[1] and Luiz A. Baccala[2]

[1] *Programa de Pós-graduação em Engenharia Elétrica, Escola de Engenharia da Universidade Presbiteriana Mackenzie, São Paulo, Brazil*

[2] *Departamento de Telecomunicações Controle, Escola Politécnica da Universidade de São Paulo, São Paulo, Brazil*

Correspondence should be addressed to Marcio Eisencraft, marcioft@mackenzie.br

Over the last couple of decades, many methods for synchronizing chaotic systems have been proposed with communications applications in view. Yet their performance has proved disappointing in face of the nonideal character of usual channels linking transmitter and receiver, that is, due to both noise and signal propagation distortion. Here we consider a discrete-time master-slave system that synchronizes despite channel bandwidth limitations and an allied communication system. Synchronization is achieved introducing a digital filter that limits the spectral content of the feedback loop responsible for producing the transmitted signal.

1. Introduction

Since Pecora and Carroll's seminal work [1], much has been written about the potential usefulness of chaotic synchronization in communication systems (e.g., [2–8]). Much of the impetus for chaotic communications has been the rationale whereby both analog and digital chaotic modulations would have the same properties as conventional spread spectrum techniques [9].

The inherent wideband character of chaotic signals, however, becomes a problem when the communication channel imposes bandwidth limitations. Because of the receiver's nonlinear nature, if any spectral component is amiss, all spectral components at the receiver become affected. Even minute gain or phase changes are enough to fully hinder synchronism [6, 10].

Many papers have approached the problem by trying to eliminate channel distortion via channel inversion [11–13]. Reasonable results have only been attained when the

channel filters are represented by low-order filters. Furthermore, these methods usually unrealistically assume perfect channel frequency response knowledge at the receiver [6].

Rulkov and Tsimring [6] and Eisencraft and Gerken [14] independently proposed a method for synchronizing transmitter and receiver using chaotic signals under bandwidth limitations. The basic idea is to employ an identical filter at both the transmitter and the receiver so as to circumvent channel impairments. An analog circuit implementation was proposed by [6].

Here we extend this method to discrete-time dynamical systems [8, 15–17]. Much of the interest in this approach lies in the ease of employing Digital Signal Processors (DSPs) or microcontrollers for their implementation. These solutions have become ubiquitous in present day technology.

This paper is organized as follows. In Section 2, the Wu and Chua [4] synchronization method is extended to discrete-time systems. This is followed by the description of an allied communication system whose performance under bandwidth limitation is simulated. Means for overcoming the former limitations are proposed in Section 3. Finally our main conclusions are summed up in Section 4.

2. Wu and Chua's Synchronization of Discrete-Time Chaotic Systems

Wu and Chua's paper [4] addresses chaotic system synchronization differently from Pecora and Carroll's [1] seminal paper. Instead of using conditional Lyapunov exponents to check the asymptotic stability of the slave system and hence the possibility of synchronism, Wu and Chua propose that the master and slave equations be written in such a way that the dynamics of the synchronization error is simple enough to permit the direct verification of its convergence to zero.

In this section we adapt their method for discrete-time master-slave systems.

Consider two discrete-time systems defined by

$$\mathbf{x}(n+1) = A\mathbf{x}(n) + \mathbf{b} + \mathbf{f}(\mathbf{x}(n)), \tag{2.1}$$

$$\mathbf{y}(n+1) = A\mathbf{y}(n) + \mathbf{b} + \mathbf{f}(\mathbf{x}(n)), \tag{2.2}$$

where $n \in \mathbb{N}$, $\{\mathbf{x}(n), \mathbf{y}(n)\} \subset \mathbb{R}^K$, $\mathbf{x}(n) = [x_1(n), x_2(n), \ldots, x_K(n)]^T$, $\mathbf{y}(n) = [y_1(n), y_2(n), \ldots, y_K(n)]^T$. The real $A_{K \times K}$ matrix and vector $\mathbf{b}_{K \times 1}$ are constants. The function $\mathbf{f}(\cdot)$, $\mathbb{R}^K \to \mathbb{R}^K$ is nonlinear.

The system described by (2.1) is autonomous and is called *master*. The one described by (2.2) depends on $\mathbf{x}(n)$ and is called *slave*.

The synchronization error dynamics between the two systems $\mathbf{e}(n) = \mathbf{y}(n) - \mathbf{x}(n)$, in this case, is given by

$$\mathbf{e}(n+1) = A\mathbf{e}(n). \tag{2.3}$$

They are said *completely synchronized* if $\mathbf{e}(n) \to 0$ as n grows. Consequently, master and slave synchronize completely if the eigenvalues λ_i of A satisfy [18]

$$|\lambda_i| < 1, \quad 1 \leq i \leq K. \tag{2.4}$$
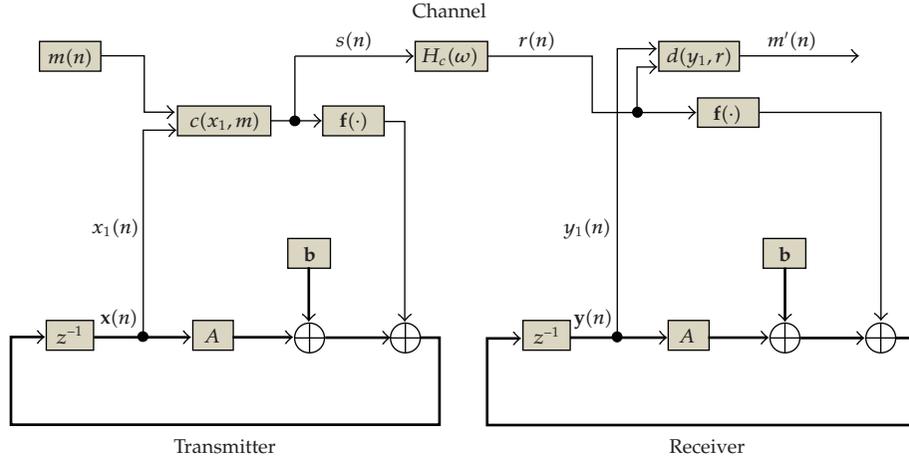
Figure 1: Block Diagram of Wu and Chua's [4] proposal adapted to discrete-time systems.

Hence, if a system can be written as in (2.1) with A satisfying condition (2.4), it is easy to set up a slave system that synchronizes with it.

Using this synchronization method, Wu and Chua [4] proposed an information transmission system using chaotic signals that lead to no errors when channel conditions are ideal. The discrete-time version of this system is shown in Figure 1.

Here we consider that $\mathbf{f}(\cdot)$ depends solely on the component $x_1(n)$ of $\mathbf{x}(n)$, and that it can be written as

$$\mathbf{f}(\mathbf{x}(n)) = \left[f(x_1(n)); 0; 0; \ldots; 0 \right]^T. \tag{2.5}$$

The communication channel is represented by a linear system with frequency response $H_C(\omega)$.

In this scheme, the information signal $m(n)$ is coded using the $x_1(n)$ component of the state vector $\mathbf{x}$ via a coding function

$$s(n) = c(x_1(n), m(n)), \tag{2.6}$$

so that the information signal can be decoded using the inverse function

$$m(n) = d(x_1(n), s(n)) = d(x_1(n), c(x_1(n), m(n))). \tag{2.7}$$

The equations governing the global system have the same form as (2.1) and (2.2). The only changes are the arguments of $f(\cdot)$:

$$\mathbf{x}(n+1) = A\mathbf{x}(n) + \mathbf{b} + \mathbf{f}(s(n)),$$

$$\mathbf{y}(n+1) = A\mathbf{y}(n) + \mathbf{b} + \mathbf{f}(r(n)), \tag{2.8}$$
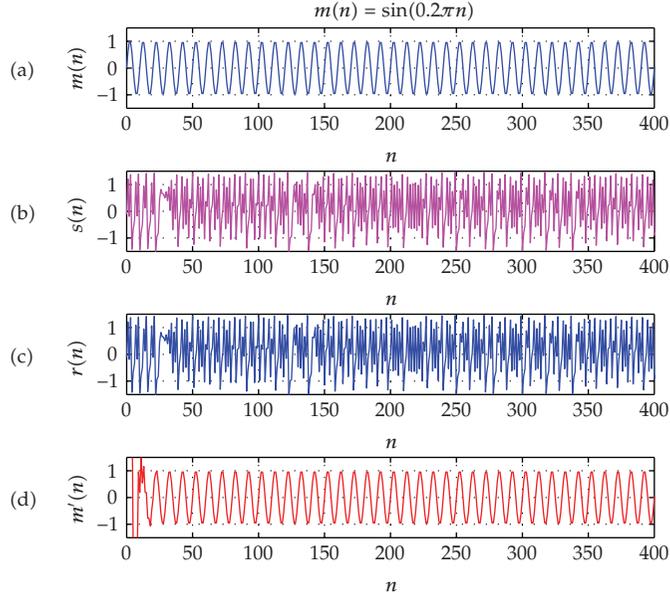
Figure 2: Simulations of the communication system shown in Figure 1 for an ideal channel: (a) message $m(n)$; (b) transmitted signal $s(n)$; (c) received signal $r(n) = s(n)$; (d) recovered message $m'(n)$.

where $r(n)$ is the signal that the channel delivers to the receiver, that is,

$$r(n) = s(n) * h_c(n), \tag{2.9}$$

$h_c(n)$ is the impulse response of the channel and "$*$" represents linear convolution.

In principle, when choosing $c(\cdot, \cdot)$ and $d(\cdot, \cdot)$ one must seek to satisfy

$$s(n) \approx x_1(n) \tag{2.10}$$

for all admissible $m(n)$ because

(i) s(n) is fed back in lieu of $x_1(n)$ at the transmitter. Condition (2.10) imposes that the dynamics is not significantly changed and consequently the system continues to generate chaotic signals;

(ii) also if secure communication is desired, condition (2.10) ensures that m(n) is not immediately apparent from s(n).

For an ideal channel, that is, $H_c(\omega) = 1$ and $s(n) = r(n)$, (2.8) become

$$\mathbf{x}(n+1) = A\mathbf{x}(n) + \mathbf{b} + f(s(n)),$$

$$\mathbf{y}(n+1) = A\mathbf{y}(n) + \mathbf{b} + f(s(n)). \tag{2.11}$$
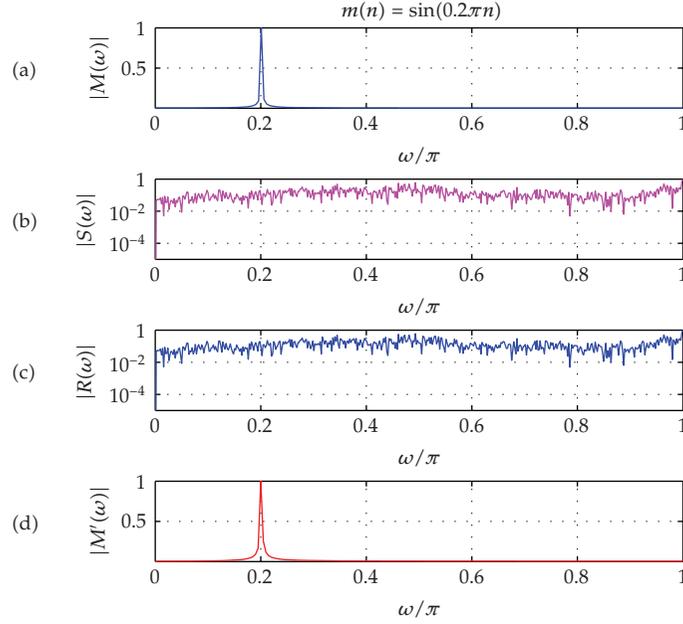
Figure 3: Normalized representation of the signals from Figure 2 in the frequency domain: (a) message; (b) transmitted signal; (c) received signal; (d) recovered message.

Once more, the synchronization error dynamics is given by (2.3) and if condition (2.4) holds, then $\mathbf{y}(n) \to \mathbf{x}(n)$ and, in particular, $y_1(n) \to x_1(n)$. Hence, using (2.7),

$$m'(n) = d\big(y_1(n), s(n)\big) \longrightarrow d(x_1(n), s(n)) = m(n). \tag{2.12}$$

Therefore when transmitter and receiver parameters are perfectly matched over an ideal channel, the message is recovered at the receiver without degradation except for a synchronization transient.

To exemplify, consider the three-dimensional generalization of the Hénon map [19, 20]:

$$
\begin{aligned}
x_1(n+1) &= -ax_1^2(n) + x_3(n) + 1, \\
x_2(n+1) &= -bx_1(n), \\
x_3(n+1) &= bx_1(n) + x_2(n).
\end{aligned}
\tag{2.13}
$$

It can be immediately written in the form of (2.1)

$$
\mathbf{x}(n+1) = \underbrace{\begin{bmatrix} 0 & 0 & 1 \\ -b & 0 & 0 \\ b & 1 & 0 \end{bmatrix}}_{A} \mathbf{x}(n) + \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{b}} + \underbrace{\begin{bmatrix} -ax_1^2(n) \\ 0 \\ 0 \end{bmatrix}}_{\mathbf{f}(\mathbf{x}(n))}. \tag{2.14}
$$
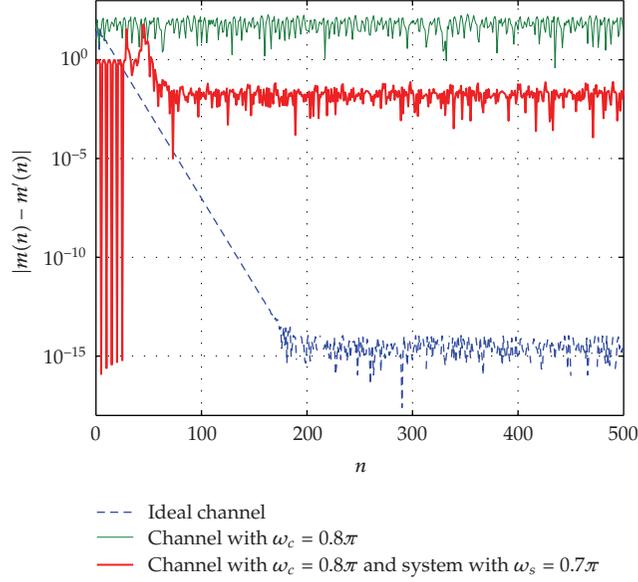
Figure 4: Demodulation error $|m(t) - m'(t)|$ for the simulated systems.

Taking $a = 1.07$ and $b = 0.3$ [19], simulations show that, for almost all initial conditions in the unity sphere, (2.13) generates chaotic orbits whose largest Lyapunov exponent is $h_1 \approx 0.23$.

In this case, the eigenvalues of A are

$$
\begin{aligned}
\lambda_1 &= 0.4084 + 0.4477i, \\
\lambda_2 &= 0.4084 - 0.4477i, \\
\lambda_3 &= -0.8169,
\end{aligned}
\tag{2.15}
$$

ensuring the satisfaction of the stability condition in (2.4). This allows assembling the communications system depicted in Figure 1.

In the simulations we chose

$$
s(n) = c(x_1(n), m(n)) = x_1(n) + 0.01m(n),
\tag{2.16}
$$

so that

$$
m'(n) = d(y_1(n), r(n)) = 100(r(n) - y_1(n)).
\tag{2.17}
$$

Furthermore, the transmitted message was taken as

$$
m(n) = \sin(0.2\pi n).
\tag{2.18}
$$
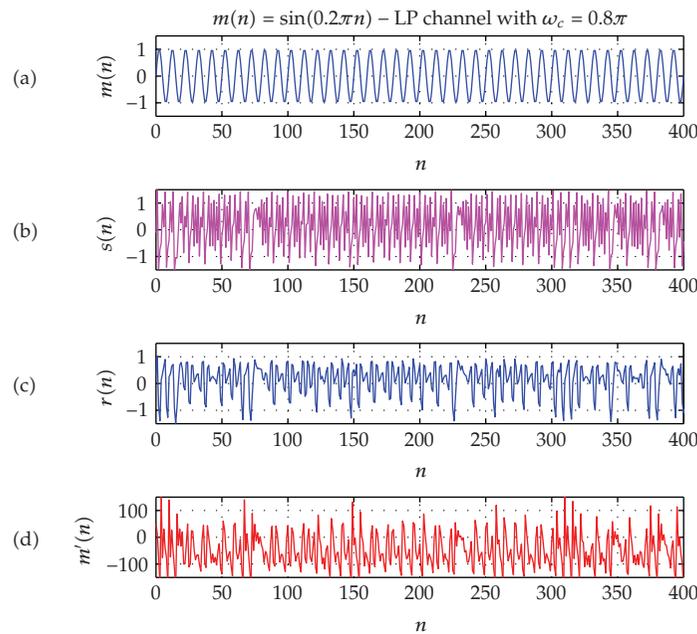
Figure 5: Simulations of the communication system shown in Figure 1 for an LP FIR channel with cutoff frequency at $\omega_c = 0.8\pi$: (a) message $m(n)$; (b) transmitted signal $s(n)$; (c) received signal $r(n)$; (d) recovered message $m'(n)$. Note that the scale used in (d) is different from the previous ones.
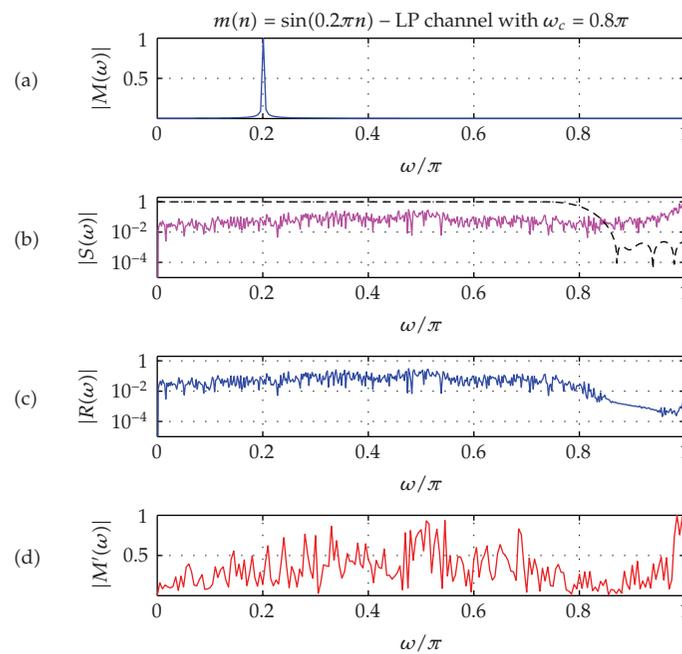

Figure 6: Normalized representation of the spectral components absolute value of signals in Figure 5: (a) message; (b) transmitted signal and amplitude frequency response of the channel (dashed line); (c) received signal; (d) recovered message.
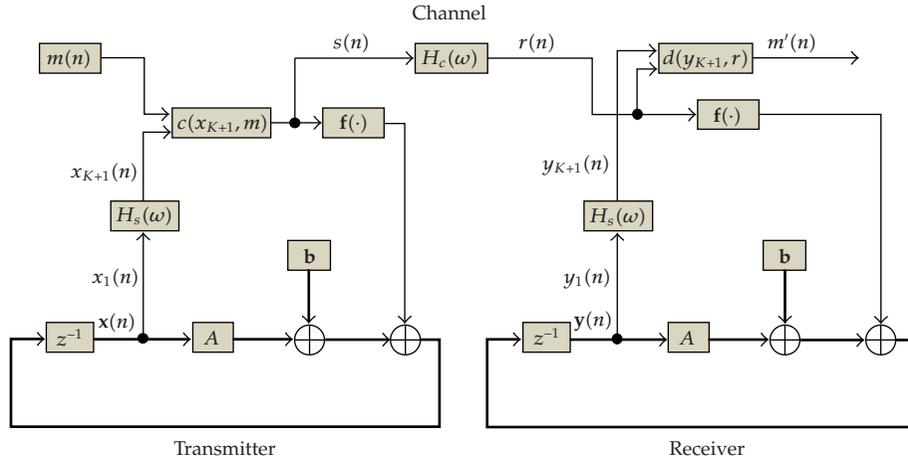
Figure 7: Proposed discrete-time communication system for bandlimited channels.
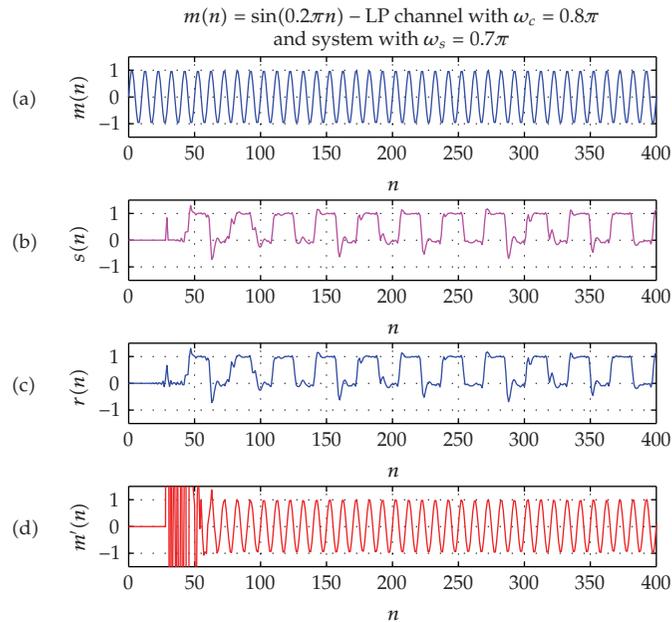
Figure 8: Simulations of the proposed communication system for an LP FIR channel with cutoff frequency at $\omega_c = 0.8\pi$ and LP FIR feedback filter $H_S(\omega)$ with cutoff frequency at $\omega_S = 0.7\pi$: (a) message $m(n)$; (b) transmitted signal $s(n)$; (c) received signal $r(n)$; (d) recovered message $m'(n)$.

Figures 2 and 3 illustrate a realization of the involved signals both in the time and in the frequency domains for an ideal channel. The spectra are normalized relatively to their maximum value.

Perfect recovery of the message once the synchronization is reached can be clearly seen in Figure 2(d) and the spectral spreading due to the chaotic signal modulation is shown in Figure 3(b). Note that the presence of $m(n)$ in the transmitted signal is not apparent neither in the time nor in the frequency domain.

The demodulation error $|m(n) - m'(n)|$ for the ideal channel case is shown by the dashed line in Figure 4. It quickly falls and remains at the numerical precision level of the computer simulation.

This situation changes completely when the channel is not ideal, that is, $H_c(\omega) \neq 1$, $\forall \omega$. In this case, $r(n) \neq s(n)$ in (2.8) and the synchronism is affected so that $m'(n) \not\to m(n)$. As an illustrative example, consider that $H_c(\omega)$ is a linear-phase Low Pass (LP) Finite Impulse Response (FIR) filter of order $N_C = 50$ with cutoff frequency at $\omega_c = 0.8\pi$ [21]. Figures 5 and 6 show a realization of the signals involved in this case. The amplitude frequency response $|H_c(\omega)|$ is shown by the dashed line in Figure 6(b). Just about 20% of the spectrum of the transmitted signal is affected, yet the message is completely lost as seen in Figures 5(d) and 6(d). Note that the bandwidth associated with the original message $m(n)$, concentrated at 0.2π, is practically unaffected.

The error in obtaining $m'(n)$ in this case is shown by the solid thinner line in Figure 4.

This example shows that the system in Figure 1 fails when channel conditions are not ideal. This makes its use impractical, in the same way as it happens with its continuous counterpart, as pointed out by [6, 14].

A mean to circumvent these difficulties is to adjust the spectrum of the transmitted chaotic signal so that it becomes less vulnerable in the presence of a given channel. This idea, proposed for continuous time systems in [6, 14], is extended for the discrete-time case in the next section.

3. Synchronizing Discrete-Time Systems in Band-Limited Channels

A way to combat the harmful effects of a bandlimited channel on the communication system of Figure 1 is to insert a filter $H_S(\omega)$ in the feedback loops of both the transmitter and the receiver so that the total transmitted signal power is contained within the channel bandwidth.

A block diagram of the proposed system is shown in Figure 7.

Considering that the filter $H_S(\omega)$ is an FIR filter of order N_S, its output $x_{K+1}(n)$ can be expressed as

$$x_{K+1}(n) = c_1 x_1(n) + c_2 x_1(n-1) + \cdots + c_{N_S+1} x_1(n - N_S), \tag{3.1}$$

where $c_1, c_2, \ldots, c_{N_S+1}$ are the filter coefficients.

Consequently, the dimensions of the difference equation systems that describes the master and the slave subsystems now are of order $K + N_S$ instead of K.

As an example, consider that the channel $H_C(\omega)$ is an LP FIR filter of order $N_C = 50$ and with cutoff frequency at $\omega_c = 0.8\pi$, as in the previous section. To combat the channel effects, we used linear-phase LP FIR filters $H_S(\omega)$ with cutoff frequency at $\omega_s = 0.7\pi$ of order $N_S = 30$. Figures 8 and 9 show a realization of the signals involved.

The information signal is unaffected by these filters and the message is fully recovered as can be seen from Figures 8(d) and 9(d). The transient time is larger than that of the previous examples due to the additional filtering effects.

The demodulation error in this case is shown by the thicker solid line in Figure 4. After the synchronization transient, the mean error is about 1%.

An important question that must be addressed is whether the generated signals remain chaotic with the introduction of the feedback filter.
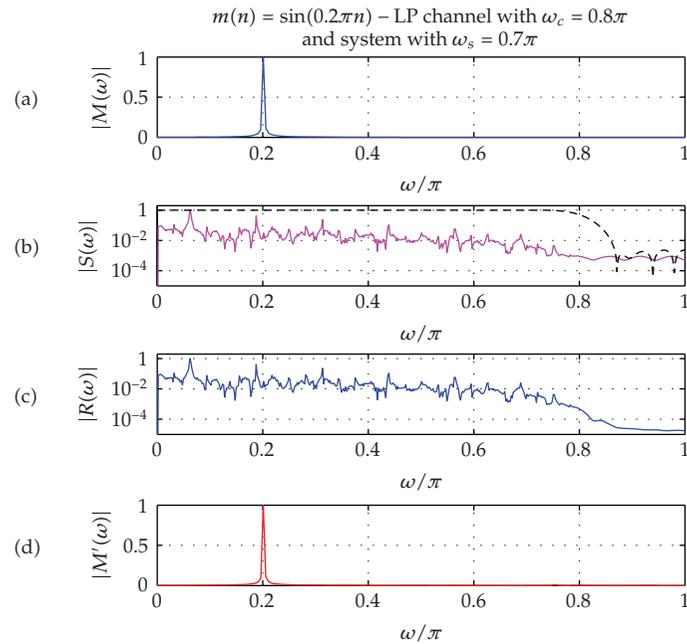
Figure 9: Normalized representation of signals in Figure 8 in the frequency domain: (a) message; (b) transmitted signal and frequency response of the channel (dashed line); (c) received signal; (d) recovered message.
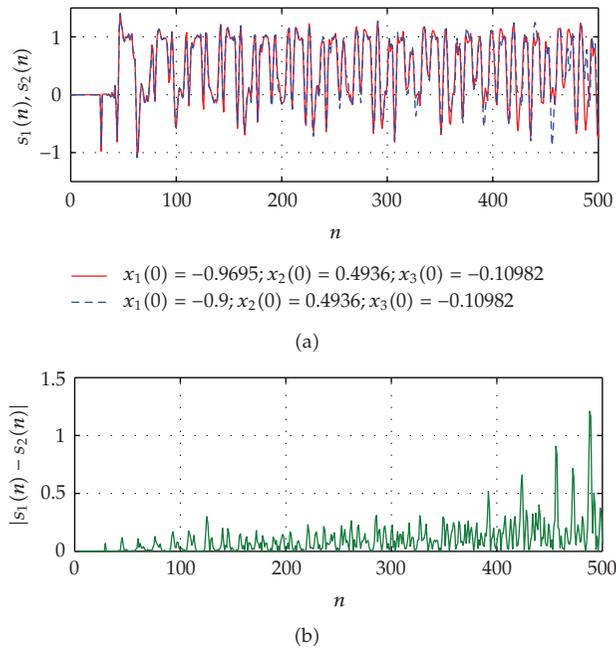

Figure 10: (a) Transmitted signals $s(n)$ for two slightly different initial conditions; (b) absolute value of the difference between the signals in (a).

In the example, the transmitted signal $s(n)$ shown in Figure 8(b) presents aperiodic behavior and sensitivity to initial conditions. This last property is illustrated in Figure 10 where signals generated using two slightly different initial conditions diverge from one another, which is evidence of chaotic signal behavior.

In a more general framework, analysis shows that when we write the system equations as in (2.8) including the feedback filter, the $K+N_S$ eigenvalues of A do not depend on the filter coefficients so that synchronism remains unaffected. However, the conditions for the signals to remain chaotic require further investigation. For low-order filters, inadequate coefficient choices may transform an unstable fixed point of the original system into a stable one and thereby destroy the chaotic attractor. These results are the subject of current investigation and lie beyond our intended scope.

4. Conclusions

In this paper, we proposed a discrete-time communication system using chaotic signals for bandlimited channels. Our system is based on the one proposed by [6, 14] who have developed their ideas within a continuous-time framework. Using discrete-time systems based on difference equations makes their implementation quite natural via today's ubiquitously available DSPs and microcontrollers.

We have shown numerical examples based on a three-dimensional generalization of the Hénon map and on linear-phase FIR filters.

It is important to note that introducing filters into feedback loops can change the dynamics of both the master and the slave systems in relevant ways. The conditions on filter coefficients for which the systems remain chaotic signal generators have the status of a nontrivial open problem. Work is in progress to simulate the use of this system to transmit digital messages and to assess its symbol error rate performance for different channels and feedback filter bandwidths.

Acknowledgments

References

[1] L. M. Pecora and T. L. Carroll, "Synchronization in chaotic systems," *Physical Review Letters*, vol. 64, no. 8, pp. 821–824, 1990.

[2] T. L. Carroll and L. M. Pecora, "Synchronizing chaotic circuits," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 4, pp. 453–456, 1991.

[3] L. M. Pecora and T. L. Carroll, "Driving systems with chaotic signals," *Physical Review A*, vol. 44, no. 4, pp. 2374–2383, 1991.

[4] C. W. Wu and L. O. Chua, "A simple way to synchronize chaotic systems with applications to secure communication systems," *International Journal of Bifurcation and Chaos*, vol. 3, no. 6, pp. 1619–1627, 1993.

[5] C. L. Koh and T. Ushio, "Digital communication method based on M-synchronized chaotic systems," *IEEE Transactions on Circuits and Systems I*, vol. 44, no. 5, pp. 383–390, 1997.

 [6] N. F. Rulkov and L. S. Tsimring, "Synchronization methods for communication with chaos over band-limited channels," *International Journal of Circuit Theory and Applications*, vol. 27, no. 6, pp. 555–567, 1999.

 [7] M. S. Baptista, E. E. Macau, C. Grebogi, Y.-C. Lai, and E. Rosa Jr., "Integrated chaotic communication scheme," *Physical Review E*, vol. 62, no. 4, pp. 4835–4845, 2000.

 [8] L. A. B. Tôrres, "Discrete-time dynamic systems synchronization: information transmission and model matching," *Physica D*, vol. 228, no. 1, pp. 31–39, 2007.

 [9] W. M. Tam, F. C. M. Lau, and C. K. Tse, *Digital Communications with Chaos: Multiple Access Techniques and Performance*, Elsevier Science, New York, NY, USA, 2006.

[10] T. L. Carroll and L. M. Pecora, "Effect of filtering on communication using synchronized chaotic circuits," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '96)*, vol. 3, pp. 174–177, Atlanta, Ga, USA, May 1996.

[11] T. L. Carroll, "Synchronizing chaotic systems using filtered signals," *Physical Review E*, vol. 50, no. 4, pp. 2580–2587, 1994.

[12] T. L. Carroll, "Communicating with use of filtered, synchronized, chaotic signals," *IEEE Transactions on Circuits and Systems I*, vol. 42, no. 2, pp. 105–110, 1995.

[13] E. E. N. Macau and C. M. P. Marinhe, "Communication with chaos over band-limited channels," *Acta Astronautica*, vol. 53, no. 4–10, pp. 465–475, 2003.

[14] M. Eisencraft and M. Gerken, "Comunicação utilizando sinais caóticos: influência de ruído e de limitação em banda," in *Proceedings of the 18th Simpósio Brasileiro de Telecomunicações (SBT '00)*, Gramado, Brazil, September 2000.

[15] Z. Yan, "Q-S synchronization in 3D Hénon-like map and generalized Hénon map via a scalar controller," *Physics Letters A*, vol. 342, no. 4, pp. 309–317, 2005.

[16] Z. Yan, "A nonlinear control scheme to anticipated and complete synchronization in discrete-time chaotic (hyperchaotic) systems," *Physics Letters A*, vol. 343, no. 6, pp. 423–431, 2005.

[17] A. Y. Aguilar-Bustos and C. Cruz-Hernández, "Synchronization of discrete-time hyperchaotic systems: an application in communications," *Chaos, Solitons & Fractals*. In press.

[18] R. P. Agarwal, *Difference Equations and Inequalities: Theory, Methods, and Application*, vol. 155 of *Monographs and Textbooks in Pure and Applied Mathematics*, Marcel Dekker, New York, NY, USA, 1992.

[19] D. L. Hitzl and F. Zele, "An exploration of the Hénon quadratic map," *Physica D*, vol. 14, no. 3, pp. 305–326, 1985.

[20] K. Stefański, "Modelling chaos and hyperchaos with 3-D maps," *Chaos, Solitons & Fractals*, vol. 9, no. 1-2, pp. 83–93, 1998.

[21] A. Oppenheim, R. Schafer, and J. Buck, *Discrete-Time Signal Processing*, Prentice-Hall, Upper Saddle River, NJ, USA, 2nd edition, 1997.

Research Article

Trajectory Sensitivity Method and Master–Slave Synchronization to Estimate Parameters of Nonlinear Systems

Elmer P. T. Cari, Edson A. R. Theodoro, Ana P. Mijolaro, Newton G. Bretas, and Luis F. C. Alberto

Departamento de Engenharia Elétrica da Escola de Engenharia de São Carlos, Universidade de São Paulo 13566-590, São Carlos, SP, Brazil

Correspondence should be addressed to Luis F. C. Alberto, lfcalberto@usp.br

A combination of trajectory sensitivity method and master-slave synchronization was proposed to parameter estimation of nonlinear systems. It was shown that master-slave coupling increases the robustness of the trajectory sensitivity algorithm with respect to the initial guess of parameters. Since synchronization is not a guarantee that the estimation process converges to the correct parameters, a conditional test that guarantees that the new combined methodology estimates the true values of parameters was proposed. This conditional test was successfully applied to Lorenz's and Chua's systems, and the proposed parameter estimation algorithm has shown to be very robust with respect to parameter initial guesses and measurement noise for these examples.

1. Introduction

Several algorithms make use of synchronization to estimate parameters of nonlinear systems based on measured data. In [1, 2], for example, a Lyapunov-based design control and synchronization were used to estimate parameters of nonlinear systems. A similar approach that combines synchronization and geometric control is shown in [3]. Other attempts, including synchronization as an auxiliary tool for parameter estimation, are found in [4–6]. In addition, parameter estimation of delay systems can be found in [7, 8].

The measured output of the real system and the calculated output of an auxiliary system, usually taken as the mathematical model of the real system, are compared. Based on the output mismatch, parameters of the model are updated. When the outputs synchronize, that is, the difference between the real system and the model is sufficiently small, the parameters of the auxiliary system are assumed to be close enough to the real measurements.

However, synchronization is not sufficient to ensure the correct parameter estimation. In [9], for example, the authors follow the same techniques used in [6] and verify using examples that even when the outputs synchronize, the model parameters are far from the real system parameters.

Some studies try to guarantee the correct parameter estimation when the real system and the model outputs synchronize. In [10], for example, the authors prove, for a class of dynamical systems and using a convenient Lyapunov Control Function, that the system globally synchronizes, and synchronization implies convergence to the true parameters if some special conditions are satisfied.

In this paper, master-slave synchronization framework combined with a trajectory sensitivity-based fitting algorithm is used to estimate parameters of nonlinear systems. Trajectory sensitivity method has important characteristics and advantages when compared to other approaches, such as Lyapunov Control Function framework. Trajectory sensitivity-based parameter estimation approach has the advantage of being easily implemented for any nonlinear system while other approaches like Lyapunov-based metholodoly require the existence of a Lyapunov control function. Moreover, Lyapunov-based design may require long time intervals of measured data in order to achieve synchronization while trajectory sensitivity approach can easily deal with short time intervals of measured data [1]. Another interesting feature of trajectory sensitivity approaches is that they provide estimates of the initial conditions (a desired estimate in some problems like the determination of the initial state of population density of a living species in evolutionary studies) that cannot be obtained by the Lyapunov-based approach [1, 11].

In spite of these advantages, it is very difficult to guarantee synchronization and convergence to the true parameters when trajectory sensitivity analysis is used as a fitting algorithm. Usually, these assumptions are not checked, and numerical tests are used to verify the robustness of the algorithm. In general, the estimation fails due to the high relative sensitivities of the trajectories with respect to parameters and initial conditions that leads to a very small convergence region of the algorithm. In this scenario, master-slave synchronization emerges as a good auxiliary tool to increase the robustness of the parameter estimation algorithm with respect to the initial guess of parameters.

Moreover, for this combined approach, there will be proposed a conditional test to ensure convergence to the true parameter values. That is, if the parameter fitting algorithm provides synchronization in a different sense, as it will explained in Section 3.1, then the convergence to the true parameters is guaranteed. Based on the trajectory sensitivity method and master-slave synchronization, some applications to estimate parameters of synchronous generator in electric power system have been developed by the authors [12, 13].

The structure of this paper is as follow: Section 2 presents a general framework to estimate parameters based on measured data and synchronization. In Section 3, there will be exhibited the trajectory sensitivity method. Following that, master-slave synchronization and trajectory sensitivity method will be combined and tested in two nonlinear systems; Sections 3.1 and 3.2. Moreover, the assumptions previously proposed to guarantee convergence to the true parameters will be checked for these examples. The results will be discussed in Section 4.

2. General Framework to Estimate Parameters Based on Measured Data and Synchronization

A common problem that appears in many applications is the necessity of estimating parameters of a system based on the information contained in measured data (time series).

Figure 1 illustrates the traditional framework to estimate parameters of a general nonlinear dynamical system based on measured data. In order to accomplish the estimation, an auxiliary system, which is usually taken as the mathematical model of the real system, is employed. Some of the measurements are chosen as inputs of the auxiliary system while the others are used as outputs for comparison. Using the stored measured data of the inputs, the outputs of the auxiliary system are calculated for an initial guess of parameters. Both the measured and the auxiliary system outputs are compared. Based on the output mismatch, the parameters are updated according to some prescribed rule.

The success of the general framework of Figure 1 to estimate parameters is guaranteed if the following assumptions hold.

(A1) Synchronization between the real system and the auxiliary system outputs should imply parameters of the auxiliary system sufficiently close to parameters of the real system.

(A2) The parameter fitting algorithm provides output synchronization.

Assumption (A1) is a necessary but not enough property, that has to be satisfied to guarantee the correctness of estimates. Assumption (A2) has to be satisfied to guarantee the convergence of the algorithm.

There are many alternatives to design parameter fitting algorithms; however, in most of the practical cases, it is very difficult to prove that assumption (A2) holds. In general, the success of the parameter estimation algorithm depends on personal experience to choose a convenient auxiliary system and an efficient parameter fitting algorithm. Usually, the effectiveness of the algorithms is numerically checked by means of a large number of tests, and their problems are usually related to two main aspects.

(1) The convergence region of the parameter estimation algorithm, that is a subset of the synchronization region, can be very small; that is, if a good guess of parameters is not available, then the parameter estimation algorithm diverges (synchronization is not achieved) or converges to wrong values (assumption (A1) is not satisfied).

(2) They may not be robust with respect to the presence of noise in the measurements (synchronization is not achieved due to noise presence).

In this paper, an unilateral coupling between the real and the auxiliary systems is used to enhance the robustness of synchronization. As a result, parameters of nonlinear systems (including chaotic systems) are correctly estimated, even for bad initial guesses of parameters, in the presence of measurements with noise. The coupling is of unilateral type [14], that is, some of the outputs of the real system are used as inputs to the auxiliary system. This type of coupling is known as master-slave synchronization. In this case, the real system is the master while the auxiliary system is the slave. Figure 2 illustrates the situation.

It is important to emphasize that the authors that use control Lyapunov function approach to adjust parameters have shown conditions to guarantee the satisfaction of condition (A2) for some class of systems [14]. In our case, the adjustment is accomplished by the sensitivity trajectory method using Newton's method; therefore it is not possible to exhibit an analytical condition to guarantee the satisfaction of condition (A2). As a consequence, we pursued another goal; that is, we offered a conditional proof of condition (A2); that is, (A1) is numerically checked, and if synchronization is achieved, then (A2) is true. Although the control Lyapunov approach can provide, for some classes of nonlinear systems, the satisfaction of assumptions (A1) and (A2), the motivation to study theses conditions under
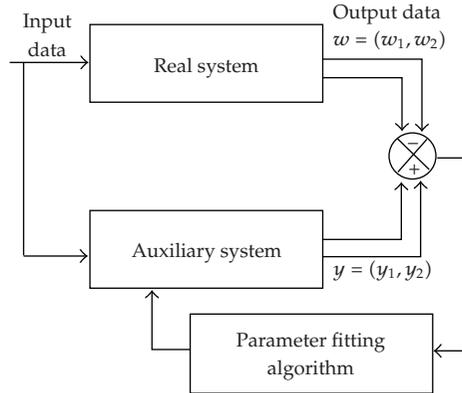
Figure 1: Framework to estimate parameters of nonlinear dynamical systems. Measured outputs of the real system are compared to the outputs of an auxiliary system. Based on the output mismatch, the fitting algorithm updates the parameters. If outputs are synchronized, parameters of the auxiliary system are assumed to be sufficiently close to parameters of the real system model.
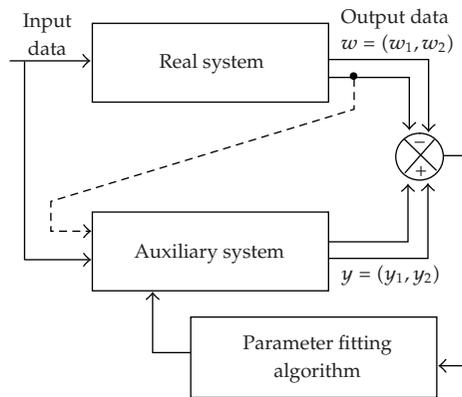
Figure 2: Parameter estimation framework with master-slave coupling between the real and the auxiliary systems. Some outputs of the real system are used as inputs of the auxiliary system.

the sensitivity trajectory method is that this method can be implemented for any nonlinear system in the form of (3.1)-(3.2), even for those we cannot exhibit a Lyapunov function.

In the next sections, this master-slave coupling approach is tested with a trajectory sensitivity-based parameter fitting algorithm.

3. Trajectory Sensitivity Method with Master-Slave Coupling

In this section, trajectory sensitivity analysis and master-slave synchronization will be used to estimate parameters of two chaotic systems. Although assumption (A2) cannot be easily checked, trajectory sensitivity analysis provides, from the practical point of view, advantages that justify its usage.

Trajectory sensitivity method can easily deal with hard nonlinearities. A very interesting extension of sensitivity method for differential algebraical equations (DAE systems) subject to nonsmooth events, like switchings, is presented in [15].

Consider a nonlinear system modeled by

$$\frac{d}{dt}x(t) = f(x(t), p, u(t)),\tag{3.1}$$

$$y(t) = g(x(t), p, u(t)),\tag{3.2}$$

where $x \epsilon R^n$ is the state vector, $y \epsilon R^m$ is the output vector, $u \epsilon R^l$ is the input vector, and $p \epsilon R^k$ is the parameter vector. Functions f and g are nonlinear, continuous, and Lipschitz with respect to x, p, and u. Let p_i be the ith component of p. We assume that f and g are differentiable with respect to every component p_i of p. In case they are not differentiable, a numerical approximation can be used to evaluate these derivatives [16]. The trajectory sensitivities $\partial x(t)/\partial p_i$ and $\partial y(t)/\partial p_i$ are computed, respectively, as

$$\frac{d}{dt}\frac{\partial x(t)}{\partial p_i} = \frac{\partial f(x(t), p, u(t))}{\partial x} \cdot \frac{\partial x(t)}{\partial p_i} + \frac{\partial f(x(t), p, u(t))}{\partial p_i},$$

$$\frac{\partial y(t)}{\partial p_i} = \frac{\partial g(x(t), p, u(t))}{\partial x} \cdot \frac{\partial x(t)}{\partial p_i} + \frac{\partial g(x(t), p, u(t))}{\partial p_i}.\tag{3.3}$$

Trajectory sensitivities quantify the variation of a trajectory with respect to small variations in parameters. This quantification is used to update the model parameters in order to minimize the distance between the outputs of the real system and the mathematical model.

The parameter fitting algorithm is formulated as an optimization problem; that is, we try to minimize the error function J, which is given by

$$J(p) = \frac{1}{2}\int_0^{T_o}(y - w)^t(y - w)dt,\tag{3.4}$$

where, w is the output vector of the real system, y is the output of the auxiliary system (3.2), and $[0, T_o]$ is the time interval considered in the analysis. In other words, the minimization searches for output synchronization.

Given an initial value $p = p^{(o)}$, this optimization problem can be solved computing the sensitivity $\partial J(p)/\partial p$ and using a least squares method:

$$G(p) = \frac{\partial J(p)}{\partial p} = \int_0^{T_o}\frac{\partial y^t}{\partial p}(y - w)dt\bigg|_{p=p^{(i)}}.\tag{3.5}$$

Expanding $G(p)$ in Taylor series around $p = p^{(i)}$ and neglecting high-order terms one has

$$G(p) \approx G\left(p^{(i)}\right) + \Gamma\Delta p,\tag{3.6}$$

where $\Gamma(p) = \partial G(p)/\partial p$. Then $\Delta p = -\Gamma^{-1} G(p^{(i)})$. Parameters are fitted for the ith iteration by $p^{(i+1)} = p^{(i)} + \Delta p^{(i+1)}$. Matrix $\Gamma(p)$ can be computed by differentiating (3.5) and neglecting higher-order terms, thus

$$\Gamma(p) \approx \int_0^T \frac{\partial y}{\partial p}^t \frac{\partial y}{\partial p} dt \bigg|_{p=p^{(i)}} \tag{3.7}$$

When the measurements are sampled at discrete time intervals, the previous integrals are replaced by summations. For more details see reference [17].

The error function indirectly depends on the parameters. Its evaluation requires the solution of a set of ordinary differential equations (the model). Usually, nonlinear dynamic models do not have closed analytical solutions; that is, the evaluation of the error function is made via numerical integration algorithms. Newton's method, which is used to achieve synchronization, demands the evaluation of trajectory sensitivity equations that are also obtained via numerical integration algorithms. As far as we know, there is no general condition to guarantee convergence of Newton's algorithm (or similar one) in this case.

The use of trajectory sensitivities in the framework described in Figure 1 has two main problems: (i) the trajectory sensitivity method is very sensitive to initial conditions (first parameter guess), and (ii) parameters with very low sensitivities (as compared to other parameters) are not numerically identifiable due to ill-conditioned calculations. These problems become worst especially when many parameters have to be simultaneously estimated and/or chaotic behavior is present. Very often, the trajectory sensitivity method leads to erroneous estimations or even divergence of the numerical algorithm.

In this section, a master-slave coupling and trajectory sensitivity-based approach are combined to estimate parameters of chaotic systems. Comparisons between the traditional trajectory sensitivity approach (presented in Figure 1) and the proposed methodology, which includes a master-slave coupling to enhance synchronization robustness, are made in order to show the advantages of the proposed methodology.

3.1. Lorenz's System

Consider Lorenz's system as the real system,

$$\dot{x}_1 = -\sigma_r x_1 + \sigma_r x_2,$$
$$\dot{x}_2 = -x_2 - x_1 x_3 + r_r x_1, \tag{3.8}$$
$$\dot{x}_3 = -b_r x_3 + x_1 x_2,$$

and suppose that at least two states can be measured. Let $w = (x_1, x_2)^T$ be the output vector of the real system.

In the traditional trajectory sensitivity approach, the following auxiliary system would be chosen to estimate parameters of Lorenz's system:

$$\dot{z}_1 = -\sigma z_1 + \sigma z_2,$$

$$\dot{z}_2 = -z_2 - z_1 z_3 + r z_1, \qquad (3.9)$$

$$\dot{z}_3 = -b z_3 + z_1 z_2.$$

This system has exactly the same structure as Lorenz's system model; the only difference is that the state variables x_i, $i = 1, 2, 3$ were replaced by z_i, $i = 1, 2, 3$, and the parameters do not have the subindex r used for the real parameters. Assuming that measurements are subject to random noise, the initial conditions of states z_i, $i = 1, 2, 3$ are unknown and have to be estimated. Thus, the extended parameter vector is $p_r = (\sigma_r, r_r, b_r, z_{1o}, z_{2o}, z_{3o})$.

Our experience shows that Lorenz's system parameters cannot be simultaneously estimated using the traditional trajectory sensitivity methodology even for very small displacements of the initial conditions and parameter values from real values. Nonlinear systems with chaotic behavior, such as Lorenz's system, have a very small convergence region, because of their high relative output sensitivities with respect to the parameters.

Figure 3 shows the sensitivities of output 1 with respect to Lorenz's system parameters. The sensitivity of output z_1 with respect to parameter σ is very small when compared with sensitivities of the z_1 output with respect to parameter b or r. Such behavior makes the approach of Figure 1 inappropriate for parameter estimation of chaotic systems. In the particular Lorenz's system case, errors as small as $\pm 1\%$ in the initial parameter and initial condition guesses lead the algorithm to nonconvergence due to lack of synchronization between the real system and the auxiliary system.

A master-slave coupling between the real system and the auxiliary system will be used to enhance the numerical robustness of the trajectory sensitivity-based parameter estimation algorithm. For this purpose, the next auxiliary system will be employed:

$$\dot{z}_1 = -\sigma z_1 + \sigma z_2, \qquad (3.10)$$

$$\dot{z}_2 = -z_2 - x_1 z_3 + r x_1, \qquad (3.11)$$

$$\dot{z}_3 = -b z_3 + x_1 z_2, \qquad (3.12)$$

where $z = (z_1, z_2, z_3)^T$ is the state space vector. In this case, the states will be taken as the outputs of the auxiliary system, that is, $y = (z_1, z_2, z_3)^T$. The auxiliary system (3.10) to (3.12) resembles Lorenz's system; the difference is that state variable z_1 was replaced by the coupling variable x_1 in some convenient positions, in order to decrease nonlinearities of the auxiliary system.

The replacing of the coupling variable is accomplished in order to eliminate the nonlinear terms in the difference system (3.13). The substitution of the coupling variable x_1 in (3.10) does not have any influence in the estimate process, for being a linear term. However, we still do not have a systematic procedure to determine the best terms to substitute.
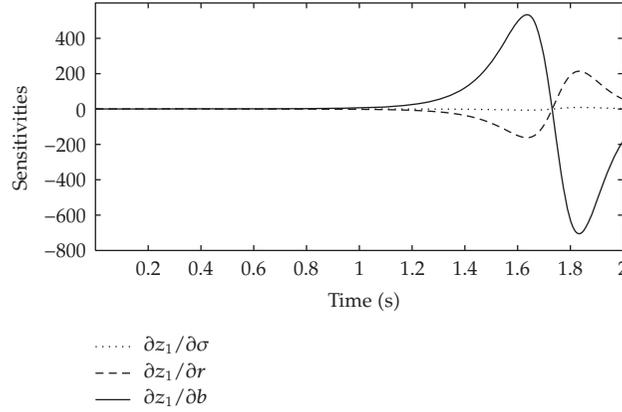
Figure 3: Z_1 output sensitivity (Lorenz's system) with respect to the parameters σ, r, and b.
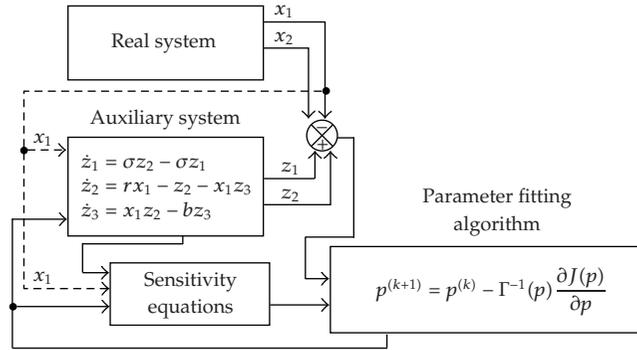
Figure 4: Master-slave trajectory sensitivity-based framework for parameter estimation of Lorenz's system parameters.

Although there is no rules in how to do this replacement, our experience shows that these modifications enlarge the stability region of the parameter estimation algorithm. Figure 4 illustrates the schematic diagram with this approach.

The error $e_i = z_i - x_i$, $i = 1, 2, 3$ is derived from the difference between the outputs of both the auxiliary and the real systems:

$$
\begin{aligned}
\dot{e}_1 &= -\sigma e_1 + \sigma e_2 + (x_2 - x_1)\sigma_e, \\
\dot{e}_2 &= -e_2 - x_1 e_3 + x_1 r_e, \\
\dot{e}_3 &= x_1 e_2 - b e_3 - x_3 b_e,
\end{aligned}
\tag{3.13}
$$

where $\sigma_e = \sigma - \sigma_r$, $r_e = r - r_r$, and $b_e = b - b_r$ are the error of parameter estimation.

It is very difficult to prove assumption (A2). However, it will be shown that assumption (A1) is satisfied, for this choice of auxiliary system, in a different sense of synchronization.

Table 1: Parameter estimation of Lorenz's system using trajectory sensitivity method including master-slave coupling with noise in the measurements.

Parameter	Initial value	Deviation (%)	Estimated value	True value	Error (%)
σ	4.60	−54%	10.034	10.0	0.34
r	9.20	−54%	19.96	20.0	−0.2
b	4.10	+54%	2.66	2.66	0.00
z_{1o}	1.90	+5%	2.07	2.00	3.5
z_{2o}	2.85	+5%	2.94	3.00	2.0
z_{3o}	3.22	−54%	6.96	7.00	−0.57

Definition 3.1. The outputs of the real system $w(t)$ and the auxiliary system $y(t)$ C^1-synchronize in the interval $[T_a, T_b]$ with precision ε if

$$\sup_{T_a \le t \le T_b} \|w(t) - y(t)\| + \sup_{T_a \le t \le T_b} \|\dot{w}(t) - \dot{y}(t)\| < \varepsilon. \tag{3.14}$$

It is possible to prove, using (3.13) and The Implicit Function Theorem, that C^1-synchronization, in the sense of Definition 3.1, implies that the estimated parameters are ε-close to the real parameters.

Although assumption (A2) cannot be easily proven, it can be easily checked at the end of iterations. Moreover, the proposed master-slave coupling brings more robustness to synchronization. Lorenz's system parameters were estimated with good accuracy even for cases where the first guess of parameter values was displaced up to ±54% from the real value. Table 1 shows the estimation results under the presence of a ±5% white Gaussian random noise in all the measurements.

The outputs of the real and the auxiliary system at the beginning and at the end of iterations are shown in Figure 5.

3.2. Chua's System

Chua's circuit is a singular example of chaotic system, because it is the simplest circuit that exhibits this kind of phenomenon. It is composed only by one inductor, two capacitors, one resistor, and one nonlinear active resistor with a three-segment picewise-linear volt-current (V-I) characteristics, called "Chua's Diode," as shown in Figure 6.

Let the dimensionless equations of Chua's circuit [18] represent the real system:

$$\dot{x}_1 = \alpha_r(x_2 - x_1 - f(x_1)), \tag{3.15}$$

$$\dot{x}_2 = x_1 - x_2 + x_3, \tag{3.16}$$

$$\dot{x}_3 = -\beta_r x_2 - \gamma_r x_3, \tag{3.17}$$

$$f(x_1) = bx_1 + \frac{1}{2}(a - b)\{|x_1 + 1| - |x_1 - 1|\}, \tag{3.18}$$

where α, β, γ, a, and b are the parameters to be estimated. We assume that states are measured; that is, $w = (x_1, x_2, x_3)^T$ is the output vector of the real system. Assuming that the
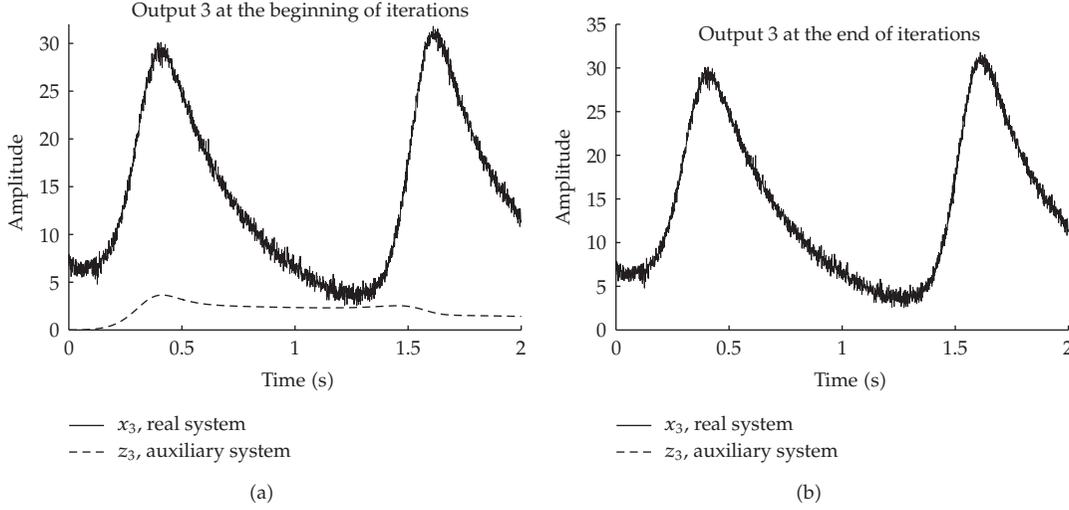
Figure 5: Comparison between the outputs of the auxiliary system at the beginning and at the end of iterations for output number 3. A white Gaussian noise with zero mean and standard deviation of 5% of the highest measured value was applied to all measurements.

initial conditions of differential equations of the model are unknown and they also have to be estimated. Thus, the extended parameter vector is $p = (\alpha, \beta, \gamma, a, b, x_{1o}, x_{2o}, x_{3o})$, whose true values are $\alpha_r = 6.5792$, $\beta_r = 10.9024$, $\gamma_r = -0.0445$, $a_r = -1.1829$, $b_r = -0.6524$, $x_{1o} = 0.15$, $x_{2o} = 0.90$, and $x_{3o} = 0.80$. Measurements of the real system were obtained by numerical integration of (3.15) to (3.18), with their true parameters and initial conditions. Random white Gaussian noise of ±5% of the peak value was added to the real measurements.

Like Lorenz's system, parameters of Chua's circuit cannot be simultaneously estimated using the traditional trajectory sensitivity approach of Figure 1 due to high relative sensitivity of trajectories with respect to parameters and initial conditions. Our experience shows that errors as small as ±1% on the initial parameter guesses lead the algorithm to nonconvergence due to lack of synchronization.

In order to overcome this difficulty, the following auxiliary system was chosen:

$$\dot{z}_1 = \alpha(z_2 - x_1 - f(x_1)) - k(z_1 - x_1),$$

$$\dot{z}_2 = x_1 - z_2 + z_3,$$

$$\dot{z}_3 = -\beta z_2 - \gamma z_3, \tag{3.19}$$

$$f(x_1) = bx_1 + \frac{1}{2}(a - b)\{|x_1 + 1| - |x_1 - 1|\}.$$

The auxiliary system resembles Chua's circuit model; the difference is that a master-slave coupling between the real and the auxiliary systems was employed; that is, state variable z_1 was replaced by the measured real variable x_1 in some convenient positions to reduce nonlinearities in the auxiliary system. Moreover, to enhance synchronization robustness, an extra term was added to (3.15), $g(z_1, x_1) = -k(z_1 - x_1)$, where $k = 10$. Figure 7 illustrates this parameter estimation scheme.
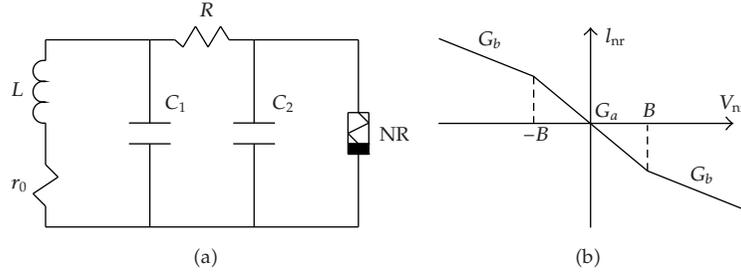
Figure 6: Chua's circuit and the three-segment odd-symmetric picewise-linear volt-current (V-I) characteristic of Chua's diode.
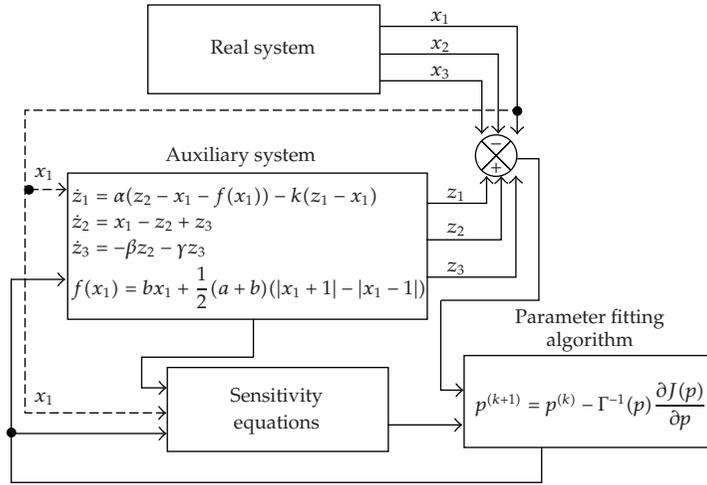
Figure 7: Master-slave trajectory sensitivity based framework for parameter estimation of Chua's circuit.

The error $e_i = z_i - x_i$, $i = 1, 2, 3$ is derived from the difference between the outputs of both the auxiliary and the real systems:

$$\dot{e}_1 = -be_1 + \alpha e_2 + \alpha_e(x_2 - x_1) - x_1(\alpha b_e - b_r \alpha_e)$$

$$- \frac{1}{2}[\alpha_e(a_r - b_r) - \alpha(a_e - b_e)]\{|x_1 + 1| - |x_1 - 1|\},$$

$$\dot{e}_2 = -e_2 + e_3,$$

$$\dot{e}_3 = -\beta e_2 - \gamma e_3 - x_3 \gamma_e - x_2 \beta_e,$$

$$(3.20)$$

where $\alpha_e = \alpha - \alpha_r$, $\beta_e = \beta - \beta_r$, $\gamma_e = \gamma - \gamma_r$, $a_e = a - a_r$, and $b_e = b - b_r$ are the error of parameter estimation.

It is practicable proved that $\mathcal{C}^1$-synchronization of outputs, in the sense of Definition 3.1, implies that parameters of the auxiliary system $p = (\alpha, \beta, \gamma, a, b)$ are ε-close

Outputs in the first iteration

(a)
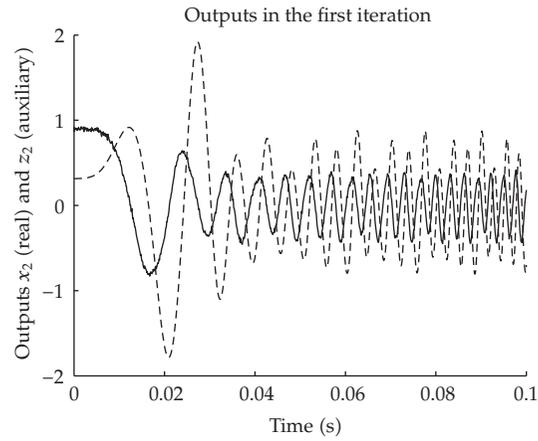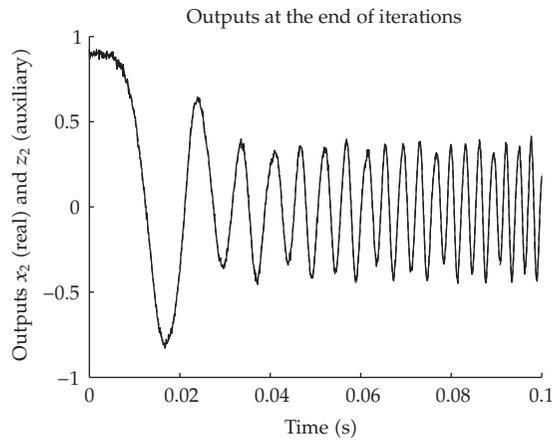
Outputs at the end of iterations

(b)

Figure 8: Comparison between outputs of real and auxiliary systems at the beginning and at the end of iterations for the output number 2. A white Gaussian noise with mean zero and standard deviation of 5% of the highest measurement value was applied to all measurement.

to the real parameter values $p_r = (\alpha_r, \beta_r, \gamma_r, a_r, b_r)$; that is, assumption (A1) is satisfied. For this purpose, the Implicit Function Theorem and (3.20) must be used.

Although assumption (A2) cannot be easily proven, it can be checked at the end of the iterations. Chua's circuit parameters were successfully estimated even for parameter initial guess deviations as large as 65% from the real values. Table 2 shows the estimation results.

The output of the real and the auxiliary system at the beginning and at the end of iterations are shown in Figure 8.

Table 2: Parameter estimation of Chua's system using trajectory sensitivity method including master-slave coupling with noise in the measurements.

Parameter	Initial value	Deviation	Estimated value	True value	Error (%)
α	2.3027	−65%	6.5435	6.5792	0.54
β	3.8158	−65%	10.8906	10.9024	0.11
γ	−0.0156	−65%	−0.0444	−0.0445	0.27
a	−0.4137	−65%	−1.1807	−1.1820	0.12
b	−0.2283	−65%	−0.6539	−0.6524	0.24
x_0	0.0525	−65%	0.1542	0.1500	2.83
y_0	0.3150	−65%	0.9009	0.9000	0.10
z_0	0.2800	−65%	0.8185	0.8000	2.31

4. Conclusions

A combination of trajectory sensitivity method and master-slave synchronization was proposed to parameter estimation of nonlinear systems. It was shown that master-slave coupling increases the robustness of the trajectory sensitivity algorithm with respect to the initial guess of parameters. Since synchronization is not a guarantee that the estimation process converges to the correct parameters, a conditional test that guarantees the new combined methodology estimates that the true values of parameters was proposed. This conditional test was successfully applied to Lorenz's and Chua's systems and the proposed parameter estimation algorithm has shown to be very robust with respect to parameter initial guesses and measurement noise for these examples.

Acknowlegments

References

[1] D. Huang and R. Guo, "Identifying parameter by identical synchronization between different systems," *Chaos*, vol. 14, no. 1, pp. 152–159, 2004.

[2] R. Konnur, "Synchronization-based approach for estimating all model parameters of chaotic systems," *Physical Review E*, vol. 67, no. 2, Article ID 027204, 2003.

[3] U. S. Freitas, E. E. N. Macau, and C. Grebogi, "Using geometric control and chaotic synchronization to estimate an unknown model parameter," *Physical Review E*, vol. 71, no. 4, Article ID 047203, 4 pages, 2005.

[4] Y. Crispin, "A fluid dynamical approach to the control, synchronization and parameter identification of chaotic systems," in *Proceedings of the American Control Conference*, vol. 3, pp. 2245–2250, May 2002.

[5] P. M. Ines and M. Joaquin, "Adaptative approximation method for joint parameter estimation and identical synchronization of chaotic systems," *Physical Review E*, vol. 72, Article ID 057202, 2005.

[6] U. Parlitz, "Estimating model parameters from time series by autosynchronization," *Physical Review Letters*, vol. 76, no. 8, pp. 1232–1235, 1996.

[7] D. Ghosh and S. Banerjee, "Adaptive scheme for synchronization-based multiparameter estimation from a single chaotic time series and its applications," *Physical Review E*, vol. 78, no. 5, Article ID 056211, 2008.

[8] D. Ghosh, "Nonlinear-observer-based synchronization scheme for multiparameter estimation," *Europhysics Letters*, vol. 84, no. 4, Article ID 40012, 4 pages, 2008.

[9] L. Li, H. Peng, X. Wang, and Y. Yang, "Comment on two papers of chaotic synchronization," *Physics Letters A*, vol. 333, no. 3-4, pp. 269–270, 2004.

[10] W. Yu, G. Chen, J. Cao, J. Lü, and U. Parlitz, "Parameter identification of dynamical systems from time series," *Physical Review E*, vol. 75, Article ID 067201, 4 pages, 2007.

[11] M. R. Gameiro and H. M. Rodrigues, "Applications of robust synchronization to communication systems," *Applicable Analysis*, vol. 79, no. 1-2, pp. 21–45, 2000.

[12] E. P. T. Cari, L. F. C. Alberto, and N. G. Bretas, "A methodology for parameter estimation of synchronous generators based on trajectory sensitivity and synchronization technique," in *Proceedings of IEEE Power Engineering Society General Meeting (PES '06)*, June 2006.

[13] E. P. T. Cari, L. F. C. Alberto, and N. G. Bretas, "Metodologia para a estimao de parmetros baseada na tcnica de sensibilidade de trajetria e conceitos de sincronizao," in *Proceedings of the Anais16th Congresso Brasileiro de Automtica*, October 2006.

[14] L. M. Pecora and T. L. Carroll, "Synchronization in chaotic systems," *Physical Review Letters*, vol. 64, no. 8, pp. 821–824, 1990.

[15] I. A. Hiskens, "Nonlinear dynamic model evaluation from disturbance measurements," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 702–710, 2001.

[16] S. M. Benchluch and J. H. Chow, "A trajectory sensitivity method for the identification of nonlinear excitation system models," *IEEE Transaction on Energy Conversion*, vol. 8, pp. 159–164, 1993.

[17] J. B. Cruz, *Feedback System*, McGraw-Hill, New York, NY, USA, 1972.

[18] L. O. Chua, C. W. Wu, A. Huang, and G. Zhong, "A universal circuit for studying and generating chaos—part I: routes to chaos," *IEEE Transactionson Circuits and Systems*, vol. 40, no. 10, pp. 732–744, 1993.

Research Article

Adaptive Step-Size Control in Simulation of Diffusive CVD Processes

Jürgen Geiser and Christian Fleck

Department of Mathematics, Humboldt-Universität zu Berlin, Unter den Linden 6,
D-10099 Berlin, Germany

Correspondence should be addressed to Jürgen Geiser, geiser@mathematik.hu-berlin.de

We present control strategies of a diffusion process for chemical vapor deposition for metallic bipolar plates. In the models, we discuss the application of different models to simulate the plasma-transport of chemical reactants in the gas-chamber. The contribution are an optimal control problem based on a PID control to obtain a homogeneous layering. We have taken into account one- and two-dimensional problems that are given with constraints and control functions. A finite-element formulation with adaptive feedback control for time-step selection has been developed for the diffusion process. The optimization is presented with efficient algorithms. Numerical experiments are discussed with respect to the diffusion processes of the macroscopic model.

1. Introduction

We motivate our studying on simulating a low-temperature low-pressure plasma that can be found in chemical vapor deposition (CVD) processes. In the last years the research and optimization in producing high-temperature films by depositing low-pressure processes have increased by using simulation tools, see [1, 2]. Theoretical models exist for deposition processes and can be modeled by coupled transport and flow equations, see [3, 4]. Further interest on standard applications to deposit titanium-nitrogen (TiN) and titanium-carbon (TiC) on metallic layers are immense, see [5]. Recently more and more focus on deposition with new material classes known as MAX-phases are becoming important, see [6, 7]. The MAX-phase are nanolayered terniar metal-carbides or -nitrids, where M is a transition metal, A is an A-group element (e.g., Al, Ga, In, Si, etc.) and X is C (carbon) or N (nitride), see [8]. Such materials combine ceramic and metallic behavior and can be implanted in the metallic bipolar plates to obtain a new material with noncorrosive and good metallic conductivity behavior.

We discuss a model for low-temperature and low-pressure plasma that can be used to implant or deposit thin layers of important materials, see [9]. This model is used for the implantation process. First, we present the process in the plasma-reactor that transport the contaminants to the wafer surface, see [10]. We deal with a continuous flow model, while we assume a vacuum- and a diffusion-dominated processes. Second, the process at the wafer-surface is modeled by the heavy particles problem with underlying drift. This model deals more with the atomic behavior and we do not allow $p = 0$, see [10].

To solve such optimization problems, we present a PID-controller (proportional, integral, differential) to control our deposition process, see [11]. We improved heuristic methods of deriving the PID parameters, while we discuss the posteriori error estimates respecting the time-step size control. Our contribution is a modified automatically step-size control and a best approximation is obtained with the time-dependent control method based on the Chien-Hrones-Reswick algorithm, see [12].

Numerical methods are described in the context of time- and spatial-discretization methods for the mesoscopic-scale model. We discussed different experiments and their convergence rates.

For the simulations we apply analytical and also numerical methods to obtain results to control the grow of thin layers.

The paper is outlined as follows. In Section 2 we present our mathematical model and a possible reduced model for the further approximations. In Section 3 we discuss the theoretical background for the simulation of CVD processes. The optimal control and their control paths based on the PID-control approach are discussed in Section 4. The software and program-tools are discussed in Section 5. The numerical experiments are given in Section 6. In the contents that are given in Section 7, we summarize our results.

2. Mathematical Model

In the following, the models are discussed in two directions of far-field and near-field problems:

(1) reaction-diffusion equations, see [13] (far-field problem);

(2) Boltzmann-Lattice equations, see [9] (near-field problem).

The modeling is considered by the Knudsen Number (Kn), which is the ratio of the mean free path λ over the typical domain size L. For small Knudsen numbers Kn $\approx 0.01 - 1.0$, we deal with a Navier-Stokes equation or with the convection-diffusion equation, see [5, 14], whereas for large Knudsen numbers Kn ≥ 1.0, we deal with a Boltzmann equation, see [4].

2.1. Modeling with Partial Differential Equations

Dynamic processes with modifications in time and space will be reshaped by partial differential equations. There is (i) the PDE-formula itself which describes the physical laws of nature that influence the process and (ii) initial and boundary conditions in which specific characteristics of the process, like boundary behavior, can be coded.

There are two types of boundary conditions, namely, Dirichlet and Neumann boundary. With the Dirichlet type the exact value of the boundary is known, however, with
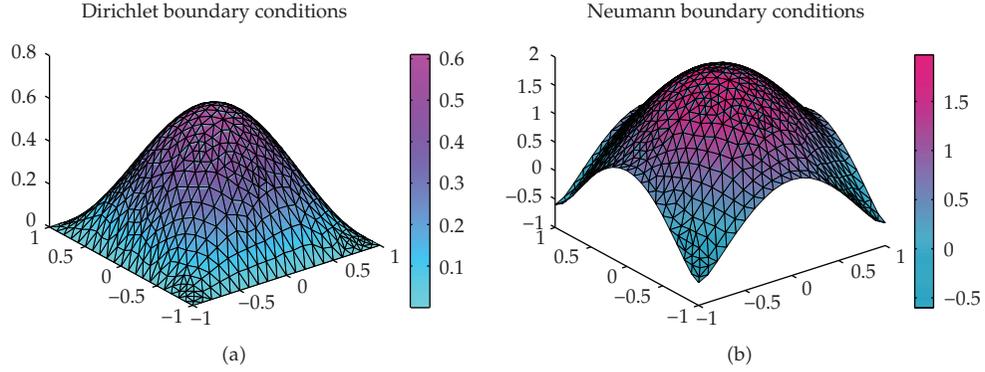
Figure 1: Dirichlet and Neumann boundary conditions.

Neumann boundaries the spatial derivation of the boundary values in normal direction is known, see an example of the boundary conditions in Figure 1.

2.2. Model for Optimal Control of the Layer

We will concentrate us on a continuum model of mass transportation and assume that the energy and momentum is conserved, see [13]. Therefore, the continuum flow of the mass can be described as diffusion reaction equation given as

$$\partial_t c - \nabla D \nabla c - R_g = 0, \quad \text{in } \Omega \times [0,T], \tag{2.1}$$

$$c(x,0) = c_0(x), \quad \text{on } \Omega, \tag{2.2}$$

$$\frac{\partial c(x,t)}{\partial n} = c_1(x,t), \quad \text{on } \partial\Omega \times [0,T], \tag{2.3}$$

where c is the molar concentration, D is the diffusion parameter, and R_g is the reaction and source term.

We modify our model equation (2.1) to a control problem with an additionally right-hand side source:

$$\partial_t c - \nabla D \nabla c = c_{\text{source}}, \quad \text{in } \Omega \times [0,T],$$
$$c(x,0) = c_0(x), \quad \text{on } \Omega,$$
$$\frac{\partial c(x,t)}{\partial n} = c_1(x,t), \quad \text{on } \partial\Omega \times [0,T], \tag{2.4}$$

where $c_{\text{source}}(x,t) = R_g$ is the discontinuous or continuous source term of the concentration c and we neglect a reaction term of this concentration.

We assume an optimal concentration at the layer

$$c_{\text{opt}}(x,t), \tag{2.5}$$

Orthogonal triangle

Equilateral triangle

(a)

(b)
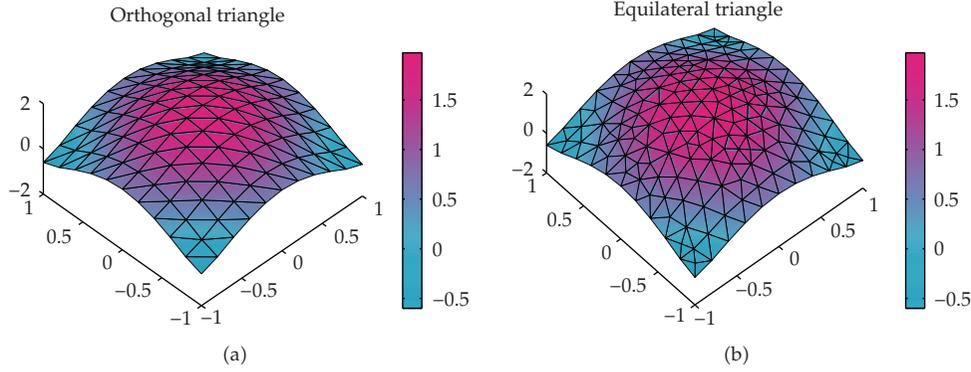
Figure 2: Spatial-discretization.

where the layer is given as $x \in \Omega_{\text{layer}}$ and our constraints are given as

$$c_{\text{source,min}} \leq c_{\text{source}} \leq c_{\text{source,max}}. \tag{2.6}$$

Additionally, we have to solve the minimization problem:

$$\min(J(c, c_{\text{source}})) := \frac{1}{2} \int_T \int_{\Omega_{\text{layer}}} |c(x,t) - c_{\text{opt}}(x,t)|^2 dx\, dt + \frac{\lambda}{2} \int_T \int_{\Omega} |c_{\text{source}}(x,t)|^2 dx\, dt, \tag{2.7}$$

where T is the time period of the process.

Remark 2.1. We choose the L_2-error to control our minimization problem. In literature, see [15, 16], there exists further control-errors, which respect the time behavior.

In a first part, we only solve the transport equation with UG software-tool (unstructed grid software, see [17]) and try to find out the optimal control of the sources to obtain the best homogeneous layer.

In a second part, we consider the optimal control problem and solve also the backward problem.

3. Theoretical Background for Simulation of Diffusive CVD Processes

In what follows we discuss the approximation methods and errors for the simulation of the CVD processes.

3.1. Approximation and Discretization

For the numerical solutions we need to apply approximation methods, for example, finite-difference methods and iterative solver methods for the nonlinear differential equations, see [18, 19].

The finite-element discretization is based on Ω_h the variational boundary value problem reduces to find $u_h \in V_h$ satisfying the initial condition $u_h(0)$ such that

$$\int_{\Omega_h} \left(\frac{\partial u_h}{\partial t} v_h + D \nabla u_h \cdot \nabla v_h \right) dx = 0, \quad \forall v_h \in V_h. \tag{3.1}$$

We define the minimal length of triangle which we get from the spatial-discretization with Δx.

This leads to the following linear semidiscretized system of ordinary differential equations:

$$M \frac{du^*}{dt} + A u^* = 0, \tag{3.2}$$

where M is the mass and A the M-matrix.

Here we have taken into account the Courant-Friedrichs-Levy- (CFL-) condition, which is given as

$$CFL = 2 D_{\max} \frac{\Delta t}{\min_{e \in E} \Delta x_e^2}, \tag{3.3}$$

where $D_{\max}$ is the maximal diffusion parameter, E is the set of the edges of the discretization. We restrict the CFL-condition to 1, if we use an explicit time-discretization and can lower the condition, if we use an implicit discretization.

For the explicit time-discretization, we apply explicit Euler or Runge-Kutta methods.

We use the explicit lower-order Runge-Kutta methods:

$$
\begin{array}{c|cc}
0 & & \\
\frac{1}{2} & \frac{1}{2} & \\
\hline
& 0 & 1
\end{array}
\tag{3.4}
$$

Furthermore we use the following Heun method (third-order):

$$
\begin{array}{c|ccc}
0 & & & \\
\frac{1}{3} & \frac{1}{3} & & \\
\frac{2}{3} & 0 & \frac{2}{3} & 0 \\
\hline
& \frac{1}{4} & 0 & \frac{3}{4}
\end{array}
\tag{3.5}
$$

The implicit time-discretization is done with implicit Euler or Runge-Kutta methods.

Here, we use the implicit trapezoidal rule:

$$
\begin{array}{c|cc}
0 & & \\
1 & \dfrac{1}{2} & \dfrac{1}{2} \\[2mm]
\hline
& \dfrac{1}{2} & \dfrac{1}{2}
\end{array}
\tag{3.6}
$$

Furthermore we use the following Gauss Runge-Kutta method:

$$
\begin{array}{c|cc}
\dfrac{1}{2}-\dfrac{\sqrt{3}}{6} & \dfrac{1}{4} & \dfrac{1}{4}-\dfrac{\sqrt{3}}{6} \\[3mm]
\dfrac{1}{2}+\dfrac{\sqrt{3}}{6} & \dfrac{1}{4}+\dfrac{\sqrt{3}}{6} & \dfrac{1}{4} \\[3mm]
\hline
& \dfrac{1}{2} & \dfrac{1}{2}
\end{array}
\tag{3.7}
$$

Remark 3.1. We apply implicit time-discretization methods for the pure diffusion part, where we apply explicit time-discretization methods for the pure convection part. Here we have to respect the CLF-condition, see [20].

3.2. Errors and Convergence Rate

For studying the errors and the convergence-rates in our test example, we have to define the following norm in two space-dimensions:

(i) discrete $L_{\max}$-norm:

$$
\mathrm{err}_{L_{\max},\Delta x,\Delta t} = \max_{i=1}^{p}\left|c_{\mathrm{num}}(x_i,T) - c_{\mathrm{ref}}(x_i,T)\right|,
\tag{3.8}
$$

(ii) discrete L_1-norm:

$$
\mathrm{err}_{L_1,\Delta x,\Delta t} = \sum_{i=1}^{p}\Delta x^2\left|c_{\mathrm{num}}(x_i,T) - c_{\mathrm{ref}}(x_i,T)\right|,
\tag{3.9}
$$

(iii) discrete L_2-norm:

$$
\mathrm{err}_{L_2,\Delta x,\Delta t} = \sqrt{\sum_{i=1}^{p}\Delta x^2\left|c_{\mathrm{num}}(x_i,T) - c_{\mathrm{ref}}(x_i,T)\right|^2},
\tag{3.10}
$$

where Δx is the spatial-step of the discretization, Δt is the time-step of the discretization, and T is the end-time of the computation. p is the number of grid points in the discretization

method. c_{num} is the numerical solution and c_{ref} is the reference solution, computed at fine spatial- and time-grids.

The numerical convergence rate are given as follows.

(i) For the spatial error, we define

$$\rho_{L_2,\Delta x_1,\Delta x_2,\Delta t} = \frac{\log\left(\mathrm{err}_{L_2,\Delta x_1,\Delta t}/\mathrm{err}_{L_2,\Delta x_2,\Delta t}\right)}{\log(\Delta x_1/\Delta x_2)}, \tag{3.11}$$

where Δx_1 is the coarse, Δx_2 is the fine spatial grid-step, and Δt is the time-grid step for both results.

(ii) For the time error, we define

$$\rho_{L_2,\Delta x,\Delta t_1,\Delta t_2} = \frac{\log\left(\mathrm{err}_{L_2,\Delta x,\Delta t_1}/\mathrm{err}_{L_2,\Delta x,\Delta t_2}\right)}{\log(\Delta t_1/\Delta t_2)}, \tag{3.12}$$

where Δt_1 is the coarse, Δt_2 is the fine-time-step, and Δx is the spatial-grid step for both results.

We often use $\Delta x_2 = \Delta x_1/2$. In this case, we have $\rho_{L_2,\Delta x_1,\Delta x_2,\Delta t} = \rho_{L_2,\Delta x_1,\Delta t}$. Further we have to choose Δx_1 with respect to the $\Delta t \in I = [0, \Delta t_{max}]$, which have maximal ρ. Thus we define $\mathrm{Arg\,Max}(\Delta x)$:

$$\mathrm{Arg\,Max}(\Delta x) := \underset{\Delta t \in I}{\arg\max} \, \rho_{L_2,\Delta x,\Delta x/2,\Delta t}. \tag{3.13}$$

4. Optimal Control Methods

Here we discuss the control of a diffusion equation with a feedback based on a PID-controller.

4.1. Forward Controller (Simple P-Controller)

The first controller we discuss is the simple P-controller, see [11]. A first idea is to control linearly the error of the solved PDE.

In Figure 3, we present the P-controller.

Our control problem is given with the control of the error to the optimal concentration of the layer and correct the source-flux:

$$\begin{aligned}
\partial_t c - \nabla D \nabla c &= c_{source}, && \text{in } \Omega \times [0,T], \\
c(x,0) &= c_0(x), && \text{on } \Omega, \\
\frac{\partial c(x,t)}{\partial n} &= c_1(x,t), && \text{on } \partial\Omega \times [0,T],
\end{aligned} \tag{4.1}$$

where $c_{source}(x,t)$ is a discontinuous source flow of the concentration c.

We assume an optimal concentration at the layer with the concentration $c_{opt}(x,t)$, where the layer is given as $x \in \Omega_{layer}$ and 0 elsewhere.
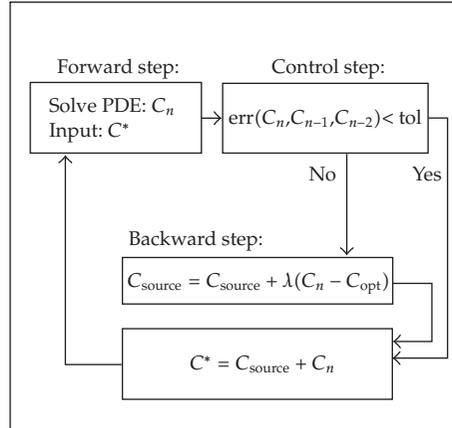
Figure 3: P-controller for the solution C.
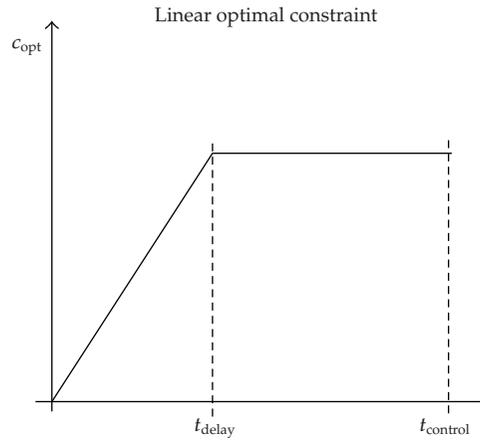
Figure 4: Linear constraint c_{opt} for the deposition process. x-axis: Time, y-axis: c (concentration).

Our constraints are bounded as

$$c_{source,min} \leq c_{source} \leq c_{source,max}. \tag{4.2}$$

Remark 4.1. Taken into account the hysteresis of the deposition process, we apply a linear increase of $c_{source,max}$ in the optimal control with respect to time, see Figure 4.

4.2. PID-Controller

The PID-controller is used to control temperature, motion, and flow. The controller is available in analog and digital forms, see [16]. The aim of the controller is to get the output (velocity, temperature, position) in the area of the constraint output, in a short time, with minimal overshoot, and with small errors.

$$K_P \cdot \text{err}(T)$$

$$K_I \int_0^T \text{err}(t)\,dt$$

$$K_D \frac{d\ \text{err}(t)}{dt}\,[T]$$

Structure of PID-control
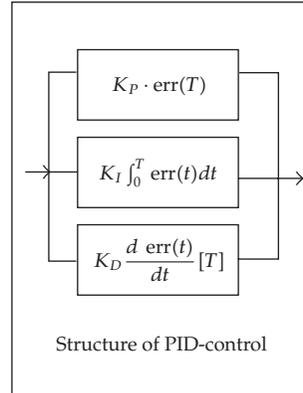
Figure 5: PID-control. Effect on control system: the main influence in a control loop, K_P reduces a large part of the overall error. K_I reduces the final error in a system. Summing even a small error over time produces a drive signal large enough to move the system toward a smaller error. K_D counteracts the K_P and K_I terms when the output changes quickly. This helps reduce overshoot and ringing.

We have three elements in the PID-control, where P is the proportional part, I the integral part, and D is the derivative part of the controller, see [16].

These terms describe three basic mathematical functions applied to the error signal, error $= C_{\text{optimal}} - C_{\text{computed}}$.

The errors represented the difference between constraint (optimal set) and computed results in the simulation.

To accelerate a PID-controller means to adjust the three multipliers K_P, K_I, and K_D adding in various amounts of these functions to get the system to behave the way you want, see [11].

Figure 5 summarizes the PID terms and their effect on a control system.

Initialization of the PID-Controller

The algorithm of the initialization of the PID-control (i.e., search K_P, K_I, K_D) is given as in Algorithm 4.2 (see [15]).

Algorithm 4.2. (1) We initialize the P-controller: $K_I = 0.0, K_D = 0.0$.

(2) The amplifying factor K_P is increased till we reached the permanent oscillations as a stability boundary of the closed control system.

(3) We obtain for K_P the critical value $K_{P,\text{crit.}}$.

(4) The period-length of the permanent oscillation is given as T_{crit}.

(5) We obtain the parameters from Table 1.

Further we compute the rest parameters as $K_I = K_P/T_n, K_D = K_P * T_v$, see [21].

4.3. Adaptive Time-Control

Often the heuristic assumptions of the PID-parameters are too coarse.

One can improve the method by applying an adaptive step-size control.

Table 1: Heuristic derivation of the control parameters (Nichols-Ziegler).

Controller	K_P	T_n	T_v
P	$0.5K_{P,\mathrm{crit}}$		
PI	$0.45K_{P,\mathrm{crit}}$	$0.85T_{\mathrm{crit}}$	
PID	$0.6K_{P,\mathrm{crit}}$	$0.5T_{\mathrm{crit}}$	$0.12T_{\mathrm{crit}}$

We discuss the step-size control with respect to our underlying error, that is, given by the computed and optimal output of our differential equation.

Based on the adaptive control, we can benefit to accelerate the control problem.

According to Hairer and Wanner [22], we apply the automatic control problem with a PID-controller.

The automatically step-size is given as (see [11])

$$\Delta t_{n+1} = \left(\frac{e_{n-1}}{e_n}\right)^{K_P} \left(\frac{\mathrm{tol}}{e_n}\right)^{K_I} \left(\frac{e_{n-1}^2}{e_n e_{n-2}}\right)^{K_D} \Delta t_n, \tag{4.3}$$

where tol is the tolerance, e_n is the error of the quantities of interest in time-step Δt.

We can control the step-size with respect to our heuristically computed K_P, K_I, and K_D parameters. Initialization of the adaptive control can be seen in Algorithm 4.3.

Algorithm 4.3. (1) Define Tolerance, Min and Max of the concentration.

(2) Apply the parameters: K_P, K_D, K_I form a first run.

(3) Optimize the computations with a first feedback.

4.4. Identification of the Control Path

In the forward problem, we computed a PDE to simulate the CVD process in the underlying domain. For our control problem, we have the behavior of two points in the underlying domain, C_{source} our source and C_x the source restricted to the response of the controlled system. To analyze the differences between given source the C_{source} and the computed source in the control process C_x, we study the step response from our system:

$$x_a(t) := \left\{C_x(t) \mid C_{\mathrm{Source}}(t) = x_{e0}\right\}. \tag{4.4}$$

Algorithm 4.4. (1) We determine the model of the control path, for example, PT1 (Proportional time 1), PT2 (Proportional time 2), see [16]. For that purpose we investigate the step response x_a, his first and second derivate [16, pages 117 and 331].

(2) We determine the parameters of our model: Kp, T1, T2.

(3) Our goal is to control the system with a controller. Also we have to determine the control-parameter [16, page 405].
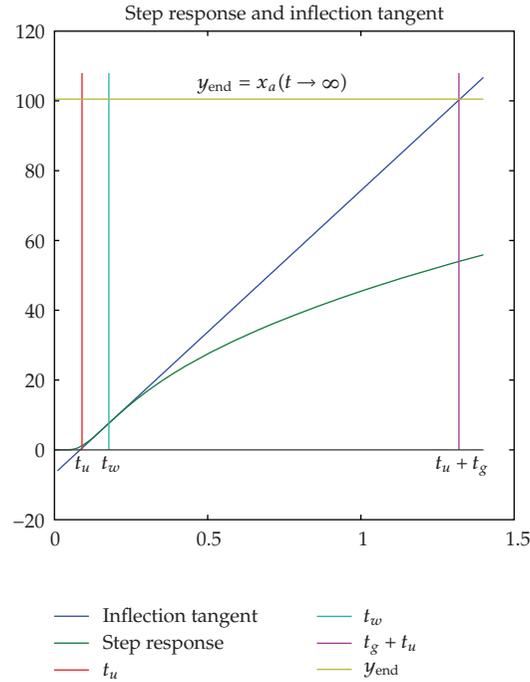
Figure 6: Step response and inflection tangent.

Identification of the Control Path

With the step response x_a, we identify the control path as a PT2 element.

Because $\lim_{t \to \infty} x_a(t) = $ constant (also $\neq 0$ nor ∞), the basic behavior of our element is P (and neither D nor I). Now we have to determine the number of time delays in the control path.

Identification of the Parameters

For such control path which has PT2 behavior, there exists a model as ordinary differential equation. Here a generalization can be obtained with four parameters and be divided by a_0 a normed form with three parameters ω_0 = angular frequency, A = attenuation, and K_S = proportional factor:

$$a_2 \cdot \ddot{x}_a(t) + a_1 \cdot \dot{x}_a(t) + a_0 \cdot x_a(t) = b_0 \cdot x_{e0},$$

$$\frac{1}{\omega_0^2} \cdot \ddot{x}_a(t) + \frac{2A}{\omega_0} \cdot \dot{x}_a(t) + x_a(t) = K_S \cdot x_{e0}. \qquad (4.5)$$

The solution of the differential equation results in the characteristic equation:

$$\frac{\alpha^2}{\omega_0^2} + 2A \cdot \frac{\alpha}{\omega_0} + 1 = 0 \quad \text{with } \alpha_{1,2} = -A \cdot \omega_0 \pm \omega_0 \cdot \sqrt{A^2 - 1}. \qquad (4.6)$$
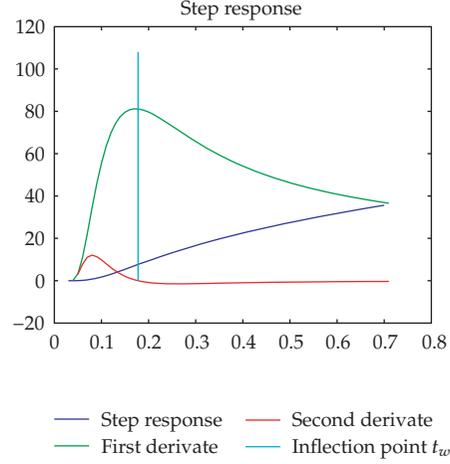
Figure 7: Empirical determination of the model of the control path. The three main cases: (1) $x_a(t = 0) \neq 0$: P, PDT$_1$, PPT$_1$, PIDT$_1$, (2) $x_a(t = 0) = 0, \dot{x}_a(t = 0) \neq 0$: I, PT$_1$, PDT$_2$, (3) $x_a(t = 0), \dot{x}_a(t = 0) = 0, \ddot{x}_a(t = 0) \neq 0$: IT$_1$, PT$_2$,....,(1) $x_a(0) = 0$, (2) $\dot{x}_a(0) = 0$, (3) $\ddot{x}_a(0) \neq 0$.

So that we have the analytical solution for the normed step response:

$$\frac{x_a(t)}{x_{e0}} = K_S \cdot \left[1 - \frac{T_1}{T_1 - T_2} \cdot e^{-t/T_1} + \frac{T_2}{T_1 - T_2} \cdot e^{-t/T_2} \right],$$

$$\text{with } \alpha_1 = -\frac{1}{T_1}, \quad \alpha_2 = -\frac{1}{T_2}. \tag{4.7}$$

We go now the inverse way to identify the system. We start with the step response. Using the inflected tangent at $(t_w, x_w, \dot{x}_a(t_w))$ and the limit $x_a(t \to \infty)$, we obtain our parameters.

Firstly, we have $K_S = x_a(t \to \infty)/x_{e0}$, $t_g = K_S \cdot (x_{e0}/\dot{x}_a(t_w))$, and $t_u = t_w - t_g \cdot (x_a(t_w)/x_a(t \to \infty))$, see Figure 6. In the second step, we obtain T_1 and T_2 by the following formulas:

$$\frac{t_u}{t_g} = \frac{\alpha^{\alpha/(1-\alpha)} \cdot (\alpha \cdot \ln(\alpha) + \alpha^2 - 1)}{\alpha - 1} - 1, \quad \frac{t_g}{T_1} = \alpha^{\alpha/(1-\alpha)}, \quad \alpha = \frac{T_2}{T_1}. \tag{4.8}$$

Control Parameters

The parameter occurrence also in the transfer function $G(s)$, which is the Laplace transformation of the differential equation. Our first choice to determine the control parameter is here the PT2 element. Alternatively, we use an approximation to a PT1 element, but here we have additional a scan-time parameter:

$$G(s) = \frac{K_S}{(1 + T_1 s)(1 + T_2 s)} \text{ PT2 element,}$$

$$G(s) \approx K_S \cdot \frac{e^{-T_u s}}{1 + s \cdot T_g} \text{ PT1 element.} \tag{4.9}$$

Table 2: Derivation of the control parameters (Chien-Hrones-Reswick and Takahashi).

PID	(PO) = 0	(PO) = 20	Takahashi
K_R	$0.6 \cdot \dfrac{t_g}{t_u \cdot K_S}$	$0.95 \cdot \dfrac{t_g}{t_u \cdot K_S}$	$\dfrac{1,2 \cdot t_g}{K_s \cdot (t_u + T)}$
T_n	$1.0 \cdot t_g$	$1.35 \cdot t_g$	$\dfrac{2 \cdot (t_u + T/2)^2}{t_u + T}$
T_v	$0.5 \cdot t_u$	$0.47 \cdot t_u$	$0.5(t_u + T)$

Table 3: MATLAB-toolbox pdetool.

MATLAB function	Description	Enhancement
Parabolic	Solve parabolic PDE (heat equation)	In the future we plan to use academical code to compute the PDE-solutions (e.g., θ-method). Furthermore we will use Comsol/Femlab
Pdetool	MATLAB toolbox to create the geometry of the FEM-structure and the boundary conditions	We have also implemented an alternative mesh, with orthogonal triangles. MATLAB always uses equilateral triangles
Refine	This function refines the geometry of the mesh. All triangles, were replaced with four new triangles	For convergence rate calculations we must guarantee that the geometry near the points, which are changed in the backward step (source), are similar
Guide	MATLAB toolbox to create graphical user interfaces (gui)	

Independent from the behavior which is being wished, we obtain values for the parameters K_P, K_I, and K_D, respectively. $K_R = K_P$, T_n and T_v are of the PID-controller. In Table 2 are given the values in dependence to the percentage overshoot (PO). There are the values of PO = 0, respectively, PO = 20. The rest can be determined by linear interpolation. An alternative schema is given by Takahashi. Here we have in addition the parameter scan-time T. This is a general form of the Nichols-Ziegler method (see Table 1). Further are $K_P = K_R$, $K_I = K_P/T_n$, and $K_D = K_P * T_v$.

5. Software and Program-Tools

MATLAB Functions

We use the MATLAB-toolbox pdetool for the time- and spatial-discretizations, where we have a finite-element method with $P1$-elements and an implicit Euler method for the time-discretization.

DEPOSIT-PID Toolbox

The PID-controller is also programmed in MATLAB. Our combined code is given in the DEPOSIT-PID toolbox and described in what follows. The DEPOSIT-PID toolbox is

Distribution of the concentration
before backward step

Distribution of the concentration
before backward step

(a)

(b)
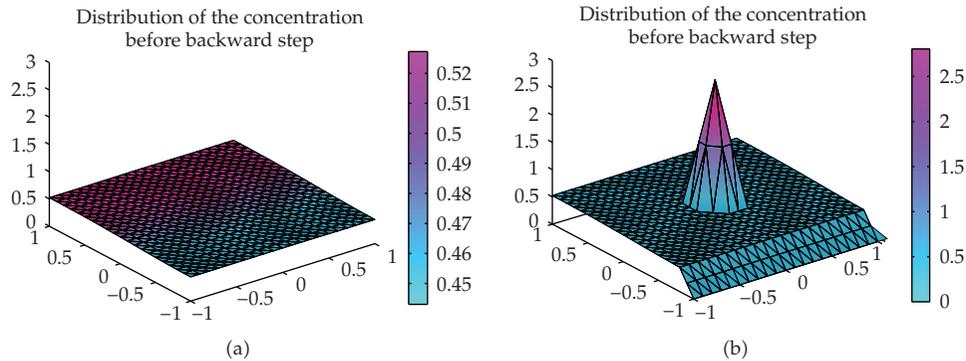
Figure 8: Backward step.

manipulated by a graphical user interface, by which simulation- and control-models are chosen and corresponding parameters can be manually adjusted.

The objective is to simulate the diffusion and deposition of the vapor in the apparatus and to obtain the optimal vapor concentration at the measuring point.

We can divide the process into three phases, namely, forward step, control step, backward step, which have a cyclic repetition. In the forward step, the diffusion takes place, which can be simulated by a time-step of the heat equation.

After that, in the control step, the actual concentration at the under boundary $(0, -1)$ can be measured (shown by computed in the graph) and compared with the optimal value (shown by optimal).

The control step is followed by the backward step: from the error in control step and the control model, the optimal alteration of the source can be computed. The vapor flows through the source point $(0, 0)$ in the apparatus and this value is shown by the SourceOutput in Figure 12. This will be simulated through the addition of SourceOutput to actual concentration at the source point (see Figure 8). Further, as a simplification, we set the concentration at the under boundary to zero, because here the gas transforms into solid matter .

In Section 4 we introduced some fundamentals in order to control the apparatus. The models and corresponding parameters can be altered by the Gui.

The Layout of the DEPOSIT-PID Gui

This Gui contains the following:

(1) short-time plot (2D) of computed, optimal, and SourceOutput;

(2) long-time plot (2D);

(3) listbox with names parameters:

    (a) textbox with actual value of the parameter chosen in [12],

    (b) textbox to change the value of the parameter chosen in [12];

(4) listbox with names parameters:

    (a) textbox with actual value of the parameter chosen in [23],

    (b) textbox to change the value of the parameter chosen in [23];
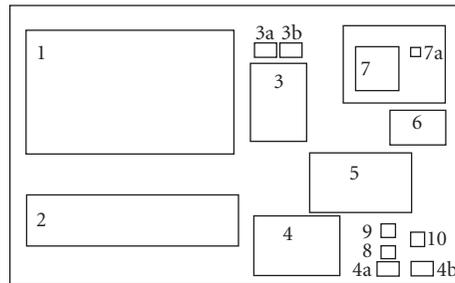
Figure 9: Layout of the DEPOSIT-PID Gui.

    (5) 3D plot of distribution;

    (6) 3D grid plot of distribution;

    (7) listbox with names of parameters:

        (a) checkbox with actual value of the parameter chosen in [24];

    (8) push button: save;

    (9) push button: reset;

  (10) radio button: start.

KONTOOL

Another software-tool, KONTOOL, is programmed to compute the numerical convergence rates of the applications. The software-tool has implemented the errors and convergence rates defined in Section 3.2. An error-analysis based on successive refinement of space and time is done and the resulting errors and convergence rates are computed. Optimal convergence rates with respect to balance the time- and spatial-grids are calculated, see Algorithm 5.2.

Remark 5.1. The software-tool can be modified and applied to arbitrary spatial- and time-discretization methods. The interface of KONTOOL needs at least the parameters of the spatial- and time-grid and the starting parameters of the underlying methods.

Algorithm 5.2 (The algorithm of the computation of the numerical convergence tableau). (1) We compute reference solutions: (a) numerically: fine time and spatial steps or (b) analytically (if there exists an analytical solution).
    (2) We apply one spatial discretization of step Δx and apply all time discretization with steps Δt, where the coarsest Δt is given by the CFL condition or till first non-numerical results as oscillations. We compute the error $c_{\mathrm{num}} - c_{\mathrm{ref}}$ in the L_2-norm.
    (3) We continue the next fine spatial steps, for example, $\Delta x/2$.
    (4) We compute the convergence tableau with time and space.

    In the next section, we discuss the numerical experiments.
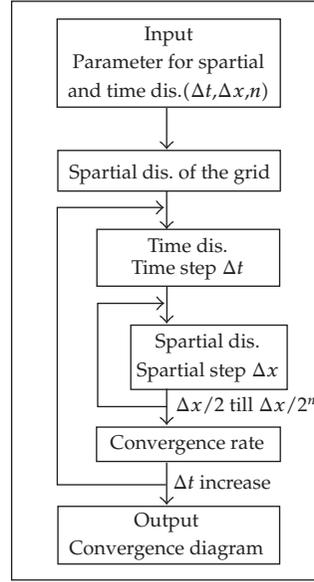
Figure 10: Convergence diagram tool.

6. Experiment for the Plasma Reactor

In this section, we present our numerical experiments for the CVD processes in a plasma reactor.

6.1. Simulation of a Diffusion Equation with Analytical Solution (Neumann Boundary Conditions)

Here we simulate a diffusion equation with Neumann boundary conditions and right-hand side 0. Our control problem has only the forward problem to solve and we consider the accuracy of our simulations.

We have the following equation:

$$\partial_t c - \beta^2 (\partial_{xx} + \partial_{yy}) c = 0, \quad \text{in } \Omega \times [0, T],$$
$$c(x, y, 0) = \cos(2x) + \cos(2y), \quad \text{on } \Omega, \qquad (6.1)$$
$$\partial_n c(x, y, t) = 0, \quad \text{on } \partial\Omega \times [0, T],$$

where c is the molar concentration, $\Omega = [-1, 1] \times [-1, 1]$, and $t \in (0, T)$. $D = \beta^2$ is the diffusion parameter of the diffusion equation.

We have the following analytical solution:

$$c_{\text{ana}}(x, y, t) = \sin(2) + \sum_{n=1}^{\infty} A_n \exp(-\beta^2 n^2 \pi^2 t)(\cos(n\pi x) + \cos(n\pi y)), \qquad (6.2)$$

Numerical solution: $\Delta x = 0.2$ (a)

Numerical solution: $\Delta x = 0.1$ (b)

Numerical solution: $\Delta x = 0.05$ (c)

Analytical solution (d)
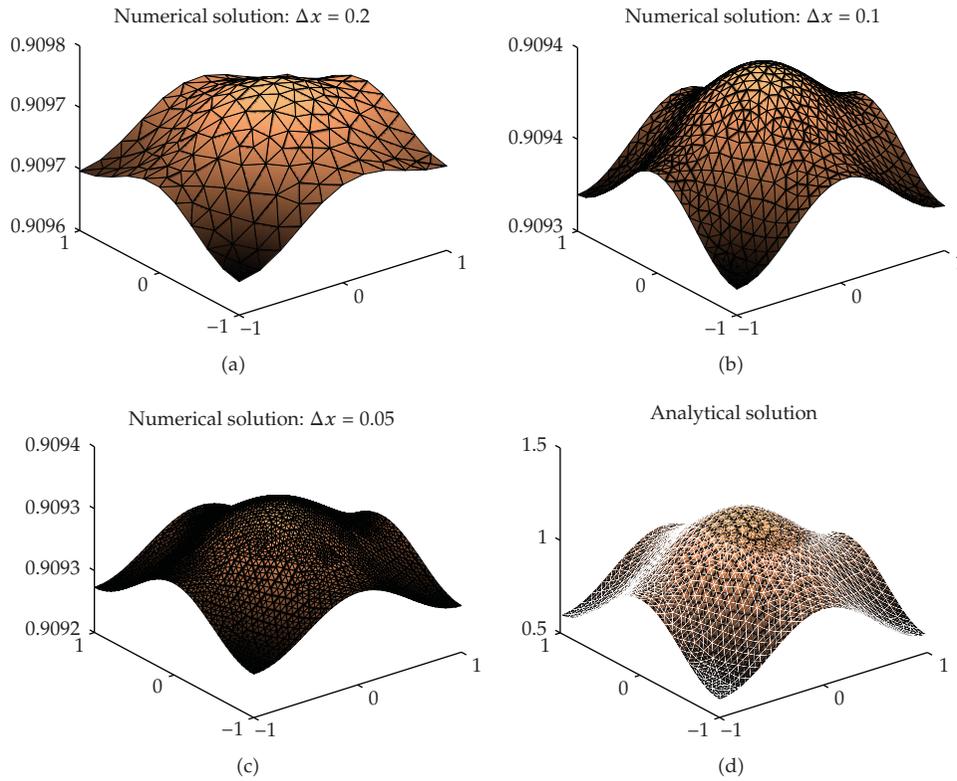
Figure 11: 2D experiment of the diffusion equation at the end-time $T = 12.0$, $\beta = 0.1$. x-axis: x-coordinate, y-axis: y-coordinate and z-axis: c (concentration).

Table 4: Offset convergence ($\Delta t = 0.1$, time-step = 100, $\beta = 0.61644$).

Δx	offset(analy.)	max(num.)	min(num.)	L_1-error	L_2-error
0.20	0.90**9**297426826	0.90**21**08407643	0.90**21**08381295	7.189e–3	5.17e–5
0.10	0.90**9**297426826	0.90**7**514994036	0.90**7**514987008	1.782e–3	3.18e–6
0.05	0.90**9**297426826	0.90**8**858511500	0.90**8**858509666	4.389e–4	1.93e–7

where $A_n = (-1)^n(-4\sin(2)/(n^2\pi^2 - 4))$. We apply the diffusion coefficient $D = 0.01$, respectively, $\beta = 0.1$. We obtain the following result after time $T = 12.0$, see Figure 11.

We see that for large Δx the numerical solution converge faster to the stable constant endsolution (offset) than solution for smaller Δx and the analytical solution. The error between the constant analytical endsolution (offset, $\sin(2) = 0.9093$) is also greater than for smaller Δx. The L_2-error is given in Table 4.

Remark 6.1. We test for the pure diffusion equation our underlying discretization methods and apply finite elements for the spatial-discretization and implicit Runge-Kutta methods for the time-discretization. In the results, we obtain decreasing errors for the different time- and spatial-steps.
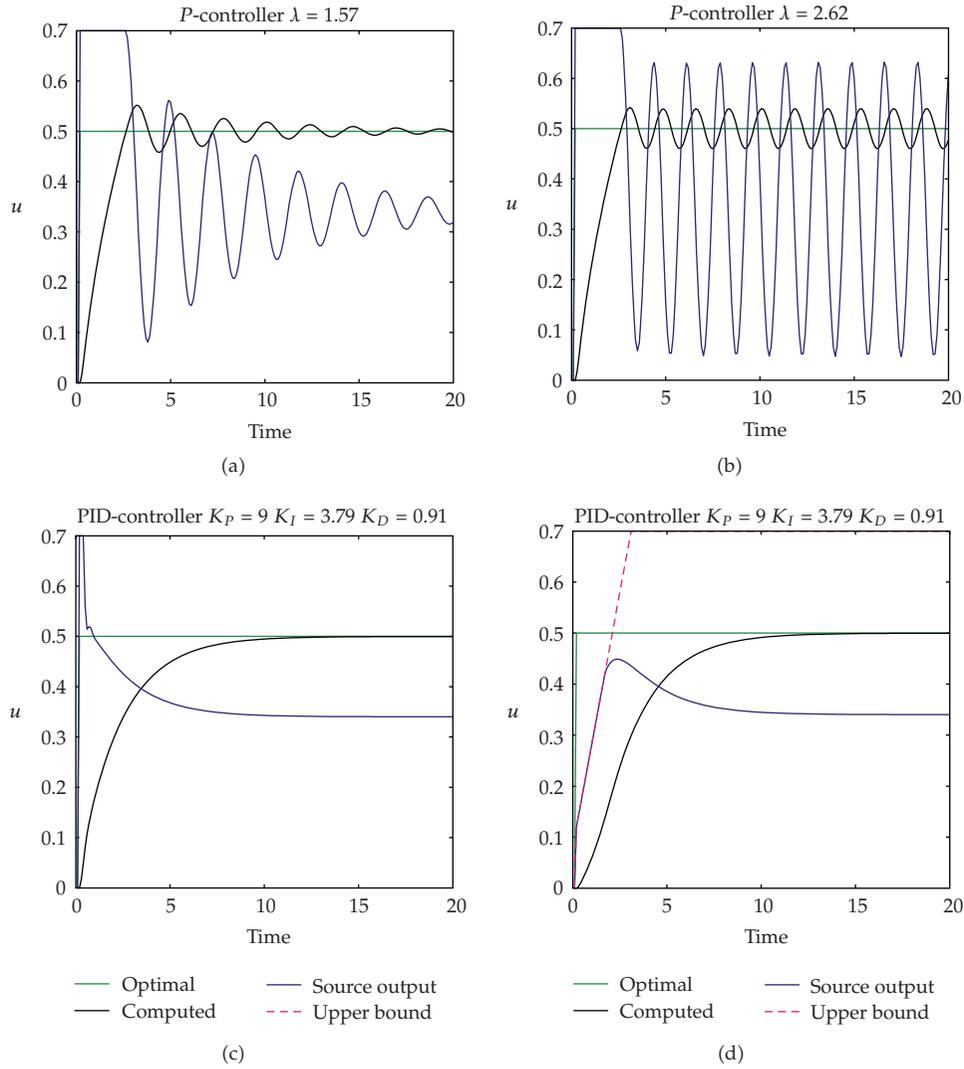
Figure 12: 2D experiment of the diffusion equation and control of a single point.

6.2. Simulation of an Optimal Control of a Diffusion Equation with Heuristic Choise of the Control Parameters

Here we simulate a first example of a diffusion equation and control the concentrations in the deposition process.

We have the following equation:

$$
\begin{aligned}
\partial_t c - \nabla D \nabla c &= f(t), &&\text{in } \Omega \times [0, T], \\
c(x, 0) &= c_0(x), &&\text{on } \Omega, \\
\frac{\partial c(x, t)}{\partial n} &= c_1(x, t), &&\text{on } \partial\Omega \times [0, T],
\end{aligned}
\tag{6.3}
$$

where c is the molar concentration, D is the diffusion parameter of the diffusion equation, and $f(t)$ is the right-hand side or diffusion source. We have the following constraint: $c_{\text{optimal}}(x_{\text{point}}, y_{\text{point}}) = 0.5$, where $(x_{\text{point}}, y_{\text{point}})$ is the control point in our domain. The parameters are given as $D = 0.01$ and $f(t) = at + b$, so we deal with a linear source, $a = 0.2$ and $b = 0.1$ are constants. In the following tests, we propose the 3 possibilities to control the optimal temperature:

(i) P-control with constant optimal constraint,

(ii) PID-control with constant optimal constraint,

(iii) PID-control with Linear optimal constraint.

The results for the control methods are given in Figure 12.

6.3. Preliminary Remark for Simulations for the Convergence Order

To determine the function Arg Max, firstly, we have to determine a practicable interval I for

$$\text{Arg Max}(\Delta x) := \arg\max_{\Delta t \in I} \rho_{L_2, \Delta x, \Delta x/2, \Delta t}. \tag{6.4}$$

One possibility is the interval

$$I_{\text{CFL}} := (0, \Delta t_{\text{CLF}}], \tag{6.5}$$

with $\Delta t_{\text{CFL}} := \Delta s^2/2D_{\max}$ from the CFL-condition. In the experiments, we find another interval, where the convergence rate function is relative stable convex:

$$I_{\text{stable}} := (0, \Delta t_{\text{stable}}], \tag{6.6}$$

where

$$\Delta t_{\text{stable}} := \max_{\Delta t}\{\rho : (0, \Delta t) \longrightarrow \mathbb{R} \text{ is convex}\}. \tag{6.7}$$

It is clear that we can in I_{stable} take some restrictions to a subinterval $I_{\text{sub}} := [\Delta t_{s\,\min}, \Delta t_{s\,\max}] \subset I_{\text{stable}}$, when we know that some $\Delta t \in I_{\text{sub}}$ with $\rho(\Delta t) > \rho(\Delta t_{s\,\min}), \rho(\Delta t_{s\,\max})$.

The function of the numerical convergence rate is discrete in the spatial-discretization variable Δx since we get a finer discretization with a bisection of Δx. With a finer discretization, a triangle is replaced by four subtriangles.

In the temporal discretization variable Δt, we are in contrast not restricted to such conditions and Δx could be chosen to any arbitrary value above 0. To get a first glance, we have selected the methods of bisection. Subsequently, we consider finer discretizations in intervals which are of special interest.

Table 5: Numerical results for the P-controller for different spatial-steps with $D = 0.1$ and $\lambda = 1.0$ as P-value for the controller.

Δx	Δt	err $= |u_{\text{num},\Delta x,\Delta t} - u_{\text{num,fine}\Delta x,\Delta t}|$	Convergence rate
0.1	0.1	0.077007	3.2531
0.1	0.05	0.016153	4.0077
0.1	0.025	0.0020085	3.9447
0.1	0.0125	0.00026087	1.8276
0.1	0.00625	0.00014699	0
0.05	0.1	0.27873	2.8591
0.05	0.05	0.076833	3.7481
0.05	0.025	0.011437	4.052
0.05	0.0125	0.001379	3.7776
0.05	0.00625	0.00020111	0
0.025	0.1	0.6564	2.4552
0.025	0.05	0.23939	3.2449
0.025	0.025	0.050505	4.008
0.025	0.0125	0.0062781	3.9999
0.025	0.00625	0.00078482	0
0.0125	0.1	1.04	2.1424
0.0125	0.05	0.4711	2.7173
0.0125	0.025	0.14327	3.7251
0.0125	0.0125	0.021668	4.0516
0.0125	0.00625	0.0026133	0
0.00625	0.1	1.2092	1.586
0.00625	0.05	0.80554	2.4179
0.00625	0.025	0.30148	3.2179
0.00625	0.0125	0.064803	4.005
0.00625	0.00625	0.0080726	0

6.4. Simulations for the Convergence Order

We consider the convergence rate ρ as a function in dependent of Δx and Δt. This function also depends on some parameters, for instance, the diffusion coefficient D and the control parameter λ. We now present the results of two chosen experiments. For the first experiment, the parameters are $D = 0.1$ and $\lambda = 1$, while we increase the propagation velocity in such a way that we increase D to 1 for the second experiment.

We observe that, for instance, in the case of spatial-discretization $\Delta x = 0.05$ and $x = 0.25$, the maximal convergence rate lies between $\Delta t = 0.05$ and $\Delta t = 0.0125$. The maximum itself lies close to 0.025. To determine the precise value of $\Delta t = \text{Arg Max}(\Delta x)$, we refine our method in these interesting intervals.

In Figure 13, we observe that there is a relatively stable convex area starting at 0. This area then switches over to an area with strong fluctuations. We now compare the stable area with the CFL-condition. The area where CFL < 1 lies in the stable convex area of the function ($\Delta t_{\text{CFL}} < \Delta t_{\text{stable}}$). For $\Delta x = 0.05$, $D = 0.1$, the CFL-area ends at $\Delta t_{\text{CFL}} = 0.0125$, the stable convex area range to about $\Delta t_{\text{stable}} = 0.2$. For our convergence diagram, we try to find for

(a)

(b)
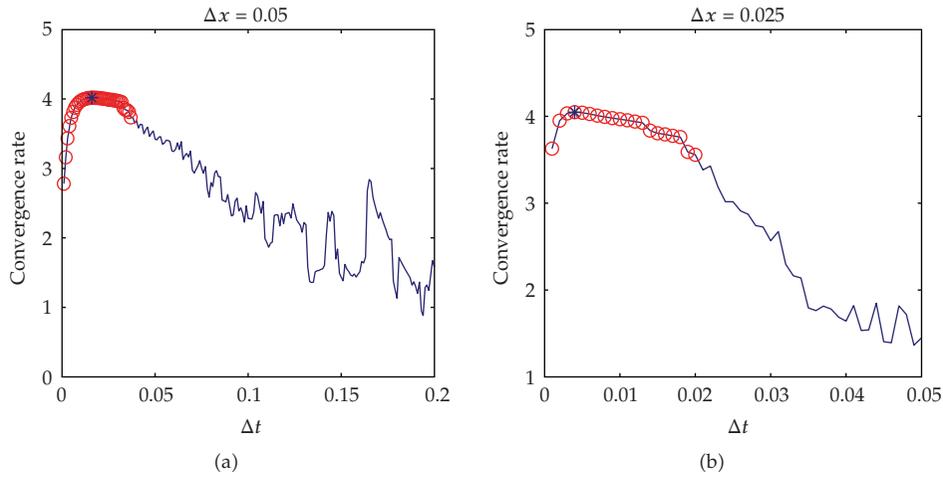
Figure 13: ρ for a P-controller, $D = 0.1$.
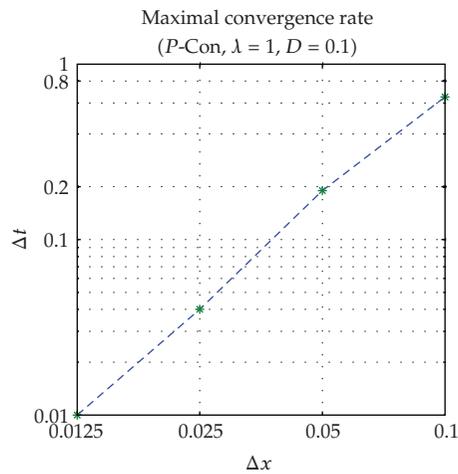
Figure 14: Convergence diagram (KONTOOL). We can see in the loglog-plot $\mathrm{Arg\,Max}(\Delta x)$ linear dependence.

every Δx the value Δt, where the convergence rate becomes maximal (ArgMax). It is clear that we only use Δt at which the convergence rate function is stable: $\Delta t < \Delta t_{\mathrm{stable}}$.

Remark 6.2. The experiment shows the linear convergence rate of the P-controller with different λ values. So we obtain a stable method with respect to the P-controller. In the examples, we apply heuristic methods to derive the control parameters for the P- and PID-controller. We show that we have reached the linear order of the underlying finite element discretization method. We have higher control errors if we did not compute the correct control parameters and the numerical errors are smaller than our control error. To prohibit this problem, we have to compute in the next example the control parameters by a feedback equation, see [25].

Maximal convergence rate
$(P\text{-Con}, \lambda = 1, D = 1)$
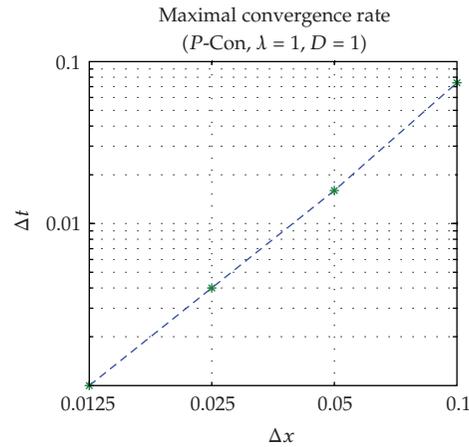
Figure 15: Convergence diagram (KONTOOL). The loglog-plot $\mathrm{Arg\,Max}(\Delta x)$ with linear dependence.

Table 6: The associated Δt, which have maximal convergence rate ($\mathrm{Arg\,Max}$), for $\Delta x = 0.1, 0.05, 0.025, 0.0125$.

Δx	$\mathrm{Arg\,Max}$	$-\log(\mathrm{Max})$
.10000	.65000	0.62
.05000	.19000	2.4
.02500	0.04	4.64
.01250	.01000	6.64

Table 7: The associated Δt, which have maximal convergence rate ($\mathrm{Arg\,Max}$), for $\Delta x = 0.1, 0.05, 0.025, 0.0125$.

Δx	$\mathrm{Arg\,Max}$	$-\log(\mathrm{Max})$
.10000	0.0700	3.84
.05000	0.0170	5.88
.02500	0.0040	7.97
.01250	0.0010	9.97

6.5. Simulation of an Optimal Control of a Diffusion Equation with Adaptive Control

In the second example, we simulate the diffusion equation and control the temperature with and adaptive control based on a PID-controller, see [11].

We have the following equation:

$$\begin{aligned}
\partial_t c - \nabla D \nabla c &= f(t), \quad \text{in } \Omega \times [0, T], \\
c(x, t) &= c_0(x), \quad \text{on } \Omega, \\
\frac{\partial c(x, t)}{\partial n} &= c_1(x, t), \quad \text{on } \partial\Omega \times [0, T],
\end{aligned} \tag{6.8}$$

where c is the molar concentration, D is the diffusion parameter of the diffusion equation, and $f(t)$ is the right hand side or source.
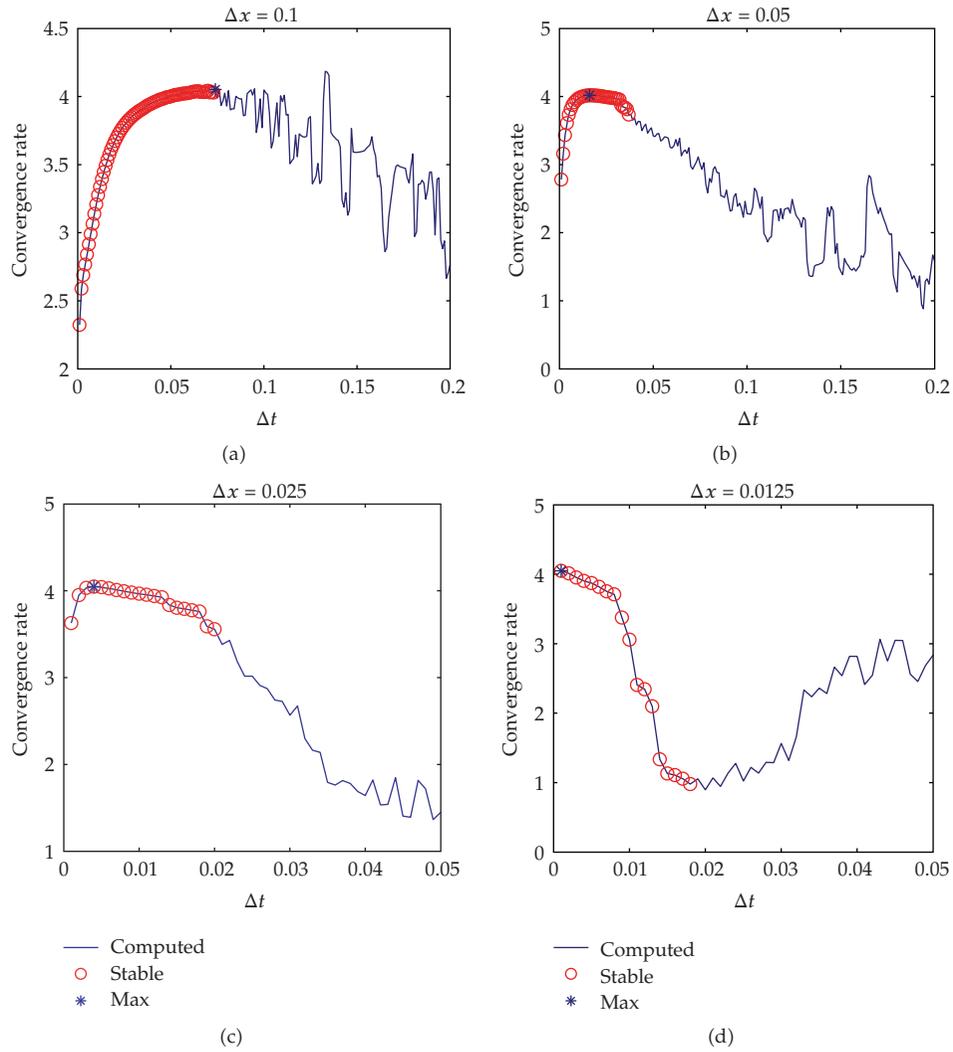
Figure 16: ρ for a P-controller, $D = 1$.

We have the following constraint:

$$c_{\text{optimal}}(x_{\text{point}}, y_{\text{point}}) = 0.5, \qquad (6.9)$$

where $(x_{\text{point}}, y_{\text{point}}) = (0, -1)$ is the control point in our domain.

The automatically step-size is given as (see [11])

$$\Delta t_{n+1} = \left(\frac{e_{n-1}}{e_n}\right)^{K_P} \left(\frac{\text{tol}}{e_n}\right)^{K_I} \left(\frac{e_{n-1}^2}{e_n e_{n-2}}\right)^{K_D} \Delta t_n, \qquad (6.10)$$

where tol = 1 is the tolerance, e_n is the error of the quantities of interest in time-step Δt.

Adaptive PID ($P_{\mathrm{krit}} = 15$, $T_{\mathrm{krit}} = 5$)

—— Adaptive PID: tol = 1, $\Delta t_{\max} = 0.1$, $\Delta t_{\min} = 0.07$
—— PID: $\Delta t = 0.1$
—— Optimal
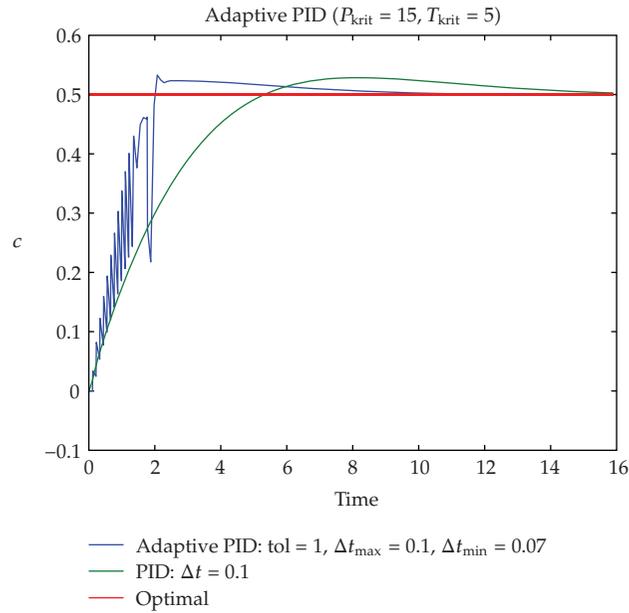
Figure 17: 2D experiment with and without the adaptive time-step control.

Adaptive PID: parameter $\Delta t_{\min}$

—— Adaptive PID: $\Delta t_{\min} = 0.0001$
- - - Adaptive PID: $\Delta t_{\min} = 0.1$
—— Optimal

(a)

Adaptive PID: parameter tol

—— Adaptive PID: tol = 0.1
- - - Adaptive PID: tol = 1
—— Optimal

(b)
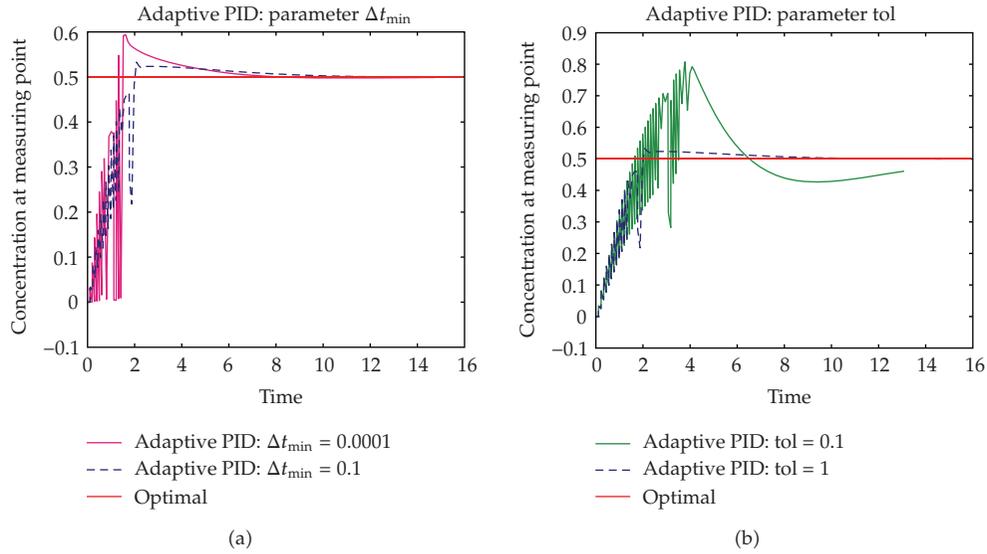
Figure 18: Adaptive PID with modified parameters.

The errors are given as

$$e_n = \frac{\|u_n - u_{n-1}\|}{\|u_n\|}, \tag{6.11}$$

where u_n is the result at time-step t^n.
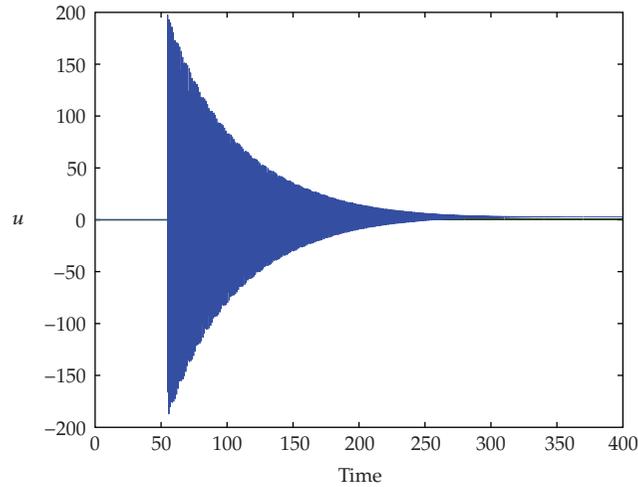
Figure 19: Error of the time-step control.

Table 8: Convergence of λ with bisection.

| λ	$a(\max)$	$a(\min)$	Time-step
790.000	0.031267612875889	0.031262218711872	**2**
592.500	0.017928791447504	0.018012608268341	3
493.750	0.009755520357640	0.009663115651239	4
444.375	0.005011636633412	0.005007035757358	5
419.687	0.002464209438841	0.002504970973550	6
407.343	0.001131628414131	0.001182533773415	7
401.171	0.000466907455284	0.000465794576191	8
398.085	0.000109260068267	0.000189637255604	9
397.314	0.000080312399314	0.000043201893613	11
396.928	0.000041628241361	0.000012118492674	12
396.735	0.000002923384293	0.000006047318776	13
396.729			**18**
396.723	−0.000017536337301	−0.000000444521435	17
396.711	−0.000016718130883	−0.000004912794617	16
396.687	−0.000002204413679	−0.000022279535981	15
396.639	−0.000015370659931	−0.000016743767093	14
396.542	−0.000042906107641	−0.000015117641002	10
395.000	−0.000163960451645	−0.000164998927189	**1**

The parameters are given as $D = 0.1$, $P_{\text{krit}} = 15$, $T_{\text{krit}} = 5$, Δt (PID-control), tol = 1 (adaptive PID-control), $\Delta t_{\max} = 0.1$ (adaptive PID-control), $\Delta t_{\min} = 0.01$ (adaptive PID-control) (see Figure 17).

Furthermore, we change the parameter tol = 1 to tol = 0.1 and $\Delta t_{\min} = 0.01$ to $\Delta t_{\min} = 0.0001$ (see Figure 18).

Remark 6.3. In Figures 17 and 18, we see an oscillating time interval at the beginning of the automatically step-size control. In the first experiments, we had only taken into account a

Minima: y_2+KonstanteA: −0.018659B: 5.1473   Maxima: y−KonstanteA: −0.019462B: 5.1428

(a)   (b)
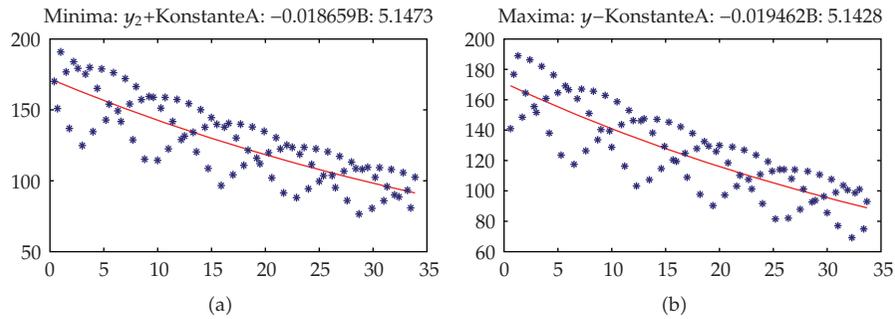
Figure 20: Exponential regression of min and max. Behavior of λ.

P_{krit} T_{krit}

K_P T_n T_v

K_P K_I K_D

(a)   (b)   (c)
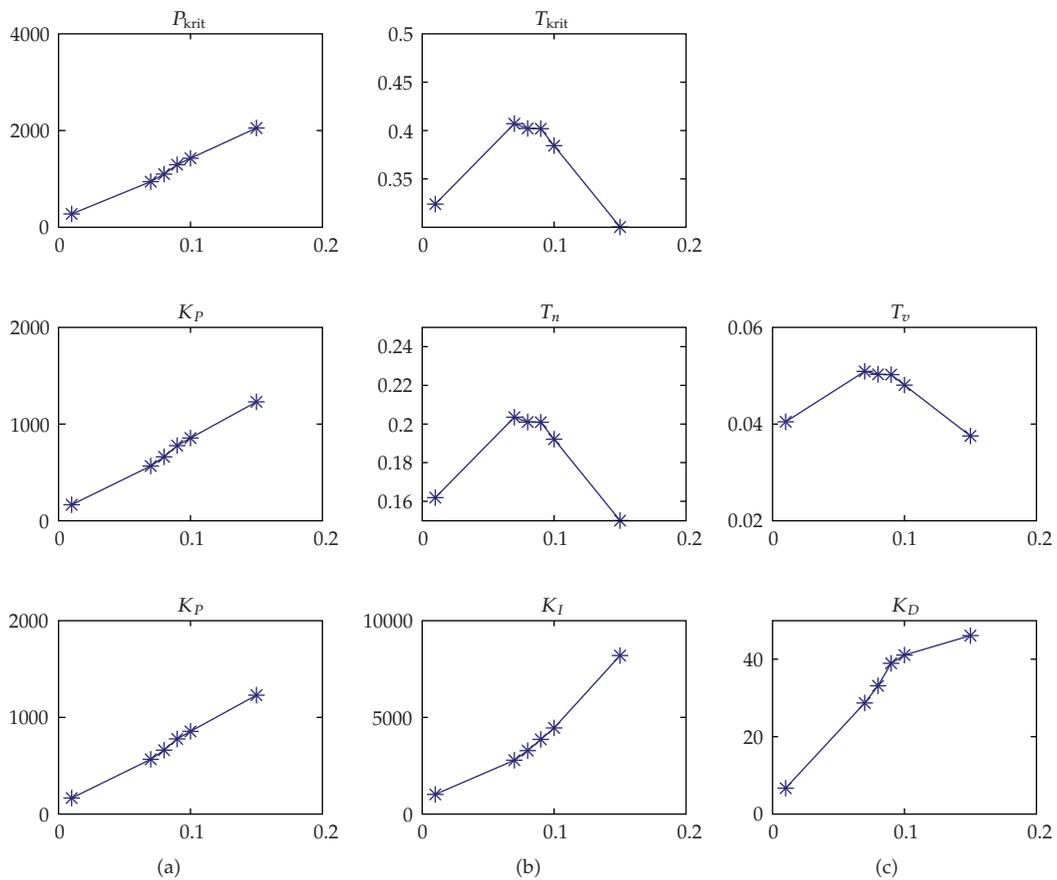
Figure 21: Δt dependence parameters determined by Nichols-Ziegler.

previous heuristically computation of the control parameters K_P, K_I, and K_D before the step-size control for the whole time-interval $0, T$. The optimal control parameters are given to the whole time-interval $0, T$. A modified algorithm to compute the control parameters for the initialization time-interval $0, t_{\text{begin}}$ and the whole time-interval $0, T$ overcome the oscillation problems.
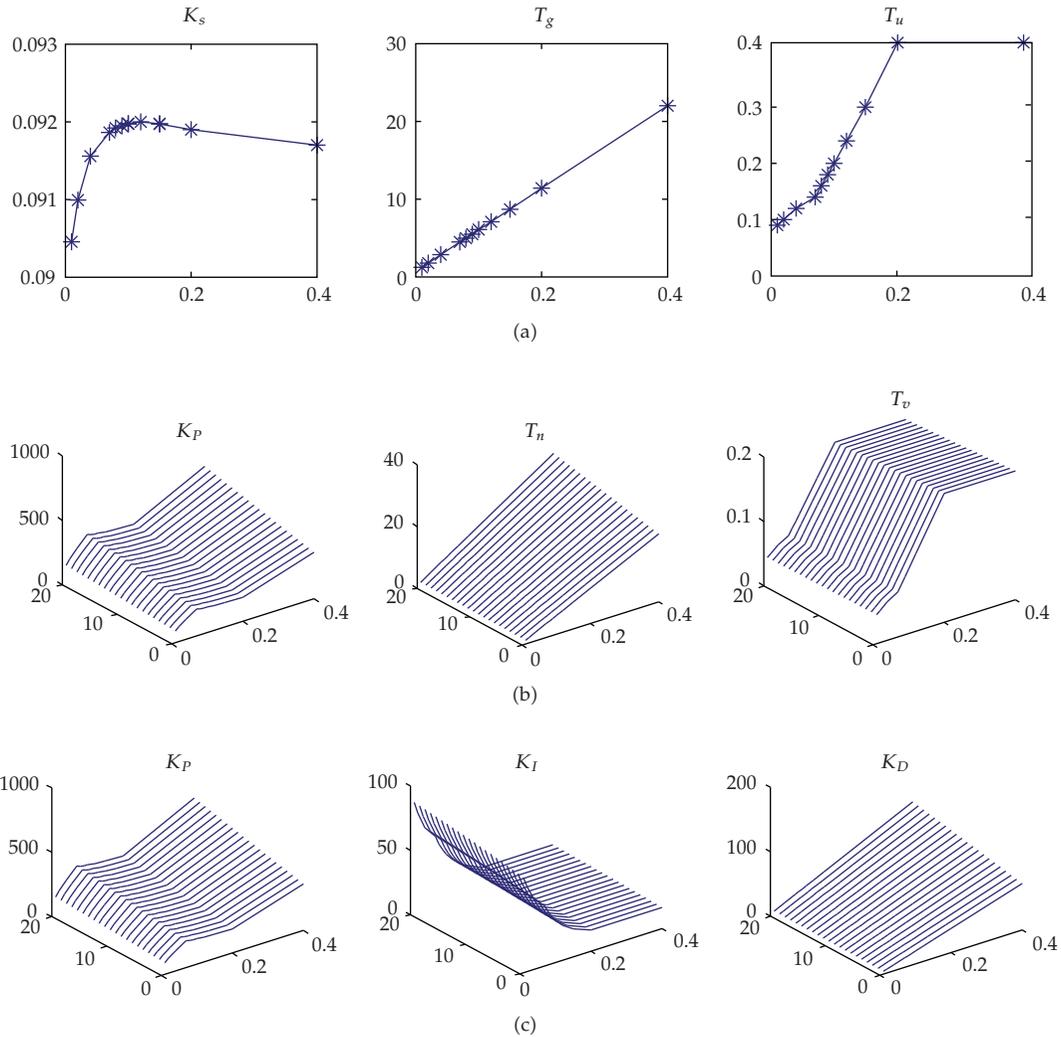
Figure 22: Δt dependence parameters determined by Chien-Hrones-Reswick with PO $= 0, \ldots, 20$.

A modified automatically step-size control, which minimize the oscillations is given in the following algorithm.

Algorithm 6.4. (1) We compute the reference control parameters $K_{P,\text{global}}$, $K_{I,\text{global}}$ and $K_{D,\text{global}}$ for the time-interval $0, T$.

(2) We apply the automatically step-size control for the global control parameters with $\text{tol}_{\text{global}}$, $\Delta t_{\text{max,global}}$ and $\Delta t_{\text{min,global}}$, which can by chosen large.

(3) We stop the computation till we reach the optimal solution and mark remember the time t_{oszill}.

(4) We compute the local control parameters $K_{P,\text{local}}$, $K_{I,\text{local}}$ and $K_{D,\text{local}}$ for the time-interval $0, t_{\text{oszill}}$.
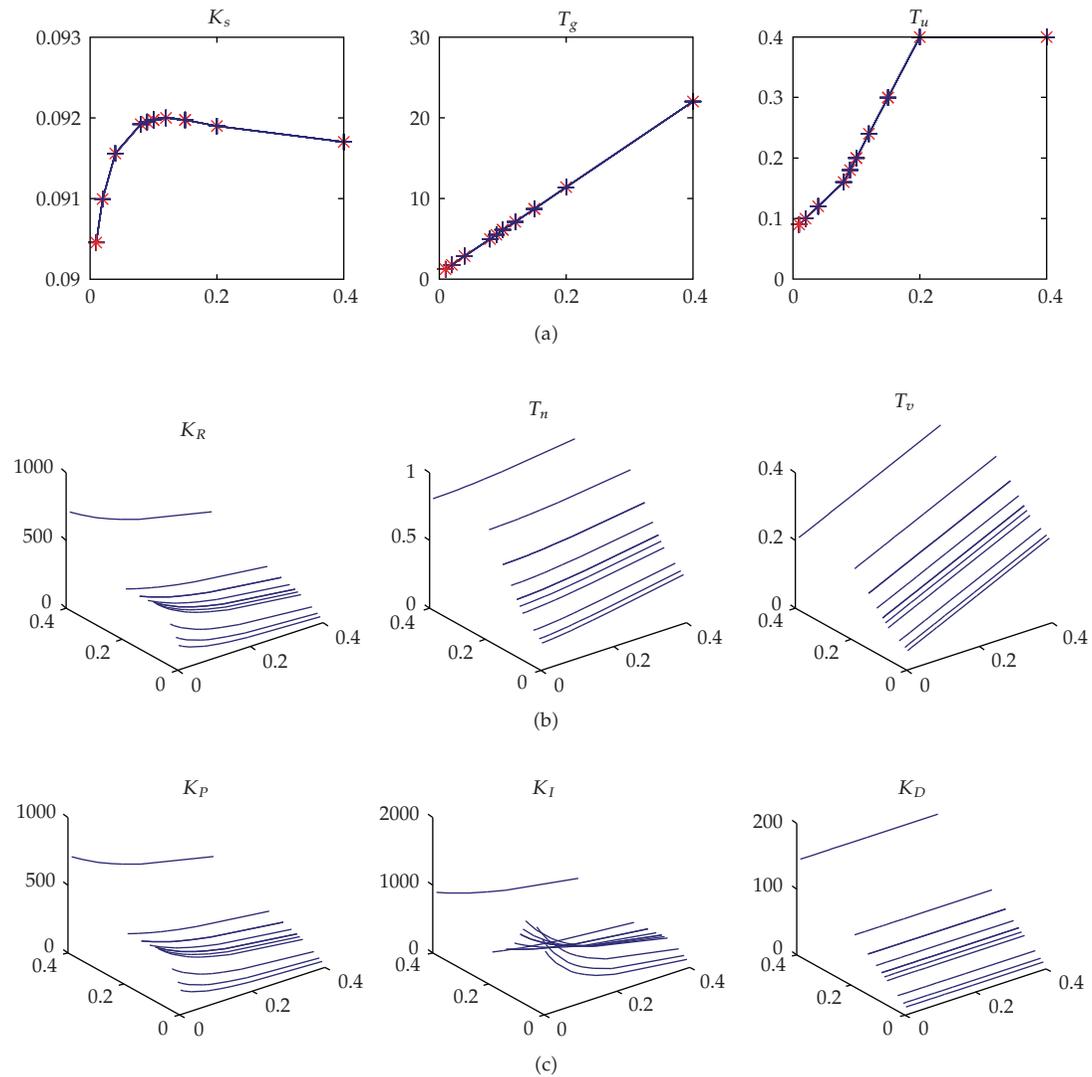
Figure 23: Δt dependence parameters determined by Takahashi.

(5) We restart the computation with the local control parameters and smaller step-size parameters $\text{tol}_{\text{local}}$, $\Delta t_{\text{max,local}}$ and $\Delta t_{\text{min,local}}$ till we reach t_{oszill} and continue the computation with the global parameters.

(6) We stop the computation if we reach $t = T$. If we obtain also high oscillation with the local parameters, we refine the local interval and go to step (3).

Remark 6.5. The modified automatically step-size control had taken into account the local behavior of the control problem. We could adapt the control parameters K_P, K_I, and K_D with respect to the local time-intervals. This modified algorithm considers a local time behavior more accurate and reduces oscillations at the initialization process.
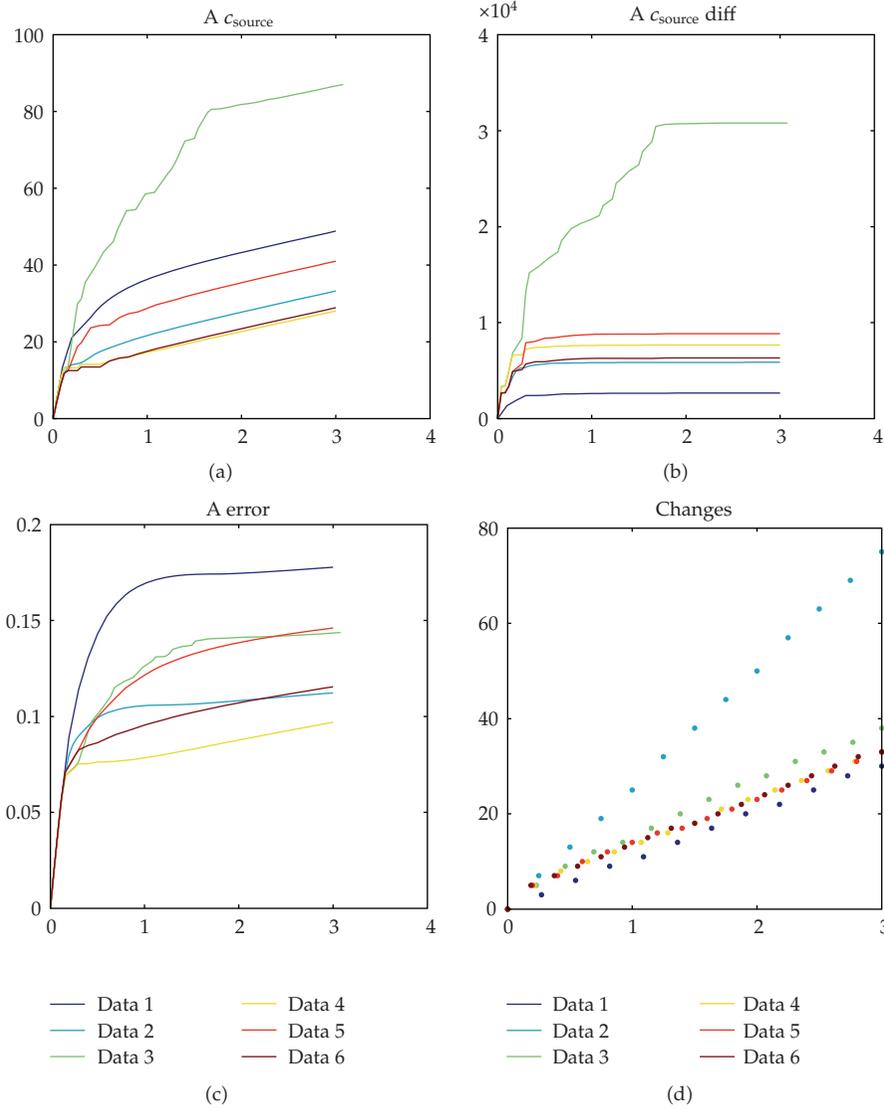
Figure 24: Parameters chosen by Chien-Hrones-Reswick: Summands of the minimization problem.

6.6. Simulation of an Optimal Control of a Diffusion Equation with Adaptive Control and Recovering of the Control Parameters

In what follows, we present the adaptive control based on Algorithm 6.4.

To be more precise the computation of global control parameters $K_{P,\text{global}} = K_P(\Delta t)$, $K_{I,\text{global}} = K_I(\Delta t)$, and $K_{D,\text{global}} = K_D(\Delta t)$ can be automatized in an interval $\Delta t \in (0, T)$.

We improve the automatically step-size, which is given in Section 4.3 to

$$(\Delta t)_{n+1} = \left(\frac{e_{n-1}}{e_n} \right)^{K_P(\Delta t)} \left(\frac{\text{tol}}{e_n} \right)^{K_I(\Delta t)} \left(\frac{e_{n-1}^2}{e_n e_{n-2}} \right)^{K_D(\Delta t)} (\Delta t)_n, \qquad (6.12)$$

where tol is the tolerance, e_n is the error of the quantities of interest in time-step n.
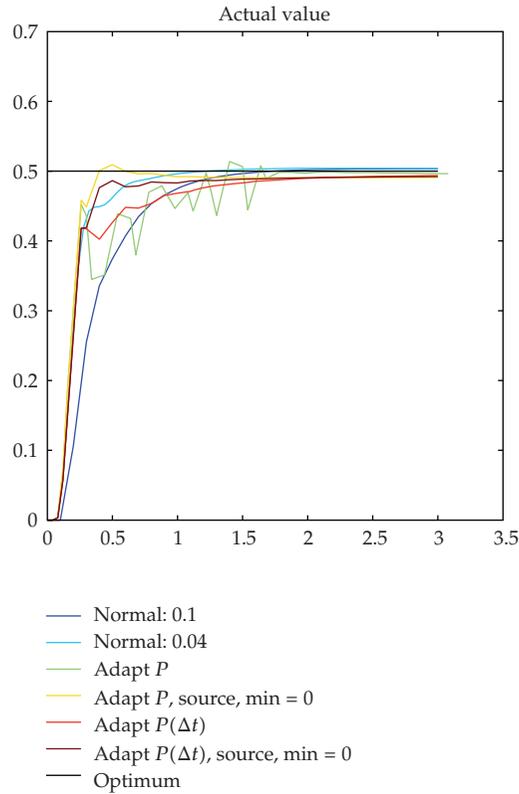
Figure 25: Parameters chosen by Chien-Hrones-Reswick: Actual value.

With approximations to each subintervals of $(0, T)$, we derive the control parameters $K_P(\Delta t)$, $K_I(\Delta t)$, and $K_P(\Delta t)$.

The improved globalized control parameters are computed in the following algorithm.

Algorithm 6.6. Computing of the control parameters $K_{P,\text{global}} = K_P(t)$, $K_{I,\text{global}} = K_I(t)$ and $K_{D,\text{global}} = K_D(t)$ for the time-interval $0, T$.

(1) Compute the critical control function with the minima and maxima. Based on the optimal oscillated function, we can derive the $K_P(T/3)$ for the interval $(0, T/3)$.

(2) Redo the step (1) for the intervals $(T/3, 2T/3)$ and $2T/3, T$ and approximate the function $K_P(t)$. Based on the $K_P(t)$ we can derive the $K_I(t)$ and $K_D(t)$ function.

(3) The optimal control parameters are used in Algorithm 6.4.

Remark 6.7. Further ideas to determine the time-dependent parameters $K_P(\Delta t)$, $K_I(\Delta t)$, and $K_D(\Delta t)$ can be done by using our identification method from Algorithm 4.4 in Section 4.4. There we have discussed two methods to determine the parameters from the control path, namely, Chien-Hrones-Reswick and Takahashi. Alternatively, we have the method from Algorithm 4.2 (Nichols-Ziegler) in Section 4.2. For the identification of the control path, we have only look to the step response. The alternative method demanded to find the critical P-value.
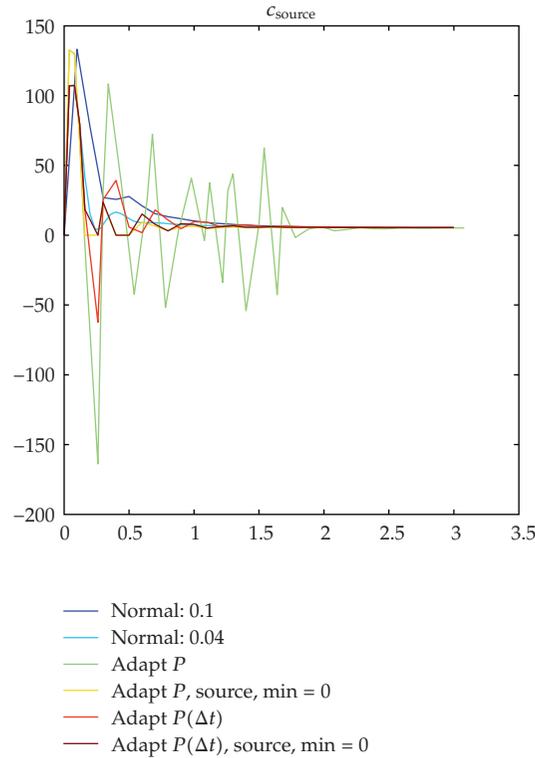
Figure 26: Parameters chosen by Chien-Hrones-Reswick: Source.

6.7. Determinate of Critical P with Exponential Regression

To initialize the PID-controller, we have to derive the K_P, K_D, and K_I parameters.

An idea is derived with the step function response, see [15].

This idea has also be included into the time-step control of the PID-controller.

In Figure 19, we can see that the maximum and minimum of the error function, done with the step function response, have a exponential behavior.

So we use the exponential regression

$$y = b \cdot \exp(a \cdot t) + \epsilon, \tag{6.13}$$

with $\epsilon \sim N(0, \sigma^2)$ normal distributed. Via log the exponential regression can be formed in a linear regression:

$$\begin{aligned} y &= b \cdot \exp(a \cdot t), \\ \log(y) &= \log(b) + a \cdot t. \end{aligned} \tag{6.14}$$

Before we have to convert the minima into the positive area.

Let $\lambda(n)$ be the actual lambda value, $\lambda(n)^+$ the next greater λ, and $\lambda(n)^-$ the next smaller one. Firstly, we use

$$
\lambda(n+1) = \begin{cases} \dfrac{(\lambda(n) + \lambda(n)^+)}{2} : & a(\max, n) < 0, \ a(\min, n) > 0 \ (\lambda \ \text{to small}), \\[2ex] \dfrac{(\lambda(n) + \lambda(n)^-)}{2} : & a(\max, n) > 0, \ a(\min, n) < 0 \ (\lambda \ \text{to great}). \end{cases} \tag{6.15}
$$

In the first step and in limit case, we use $\lambda(n+1) = 2 \cdot \lambda(n)$, respectively, $\lambda(n+1) = \lambda(n)/2$. In the first experiment, we start with $\lambda = 359$, which we have found with manual tests. The automatic found that the value 396.729 the a-value of $\lambda = 3.59$ have 3 zeros, the new less than five.

Remark 6.8. The adaptive step-size control of the PID-controller had taken into account the optimization in each local time-interval. We consider a larger time-interval and derive the optimal control parameters $K_{P,\text{global}}$, $K_{I,\text{global}}$, and $K_{D,\text{global}}$. Based on this parameterization, we apply the adaptation in time to be optimal in such a large time-interval.

6.8. Results of Our Experiments to Find the Time-Dependent Parameters

To automatize the adaptive time-step algorithm, we have to compute the step response and derive the K_P, K_I, and K_D parameters of this process. Here, we present the experiments in Figure 21, which had taken into account the critical values P_{crit} and T_{crit} to derive the optimal K_P, K_D, and K_I parameters based on Algorithm 6.6.

A further algorithm to derive the control parameters is done by Chien et al., see [12]. The results are given in Figure 23.

The K_P, K_D, and K_I parameters are derived with Algorithm 6.6.

6.9. Experiments with Time-Dependent Parameters

In these experiments, we concentrate on the Chien-Hrones-Reswick method, see [12].

To automatize the adaptive time-step algorithm, we have to compute the step response and derive the K_P, K_I, and K_D parameters of this process. The minimization problem is presented in Figure 24, which had taken into account the critical parameters.

In Figure 25, we see the step function response and the best fit of this function with the adaptive time control method. Because of the sensitivity to the initial problem, we can overcome the oscillation with the Chien-Hrones-Reswick method.

In Figure 26, we see the comparison of our different adaptive methods. Here the best method is the benefit of the time adaptive control method based on Algorithm 6.6.

Remark 6.9. The benefit of the time-dependent control parameters $K_P(\Delta t)$, $K_D(\Delta t)$, and $K_I(\Delta t)$ are important in the initialization of the control problem. Based on the fine-time scales globally chosen, parameters neglect the local time behavior. In the experiments, we found out the accelerated relaxation behavior and the fast control to the optimal value. Such adaptive algorithms to optimize the control parameters can improve the control methods and had taken into account the local-time scales.

7. Conclusions and Discussions

We present a continuous or kinetic model, due to the fare-field or near-field effect of our deposition process. We discuss the PID-controller to automatize our deposition process. Due to heuristic methods of deriving the PID parameters, we discuss aposteriori error estimates to automatize the time-stepping methods. A modified automatically step-size control is discussed and the best approximations are obtained with the time-dependent control method based on the Chien-Hrones-Reswick algorithm. For the mesoscopic-scale model, we discussed different experiments and their convergence rates. In future, we will analyze the validity of the models with physical experiments.

References

[1] V. Hlavacek, J. Thiart, and D. Orlicki, "Morphology and film growth in CVD reactions," *Journal de Physique IV*, vol. 5, pp. 3–44, 1995.

[2] H. Rouch, "MOCVD research reactor simulation," in *Proceedings of the COMSOL Users Conference*, pp. 1–7, Paris, France, November 2006.

[3] M. A. Lieberman and A. J. Lichtenberg, *Principle of Plasma Discharges and Materials Processing*, John Wiley & Sons, New York, NY, USA, 2nd edition, 2005.

[4] M. Ohring, *Materials Science of Thin Films*, Academic Press, San Diego, Calif, USA, 2nd edition, 2002.

[5] S. Middleman and A. K. Hochberg, *Process Engineering Analysis in Semiconductor Device Fabrication*, McGraw-Hill, New York, NY, USA, 1993.

[6] M. W. Barsoum and T. El-Raghy, "Synthesis and characterization of a remarkable ceramic: Ti_3SiC_2," *Journal of the American Ceramic Society*, vol. 79, no. 7, pp. 1953–1956, 1996.

[7] C. Lange, M. W. Barsoum, and P. Schaaf, "Towards the synthesis of MAX-phase functional coatings by pulsed laser deposition," *Applied Surface Science*, vol. 254, no. 4, pp. 1232–1235, 2007.

[8] P. Eklund, A. Murugaiah, J. Emmerlich, et al., "Homoepitaxial growth of Ti-Si-C MAX-phase thin films on bulk Ti_3SiC_2 substrates," *Journal of Crystal Growth*, vol. 304, no. 1, pp. 264–269, 2007.

[9] T. K. Senega and R. P. Brinkmann, "A multi-component transport model for non-equilibrium low-temperature low-pressure plasmas," *Journal of Physics D*, vol. 39, no. 8, pp. 1606–1618, 2006.

[10] J. Geiser and M. Arab, "Modelling, Optimzation and Simulation for a Chemical Vapor Deposition," *Journal of Porous Media, Begell House Inc., Redding, USA*, June 2008.

[11] A. M. P. Valli, G. F. Carey, and A. L. G. A. Coutinho, "Control strategies for timestep selection in simulation of coupled viscous flow and heat transfer," *Communications in Numerical Methods in Engineering*, vol. 18, no. 2, pp. 131–139, 2002.

[12] K. L. Chien, J. A. Hrones, and J. B. Reswick, "On the automatic tuning of generalized passive systems," *Transactions of the ASME*, vol. 74, pp. 175–185, 1952.

[13] M. K. Gobbert and C. A. Ringhofer, "An asymptotic analysis for a model of chemical vapor deposition on a microstructured surface," *SIAM Journal on Applied Mathematics*, vol. 58, no. 3, pp. 737–752, 1998.

[14] H. H. Lee, *Fundamentals of Microelectronics Processing*, McGraw-Hill, New York, NY, USA, 1990.

[15] N. B. Nichols and J. G. Ziegler, "Optimum settings for automatic controllers," *Transactions of the ASME*, vol. 64, pp. 759–768, 1942.

[16] H. Lutz and W. Wendt, *Taschenbuch der Regelungstechnik*, Harri-Deutsch, Frankfurt, Germany, 6th edition, 2005.

[17] P. Bastian, K. Birken, K. Johannsen, et al., "UG—a flexible software toolbox for solving partial differential equations," *Computing and Visualization in Science*, vol. 1, no. 1, pp. 27–40, 1997.

[18] J. Geiser, "Discretization methods with embedded analytical solutions for convection-diffusion dispersion-reaction equations and applications," *Journal of Engineering Mathematics*, vol. 57, no. 1, pp. 79–98, 2007.

[19] R. J. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, UK, 2002.

[20] W. Hundsdorfer and J. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, vol. 33 of *Springer Series in Computational Mathematics*, Springer, Berlin, Germany, 2003.

[21] J. Lee and Th. F. Edgar, "Continuation method for the modified Ziegler-Nichols tuning of multiloop control systems," *Industrial and Engineering Chemistry Research*, vol. 44, no. 19, pp. 7428–7434, 2005.

[22] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations. II: Stiff and Differential-Algebraic Problems*, vol. 14 of *Springer Series in Computational Mathematics*, Springer, Berlin, Germany, 2nd edition, 1996.

[23] I. Faragó and Á. Havasi, "On the convergence and local splitting error of different splitting schemes," *Progress in Computational Fluid Dynamics*, vol. 5, no. 8, pp. 495–504, 2005.

[24] K.-J. Engel and R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, vol. 194 of *Graduate Texts in Mathematics*, Springer, New York, NY, USA, 2000.

[25] T. Yamaguchi and K. Shimizu, "Asymptotic stabilization by PID control: stability analysis based on minimum phase and high-gain feedback," *Electrical Engineering in Japan*, vol. 156, no. 1, pp. 44–53, 2006.

Research Article

A New Mechanical Model for Particle Transport by Surface Waves and Applications

Minvydas Ragulskis,[1] Edita Sakyte,[1] Jesús M. Seoane,[2] and Miguel A. F. Sanjuán[2]

[1] *Department of Mathematical Research in Systems, Kaunas University of Technology, Studentu 50-222, 51638 Kaunas, Lithuania*

[2] *Nonlinear Dynamics, Chaos and Complex Systems Group, Department of Physics, University Rey Juan Carlos, Tulipán s/n, 28933 Móstoles, Madrid, Spain*

Correspondence should be addressed to Miguel A. F. Sanjuán, miguel.sanjuan@urjc.es

We present a study of the behavior of a ball under the influence of gravity on a platform. A propagating surface wave travels on the surface of the platform while the platform remains motionless. This is a modification of the classical bouncing ball problem and describes the transport of particles by surface waves. Phase and velocity maps cannot be expressed in the explicit form due to implicit formulations, and no formal analytical analyses is possible. Numerical analysis shows that the transition to chaos is produced via a period doubling route which is a common property for classical bouncers. These numerical analysis have been carried out for the conservative and for the viscous cases and also for elastic and for inelastic collisions. The bouncing process can be sensitive to the initial conditions and can be useful for control techniques which can dramatically increase the effectiveness of particle transport in practical applications. Finally, we also consider the mechanical model of a particle sliding on a surface which is also important because it has important physical implications such as the transportation of thin films in biomedical applications, among others.

1. Introduction

A particle falling down, in a constant gravitational field, on a moving platform is called a bouncing ball problem, or a bouncer. This model was suggested more than thirty years ago [1, 2] as an alternative to the Fermi-Ulam model [3] of cosmic ray acceleration [4]. In the ensuing years many approaches to the bouncer model have been studied theoretically and experimentally [5–8]. It has been proved to be a useful system for experimentally exploring several new nonlinear effects [9, 10]. Moreover, it has been implemented into a number of engineering applications [11, 12].

The bouncer model can be briefly characterized by the following basic statements. (i) Maps derived for the bouncer model can be exactly iterated for any time function describing the moving platform [7, 9] (though usually the platform is assumed to oscillate with a single frequency). (ii) The ball-platform collisions can be characterized by a coefficient of restitution α changing from $\alpha = 1$ for a perfectly elastic case to $\alpha = 0$ for a completely inelastic situation. (iii) The chaotic bouncer can be easily used to relate theoretical predictions to experimental results, [9, 10] what makes it a paradigm model in nonlinear dynamics.

In this paper we assume that a particle is falling down in a constant gravitational field on a stationary platform. A propagating surface wave travels on the surface of the platform while the platform remains motionless. Such a model can be used to describe the transport of particles by propagating surface waves, which is an important problem with numerous applications. Powder transport by piezoelectrically excited ultrasonic surface waves [13], manipulation of bioparticles using traveling wave electrophoresis [14, 15], and conveyance of submerged buoys in coastal waters [16] are just a few examples of problems involving the interaction between propagating waves and transported bouncing particles.

This paper is organized as follows. In Section 2 we present a complete description of our model, the bouncer system. Section 3 presents numerical simulations showing that the modified bouncer model possesses such an inherent chaotic dynamics. These results are carried out for both, the conservative and the dissipative cases. The case of a particle sliding on the surface is fully analyzed in Section 4. Conclusions and discussions of the main results of this paper are presented in Section 5.

2. Model Description

We consider the two-dimensional system shown in Figure 1, where the surface of an elastic plate is represented by a solid line which coincides with the x-axis in the state of equilibrium. A point of the surface in the state of equilibrium $(x, 0)$ is translated to coordinates (X, Y) when a wave process takes place. This translation is sensitive to time t and coordinate x:

$$X = x + \eta(x, t),$$
$$Y = \zeta(x, t), \tag{2.1}$$

where the functions $\eta(x, t)$ and $\zeta(x, t)$ determine deflections from the state of equilibrium.

Explicitly, the longitudinal and transverse displacements of the medium at the surface of flat boundary with travelling Rayleigh wave can be expressed like [17]

$$u_x = \frac{\omega}{\chi} \sqrt{\frac{2\rho(1+v)}{E}} \left(1 - \frac{\sqrt{1-\chi^2}\sqrt{2(1-v) - \chi^2(1-2v)}}{(1 - 0.5\chi^2)\sqrt{2(1-v)}} \right) C \sin(\omega t \mp kx),$$

$$u_y = \frac{\omega}{\chi} \sqrt{\frac{2\rho(1+v)}{E}} \left(1 - \frac{\sqrt{2(1-v) - \chi^2(1-2v)}}{\sqrt{2(1-v)}} \right) 0.5\chi^2 C \cos(\omega t \mp kx), \tag{2.2}$$
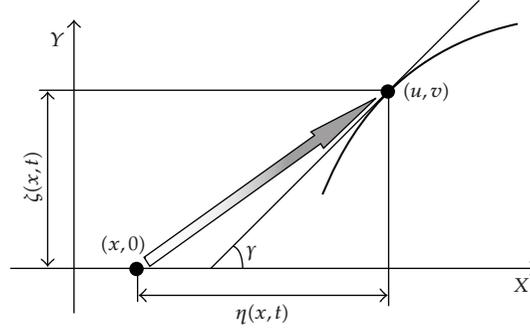
Figure 1: Schematic diagram illustrating the collision between the particle and the surface.

where u_x and u_y are the longitudinal and transverse displacements, x the coordinate of the surface point of the medium before the wave process took place; C is a constant, k the wavenumber and ρ the density. χ can be found from the following algebraic equation:

$$\chi^6 - 8\chi^4 + 8\left(3 - \frac{1 - 2\nu}{1 - \nu}\right)\chi^2 - 16\left(1 - \frac{1 - 2\nu}{2(1 - \nu)}\right) = 0, \tag{2.3}$$

and the angular velocity ω can be found from the following transcendental equation:

$$\left(k^2 + \beta^2\right)^2 \cosh(\alpha a)\sinh(\beta a) - 4k^2 \alpha\beta \cosh(\beta a)\sinh(\alpha a) = 0, \tag{2.4}$$

where $\alpha = \sqrt{(v_l k^2 - \omega^2)/v_l^2}$, $\beta = \sqrt{(v_t k^2 - \omega^2)/v_t^2}$, $v_l = \sqrt{(\lambda + 2\mu)/\rho}$ and $v_t = \sqrt{\mu/\rho}$, ν the Poisson's ratio and λ the first Lamé constant. It can be noted that the ratio between the amplitudes of transverse and longitudinal deformations depends on ν. In usual elastic media it is quite normal that the transverse displacement is about 1.5 times larger than the longitudinal displacement [18]. The motion of a point in the medium is an ellipse. Also, the direction of the velocity of the particles at the peaks of the wave is opposite to the direction of wave propagation.

Rayleigh waves are dispersive due to a dependence of the wave's speed on its wavelength. Typical example is Rayleigh waves in the Earth where waves with a higher frequency travel more slowly than those with a lower frequency. Rayleigh waves thus often appear spread out on seismograms recorded at distant earthquake recording stations [19]. Surface acoustic waves (SAWs) generated by SAW devices on rough anisotropic materials also experience considerable dispersion [20]. On the other hand, film waves generated on a surface of a finite liquid bed [21] can be characterized by a single frequency wave component. Therefore we concentrate on a one single frequency steady-state Rayleigh wave propagation and disregard dispersion.

Whenever a traveling nondispersive Rayleigh surface wave occurs in a medium, it can be characterized by a retrograde elliptic motion of the particles of that medium:

$$\eta(x, t) = a\,\sin(\omega t - kx),$$
$$\zeta(x, t) = b\,\cos(\omega t - kx), \tag{2.5}$$

where a and b are longitudinal and transverse amplitudes of the oscillations; ω is the angular frequency, and k is the wave number. Remind that in a usual elastic medium it is quite normal for the transverse displacement to be about 1.5 times larger than the longitudinal displacement [18].

The coordinates of the particle are denoted as (u, v). Assume that the particle is in contact with the surface at time moment t, then the following constrain takes place:

$$v = \zeta(x, t), \tag{2.6}$$

where x is to be found from the following algebraic equality (where u and t are given and x is the unknown):

$$x + \eta(x, t) = u. \tag{2.7}$$

In other words, the instantaneous shape of the surface cannot be described by an explicit function. Nevertheless, the tangent to the surface at the point with abscise u can be expressed explicitly:

$$\tan \gamma = \frac{\partial \zeta(x, t) / \partial x}{1 + \partial \eta(x, t) / \partial x}, \tag{2.8}$$

where γ is the angle between the tangent and the x-axis. Instantaneous velocities (x- and y-components) of the point of the surface in contact with the particle can be expressed as $\partial \eta(x, t) / \partial t$ and $\partial \zeta(x, t) / \partial t$ accordingly.

The governing equations of motion of a particle in a free flight mode are

$$m\ddot{u} + h\dot{u} = 0,$$
$$m\ddot{v} + h\dot{v} = -mg, \tag{2.9}$$

where top dots denote full derivative by time, m is the mass of the particle, h is the coefficient of viscous damping of the media above the surface, and g is the free fall acceleration. Initial conditions $u(t_0) = u_0;\ \dot{u}(t_0) = \dot{u}_0;\ v(t_0) = v_0;\ \dot{v}(t_0) = \dot{v}_0$ yield partial solutions:

$$u(t) = u_0 + \dot{u}_0 \frac{m}{h} \left(1 - \exp\left(-\frac{h}{m}(t - t_0) \right) \right);$$

$$\dot{u}(t) = \dot{u}_0 \exp\left(-\frac{h}{m}(t - t_0) \right);$$

$$v(t) = v_0 + \left(\dot{v}_0 + \frac{mg}{h} \right) \frac{m}{h} \left(1 - \exp\left(-\frac{h}{m}(t - t_0) \right) \right) - \frac{mg}{h}(t - t_0); \tag{2.10}$$

$$\dot{v}(t) = \dot{v}_0 \exp\left(-\frac{h}{m}(t - t_0) \right) - \frac{mg}{h} \left(1 - \exp\left(-\frac{h}{m}(t - t_0) \right) \right).$$

The free flight stage continues until the particle collides with the surface. Unfortunately, it is impossible to determine the explicit time moment of the collision due to the

fact that the instantaneous shape of the surface cannot be expressed by an explicit function. Instead, one has to use iterative numerical techniques in order to determine the exact moment of the bounce.

Localization of the root (the time moment of the collision) is performed using a time marching technique starting from the initial conditions until

$$v(t_0 + i \cdot \Delta t) < \zeta(x_i, t_0 + i \cdot \Delta t) \tag{2.11}$$

where Δt is the time step; $i = 1, 2, \ldots, r$; r is the step number for which (4.8) is satisfied for the first time, and x_i is the solution of (4.4) at fixed i:

$$x_i + \eta(x_i, t_0 + i \cdot \Delta t) = u(t_0 + i \cdot \Delta t); \tag{2.12}$$

and $u(t_0 + i \cdot \Delta t)$, $v(t_0 + i \cdot \Delta t)$ are determined by (4.7). Solution of (4.9) also requires an iterative numerical algorithm.

When the root $\widehat{t}$ is localized in the interval $t_0 + (r-1) \cdot \Delta t < \widehat{t} \le t_0 + r \cdot \Delta t$, one needs to fine down the value of $\widehat{t}$ using an iterative computational algorithm. This iterative algorithm can be a most simple bisection method, though more sophisticated algorithms comprising the golden section rule or Newton's iterations for example can be used instead until the desirable accuracy is achieved. As the collision moment $\widehat{t}$ is fined down in every iteration, the coordinate $\widehat{x}$ (corresponding to the collision point $\widehat{u} : \widehat{x} + \eta(\widehat{x}, \widehat{t}) = \widehat{u}$) is also made more precise. Initially, $x_{r-1} < \widehat{x} \le x_r$; every iteration helps to reach a better accuracy.

Such iterative method of determination of the collision moment leads us to the important conclusion that phase and velocity maps cannot be expressed in an explicit form, and no formal analytical analysis is possible.

Nevertheless, the geometrical coordinates of the point of collision are $(\widehat{u}; \zeta(\widehat{x}, \widehat{t}))$ and can be reconstructed using computational techniques. Velocities of the particle just before the collision are $\dot{u}(\widehat{t})$ and $\dot{v}(\widehat{t})$. Similarly, instantaneous velocities of the surface in contact with the particle are $\eta'_t(\widehat{x}, \widehat{t})$ and $\zeta'_t(\widehat{x}, \widehat{t})$.

Projections of the particle's velocities just before the collision to the normal and to the tangent to the surface at the contact point can be expressed in the following form:

$$\begin{aligned} \widehat{P}_n &= -\dot{u}\left(\widehat{t}\right)\sin\gamma + \dot{v}\left(\widehat{t}\right)\cos\gamma, \\ \widehat{P}_t &= \dot{u}\left(\widehat{t}\right)\cos\gamma + \dot{v}\left(\widehat{t}\right)\sin\gamma, \end{aligned} \tag{2.13}$$

where the angle γ is determined from (4.5) at the point of collision.

Analogously, projections of velocities of the point of the surface in contact with the particle to the normal and to the tangent take the following form:

$$\begin{aligned} \dot{S}_n &= -\eta'_t\left(\widehat{x}, \widehat{t}\right)\sin\gamma + \zeta'_t\left(\widehat{x}, \widehat{t}\right)\cos\gamma, \\ \dot{S}_t &= \eta'_t\left(\widehat{x}, \widehat{t}\right)\cos\gamma + \zeta'_t\left(\widehat{x}, \widehat{t}\right)\sin\gamma. \end{aligned} \tag{2.14}$$

Then, the velocities of the particle just after the collision (in the normal and tangent directions) are

$$
\dot{P}_n = (1 + \alpha)\dot{S}_n - \alpha\widehat{\dot{P}}_n,
$$
$$
\dot{P}_t = \beta\dot{S}_n - (1 + \beta)\widehat{\dot{P}}_t,
$$

(2.15)

where α is the coefficient of restitution for the collision in the normal direction. This constant is a measure of the energy loss at each impact. For elastic collisions $\alpha = 1$, and $\alpha < 1$ for inelastic collisions. Coefficient β determines the friction between the particle and the surface at the moment of collision. There is no friction between the particle and the surface when $\beta = 0$. The utmost value $\beta = 1$ represents the situation when the projection of the particle's velocity (immediately after the impact) and the projection of the surface's point velocity to the tangent are equal.

The free flight stage starts over again immediately after the collision, and the initial conditions are

$$
u\left(\widehat{t}\right) = \widehat{u},
$$
$$
\dot{u}\left(\widehat{t}\right) = -\dot{P}_n \sin\gamma + \dot{P}_t \cos\gamma,
$$
$$
v\left(\widehat{t}\right) = \zeta\left(\widehat{x}, \widehat{t}\right),
$$
$$
\dot{v}\left(\widehat{t}\right) = \dot{P}_n \cos\gamma + \dot{P}_t \sin\gamma.
$$

(2.16)

The presented model is a modification of the classic bouncer model which can be derived assuming $\zeta(x,t) = b\cos(\omega t)$ and $\eta(x,t) = 0$. In that case $u = x$, and the model becomes explicit.

3. Complex Dynamics in the Modified Bouncer Model

We will demonstrate that the modified bouncer model possesses such an inherent feature as chaotic dynamics. Moreover, we will show that the sensitivity to initial conditions can be exploited for the control of the process of conveyance. We will show these results for the conservative or nonviscous case ($h = 0$) and for the viscous case ($h \neq 0$).

3.1. Nonviscous Case

We take $h = 0$, for which the media above the surface is non-viscous, $\alpha = 1$ (elastic collisions) and $\beta = 0$ (no damping generated by sliding).

The dynamics of a bouncing particle on a surface of a propagating wave is very sensitive to the initial conditions if the dynamics is Hamiltonian. Apparently, it is possible to find such a set of initial conditions which lead to regular and periodic dynamics. This is illustrated in Figure 2 where collision heights $v(\widehat{t})$ are plotted versus initial velocity $\dot{u}(0)$. 50 successive collisions are used for every discrete value of $\dot{u}(0)$ to produce this diagram. The
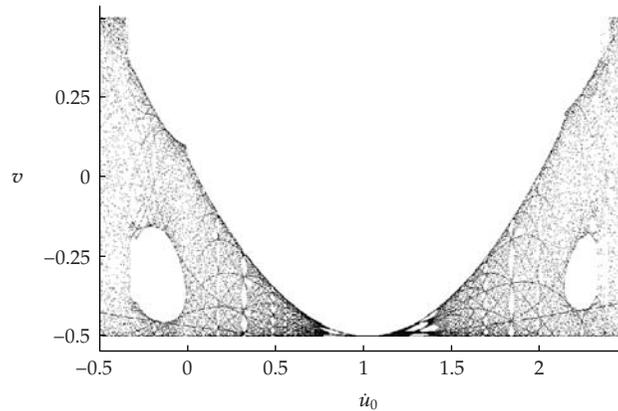
Figure 2: Collision heights $v(\hat{t})$ plotted versus initial velocity $\dot{u}(0)$ for $\dot{u}(0) = \omega/k$ at $\omega = 1$, $k = 1$, $m = 0.5$, $g = 1$, $a = b/1.5$, $u(0) = \pi$, $v(0) = 0$, and $\dot{v}(0) = -0.5$.
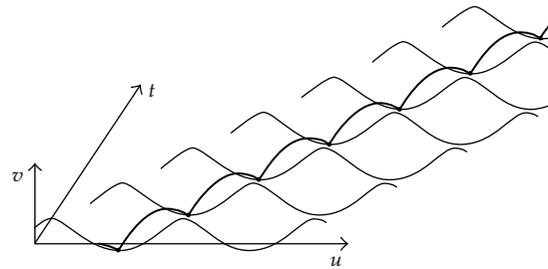
Figure 3: Period 1 trajectory of a bouncing particle for $\omega = 1$, $k = 1$, $m = 0.5$, $g = 1$, $a = b/1.5$, $u(0) = \pi$, $\dot{u}(0) = 1$, $v(0) = 0$, and $\dot{v}(0) = -0.5$.
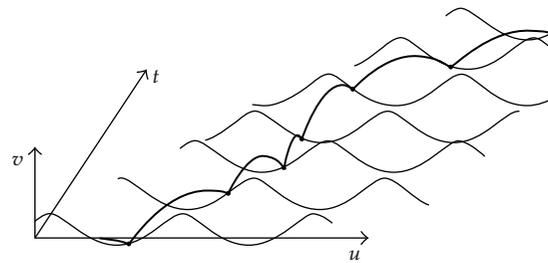
Figure 4: Chaotic trajectory of a bouncing particle for $\omega = 1$, $k = 1$, $m = 0.5$, $g = 1$, $a = b/1.5$, $u(0) = \pi$, $\dot{u}(0) = 2$, $v(0) = 0$, and $\dot{v}(0) = -0.5$.

initial condition $\dot{u}(0) = \omega/k$ produces a period 1 motion at $\omega = 1$, $k = 1$, $m = 0.5$, $g = 1$, $a = b/1.5$, $u(0) = \pi$, $v(0) = 0$ and, $\dot{u}(0) = -0.5$. This is illustrated in Figure 3.

We plot the trajectory of the particle in 3D for better visual interpretation. Collision moments are marked as black dots. At every moment of collision we also plot the instantaneous shape of the surface (one can note that the instantaneous shape of the surface is not harmonic). One can clearly see the difference in the complexity of the particle dynamics at $\dot{u}(0) = 1$ (see Figure 3) and $\dot{u}(0) = 2$ (see Figure 4).

(a)

(b)
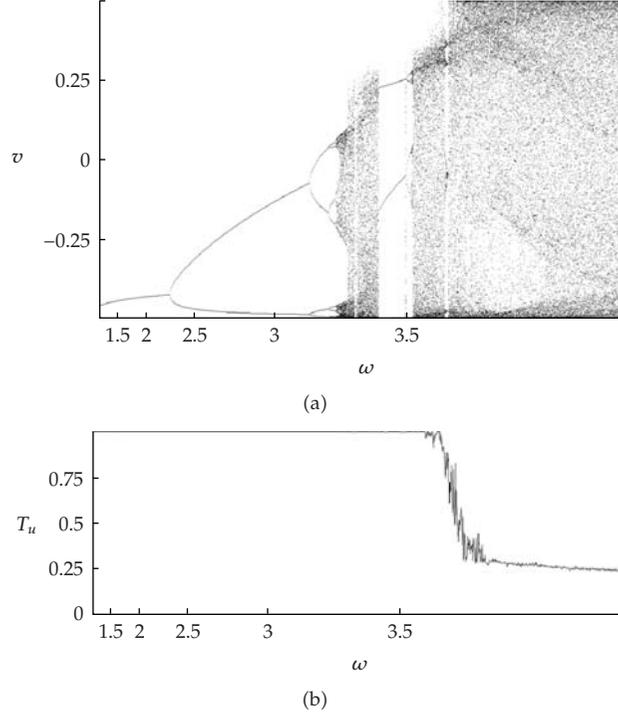
Figure 5: Transport of particles at increasing wave speeds (elastic collisions, viscous media over the surface). Reduced impact representation (a) shows the transition to chaos via a period doubling route. Note that impact heights are distributed in the interval $[-0.5, 0.5]$. Nondimensional longitudinal particle's transport velocity T_u drops down at higher wave speeds due to the viscosity of the media above the surface (b). System's parameters are $\alpha = 1$; $\beta = 0$; $\eta(x,t) = (2/3)\sin(\omega t - x)$; $\zeta(x,t) = (1/2)\cos(\omega t - x)$; $h = 0.1$; $m = 0.5$; $g = 9.81$.

3.2. Viscous Case

For this case, we assume that collisions are completely elastic ($\alpha = 1$), and there is no tangential friction between the particle and the surface ($\beta = 0$), but the media above the surface is viscous fixing the value of $h = 0.1$ as in [22]. We use the reduced impact representation, where the height of the bouncing ball is sampled at each impact with the surface (impact sampling). Since the system is dissipative, we plot the bouncing process after the initial transients cease down (see Figure 5(a)). We skip 1500 successive bounces before starting to plot the collision heights $v(t)$ for every discrete value of ω. Parameter ω is varied following the rule $\omega_i = 1 + (3/\ln 21)\ln(1 + 20i/1024)$; $i = 1, \ldots, 1024$, which helps to expand the cascade of period doubling bifurcations. The control parameter in our case is not the amplitude of the platform's oscillation but the velocity of the wave propagation; the collision height is used instead of collision velocity for a reduced impact representation. Moreover, the media above the surface of the plate is viscous. It appears that the transition to chaos via a period doubling route observed for a classical bouncer [23] is observed also for a particle bouncing on a surface of a stationary platform with a propagating wave traveling on its surface.

A phenomenological model could be used to exemplify the bifurcation diagram presented in Figure 5(a). The logistic map [24] is probably the simplest model ever used to

study the transition to chaos via a period doubling route. Simple computational experiments with appropriately chosen parameter values of the logistic map would illustrate the universality of the bifurcation diagram in Figure 5(a).

An important parameter characterizing the effectiveness of the transport is the average longitudinal velocity of the particle $\bar{u}$. We average it over a long period of time after the initial transients cease down. In order to calculate a nondimensional quantity we divide it from the velocity of the traveling wave $T_u = k\bar{u}/\omega$. Thus, the average velocity of conveyance is equal to the velocity of the traveling wave if T_u is equal to 1 as shown in Figure 5(b).

It is interesting to observe that the particle is transported with the average velocity of the traveling wave until the period 3 bouncing mode after a cascade of period doubling bifurcations (see Figure 5(a)). The particle's average transportation velocity drops down only when the period 3 bouncing mode experiences its own cascade of period doubling bifurcations. External damping forces acting to the particle prevent its motion with the average wave's velocity in the direction of the wave propagation when this velocity becomes large enough (even though the collisions are elastic). Also, the bouncing process is insensitive to initial conditions–eventually it converges to the one and only attractor shown in Figure 5(a) (at fixed ω).

Figures 6(a)–6(d) show different dynamical behaviors of the transient processes for the elastic case ($\alpha = 1$) once we fixed the parameter values as follows: $\beta = 0$, $k = 1$, $m = 0.5$, $g = 9.81$, $b = 0.5$, and $a = b/1.5$. We obtain both, periodic and chaotic motions depending on the value of the parameter ω. Figures 6(a)–6(c) show for $\omega = 2$, $\omega = 3$, and $\omega = 3.45$ period 1, period 2, and period 3 processes, respectively. Chaotic bouncing for the value $\omega = 4$ is shown in Figure 6(d).

The situation becomes different when collisions are inelastic, as shown in [22]. In the presence of inelastic collisons, a vanishing bouncing process takes place (complete chattering [23]) when the particle sets into the state of rest on a slope of the propagating wave. The term *complete chattering* is used in literature to describe the process when the time interval between inelastic bounces tends to zero and the ball finally "sticks" to the surface of the oscillating platform. As we mentioned previously, a complete description of this phenomenon is given in [22].

4. The Sliding Particle Model

In this section we thoroughly analyze the case in which the particle is sliding on the surface instead of the case in which it is falling down on it, analyzed previously. Our motivation is the following. Conveyance of particles and bodies by propagating waves is an important scientific and engineering problem with numerous applications. Manipulation of bioparticles and gene expression profiling using traveling wave dielectrophoresis [14, 25, 26], segregation of particles in suspensions subject to ac electric fields [27], transport of sand particles and oil spills in coastal waters [28, 29], powder transport by piezoelectrically excited ultrasonic waves [13, 22], transportation of thin films in biomedical applications [21] are just a few examples of problems involving interaction between propagating waves and transported objects.

We now describe, as in the bouncer model, the equations of motions of our sliding particle model.

It is assumed that a mass particle is in contact with the deformed surface at a point (u, v) at a time moment t (Figure 7). A point of the surface in the equilibrium state $(x, 0)$ is

(a)

(b)

(c)

(d)
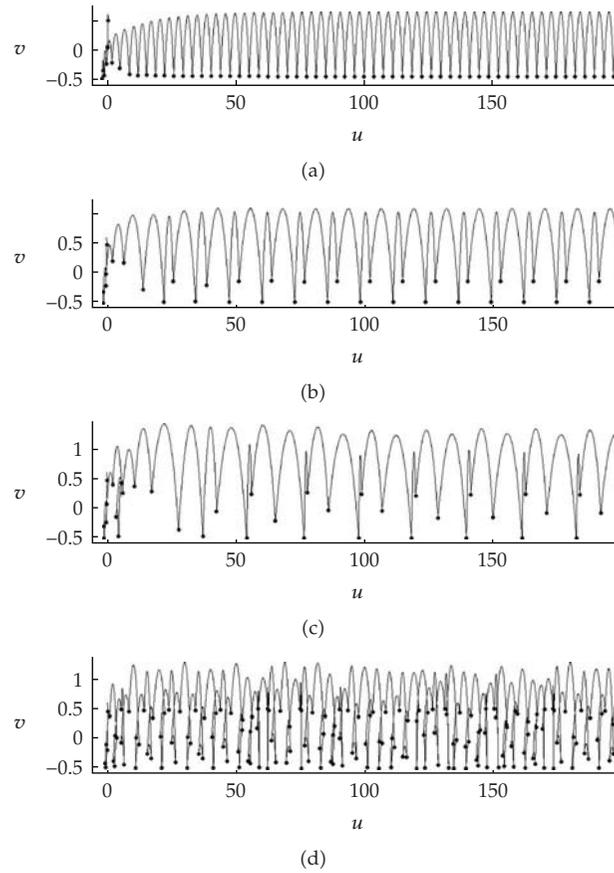
Figure 6: Transient processes for parameter values as follows: $h = 0.1$, $\alpha = 1$, $\beta = 0$, $k = 1$, $m = 0.5$, $g = 9.81$, $b = 0.5$, and $a = b/1.5$. We observe the following behaviors: (a) period 1 process for $\omega = 2$, (b) period 2 process for $\omega = 3$, (c) period 3 process for $\omega = 3.45$, and (d) chaotic bouncing for $\omega = 4$, respectively.

translated to coordinates (u, v) at time moment t. This translation is sensitive to time t and coordinate x:

$$u = x + \eta(x, t),$$

$$v = \zeta(x, t),$$

$$(4.1)$$

where $\eta(x, t)$ and $\zeta(x, t)$ are predefined functions.

The condition that the particle is located on the surface leads to the following constraint:

$$v = \zeta(x, t), \tag{4.2}$$
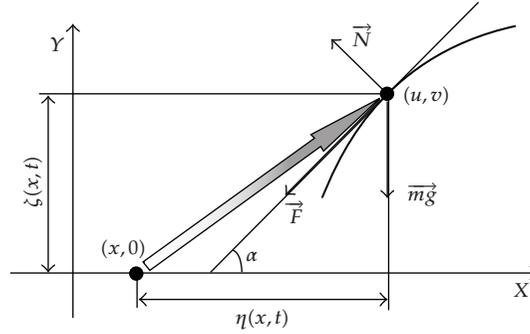
Figure 7: A geometric scheme of the dynamical system showing the particle sliding on the surface.

where x is to be found from the following algebraic equality (in which u is given and x is unknown):

$$x + \eta(x, t) = u. \tag{4.3}$$

Though the instantaneous shape of the oscillating surface cannot be described by an explicit function, the tangent to the surface at the point (u, v) can be expressed as

$$\tan \alpha = \frac{\zeta'_x(x, t)}{1 + \eta'_x(x, t)}. \tag{4.4}$$

Instantaneous velocities of the surface's point (u, v) in the direction of x- and y-axis can be expressed as follows:

$$
\begin{aligned}
\dot{u}|_{x=\text{const}} &= \eta'_t(x, t), \\
\dot{v}|_{x=\text{const}} &= \zeta'_t(x, t),
\end{aligned}
\tag{4.5}
$$

where dots denote derivatives by t.

When a mass particle slides on the surface, it does not necessarily move in contact with one point of the surface. Therefore x is no longer a constant. Thus,

$$
\begin{aligned}
\dot{u} &= \dot{x}\left(1 + \eta'_x\right) + \eta'_t, \\
\ddot{u} &= \ddot{x}\left(1 + \eta'_x\right) + \dot{x}^2 \eta''_{xx} + 2\dot{x}\eta''_{xt} + \eta''_{tt}.
\end{aligned}
\tag{4.6}
$$

The condition that the mass particle continuously slides on the surface brings another constraint into force (the relative velocity in the normal direction to the surface at the contact point must be zero):

$$\tan \alpha = \frac{\dot{v} - \zeta'_t(x, t)}{\dot{u} - \eta'_t(x, t)}. \tag{4.7}$$

Equation (4.4) with (4.7) in force yields

$$\dot{v} = \frac{\dot{u} - \eta'_t}{1 + \eta'_x} \zeta'_x + \zeta'_t, \tag{4.8}$$

which together with (4.6) produces the following relationship:

$$\dot{v} = \dot{x}\zeta'_x + \zeta'_t. \tag{4.9}$$

Differentiation of (4.9) yields

$$\ddot{v} = \ddot{x}\zeta'_x + \dot{x}^2\zeta''_{xx} + 2\dot{x}\zeta''_{xt} + \zeta''_{tt}. \tag{4.10}$$

Then the relative sliding velocity of the particle on the surface v_{12} can be expressed as

$$v_{12} = \left(\dot{u} - \eta'_t\right)\cos\alpha + \left(\dot{v} - \zeta'_t\right)\sin\alpha = \frac{\dot{x}}{\sqrt{1 + \tan^2\alpha}}\left(1 + \eta'_x + \zeta'_x\tan\alpha\right) = \dot{x}\sqrt{\left(1 + \eta'_x\right)^2 + \zeta'^{\,2}_x}. \tag{4.11}$$

The condition of dynamic equilibrium leads to the following system of equations:

$$m\ddot{u} + N\sin\alpha + F\cos\alpha = 0,$$
$$m\ddot{v} + mg + F\sin\alpha = N\cos\alpha, \tag{4.12}$$

where m is the mass of the particle; N is the reaction force at the contact point; g is the gravity acceleration; F is the friction force between the mass particle and the surface. The system of equations in (4.12) is in force when $N > 0$. Otherwise the particle jumps off the oscillating surface.

It is assumed that the friction force is linear. Thus F can be expressed like

$$F = hv_{12}, \tag{4.13}$$

where h is the coefficient of linear friction.

Finally, the governing equation of motion can be derived from (4.12). Elementary transformations and substitutions lead to the following explicit differential equation:

$$B_1(x,t) \cdot \ddot{x} + B_2(x,t) \cdot \dot{x} + B_3(x,t) + B_4(x,t) \cdot (\dot{x})^2 = 0, \tag{4.14}$$

where

$$B_1(x,t) = m\left(1 + \eta'_x + \frac{(\zeta'_x)^2}{1 + \eta'_x}\right),$$

$$B_2(x,t) = 2m\left(\eta''_{xt} + \frac{\zeta'_x \zeta''_{xt}}{1 + \eta'_x}\right) + h\left(1 + \eta'_x + \frac{(\zeta'_x)^2}{1 + \eta'_x}\right),$$

$$B_3(x,t) = m\left(\eta''_{tt} + g\frac{\zeta'_x}{1 + \eta'_x} + \frac{\zeta'_x \zeta''_{tt}}{1 + \eta'_x}\right),$$

$$B_4(x,t) = m\left(\eta''_{xx} + \frac{\zeta'_x \zeta''_{xx}}{1 + \eta'_x}\right).$$

(4.15)

A major obstacle is eliminated, and direct numerical time marching techniques can be used for integration of (4.14)–computation of u and v is straightforward if the coordinate x is given at time t (4.1). Existence of a stability of the dynamic equilibrium can be analyzed explicitly.

But before proceeding with the analysis of dynamic equilibrium the following observation can be done. If kinematic relationships describing a traveling Rayleigh wave are in force, the change of variables

$$z = \omega t - kx$$

(4.16)

transforms (4.14) to the following autonomous form:

$$C_1(z) \cdot \ddot{z} + C_2(z) \cdot \dot{z} + C_3(z) + C_4(z) \cdot (\dot{z})^2 = 0,$$

(4.17)

where

$$C_1(z) = -\frac{m}{k}\left(1 - ka\cos(z) + \frac{k^2 b^2 \sin^2(z)}{1 - ka\cos(z)}\right),$$

$$C_2(z) = \frac{h}{m}C_1(z),$$

$$C_3(z) = \frac{\omega\, h}{k}\left(1 - ka\cos(z) + \frac{k^2 b^2 \sin^2(z)}{1 - ka\cos(z)}\right) + mg\frac{kb\sin(z)}{1 - ka\cos(z)},$$

$$C_4(z) = -m\left(a\sin(z) + \frac{kb^2\sin(z)\cos(z)}{1 - ka\cos(z)}\right).$$

(4.18)

An important conclusion can be done. Dynamics of a particle sliding on the surface of a propagating Rayleigh wave cannot be chaotic. This is due to the fact that the governing equation of motion is a second-order autonomous ordinary differential equation with smooth parameter functions [30].

Equation (4.8) yields the dynamic equilibrium which represents a motion of the particle on a slope of the propagating wave with the velocity of its propagation:

$$\ddot{u} = 0,$$

$$\dot{u} = \frac{\omega}{k},$$

$$u = \frac{\omega}{k} \cdot t - \psi, \tag{4.19}$$

where ψ is a constant. Then, it follows from (4.4) that

$$x + a\sin(\omega t - kx) = \frac{\omega}{k} \cdot t - \psi. \tag{4.20}$$

The term $a\sin(\omega t - kx)$ is bounded, therefore (4.20) will be in force when

$$x = \frac{\omega}{k} \cdot t - \theta, \tag{4.21}$$

where θ is a constant satisfying the equality $-\theta + a\sin(k\theta) = -\psi$. Moreover, conditions of existence of the dynamic equilibrium are similar in terms of x or u:

$$\dot{x} = \frac{\omega/k - \eta_t'}{1 + \eta_x'} = \frac{\omega}{k},$$

$$\ddot{x} = \frac{-\dot{x}^2\eta_{xx}'' - 2\dot{x}\eta_{xt}'' - \eta_{tt}''}{1 + \eta_x'} = 0. \tag{4.22}$$

As mentioned earlier, the explicit governing equation is formulated in terms of x, not the coordinate of the contact point u. First, coordinates of the unstable saddle point are determined. Then coordinates of the same saddle point are calculated in the frame $(\omega t - ku)$; $\dot{u}$ using the relationship in (4.3). Forward and reverse time marching techniques are used to construct basin boundaries of attractors when partial solutions of (4.14) are sought from the infinitesimal surrounding of the saddle point.

The described computational technique is used to construct basin boundaries of the system's attractors (Figure 8). Solutions in terms of u (forward and reverse) are visualized only. It can be noted that two stable attractors can coexist–a stable equilibrium point and a stable limit cycle. Shaded regions in Figure 8 correspond to a basin (attracting set of initial conditions) of stable equilibrium points, while white region corresponds to a basin of the limit cycle. The phase plane in Figure 8 is periodic by 2π and can be visualized in cylindrical coordinates, but the plane representation is clearer.

A special attention should be paid to dashed line intervals on basin boundaries. Equation (4.14) describes a motion of a particle on the surface of a propagating wave. This governing equation of motion holds until the reaction force N in (4.12) is positive. Whenever N gets equal or lower than zero, the particle looses a contact with the surface and starts a free fly in a gravitational field until it bounces on the surface again. Therefore, the moment when the particle looses the contact with, the surface is detected, and the trajectory is marked by
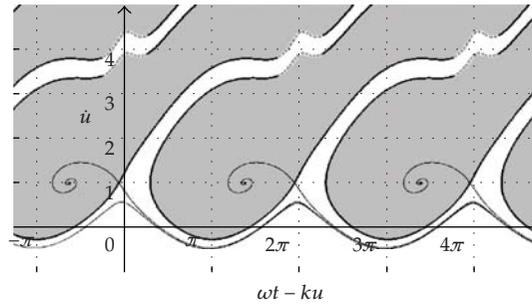
Figure 8: Basin boundaries at $b = 0.5$; $a = b = 1.5$; $m = 0.4$; $h = 0.1$; $\omega = k = 1$, shaded regions illustrate the basin of attraction of stable equilibrium points.
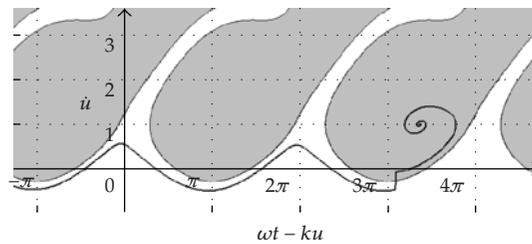

Figure 9: Illustration of the attractor control strategy: limit cycle is represented as a periodic trajectory in frame $(\omega t - ku; \dot{u})$; small external impulse kicks the trajectory to basin boundary of stable equilibrium point where the particle eventually settles down.

a dashed line. It can be noted that such motions occur only at relatively high particle velocities (Figure 8).

Conveyance of a particle by a propagating Rayleigh wave is a nonlinear problem, so such effects as the coexistence of stable attractors should not be astonishing. Stable equilibrium point type attractor in Figure 8 corresponds to a surf-type motion on a slope of a propagating wave; stable limit cycle corresponds to a motion with an average velocity much lower than the velocity of the propagating wave. Coexistence of attractors (a stable equilibrium point and a stable limit cycle) enables development of motion control strategies based on a small external impulses which can bring the system from the regime of motion with small average velocity into motion with the propagating wave's velocity [21]. Such attractor control strategy is illustrated in Figure 9 where the particle first oscillates in the limit cycle, and then a small external impulse kicks it to the basin of attraction of the stable focus point.

It can be noted that the up-mentioned control strategy can be implemented only when the stable equilibrium point and the stable limit cycle coexist. Thus, it would be impossible to transport a sand particle with the velocity of the propagating wave by an acoustic surface Rayleigh wave. Nevertheless, such attractor control strategies could be implemented for transportation of biomedical objects on the surface of an undulation film [21]. The sliding particle model presented in this section also exhibits a very rich dynamics as in the case of the bouncer model. In particular, the sliding particle model should also have the sensitivity to the initial conditions for certain sets of parameter values as occuring in the bouncer model

(see [22]). The sensitivity of transient processes to initial conditions takes place for both, the bouncer model and the sliding particle model.

5. Conclusions and Discussion

Transport of particles by surface waves is an important scientific and engineering problem, with numerous practical applications, including MEMS (micro-electro-mechanical systems) used to manipulate objects like particles or cells. We show that this problem is a modification of the classical bouncer model which is considered as a paradigm model in nonlinear physics. The formulations of our model are implicit, thus phase and velocity maps cannot be expressed in explicit form.

Chaotic dynamics of a conveyed particle is not an unexpected fact due to the complexity of the constitutive model. More surprising is the rich dynamical behavior in models comprising dissipative dynamics, elastic and inelastic collisions. It appears that the transition to chaos via a period doubling route is a universal property for bouncers and is observed in our model of particles transport in both, conservative and viscous media. Moreover, the sensitivity to initial conditions can be useful for control techniques which can dramatically increase the effectiveness of particles transport by surface waves. These results are relevant in the sense that we have also found the sensitivity to the initial conditions for the sliding particle model, which may have important applications in practical implementations as powder transport by piezoelectrically excited ultrasonic waves, transport of sand particles, among others.

Though the numerical analysis was concentrated on the dimensionless system only, theoretical and experimental investigation of dry particle conveyance and its control is a definite object for future research.

Acknowledgments

References

[1] G. M. Zaslavsky, "The simplest case of a strange attractor," *Physics Letters A*, vol. 69, no. 3, pp. 145–147, 1978.
[2] L. D. Pustilnikov, "On Ulam's problem," *Theoretical and Mathematical Physics*, vol. 57, p. 1035, 1983.
[3] M. A. Lieberman and A. J. Lichtenberg, "Stochastic and adiabatic behavior of particles accelerated by periodic forces," *Physical Review A*, vol. 5, no. 4, pp. 1852–1866, 1972.
[4] E. Fermi, "On the origin of the cosmic radiation," *Physical Review*, vol. 75, no. 8, pp. 1169–1174, 1949.
[5] B. V. Chirikov, "A universal instability of many-dimensional oscillator systems," *Physics Reports*, vol. 52, no. 5, pp. 264–379, 1979.
[6] A. J. Lichtenberg, M. A. Lieberman, and R. H. Cohen, "Fermi acceleration revisited," *Physica D*, vol. 1, no. 3, pp. 291–305, 1980.
[7] P. Pieranski and J. Malecki, "Noisy precursors and resonant properties of the period-doubling modes in a nonlinear dynamical system," *Physical Review A*, vol. 34, no. 1, pp. 582–590, 1986.
[8] R. M. Everson, "Chaotic dynamics of a bouncing ball," *Physica D*, vol. 19, no. 3, pp. 355–383, 1986.
[9] K. Wiesenfeld and N. B. Tufillaro, "Suppression of period doubling in the dynamics of a bouncing ball," *Physica D*, vol. 26, no. 1–3, pp. 321–335, 1987.

[10] Z. J. Kowalik, M. Franaszek, and P. Pierański, "Self-reanimating chaos in the bouncing-ball system," *Physical Review A*, vol. 37, no. 10, pp. 4016–4022, 1988.

[11] M.-O. Hongler, P. Cartier, and P. Flury, "Numerical study of a model of vibro-transporter," *Physics Letters A*, vol. 135, no. 2, pp. 106–112, 1989.

[12] M.-O. Hongler and J. Figour, "Periodic versus chaotic dynamics in vibratory feeders," *Helvetica Physica Acta*, vol. 62, no. 1, pp. 68–81, 1989.

[13] M. Mracek and J. Wallaschek, "A system for powder transport based on piezoelectrically excited ultrasonic progressive waves," *Materials Chemistry and Physics*, vol. 90, no. 2-3, pp. 378–380, 2005.

[14] M. S. Talary, J. P. H. Burt, J. A. Tame, and R. Pethig, "Electromanipulation and separation of cells using travelling electric fields," *Journal of Physics D*, vol. 29, no. 8, pp. 2198–2203, 1996.

[15] C.-F. Chou, J. O. Tegenfeldt, O. Bakajin, et al., "Electrodeless dielectrophoresis of single- and double-stranded DNA," *Biophysical Journal*, vol. 83, no. 4, pp. 2170–2179, 2002.

[16] R. H. Plaut, A. L. Farmer, and M. M. Holland, "Bouncing-ball model of 'dry' motions of a tethered buoy," *Journal of Vibration and Acoustics*, vol. 123, no. 3, pp. 333–339, 2001.

[17] J. D. Achenbach, *Wave Propagation in Elastic Solids*, Elsevier, New York, NY, USA, 1984.

[18] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, Pergamon Press, Oxford, UK, 1986.

[19] K. Aki and P. G. Richards, *Quantitative Seismology*, Freeman, New York, NY, USA, 1980.

[20] C. M. Flannery and H. Von Kiedrowski, "Dispersion of surface acoustic waves on rough anisotropic materials," in *Proceedings of the IEEE Ultrasonics Symposium*, vol. 1, pp. 583–586, 2001.

[21] M. Ragulskis and K. Koizumi, "Applicability of attractor control techniques for a particle conveyed by a propagating wave," *Journal of Vibration and Control*, vol. 10, no. 7, pp. 1057–1070, 2004.

[22] M. Ragulskis and M. A. F. Sanjuán, "Transport of particles by surface waves: a modification of the classical bouncer model," *New Journal of Physics*, vol. 10, Article ID 083017, 2008.

[23] J. M. Luck and A. Mehta, "Bouncing ball with a finite restitution: chattering, locking, and chaos," *Physical Review E*, vol. 48, no. 5, pp. 3988–3997, 1993.

[24] R. M. May, "Simple mathematical models with very complicated dynamics," *Nature*, vol. 261, no. 5560, pp. 459–467, 1976.

[25] C.-F. Chou, J. O. Tegenfeldt, O. Bakajin, et al., "Electrodeless dielectrophoresis of single- and double-stranded DNA," *Biophysical Journal*, vol. 83, no. 4, pp. 2170–2179, 2002.

[26] L. Cui and H. Morgan, "Design and fabrication of travelling wave dielectrophoresis structures," *Journal of Micromechanics and Microengineering*, vol. 10, no. 1, pp. 72–79, 2000.

[27] A. D. Dussaud, B. Khusid, and A. Acrivos, "Particle segregation in suspensions subject to high-gradient ac electric fields," *Journal of Applied Physics*, vol. 88, no. 9, pp. 5463–5473, 2000.

[28] W. N. Hassan and J. S. Ribberink, "Transport processes of uniform and mixed sands in oscillatory sheet flow," *Coastal Engineering*, vol. 52, no. 9, pp. 745–770, 2005.

[29] S. D. Wang, Y. M. Shen, and Y. H. Zheng, "Two-dimensional numerical simulation for transport and fate of oil spills in seas," *Ocean Engineering*, vol. 32, no. 13, pp. 1556–1571, 2005.

[30] R. C. Hilborn, *Chaos and Nonlinear Dynamics*, Oxford University Press, New York, NY, USA, 1994.

Research Article

Nonlinear Dynamics and Chaos in a Fractional-Order HIV Model

Haiping Ye[1, 2] and Yongsheng Ding[2, 3]

[1] *Department of Applied Mathematics, Donghua University, Shanghai 201620, China*
[2] *College of Information Sciences and Technology, Donghua University, Shanghai 201620, China*
[3] *Engineering Research Center of Digitized Textile and Fashion Technology, Ministry of Education,*
   Donghua University, Shanghai 201620, China

Correspondence should be addressed to Yongsheng Ding, ysding@dhu.edu.cn

We introduce fractional order into an HIV model. We consider the effect of viral diversity on the human immune system with frequency dependent rate of proliferation of cytotoxic T-lymphocytes (CTLs) and rate of elimination of infected cells by CTLs, based on a fractional-order differential equation model. For the one-virus model, our analysis shows that the interior equilibrium which is unstable in the classical integer-order model can become asymptotically stable in our fractional-order model and numerical simulations confirm this. We also present simulation results of the chaotic behaviors produced from the fractional-order HIV model with viral diversity by using an Adams-type predictor-corrector method.

1. Introduction

An important part of the human immune response against viral infections is cytotoxic T lymphocytes (CTLs) [1]. They recognize and kill cells which are infected by virus. There are many immune models describing the virus dynamics with CTL immune response. Nowak and Bangham [2, 3] proposed an ODE model which explores the relation among CTL immune responses, virus load, and virus diversity. In [2], a rate of specific CTL (Z_j) proliferation in response to the corresponding specific infected cells (I_j) depends on the mass action law cI_jZ_j. This model has been important in the field of mathematical modelling of HIV infection. In their model, there is no interaction among different types of CTL (Z_j). Iwami et al. [4] assumed that the correlation is incorporated as a function of the frequency that the specific CTLs (Z_j) encounter in the specific infected cells (I_j). In a similar manner, they considered the rate of elimination of specific infected cells (I_j) by the specific CTLs (Z_j) to be proportional to this frequency. However, these models do not take into account

the fractional order derivatives that have been extensively applied in many fields (e.g., [5–17] and the reference cited therein). Recently many mathematicians and applied researchers have tried to model real processes using the fractional order differential equations (FODE) [16]. In biology, it has been deduced that the membranes of cells of biological organism have fractional order electrical conductance [13] and then, they are classified into group of noninteger order models. Also, it has been shown that modelling the behavior of brainstem vestibule-oculomotor neurons by FODE has more advantages than classical integer order modelling [8].

Particular emphasis is that a major difference between fractional order models and integer order models is that fractional order models possess memory [5, 12], while the main features of immune response involve memory [18]. Hence, we attempt to model HIV infection with immune response using a fractional order system. Our presentation is based on the immune model of HIV infection which is developed by Iwami et al. [4]. For the one-virus model, we carry out a detailed analysis on stability of equilibrium. Our analysis shows that the interior equilibrium which is unstable in the classical integer order model can become asymptotically stable in our fractional order model. We also find that chaos does exist in the fractional order HIV model with viral diversity.

2. Model Derivation

We first give the definition of fractional order integration and fractional order differentiation [14, 16]. For the concept of fractional derivative we will adopt Caputo's definition which is a modification of the Riemann-Liouville definition and has the advantage of dealing properly with initial value problems.

Definition 2.1. The fractional integral of order $\alpha > 0$ of a function $f : R^+ \to R$ is given by

$$I^\alpha f(x) = \frac{1}{\Gamma(\alpha)} \int_0^x (x - t)^{\alpha - 1} f(t) dt \tag{2.1}$$

provided the right side is pointwise defined on R^+.

Definition 2.2. The Caputo fractional derivative of order $\alpha \in (n-1, n)$ of a continuous function $f : R^+ \to R$ is given by

$$D^\alpha f(x) = I^{n-\alpha} D^n f(x), \quad D = \frac{d}{dt}. \tag{2.2}$$

Now we introduce fractional order into the ODE model by Iwami et al. [4]. The new system is described by the following set of FODE:

$$D^{q_1} T = \lambda - dT - \sum_{l=1}^n \beta_l' T V_l,$$

$$D^{q_{2j}} I_j = \beta_j' T V_j - a I_j - b Z_j \frac{I_j}{T + \sum_{l=1}^n I_l},$$

$$D^{q_{3j}} Z_j = cZ_j \frac{I_j}{T + \sum_{l=1}^{n} I_l} - \delta Z_j,$$

$$D^{q_{4j}} V_j = kaI_j - uV_j \quad (j = 1, 2, \ldots, n),$$

$$(2.3)$$

where $T(t)$ represents the concentration of uninfected cells at time t, $I_j(t)$ represents the concentration of infected cells with a virus particle of type j, $V_j(t)$ the concentration of free virus particle of type j, and $Z_j(t)$ denotes the magnitude of the specific CTL response against variant j. Here, $0.95 \leq q_1, q_{2j}, q_{3j}, q_{4j} \leq 1 \, (j = 1, 2, \ldots, n)$ are restricted such that fractional derivative can be approximately described the rate of change in number.

Following [4], uninfected cells are assumed to be generated at a constant rate λ. Uninfected cells, infected cells, free viruses, and CTLs decline at rates d, a, u, and δ, respectively. The total number of virus particles produced from one cell is k. The rate of CTL proliferation in response to antigen is given by $cZ_j I_j / (T + \sum_{l=1}^{n} I_l)$ and the specific infected cells are killed by specific CTLs at rate $bZ_j I_j / (T + \sum_{l=1}^{n} I_l)$, while infected cells are produced from uninfected cells and free virus at rate $\beta_l' TV_l$. That is, a rate of specific CTL(Z_j) proliferation in response to the corresponding specific infected cells (I_j) depends on the frequency, instead of the mass action law.

To simplify the model, it is reasonable to assume that the decay rate of free virus, u, is much larger than that of the infected cells, a, and this system describes the qualitative dynamics of the asymptomatic phase of HIV infection. Thus, we may introduce as a good approximation that the virus is in steady state (i.e., $D^{q_{4j}} V_j = 0$) and hence $V_j = kaI_j / u$ (see [4, 19]). This leads to the following simplified system of FODE:

$$D^{q_1} T = \lambda - dT - \sum_{l=1}^{n} \beta_l TI_l,$$

$$D^{q_{2j}} I_j = \beta_j TI_j - aI_j - bZ_j \frac{I_j}{T + \sum_{l=1}^{n} I_l},$$

$$(2.4)$$

$$D^{q_{3j}} Z_j = cZ_j \frac{I_j}{T + \sum_{l=1}^{n} I_l} - \delta Z_j \quad (j = 1, 2, \ldots, n),$$

where $\beta_j = ka\beta_j'/u$.

3. One-Virus Model

In this section, we discuss in detail an important special case of model (2.4) and perform an equilibrium and stability analysis for this special case. We consider the one-virus model ($n = 1$) and assume that $q_1 = q_{21} = q_{31} = \alpha \, (0.95 \leq \alpha \leq 1)$. This one-virus model is described by the following system of FODE:

$$D^\alpha T = \lambda - dT - \beta_1 TI_1,$$

$$D^\alpha I_1 = \beta_1 TI_1 - aI_1 - \frac{bZ_1 I_1}{T + I_1},$$

$$(3.1)$$

$$D^\alpha Z_1 = \frac{cZ_1 I_1}{T + I_1} - \delta Z_1.$$

To evaluate the equilibria, let

$$D^{\alpha}T = 0,$$
$$D^{\alpha}I_1 = 0, \tag{3.2}$$
$$D^{\alpha}Z_1 = 0.$$

Then system (3.1) has three equilibria: the uninfected equilibrium $E_H = (\lambda/d, 0, 0)$, the boundary equilibrium $E_I = (T^*, I_1^*, 0)$, where

$$T^* = \frac{a}{\beta_1}, \qquad I_1^* = \frac{\lambda}{a} - \frac{d}{\beta_1}, \tag{3.3}$$

and the interior equilibrium $E_c = (\widehat{T}, \widehat{I}_1, \widehat{Z}_1)$, where

$$\widehat{T} = \frac{-d + \sqrt{d^2 + 4\lambda\widehat{\beta}}}{2\widehat{\beta}}, \qquad \widehat{I}_1 = \frac{\delta}{c - \delta}\widehat{T},$$

$$\widehat{Z}_1 = \frac{c\widehat{T}}{b(c - \delta)}\left(\beta_1\widehat{T} - a\right), \qquad \widehat{\beta} = \frac{\delta\beta_1}{c - \delta}. \tag{3.4}$$

Following the analysis in [4], we introduce a basic reproduction number which is defined by

$$R_0 = \frac{\lambda\beta_1}{ad}. \tag{3.5}$$

Denote $R_+^3 = \{(T, I_1, Z_1) \in R^3 \mid T \geq 0, I_1 \geq 0, Z_1 \geq 0\}$ and we always assume that $c > \delta$. Note that $D^{\alpha}Z_1 < 0$ always holds true if $c \leq \delta$. By generalized mean value theorem [15], we get $Z_1(t)$ is decreasing if $c \leq \delta$.

Next we will discuss the existence and stability of the equilibria of the model (3.1).

Theorem 3.1. (a) *The uninfected equilibrium E_H is locally asymptotically stable (LAS) if $R_0 \in (0, 1)$ and unstable if $R_0 > 1$.*

(b) *If $R_0 > 1$, then the boundary equilibrium E_I exists. This equilibrium is LAS if $1 < R_0 < (a\delta/d(c - \delta)) + 1$ and unstable if $R_0 > (a\delta/d(c - \delta)) + 1$.*

(c) *If $R_0 > (a\delta/d(c - \delta)) + 1$, then E_c exists in Int R_+^3, where Int R_+^3 is the interior of R_+^3.*

Proof. (a) The Jacobian matrix $J(E_H)$ for system (3.1) evaluated at E_H is given by

$$J(E_H) = \begin{pmatrix} -d & -\dfrac{\beta_1\lambda}{d} & 0 \\ 0 & \dfrac{\beta_1\lambda}{d} - a & 0 \\ 0 & 0 & -\delta \end{pmatrix}. \tag{3.6}$$

E_H is locally asymptotically stable if all of the eigenvalues p of the Jacobian matrix $J(E_H)$ satisfy the following condition [6, 17]:

$$\left|\arg(p)\right| > \frac{\alpha\pi}{2}. \tag{3.7}$$

The eigenvalues of $J(E_H)$ are $-d$, $\beta_1\lambda/d - a$, $-\delta$. It is clear that E_H is LAS if $R_0 < 1$ and is unstable if $R_0 > 1$.

(b) If $R_0 > 1$, then the existence of E_I is obvious.

The Jacobian matrix $J(E_I)$ for system (3.1) evaluated at E_I is given by

$$J(E_I) = \begin{pmatrix} -d - \beta_1 I_1^* & -\beta_1 T^* & 0 \\ \beta_1 I_1^* & 0 & -\dfrac{bI_1^*}{T^* + I_1^*} \\ 0 & 0 & \dfrac{cI_1^*}{T^* + I_1^*} - \delta \end{pmatrix}. \tag{3.8}$$

For $J(E_I)$ given by (3.8), the characteristic equation becomes

$$\left[p^2 + (d + \beta_1 I_1^*)p + \beta_1^2 I_1^* T^*\right]\left(p - \frac{cI_1^*}{T^* + I_1^*} + \delta\right) = 0. \tag{3.9}$$

and hence all the eigenvalues are

$$p_{1,2} = \frac{-(d + \beta_1 I_1^*) \pm \sqrt{(d + \beta_1 I_1^*)^2 - 4\beta_1^2 I_1^* T^*}}{2},$$

$$p_3 = \frac{cI_1^*}{T^* + I_1^*} - d. \tag{3.10}$$

If $R_0 > 1$, then $T^* > 0$, $I_1^* > 0$, and $p_{1,2}$ have negative real parts. Furthermore, if $1 < R_0 < (a\delta/d(c - \delta)) + 1$, then $p_3 < 0$ and E_I is LAS. If $R_0 > (a\delta/d(c - \delta)) + 1$, then $p_3 > 0$ and E_I is unstable.

(c) If $R_0 > (a\delta/d(c - \delta)) + 1$, then we obtain $\widehat{Z}_1 > 0$. Thus, E_c exists in Int R_+^3. Therefore, the proof is complete. □

To discuss the local stability of the interior equilibrium E_c, we consider the linearized system of (3.1) at E_c. The Jacobian matrix at E_c is given by

$$J(E_c) = \begin{pmatrix} -d - \beta_1 \widehat{I}_1 & -\beta_1 \widehat{T} & 0 \\ \beta_1 \widehat{I}_1 + \dfrac{b\widehat{Z}_1 \widehat{I}_1}{(\widehat{T} + \widehat{I}_1)^2} & \beta_1 \widehat{T} - a - \dfrac{b\widehat{Z}_1 \widehat{T}}{(\widehat{T} + \widehat{I}_1)^2} & -\dfrac{b\widehat{I}_1}{\widehat{T} + \widehat{I}_1} \\ -\dfrac{c\widehat{Z}_1 \widehat{I}_1}{(\widehat{T} + \widehat{I}_1)^2} & \dfrac{c\widehat{Z}_1 \widehat{T}}{(\widehat{T} + \widehat{I}_1)^2} & \dfrac{c\widehat{I}_1}{\widehat{T} + \widehat{I}_1} - \delta \end{pmatrix}. \tag{3.11}$$

For convenience, we denote $\hat{T} = T$, $\hat{I}_1 = I$, $\hat{Z}_1 = Z$, and $\beta_1 = \beta$. In view of the above assumptions and using

$$\beta T - a - \frac{bZ}{T+I} = 0, \qquad \lambda - dT - \beta TI = 0, \qquad \frac{cI}{T+I} = \delta, \qquad \frac{T}{T+I} = \frac{c-\delta}{c}, \qquad (3.12)$$

$J(E_c)$ can now be written as follows:

$$J(E_c) = \begin{pmatrix} -d - \beta I & -\beta T & 0 \\ \beta I + \dfrac{\delta}{c}(\beta T - a) & \dfrac{\delta}{c}(\beta T - a) & -\dfrac{b\delta}{c} \\ -\dfrac{\delta}{b}(\beta T - a) & \dfrac{c-\delta}{b}(\beta T - a) & 0 \end{pmatrix}. \qquad (3.13)$$

Then the characteristic equation of the linearized system of (3.1) is

$$\Phi(p) = p^3 + a_1 p^2 + a_2 p + a_3 = 0, \qquad (3.14)$$

where

$$a_1 = d + \frac{\delta a}{c} + \frac{\delta^2 \beta T}{c(c-\delta)},$$

$$a_2 = \beta T \left(\beta I + \frac{\delta}{c}(\beta T - a) \right) + \frac{\delta}{c}(c-\delta)(\beta T - a) - \frac{\delta}{c}(d + \beta I)(\beta T - a), \qquad (3.15)$$

$$a_3 = \frac{\delta}{c}(c-\delta)(\beta T - a)(d + \beta I) + \frac{\beta \delta^2}{c} T(\beta T - a).$$

Proposition 3.2. *The interior equilibrium E_c is LAS if all of the eigenvalues p of $J(E_c)$ satisfy $|\arg(p)| > \alpha \pi/2$.*
Denote

$$D(\Phi) = - \begin{vmatrix} 1 & a_1 & a_2 & a_3 & 0 \\ 0 & 1 & a_1 & a_2 & a_3 \\ 3 & 2a_1 & a_2 & 0 & 0 \\ 0 & 3 & 2a_1 & a_2 & 0 \\ 0 & 0 & 3 & 2a_1 & a_2 \end{vmatrix} \qquad (3.16)$$

$$= 18a_1 a_2 a_3 + (a_1 a_2)^2 - 4a_3 a_1^3 - 4a_2^3 - 27a_3^2.$$
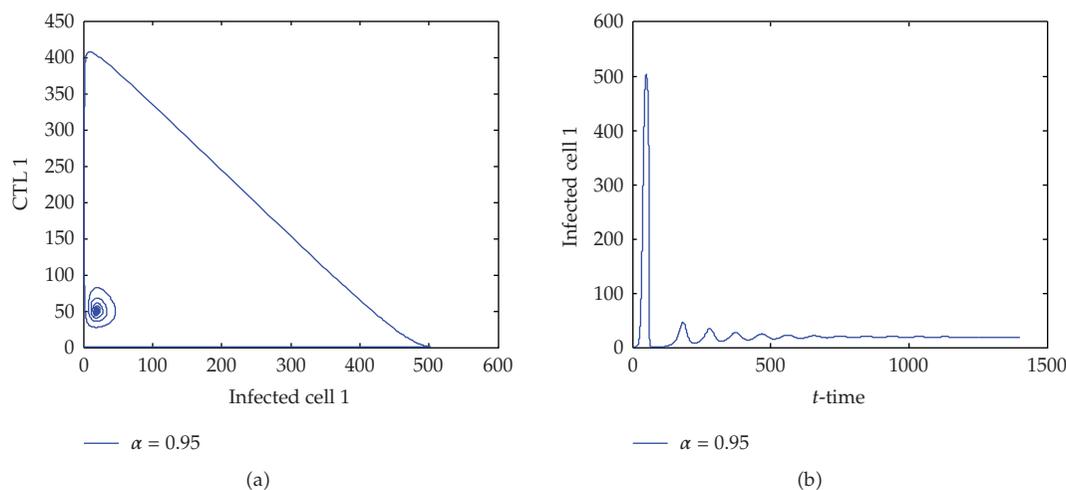
Figure 1: Numerical solutions of system (3.1). The plots show that trajectories of system (3.1) approach to the interior equilibrium for $\alpha = 0.95$.

Using the results of [5, 20], we have the following proposition.

Proposition 3.3. *One assumes that E_c exists in* Int R_+^3.

(i) *If the discriminant of $\Phi(p)$, $D(\Phi)$ is positive and Routh-Hurwitz conditions are satisfied, that is,*

$$D(\Phi) > 0, \qquad a_1 > 0, \qquad a_3 > 0, \qquad a_1 a_2 > a_3, \tag{3.17}$$

then the interior equilibrium E_c is LAS.

(ii) *If $D(\Phi) < 0$, $a_1 > 0$, $a_2 > 0$, $a_1 a_2 = a_3$, $\alpha \in [0,1)$, then the interior equilibrium E_c is LAS.*

(iii) *If $D(\Phi) < 0$, $a_1 < 0$, $a_2 < 0$, $\alpha > 2/3$, then the interior equilibrium E_c is unstable.*

In our first example we set $\lambda = 10$, $d = 0.02$ which are chosen according to [21] and set $a = \delta = 0.04$, $c = b = 0.8$, $\beta_1 = 4 \times 10^{-4}$ which come from [4]. With these parameter values, $R_0 = 5 > (a\delta/d(c-\delta))+1$, $D(\Phi) = -9.4073 \times 10^{-7} < 0$, $a_1 a_2 < a_3$. By Proposition 3.2., we obtain the interior equilibrium $E_c = (362.0335, 19.0544, 49.9289)$ is LAS when $\alpha < 0.9916$. Numerical simulations show that trajectories of system (3.1) approach to the interior equilibrium (see Figures 1(a) and 1(b)). However, when $\alpha = 1$ (that is the case of classical integer order), E_c is unstable by the Routh-Hurwitz criterion(see Figures 2(a) and 2(b)).

4. Two-Virus Model

In this section, we consider viral diversity. We examine the two-virus model using numerical simulations. By examining the behavior of this simpler model we hope to get an idea as to

(a)   (b)
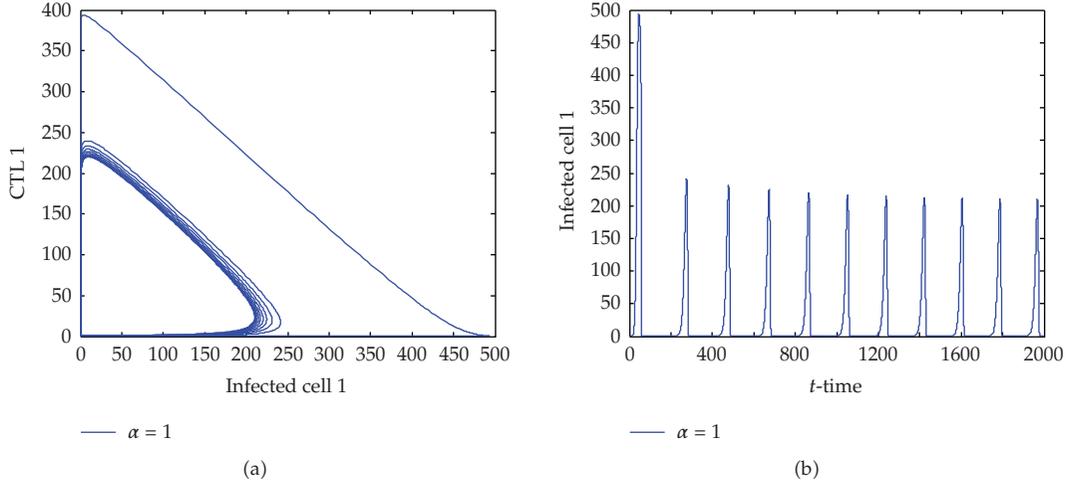
Figure 2: Numerical solutions of system (3.1). The plots show that the interior equilibrium is unstable for $\alpha = 1$.

how the more general models in system (2.4) may behave. The two-virus model is given by the following system of FODE:

$$D^{q_1}T = \lambda - dT - \beta_1 T I_1 - \beta_2 T I_2,$$

$$D^{q_{21}}I_1 = \beta_1 T I_1 - aI_1 - bZ_1 \frac{I_1}{T + I_1 + I_2},$$

$$D^{q_{22}}I_2 = \beta_2 T I_2 - aI_2 - bZ_2 \frac{I_2}{T + I_1 + I_2}, \tag{4.1}$$

$$D^{q_{31}}Z_1 = cZ_1 \frac{I_1}{T + I_1 + I_2} - \delta Z_1,$$

$$D^{q_{32}}Z_2 = cZ_2 \frac{I_2}{T + I_1 + I_2} - \delta Z_2,$$

with initial value condition

$$T(0) = T_0, \qquad I_i(0) = I_{i0}, \quad Z_i(0) = Z_{i0}, \quad i = 1, 2, \tag{4.2}$$

where $0.95 \leq q_1, q_{2j}, q_{3j}\ (j = 1, 2) \leq 1$.

To find numerical solution to (4.1) and (4.2) in the interval $[0, T]$, we reduce the systems (4.1) and (4.2) to a set of fractional integral equations, by using an equivalence (see [16, Theorem 3.24])

$$D^\alpha X = f(X) \Longleftrightarrow X(t) = X(0) + I^\alpha f(X). \tag{4.3}$$

Then we apply the generalized Adams-type predictor-corrector method or, more precisely, Predict, Evaluate, Correct, Evaluate (PECE) methods (see [22, 23]).
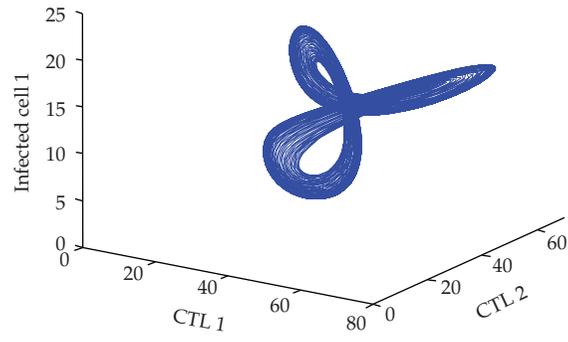
(a)
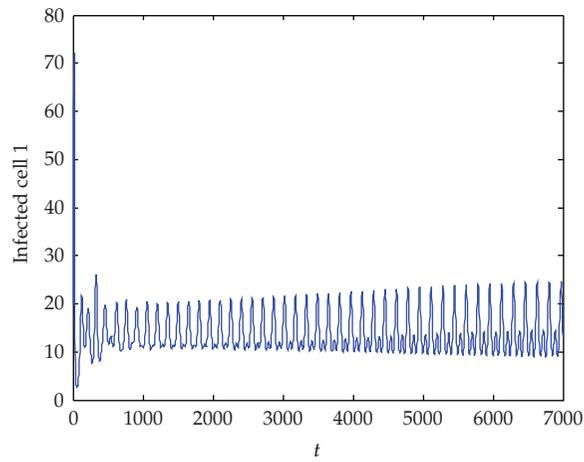
(b)
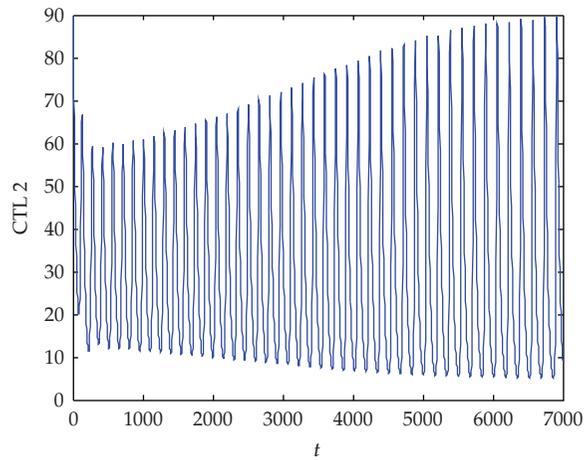
(c)

Figure 3: Numerical solutions of system (4.1) for $\alpha = [0.95, 0.95, 1, 1, 1]$. (a) A strange attractor in the Z_1-Z_2-I_1 phase. (b) Infected cell 1. (c) CTL 2.
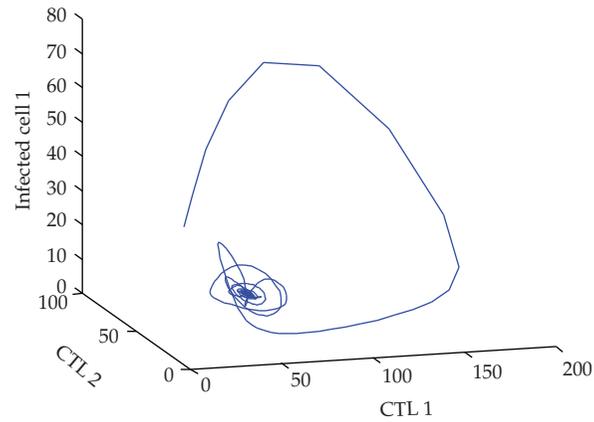
(a)
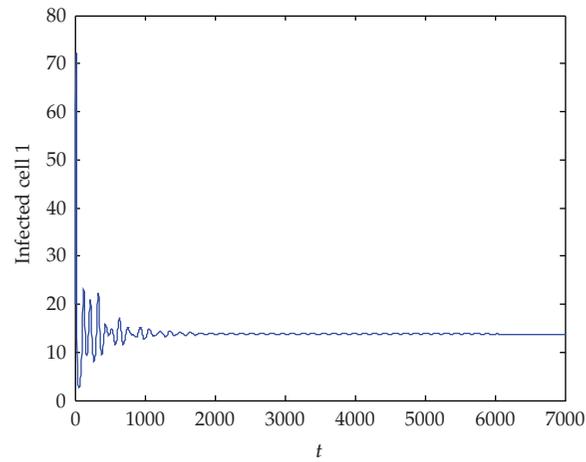
(b)
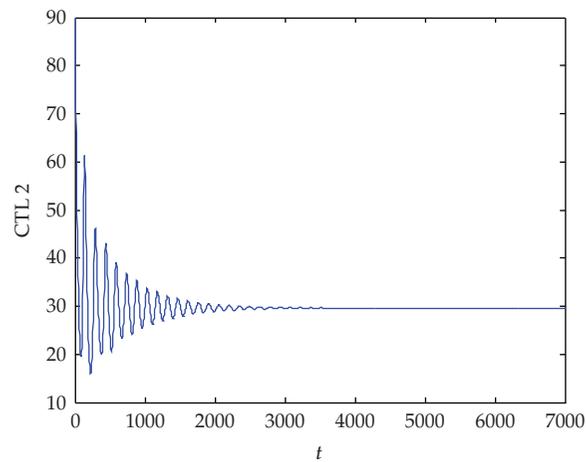
(c)

Figure 4: Numerical solutions of system (4.1) for $\alpha = [0.95, 0.95, 0.95, 1, 1]$. (a) Z_1-Z_2-I_1 phase. (b) Infected cell 1. (c) CTL 2.

For notational convenience, we denote $\alpha = (q_1, q_{21}, q_{22}, q_{31}, q_{32})$. We carry out numerical simulations for system, (4.1) and (4.2) with parameters $\lambda = 10$, $b = c = 0.8$, $a = 0.031$, $\delta = 0.03$, $d = 0.02$, $\beta_1 = 4 \times 10^{-4}$, and $\beta_2 = 2.8 \times 10^{-4}$ for the step size 0.07. Numerical solutions of systems (4.1) and (4.2) support that the system exhibits a chaotic behavior and systems (4.1) and (4.2) have a strange attractor in Int R_+^5 for $\alpha = [0.95, 0.95, 1, 1, 1]$ (see Figures 3(a)–3(c)). It is clear that chaos does exist in our fractional order model with viral diversity as in the case of integer order model. The effect of viral diversity and the frequency dependence results in collapse of the immune system and make the behavior of the system dynamics complex [4]. However, as the value of some component or more components of the order α further decreases, for example, $\alpha = [0.95, 0.95, 0.95, 1, 1]$, the chaotic motion disappears and the systems (4.1) and (4.2) stabilize to a fixed point (see Figures 4(a)–4(c)).

5. Conclusions

In this paper, we have proposed a fractional order HIV model, as a generalization of an integer order model, developed by Iwami et al. [4]. The premise of the proposed model is the fact that fractional order models possess memory while the main features of immune response involve memory. It is an attempt to incorporate fractional order into the mathematical model of HIV-immune system dynamics and it is still an interesting exercise to determine, mathematically, how the order of a fractional differential system affects the dynamics of system.

In the case of one-virus model, the fractional order system has an interior equilibrium under some restriction. By using stability analysis on fractional order system, we obtain sufficient condition on the parameters for the stability of the interior equilibrium. Our analysis shows that the interior equilibrium which is unstable in the classical integer order model can become asymptotically stable in our fractional order model. Note that the interior equilibrium is globally asymptotically stable (GAS) (see [24]) if the terms associated with immune reactions are given by cZ_1I_1 and bZ_1I_1 instead of $cZ_1I_1/(T + I_1)$ and $bZ_1I_1/(T + I_1)$ in (3.1). That is, the interior equilibrium of the one-virus model can become unstable because of the frequency dependence (see [4]). However, in our fractional order model with the frequency dependence, the interior equilibrium can also become asymptotically stable if the order $\alpha < 0.9916$.

We then consider viral diversity. If the terms associated with immune reaction depend on the mass action law instead of frequency, an interior equilibrium in [24] is GAS. Similar to the integer order model in [4], we find that strange chaotic attractors can be obtained under fractional order model with frequency dependence. That is, the effect of viral diversity and the frequency dependence results in collapse of the immune system and make the behavior of the system dynamics complex. However the chaotic motion may disappear and the fractional order system stabilizes to a fixed point if the value of the order α decreases. The specific biological meaning is deserved to further study.

Acknowledgments

References

[1] M. A. Nowak, R. M. May, and K. Sigmund, "Immune responses against multiple epitopes," *Journal of Theoretical Biology*, vol. 175, no. 3, pp. 325–353, 1995.

[2] M. A. Nowak and C. R. M. Bangham, "Population dynamics of immune responses to persistent viruses," *Science*, vol. 272, no. 5258, pp. 74–79, 1996.

[3] M. A. Nowak and R. M. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology*, Oxford University Press, Oxford, UK, 2000.

[4] S. Iwami, S. Nakaoka, and Y. Takeuchi, "Frequency dependence and viral diversity imply chaos in an HIV model," *Physica D*, vol. 223, no. 2, pp. 222–228, 2006.

[5] E. Ahmed and A. S. Elgazzar, "On fractional order differential equations model for nonlocal epidemics," *Physica A*, vol. 379, no. 2, pp. 607–614, 2007.

[6] E. Ahmed, A. M. A. El-Sayed, and H. A. A. El-Saka, "Equilibrium points, stability and numerical solutions of fractional-order predator-prey and rabies models," *Journal of Mathematical Analysis and Applications*, vol. 325, no. 1, pp. 542–553, 2007.

[7] M. P. Lazarević, "Finite time stability analysis of PD^α fractional control of robotic time-delay systems," *Mechanics Research Communications*, vol. 33, no. 2, pp. 269–279, 2006.

[8] T. J. Anastasio, "The fractional-order dynamics of brainstem vestibulo-oculomotor neurons," *Biological Cybernetics*, vol. 72, no. 1, pp. 69–79, 1994.

[9] T. T. Hartley, C. F. Lorenzo, and H. K. Qammer, "Chaos in a fractional order Chua's system," *IEEE Transactions on Circuits and Systems I*, vol. 42, no. 8, pp. 485–490, 1995.

[10] B. Bonilla, M. Rivero, L. Rodríguez-Germá, and J. J. Trujillo, "Fractional differential equations as alternative models to nonlinear differential equations," *Applied Mathematics and Computation*, vol. 187, no. 1, pp. 79–88, 2007.

[11] A. M. A. El-Sayed, A. E. M. El-Mesiry, and H. A. A. El-Saka, "On the fractional-order logistic equation," *Applied Mathematics Letters*, vol. 20, no. 7, pp. 817–823, 2007.

[12] R. Hilfer, Ed., *Applications of Fractional Calculus in Physics*, World Scientific, River Edge, NJ, USA, 2000.

[13] K. S. Cole, "Electric conductance of biological systems," in *Proceedings of the Cold Spring Harbor Symposia on Quantitative Biology*, pp. 107–116, Cold Spring Harbor, NY, USA, January 1993.

[14] I. Podlubny, *Fractional Differential Equations*, vol. 198 of *Mathematics in Science and Engineering*, Academic Press, San Diego, Calif, USA, 1999.

[15] Z. M. Odibat and N. T. Shawagfeh, "Generalized Taylor's formula," *Applied Mathematics and Computation*, vol. 186, no. 1, pp. 286–293, 2007.

[16] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo, *Theory and Applications of Fractional Differential Equations*, vol. 204 of *North-Holland Mathematics Studies*, Elsevier, Amsterdam, The Netherlands, 2006.

[17] D. Matignon, "Stability results for fractional differential equations with applications to control processing," in *Computational Engineering in Systems Applications*, vol. 2, pp. 963–968, IMACS IEEE-SMC, Lille, France, 1996.

[18] J. Velasco-Hernández, J. García, and D. Kirschner, "Remarks on modeling host-viral dynamics and treatment," in *Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction to Models, Methods, and Theory*, vol. 125 of *The IMA Volumes in Mathematics and Its Applications*, pp. 287–308, Springer, New York, NY, USA, 2002.

[19] R. R. Regoes, D. Wodarz, and M. A. Nowak, "Virus dynamics: the effect of target cell limitation and immune responses on virus evolution," *Journal of Theoretical Biology*, vol. 191, no. 4, pp. 451–462, 1998.

[20] E. Ahmed, A. M. A. El-Sayed, and H. A. A. El-Saka, "On some Routh-Hurwitz conditions for fractional order differential equations and their applications in Lorenz, Rössler, Chua and Chen systems," *Physics Letters A*, vol. 358, no. 1, pp. 1–4, 2006.

[21] A. S. Perelson, D. E. Kirschner, and R. De Boer, "Dynamics of HIV infection of CD4$^+$ T cells," *Mathematical Biosciences*, vol. 114, no. 1, pp. 81–125, 1993.

[22] K. Diethelm, N. J. Ford, and A. D. Freed, "A predictor-corrector approach for the numerical solution of fractional differential equations," *Nonlinear Dynamics*, vol. 29, no. 1–4, pp. 3–22, 2002.

[23] K. Diethelm, N. J. Ford, and A. D. Freed, "Detailed error analysis for a fractional Adams method," *Numerical Algorithms*, vol. 36, no. 1, pp. 31–52, 2004.

[24] Y. Iwasa, F. Michor, and M. Nowak, "Some basic properties of immune selection," *Journal of Theoretical Biology*, vol. 229, no. 2, pp. 179–188, 2004.

Research Article

Chaotic Image Encryption Design Using Tompkins-Paige Algorithm

Shahram Etemadi Borujeni[1, 2] and Mohammad Eshghi[1]

[1] *Computer Engineering Department, Faculty of Computer and Electrical Engineering,*
   Shahid Beheshti University, Evin, Tehran 1983963113, Iran
[2] *Computer Engineering Department, Faculty of Engineering, University of Isfahan,*
   Isfahan 8174673441, Iran

Correspondence should be addressed to Shahram Etemadi Borujeni, etemadi@eng.ui.ac.ir

In this paper, we have presented a new permutation-substitution image encryption architecture using chaotic maps and Tompkins-Paige algorithm. The proposed encryption system includes two major parts, chaotic pixels permutation and chaotic pixels substitution. A logistic map is used to generate a bit sequence, which is used to generate pseudorandom numbers in Tompkins-Paige algorithm, in 2D permutation phase. Pixel substitution phase includes two process, the tent pseudorandom image (TPRI) generator and modulo addition operation. All parts of the proposed chaotic encryption system are simulated. Uniformity of the histogram of the proposed encrypted image is justified using the chi-square test, which is less than $\chi^2(255, 0.05)$. The vertical, horizontal, and diagonal correlation coefficients, as well as their average and RMS values for the proposed encrypted image are calculated that is about 13% less than previous researches. To quantify the difference between the encrypted image and the corresponding plain-image, three measures are used. These are MAE, NPCR, and UACI, which are improved in our proposed system considerably. NPCR of our proposed system is exactly the ideal value of this criterion. The key space of our proposed method is large enough to protect the system against any Brute-force and statistical attacks.

1. Introduction

In any communication system, including satellite and internet, it is almost impossible to prevent unauthorized people from eavesdropping. When information is broadcasted from a satellite or transmitted through the internet, there is a risk of information interception. Security of image and video data has become increasingly important for many applications including video conferencing, secure facsimile, medical and military applications. Two main groups of technologies have been developed for this purpose. The first group is content

protection through encryption, for which a key is required for proper decryption of the data. The second group is digital watermarking, which aims to embed a message into the multimedia data. These two technologies could be used complementary to each other [1, 2].

In secured communications using encryption, which is the focus of the present work, the information under consideration is converted from the intelligible form to an unintelligible structure using certain operations at the transmitter. Data encryption is mainly scrambling the content of data, such as text, image, audio, and video to make the data unreadable, invisible or incomprehensible during transmission. The unintelligible or encrypted form of the information is then transmitted through the insecure channel, that is, internet, to the destination. At the intended recipient side, however, the information is again converted back to an understandable form using decryption operation and thus the information is conveyed securely. It should be noted that the same keys guide both these encryption and decryption operations. Such encryption system is grouped under private key cryptography [1, 3].

In particular, an image-scrambling scheme transforms an image into another unintelligible image, based on keys only known to the senders and the receivers. The fundamental techniques to encrypt a block of pixels are substitution and permutation. Substitution replaces a pixel with another one; permutation changes the sequence of the pixels in a block to make them unreadable.

In recent years, chaotic maps have been employed for image encryption. Most chaotic image encryptions (or encryption systems) use the permutation-substitution architecture. These two processes are repeated for several rounds, to obtain the final encrypted image. For example, in [4], Fridrich suggested a chaotic image encryption method composed of permutation and substitution. All the pixels are moved using a 2D chaotic map. The new pixels moved to the current position are taken as a permutation of the original pixels. In the substitution process, the pixel values are altered sequentially. Chen et al. employed a three-dimensional (3D) Arnold cat map [5] and a 3D Baker map [6] in the permutation stage. Guan et al. used a 2D cat map for pixel position permutation and the discretized Chen's chaotic system for pixel value masking [7]. Lian et al. [8] used a chaotic standard map in the permutation stage and a quantized logistic map in the substitution stage. The parameters of these two chaotic maps are determined by a key stream generated in each round. Mao et. al. construct a new image encryption scheme based on the extended chaotic Baker map [6]. Zhang et al. first permute the pixels of images with discrete exponential chaotic map, and then use "XOR plus mod" operation for substitution [9]. Gao et al. present the image encryption algorithm based on a new nonlinear chaotic algorithm using a power function and a tangent function instead of a linear function. It also uses a chaotic sequence generated by a nonlinear chaotic algorithm to encrypt image data using XOR operation [10]. Zhou et al. propose a parallel image encryption algorithm using discretized kolmogorov flow map. All the pixels are first permuted with a discretized chaotic map and then encrypted under the cipher block chain mode [11].

There are however some other chaotic image encryption systems with different structures. For example, Pisarchik and Zanin suggested an algorithm to convert image pixels to chaotic maps coupled to form a chaotic map lattice. The encrypted image is obtained by iterating the chaotic map lattice with secret system parameters and number of cycles [12]. Pareek et al. extended the concept of their text encryption to image encryption by using two logistic maps and a key [13].

In this paper, a new permutation-substitution architecture using chaotic maps and Tompkins-Paige algorithm is proposed. Our designed technique for speech scrambling [14] is extended to two-dimensional (2D) permutation and is applied to image permutation [15]. We have improved our work by using chaotic maps and adding a substitution part to an image encryption system. In the permutation phase, a logistic map is used to generate a bit sequence, which is used to generate Pseudorandom numbers in Tompkins-Paige algorithm. A tent map is also used in the substitution phase to product a Pseudorandom image that is used to mix it with the permuted image. The permutation and substitution operations need two different keys, *Key-P* and *Key-S*, respectively. Satisfactory security performance of the proposed system is achieved in only one round and therefore the total encryption time is short .

The paper is organized as follows. In Section 2, principles of chaotic cryptography including chaotic maps and chaotic encryption are introduced. The proposed chaotic encryption systems using logistic random bit sequence generator, Tompkins-Paige algorithm, and tent Pseudorandom image generator are described in Section 3. In Section 4, simulation results of the proposed image encryption systems are presented. Finally, the security analysis is explained in Section 5.

2. Principle of Chaotic Cryptography

The word cryptography refers to the science of keeping secrecy of information exchanged between a sender and a receiver over an insecure channel. The objective is achieved by data encryption so that only individuals who have the key can decrypt it. The key K, in a typical encryption system, determines the transformation from the set of all possible samples, to the set of all possible permuted samples. An encryption system is a finite set T of transformations from a finite sample space M onto a permuted sample space C. It means that each of the transformations in T must be reversible, so that if a sample m is transformed into the permuted sample c by transformation t, $c = t(m)$, then the sample m is $m = t^{-1}(c)$, where t^{-1} is the inverse transform of t [16, 17].

In practice, we need to transmit a reasonable amount of information, which requires a large sample space and that in turn implies a large number of keys. The distribution of a large number of keys is liable to cause horrendous management problems. In a practical system, a cryptanalyst will have to worry about time and facilities. Often, the time taken to solve a permuted sample will be of utmost importance. It is quite likely that the samples need to be secret for a limited period of time, referred to as required cover time. Thus, it is certainly possible for a theoretically insecure system to provide adequate practical security [1]. If we set the cryptanalyst a task requiring a large amount of storage, or sufficiently large number of operations, then we may regard our system as practically secure.

With the desirable properties of ergodicity and high sensitivity to initial conditions and control parameters, chaotic maps are suitable for various data encryption schemes. In particular, chaotic maps are easy to be implemented using microprocessors or personal computers. Therefore, chaotic encryption systems generally have high speed with low cost, which makes them better candidates than many traditional ciphers for multimedia data encryption. There are two types of chaotic encryption systems: chaotic stream encryption systems, and chaotic block encryption systems. In chaotic stream encryption systems, a key stream is produced by a chaotic map, which is used to encrypt a plain-text bit by bit. A chaotic block encryption system, on the other hand, transforms a plain-text block by block with some chaotic maps [8].

2.1. Chaotic Maps

In this subsection, we consider nonlinear and chaotic one-dimensional maps $f : S \longrightarrow S$, where $S \subset R$. The set S is $S = [0, 1]$. The one-dimensional dynamical system can be defined by a difference equation similar to

$$x_{k+1} = f(x_k), \quad k = 0, 1, 2, \ldots, x_k \in S, \tag{2.1}$$

where the variable k stands for time. A dynamical system consists of a set of possible states, together with a deterministic rule, which means that the present state can be determined uniquely from the past states. The orbit of x under f is the set of points $\{x, f(x), f^2(x), \ldots, f^n(x)\}$, where $f^2(x) = f(f(x))$ and $f^n(x)$ means n times iterating of the function $f(x)$. The starting point x for the orbit is called the initial value of the orbit. A chaotic orbit is one that forever continues to experience the unstable behavior that an orbit exhibits near a source, but that is not itself fixed or periodic [18]. Two well known one-dimensional chaotic maps are tent and logistic maps.

The iterative relation of the tent map is given by [18].

$$x_{k+1} = \begin{cases} \dfrac{x_k}{p}, & \text{if } 0 \leq x_k \leq p, \\[2mm] & \quad \text{where } x_k \in [0, 1], \\[2mm] \dfrac{1 - x_k}{1 - p}, & \text{if } p \leq x_k \leq 1, \end{cases} \tag{2.2}$$

where x_0 is the initial condition and p is the control parameter. The tent map is chaotic if p is in the range of $(0, 1)$ and $p \neq 0.5$. Figure 1(a) shows a sample return map of a tent map.

Logistic map is a one-dimensional quadratic map defined by

$$x_{k+1} = ax_k(1 - x_k), \quad x_k \in [0, 1], \tag{2.3}$$

where a is the control parameter, and x_0 is the initial condition. The control parameter a should be taken in the range of $[3.6, 4]$ to keep the logistic map chaotic. Figure 1(b) shows a sample return map of a logistic map.

The logistic equation involves two multiplications and one subtraction per iteration, while the tent equation includes one division and on average one subtraction. Meanwhile, the tent map has better chaotic behavior than the logistic map. As mentioned above, the range of control parameter (p) of the tent map is about twice the range of the logistic map. However, hardware implementation of the logistic map is simpler. We have used both these in our proposed system to improve security.

2.2. Chaotic Encryption Scheme

Due to the tight relationship between chaos and cryptography, the use of chaotic maps to construct an encryption system has been widely investigated [19]. There are three typical ways of using chaos in an image encryption.
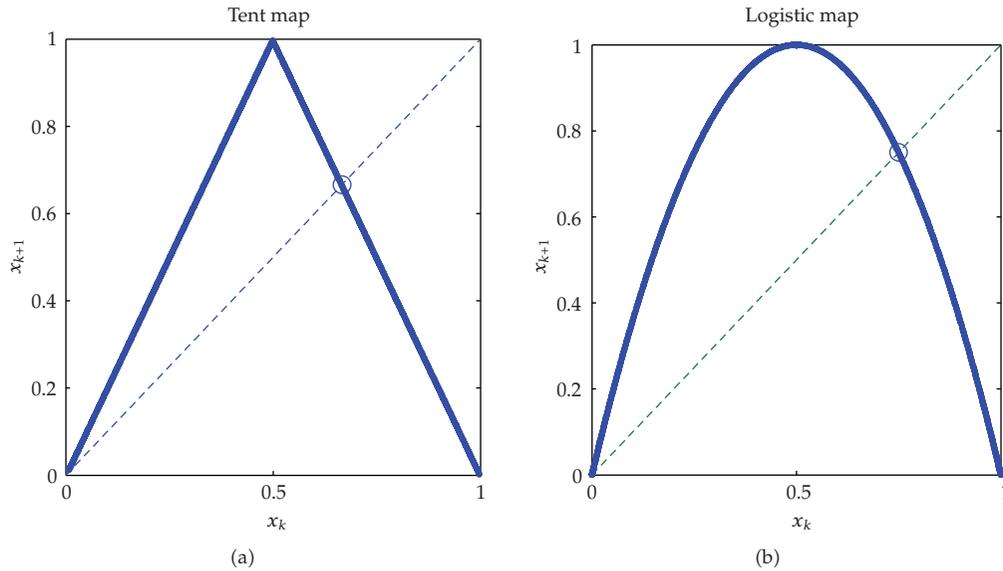
Figure 1: (a) Tent return map ($p = 0.5$). (b) Logistic return map ($a = 4$).

(1) Using chaos as a source to generate Pseudorandom bits with desired statistical properties to realize a secret permutation operation [6, 7, 20].

(2) Using chaos as a source to generate Pseudorandom pixels with desired statistical properties to realize a secret substitution operation [5, 21–23].

(3) Using two chaotic maps in both permutation and substitution [8, 11, 12].

The fundamental techniques to encrypt a block of symbols are confusion and diffusion. Confusion can make ambiguous the relationship between the plain-text and the cipher-text. Diffusion can spread the change throughout the whole cipher-text. Substitution, which replaces a symbol with another one, is the simplest type of confusion, and permutation that changes the sequence of the symbols in the block is the simplest method of diffusion. These techniques together are still the foundations of encryption [3].

2.2.1. Chaotic Permutation

In designing private key cryptographic techniques, permutation methods are considered as important building blocks in conjunction with Pseudorandom sequence generators for selecting a specific permutation key. First, a *Key-P* is entered as a binary number equivalent to the given key. Then, a 1Dimentional chaotic map generates a random bit-string. Subsequently, a permutation matrix for the system is calculated.

A permutation matrix is an identity matrix with the rows and columns interchanged. It has a single 1 in each row and column; all the other elements are 0. For example,

$$P = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{2.4}$$

Any vector A is multiplied by the permutation matrix in order to rearrange its elements. For example from 1-2-3-4 to 4-1-3-2 as shown in

$$P \cdot A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} a_4 \\ a_1 \\ a_3 \\ a_2 \end{bmatrix}. \tag{2.5}$$

For simplicity, the 4×4 matrix p could be expressed as a 4-element vector Q as shown in

$$Q = \begin{bmatrix} 4 & 1 & 3 & 2 \end{bmatrix}. \tag{2.6}$$

Therefore, the $L \times L$ permutation matrix of the L elements could be expressed as an L-element vector Q for simplicity (2.7), where $1 \leq k_i \leq L$ [16].

$$Q = \begin{bmatrix} k_1 & k_2 & k_3 & k_4 & \cdots & k_L \end{bmatrix}. \tag{2.7}$$

Each element in the vector Q shows the new position of each element in the permuted vector. The elements of the vector are rearranged in a new order according to K_i's in the simplified permutation matrix, Q. Practically, the element at the k_ith position is moved to the ith position, respectively. By permutation P on a set of L elements, A,

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & \cdots & a_L \end{bmatrix}^T,$$

$$P \cdot A = P \cdot \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & \cdots & a_L \end{bmatrix}^T \tag{2.8}$$

$$= \begin{bmatrix} d_1 & d_2 & d_3 & d_4 & \cdots & d_L \end{bmatrix}^T,$$

where $d_i = a_{k_i}$, for $i = 1, 2 \ldots, L$ (as an example: if $k_1 = 4$ then $d_1 = a_4$).

2.2.2. Chaotic Substitution

In cryptography, a substitution cipher is a method of encryption by which blocks of plain text are replaced with cipher-text according to a regular system; the blocks may be single or several letters. The receiver deciphers the text by performing an inverse substitution. Substitution ciphers can be compared with permutation ciphers. In a permutation cipher, the blocks of the plain-text are rearranged in a different and usually quite complex order, but the blocks themselves are left unchanged. By contrast, in a substitution cipher, the blocks of the plain-text are retained in the same sequence as in the cipher-text, but the blocks themselves are altered.

A permutation-only encrypted system is insecure against attacks [24]. To improve the security, substitution process is added to the encryption system. The substitution could be one of simple operations such as XOR, XNOR, shift, Add, and/ or a combination of these simple operations. Chaotic map is used as generation of Pseudorandom image for substitution.
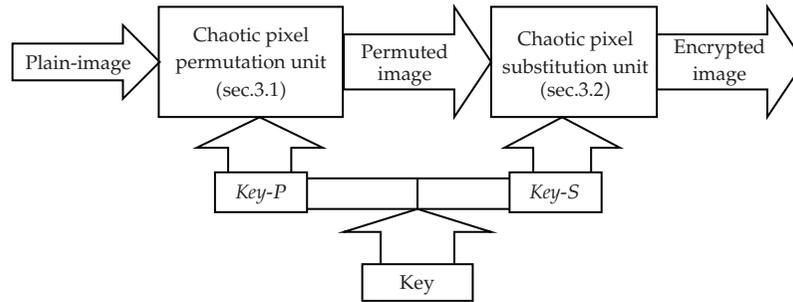
Figure 2: Block diagram of the chaotic image encryption system.

Actually, chaotic image with a size equal to plain-image is generated. All pixels of permuted image and new chaotic image are combined with modular addition. Substitute operation decreases the correlation between blocks or samples in text and makes its histogram uniform.

3. Design of a Chaotic Image Encryption System

The block diagram of the proposed chaotic image encryption system is illustrated in Figure 2. This system includes two major units, chaotic pixels permutation unit and chaotic pixels substitution unit. Two different dynamical systems, that is, logistic and tent maps are also considered to generate a more complicated key and consequently a highly secure encryption system. In this paper, logistic map is used as a Pseudorandom bit generator while tent map is utilized to generate a Pseudorandom image generator. Pixels of a plain-image are rearranged by the permutation unit. The permutation unit uses a chaotic bit generator and Tompkins-Paige algorithm, to implement a $1D$ and $2D$ image permutation. The pixels of the permutated image are then changed in the chaotic pixels substitution unit. The substitution unit is used for the modular addition of the permuted image with a Pseudorandom image. A key that is used for the encryption system includes *Key-P* and *Key-S*. *Key-P* is used as the initial value and control parameter of a logistic map, which is utilized to generate random bit sequences. *Key-S* is applied to the tent map, as an initial value and the control parameter. Tent map is used to generate a Pseudorandom image. More details of each unit are explained in the following sections.

3.1. Chaotic Pixel Permutation Unit

It is assumed that L pixels are expected to be permuted. The number of possible permutations for L pixels is $L!$, however not all permutations can be used. The Hamming distance is the number of elements moved by the permutation. The more number of elements moved by the permutation, the larger the Hamming distance. Meanwhile, permutation matrixes that are close to any circularly shifted versions of the identity order produce a permuted sample of high closeness to the original sample. Since typically, we might wish to have a choice of about N permutation matrixes, the number of key bit, M, should be selected such that $N < 2^M - 1$. In the encryption system, the number of pixels to be permuted L is assumed to be 128, since the image size 128. Therefore, the number of all possible permutations is $128!(3.85E + 215)$ [25, 26].
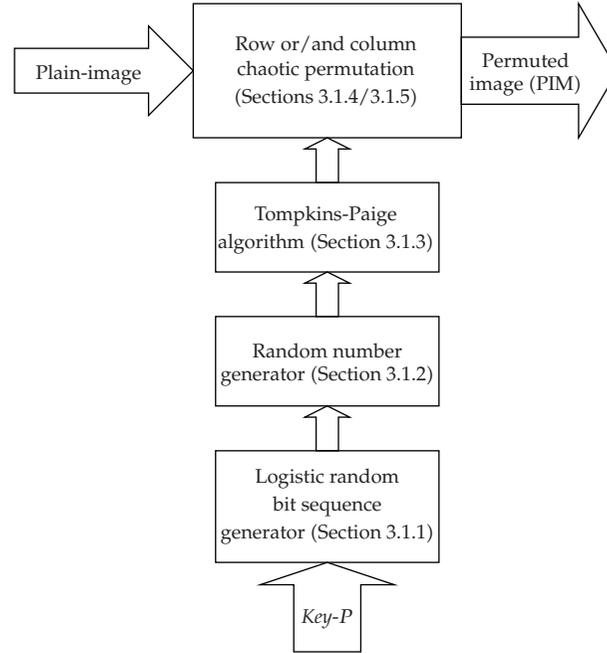
Figure 3: Block diagram of the chaotic pixel permutation unit.

There are three steps in the design of the permutation subsystem, which are explained as follows. First, a 90 *bit* key is entered as a binary number equivalent to the given sub key *Key-P*. It is used as the initial value (26 *bits*) and control parameter (64 *bits*) of a logistic random bit generator. Then, a random bit string is generated. Subsequently, integer Pseudorandom numbers are calculated according to its range and Tompkins-Paige algorithm. Finally, Tompkins-Paige algorithm is applied to provide a permutation matrix for permutation of the pixels.

The block diagram of the system is illustrated in Figure 3. Three subsections, logistic random bit sequence (LRBS) generator, Pseudorandom number calculator, and Tompkins-Paige algorithm perform pixel permutation. An LRBS generator is needed as a first stage of the permutation matrix generator. An initial key is used as an initial value and control parameter of the logistic map. A bit-string is generated by LRBS and the integer Pseudorandom numbers are calculated by (3.1). The Tompkins-Paige algorithm is used to generate the target permutation matrix, which is obtained from repetition of some simple permutations.

Chaotic pixel permutation is used as the target permutation matrix to implement 1D and 2D image permutation [27].

3.1.1. Logistic Random Bit Sequence (LRBS) Generator

Generation of the chaotic random bit sequence is done as follows [28]. An appropriate chaotic map is selected. The logistic map is general and simple as mentioned in Section 2. The probability density function of logistic is not uniform, but by introducing a proper threshold level, the output of the bit sequence becomes uniform. The control parameter and initial

value of the map is determined. Then, a real value is generated by each iteration, which is converted into a bit by a single level threshold function. The threshold value is calculated using a computer simulation. Different values are considered and the occurrence of 0 and 1 is examined. The threshold is selected to 0.6 such that the probability of frequencies of 0 (f_0) and frequencies of 1 (f_1) is approximately equal. A sample histogram of a logistic map with $a = 3.9$ and $x_0 = 0.5$ is shown in Figure 4. The initial 90 *bit* Key- *P* contains a 26 *bit* initial value and a 64 bit control parameter. A string of 768*bits*, $\{b_0 \sim b_{767}\}$ is generated in 768 iterations of LRBS. If the real output of logistic map in the specific iteration is less than 0.6, the output bit of LRBS is 0, otherwise it is 1.

3.1.2. Pseudorandom Number Generator

Let b_i ($i = 0, 1, 2, \ldots$) be the ith output bit of the LRBS, which is generated according to the initial key, *Key-P*. $L - 1$ integer Pseudorandom numbers, g_i's ($i = 1, 2, \ldots, L - 1$) are calculated using these b_i's, as shown in [26]

$$
\begin{aligned}
g_1 &= 1, \\
g_2 &= \left\lfloor \frac{(2b_0 + b_1)(1)}{(2^2 - 1)} \right\rfloor + 1, \\
g_3 &= \left\lfloor \frac{(2b_2 + b_3)(2)}{(2^2 - 1)} \right\rfloor + 1, \\
g_i &= \left\lfloor \frac{(2^{j-1}b_k + 2^{j-2}b_{k+1} + \cdots + b_{k+j-1})(i-1)}{(2^j - 1)} \right\rfloor + 1,
\end{aligned}
\tag{3.1}
$$

where

$$
j = \lfloor \log_2 i \rfloor + 1, \qquad k = \sum_{s=2}^{i-1} (\lfloor \log_2 s \rfloor + 1).
\tag{3.2}
$$

$\lfloor x \rfloor$ denotes the floor of x. Since the number of permuted pixels ($L = 128$) is equal to the image size, 127 integers (g_i) and 768 *bits* (b_i) are required. It is obvious that the maximum value of every g_i is i, ($g_i < i + 1$).

3.1.3. Tompkins-Paige Algorithm

Tompkins-Paige algorithm gives a one-to-one correspondence between the integers and the permutation. As an example, the simple permutation of nine elements of *order* 7 and *degree* 3 is shown in the second row. The last 7 elements are disturbed and an end-around shift of 3 elements to the left are performed

$$
\begin{bmatrix}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
1 & 2 & 6 & 7 & 8 & 9 & 3 & 4 & 5
\end{bmatrix}.
\tag{3.3}
$$
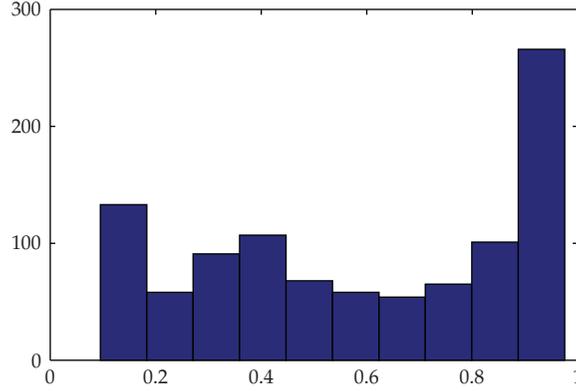
Figure 4: A sample histogram of a logistic map with $a = 3.9$ and $x_0 = 0.5$.

Therefore, the simple permutation of *order* $L - m_1 + 1$ and *degree* $m_2 - m_1$ is generally defined as shown in the matrix [26]

$$
\begin{bmatrix}
1 & 2 & \cdots & m_1 - 1 & m_1 & m_1 + 1 & \cdots & m_2 - 1 & m_2 & m_2 + 1 & \cdots & L \\
1 & 2 & \cdots & m_1 - 1 & m_2 & m_2 + 1 & \cdots & L & m_1 & m_1 + 1 & \cdots & m_2 + 1
\end{bmatrix}. \tag{3.4}
$$

In the above example, $m_1 = 3$, $m_2 = 6$, and $L = 9$.

In this example, the target permutation of these 9 elements can be obtained through compounding 8 simple permutations of *orders* 9 to 2, and for each order an associated degree. In each simple permutation, degree should be less than its order. In General, the target permutation on L elements is the result of compounding $L - 1$ simple permutations with order of L to 2 and degree of g_i, where g_i is less than the corresponding order in each simple permutation.

In this paper, the Tompkins-Paige algorithm is applied to $L = 128$ elements with order of 128 to 2 and degree of g_i. Finally, the 128×128 permutation matrix of the 128 elements is expressed in a 128 elements vector [27].

As mentioned earlier, $L = 128$ is the number of pixels, to be permuted and g_i's are calculated using (3.1), where, $g_i < i + 1$.

3.1.4. 1D Chaotic Pixel Permutation

The main idea behind the present work is that an image can be viewed as an arrangement of 2D pixels [29]. The intelligible information present in an image is due to the correlations among the pixels in a given arrangement. This perceivable information can be reduced by decreasing the correlation among the pixels using certain random permutation techniques.

The image can be seen as a 2D array of pixels, each with 256 gray scales. In pixel permutation techniques the pixels taken from the image are permuted with the key chosen from the key space. There are two options in 1D permutation process: row permutation or column permutation.

In row permutation, according to Figure 3, the *Key-P* is used to generate the permutation matrix. The pixels of all rows are rearranged with respect to the permutation

matrix/vector, as explained in Section 2.2.1. The result of these permutations is discussed in Section 4.

In column permutation, the pixels of all columns are rearranged with respect to the permutation matrix/vector according to the *Key-P*. The result of these permutations is also presented in Section 4.

3.1.5. 2*D Chaotic Pixel Permutation*

We extend the basic concept of 1*D* Chaotic pixel permutation in order to design a 2*D* permutation method. Here row and column permutation are applied simultaneously. In the 2*D* permutation of the image, the permutation matrix of the rows and columns could be either identical, using the identical *Key-P*, or different, using the dissimilar *Key-P*.

In identical permutation approach, the pixels of all rows are first rearranged, with respect to the permutation matrix/vector and the pixels of all columns are then rearranged with respect to the same permutation matrix/vector. The encrypted images that appear as a random noisy image are shown in Section 4.

In a different permutation approach, the pixels of all rows are first rearranged, with respect to the first permutation matrix/vector and the pixels of all columns are then rearranged, with respect to the second permutation matrix/vector. The encrypted images that appear as a random noisy image are also shown in Section 4.

3.2. *Chaotic Pixel Substitution Unit*

In the permutation part of the system, pixel positions are displaced without changing their gray level values. Hence, the histogram of the permuted image is similar to the histogram of the plain-image. The permuted image however cannot resist against "statistical" and "known plain text" attacks [6]. To improve the security of the proposed encryption system, its histogram needs to approximate the uniform distribution. This improvement is done using a substitution scheme, shown in Figure 5.

There are two main subunits in the substitution unit, tent Pseudorandom image generator and modulo addition. A 128 *bits* key (*Key-S*) is entered to the tent map, 64 *bits* as an initial value of the map and the remaining bits as the control parameter. Tent map is used to generate a Pseudorandom image. In each iteration of the tent map, a Pseudorandom number between 0 to 1 is generated. For each 128×128 image, 16 384 iterations are required. Since a 256 gray-scale image is selected, the Pseudorandom number should be linearly transformed to the range of $[0, 255]$. Then, the permuted image is modularly added with random image pixel-wise in mod 256. The substitution procedure decreases a correlation between pixels and makes the histogram more uniform.

3.2.1. Tent Pseudorandom Image (TPRI) Generator

There are two options to generate a chaotic Pseudorandom image. A chaotic random generator along with a simple threshold detector similar to Section 3.1 could be utilized. Afterward, every 8 *bits* stream should be converted to a gray scale of a pixel. As a second option, a chaotic random generator along with a linear transform may be used. The transformer is employed to convert a real range of $[0, 1]$ to an integer range of $[0, 255]$
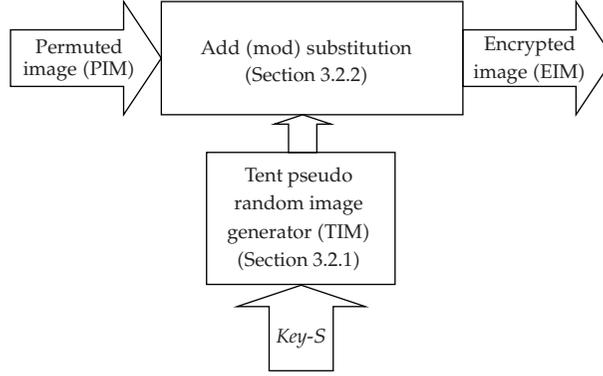
Figure 5: Block diagram of the chaotic pixel substitution unit.

linearly. It is completed by introducing 255 threshold levels. As the latter seems to be faster, modification of chaotic random generator has been done as follows:

(1) Tent map is chosen as a chaotic system instead of a logistic map, since its probability density function (PDF) is uniform and implementation is almost simple.

(2) Control parameter and initial condition of the map is determined by *Key-S*. Each of them is defined with 64 *bits* and a simple linear transformation.

(3) Real values of chaotic sequences are generated by iterations of the map: $x_0, x_1, x_2, \ldots, x_{(nxn)}$ where n is the image size.

(4) 255 threshold levels in the range $[0, 1)$ are defined and a gray scale of pixels from 0 through 255 are attributed to them, respectively.

TPRI output seems to be a noisy image and its histogram is uniform.

3.2.2. Modulo Addition for Chaotic Pixel Substitution

It is desirable to decrease intelligibility of the encryption image. That is achievable with a substitute operation such that the final histogram becomes uniform and correlation between pixels is reduced. The permuted image could therefore be mixed with a noise image, TPRI. Modulo addition/subtraction is more suitable than XOR/XNOR operation. In this research, modulo 256 additions are performed. At the encryption side, the 2*D* permuted image, called *PIM*, is added modularly with the TRPI Image, called *TIM*, pixel wise to generate the encrypted image, called *EIM*. This is shown in

$$\text{EIM}_{i,j} = \text{PIM}_{i,j} + \text{TIM}_{i,j} \bmod 256, \tag{3.5}$$

where i and j are the coordinates of the pixels in the range of $[0, 255]$.

In the decryption side, to recover the 2*D* encrypted image, the same TRPI Image should be modularly subtracted from the 2*D* encrypted image in mod 256. After that, the product image will be depermuted to retrieve a plain image. The simulation of the proposed encryption system is investigated in Section 4.
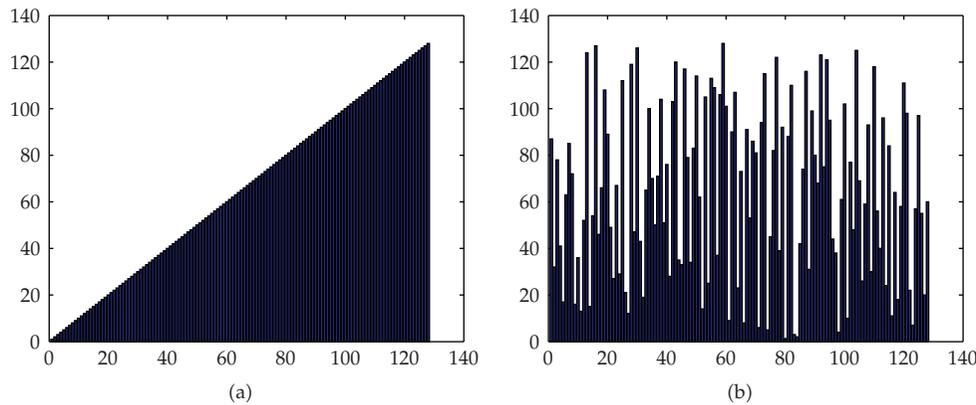
Figure 6: A sample of identity and permutation matrix.

4. Simulation of the Chaotic Image Encryption System

The proposed chaotic image encryption along with individual permutation and substitution has been simulated using MATLAB tools. In order to verify the exact operation of the proposed encryption system, and according to the process map of the system, that is, Figures 2, 3, and 5, the proposed chaotic image encryption has been coded and simulated. A 128×128 Lena image with 256 gray scales is used as a plain-image. The results obtained by the Lena gray scales image are demonstrated.

The proposed encryption system includes two major units, chaotic pixels permutation unit and chaotic pixels substitution unit. Three processes called logistic random bit sequence (LRBS) generator, Pseudorandom number calculator, and Tompkins-Paige algorithm are used to perform the pixel permutations.

First, the logistic map to generate a string of bits uses a *Key-P* with 90 *bits*. Since the chaotic range of the initial parameter is about 0.4, 26 *bits* of *Key-P* are used as an initial value. 64 *bits* are considered as a control parameter. Then 127 integer Pseudorandom numbers, degree, are calculated and used as the degree of permutation. The Tompkins-Paige algorithm is then used to find the target permutation by multiplication of 127 simple permutations. A sample of identity and permutation matrix of a sample key is presented in Figure 6.

Subsequently, as shown in Figure 5 and explained in Section 3.2, *Key-S* (128 *bits)* is entered as an input variable to the system, 64 *bits* as an initial value of the map and 64 *bits* as the control parameter. A Pseudorandom image of size 128 × 128 with 256 gray scales is generated using the tent map. The pixels of the Pseudorandom image are then modularly added with the pixels of permutated image of the previous phase to generate the final encrypted image.

Afterward, an 128 × 128 image with 256 gray scales (Figure 7) is used as a plain-image and applied to the proposed encryption system. The output of each stage is shown next. The results of row and column permutation unit are shown in Figures 8 and 9, respectively.

In Figure 10, the result of the 2D permutation of plain-image is illustrated, where the permutation matrixes of the rows and columns are identical using the identical *Key-P*'s. Figure 11 shows another 2D permutation with different permutation matrixes, using different *Key-P*'s. The encrypted images of Figures 10 and 11 approximate Pseudorandom noisy images.

(a)   (b)
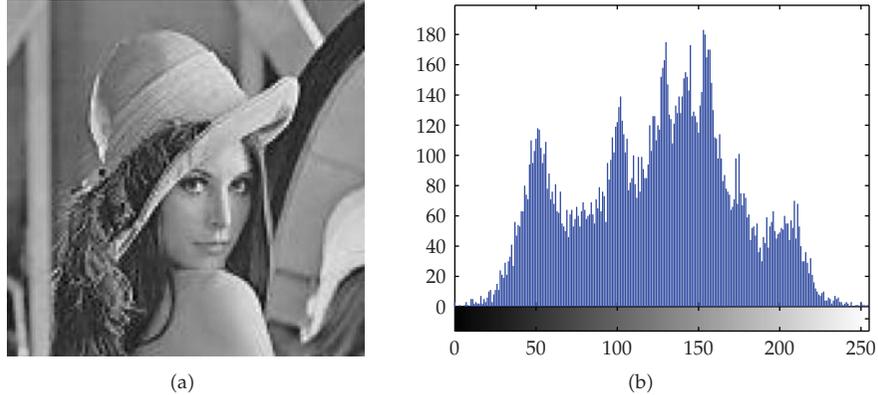
Figure 7: 128 × 128 Lena image with 256 gray scales and its image histogram.
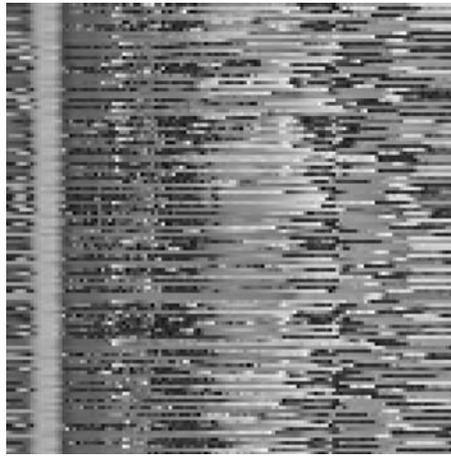
Figure 8: Row chaotic pixel permutation of Lena image.

A 128 *bits* key is then entered to the tent map, 64*bits* as an initial value of the map and 64*bits* as the control parameter. Tent map is used to generate a Pseudorandom image. In each iteration of the tent map, a Pseudorandom number between 0 to 1 is generated. The Pseudorandom number should be linearly transformed to a range of $[0, 255]$, since a 256 gray scale image is desired. Then, the permuted image is modularly added with Pseudorandom image pixel-wise in mod 256. Figure 12 illustrates an example of a 128 × 128 tent Pseudorandom image with 256 gray scales and its histogram. It is similar to a noisy image.

Finally, the 2D permuted image is modularly added with the TRPI Image pixel-wise in mod 256. The results of the modularly addition stage and its histogram are depicted in Figure 13. The final histogram clearly appears uniform.

5. Security Analysis

In this section, the performance of the proposed chaotic image encryption system is analyzed. The security analysis presented in this section is based on the performance of only one
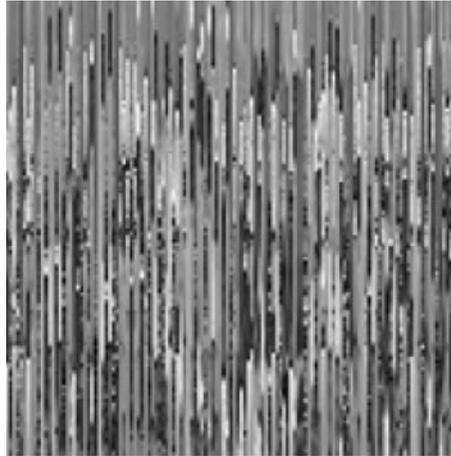
Figure 9: Column chaotic pixel permutation of Lena image.
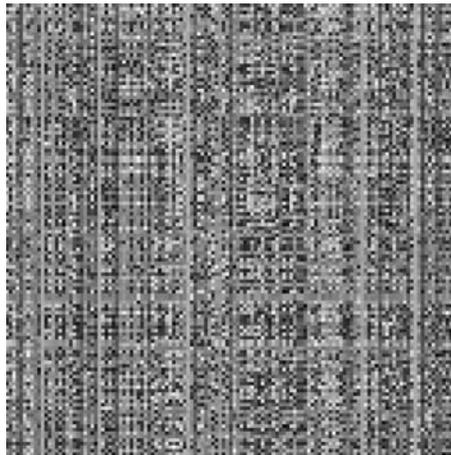
Figure 10: Row-Column chaotic 2D pixels permutation of Lena image (identical permutation matrixes for row and column).

round of operation of the proposed encryption system including a 2D permutation and a substitution. However, to improve the security of the proposed algorithm, more than one iteration can be applied with different keys. The first criterion for this security analysis is the chi-square test of histogram of each encrypted image. The second criterion is the correlation coefficients of pixels in the encrypted image in the vertical, horizontal, and diagonal directions. The third criterion is the difference between each encrypted and corresponding plain-image, which is measured by mean absolute difference, number of pixel change rate, and unified average changing intensity. The fourth criterion in this security analysis is key space.

5.1. Histogram

The histogram of the plain-image is illustrated in Figure 7. The histograms of all permuted images shown in Figures 8 to 11 are similar to the histogram of the plain-image. The
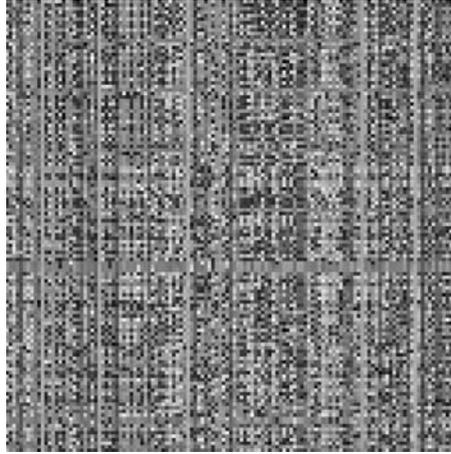
Figure 11: Row-Column chaotic 2D pixels permutation of Lena image (different permutation matrixes for row and column).

(a)   (b)
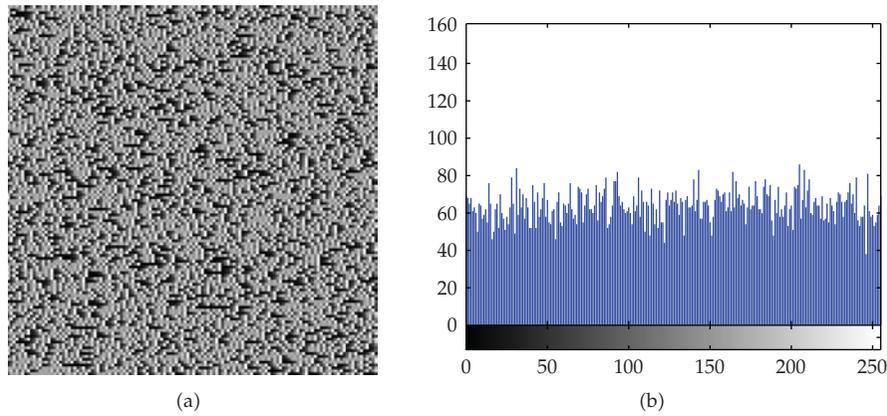
Figure 12: 128 × 128 tent Pseudorandom image with 256 gray scales and its image histogram.

histogram of the encryption system has to approximate the uniform distribution. The result of the encryption system and its histogram are illustrated in Figure 13. The histogram is approximated by a uniform distribution. The uniformity is justified by the chi-square test [30] in

$$\chi^2 = \sum_{k=1}^{256} \frac{(v_k - 64)}{64},$$

(5.1)

where k is the number of gray levels (256), v_k is the observed occurrence frequencies of each gray level (0–255), and the expected occurrence frequency of each gray level is 64. Assuming a significant level of 0.05, $\chi^2(255, 0.05) = 293$. Chi-square value for the final encrypted image of the proposed system is 290, $\chi^2(\text{test}) = 290$. This implies that the null hypothesis is not rejected and the distribution of the encrypted histogram is uniform, $\chi^2(\text{test}) < \chi^2(255, 0.05)$ [30].

(a)

(b)
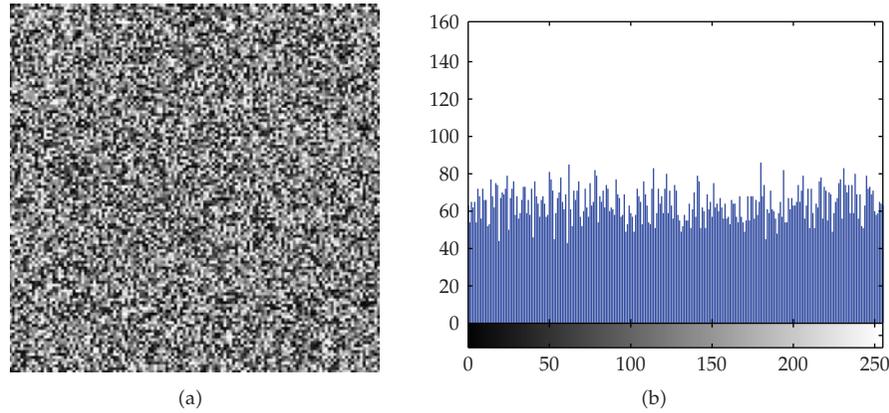
Figure 13: Proposed Chaotic Encrypted Image of Lena with 256 gray scales and its image histogram.

5.2. Correlation Coefficient

The proposed chaotic image encryption system should be resistant to statistical attacks. Correlation coefficients of pixels in the encrypted image should be as low as possible [12, 31]. Horizontal, vertical, and diagonal correlation coefficients (r_{xy}) of two adjacent pixels can be calculated using the following equations:

$$r_{xy} = \frac{\text{COV}(x, y)}{\sqrt{D(x)}\sqrt{D(y)}},$$

$$D(x) = \frac{1}{N} \sum_{i=1}^{N} \left(x_i - \frac{1}{N} \sum_{i=1}^{N} x_i \right)^2, \tag{5.2}$$

$$\text{COV}(x, y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - E(x))(y_i - E(y)),$$

where x and y are gray-scale values of two adjacent pixels in the image and E denotes the expectation operator shown in

$$E(z) = \frac{1}{N} \sum_{i=1}^{N} z_i. \tag{5.3}$$

About a thousand pairs of two adjacent (in vertical, horizontal, and diagonal direction) pixels are randomly selected from the encrypted image, and the correlation coefficients are calculated, respectively. The results are shown in Table 1. It is clear that the correlation coefficients of the proposed encrypted image (Figure 13) in all three directions are smaller than the correlation coefficients of the proposed permuted image (Figure 7). Correlation coefficients of tent Pseudorandom image (Figure 12) are also small.

Meanwhile, the correlation coefficients of the proposed methods (Table 1: column 5, Table 2: column 6) are compared with results of four other papers [6, 9, 10, 32], which are

Table 1: Comparison of correlation coefficients of the proposed methods.

Correlation coefficient	Plain-image (Figure 7)	Proposed permuted image (Figure 10)	Proposed tent Pseudorandom image (Figure 12)	Proposed encrypted image (Figure 13)
Horizontal (H)	0.798	0.043	0.149	0.005
Vertical (V)	0.867	0.271	0.033	0.011
Diagonal (D)	0.769	0.054	0.051	0.023
$(H^2 + V^2 + D^2)^{0.5}$	1.407	0.280	0.161	0.026
Average (H, V, D)	0.811	0.123	0.078	0.013

Table 2: Comparison of correlation coefficient of the proposed method and the other methods.

Correlation coefficient	Mao et al. [6]	Zhang et al. [9]	Gao et al. [10]	Zhou et al. [32]	Proposed encrypted image (Figure 13)
Horizontal (H)	0.045	0.082	0.016	0.012	0.005
Vertical (V)	0.028	0.040	0.065	0.027	0.011
Diagonal (D)	0.021	0.005	0.032	0.007	0.023
$(H^2 + V^2 + D^2)^{0.5}$	0.057	0.091	0.074	0.030	0.026
Average (H, V, D)	0.031	0.042	0.038	0.015	0.013

shown in Table 2. As shown, the average correlation coefficient of the proposed system is less than all of the other methods.

5.3. Difference between Encrypted and Plain-Images

The encrypted image should be significantly different to the original one. To quantify this requirement, three measures are used: mean absolute error (MAE), the number of pixel change rate (NPCR), and unified average changing intensity (UACI) [1, 30].

The performance of each stage of the difference between permuted/encrypted and plain-images is measured by the mean absolute error (MAE) criterion in

$$\text{MAE} = \frac{1}{L \times L} \sum_{j=1}^{L} \sum_{i=1}^{L} |a_{ij} - b_{ij}|, \tag{5.4}$$

where L, size of image, is equal to 128. The parameters a_{ij} and b_{ij} are gray-scale values of pixels in plain and encrypted images, respectively. The larger the MAE value, the better the encryption security. According to the selected key, the results are shown in Table 3. It is illustrated that MAE of the proposed column permutated image is about 23, while the MAE for row permuted image and 2D permuted image are about 27. The MAE of proposed encrypted image is about 35 that is 26 percent more than MAE of row and 2D permutation. It is obvious that substitution and permutation are more secure than only-permutation encryption systems.

Table 3: A comparison of MAE of different methods.

Proposed Methods	MAE
Row Permutation	27.48
Column Permutation	22.99
2D Permutation	27.84
Substitution and Permutation	35.13

Table 4: Comparison of NPCR and UACI criteria of proposed method and the others.

Criteria (expected value)	Mao et al. [6] first round	Zhang et al. [9] 2nd round	Gao et al. [10]	Zhou et al. [32] 2nd round	Proposed method first round
NPCR (99.61%)	37%	21.5%	NA	25.0%	99.7%
UACI (33.46%)	9%	2.5%	NA	8.5%	29.3%

The NPCR is the percentage of corresponding pixels with different gray levels in two images. Let $C_1(i, j)$ and $C_2(i, j)$ be the gray level of the pixels at the ith row and jth column of two W × H images. The NPCR of these two images is defined in

$$\text{NPCR} = \frac{\sum_{i,j} D(i, j)}{W \times H} \times 100\%, \tag{5.5}$$

where $D(i, j)$ is defined as

$$D(i, j) = \begin{cases} 0, & \text{if } C_1(i, j) = C_2(i, j), \\ 1, & \text{if } C_1(i, j) \neq C_2(i, j). \end{cases} \tag{5.6}$$

Another measure, UACI, is defined as the average intensity difference in a gray level of corresponding pixels and is defined as

$$\text{UACI} = \frac{1}{W \times H} \left[\sum_{i,j} \frac{C_1(i, j) - C_2(i, j)}{2^L - 1} \right]. \tag{5.7}$$

Considering two Pseudorandom images, the expected value of NPCR is found to be 99.61%. The proposed method is evaluated using this criterion and NPCR of Figure 13 is 99.7%. In the case of two Pseudorandom images, the expected value of UACI can be computed as 33.46%, assuming each gray level is coded with 8 *bits*. The proposed method is evaluated using this criterion too and UACI of image in Figure 13 is 29.3%.

As shown in Table 4 our proposed method with 0.09% difference to the expected value of NPCR is improved compared to the other reported methods. It also shows that, our method has advantage with respect to UACI criteria, with a difference of about 10% to its expected value.

Table 5: Comparison of Key length of proposed method and the others.

Key	Mao et al. [6]	Zhang et al. [9]	Gao et al. [10]	Zhou et al. [32]	Proposed Method
Length (Bin.)	2^{128}	NA	2^{150}	2^{112}	2^{218}
Length (Dec.)	10^{38}	NA	10^{45}	10^{33}	10^{65}

5.4. Key Space Analysis

Key space should be sufficiently large to make brute-force attack infeasible. Key space is the total number of different keys that can be used in the encryption system. The keys of the proposed system in this paper consist of permutation key, *Key-P* = 90 *bits,* and substitution key, *Key-S* = 128 *bits*. Each key includes initial value and control parameter of corresponding chaotic maps. Only 26 *bits* are used as parameter of logistic map, since the chaotic range of logistic map is about 40% of a chaotic range of tent map. The total key length is 218 *bits*, which contain three equal 64 *bits* plus 26 *bits*. Therefore, the key space is 2^{218}, that is, $4.12 \times 10^{+65}$. It is shown that, the key space is large enough to resist the proposed system against any brute-force attack. Comparison of the key length in our proposed method with the others is shown in Table 5.

It is possible to increase the number of bits for total key in hardware implementation. However, by increasing the key length, volume of hardware is increased and consequently speed of the system is decreased. With respect to the speed of the today's computers, the key space size should be more than $2^{100} = 10^{30}$ in order to avoid brute-force attacks [33].

6. Conclusion

In this paper, we presented a new permutation-substitution image encryption architecture using chaotic maps and Tompkins-Paige algorithm. The proposed encryption system included two major parts, chaotic pixels permutation and chaotic pixels substitution. A logistic map was used to generate a bit sequence, which was in turn used to generate Pseudorandom numbers in Tompkins-Paige algorithm, in pixel permutation phase. Pixel substitution phase, included two processes, the tent Pseudorandom image (TPRI) generator and modulo addition operation. A tent map was used to produce a Pseudorandom image that was mixed with the permuted image.

The permutation and substitution operations needed two different keys, *Key-P* and *Key-S*, respectively. The total key length was 218 *bits*. Therefore, the key space was 2^{218}, that is, $4.12 \times 10^{+65}$, which was large enough to protect the system against any brute-force attacks.

The image was a 2*D* array of pixels, each with 256 gray scales. The 2*D* permutation was designed by permutation of rows and columns simultaneously. To improve security of the proposed encryption system, the histogram needed to become uniform. This was achieved by pixel substitution. There were two main parts for pixel substitution here, tent Pseudorandom image generator and modulo addition operation. A *Key-S* was entered in to the tent map to generate a Pseudorandom image with uniform histogram. Subsequently, pixels of permuted image were modularly added to pixels of random image with uniform distortion.

All parts of the proposed chaotic encryption system were simulated using a computer code. The histogram of the encrypted image was approximated a uniform distribution. The uniformity was justified by the chi-square test. Chi-square value shows that the distribution of the histogram of the encrypted image is uniform. The vertical, horizontal, and

diagonal correlation coefficients, as well as their average and RMS values for the proposed encrypted image were calculated. The individual values and their average and RMS values of correlation coefficients were lower than the corresponding values from previous research by a factor between 13% to 70%. Therefore, the proposed encryption system was resistant against any statistical attack.

To quantify the difference between encrypted image and corresponding plain-image, three measures were used: mean absolute error (MAE), number of pixel change rate (NPCR), and unified average changing intensity (UACI). It was concluded that the NPCR and UACI criteria of the proposed system were satisfactory when compared to other research results as was the security performance of the proposed system. All these results were obtained in only one round of encryption process.

Acknowledgment

References

[1] Y. V. Mitra, S. Rao, and S. R. M. Prasanna, "A new image encryption approach using combinational permutation techniques," *International Journal of Computer Science*, vol. 1, no. 2, pp. 127–131, 2006.

[2] D. Van de Ville, W. Philips, R. Van de Walle, and I. Lemahieu, "Image scrambling without bandwidth expansion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 6, pp. 892–897, 2004.

[3] M. Yang, N. Bourbakis, and S. Li, "Data-image-video encryption," *IEEE Potentials*, vol. 23, no. 3, pp. 28–34, 2004.

[4] J. Fridrich, "Image encryption based on chaotic maps," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 1105–1110, 1997.

[5] G. Chen, Y. Mao, and C. K. Chui, "A symmetric image encryption scheme based on 3D chaotic cat maps," *Chaos, Solitons & Fractals*, vol. 21, no. 3, pp. 749–761, 2004.

[6] Y. Mao, G. Chen, and S. Lian, "A novel fast image encryption scheme based on 3D chaotic baker maps," *International Journal of Bifurcation and Chaos*, vol. 14, no. 10, pp. 3613–3624, 2004.

[7] Z.-H. Guan, F. Huang, and W. Guan, "Chaos-based image encryption algorithm," *Physics Letters A*, vol. 346, no. 1–3, pp. 153–157, 2005.

[8] S. Lian, J. Sun, and Z. Wang, "A block cipher based on a suitable use of the chaotic standard map," *Chaos, Solitons & Fractals*, vol. 26, no. 1, pp. 117–129, 2005.

[9] L. Zhang, X. Liao, and X. Wang, "An image encryption approach based on chaotic maps," *Chaos, Solitons & Fractals*, vol. 24, no. 3, pp. 759–765, 2005.

[10] H. Gao, Y. Zhang, S. Liang, and D. Li, "A new chaotic algorithm for image encryption," *Chaos, Solitons & Fractals*, vol. 29, no. 2, pp. 393–399, 2006.

[11] Q. Zhou, K.-W. Wong, X. Liao, T. Xiang, and Y. Hu, "Parallel image encryption algorithm based on discretized chaotic map," *Chaos, Solitons & Fractals*, vol. 38, no. 4, pp. 1081–1092, 2008.

[12] A. N. Pisarchik and M. Zanin, "Image encryption with chaotically coupled chaotic maps," *Physica D*, vol. 237, no. 20, pp. 2638–2648, 2008.

[13] N. K. Pareek, V. Patidar, and K. K. Sud, "Image encryption using chaotic logistic map," *Image and Vision Computing*, vol. 24, no. 9, pp. 926–934, 2006.

[14] M. S. Ehsani and S. E. Borujeni, "Fast Fourier transform speech scrambler," in *Proceedings of the 1st International IEEE Symposium on Intelligent Systems*, vol. 1, pp. 248–251, 2002.

[15] S. E. Borujeni and A. Zakerolhoseini, "Permutation based image encryption using pseudo random number generator and Tompkins-Paige algorithm," in *Proceedings of the International Conference on Robotics, Vision, Information and Signal Processing*, pp. 478–481, Penang, Malaysia, 2007.

[16] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, pp. 656–715, 1949.

[17] D. R. Stinson, *Cryptography: Theory and Practice*, CRC Press Series on Discrete Mathematics and Its Applications, Chapman & Hall/CRC, Boca Raton, Fla, USA, 2nd edition, 2002.

[18] K. T. Alligood, T. D. Sauer, and J. A. Yorke, *Chaos: An Introduction to Dynamical Systems*, Textbooks in Mathematical Sciences, Springer, New York, NY, USA, 1997.

[19] B. Furht and D. Kirovski, *Multimedia Security Handbook*, CRC Press, Boca Raton, Fla, USA, 2005.

[20] J.-C. Yen and J.-I. Guo, "A new chaotic key-based design for image encryption and decryption," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '00)*, vol. 4, pp. 49–52, Geneva, Switzerland, May 2000.

[21] L. P. L. de Oliveira and M. Sobottka, "Cryptography with chaotic mixing," *Chaos, Solitons & Fractals*, vol. 35, no. 3, pp. 466–471, 2008.

[22] S. Lian, J. Sun, and Z. Wang, "Security analysis of a chaos-based image encryption algorithm," *Physica A*, vol. 351, no. 2–4, pp. 645–661, 2005.

[23] W. Yuanzhi, R. Guangyong, J. Julang, Z. Jian, and S. Lijuan, "Image encryption method based on chaotic map," in *Proceedings of the 2nd IEEE Conference on Industrial Electronics and Applications (ICIEA '07)*, pp. 2558–2560, 2007.

[24] S. Li, G. Chen, and X. Zheng, "Chaos-based encryption for digital images and videos," in *Multimedia Security Handbook*, pp. 133–167, CRC Press, Boca Raton, Fla, USA, 2004.

[25] S. E. Borujeni, "Speech encryption based on fast Fourier transform permutation," in *Proceedings of the 7th IEEE International Conference on Electronics, Circuits and Systems (ICECS '00)*, vol. 1, pp. 290–293, 2000.

[26] K. Sakurai, K. Koga, and T. Muratani, "A speech scrambler using the fast Fourier transform technique," *IEEE Journal on Selected Areas in Communications*, vol. 2, no. 3, pp. 434–442, 1984.

[27] G. Polya, *Applied Combinatorial Mathematics*, Krieger, 1981.

[28] H.-P. Xiao and G.-J. Zhang, "An image encryption scheme based on chaotic systems," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 2707–2711, Dalian, China, 2006.

[29] J. Zou, C. Xiong, D. Qi, and R. K. Waro, "The application of chaotic maps in image encryption," in *Proceedings of the 3rd International IEEE Northeast Workshop on Circuits and Systems Conference (NEWCAS '05)*, pp. 331–334, Quebec City, Canada, 2005.

[30] H. S. Kwok and W. K. S. Tang, "A fast image encryption system based on chaotic maps with finite precision representation," *Chaos, Solitons & Fractals*, vol. 32, no. 4, pp. 1518–1529, 2007.

[31] S. E. Borujeni and M. Eshghi, "Design and simulation of encryption system based on PRNG and Tompkins-Paige permutation algorithm using VHDL," in *Proceedings of the International Conference on Robotics, Vision, Information and Signal Processing*, pp. 63–67, Penang, Malaysia, 2007.

[32] F. Zhou, G. Cao, and B. Li, "Design of digital image encryption algorithm based on compound chaotic system," *Journal of Harbin Institute of Technology*, vol. 14, supplement 2, pp. 30–33, 2007.

[33] G. Alvarez and S. Li, "Some basic cryptographic requirements for chaos-based cryptosystems," *International Journal of Bifurcation and Chaos*, vol. 16, no. 8, pp. 2129–2151, 2006.

Research Article

An LPV Fractional Model for Canal Control

Rubén Martínez-González,[1] Yolanda Bolea,[1] Antoni Grau,[1] and Herminio Martínez-García[2]

[1] *Automatic Control Department, Technical University of Catalonia, Pau Gargallo 5, 08028 Barcelona, Spain*
[2] *Electronics Engineering Department, Technical University of Catalonia, c/Urgell 187, 08036 Barcelona, Spain*

Correspondence should be addressed to Antoni Grau, antoni.grau@upc.edu

An LPV rational order control model of an irrigation canal is derived from system identification experiments. This model is experimentally obtained by using the described LPV fractional identification procedure. This procedure consists of the identification of a rational order model in each operation point in an experimental test canal. Global LPV model is obtained from polynomial interpolation of local model parameters. Validation results demonstrate that rational order models are more accurate than integer order models. Therefore rational order control models have an important role to play in management and efficient use of water resources.

1. Introduction

Water is becoming a precious and very scarce resource in many countries due to the increase of industrial and agricultural demands, as well as population growth. Irrigation is the main water consuming activity in the world, as it represents about 80% of the available fresh water consumption. There is growing interest for the application of advanced management methods that prevent wastage and facilitate the efficient use of this vital resource [1].

Unfortunately, for control design purposes control techniques and their implementation are directly proportional to the complexity of proposed control models. Then, it is essentially a noncomplex and simple control model that represents in a precise way water behavior of open-flow canals. However, this type of systems corresponds to long distributed systems with complex dynamics. Furthermore, these systems involve mass energy transport phenomena which behave as intrinsically distributed parameter systems, and their characteristics are very complex such as the variation of parameters with operation points, large delays that vary with operation point, and numerous interactions between different consecutive subsystems and strong nonlinearity. Their complete dynamics is

represented by nonlinear partial differential hyperbolic equations (PDEs) that depend on the time as well as the spatial coordinates: Saint-Venant's equations. This equation system has unknown analytical solution in real geometry and it has to be solved numerically (characteristic method, Preissman implicit scheme, etc.) [2].

Resulting time consuming simulation models are therefore suitable for scientific purposes but they are too complex for on-line applications and control needs. Moreover, linearizations or simplifications of Saint-Venant's equations are currently studied by irrigation control research community [3]. Distributed parameters systems, considered as systems with a very large number of states could be approximated with low-order linear time invariant (LTI) models in order to use classical linear control design tools, as an usual practice in control engineering. There are two main approaches that are followed to obtain a linear model for irrigation main canals: the use of linearized Saint-Venant equations [4, 5] and the use of identification methods [6–8]. In case of open canal hydraulic system, identification is a classical method because their operational data are widely available and resulting models are suitable for design control.

Normally, classical identification methods [9] are used to obtain LTI discrete models which describe dynamics of irrigation water. However, in such systems LTI models lose information about these characteristics (non-linearity, coupling between pools, dynamics parameters changing over operation time in a wide range variationĚ). Then, a simplified control model structure that still preserve their information is needed. Such a structure can be provided by linear parameter varying (LPV) models consisting of a linear lumped parameter model in which parameters are not constant, but they depend on external parameters and/or system states and/or operating conditions of the system.

One of the main motivations for using LPV gain scheduling control versus classical gain scheduling control is that the former, as opposed to the latter, rigorously guarantees system stability [10]. Gain scheduling control is a heuristic method that consists in dividing the parameter space into small regions, in which the plant is observed as an LTI system, and LTI controllers are designed for every fixed set of parameters to achieve a synthetic controller with the use of interpolation or other techniques as switching techniques or fuzzy control. Heuristic gain scheduling controllers normally guarantee control system stability when parameters perform a slow variation [11] but sometimes may lead to unstability or chaotic behavior [12]. Furthermore, benefits of using gainscheduling techniques instead of robust control are obvious in this type of systems because of conservative results of robust control since model errors are partly due to non-linear effects and partly to the strong unknown perturbations considered as uncertainties [13]. Then, it is convenient to identify an LPV model for control canal purposes. Mainly, there are two approaches of identifying LPV models: since an LPV model is essentially a parameterized family of LTI models, a first identification approach is to collect data enough at each operating point to identify its corresponding LTI model [14]. Identified LTI coefficients are used to interpolate LPV coefficients as polynomial functions of scheduling variables.

Alternatively, a second approach that can be carried out in "one shot" , by assuming a linear dependence of parameters with operating points. Here, according to [15], identification problem can be reduced to a linear regression that may be solved using an extended regressor in the Least Mean Square (LMS) algorithm. In general, both methods lead to similar models. These identified LTI integer models do fit good enough with the dynamics of the canal system in each operating point in order to lineary control the system in such points. But, due to (i) that recently some control researchers have used fractional control methods for canal control purposes with satisfactory results [16] and (ii) noninteger models describe completely the

behaviour of distributed systems [17], such as irrigation canals, in this article authors have carried out an LPV fractional identification using the former mentioned LPV approach. This fractional identification approach has been developed to model an irrigation prototype canal. Some properties of fractional calculus are applied in order to obtain a noninteger order model in each operation point.

2. LPV Noninteger Order Modeling for Irrigation Canal Pool

The last two decades have witnessed considerable development in the use of fractional differentiation in various fields. Fractional control is now mature enough and is widely used to design control for representing systems that present diffusive phenomena, electromechanical diffusion, and transport phenomena. This last phenomenon corresponds to the case of irrigation pools. In this section, LPV identification methodology used for the experimental modelling of a pilot canal plant is described.

2.1. Pilot Canal Plant Description

An experimental canal prototype (this experimental test canal is a part of a more complex laboratory research canal available at Automatic Control Dept, UPC, Barcelona) is used in the research presented in this paper (Figure 1). This plant consists on two tanks, P_1 and P_2 (Figure 2), with a top side view shown in Figure 3. On one side of pool P_1 there is a pump (B_2, $1.3\,kl/h$) to empty the pool. The output-flowing liquid of B_2 is collected in P_2, where there is a second pump (B_3, $1.3\,kl/h$) to empty the pool. The output-flowing liquid of B_3 is collected in a reservoir, R, located under P_2. The reservoir supplies flow to the pool P_1 by another pump (B_1, $3.8\,kl/h$). In fact, the plant is a closed system, where the liquid that arrives to the reservoir from the pool P_2 returns to the pool P_1 via the pump. Lengthening the water path, tank plant is easily converted into a canal plant. The water path can vary placing methacrylate plates along the structure, (Figure 3). Here, the plates are separated $2\,cm$ away creating a zigzag path. Then, pools are enlarged from $2\,m$ to $12\,m$ long, $15\,cm$ wide, and the maximum allowed level is $25\,cm$. To know pools' levels after the zigzag path, that is, the pool level at the end of their path, two ultrasonic level sensors, y_1 and y_2, with a precision of $1\,mm$ are used. The sensors are attached to the canal metallic structure.

2.2. Preliminary Definitions in Fractional Modeling

The mathematical definition of fractional derivatives has been the subject of several different approaches [17]. In this paper the following definition of fractional discrete derivative,

$$\Delta^\alpha y_k = \sum_{j=0}^{k} w_j^\alpha y_{k-j} \quad 0 < \alpha < 1, \tag{2.1}$$

where

$$w_j^\alpha = (-1)^j \binom{\alpha}{j} \tag{2.2}$$

will be used; α is the order of the fractional difference.

Figure 1: Frontal view of the experimental prototype canal.
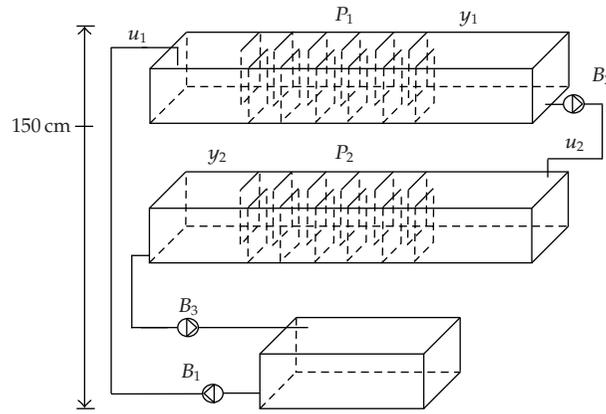
Figure 2: Full structure of the plant.

The fractional order models are clasificated in commensurable and non-commensurable order models. In this work, commensurable models are used.

Definition 2.1. A system is of commensurable order if it can be represented by a differential equation where all the orders of derivation are integers multiple of an order basis, α, that is, systems where the next condition is fulfilled:

$$a_n \Delta^{\gamma_n} y(t) + a_{n-1} \Delta^{\gamma_{n-1}} y(t) + \cdots + a_0 \Delta^{\gamma_0} y(t) = b_m \Delta^{\beta_m} u(t) + b_{m-1} \Delta^{\beta_{m-1}} u(t) + \cdots + b_0 \Delta^{\beta_0} u(t)$$

$$\gamma_k, \beta_k = k\alpha, \quad \alpha \in \mathbb{R}^+.$$

(2.3)

So, the differential equation (2.3) can be written as follow:

$$\sum_{k=0}^{n} a_k \Delta^{k\alpha} y(t) = \sum_{k=0}^{m} b_k \Delta^{k\alpha} u(t).$$

(2.4)
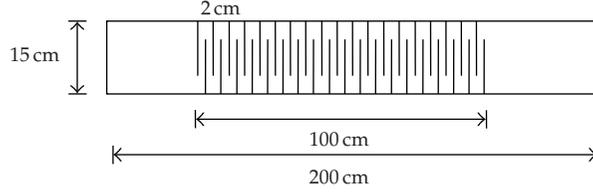
Figure 3: Top side view of the tank, converted into a pool.

Definition 2.2. A system is of rational order, if it is a commensurable order system and besides fulfills the condition of $\alpha = 1/q$ for all $q \in \mathbb{N} \mid q \neq 0$.

From the previous definition and based on the property of "q", *an integer order system is a particular case of rational order systems, where q = 1 .*

Consider the fractional discrete linear system, described by the state-space equations

$$\Delta^{\alpha} x_{k+1} = Ax_k + Bu_k; \quad k \in \mathbb{Z}^+,$$
$$y_k = Cx_k, \quad (2.5)$$

where $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $y_k \in \mathbb{R}^p$ are the state, input, and output vectors and $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$, $C \in \mathbb{R}^{p \times n}$. Using Definition 2.1, equations (2.5) can be written in the form

$$x_{k+1} + \sum_{j=1}^{k+1} w_j^{\alpha} x_{k-j+1} = Ax_k + Bu_k, y_k = Cx_k. \quad (2.6)$$

2.3. LPV Fractional Identification Methodology

LPV identification method used in this article is a two-step procedure where (1) nonfractional models are identified at several different equilibrium (operating condition) by classical methods [9]; (2) a global multimodel is obtained by interpolating among the local nonfractional models [14]. In this paper, a nonlinear leastsquares estimation method, based on Levenberg-Marquardt [18, 19], is used to obtain the parameters of the rational identified model in each operation point [20]. Local identification method forces rational local models to fit the system separately and locally. This local identification procedure (in each operation point) is standard and it can be itemized as follows: (1) design of the experiment and collection of input-output data in each operation mode from the process to be identified; (2) model structure selection in each operation point; (3) parameter estimation in each operation point; (4) model validation in each operation point.

As the LPV model is interpolated between local rational models, varying parameters of LPV model can be locally interpreted as parameters of the interpolated rational model. Varying parameters in each operation point are interpolated in a polynomical way. This polynomial depends on a scheduling parameter vector $\theta \in \mathbb{R}_+^2$, in this case $\theta = [u_1, u_2]$, that corresponds to the integral of pump activation in each canal that changes in their operating ranges. These values correspond to the upstream levels and are proportional to the upstream flow of each pool. Once the LPV model is obtained, it is validated globally.

Table 1: Operation points for pool P_1.

Pool P_1	Operation range [cm]; $u_2 = 0.5$
$OP_{1_{P_1}}$	$u_1 \in [0.0000, 1.5398]$
$OP_{2_{P_1}}$	$u_1 \in [1.5398, 3.1241]$
$OP_{3_{P_1}}$	$u_1 \in [3.1241, 4.6063]$
$OP_{4_{P_1}}$	$u_1 \in [4.6063, 6.3979]$
$OP_{5_{P_1}}$	$u_1 \in [6.3979, 8.3671]$

Table 2: Operation points for pool P_2.

Pool P_2	Operation range [cm]; $u_1 = 0.5$
$OP_{1_{P_2}}$	$u_2 \in [0.0000, 0.9396]$
$OP_{2_{P_2}}$	$u_2 \in [0.9396, 1.8679]$
$OP_{3_{P_2}}$	$u_2 \in [1.8679, 2.8067]$
$OP_{4_{P_2}}$	$u_2 \in [2.8067, 3.7535]$
$OP_{5_{P_2}}$	$u_2 \in [3.7535, 4.6989]$
$OP_{6_{P_2}}$	$u_2 \in [4.6989, 5.6261]$
$OP_{7_{P_2}}$	$u_2 \in [5.6261, 6.5445]$
$OP_{8_{P_2}}$	$u_2 \in [6.5445, 7.4333]$

In this paper, this system identification procedure is used to obtain a reliable dynamic model of a main irrigation canal when the design of a model-based control system is requested.

3. Experiment Design and Model Structure Selection

For identification of the pilot canal system different experiments have been carried out. These canal pools are operated by means of a downstream water level regulation method. Available measurements are downstream water levels (y_1 for pool P_1 and y_2 for pool P_2) and pump voltage (u_{P_1} for pump B_1 and u_{P_2} for pump B_2). Then, for the identification of the control model canal, as output variables, downstream levels are used, and as input variables integral pump voltage variables (u_1 and u_2) are used. According to literature [5, 21, 22], this model obtained after identification corresponds to a first-order model with delay with an integrator or to a second order model with delay with an integrator, depending on the geometry of the pool.

The appearance of integrator pole, or in other words, the fact that a reach has similarities with a swimming pool or a tank, is not a real surprise and is, in some case, expected. As mentioned before, this pole appears clearly in the uniform case regime and has been successfully included in several simplified models proposed in other works (Integrator Delay (ID) model [23], Integrator Delay Zero (IDZ) model [4], etc.). It is known that the identification of a system with integrators is very erratic about the exact localization of its poles. For this reason, the identified model relates the downstream levels (model outputs) and the integral of pump voltages (model inputs: u_1 for pool P_1 and u_2 for pool P_2).

3.1. Experiment Design

To obtain data containing the maximum information about the canal pools dynamic behaviour, pools must be excited with a persistent input signal that contains the largest number of frequencies representative of the system dynamics [9]. Then a pseudorandom binary sequence (PRBS) is a kind of signal that fulfills these conditions. Since these signals are suitable to identify linear systems and our system is nonlinear and timevarying, a PBRS is used in each operating point within the working range of the system. These signals are integrated (because the system has implicitly an integrator [5]) generating the input for the identification process, u_1 and u_2.

The sampling time T was selected to be 0.5 second because it is enough dueto the system dynamics. Pools act in different operating points. As the pool dynamics are different (due to their input pumps) five points have been selected for pool P_1 ($OP_{k_{P_1}}$, $k = 1, \ldots, 5$) and eight points for pool P_2 ($OP_{k_{P_2}}$, $k = 1, \ldots, 8$); see Tables 1 and 2.

3.2. Model Structure Selection

The model structure selection constitutes one of the most important and difficult decisions in system identification procedure because model complexity influences the accuracy of the description of the real process and the control schemes. Saint-Venant equations [24] represent the dynamics of an open flow canal in a precise and complete manner. This pair of partialdifferential equations constitutes a nonlinear hyperbolic system, which has no analytic solution for arbitrary geometry. However, such equations are not useful for designing a controller using linear theory as already noticed by [4, 25]. In these references, a simplified control-oriented model methodology is proposed that describes an n-pool canal system. In this methodology each pool is modeled around a given operating point using the transfer function matrices:

$$\begin{bmatrix} Y_1(s) \\ Y_2(s) \end{bmatrix} = \begin{bmatrix} P_{11}(s) & P_{12}(s) \\ P_{21}(s) & P_{22}(s) \end{bmatrix} \begin{bmatrix} Q_1(s) \\ Q_2(s) \end{bmatrix}, \tag{3.1}$$

where $Y_1(s)$ and $Y_2(s)$ are the upstream and downstream water levels of pools, respectively, and $Q_1(s)$ and $Q_2(s)$ are the upstream and downstream flow levels of pools considered here. $P_{12} = 0$ and $P_{21} = 0$ because, normally, control models do not take into account the strong coupling between canals because SISO controllers and decouplers are used [26], and the model structure for each pool:

$$P_{ij}(s) = \frac{1}{s} \frac{k}{S_{ij}s^2 + M_{ij}s + 1} e^{-\tau_{ij}s}, \tag{3.2}$$

where $i = j$ and $i = 1, 2$ are transfer functions relating downstream flows with upstream levels. Additionally, there is a relationship between discharge flow and pump voltage. The upstream flow of each pool can be related with its upstream level equivalent to the integral
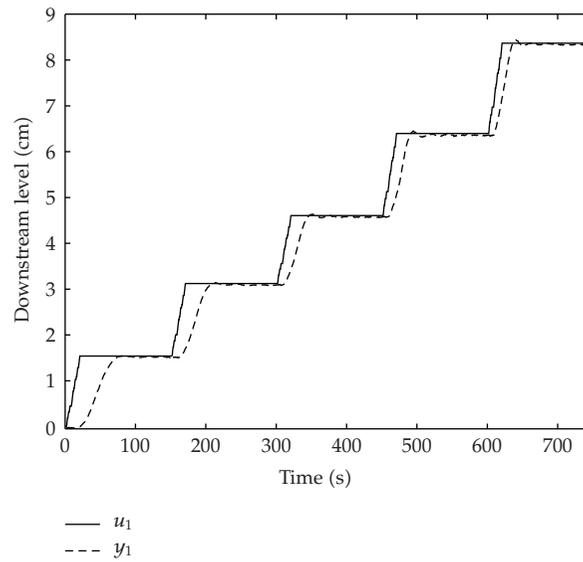
Figure 4: Downstream level for pool P_1, y_1 [cm], and pump input voltage integral, u_1 [cm].
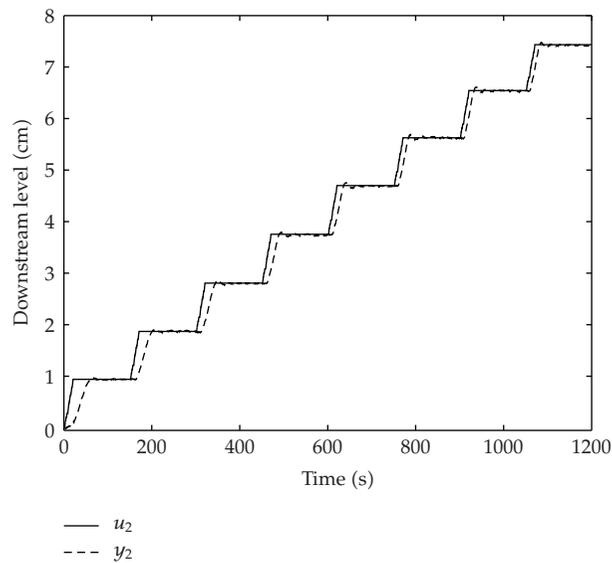
Figure 5: Downstream level for pool P_2, y_2 [cm], and pump input voltage integral, u_2 [cm].

of pump voltages, respectively in a linear way. The following additional relationship should be considered [27]:

$$Q_i(s) = \alpha_i U_i(s). \tag{3.3}$$

The second-order system behaviour can be clearly observed in Figures 4 and 5 when the integral of pump voltage is used as input of the identification model. As it is studied

in literature, in backwater part of each pool the dynamics are complicated: waves move up and down and reflect against the boundaries. However, at low frequencies, the water level "integrates" flow variations in the backwater part. In other words, the backwater can be considered to behave as an integrator or reservoir for low frequencies, and for this reason the integrator is included in the control model.

In order to identify the canal system, the continuous model is discretized by using zero-order hold method. Furthermore we assume that control model is LPV (as it is explained in Section 2.3). For each operating point in each pool ($OP_{k_{P_1}}$ for pool P_1 and $OP_{k_{P_2}}$ for pool P_2, see Tables 1 and 2), the discretized model can be expressed as

$$P_d(z,\theta) = \frac{a_3(\theta)z + a_4(\theta)}{z^{-2} + a_1(\theta)z^{-1} + a_2(\theta)} z^{-\tau(\theta)/T}. \tag{3.4}$$

Observing and analyzing the PRBS responses obtained at each operation point (see Figures 4 and 5) in our prototype canal, the canal dynamics can be represented by a seconds order equation with delay, as it is often used in the literature by Hayami model in linear and integer control [1]. As canals are systems that vary according to the operation point, an LPV Hayami model is more suitable [28]. Besides, as canals are nonlinear systems and with distributed parameters, fractional control models are suitable because they yield a more accurate behavior representation. It is desirable to hold the maximum degree of the dynamical equation (second order). So, our models in each operation point are of n-rational order with $n\alpha = 2$. Then, as defined by (2.5), the proposed model structure for $\alpha = 0.5$ and $n = 4$ is

$$\Delta^{0.5}x_{k+1} = A_{0.5}(\theta)x_k + B_{0.5}(\theta)u_k,$$
$$y_k = C_{0.5}(\theta)x_k, \tag{3.5}$$

where $x_k \in \mathbb{R}^4$, $u_k \in \mathbb{R}$, $y_k \in \mathbb{R}$ and

$$A_{0.5}(\theta) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -b_4(\theta) & -b_3(\theta) & -b_2(\theta) & -b_1(\theta) \end{bmatrix}, \tag{3.6}$$

$B_{0.5}(\theta) = [0\ 0\ 0\ b_4(\theta)]^T$ and $C_{0.5}(\theta) = [1\ b_7(\theta)\ b_6(\theta)\ b_5(\theta)]$.

For $\alpha = 0.25$ and $n = 8$, the proposed model structure is

$$\Delta^{0.25}x_{k+1} = A_{0.25}(\theta)x_k + B_{0.25}(\theta)u_k,$$
$$y_k = C_{0.25}(\theta)x_k, \tag{3.7}$$

Table 3: Model parameters obtained by identification in each operating point $OP_{k_{P_1}}$: pool P_1

Parameters	$OP_{1_{P_1}}$	$OP_{2_{P_1}}$	$OP_{3_{P_1}}$	$OP_{4_{P_1}}$	$OP_{5_{P_1}}$
τ	17	10	8	6	5
a_1	−1.9527	−1.9287	−1.9008	−1.8960	−1.8659
a_2	0.9537	0.9309	0.9051	0.9011	0.8747
a_3	0.0010	0.0022	0.0043	0.0051	0.0088
a_4	0.0203	−0.0261	−0.0409	0.0063	−0.0028
b_1	−0.1313	−0.1144	−0.0757	−0.0667	−0.0080
b_2	0.0370	0.0333	0.0261	0.0226	0.0212
b_3	0.0019	0.0052	0.0104	0.0139	0.0198
$b_4 (\times 10^{-3})$	0.0995	0.1224	0.1793	0.1969	0.3952
b_5	633.48	606.00	553.53	544.37	229.55
b_6	−345.72	−352.92	−335.57	−363.27	−179.48
b_7	55.0820	74.408	82.686	95.255	62.955
c_1	−2.1318	−1.7484	−2.1485	−2.2626	−2.0778
c_2	2.0729	1.8516	2.1183	2.3379	2.0016
c_3	−1.1719	−1.3840	−1.2185	−1.4044	−1.1296
c_4	0.4176	0.7331	0.4440	0.5309	0.4104
c_5	−0.0948	−0.2552	−0.1036	−0.1270	−0.0985
c_6	0.0133	0.0560	0.0151	0.0185	0.0162
c_7	−0.0010	−0.0071	−0.0013	−0.0015	−0.0021
$c_8 (\times 10^{-3})$	0.0464	0.6397	0.0990	0.1061	0.2949
c_9	−1.9148	−1.5150	0.9473	0.9293	1.8894
c_{10}	−2.0112	−1.5179	0.8899	0.8545	1.1611
c_{11}	−1.3216	−0.8523	0.9887	0.9411	0.6570
c_{12}	0.3424	0.5942	1.4846	1.4310	0.9577
c_{13}	2.2736	2.1238	2.0835	2.0224	1.8793
c_{14}	2.1646	1.7134	1.2674	1.1950	1.5309
c_{15}	−3.1753	−2.8745	−2.6283	−2.6231	−2.6879

where $x_k \in \mathbb{R}^8$, $u_k \in \mathbb{R}$, $y_k \in \mathbb{R}$ and

$$
A_{0.25}(\theta) = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \\ -c_8(\theta) & -c_7(\theta) & \cdots & -c_1(\theta) \end{bmatrix}_{8 \times 8},
$$

$$
B_{0.25}(\theta) = \begin{bmatrix} 0 & \cdots & 0 & c_8(\theta) \end{bmatrix}_{8 \times 1}^T,
$$

$$
C_{0.25}(\theta) = \begin{bmatrix} 1 & c_{15}(\theta) & \cdots & c_9(\theta) \end{bmatrix}_{1 \times 8}
$$

(3.8)

$a(\theta) \in \mathbb{R}^4$, $b(\theta) \in \mathbb{R}^7$, $c(\theta) \in \mathbb{R}^{15}$, and $\tau(\theta)$ are the coefficients to be determined in operation points proposed in Tables 1 and 2. As it can be appreciated in (3.4), (3.5), and (3.7), both canals have been considered uncoupled, a widely common practice in literature [26].

Table 4: Model parameters obtained by identification in each operating point $OP_{k_{P_2}}$: pool P_2

Parameters	$OP_{1_{P_2}}$	$OP_{2_{P_2}}$	$OP_{3_{P_2}}$	$OP_{4_{P_2}}$	$OP_{5_{P_2}}$
τ	11	10	9	8	7
a_1	−1.9419	−1.9244	−1.9115	−1.8951	−1.8830
a_2	0.9434	0.9270	0.9153	0.9005	0.8904
a_3	0.0015	0.0026	0.0038	0.0055	0.0074
a_4	−0.0188	−0.0190	0.0075	0.0147	0.0205
b_1	−0.1515	−0.1931	−0.1361	−0.0535	−0.0244
b_2	0.0377	0.0484	0.0356	0.0178	0.0100
b_3	0.0032	0.0017	0.0069	0.0154	0.0199
$b_4(\times 10^{-3})$	0.1347	0.3226	0.2270	0.1694	0.1214
b_5	603.9000	117.29	307.20	726.79	989.96
b_6	−299.6300	−72.221	−204.48	−496.96	−734.07
b_7	51.8040	15.713	48.386	122.65	207.87
c_1	−2.0539	−1.8107	−2.1661	−2.3208	−2.3517
c_2	1.9327	1.6503	2.1372	2.4464	2.5364
c_3	−1.0584	−0.9557	−1.2253	−1.4919	−1.5938
c_4	0.3661	0.4019	0.4440	0.5688	0.6308
c_5	−0.0808	−0.1250	−0.1032	−0.1363	−0.1579
c_6	0.0111	0.0283	0.0151	0.0197	0.0241
c_7	−0.0008	−0.0042	−0.0014	−0.0016	−0.0021
$c_8(\times 10^{-3})$	0.0434	0.4807	0.1050	0.1112	0.1667
c_9	2.1585	−1.5600	−0.6518	−0.1523	1.1259
c_{10}	1.4147	−1.7388	−0.9192	−0.5187	0.0930
c_{11}	0.8165	−1.0839	−0.4348	−0.2286	−0.4165
c_{12}	1.0115	0.6744	1.1870	1.1744	0.5429
c_{13}	1.7673	2.4956	2.8789	2.7318	2.4285
c_{14}	0.9012	1.3293	1.2377	1.1205	1.6084
c_{15}	−2.2559	−2.7775	−2.8406	−2.7515	−2.8369

Parameters of models (3.4)–(3.7) in each operation points and pools are independently identified. To test the improvement of these rational order models (3.5) and (3.7) with respect to the LTI model with delay (3.4) in each pool, a parametric estimation of each model has been carried out. This estimation consists in the computation of parameter vectors $a(\theta)$ and $\tau(\theta)$ for integer model, $b(\theta)$ and $c(\theta)$ for non-integer models ($\alpha = 0.5$ and $\alpha = 0.25$, resp.).

The estimation method used in this work is the previously mentioned in Section 2.3 (see [18, 19]). This methodology guarantees robust convergence, even when the parameters are initialized with values far from the optimal value.

In the case of integer model, there exists a delay which is estimated using correlation analysis [9], providing an estimation of the canal impulse response with regard to the integral of pump activation. This method computes intervals for the delay with a given confidence, and only the nominal values are chosen.

Parameters of models obtained in both pools, P_1 and P_2, are gathered in Tables 3, 4, and 5, respectively.

Parameters are estimated experimentally by applying the set of input PRBSs, explained in Section 3.1, sweeping all the operating points in each pool (Figures 4 and 5).
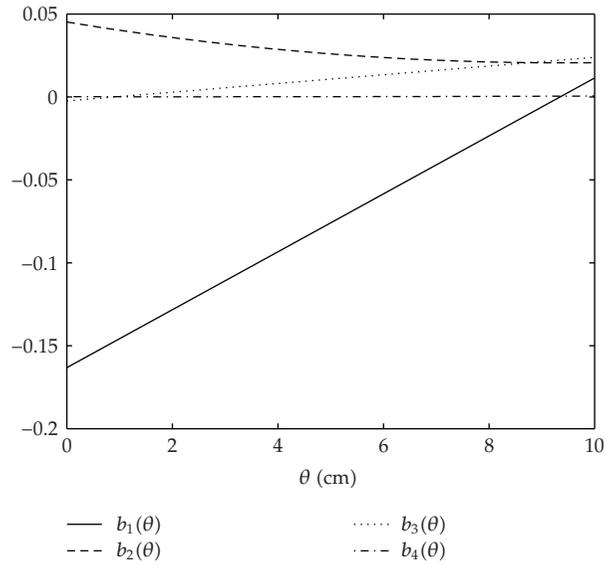
Figure 6: Polynomial aproximations of $b_1(\theta) - b_4(\theta)$ in pool P_1.
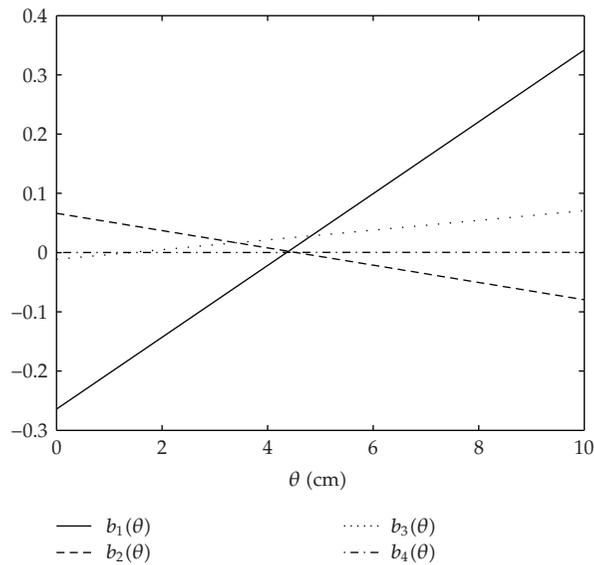
Figure 7: Polynomial aproximations of $b_1(\theta) - b_4(\theta)$ in pool P_2.

Each linear varying parameter depends on the gain scheduling variable $\theta = [u_1, u_2]$. Hence, it is assumed that the variation of parameters $a(\theta)$ and $\tau(\theta)$ for integer model and $b(\theta)$ and $c(\theta)$ for non-integer models according to the scheduling variable θ can be approximated by a polynomial function of θ, where $\theta = u_1$ with $u_2 = 0.5$ cm for pool P_1, and $\theta = u_2$ with $u_1 = 0.5$ cm for pool P_2.

Table 5: Model parameters obtained by identification in each operating Point $OP_{k_{P_2}}$: pool P_2 (cont.)

Parameters	$OP_{6_{P_2}}$	$OP_{7_{P_2}}$	$OP_{8_{P_2}}$
τ	5	4	3
a_1	−1.8826	−1.8804	−1.8697
a_2	0.8907	0.8891	0.8791
a_3	0.0081	0.0088	0.0094
a_4	0.0309	0.0447	0.0720
b_1	0.1722	0.2216	0.0919
b_2	−0.0398	−0.0470	−0.0235
b_3	0.0434	0.0509	0.0437
$b_4 (\times 10^{-3})$	0.3951	0.4421	0.2546
b_5	722.21	566.36	757.35
b_6	−515.24	−476.59	−674.80
b_7	133.11	136.39	203.67
c_1	−2.2296	−2.3769	−2.4777
c_2	2.2776	2.6077	2.8311
c_3	−1.3542	−1.6760	−1.8925
c_4	0.5203	0.6903	0.8029
c_5	−0.1375	−0.1888	−0.2194
c_6	0.0282	0.0369	0.0390
c_7	−0.0050	−0.0059	−0.0048
$c_8 (\times 10^{-3})$	0.7180	0.7576	0.4943
c_9	5.6247	3.8395	13.4920
c_{10}	2.3606	2.4912	−11.8110
c_{11}	−2.0125	−1.2561	−1.2178
c_{12}	−3.8188	−3.8690	7.8987
c_{13}	0.1248	−0.6176	−8.5361
c_{14}	5.8407	6.5926	8.5167
c_{15}	−4.1413	−4.4149	−4.4914

Table 6: Values of p for each $b_i(\theta)$: pool P_1.

Coefficients	p_1	p_2	p_3
b_1	0	0.01746	−0.1632
b_2	0.0002788	−0.005239	0.04517
b_3	0	0.002626	−0.002383
b_4	6.889×10^{-6}	-2.84×10^{-5}	0.0001364

For instance, for non-integer model $\alpha = 0.5$. Figures 6 and 7 graphically depict polynomial approximations of $b_1(\theta) - b_4(\theta)$ in both pools that correspond to the following functions:

$$b_i(\theta) = p_1\theta^2 + p_2\theta + p_3, \tag{3.9}$$

where the values of p_j $(j = 1, \ldots, 3)$ are shown in Tables 6 and 7.

Table 7: Values of p for each $b_i(\theta)$: pool P_2.

Coefficients	p_1	p_2	p_3
b_1	0	0.0606	−0.1632
b_2	0	−0.01459	0.06633
b_3	0	0.008233	−0.01152
b_4	0	2.42×10^{-5}	0.0001565

Table 8: Mean absolute error (MAE) in every operation point: pool P_1.

Operation points	$\widehat{y}_i(\alpha = 1)$	$\widehat{y}_i(\alpha = 0.5)$	$\widehat{y}_i(\alpha = 0.25)$
$OP_{1_{P_1}}$	0.0294	0.0152	0.0148
$OP_{2_{P_1}}$	0.0228	0.0145	0.0067
$OP_{3_{P_1}}$	0.0164	0.0120	0.0163
$OP_{4_{P_1}}$	0.0191	0.0168	0.0180
$OP_{5_{P_1}}$	0.0187	0.0172	0.0179

Table 9: Mean absolute error (MAE) in every operation point: pool P_2.

Operation points	$\widehat{y}_i(\alpha = 1)$	$\widehat{y}_i(\alpha = 0.5)$	$\widehat{y}_i(\alpha = 0.25)$
$OP_{1_{P_2}}$	0.0160	0.0093	0.0074
$OP_{2_{P_2}}$	0.0133	0.0095	0.0073
$OP_{3_{P_2}}$	0.0121	0.0098	0.0188
$OP_{4_{P_2}}$	0.0133	0.0123	0.0106
$OP_{5_{P_2}}$	0.0117	0.0114	0.0108
$OP_{6_{P_2}}$	0.0129	0.0124	0.0093
$OP_{7_{P_2}}$	0.0128	0.0125	0.0099
$OP_{8_{P_2}}$	0.0101	0.0097	0.0077

4. Model Validation

Model validation is the core of the identification problem because it makes possible to evaluate the model quality, that is, if the method fits the measured experimental data with accuracy enough, if it is valid for its purpose, and if the model describes correctly the real process [9]. Figures 8 and 10 show the performance in all the operation points for rational models as well as for integer model in pools P_1 and P_2, respectively. Globally, in Figures 8 –11 it can be appreciated that rational models track better measured downstream level in transitory case and also in permanent regime case than integer models.

In order to assess how suitable models respect validation data set, mean absolute error (MAE) is quantified as

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{y}_i(\alpha) - y_i\right| = \frac{1}{n}\sum_{i=1}^{n}|e_i|. \tag{4.1}$$

As its name suggests, the mean absolute error is an average of absolute errors $e_i = \widehat{y}_i(\alpha) - y_i$, where $\widehat{y}_i(\alpha)$ is the prediction value and y_i the real value. The values of MAE for operating points in each pool are shown in Tables 8 and 9, being $\widehat{y}_i(\alpha = 1)$ the integer case (3.4) and
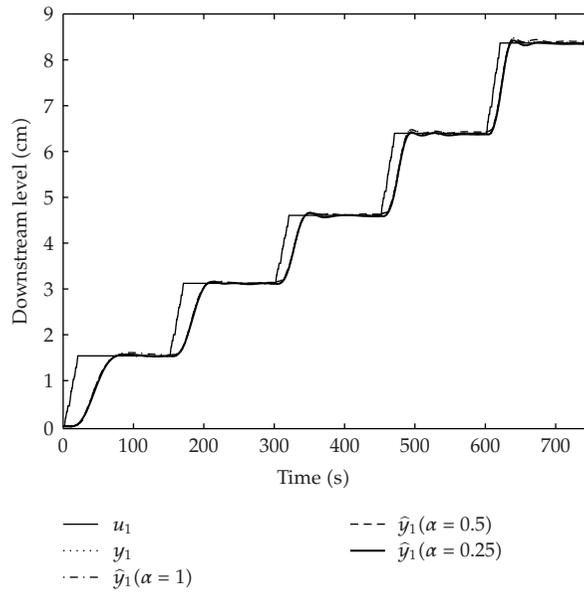
Figure 8: Model output in pool P_1.
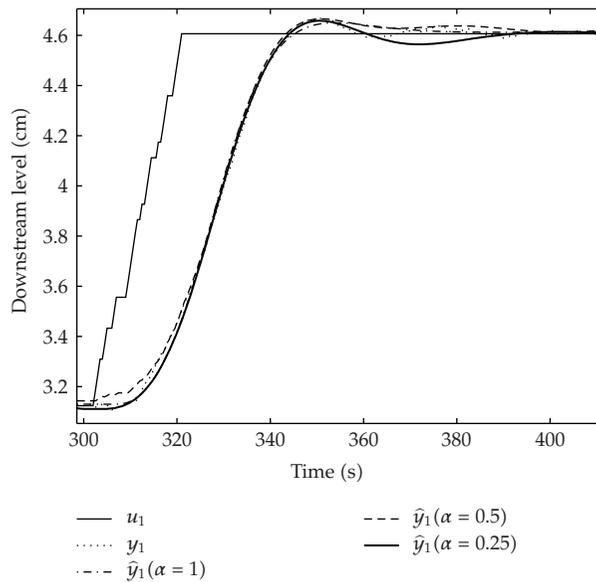
Figure 9: Model output in operation point $OP_{3_{P_1}}$ in pool P_1.

$\widehat{y}_i(\alpha = 0.5)$ and $\widehat{y}_i(\alpha = 0.25)$ the rational models (3.5) and (3.7), respectively. As it can be observed, most of errors in the integer case are higher than errors in the rational case, indicating that rational models give an improvement in the accuracy in each control model.

However, the lower the value of α is, the higher is the number of coefficients to be determined (see Tables 3, 4, and 5).
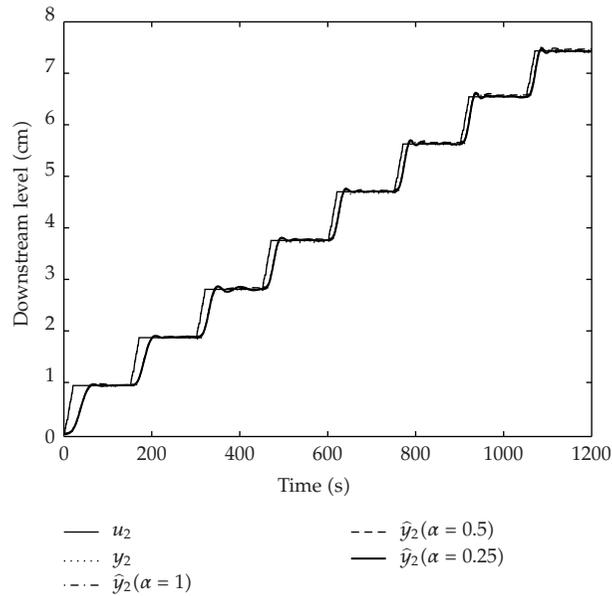
Figure 10: Model output in pool P_2.
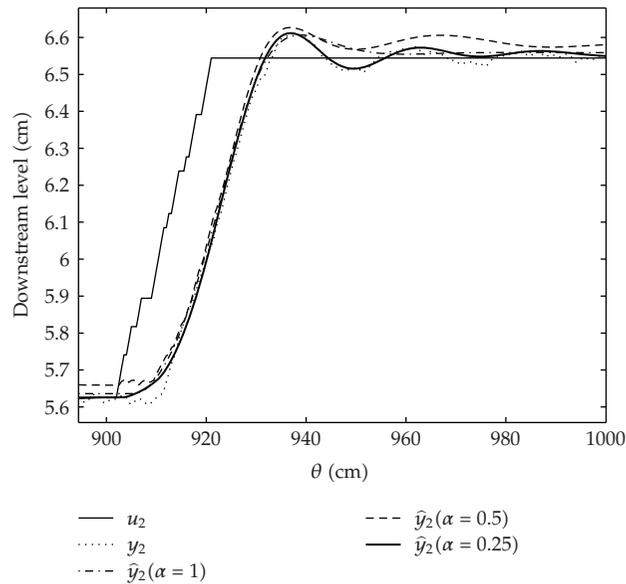
Figure 11: Model output in operation point $OP_{7_{P_2}}$ in pool P_2.

5. Conclusions

In this article, an LPV rational order model-based control-oriented system identification procedure for irrigation canals has been developed. This identification procedure has been applied in an experimental prototype canal. In this case, rational local models for an irrigation pool in different operation points have been obtained and interpolated to reach the complete

model: the LPV rational model. Resulting LPV rational order control model normally describes the plant with a lower error than the corresponding LPV integer order control model. The lower the α value (degree of the rational order models) is, the lower the error is. Nevertheless, there exists a relevant trade-off between α values and model complexity for control purposes, because the lower the α values are, the higher is the number of coefficients to be computed. This amount of data increases controller computational complexity but on the other hand controller design techniques become simpler.

Acknowledgments

References

[1] X. Litrico, "Modélisation et commande robuste de systèmes barrage-rivière," *La Houille Blanche*, no. 3-4, pp. 86–90, 2001.

[2] J. A. Cunge, F. M. Holly, and A. Verwey, *Practical Aspects of Computational River Hydraulics*, Pitman Advanced Publishing, London, UK, 1980.

[3] X. Litrico and V. Fromion, "Variations autour de la Performance Optimale de Systèmes Hydrauliques à Surface Libre," in *Colloque Automatique et Agronomie*, pp. 197–207, INRA, Montpellier, France, 2003.

[4] X. Litrico and V. Fromion, "Analytical approximation of open-channel flow for controller design," *Applied Mathematical Modelling*, vol. 28, no. 7, pp. 677–695, 2004.

[5] X. Litrico and V. Fromion, "Frequency modeling of open-channel flow," *Journal of Hydraulic Engineering*, vol. 130, no. 8, pp. 806–815, 2004.

[6] Y. Bolea, J. Blesa, and V. Puig, "LPV modelling and identification of an open canal for control," in *Proceedings of the European Control Conference (ECC '07)*, Kos, Greece, 2007.

[7] S. K. Ooi, M. P. M. Krutzen, and E. Weyer, "On physical and data driven modelling of irrigation channels," *Control Engineering Practice*, vol. 13, no. 4, pp. 461–471, 2005.

[8] E. Weyer, "System identification of an open water channel," *Control Engineering Practice*, vol. 9, no. 12, pp. 1289–1299, 2001.

[9] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1987.

[10] D. J. Leith and W. E. Leithead, "Survey of gain-scheduling analysis and design," *International Journal of Control*, vol. 73, no. 11, pp. 1001–1025, 2000.

[11] W. J. Rugh and J. S. Shamma, "Research on gain scheduling," *Automatica*, vol. 36, no. 10, pp. 1401–1425, 2000.

[12] D. J. Leith, R. N. Shorten, W. E. Leithead, O. Mason, and P. Curran, "Issues in the design of switched linear control systems: a benchmark study," *International Journal of Adaptive Control and Signal Processing*, vol. 17, no. 2, pp. 103–118, 2003.

[13] X. Litrico and D. Georges, "Robust continuous-time and discrete-time flow control of a dam-river system. (II) Controller design," *Applied Mathematical Modelling*, vol. 23, no. 11, pp. 829–846, 1999.

[14] J. Abonyi and R. Babuska, "Local and global identification and interpretation of parameters in Takagi-Sugeno fuzzy models," in *Proceedings of the IEEE International Conference on Fuzzy Systems*, vol. 2, pp. 835–840, San Antonio, Tex, USA, 2000.

[15] B. Bamieh and L. Giarré, "Identification of linear parameter varying models," *International Journal of Robust and Nonlinear Control*, vol. 12, no. 9, pp. 841–853, 2002.

[16] V. Feliu, R. Rivas, and F. J. Castillo, "Fractional robust control to delay changes in main irrigation canals," in *Proceedings of the 16th IFAC World Congress*, Prague, Czech Republic, 2005.

[17] I. Podlubny, *Fractional Differential Equations*, vol. 198 of *Mathematics in Science and Engineering*, Academic Press, San Diego, Calif, USA, 1999.

[18] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM Journal on Applied Mathematics*, vol. 11, pp. 431–441, 1963.

[19] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numerical Analysis*, G. A. Watson, Ed., vol. 630 of *Lecture Notes in Mathematics*, pp. 105–116, Springer, Berlin, Germany, 1978.

[20] R. Malti, S. Victor, O. Nicolas, and A. Oustaloup, "System identification using fractional models: state of the art," in *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (DETC '07)*, pp. 295–304, Las Vegas, Nev, USA, 2007.

[21] Y. Bolea, V. Puig, J. Blesa, M. Gómez, and J. Rodellar, "An LPV model for canal control," in *Proceedings of the 10th IEEE International Conference on Methods and Models in Automation and Robotics (MMAR '04)*, Miedzyzdroje, Poland, 2004.

[22] E. F. Camacho, *Model Predictive Control*, Springer, New York, NY, USA, 1999.

[23] J. Schuurmans, *Control of water levels in open-channels*, Ph.D. thesis, Delft University of Technology, Delft, The Nederlands, 1997.

[24] V. T. Chou, *Open Channel Hydraulics*, MacGraw-Hill, New York, NY, USA, 1959.

[25] J. Schuurmans, O. H. Bosgra, and R. Brouwer, "Open-channel flow model approximation for controller design," *Applied Mathematical Modelling*, vol. 19, no. 9, pp. 525–530, 1995.

[26] X. Litrico, V. Fromion, and J.-P. Baume, "Tuning of robuts distant downstream PI controllers for an irrigation canal pool: (II) implementation issues," *Journal of Irrigation and Drainage Engineering*, vol. 132, no. 4, pp. 369–379, 2006.

[27] R. S. Sánchez-Peña, Y. Bolea, and V. Puig, "MIMO Smith predictor: global and structured robust performance analysis," *Journal of Process Control*, vol. 19, no. 1, pp. 163–177, 2009.

[28] Y. Bolea, V. Puig, and J. Blesa, "Gain-scheduled smith PID controllers for LPV systems with time varying delay: application to an open-flow canal," in *Proceedings of the 17th IFAC World Congress*, Seoul, South Korea, 2008.

Research Article

Flux Reflection Model of the Ferroresonant Circuit

Kruno Miličević, Ivan Flegar, and Denis Pelin

Faculty of Electrical Engineering, University of Osijek, Kneza Trpimira 2b, 31000 Osijek, Croatia

Correspondence should be addressed to Kruno Miličević, kruno.milicevic@etfos.hr

The paper presents a linear model of ferroresonant circuit with flux reflection. The proposed model—flux reflection model—derives from observations of typical flux waveforms of nonlinear coil during ferroresonant steady states. Simulation results of the flux reflection model are compared with simulation results of the usual nonlinear model as well as with measurements carried out on the physical model of the ferroresonant circuit. The flux reflection model enables a novel comprehension of the ferroresonant circuit behavior and simplifies the modeling of the nonlinear coil in the ferroresonant circuit.

1. Introduction

A ferroresonant circuit is a circuit which comprises a linear capacitor in series with a nonlinear coil, driven by a sine-wave voltage. The ferroresonant circuit can exhibit several bifurcations and a chaotic steady state [1–5]. Common research procedures comprise the measurements of ferroresonant circuit and the computer simulation of the mathematical model based on state-equations of the circuit.

The nonlinearity of the model caused by the nonlinear magnetization characteristic of the coil is the driving force on the route to the chaotic steady state, which is marked by the bifurcations that are initiated by varying values of a circuit parameter. Complex models, that comprehend more nonlinear elements, are needed to match results of measurements and simulation completely [6, 7]. In order to obtain the complex behavior of the circuit, the nonlinear element that comprises the magnetization nonlinearity is only necessary. Therefore, the model comprises the magnetization nonlinearity $i_L(\varphi)$, implying the core saturation of the coil. The winding resistance of the coil is neglected, and the resistor, that represents the core losses of the coil, is assumed to be linear. State equations of the ferroresonant circuit, Figure 1, are

$$\frac{d\varphi}{dt} = u_L = \widehat{U}\sin\omega t - u_C,$$
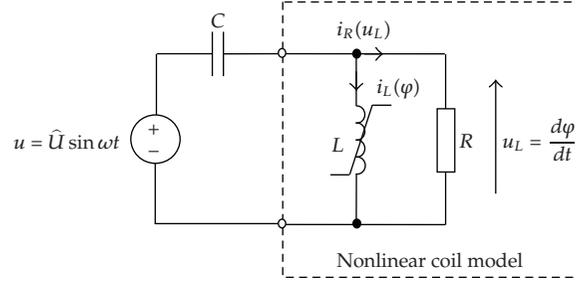
Figure 1: Ferroresonant circuit.

$$\frac{du_C}{dt} = \frac{1}{C}\left[\frac{1}{R}\left(\widehat{U}\sin\omega t - u_C\right) + i_L(\varphi)\right].$$

(1.1)

The magnetization nonlinearity $i_L(\varphi)$ is odd-symmetric and monotonically increasing. It can be presented in various forms, for example, in a polynomial form [8, 9] or piecewise linear form [10, 11]. Thereby, the polynomial form is obtained often by an interpolation of a piecewise linear form. The magnetization nonlinearity is based on measurements carried out on a nonlinear coil regarding the above-mentioned assumptions made about properties of the coil. For example, one ought to rely on the results of standard measurements carried out on the nonlinear coil by the manufacturer [11, 12]. Because of a limited thermal endurance of the coil, measurements cannot be carried out by applying the sine-wave voltage for the full range of flux and voltage peak values which occur in some operation modes of ferroresonant circuit. However, the importance of this problem is often ignored by a simple extrapolation of magnetization nonlinearity obtained by measurements.

Further problem of modeling based on standard measurements is that the parameters of model depend on the number of measurements, that is, on the number of measured values on U-I characteristic. Namely, a piecewise linear form of the magnetization nonlinearity, that is obtained by a small number of measurements, would not be as smooth as necessary; in the case of polynomial form of the magnetization nonlinearity, a small number of measurements could result with a significant interpolation error.

In the paper a novel kind of modeling the magnetization characteristic of the coil in the ferroresonant circuit is presented. The modeling will be based on a characteristic behavior that nonlinear coil exhibits in a ferroresonant circuit. The behavior will be identified from the flux waveforms obtained by measurements and by computer simulation.

The preliminary purpose of obtained novel model of the ferroresonant circuit is to enable a new comprehension of the circuit behavior and to simplify the modeling of nonlinear coil in the ferroresonant circuit by reducing the number of model parameters.

2. Measurements and Nonlinear Model

In order to notice the characteristic behavior of a coil in ferroresonant circuit, the ferroresonant steady states are obtained by measurements carried out on a ferroresonant circuit that is realized in laboratory and by computer simulation carried out on a typical nonlinear model of the ferroresonant circuit.

The ferroresonant circuit realized in the laboratory is composed of the capacitor $C = 20\,\mu F$ and the primary winding of the toroidal iron-cored two-windings transformer used as a nonlinear coil. The transformer was designed for the nominal apparent power of $200\,VA$ and for the nominal primary voltage of $30\,V$. The core is strip-wounded, made of Ni-Fe alloy (Trafoperm N3). The autotransformer of $10\,kVA$ nominal apparent power is used as a variable voltage source in all experiments.

To simplify calculations, throughout the rest of the paper all the variables and parameters of the ferroresonant circuit are expressed in relation to reference quantities, that is, in a per-unit system:

$$\overline{\varphi} = \varphi \frac{\omega}{U_{\text{ref}}}, \qquad \overline{u}_C = \frac{u_C}{U_{\text{ref}}}, \qquad \overline{\omega} = \frac{\omega}{\omega_{\text{ref}}}, \qquad \overline{i}_L = \frac{i_L}{I_{\text{ref}}}, \qquad \widehat{\overline{U}} = \frac{\widehat{U}}{U_{\text{ref}}}, \tag{2.1}$$

$$U_{\text{ref}} = 31.2\,\text{V}, \qquad I_{\text{ref}} = 0.19\,\text{A}, \qquad \omega_{\text{ref}} = 314\,\text{s}^{-1}.$$

The simulation is carried out using a typical nonlinear model of ferroresonant circuit, with the polynomial form of magnetization nonlinearity:

$$\frac{d\overline{\varphi}}{dt} = \overline{u}_L = \widehat{\overline{U}} \sin \overline{\omega} t - \overline{u}_C,$$

$$\frac{d\overline{u}_C}{dt} = \frac{1}{\overline{C}} \left[\frac{1}{\overline{R}} \left(\widehat{\overline{U}} \sin \overline{\omega} t - \overline{u}_C \right) + i_L(\overline{\varphi}) \right], \tag{2.2a}$$

$$\overline{\omega} = 1, \qquad \overline{C} = 1,$$

$$\overline{i}_L(\overline{\varphi}) = f(\overline{\varphi}) \operatorname{sign}(\overline{\varphi}),$$

$$f(\overline{\varphi}) = \sqrt{0.034 \cdot \overline{\varphi}^2 + 5.54 \cdot 10^{-3} \cdot \overline{\varphi}^{20} + 1.05 \cdot 10^{-5} \cdot \overline{\varphi}^{38}}, \tag{2.2b}$$

$$\overline{R} = 2.$$

The magnetization characteristic $\overline{i}_L(\overline{\varphi})$ and the iron-core losses $\overline{R}$ are derived from measured P-U and U-I characteristic of the nonlinear coil [12].

During measurements and simulation the steady states are obtained by varying the amplitude $\widehat{\overline{U}}$ of source voltage within the range:

$$0 < \widehat{\overline{U}} \leq 3. \tag{2.3}$$

The step of the amplitude variation is $\Delta\widehat{\overline{U}} = 0.05$ in simulation, as well as in measurements. Simulation is carried out by using the 4th order Runge-Kutta method with 10000 integration steps per period, $j = 1, \ldots, 10000$.

Characteristic steady states and bifurcations obtained by simulation and measurements are shown in Tables 1 and 2, respectively. There is no significant disagreement between results of simulation and measurements with the exception of the pitchfork bifurcation and the period-four steady state, which are not noticed in measurements.

Table 1: Steady states and *bifurcations* of the nonlinear model of ferroresonant circuit obtained by simulation.

Increasing $\widehat{\overline{U}}$	Decreasing $\widehat{\overline{U}}$	Steady states and bifurcations
$0 < \widehat{\overline{U}} < 1.05$	$0 < \widehat{\overline{U}} < 0.75$	Monoharmonic steady state
$\widehat{\overline{U}} = 1.05$	—	*Forward ferroresonant jump*
—	$\widehat{\overline{U}} = 0.75$	*Reverse ferroresonant jump*
$1.05 \le \widehat{\overline{U}} < 1.7$	$0.75 \le \widehat{\overline{U}} < 1.7$	Odd higher harmonic steady state
	$\widehat{\overline{U}} = 1.7$	*Pitchfork bifurcation*
	$1.7 \le \widehat{\overline{U}} < 2.4$	Even and odd higher harmonic steady state
	$\widehat{\overline{U}} = 2.4$	*Period-doubling bifurcation*
	$2.4 \le \widehat{\overline{U}} < 2.55$	Period-two steady state
	$\widehat{\overline{U}} = 2.55$	*Period-doubling bifurcation*
	$2.55 \le \widehat{\overline{U}} < 2.65$	Period-four steady state
	$2.65 < \widehat{\overline{U}} < 3$	Chaotic steady state

Table 2: Steady states and *bifurcations* of the ferroresonant circuit realized in the laboratory obtained by measurements.

Increasing $\widehat{\overline{U}}$	Decreasing $\widehat{\overline{U}}$	Steady states and bifurcations
$0 < \widehat{\overline{U}} < 1$	$0 < \widehat{\overline{U}} < 0.7$	Monoharmonic steady state
$\widehat{\overline{U}} = 1$	—	*Forward ferroresonant jump*
—	$\widehat{\overline{U}} = 0.7$	*Reverse ferroresonant jump*
$1 \le \widehat{\overline{U}} < 1.4$	$0.7 \le \widehat{\overline{U}} < 1.4$	Odd higher harmonic steady state
	$\widehat{\overline{U}} = 1.4$	*Pitchfork bifurcation*
	$1.7 \le \widehat{\overline{U}} < 2.2$	Even and odd higher harmonic steady state
	$\widehat{\overline{U}} = 2.2$	*Period-doubling bifurcation*
	$2.2 \le \widehat{\overline{U}} < 2.6$	Period-two steady state
		Period-doubling bifurcation
Not obtained		Period-four steady state
	$2.6 < \widehat{\overline{U}} < 3$	Chaotic steady state

By that means, Figure 3 implies that the peak value of flux $\widehat{\overline{\Phi}}$ is in all polyharmonic steady states limited by the value $\overline{\Phi}_{\text{sat}} \approx 1.5$, called here as a flux saturation value:

$$|\varphi| \le \overline{\Phi}_{\text{sat}}. \tag{2.4}$$

From simulation as well as measurements it is observed that during all steady states the slope of flux, immediately before and after the moment each time the flux reaches the peak value $\widehat{\overline{\Phi}}$, changes the sign only, that is, holding the same absolute value nearly. As it is seen in Figure 2, for instance, at the moment t_1, we assume

$$\left.\frac{d\overline{\varphi}}{dt}\right|_{t_1-0} \approx -\left.\frac{d\overline{\varphi}}{dt}\right|_{t_1+0}. \tag{2.5}$$

This property is named here as a flux reflection.
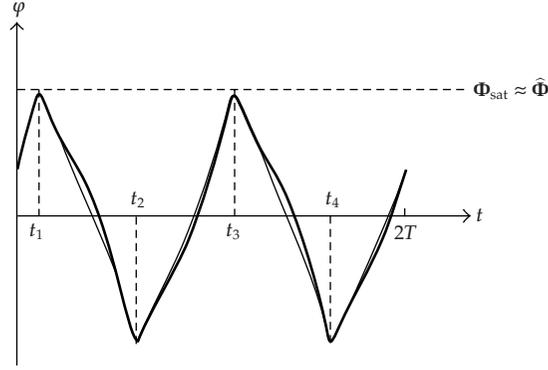
Figure 2: Typical ferroresonant steady state waveform of flux obtained by measurements (bold line) and simulation (thin line) for source voltage amplitude $\widehat{\overline{U}} = 1.1$.


(a)

(b)
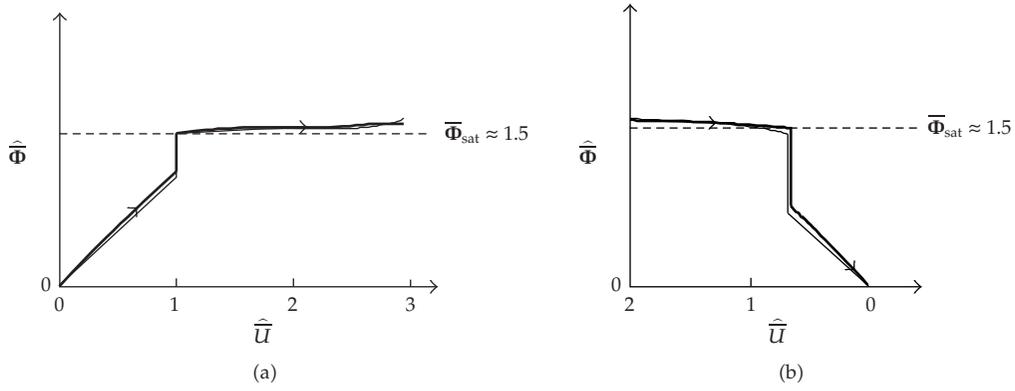
Figure 3: Peak values of flux obtained by measurements (bold line) and by computer simulation (thin line): (a) by increasing the source voltage amplitude $\widehat{\overline{U}}$, (b) by decreasing the source voltage amplitude $\widehat{\overline{U}}$.

3. Flux Reflection Model

Novel presentation of magnetization nonlinearity is based on observed limitation (2.4) and reflection (2.5) of flux values. Thereby, the magnetization characteristic for values $|\varphi| < \overline{\Phi}_{\text{sat}}$ is approximated by a linear inductance $\bar{i}_L(\overline{\varphi}) \approx (1/\overline{L})\overline{\varphi}$. In a case of the polynomial form of magnetization nonlinearity (2.2b) the linear inductance $\overline{L} = 5.4$ is approximated as shown on Figure 4.

By the reducing of the nonlinear magnetization characteristic to the linear inductance $\overline{L}$, the state-equations of the ferroresonant circuit, (2.2a) and (2.2b), become linear between two flux reflections, that is, for flux values $|\varphi| < \overline{\Phi}_{\text{sat}}$:

$$\frac{d\overline{\varphi}}{dt} = \overline{u}_L = \widehat{\overline{U}} \sin \overline{\omega} t - \overline{u}_C,$$

$$\frac{d\overline{u}_C}{dt} = \frac{1}{\overline{C}} \left[\frac{1}{\overline{R}} \left(\widehat{\overline{U}} \sin \overline{\omega} t - \overline{u}_C \right) + \frac{1}{\overline{L}} \overline{\varphi} \right],$$

$$\overline{\omega} = 1, \quad \overline{C} = 1, \quad \overline{R} = 2, \quad \overline{L} = 5.4.$$

(3.1)
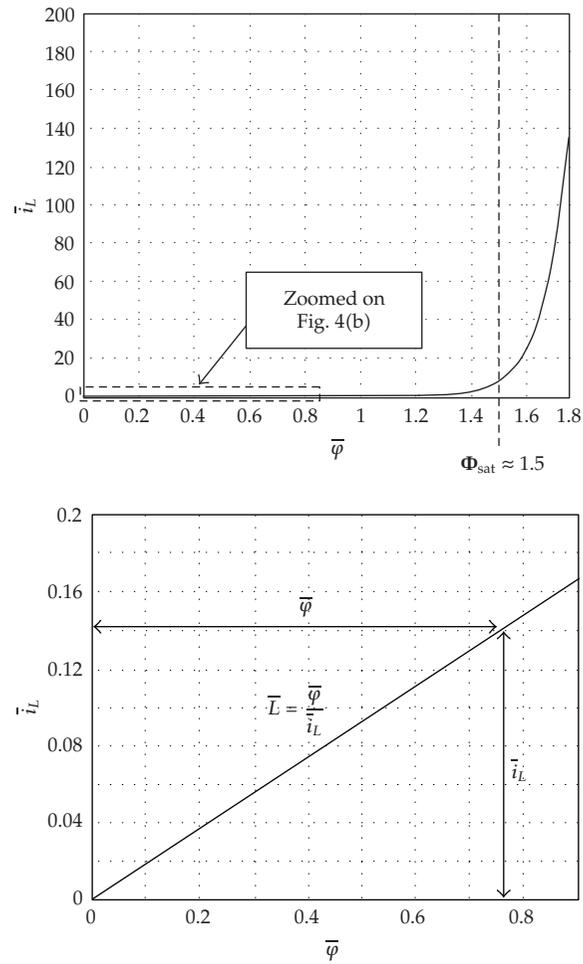
Figure 4: (a) Polynomial form of the magnetization nonlinearity. (b) Zoomed linear part of the magnetization nonlinearity.
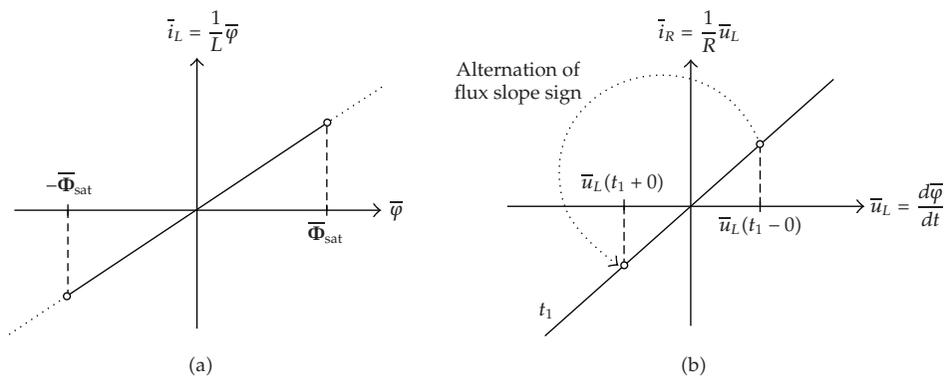

Figure 5: Magnetization characteristic $\bar{i}_L(\overline{\varphi})$ of the flux reflection model and the alternation of flux slope sign shown on $\bar{i}_R(\overline{u}_L)$ characteristic.
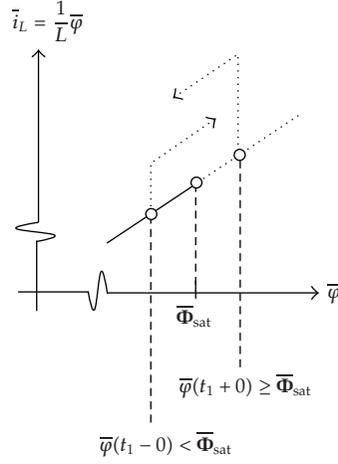
Figure 6: Part of the magnetization characteristic zoomed near the saturation value $\overline{\Phi}_{\text{sat}}$.

Thereby, the observed properties of a ferroresonant circuit—the limitation of flux values (2.4) and the flux reflection (2.5)—can be realized easily, using the following pseudocode:

$$
\begin{aligned}
&FOR\ each\ integration\ step\ j \\
&\quad calculate\ \left.\frac{d\overline{\varphi}}{dt}\right|_{j+1}\ as\ defined\ by\ used\ iterative\ method \\
&\quad IF\ \left|\overline{\varphi}(j)\right| > \overline{\Phi}_{\text{sat}}\ THEN \\
&\quad\quad set\ \left.\frac{d\overline{\varphi}}{dt}\right|_{j+1}\ to\ be\ equal\ to\ -\left.\frac{d\overline{\varphi}}{dt}\right|_{j} \\
&\quad END\ IF \\
&END\ FOR
\end{aligned}
\tag{3.2}
$$

The proposed flux reflection model, defined by state-equations (3.1) and pseudocode (3.2), could be comprehended as a piecewise linear model also. However, contrary to a typical piecewise linear model, this model comprises only one set of state-equations, that is, the magnetization characteristic $\overline{i}_L(\overline{\varphi})$ with only one linear segment. As it is shown on Figures 5 and 6 and determined by the pseudocode (3.2), the segment is limited by the saturation value $\overline{\Phi}_{\text{sat}}$ which triggers the alternation of flux slope sign and, at this way, causes the change of flux trend. For instance, the open arrows on Figure 6 indicate the increasing of flux value before the flux reflection and the decreasing of flux values after the flux reflection. Figures 5 and 6 show the flux reflection for the case $\overline{\varphi} \geq \overline{\Phi}_{\text{sat}}$ only. The depiction of the flux reflection for the case $\overline{\varphi} \leq -\overline{\Phi}_{\text{sat}}$ can be derived from the mirror symmetry of flux reflection model easily.

Table 3: Steady states and *bifurcations* of the flux reflection model of ferroresonant circuit.

Increasing $\widehat{\overline{U}}$	Decreasing $\widehat{\overline{U}}$	Steady states and bifurcations
$0 < \widehat{\overline{U}} < 1.15$	$0 < \widehat{\overline{U}} < 0.8$	Monoharmonic steady state, Figure 7(a)
$\widehat{\overline{U}} = 1.15$	—	*Forward ferroresonant jump*
—	$\widehat{\overline{U}} = 0.8$	*Reverse ferroresonant jump*
$1.15 \leq \widehat{\overline{U}} < 1.7$	$0.8 \leq \widehat{\overline{U}} < 1.7$	Odd higher harmonic steady state, Figure 7(b)
	$\widehat{\overline{U}} = 1.7$	*Pitchfork bifurcation*
	$1.7 \leq \widehat{\overline{U}} < 2.25$	Even and odd higher harmonic steady state, Figure 7(c)
		Period-doubling bifurcation
		Period-two steady state
Not obtained		*Period-doubling bifurcation*
		Period-four steady state
	$2.25 \leq \widehat{\overline{U}} \leq 3$	Chaotic steady state, Figure 7(d)

4. Comparison of the Flux Reflection Model and the Nonlinear Model

The flux reflection model can be considered as a model of the ferroresonant circuit only if the flux reflection model reveals a complex behavior of the ferroresonant circuit. Therefore, the steady states of the flux reflection model are obtained in the same way as for the ferroresonant circuit realized in laboratory and for the nonlinear model in Section 2, that is, by varying the amplitude $\widehat{\overline{U}}$ with the step of the amplitude variation $\Delta\widehat{\overline{U}} = 0.05$ within a range:

$$0 < \widehat{\overline{U}} \leq 3. \tag{4.1}$$

Between two flux reflections the flux reflection model is described by linear state equations (3.1). Thus, these equations can be solved analytically. However, as it is shown in the appendix in detail, at the start of transient state, that is, in moment $t = 0$, it is not possible to determine explicitly the ranges of parameter values and values of initial conditions at which a particular steady state would arise; it is only possible to solve (3.1) for a given set of parameter values and values of initial conditions.

In the paper the steady states solutions are obtained in the same way as for the nonlinear model in Section 2, that is, by simulation that is carried out numerically, using the 4th order Runge-Kutta method with 10000 integration steps per period, $j = 1, \ldots, 10000$.

The route to the chaotic steady state with bifurcations and corresponding steady states is obtained by varying the source voltage amplitude $\widehat{\overline{U}}$, (3.1), as shown on Table 3. Figure 7 shows the flux waveforms corresponding to each of steady states of the flux reflection model shown in Table 3.

The results of simulation based on the nonlinear model and results of measurements, shown on Tables 1 and 2, differ from results of simulation based on the flux reflection model shown on Table 3. For instance, the period-two steady state as well as period-four steady state and corresponding period-doubling bifurcations, are not obtained by simulation based on flux reflection model. The reason for this could be the approximation of the maximum value of flux by a constant saturation value $\overline{\Phi}_{sat}$. Namely, on Figure 3 during the polyharmonic steady states the maximum value of flux is constant only nearly, that is, it is not equal to the chosen saturation value ($\overline{\Phi}_{sat} = 1.5$).
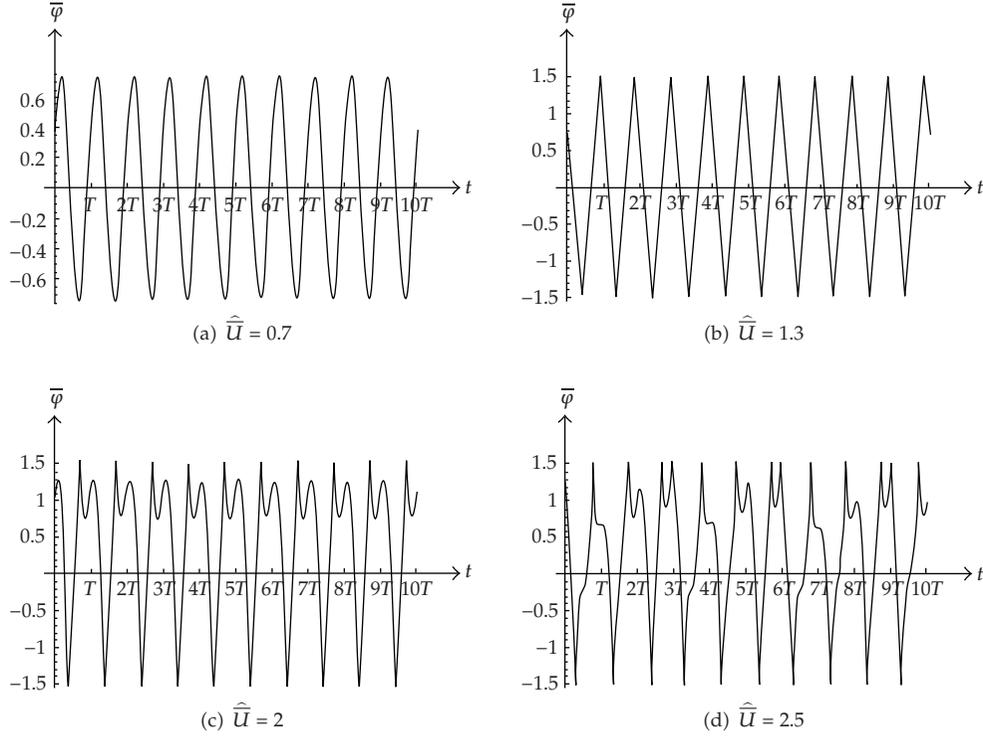
Figure 7: Flux waveforms of the flux reflection model: (a) monoharmonic steady state, (b) odd higher harmonic steady state, (c) even and odd higher harmonic steady state, and (d) chaotic steady state.

The main shortcoming of the flux reflection model is the inherent limitation of inductive component of the coil current $\bar{i}_L$

$$\left| \bar{i}_L(t) \right| \leq \left| \bar{i}_L \left[\overline{\varphi}(t_1 + 0) \right] \right| \approx \left| \bar{i}_L \left(\overline{\Phi}_{sat} \right) \right| = \left| \frac{1}{L} \overline{\Phi}_{sat} \right| \tag{4.2}$$

which does not exist in the actual ferroresonant circuit.

In spite of that, the proposed model reveals a complex steady state behavior of the ferroresonant circuit. Besides of a new comprehension of the circuit behavior, the flux reflection model enables a simple modeling of the nonlinear coil. It is sufficient to carry out the simple measurements to determine the value of linear inductance L and the saturation value Φ_{sat}.

5. Additional Simulation

The additional simulation is carried out in order to investigate the sensitivity of flux reflection model on the value of linear inductance L and on the saturation value Φ_{sat}. The step of the source voltage amplitude variation is $\Delta \widehat{\overline{U}} = 0.05$. Simulation is carried out by using the 4th order Runge-Kutta method with 10000 integration steps per period.
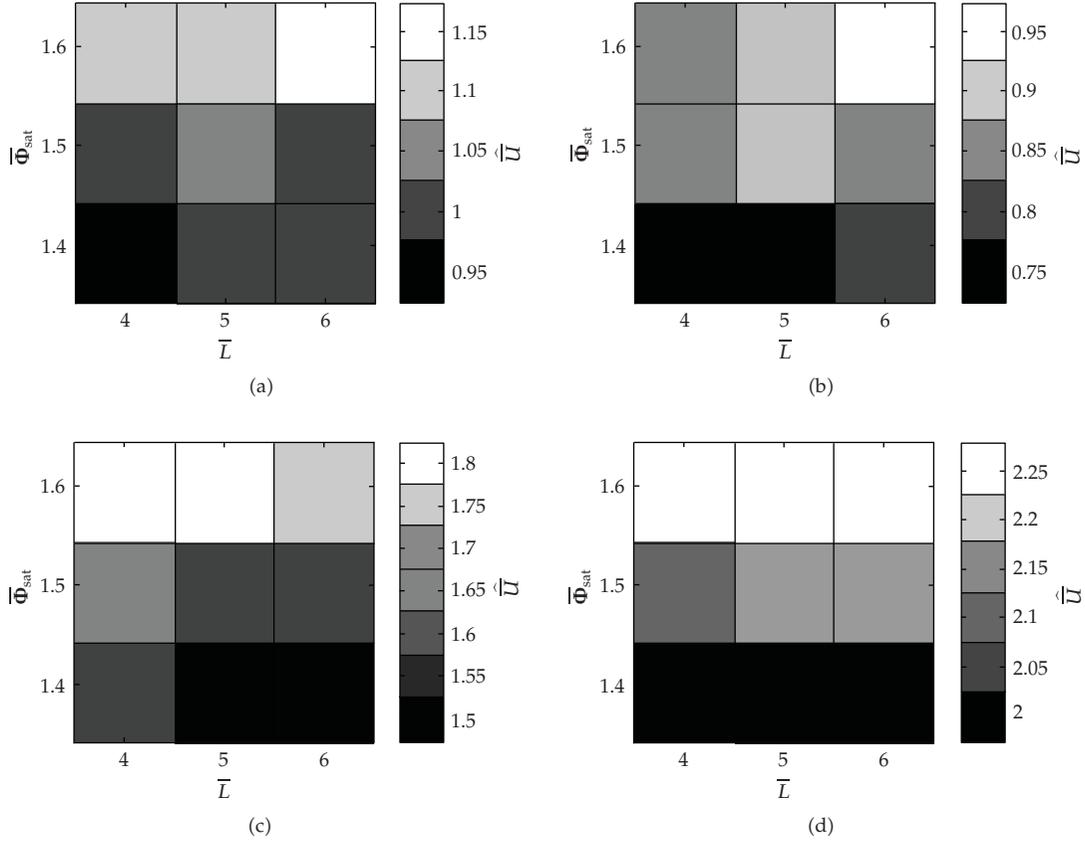
Figure 8: Influence of saturation value Φ_{sat} and linear inductance L on the initiation of (a) forward ferroresonant jump, (b) reverse ferroresonant jump, (c) pitchfork bifurcation, and (d) chaotic steady state.

Figure 8 shows results of simulation. For easier visualisation the diagrams are constructed using gray scaled squares where different shades of gray are employed to represent values of source voltage amplitude $\widehat{U}$ at which the particular bifurcation and steady state, respectively, has occurred. For instance, the bottom left square on Figure 8(c) (■) indicates that the pitchfork bifurcation for the parameter values $\overline{\Phi}_{sat} = 1.4$, $\overline{L} = 4$, occur at the source voltage amplitude $\widehat{U} = 1.55$.

The results of simulation reveal that the flux reflection model is more sensitive to the variation of the saturation value Φ_{sat}, than to the value of linear inductance L. For instance, for the saturation value $\overline{\Phi}_{sat} = 1.4$ and linear inductance values $\overline{L} = 4, 5, 6$, the reverse ferroresonant jump occurs at the source voltage amplitude values $\widehat{U} = 0.85$ (■) and $\widehat{U} = 0.90$ (▢), as it is presented by middle row on Figure 8(b). For the linear inductance value $\overline{L} = 5$ and the saturation values $\overline{\Phi}_{sat} = 1.4; 1.5; 1.6$, the reverse ferroresonant jump occurs for a wider range of source voltage amplitude values, $\widehat{U} = 0.75$ (■) and $\widehat{U} = 0.90$ (▢), as it is presented by middle column on Figure 8(b).

Therefore, it is more important to know the saturation value Φ_{sat} precisely than the value of linear inductance L.

6. Conclusions

The flux reflection model preserves the basic properties of the magnetization characteristic: linearity and saturation. Despite the noticed shortcomings, the flux reflection model reveals a complex steady state behavior characteristic for the ferroresonant circuit. It enables a new comprehension of the behavior of the ferroresonant circuit and simplifies the modeling of the nonlinear coil in the ferroresonant circuit.

Future work will address the simplification of the identification of parameters of a flux reflection model in order to avoid the identification by using the waveforms of flux values during ferroresonant steady states. Namely, in a flux reflection model a coil is defined by a linear inductance value and by a saturation value, which could be determined by standard measurements experimentally or by number of windings and by core parameters of coil analytically.

Furthermore, in order to expand the purpose of the model from the presented new comprehension of the circuit behavior to a wide application of the model, it will be necessary to identify advantages and disadvantages of the flux reflection model precisely by a more detailed comparison with usual models of ferroresonant circuit.

Appendix

On Analytical Conditions for Steady states

State-equations (3.1) can be expressed as a second-order linear differential equation:

$$\frac{d^2\varphi}{dt^2} + 2\alpha\frac{d\varphi}{dt} + \omega_0^2\varphi = \omega\widehat{U}\cos\overline{\omega}t, \quad \alpha = \frac{1}{2RC}, \quad \omega_0 = \frac{1}{\sqrt{LC}}. \tag{A.1}$$

If $\alpha < \omega_0$, as it is the case for parameter values used in the paper $\overline{C} = 1, \overline{R} = 2, \overline{L} = 5.4$, the solution of the equation can be written as:

$$\varphi(t) = Ke^{-\alpha t}\cos(\omega_d t + \theta) + \widehat{\Phi}\cos(\omega t - \psi),$$

$$\widehat{\Phi} = \omega\widehat{U}\frac{1}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\alpha^2\omega^2}}, \quad \psi = \text{arctg}\frac{2\alpha\omega}{\omega_0^2 - \omega^2}, \quad \omega_d = \sqrt{\omega_0^2 - \alpha^2}. \tag{A.2}$$

The constants K and θ are determined by values of initial conditions, $\varphi(0)$ and $d\varphi/dt|_0$. Equation, (A.1), as well as its solution, (A.2), is valid for the flux reflection model only between the reflections, that is, for flux values $|\varphi| < \Phi_{\text{sat}}$. In the moment t_1 the flux φ reaches the saturation value $\pm\Phi_{\text{sat}}$ and the flux slope $d\varphi/dt$ is alternated, (2.5). After the moment of reflection, the solution, (A.2), is valid again, but through the alternation of flux slope $d\varphi/dt$, the initial conditions are changed and, consequently, the constants of the solution.

Let us denote the solution before the first reflection in moment t_1 as

$$\varphi_1(t) = K_1 e^{-\alpha t}\cos(\omega_d t + \theta_1) + \widehat{\Phi}\cos(\omega t - \psi), \quad t_1 > t > 0. \tag{A.3}$$

Table 4: Properties of steady states.

Property (a)	Property (b)	Property (c)	Steady states
$k = 1$	Yes	Yes	Monoharmonic steady state
$k = 1$	Yes	No	Odd higher harmonic steady state
$k = 1$	No	No	Even and odd higher harmonic steady state
$k > 1, k \in N$	No	No	Period-k steady state
$k \to \infty$	No	No	Chaotic steady state

The constants K_1 and θ_1 are determined by given values of initial conditions, $\varphi_1(0)$ and $d\varphi_1/dt|_0$. Generally, the solution after n reflections in moments $t_1, t_2, \ldots, t_{n-1}, t_n$ and before the reflection in moment t_{n+1} is

$$\varphi_{n+1}(t) = K_{n+1}e^{-\alpha(t-t_n)}\cos\left[\omega_d(t - t_n) + \theta_{n+1}\right] + \widehat{\Phi}\cos\left[\omega(t - t_n) - \psi\right],$$

$$t_{n+1} > t > t_n > t_{n-1} > \cdots > t_2 > t_1.$$

$(A.4)$

Constants K_{n+1} and θ_{n+1} are determined using the initial conditions $d\varphi_{n+1}/dt|_{t_n}$ and $\varphi_{n+1}(t_n)$, which depend on values of solution $\varphi_n(t)$ and its derivative $d\varphi_n/dt$ in the moment of reflection t_n:

$$\varphi_{n+1}(t_n) = \varphi_n(t_n), \qquad \frac{d\varphi_{n+1}}{dt}\bigg|_{t_n} = -\frac{d\varphi_n}{dt}\bigg|_{t_n}. \qquad (A.5)$$

Each moment of reflection t_m $(m \in N, t_0 = 0)$ is determined by the condition

$$\left|\varphi_m(t_m)\right| = \Phi_{\text{sat}}$$

$$\longrightarrow \left|K_m e^{-\alpha(t_m - t_{m-1})}\cos\left[\omega_d(t_m - t_{m-1}) + \theta_m\right] + \widehat{\Phi}\cos\left[\omega(t_m - t_{m-1}) - \psi\right]\right| = \Phi_{\text{sat}}$$

$(A.6)$

which cannot be expressed explicitly regarding the moment t_m, that is, moments of reflection can be determined only approximately.

Characteristic steady state is established if the following conditions are met according to Table 4 $(m \in N, t_0 = 0)$:

(a) periodicity:

$$\varphi_m(t_{m-1}) = \varphi_{m+2k-1}(t_{m+2k-1}),$$

$$\frac{d\varphi_m}{dt}\bigg|_{t_{m-1}} = \frac{d\varphi_{m+2k-1}}{dt}\bigg|_{t_{m+2k-1}}, \qquad t_{m+2k-1} - t_{m-1} = k\frac{2\pi}{\omega},$$

$(A.7)$

(b) odd-symmetry:

$$\varphi_m(t_{m-1}) = -\varphi_{m+1}(t_m),$$

$$\left.\frac{d\varphi_m}{dt}\right|_{t_{m-1}} = -\left.\frac{d\varphi_{m+1}}{dt}\right|_{t_m}, \qquad t_m - t_{m-1} = \frac{1}{2}\frac{2\pi}{\omega}, \tag{A.8}$$

(c) peak flux value lower as saturation value:

$$\widehat{\Phi} < \Phi_{\text{sat}}. \tag{A.9}$$

However, at the start of transient state, that is, in moment $t = 0$, it is impossible to determine explicitly the ranges of parameter values and values of initial conditions, $\varphi_1(0)$ and $d\varphi_1/dt|_0$, at which a particular steady state would arise because a solution $\varphi_{n+1}(t)$, (A.4), can be obtained recursively only, starting from first solution $\varphi_1(t)$, (A.3), and using approximately determined values of reflection moments $t_1, t_2, \ldots, t_{n-1}, t_n$, (A.6), in each step of recursion.

References

[1] C. Hayashi, *Nonlinear Oscillations in Physical Systems*, McGraw-Hill Electrical and Electronic Engineering Series, McGraw-Hill, New York, NY, USA, 1964.

[2] C. Kieny, "Application of the bifurcation theory in studying and understanding the global behavior of a ferroresonant electric power circuit," *IEEE Transactions on Power Delivery*, vol. 6, no. 2, pp. 866–872, 1991.

[3] B. Lee and V. Ajjarapu, "Period-doubling route to chaos in an electrical power system," *IEE Proceedings: Generation, Transmission and Distribution*, vol. 140, no. 6, pp. 490–496, 1993.

[4] B. A. Mork and D. L. Stuehm, "Application of nonlinear dynamics and chaos to ferroresonance in distribution sysytems," *IEEE Transactions on Power Delivery*, vol. 9, no. 2, pp. 1009–1017, 1994.

[5] S. Mozaffari, M. Sameti, and A. C. Soudack, "Effect of initial conditions on chaotic ferroresonance in power transformers," *IEE Proceedings: Generation, Transmission and Distribution*, vol. 144, no. 5, pp. 456–460, 1997.

[6] M. R. Iravani, A. K. S. Chaudhary, W. J. Giesbrecht, et al., "Modeling and analysis guidelines for slow transients—part III: the study of ferroresonance," *IEEE Transactions on Power Delivery*, vol. 15, no. 1, pp. 255–265, 2000.

[7] J. A. Martinez-Velasco and B. A. Mork, "Transformer modeling for low frequency transients: the state of the art," in *Proceedings of the IPST International Conference on Power Systems Transients*, New Orleans, La, USA, June 2003.

[8] B. A. T. Al Zahawi, Z. Emin, and Y. K. Tong, "Chaos in ferroresonant wound voltage transformers: effect of core losses and universal circuit behaviour," *IEE Proceedings: Science, Measurement and Technology*, vol. 145, no. 1, pp. 39–43, 1998.

[9] D. A. N. Jacobson, P. W. Lehn, and R. W. Menzies, "Stability domain calculations of period-1 ferroresonance in a nonlinear resonant circuit," *IEEE Transactions on Power Delivery*, vol. 17, no. 3, pp. 865–871, 2002.

[10] L. O. Chua, M. Hasler, J. Neirynck, and P. Verburgh, "Dynamics of a piecewise-linear resonant circuit," *IEEE Transactions on Circuits and Systems*, vol. 29, no. 8, pp. 535–547, 1982.

[11] W. L. A. Neves and H. W. Dommel, "On modelling iron core nonlinearities," *IEEE Transactions on Power Systems*, vol. 8, no. 2, pp. 417–425, 1993.

[12] I. Flegar, D. Fischer, and D. Pelin, "Identification of chaos in a ferroresonant circuit," in *Proceedings of the International Conference on Electric Power Engineering (PowerTech '99)*, pp. 1–5, Budapest, Hungary, August-September 1999.