Algebraic Method for the Reconstruction of Partially Observed Nonlinear Systems Using Differential and Integral Embedding

Artur Karimov [1][ID], Erivelton G. Nepomuceno [2][ID], Aleksandra Tutueva [3][ID] and Denis Butusov [1,*][ID]

[1]   Youth Research Institute, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg 197376, Russia; aikarimov@etu.ru
[2]   Control and Modelling Group (GCOM), Department of Electrical Engineering, Federal University of São, João del-Rei, São João del-Rei MG 36307-352, Brazil; nepomuceno@ufsj.edu.br
[3]   Department of Computer Aided Design, Saint Petersburg Electrotechnical University "LETI", Saint Petersburg 197376, Russia; avtutueva@etu.ru
*   Correspondence: dnbutusov@etu.ru; Tel.: +7-950-008-7190

Abstract: The identification of partially observed continuous nonlinear systems from noisy and incomplete data series is an actual problem in many branches of science, for example, biology, chemistry, physics, and others. Two stages are needed to reconstruct a partially observed dynamical system. First, one should reconstruct the entire phase space to restore unobserved state variables. For this purpose, the integration or differentiation of the observed data series can be performed. Then, a fast-algebraic method can be used to obtain a nonlinear system in the form of a polynomial dynamical system. In this paper, we extend the algebraic method proposed by Kera and Hasegawa to Laurent polynomials which contain negative powers of variables, unlike ordinary polynomials. We provide a theoretical basis and experimental evidence that the integration of a data series can give more accurate results than the widely used differentiation. With this technique, we reconstruct Lorenz attractor from a one-dimensional data series and B. Muthuswamy's circuit equations from a three-dimensional data series.

Keywords: nonlinear systems; nonlinear identification; system reconstruction; Buchberger–Möller algorithm; Laurent polynomials; nonlinear regression; memristor; chaotic system

1. Introduction

The problem in nonlinear dynamical system reconstruction from data series can be described as follows: For a given data series, a system must be found that produces data series that are close in a certain sense to the initial series. Obtaining system models using a limited set of observations is crucial for the reverse engineering of gene regulatory networks [1], analysis of chemical oscillations [2], studies on biological circadian rhythms [3], reconstruction of neuronal models [4], etc.

When systems described by ordinary differential equations (ODEs) are considered, one can specify the problem statement as follows: For a given data series $\{\mathbf{x}_i\}$ and its time derivative $\{\dot{\mathbf{x}}_i\}$ a system of ODEs must be reconstructed as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \tag{1}$$

where $\mathbf{f}$ is a vector function. Usually, some information about $\mathbf{f}$ is known apriori, for instance, the possible elementary operations constituting it. These operations usually construct a group or, more specifically, a ring. One such ring is a polynomial ring which includes a set of polynomials in several variables with coefficients belonging to a certain field such as a field of real or complex numbers [5,6].

Dynamical systems with polynomial derivative functions are often referred to as polynomial dynamical systems (PDSs). In these systems, $\mathbf{f}(\mathbf{x})$ is considered as

$$\dot{\mathbf{x}} = \mathbf{h} \otimes \mathbf{t}(\mathbf{x}), \tag{2}$$

where $\mathbf{h}$ is the vector of unknown coefficients by the known nonlinear terms (monomials) $\mathbf{t}$ and $\otimes$ denotes element-wise multiplication. The problem usually consists of finding all the entries in $\mathbf{h}$, taking into account that the system can be sparse and many of these coefficients can equal zero. The set of monomials $\mathbf{t}$ is a-priori known. Nonlinear dynamical models of Type (2) are used quite often, examples can be found in the classical works of E. Lorenz, O. Rossler, and J.C. Sprott [7]. System reconstruction in terms of PDS has been investigated since the early 1980s by many authors, for example, by J. P. Crutchfield and B. S. McNamara [8], who proposed one of the classical definitions of the problem, and J. Cremers and A. Hübler, who used Legendre polynomials as a basis for the derivative function construction [9]. Later, Taylor-like expansions were introduced by J. L. Breeden et al. [10]. Many alternatives to PDS models have been proposed, for example, neuronal networks, also including nonlinearities and coefficients found from the experimental data [11]. A neuronal network can sometimes give more precise results and is more preferable in cases of complex systems. The disadvantage of the neuronal technique is that it yields large-scale and computationally costly models. Neuronal networks and the PDS approach can also be combined, for example, for gene regulatory network identification.

The problem of finding a reliable algorithm for determining coefficients of PDS (2) is difficult, especially in case(s) of noisy and incomplete data. Determining coefficient vector $\mathbf{h}$ can be considered as the optimization problem

$$\mathbf{h}^* = \arg\min \|\mathbf{v} - \mathbf{f}(\mathbb{X})\|, \tag{3}$$

where $\mathbf{v}$ is the set of observed derivatives $\mathbf{v} = \{\dot{\mathbf{x}}_i\}$, and $\mathbf{f}(\mathbb{X}) = \{\mathbf{f}(\mathbf{x}_i)\}$ is a set of function values evaluated on $\{\mathbf{x}_i\}$. The size of the unknown vector $\mathbf{h}$ is equal to $M \times L$ where M is the system dimension and L is a number of available monomials.

The Problem (3) can be treated as a problem of nonlinear regression. Several researchers, for example, H. Iba and D. P. Searson et al. [12,13] developed specialized versions of genetic algorithms for solving it. The main disadvantage of genetic algorithms is that their run times are, in a general case, much longer in comparison to other approaches. Meanwhile, classical optimization methods such as Newton or quasi-Newton methods are inefficient on this problem because the optimization problem is discontinuous. When one or several coefficients become equal to zero, the nonlinear regression result changes sufficiently.

Fortunately, Problem (3) can also be considered as the linear regression problem. Indeed, consider $\|\cdot\|$ as 2-norm, and notice that since the set $\mathbb{X}$ is already known, the function values $\mathbf{f}(\mathbb{X})$ can be evaluated. Therefore, the least squares method (LSM) can be applied in order to obtain $\mathbf{h}^*$:

$$\mathbb{E}^{\mathrm{T}}\mathbb{E}\,\mathbf{h}^* = \mathbb{E}^{\mathrm{T}}\mathbf{v}, \tag{4}$$

where $\mathbb{E} = \mathbf{f}(\mathbb{X})$.

The disadvantage of LSM is that Problem (3) is often ill-conditioned since some of the monomials vanish on the given data series and cause the singularity of the matrix $\mathbb{E}^{\mathrm{T}}\mathbb{E}$. To avoid this, Laubenbacher and Stigler proposed division by Gröbner basis to simplify the initial set of monomials used for regression [1], but this approach is not applicable to noisy data series.

Even if monomials do not vanish or the monomial set was somehow simplified to remove all vanishing elementary functions, the matrix $\mathbb{E}^{\mathrm{T}}\mathbb{E}$ usually has a large condition number which makes Problem (3) difficult to solve. Recently, D. Yogatama and N. Smith considered the application of L_1-regularized LSM to these types of problems, which converge better than the classical version of LSM [14]. Later, H. Kera and Y. Hasegawa proposed excluding vanishing polynomials using the

approximate Buhberger-Müller algorithm and, then, deleting minor terms from regression using a specialized procedure [15]. The latter algorithm, further referred to as the Kera-Hasegawa (KH) method, underlies the reconstruction procedure developed in this paper.

From a physical point of view, some state variables needed to perform the reconstruction procedure can be unknown. The simplest case is when only system states are known and the derivatives are not, so numerical differentiation should be applied. Howeve, if the system is partially observed, phase space reconstruction must be performed. For example, when a circuit with a memristive element is considered, one of its state variables, magnetic flux or charge in a memristive device, cannot be directly measured in a macroscopic scale [16]. The problem is more complicated in the domain of physical chemistry where only one variable is usually observed, for example, in Belousov-Zhabotinsky reaction this is Br-ISE electric potential [17]. Using the well-known Takens' embedding theorem [17] we can reconstruct the discrete-time system phase space using a shifted time series. For continuous-time reconstruction, we can use differentiation or integration. The latter technique is used quite rarely because it can cause the emergence of linear trends in integrated data but we will show how to avoid this and why it is more preferable than differentiation.

The main aim of this paper is to establish an efficient and reliable technique for identifying chaotic circuits, especially, ones with nonlinear elements with unknown parameters. The problem arises from the fact that real circuits even with simple nonlinearities and known parameters need certain parameter optimization for simulation despite the theoretical prediction that the circuit can accurately correspond to the ODE [18]. When the circuit contains a nonlinear element, the problem becomes more complicated since the relevant model of this element may also be different from the prediction. For example, it has recently been shown that original HP models are irrelevant to the actual resistance switching devices with memory effect [16]. The relevant model proposed in this work [16] contains exponents, hyperbolic trigonometric functions and negative powers.

This paper is organized as follows: First, we describe how PDS can be reconstructed from complete data using the extended Kera-Hasegawa method; second, we explain how these results can be applied to data series lacking some state variables; third, we carry out computational experiments showing the validity of the approach, to be more precise, we reconstruct the Lorenz system from 1-dimensional data series and a 4-dimensional memristive circuit proposed by B. Muthuswamy [19] using incomplete 3-dimensional data and we show that the accuracy of numerical integration and differentiation plays an important role in the success of reconstruction, even though the reconstruction method is noise-tolerant; finally, we discuss the limitations of the proposed approach and give a conclusion.

2. Algebraic Method for ODE Reconstruction

In this section, we describe an algebraic approach to the system reconstruction method based on subsequent application of the approximate Buchberger–Möller method, the least-squares method and deleting minor terms method, close to the method proposed by Kera and Hasegawa, but we extend this approach to the Laurent polynomials which can have negative powers of monomials in comparison with ordinary polynomials with nonnegative powers only.

2.1. Polynomial Rings, Ideals, ABM Algorithm

Consider a nonlinear function in Equation (1) as

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} \sum\limits_{i=1}^{L} h_{1i} t_i(\mathbf{x}) \\ \vdots \\ \sum\limits_{i=1}^{L} h_{Mi} t_i(\mathbf{x}) \end{pmatrix}, \tag{5}$$

where M is dimension, L is the number of elementary monomials t_i, and h_{ji} are the coefficients of each elementary monomial, and $\mathbf{x} = (x_1, x_2, \ldots x_M)^{\mathrm{T}}$. Each component of Equation (5) represented in the form $\sum_{i=1}^{L} h_{1i} t_i(\mathbf{x})$ is a polynomial. The monomials have general form as

$$t_i(\mathbf{x}) = \prod_{i=j}^{M} x_j^{\alpha_{ij}} \tag{6}$$

where α_{ij} are the powers of each variable. The polynomial functions

$$f_k(\mathbf{x}) = \sum_{i=1}^{L} h_{ki} t_i(\mathbf{x}) \tag{7}$$

form a *polynomial ring* $\mathbb{P}$ over the field $\mathbb{F}$, where operations of addition and multiplication are defined as usual. More precisely, $\mathbb{P}$ is an Abelian group under addition since it is associative, commutative, has an additive inverse and has zero element $f(\mathbf{x}) = 0$, $\mathbb{P}$ is monoid under multiplication since it is associative and has an identity element $f(\mathbf{x}) = 1$ which is exactly

$$f(\mathbf{x}) = 1 \cdot \prod_{i=j}^{M} x_j^0,$$

also this multiplication is distributive. The field over which the ring is defined determines the properties of the coefficients $h_{ki} \in \mathbb{F}$. In this paper, we consider $\mathbb{P}$ over a field of real numbers $\mathbb{R}$.

When $\alpha_{ij} \in \mathbb{N}$, the functions of Equation (6) form ordinary polynomials, or Taylor-like expressions, e.g., for the two-variable function it can be

$$f(x, y) = h_{11} + h_{12}x + h_{13}y + h_{14}x^2 + h_{15}xy + h_{16}y^2 + \cdots \tag{8}$$

When $\alpha_{ij} \in \mathbb{Z}$, the functions for Equation (5) are called Laurent polynomials where powers are conventionally defined within a symmetric interval $\alpha_{ij} \in [-M, M]_{\mathbb{Z}}$, but we will not require this property further. The terms in Equation (8) can be ordered in various ways. We will consider a degree-lexicographic order, where we primarily consider the power of the entire monomial defined as

$$\alpha_i = \sum_{j=1}^{M} \alpha_{ij}.$$

Second, we consider the powers of variables selected alphabetically organized in ascending order of absolute value of power. This may be exemplified for a set of Laurent monomials of 2 variables and order in $[-1; 1]$ as

$$\mathbb{T}(x, y) = \left\{ x^{-1}, \ y^{-1}, 1, xy^{-1}, x^{-1}y, \ x, y \right\}.$$

Originally, only positive powers are considered in degree-lexicographic order, so we introduce this equation here by analogy; for discussion, see Section 4.

The *order ideal O* of the polynomial ring $\mathbb{P}$ is a finite set of monomials in $\mathbb{T}$ if it is closed under divisors, i.e., satisfying the condition: for each monomial $t \in O$ all monomials t_j dividing it also belong to O. A "border" of the order ideal ∂O in an ordinary polynomial ring is a set of monomials in $\mathbb{T}$ defined as

$$\partial O = (x_1 O \cup x_2 O \cup \ldots \cup x_M O) - O, \tag{9}$$

where "minus" denotes set subtraction. If the monomial set $O = \{1, x, y\}$ is an order ideal in an ordinary polynomial ring, its border is

$$\partial O = \left\{ x^2, xy, y^2 \right\}.$$

In Laurent polynomial ring, a border is defined similarly to Equation (9) but also includes negative powers $x_i^{-1}, i = 1 \dots M$:

$$\partial O = \left(x_1 O \cup x_2 O \cup \dots \cup x_M O \cup x_1^{-1} O x_2^{-1} O \cup \dots \cup x_M^{-1} O \right) - O.$$

The *vanishing ideal* of a set of data points $\mathbb{X}$ is a set of all polynomials vanishing (i.e., equals to zero) on $\mathbb{X}$:

$$\mathbb{I}(\mathbb{X}) = \{ p \in \mathbb{P} \,|\, p(\mathbf{x}) = 0, \ \forall x \in \mathbb{X} \}. \tag{10}$$

The vanishing ideal $\mathbb{I}(\mathbb{X})$ can be spanned by a finite set of polynomials called the *basis* of $\mathbb{I}(\mathbb{X})$. Let $O = \{ t_1 \dots t_L \}$ be an order ideal with a border $\partial O = \{ b_1 \dots b_v \}$. Let $G = \{ g_1 \dots g_v \} \subset \mathbb{P}$ be a set of polynomials, and there is an ideal $\mathbb{I} \subseteq \mathbb{P}$. The set G is an *O-border prebasis* if the polynomials in it have the form

$$g_j = b_j - \sum_{i=1}^{L} \alpha_{ij} t_i \text{ for } j \in [1 \dots v], \ \alpha_{ij} \in \mathbb{F}. \tag{11}$$

The set G is an *O-border basis of* $\mathbb{I}$ if G generates $\mathbb{I}$ and if the residue classes of the terms in O constitute a vector basis of $\mathbb{P} - \mathbb{I}$. Several other bases are known, such as Gröbner basis and Macaulay basis. To find a border basis, Buchberger–Möller (BM) algorithms can be used [15]. Border basis G is called approximate if $\|g(\mathbb{X})\|_2 < \epsilon$, where $g \in G$ and ϵ is a small positive number. Modification of BM algorithm, known as the approximate Buchberger–Möller (ABM) algorithm can be used to compute an approximate border basis on a noisy dataset $\mathbb{X}$.

We also define the evaluation map $\mathcal{E}$, which transforms a given dataset $\mathbb{X} = \{ \mathbf{x}_1 \dots \mathbf{x}_N \}$ into a matrix $\mathcal{E}_{\mathbb{T}}(\mathbb{X}) = \mathbb{E}_{\mathbb{T}}$, where

$$\mathbb{E}_{\mathbb{T}} = \begin{pmatrix} t_1(\mathbf{x}_1) & \dots & t_L(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ t_1(\mathbf{x}_N) & \dots & t_L(\mathbf{x}_N) \end{pmatrix}.$$

The ABM algorithm can be described as follows. Suppose, we have a dataset $\mathbb{X}$, a tolerance ϵ, and set of monomials $\sigma \subset \mathbb{T}$ in degree-lexicographic order. We should obtain a border basis G and an order ideal O. We start with degree $d = 1$ and $O = \{1\}$. Let $L \subset \sigma$ initially contain all terms of degree 1. Then, the process starts while the order d is not greater than the maximal order defined as max degσ and L is not empty. At the current step, all monomials of degree $d - 1$ have been already processed. We estimate the vector A on the next monomial in σ as

$$A = \mathcal{E}_t(\mathbb{X}), t \in L, \tag{12}$$

and, then, remove it from L. Let $\lambda_{\min}$ be the smallest eigenvalue of the matrix $A^T A$

$$\lambda_{\min} = \lambda_i, \ i = \arg \min \left| \text{eig} \left(A^T A \right) \right|,$$

and if it satisfies the condition

$$\sqrt{\lambda_{\min}} > \epsilon,$$

then t is appended to O and all the polynomials of order $d + 1$ are added to ∂O. A polynomial with coefficients defined by an eigenvector of $A^T A$ corresponding to $\lambda_{\min}$ is added to G.

Application of ABM algorithm to the dataset $\mathbb{X}$ and the set of initial monomials σ allows obtaining O and ensures that any polynomial

$$p(\mathbf{x}) = \sum_{i=1}^{L} h_{ki} o_i(\mathbf{x}_k),$$

where $o_i \in O$, does not vanish on $\mathbb{X}$, even though O can be, strictly speaking, not always an order ideal [15].

2.2. KH Algorithm for Polynomial Function Reconstruction

Consider a dataset $\mathbb{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}$ which can be represented as observation vectors $\mathbf{x}_i$ for i-th state variable, M is the system dimension. Each data vector has N observation points. Using the order ideal O as a set of basis monomials, one can evaluate them on a given dataset $\mathbb{X}$ and then build an interpolation, using derivatives set $\mathbb{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_M\}$ which is also given or can be derived from $\mathbb{X}$ by the relation $\mathbf{v}_i = \dot{\mathbf{x}}_i$:

$$\mathbf{v}_k = \sum_{i=1}^{L} h_{ki} o_i(\mathbf{x}_k). \tag{13}$$

In each dimension, Equation (13) can be reconstructed using the least squares method. Denote $\mathbb{E} = \mathcal{E}_O(\mathbb{X})$ and transform Equation (13) into a linear equation:

$$\mathbf{v}_k = \mathbb{E}\mathbf{h}_k. \tag{14}$$

Equation (14) is usually considered as overdetermined, since the number of estimated points can be arbitrary and, in the presence of noise, a large number of observations are a necessary condition for obtaining a reliable solution. The overdetermined Equation (14) can be solved using the ordinary least squares method, obtaining the minimum

$$\mathbf{h}^* = \operatorname{argmin}\|\mathbf{v}_k - \mathbb{E}\mathbf{h}_k\|^2 \tag{15}$$

by analytical Formula (4). Due to the presence of noise and since it is supposed that the vector $\mathbf{h}_k$ is sparse, L_1-regularization is recommended to obtain better results:

$$\mathbf{h}^* = \operatorname{argmin}\left(\frac{1}{2}\|\mathbf{v}_k - \mathbb{E}\mathbf{h}_k\|^2 - \lambda|\mathbf{h}_k|\right), \tag{16}$$

where λ is a regularization parameter. Increasing λ provides a sparser solution. Problem (16) is also known as a basis pursuit denoising (BPDN) problem, and several special optimization methods have been proposed for solving it, e.g., Nesterov method [20,21]. If it is assumed that $\mathbf{h}^*$ is dense, L_2-regularization or Tikhonov regularization [22] can be applied:

$$\mathbf{h}^* = \operatorname{argmin}\left(\frac{1}{2}\|\mathbf{v}_k - \mathbb{E}\mathbf{h}_k\|^2 - \lambda\|\mathbf{h}_k\|^2\right), \tag{17}$$

with a known solution in terms of matrices:

$$\mathbf{h}^* = \left(\mathbb{E}^{\mathrm{T}}\mathbb{E} + \lambda N I\right)^{-1}\left(\mathbb{E}^{\mathrm{T}}\mathbf{v}\right), \tag{18}$$

where I is an identity matrix [23].

Kera and Hasegawa [15] proposed removing minor terms from the interpolation polynomial

$$H_k(\mathbf{x}) = \sum_{i=1}^{L} h_{ki} o_i(\mathbf{x}) \tag{19}$$

using a special procedure based on an estimation

$$w_i = \|\mathcal{E}_{O_i}(\mathbb{X})h_{ik}\|.$$

For each term in a polynomial $h_{ki} o_i(\mathbf{x})$ its weight (20) is estimated, and terms with minimal weight are subsequently removed from the interpolation while the error of approximation remains within a given tolerance η:

$$\frac{1}{N}\|\mathbf{v}_k - H_k(\mathbb{X})\| \leq \eta. \tag{20}$$

Following Kera and Hasegawa, we call this procedure delMinorTerms. It ensures that the solution is sparse enough even if the data is contaminated with noise.

The Kera–Hasegawa (KH) PDS reconstruction method finds a set of interpolating polynomials $\mathbb{H}$ in $\mathbb{P} = (\mathbb{T} \times \mathbb{R})$ of type (19), where $\times$ is the Cartesian multiplication, satisfying $(\mathbb{H} \subset \mathbb{P}:\ H_k(\mathbb{X}) \approx \mathbf{v}_k)$ for a given $\mathbb{X}$ and $\mathbb{V}$. Firstly, it performs the ABM algorithm to remove monomials from $\mathbb{T}$ vanishing on $\mathbb{X}$, comprising an order ideal O of $\mathbb{T}$. Next, it applies L_1-regularized LSM to find an interpolation polynomial on $O \times \mathbb{R}$ and then runs delMinorTerms to ensure that the interpolated polynomials are sparse while the accuracy criterion Equation (20) is satisfied.

2.3. Phase Space Reconstruction Using Integration Embedding

The KH method can be performed only if $\mathbb{X}$ and $\mathbb{V}$ are complete, i.e., the dimension M of the reconstructed system corresponds to dimensions of $\mathbb{X}$ and $\mathbb{V}$. In the general case, we can observe only one or several state variables, and thus we need to reconstruct the entire phase space. The well-known F. Takens' theorem [23] states that having a time-delayed series of a single-variable signal $\{x[n], x[n-1], x[n-2], \ldots, x[n-(M-1)]\}$, we can reconstruct M-dimensional phase space due to time delay embedding, and, similarly, N. Packard et al. [24] proposed reconstructing M-dimensional phase using $M-1$ subsequent derivatives

$$\left\{ x[n], \frac{dx[n]}{dt}, \ldots, \frac{d^{M-1}x[n]}{dt^{M-1}} \right\}, \tag{21}$$

due to differential embedding. Recently, J. Lekscha and R. Donner [25] have shown that reconstruction is possible in the case of non-uniformly sampled and noisy time series as well, which is an important result for practical use.

Unlike the differential embedding approach, integral embedding did not become popular in scientific literature, despite being more useful in several practical cases.

Lemma 1. *M-dimensional phase space can be reconstructed due to integral embedding, and a trajectory $\mathcal{F}$ of the system F reconstructed from time series $x[n]$ can be written as:*

$$\left\{ \underbrace{\int \ldots \int}_{M-1} x[n]dt^{M-1} + C_{M-1}(t), \ldots, \int x[n]dt + C_1, x[n] \right\}, \tag{22}$$

where $C_{M-1}(t)$ is a residual polynomial whose general form is:

$$C_K(t) = C_K + C_{K-1}t + C_{K-2}t^2 + \ldots + C_1 t^{K-1}, \tag{23}$$

where K is a natural number.

Proof. A system of ODEs in the Cauchy normal form is usually organized so as

$$\frac{d}{dt}x_j = x_{j+1},$$

where the state variable x_j is related to the variable of the next index through differentiation. It can also be written through the integration

$$x_j = \int x_{j+1}dt + C_j,\ j \in [1, M-1], \tag{24}$$

where C_j is a constant depending on the initial condition. Equation (24) can be used for subsequent integration of state variables for reconstructing the trajectory in a full phase space when variables $x_1 \ldots x_{j-1}$ are known. After K integrations of a time series given by x_1, we obtain

$$x_K = \underbrace{\int \ldots \int}_{K} x_1 dt + C_K + C_{K-1}t + C_{K-2}t^2 + \ldots + C_1 t^{K-1},$$

from which Equations (22) and (23) follow obviously. □

Corollary 1. *Since the polynomial residual Equation (23) grows very fast if t^{K-1} is large enough, it is undesirable if we consider a computer implementation of the reconstruction algorithm. Otherwise, the accuracy of the data representation may suffer. We propose the following procedure for obtaining x_j. First, we choose zero initial condition to detect and integrate x_{j+1} on a given time interval $[t_0, t_1]$:*

$$x_j^* = \int_{t_0}^{t_1} x_{j+1} dt.$$

Then, we estimate its mean value $\langle x_j \rangle$, and supposing $C_j = \langle x_j \rangle$, subtract it obtaining the desired solution $x_j = x_j^* - C_j$. Another way is to de-trend the second integral containing $C_j t$ with a specialized procedure, e.g., detrend function in a MATLAB environment.

The system reconstructed using integration embedding is equivalent to the system reconstructed using differential embedding up to $M - 1$ differentiations. More strictly, this can be formulated as Lemma 2.

Lemma 2. *Suppose, there is a map $G(t, y) : \mathbb{R} \times \mathbb{R}^M \to \mathbb{R}^M$, $t \in \mathbb{R}$, $y \in \mathbb{R}^M$ which projects $\mathcal{G}$ onto the phase space $\mathbb{R}^M$, can be obtained via subsequent differentiation of the one-dimensional trajectory $\mathcal{G}_1$ corresponding to the state variable y_1. Then, then a dynamical system $(t, x) : \mathbb{R} \times \mathbb{R}^M \to \mathbb{R}^M$, $x \in \mathbb{R}^M$ exists which projects $\mathcal{F}$ onto the phase space $\mathbb{R}^M$ can be obtained via subsequent integration of the one-dimensional trajectory $\mathcal{F}_M = \mathcal{G}_1$ corresponding to the state variable x_M.*

Proof. The following ODE describes a dynamical system $G(t, \mathbf{y})$ due to the condition that subsequent differentiation of one-dimensional trajectory y_1 generates $\mathcal{G}$

$$\begin{cases} \frac{d}{dt} y_1 = y_2, \\ \quad \vdots \\ \frac{d}{dt} y_{M-1} = y_M, \\ \frac{d}{dt} y_M = g(t, \mathbf{y}). \end{cases} \tag{25}$$

Since $\mathcal{G}_1 = \mathcal{F}_M$ the corresponding state variables are equal: $y_1 = x_M$. The condition that the projection $\mathcal{F}$ of the map F onto the entire phase space can be obtained via subsequent integration of its one-dimensional component $\mathcal{F}_M$ can be expressed in the form of Equation (24). Consequently, the following ODE describes $F(t, \mathbf{x})$:

$$\begin{cases} \frac{d}{dt} x_1 = x_2, \\ \quad \vdots \\ \frac{d}{dt} x_{M-1} = x_M, \\ \frac{d}{dt} x_M = f(t, \mathbf{x}), \end{cases}$$

where $f(t, \mathbf{x})$ is a differentiable function. It is related with y_1 in the following way: $y_1 = \int f(t, \mathbf{x}) dt + C$.
□

Corollary 2. *A dynamical systemF which can be reconstructed via integration embedding exists if the system G which can be reconstructed via differential embedding exists.*

The ODE representation of a partially observed system with K observed variables and M-dimensional entirely phase space can be represented in the following way:

$$\begin{cases} \frac{d}{dt}x_1 = g_1(\mathbf{x}), \\ \quad \vdots \\ \frac{d}{dt}x_K = g_K(\mathbf{x}), \\ \frac{d}{dt}x_{K+1} = x_K, \\ \quad \vdots \\ \frac{d}{dt}x_M = x_{M-1}, \end{cases} \tag{26}$$

where $f_j(\mathbf{x})$ is a scalar function. Missing derivatives of orders $\{K+1,\ldots,M\}$ can be found numerically. Finite-difference-based numerical differentiation is the most common approach, but some other techniques can be used as well, e.g., numerical differentiation based on Legendre polynomials [25]. If the known state variables are denoted $x_{K+1},\ldots,x_M$, the system reconstructed via integration reads

$$\begin{cases} \frac{d}{dt}x_1 = x_2, \\ \quad \vdots \\ \frac{d}{dt}x_K = x_{K+1}, \\ \frac{d}{dt}x_{K+1} = f_{K+1}(\mathbf{x}), \\ \quad \vdots \\ \frac{d}{dt}x_M = f_M(\mathbf{x}), \end{cases} \tag{27}$$

which is an alternative form of Equation (26) with other variables. □

The number of dimensions M can be selected concerning its physical sense or with one of the several algorithms, e.g., false nearest neighbor (FNN) algorithm [26,27], or estimating one of the fractal dimensions or entropies [28].

The embedded integration approach is more useful in some practical cases due to the following two theorems. Theorem 1 considers a case appearing in many practical applications when the differential embedding approach cannot give a valid reconstruction due to the non-uniqueness of the solution corresponding to the reconstructed trajectory $\mathcal{F}$.

Theorem 1. *A function f (t) determined on $\mathbb{R}$ exists, satisfying the following conditions:*

(1)   *$f(t)$ is a function that produces time series $\boldsymbol{u} = \{f(t_i)\}$ being a one-dimensional projection of a M-dimensional trajectory $\mathcal{F}$ of the dynamical system F;*

(2)   *$f(t)$ is continuous and satisfies the Lipchitz condition*

$$\left| f(t_1) - f(t_2) \right| \leq K|t_1 - t_2|$$

   where K > 0 is the Lipchitz constant;

(3)   *an M-dimensional trajectory $\hat{\mathcal{F}}$ can be reconstructed from time series $\boldsymbol{u}$ using differentiation which does not satisfy the condition of the uniqueness of the solution;*

(4)   *an M-dimensional trajectory $\widetilde{\mathcal{F}}$ can be reconstructed from time series $\boldsymbol{u}$ using integration that satisfies the condition of uniqueness of the solution and thus is topologically valid.*

Proof. Consider a continuous function that satisfies the following conditions. First, there exists an infinite set of time points at which the function f and its derivatives equal an arbitrary number $a \in \mathbb{R}$:

$$\exists \boldsymbol{\tau} : \ f(t_0) = a, \ f'(t_0) = a, \ldots, \ f^{M-1}(t_0) = a, \ t_0 \in \boldsymbol{\tau}. \tag{28}$$

At other time points the function f differs from a:

$$f(t_j) \neq a, \ f'(t_j) \neq a, \ldots, \ f^{M-1}(t_j) \neq a, \ t_0 \notin \boldsymbol{\tau}. \tag{29}$$

Second, for a sufficiently large number $Q \in \mathbb{N}$: $M \leq Q$ there exist two time points t_i and t_j, where the function changes its value relatively to a:

$$f \exists (t_j, t_i) : f^{(q)}(t_i) \langle a; \ f^{(q)}(t_j) \rangle a, \tag{30}$$

$q \in [0; Q]$. Then, under this conditions, reconstruction by differentiation yields in a phase space where the point $\mathbf{x}_a = (a, a, \ldots, a)^{\mathrm{T}}$ is a point at which at least two trajectory loops are connected and therefore the uniqueness of the solution is violated: starting from this point and having no apriori information of which branch can be selected we can neither predict the next state of the system nor determine the previous state of the system. Therefore, the trajectory $\hat{\mathcal{F}}$ reconstructed from time series $\mathbf{u}$ using differentiation does not satisfy the condition of uniqueness of the solution. Meanwhile, due to Lemma 1 and Lemma 2, it is possible to obtain M-dimensional trajectory $\widetilde{\mathcal{F}}$ from time series $\mathbf{u}$ using integration. Due to Corollary 1 it is possible to select such integration constants C_j when $\mathbf{u}$ and its integrals are not equal to a so it may satisfy the condition of the uniqueness of the solution and thus $\widetilde{\mathcal{F}}$ is topologically valid. Figure 1 illustrates the phase space geometry. □
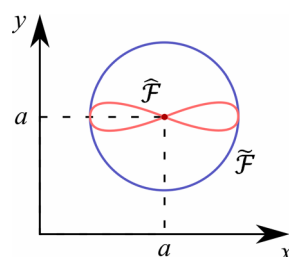

Figure 1. Illustration of Theorem 1. The red line corresponds to the double-loop trajectory $\hat{\mathcal{F}}$ obtained by the differential approach, the blue line corresponds to cyclic trajectory $\widetilde{\mathcal{F}}$ obtained by the integration approach. Here, x and y are phase variables.

Time series $\mathbf{u} = \{f(t_i)\}$ satisfying conditions 2–4 of the Theorem 1 when $a = 0$ are rather common in biology, medicine, genetics and neuroscience.

Theorem 2 is dedicated to noise-amplifying properties of differentiation operators. In the presence of additive white Gaussian noise, when the sampling frequency is high enough and the spectral density of the signal is mostly low-frequency, differentiation tends to decrease the signal-to-noise ratio (SNR), while integration does not.

Theorem 2. *Denote the Fourier transform of a signal $u(t)$ as $U(\omega)$. Let the power spectrum of u lie almost entirely in a frequency interval $[\omega_0, \ \omega_1]$. Let $\xi(t)$ be the additive white Gaussian noise (AWGN) contaminating u. Then:*

(1)   after differentiation, the SNR of $\frac{d}{dt}u$ will be smaller than the SNR of u;
(2)   after integration, the SNR of $\int u dt$ will be at least not smaller than the SNR of u.

Proof. Recall, the signal-to-noise ratio is determined as

$$\text{SNR} = \frac{A_u^2}{A_\xi^2},$$

where A_u is root-mean-square (RMS) amplitude of the signal with its square equal to

$$A_u^2 = \lim_{T \to \infty} \int_0^T \frac{[u(t)]^2}{T} dt,$$

where T is time interval, and A_ξ is RMS amplitude of the noise.

Denote the sampling frequency as ω_s. Denote the Fourier transform of the noise $\xi(t)$ as $\Xi(\omega)$. Then, the power spectral density of the noise calculated within frequency interval $\omega \in [0; \omega_s/2]$ is

$$S_{\xi\xi}(\omega) = \lim_{T \to \infty} \frac{|\Xi(\omega)|^2}{T}$$

According to Parseval's theorem,

$$\int_0^T [\xi(t)]^2 dt = \int_0^{\frac{\omega_s}{2}} |\Xi(\omega)|^2 dt,$$

and the same is true for the signal. Consequently, from Equation (31) and Equations in (32) SNR may be expressed through power spectra as

$$\text{SNR} = \lim_{T \to \infty} \frac{\int_0^{\frac{\omega_s}{2}} S_{uu}(\omega) d\omega}{\int_0^{\frac{\omega_s}{2}} S_{\xi\xi}(\omega) d\omega}$$

Note that AWGN has a uniform power spectral density, i.e.,

$$\forall \omega \in \left[0; \frac{\omega_s}{2}\right] : S_{\xi\xi}(\omega) = \langle S_{\xi\xi} \rangle.$$

The condition that the power spectrum of u lie *almost entirely* in a frequency interval $[\omega_0, \omega_1]$ means that its power spectrum $S_{uu}(\omega)$ satisfies the condition:

$$\int_0^{\omega_1} S_{uu}(\omega) d\omega > q \int_0^{\omega_s/2} S_{uu}(\omega) d\omega,$$

where $0 < q < 1$ is a positive real number close to 1, and ω_1 is a frequency up to which almost the whole spectrum is contained, usually, $\omega_1 < \frac{\omega_s}{2...10}$. The gap between ω_1 and $\omega_s/2$ is almost occupied by the noise only, see Figure 2a. Consequently, in this case, a formula for SNR can be rewritten as:

$$\text{SNR} = \lim_{T \to \infty} \frac{\int_0^{\omega_1} S_{uu}(\omega) d\omega}{\int_0^{\frac{\omega_s}{2}} S_{\xi\xi}(\omega) d\omega} \tag{31}$$

Consider integration and differentiation acting at u. Taking Fourier transform, obtain

$$\begin{cases} \int_0^T u(t) dt \doteq \frac{1}{j\omega} U(\omega) \\ \frac{d}{dt} u(t) \doteq j\omega U(\omega) \end{cases},$$

where j is the imaginary unit. The same is true for the noise. Denote $\frac{d}{dt}u(t) = \dot{u}(t)$, $\int_0^T u(t)dt = \mathcal{U}(t)$, $\frac{d}{dt}\xi(t) = \dot{\xi}(t)$, $\frac{d}{dt}\xi(t) = \Xi(t)$. Then, corresponding spectral densities are:

$$
\begin{cases}
S_{\dot{u}\dot{u}}(\omega) = \omega^2 S_{uu}(\omega), \\
S_{\mathcal{U}\mathcal{U}}(\omega) = \frac{1}{\omega^2} S_{uu}(\omega), \\
S_{\dot{\xi}\dot{\xi}}(\omega) = \omega^2 S_{\xi\xi}(\omega), \\
S_{\Xi\Xi}(\omega) = \frac{1}{\omega^2} S_{\Xi\Xi}(\omega).
\end{cases}
\tag{32}
$$

From Equation (31) and Equations in (32), SNR of the differentiated signal is:

$$
\mathrm{SNR_D} = \lim_{T\to\infty} \frac{\int_0^{\omega_1} \omega^2 S_{uu}(\omega)d\omega}{\int_0^{\frac{\omega_s}{2}} \omega^2 S_{\xi\xi}(\omega)d\omega} = \lim_{T\to\infty} \frac{\int_0^{\omega_1} \omega^2 S_{uu}(\omega)d\omega}{\frac{\omega_s^3}{24}\left\langle S_{\xi\xi}\right\rangle}.
$$

Due to Hölder's inequality, and the fact that both ω^2 and $S_{uu}(\omega)$ are nonnegative,

$$
\mathrm{SNR_D} \le \lim_{T\to\infty} \frac{\frac{\omega_1^3}{3}\int_0^{\omega_1} S_{uu}(\omega)d\omega}{\frac{\omega_s^3}{24}\left\langle S_{\xi\xi}\right\rangle} = \left(\frac{2\omega_1}{\omega_s}\right)^3 \mathrm{SNR}.
$$

Since $\frac{2\omega_1}{\omega_s} < 1$, the $\mathrm{SNR_D}$ is never greater than the SNR of the original signal under the above made assumptions.

For the integration operator, from Equation (31) and Equations in (32),

$$
\mathrm{SNR_I} = \lim_{T\to\infty} \frac{\int_0^{\omega_1} \frac{1}{\omega^2} S_{uu}(\omega)d\omega}{\int_0^{\frac{\omega_s}{2}} \frac{1}{\omega^2} S_{\xi\xi}(\omega)d\omega} = \lim_{T\to\infty} \frac{\int_0^{\omega_1} \frac{1}{\omega^2} S_{uu}(\omega)d\omega}{-\frac{2}{\omega_s}\left\langle S_{\xi\xi}\right\rangle}.
$$

Again, due to Hölder's inequality,

$$
\mathrm{SNR_I} \le \lim_{T\to\infty} \frac{\frac{1}{\omega_1}\int_0^{\omega_1} S_{uu}(\omega)d\omega}{\frac{2}{\omega_s}\left\langle S_{\xi\xi}\right\rangle} = \frac{\omega_s}{2\omega_1}\mathrm{SNR}.
$$

Therefore, while the $\mathrm{SNR_I}$ is not obligatorily greater than the SNR of the original signal, it is not necessarily less than it in contrast to the case of using differentiation. □
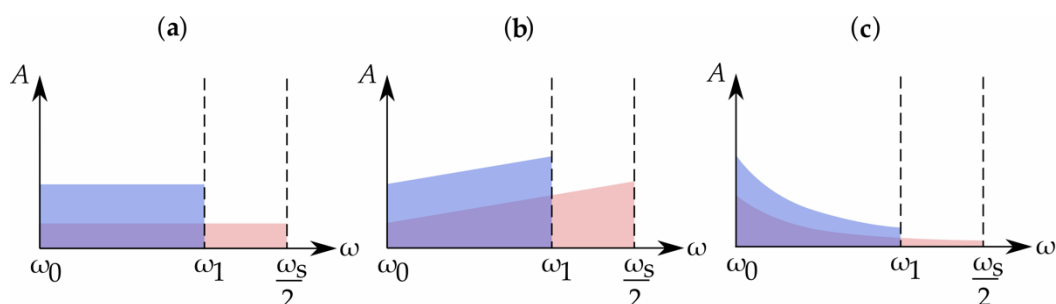

Figure 2. Illustration of Theorem 2. Blue plot corresponds to the Fourier transform of the signal $U(\omega)$ and red plot corresponds to the Fourier transform of the noise $\Xi(\omega)$. (**a**) Original signal and noise; (**b**) after differentiation; and (**c**) after integration.

Theorem 2 is in good correspondence with the signal processing theory [29] from which it is known that differentiation operator relates to a high-pass filter, and amplifies high-frequency noise, while integration operator relates to a low-pass filter.

2.4. Description of the Proposed Reconstruction Technique

The overall proposed technique is described as follows:

First, we consider the time series set $\mathbb{U}$ of a partially observed dynamical system and its derivative $\mathbb{W}$. Using both integral or differential embedding, preferring the first one due to Theorems 1 and 2, we reconstruct the phase space and the M-dimensional phase space trajectory $\mathbb{X}$. In this paper we suppose that M is already known. If not, the FNN algorithm or other dimension-finding algorithms can be applied. For the reasons why we do not reject differential embedding, see the Limitations section.

Then, we generate a degree-lexicographically organized set of monomials σ of order $\mathcal{D} = [d_-; d_+]$, where the power of each monomial lies within $\mathcal{D}$ and powers of each term in monomial also lies within $\mathcal{D}$. From a practical point of view, we assume that $d_- \leq 0$ and $d_+ \geq 0$. The algorithm for generating σ is given in pseudocode in Listing 1.

The subroutine *generateV* returns a special indexing matrix V:

$$
V = \begin{pmatrix}
\text{Dpos} \begin{cases} \overbrace{\begin{matrix} 1 & 0 & \dots & 0 \end{matrix}}^{M} \\ \vdots & \vdots \\ 1 & 0 & & \ddots \end{cases} \\
\text{Dneg} \begin{cases} -1 & 0 \\ \vdots & \vdots \\ -1 & 0 & \dots & 0 \end{cases}
\end{pmatrix}
$$

The subroutine *adds* produces a column-wise sum of V as a candidate vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)^{\mathrm{T}}$ of monomial terms powers:

$$
t(\mathbf{x}) = \prod_{i=1}^{M} x_i^{\alpha_i}. \tag{33}
$$

The monomial $t(\mathbf{x})$ is added to σ if it is not already contained in σ. The order of multiplication is neglected, e.g., the term $x_1 x_2 x_1$ is considered similar to the term $x_1^2 x_2$.

The subroutine *shiftForward* produces the next state of matrix $V_{k+1} = V_k \oplus \mathbf{1}$ considering V as a column-type of D-digit number in M-ary numeral system, where each digit has M states defined by the position of the non-zero element in each row, and where the first row is the least significant digit of the number, and $V_k \oplus \mathbf{1}$ denotes summation with the column vector $\mathbf{1} = (1, 0, \dots, 0)^{\mathrm{T}}$.

Box 1. A pseudocode for a *deglexord* routine generating degree-lexicographically ordered Laurent monomials

```
begin
  Dneg = −d₋
  Dpos = 0
  σ = ∅
  while (Dpos ≤ dmax) and (Dneg ≥ 0)
      D = Dneg + Dpos
      V = generateV(Dneg, Dpos)
      while the M-th column of V is not filled
          σ = adds(V, σ)
          shiftForward(V)
      end
      σ = adds(V, σ)
      if Dpos = d₊
          Dneg = Dneg −1
      end
      if Dpos < d₊
          Dpos = Dpos +1
      end
  end
end
```

When the degree-lexicographically ordered set of monomials σ is computed, the Kera–Hasegawa method is applied to σ, $\mathbb{V}$, and $\mathbb{X}$. The obtained reconstructed PDS corresponds to the original dynamical system if the noise level in the data is sufficiently low and the system is representable in a form of PDS in a given set of coordinates.

3. Experimental Results

In this section, we first illustrate the advantages of using integral embedding using two examples and, then, show how the proposed approach can be used for the reconstruction of Lorenz attractor from a one-dimensional time series and chaotic memristive circuit equations from a three-dimensional time series.

3.1. Illustrations of the Theorems

The following example illustrates the Theorem 1.

Example 1. *Consider the ODE:*

$$\begin{cases} \dot{x} = y, \\ \dot{y} = -x, \\ \dot{z} = f(x,y), \end{cases} \tag{34}$$

where $f(x,y) = 2\,sign(x)\,\frac{y}{x^3}e^{-\frac{1}{x^2}}$. *The solution of (30) is*

$$\begin{cases} x = \sin(t), \\ y = \cos(t), \\ z = sign(\sin t)e^{-\frac{1}{\sin^2 t}}, \end{cases} \tag{35}$$

The function $z(t)$ *in (31) satisfies the conditions of the Theorem 1. Its unresolvable point is* $x_0 = (0,0,0)^T$. *Figure 3 shows two variants of 3-dimensional phase spaces, reconstructed via differentiation and integration from time series* $\boldsymbol{u} = \{z(t_i)\}$.
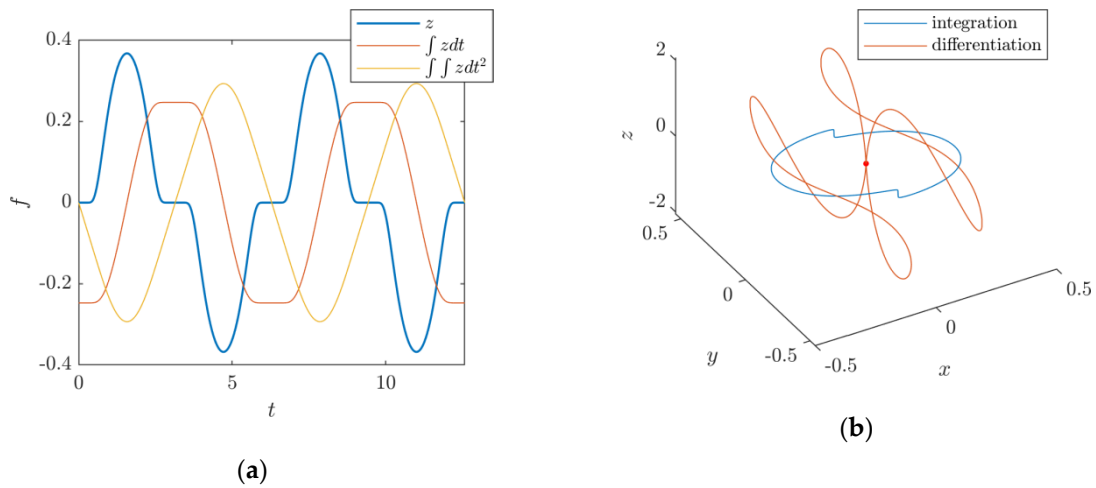
(a)

(b)

Figure 3. Example of phase space reconstruction. (**a**) Time series z and its two integrals; (**b**) comparison of an integration-based approach (blue trajectory) and a differentiation-based approach (red trajectory). The red dot at $\mathbf{x}_0 = (0,0,0)^{\mathrm{T}}$ denotes a singular point unresolvable in sense of FNN where the uniqueness condition is violated.

Example 2 shows how embedded differentiation and integration approaches can be applied to ECG signals. In this example, the conditions of both Theorems are satisfied, i.e., differentiation produces noisy trajectory reconstruction with a singular point, while integration does not have these drawbacks.

Example 2. *Consider the ECG signal from ECG-ID Database [30,31], as shown in Figure 4a.*

(a)

(b)
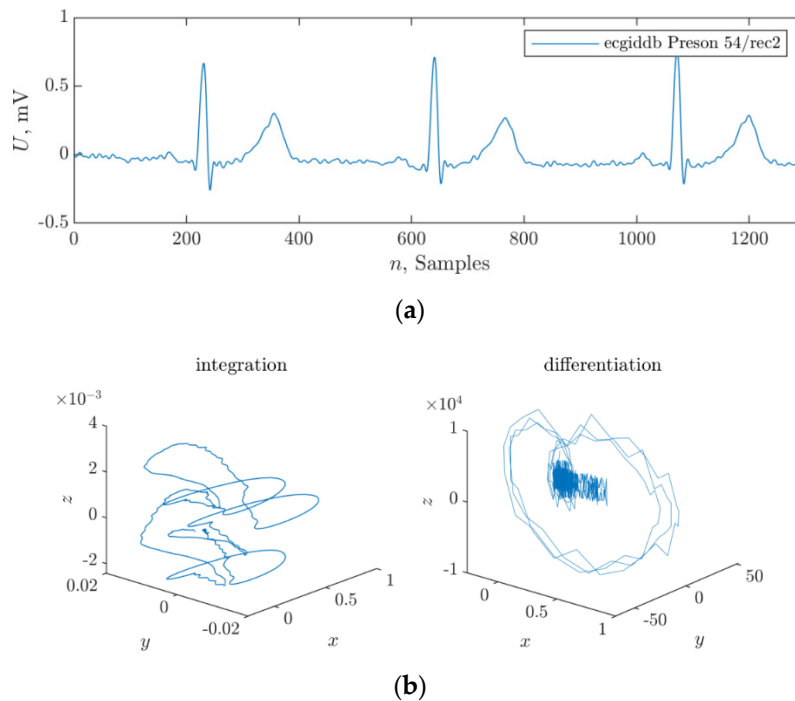
Figure 4. Example of the phase space reconstruction. (**a**) ECG time series x; (**b**) comparison of the integration-based approach (left) and the differentiation-based approach (right).

After applying the embedded differentiation approach, we obtain a trajectory that does not satisfy the uniqueness criterion and is contaminated with noise. After applying embedded integration approaches, we obtain a rather a smooth trajectory with no singular points.

3.2. Reconstruction of Lorenz Attractor

Consider the classical Lorenz attractor

$$\begin{cases} \dot{x} = -sx + sy, \\ \dot{y} = -zx + rx - y, \\ \dot{z} = xy - bz, \end{cases} \tag{36}$$

with a standard parameter set: $s = 10$, $r = 28$, $b = 8/3$ [32]. Suppose, only one variable z can be observed, so we should transform the Lorenz system to obtain a model in new coordinates:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_3, \\ \dot{x}_3 = -sx_1^3 - x_1^2 x_2 + bs(r-1)x_1 - b(1+s)x_2 - (1+b+s)x_3 + \frac{x_2 x_3}{x_1} + \frac{(s+1)x_2^2}{x_1}, \end{cases} \tag{37}$$

Equation (37) cannot be reconstructed into ordinary polynomial functions due to the negative powers of variable x_1, but it is possible to reconstruct it as the Laurent polynomial function. The application of the proposed algorithm can be done using both differential and integration approaches.

The *deglexord* procedure (see Listing 1) generates a set of $L = 50$ monomials or degrees $\mathcal{D} = [-1; 3]$. To ensure that the system of equations is overdetermined, we should take not less than 51 points in $\mathbb{X}$. We are never sure that any term will vanish on $\mathbb{X}$ and will be removed by the ABM algorithm. An example of the reconstructed system is:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_3, \\ \dot{x}_3 = -10x_1^3 + 720x_1 - x_1^2 x_2 - 29.3333x_2 - 13.6667x_3 + 11x_1^{-1}x_2^2 + x_1^{-1}x_2 x_3, \end{cases} \tag{38}$$

To reach a high quality of reconstruction, we generate data using the function *ode113* with parameters $RelTol = 10^{-13}$ and $AbsTol = 10^{-15}$ and constant stepsize $h = 10^{-4}$ and simulation time $T = 45$ in MATLAB 2019b environment (Campus license No. 40502181, ETU "LETI") and use the 4th order integration formula. Note that Equation (37) can be reconstructed from the data obtained from the original Lorenz system using the time series corresponding to the variable z and its two integrals.

Two projections of the attractor Equation (37) and its reconstructed version are given in Figure 5. The initial conditions were $(0.1, 0, -0.1)^{\mathrm{T}}$, the parameter $\eta = 10^{-2}$.
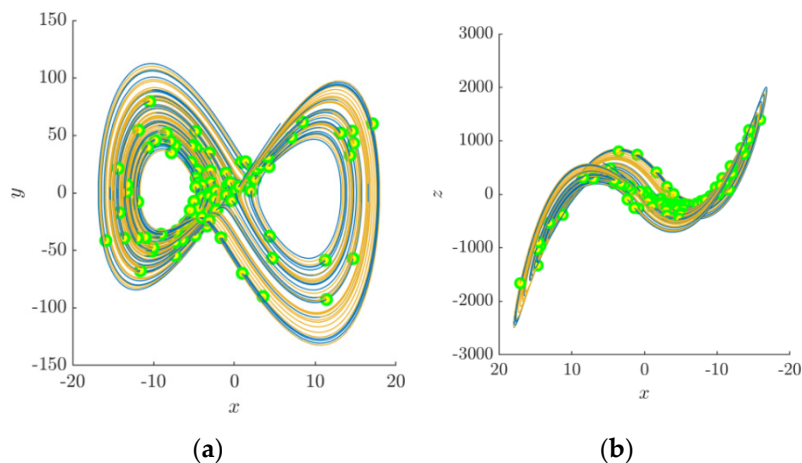
Figure 5. The trajectory of the modified Lorenz system Equation (37) is projected onto (**a**) *x-y* plane and (**b**) *x-z* plane. The green dots denote 95 data points, the blue line depicts the original trajectory, and the orange line depicts a trajectory of the reconstructed system.

Truncation and round-off errors can be considered as a sort of noise [33] which makes it possible to compare differential embedding and integral embedding in terms of sensitivity to the reconstruction error induced by the stepsize. The results are given in Figure 6.
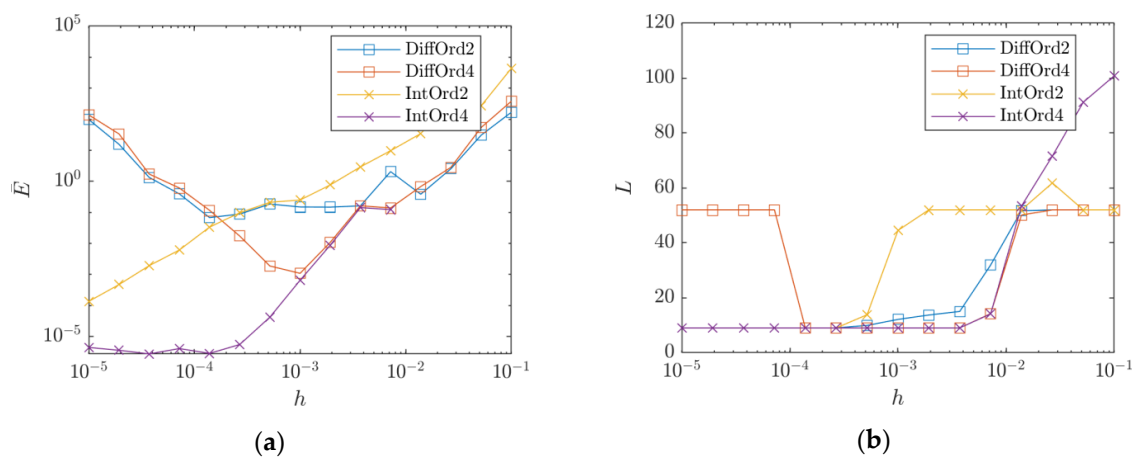
Figure 6. Results of the reconstruction of system Equation (37) using integration and differentiation of orders 2 and 4 with respect to the stepsize. (**a**) mean error $\overline{E} = \|\mathbf{v}_k - H_k(\mathbb{X})\|$ by all 3 dimensions; (**b**) overall number of terms L in the reconstruction, where $L = 9$ corresponds to the original system (37). Square markers correspond to the differentiation-based approach; crosses correspond to the integration-based approach. We tested 15 stepsizes, 10 experiments were carried out with each stepsize.

In Figure 6, we denote the phase space reconstruction method based on differentiation of the 2nd order of accuracy as *DiffOrd2* (central differences) and of the 4th order of accuracy as *DiffOrd4* (4 order finite differences) [34]. The method based on of 2nd order integration is denoted as *IntOrd2* (trapezoidal rule) and the method of 4th order of accuracy is denoted as *IntOrd4* (Simpson's rule). Two plots summarize the results on mean error $E = \|\mathbf{v}_k - H_k(\mathbb{X})\|$ across three state variables and overall number of terms L.

We see from both plots that for small stepsizes, the integration based reconstruction method performs well while the differentiation approach suffers from numerical noise following with the proof of Theorem 2. The higher the sampling frequency is, the more noise is added into the data and the algorithm is more likely to fail. However, large stepsizes are more likely to corrupt the integration-based

approach, but in this case this happens at $h > 10^{-2}$ when both methods give inaccurate results even on 4th order. The reconstruction techniques based on second-order formulae became inaccurate at $h > 2.5 \cdot 10^{-4}$, the fourth-order based formulae became inaccurate at $h > 3.5 \cdot 10^{-3}$. By "inaccurate" we mean that Equation (37) was reconstructed with some missing or excessive terms.

3.3. Reconstruction of Muthuswamy's Memristive Circuit

Replacing the nonlinear negative resistor (Chua diode) in the well-known Chua's circuit with the flux-controlled memristor one obtains a circuit able to exhibit the chaotic behavior [19]. The circuit proposed by B. Muthuswamy includes five elements (Figure 7): a linear passive inductor, two linear passive capacitors, a linear passive resistor and a nonlinear active memristor.
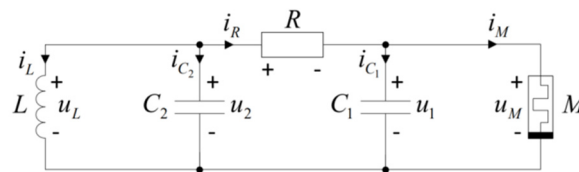
Figure 7. Schematics of the five-element chaotic circuit [19].

After scaling and transformation, we obtain the following system of ODEs:

$$\begin{cases} \dot{x}_1 = p_1 x_2, \\ \dot{x}_2 = p_2(x_3 - x_2) - p_3(\alpha + 3\beta x_1^2)x_2, \\ \dot{x}_3 = p_4(x_2 - x_3) - p_5 x_4, \\ \dot{x}_4 = p_6 x_3. \end{cases} \tag{39}$$

Parameters for chaotic behavior are $\alpha = -0.667 \cdot 10^{-3}$; $\beta = 0.029 \cdot 10^{-3}$; $p_1 = 2.594 \cdot 10^{-3}$; $p_2 = 73.53 \cdot 10^3$; $p_3 = 147 \cdot 10^6$; $p_4 = 7.353 \cdot 10^3$; $p_5 = 14.7 \cdot 10^6$; $p_6 = 55.55$. We simulated the system by the function *ode113* with parameters *RelTol* $= 10^{-13}$ and *AbsTol* $= 10^{-15}$ and constant stepsize $h = 10^{-8}$ and final time $T = 0.01$. In this experiment we used data obtained from the original Equation (39) and reconstructed the state variable x_1 by integrating variable x_2. This is motivated by two facts: first, variable x_1 stands for the magnetic flux between the terminals of the memristive device [19], and therefore cannot be measured on a macroscopic scale, so it is assumed to be unobservable. Second, integrating x_2 is very natural to the system (39) due to the relation $\dot{x}_1 = p_1 x_2$ and therefore, needs no complicated transformations.

The degrees of terms in the monomials were placed within the interval $\mathcal{D} = [0; 3]$, the tolerance parameter of AMB algorithm was $\epsilon = 10^{-2}$. The reconstructed equations are

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = 24519\, x_2 + 73530\, x_3 - 0.086055\, x_1^2 x_2, \\ \dot{x}_3 = 7353\, x_2 - 7353\, x_3 - 14700000\, x_4, \\ \dot{x}_4 = 55.55\, x_3. \end{cases} \tag{40}$$

Note, that the Equations in (40) are a rescaled version of the Equations in (39) when $\hat{p}_1 = 1$ instead of p_1 in the original equation.

Two projections of the attractor of original and reconstructed systems are given in Figure 8. The initial conditions were $(0, 0.1, 0.1, 0)^T$. We compared three integration methods by changing the tolerance parameter η in the *delMinorTerms* algorithm in the range $\eta \in [10^{-3}, 10^4]$. Figure 9 depicts the result of the experiment. Using higher-order integration formulae allows for the achievement of more precise and more compact solutions. This experiment shows the difficulty of choosing η: its low values force the algorithm to generate equations with many excessive terms making the least-squares solution more precise. But, as one can see in Figure 9, the accuracy of the proper reconstruction is rather high

due to high absolute values of the state variables (see Figure 6) and thus truncation and numerical errors in the reconstructed data for x_1 are large as well.
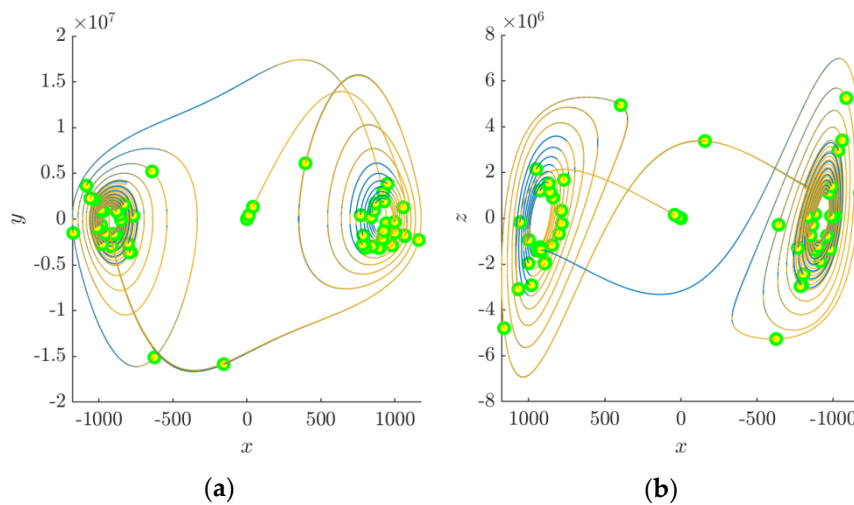

Figure 8. The trajectory of the system Equations in (40) projected onto (**a**) *x-y* plane and (**b**) *x-z* plane. The green Dots denote 53 data points, the blue line depicts the original trajectory, and the orange line depicts a trajectory of the reconstructed system scaled to fit the original trajectory. The equations were reconstructed with high accuracy.
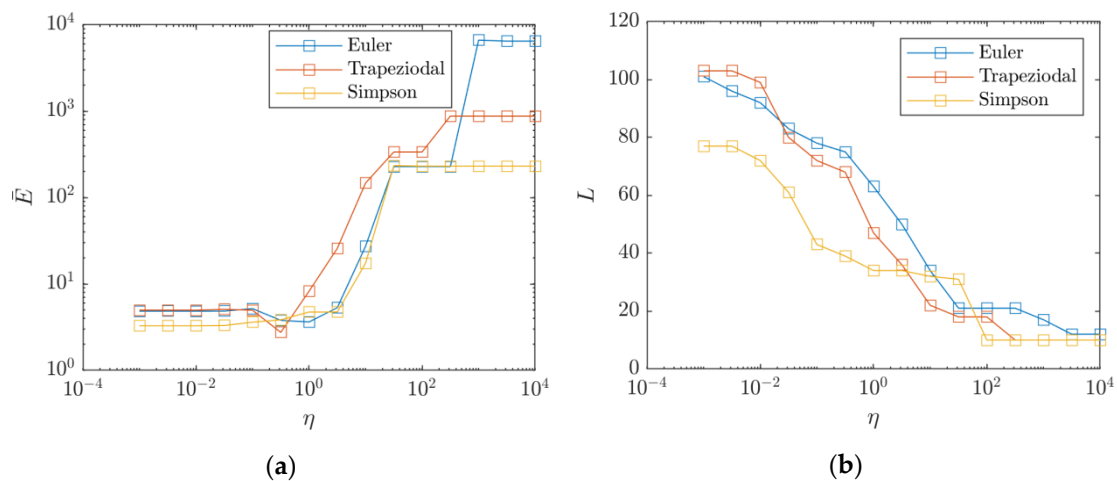
Figure 9. Results of the reconstruction of system Equations in (39) using integration of orders 1, 2 and 4 (Euler, Trapezoidal and Simpson's respectively) in accordance to the *delMinorTerms* tolerance. (**a**) Mean error $\bar{E}$ by 4 dimensions; (**b**) overall number of terms L in the reconstruction, where $L = 8$ corresponds to the original system Equations in (39). We tested 15 values of η, 10 experiments were carried out with each value.

To summarize, the experiment shows that higher-order integration is more preferable and value η should be adjusted with respect to the problem properties.

4. Discussion and Conclusions

Two main findings are highlighted in this paper. First, we show that the integral embedding approach has several advantages over the differential approach often used in cases of incompletely observed systems. Second, we extend the Kera–Hasegawa method to Laurent polynomials and show by example its performance.

Still, some limitations of this method exist. The first, and the most prominent one being, is that not each system can be represented as a PDS, and furthermore, there are some PDSs that do not allow transformation to the form of Equation (22) or Equation (24) or their combinations.

Consider the following system with an unobservable variable z:

$$\begin{cases} \dot{x} = y, \\ \dot{y} = -cx + by - byz^2, \\ \dot{z} = -y - az + yz, \end{cases} \tag{41}$$

where a, b, c are parameters. The natural way to reconstruct Equations (41) in this case is to transform it into

$$\begin{cases} \dot{x} = y, \\ \dot{y} = u, \\ \dot{u} = -cy + bu - \frac{u(by - cx - u)}{y} \pm 2by \sqrt{\frac{by - cx - u}{by}} \left(-y \pm (a - y) \sqrt{\frac{by - cx - u}{by}} \right), \end{cases} \tag{42}$$

where a new variable u is reconstructed from the observable variable y through differentiation. Equation (42) cannot be represented in terms of polynomial functions and, moreover, needs a special technique to resolve the ambiguity in $\pm$ sign appearing when z variable is expressed from the second Equation of (41).

The second limitation of the proposed reconstruction method can be exemplified by Muthuswamy's memristive circuit. In many cases, the algorithm found a PDS different from Equations in (40), but which is also correct on limited data series $\mathbb{X}$, $\mathbb{V}$ and has a stable attractor similar to the attractor of the system Equations in (40):

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = 24519\, x_2 \; + \; 73530\, x_3 \; - \; 0.086055\, x_1^2\, x_2, \\ \dot{x}_3 = \; 35748530.2895\, x_1 + \; 5895.007\, x_2 - \; 7353\, x_3 - \; 12770094.9573\, x_4 - \; 41.8226\, x_1^3, \\ \dot{x}_4 = 55.55\, x_3. \end{cases} \tag{43}$$

Moreover, the third Equation in (43) is parametric, so infinitesimally many variants of it exist. So, if similar data can be produced by several PDSs, the method may find any one of them.

The third limitation of the proposed reconstruction method was found in both experiments with the Lorenz system and Muthuswamy's memristive circuit, i.e., it is sensitive to algorithm parameters, such as stepsize, the order of accuracy of the reconstruction method, its type (differentiation or integration), and tolerance of *delMinorTerms* algorithm η. Obviously, sensitivity to the AMB algorithm tolerance ϵ also takes place, but we did not investigate its influence.

The fourth limitation is that the dimension of search grows dramatically when the number of terms or system dimension grows. For example, four-dimensional PDS in complete Laurent polynomials containing degrees in $\mathcal{D} \in [-3, 3]$ can be constructed out of 471 unique monomials, five-dimensional PDS out of 1281 unique monomials, and six-dimesional PDS out of 3067 unique monomials. Even if ABM algorithm sufficiently reduces these numbers, it is likely to obtain very poorly conditioned large-scale regression problems, and it is not obvious whether the algorithm remains useful in these cases. This question needs further investigation.

Laurent monomial ordering, *delgexord,* and BM algorithm for Laurent monomials also need the development of a stricter theoretical basement, as in the case of well-established ordinary monomials [35]. Moreover, from Equation (4) it is clear that the reconstruction approach based on the LSM can also handle not only PDSs but functions containing other elementary nonlinearities as well (e.g., trigonometric functions). This also needs further development.

References

1. Laubenbacher, R.; Stigler, B.A. Computational algebra approach to the reverse engineering of gene regulatory networks. *J. Theor. Biol.* **2004**, *229*, 523–537. [CrossRef] [PubMed]

2. Letellier, C.; Maquet, J.; Labro, H.; Sceller, L.L.; Gouesbet, G.; Argoul, F.; Arnéodo, A. Analyzing Chaotic Behavior in a Belousov− Zhabotinskyi Reaction by Using a Global Vector Field Reconstruction. *J. Phys. Chem. A* **1998**, *102*, 10265–10273. [CrossRef]

3. Akman, O.E.; Watterson, S.; Parton, A.; Binns, N.; Millar, A.J.; Ghazal, P. Digital clocks: Simple Boolean models can quantitatively describe circadian systems. *J. R. Soc. Interface* **2012**, *9*, 2365. [CrossRef] [PubMed]

4. Gerhard, F.; Kispersky, T.; Gutierrez, G.J.; Marder, E.; Kramer, M.; Eden, U. Successful reconstruction of a physiological circuit with known connectivity from spiking activity alone. *PLoS Comput. Biol.* **2013**, *9*, e1003138. [CrossRef]

5. Bansal, M.; Della, G.G.; Bernardo, D.D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics* **2006**, *22*, 815. [CrossRef]

6. Li, J.; Zhang, X.S. An optimization model for gene regulatory network reconstruction with known biological information. *Optim. Syst. Biol.* **2007**, *7*, 35.

7. Sprott, J.C. Some simple chaotic flows. *Phys. Rev. E* **1994**, *50*, R647. [CrossRef]

8. Crutchfield, J.P.; McNamara, B.S. Equations of motion from a data series. *Complex Syst.* **1987**, *1*, 417–452.

9. Cremers, J.; Hübler, A. Construction of differential equations from experimental data. *Z. Für Nat. A* **1987**, *42*, 797–802.

10. Breeden, J.L.; Dinkelacker, F.; Hübler, A. Noise in the modeling and control of dynamical systems. *Phys. Rev. A* **1990**, *42*, 5827–5836. [CrossRef]

11. Aguirre, L.A.; Letellier, C. Modeling nonlinear dynamics and chaos: A review. *Math. Probl. Eng.* **2009**, *2009*, 238960. [CrossRef]

12. Iba, H. Inference of differential equation models by genetic programming. *Inf. Sci.* **2008**, *178*, 4453. [CrossRef]

13. Searson, D.P.; Leahy, D.E.; Willis, M.J. GPTIPS: An open source genetic programming toolbox for multigene symbolic regression. In Proceedings of the International multiconference of engineers and computer scientists, Hong Kong, 17–19 March 2010; Volume 1, pp. 77–80.

14. Yogatama, D.; Smith, N. Making the most of bag of words: Sentence regularization with alternating direction method of multipliers. In Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 21–26 June 2014; Volume 32, p. 656.

15. Kera, H.; Hasegawa, Y. Noise-tolerant algebraic method for reconstruction of nonlinear dynamical systems. *Nonlinear Dyn.* **2016**, *85*, 675–692. [CrossRef]

16. Linn, E.; Siemon, A.; Waser, R.; Menzel, S. Applicability of well-established memristive models for simulations of resistive switching devices. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2014**, *61*, 2402–2410. [CrossRef]

17. Abarbanel, H.D.; Brown, R.; Sidorowich, J.J.; Tsimring, L.S. The analysis of observed chaotic data in physical systems. *Rev. Mod. Phys.* **1993**, *65*, 1331. [CrossRef]

18. Karimov, T.; Butusov, D.; Andreev, V.; Karimov, A.; Tutueva, A. Accurate synchronization of digital and analog chaotic systems by parameters re-identification. *Electronics* **2018**, *7*, 123. [CrossRef]

19. Muthuswamy, B. Implementing memristor based chaotic circuit. *Int. J. Bifurc. Chaos* **2010**, *20*, 1335–1350. [CrossRef]

20. Nesterov, Y. A Method of Solving a Convex Programming Problem with Convergence Rate O(1/k2). In *Sov. Math. Dokl*; USSR: Moscow, Russia, 1983; Volume 27, pp. 372–376.

21. Shi, Y.; Zhu, X.X.; Yin, W.; Bamler, R. A fast and accurate basis pursuit denoising algorithm with application to super-resolving tomographic SAR. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6148–6158. [CrossRef]

22. Tikhonov, A.N.; Arsenin, V.Y. *Solution of Ill-posed Problems*; Winston & Sons: Hoboken, NJ, USA, 1977.

23. Takens, F. *Detecting Strange Attractors in Turbulence. Lecture Notes in Mathematics*; Springer Science and Business Media: Berlin/Heidelberg, Germany, 1980; pp. 366–381.

24. Packard, N.H.; Crutchfield, J.P.; Farmer, J.D.; Shaw, R.S. Geometry from a time series. *Phys. Rev. Lett.* **1980**, *45*, 712–716. [CrossRef]

25. Lekscha, J.; Donner, R.V. Phase space reconstruction for non-uniformly sampled noisy time series. *Chaos: Interdiscip. J. Nonlinear Sci.* **2018**, *28*, 085702. [CrossRef]

26. Kennel, M.B.; Brown, R.; Abarbanel, H.D. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **1992**, *45*, 3403. [CrossRef] [PubMed]

27. Rhodes, C.; Morari, M. The false nearest neighbors algorithm: An overview. *Comput. Chem. Eng.* **1997**, *21*, S1149–S1154. [CrossRef]

28. Bradley, E.; Kantz, H. Nonlinear time-series analysis revisited. *Chaos: Interdiscip. J. Nonlinear Sci.* **2015**, *25*, 097610. [CrossRef] [PubMed]

29. Oppenheim, A.V.; Buck, J.R.; Schafer, R.W. *Discrete-Time Signal Processing*; Prentice Hall: Upper Saddle River, NJ, USA, 2001; Volume 2.

30. Goldberger, A.L.; Amaral, L.A.N.; Glass, L.; Hausdorff, J.M.; Ivanov, P.Ch.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.-K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **2003**, *101*, e215–e220. [CrossRef]

31. Lugovaya, T.S. Biometric Human Identification Based on Electrocardiogram. Master's Thesis, Faculty of Computing Technologies and Informatics, Electrotechnical University "LETI", Saint-Petersburg, Russia, June 2005.

32. Leonov, G.A.; Pogromsky, A.Y.; Starkov, K.E. The dimension formula for the Lorenz attractor. *Phys. Lett. A* **2011**, *375*, 1179–1182. [CrossRef]

33. Hinamoto, T.; Lu, W.S. Digital Filter Design and Realization. *River Publ.* **2017**, 384.

34. Jordan, C.; Jordán, K. Calculus of finite differences. *Am. Math. Soc.* **1965**, *33*, 15. [CrossRef]

35. Rahkooy, H.; Zafeirakopoulos, Z. Using resultants for inductive Gröbner bases computation. *Acm Comm. Comput. Algebra* **2011**, *45*, 135–136. [CrossRef]