*International Journal of Innovative Computing, Information and Control*
Volume **12**, Number **5**, October **2016**
ICIC International ©2016 ISSN 1349-4198
pp. 1703–1719

# DISCRETE AND EXACT GENERAL SOLUTION FOR NONLINEAR AUTONOMOUS ORDINARY DIFFERENTIAL EQUATIONS

Paulo Marcelo Tasinaffo, Rubens dos Santos Guimarães
Luiz Alberto Vieira Dias, Válter Strafacci Júnior
and Felipe Rafael Motta Cardoso

Computer Science Division
Technological Institute of Aeronautics
Praça Marechal Eduardo Gomes, 50 Vila das Acácias, São José dos Campos/SP 12228-900, Brazil
tasinaffo@ita.br; rubens.guimaraes@uol.com.br; l_vdias@yahoo.com.br
{ vstrafacci; felipe.reeves07 }@gmail.com

Received November 2015; revised August 2016

Abstract. *A discrete general solution and exact for nonlinear autonomous ordinary differential equations is proposed in this paper. Seemingly simple, the solution is based on Euler integrator. However, we used mean derivative functions instead of instantaneous derivative functions on this integrator. Empirically determining the mean derivative functions is not a trivial task. Therefore, it is necessary to solve a nonlinear parameter estimation problem by using some universal approximator of functions such as an artificial neural network or a fuzzy inference system. This article proposes to mathematically demonstrate these facts.*

**Keywords:** Autonomous nonlinear ordinary differential equations, Neural numerical integrators, Neural networks, Mamdani-type fuzzy inference systems

1. **Introduction.** Informally, one can say that there is an exact and accurate relationship between the continuous space and the discrete space, when the context of ordinary differential equations system (ODES) is only considered. An informal representation for this is that there is a precise and exact transformation $\Psi[\,.\,]$ between the continuous and the discrete space that can be represented by $\Psi[\dot{y} = f(y)] \rightarrow {}^{k+1}y^i = \tan_{\Delta t} {}^k\alpha^i \cdot \Delta t + {}^k y^i$, where the mean derivative $\tan_{\Delta t} {}^k\alpha_j^i$ is defined by $\tan_{\Delta t} {}^k\alpha_j^i = \frac{{}^{k+1}y_j^i - {}^k y_j^i}{\Delta t}$, for $j = 1, 2, \ldots, n$. The inverse of this transformation is also valid, that is, there is an inverse transformation $\Psi^{-1}[\,.\,]$ which relates the discrete universe to the continuous one given by $\Psi^{-1}\left[{}^{k+1}y^i = \tan_{\Delta t} {}^k\alpha^i \cdot \Delta t + {}^k y^i\right] \rightarrow \dot{y} = f(y)$. The inverse transformation $\Psi^{-1}[\,.\,]$ states that for any discrete sequence of points in the discrete space, there is a continuous instantaneous derivative function in the continuous space (existence theorem). Initially, this article aims to formally prove these facts. However, it remains how to determine in practice the mean derivative functions. A brief description of how to perform this calculation is also suggested in the following. It can be obtained by using tools used in artificial intelligence. So, one can see that the problem of empirically determining the mean derivative functions may be reduced to a nonlinear parameter estimation problem involving supervised learning using input/output training patterns.

In the past few decades, artificial intelligence has received considerable attention due to the extensive applications in signal processing, target tracking, optimization, pattern recognition, and associative memories [23-26]. Three of the most successful approaches are artificial neural networks, fuzzy logic and genetic algorithm that allow, for example, the recognition of writing, image processing, modeling of nonlinear dynamic systems, and applications in control theory. A good introduction to artificial neural networks theory can be found in [5,14,22]. An extremely important starting point in the study of artificial neural networks or fuzzy inference systems is that they are considered universal approximator of functions [3,4,6,8,16]. Thus, the artificial neural networks (ANN) have long been used in the modeling of nonlinear dynamical systems in recent decades, as they have a high ability to approximate nonlinear mappings. Several studies have been developed in this area by using NARMAX methodology (Non linear Autoregressive Moving Average with eXogenous inputs) with a subsequent application in control [2,7,11,12]. A brief description of NARMAX methodology can be found in Appendix B. In addition to this, other methodologies have also been developed in recent years to represent nonlinear dynamical systems, including the instantaneous derivative methodology [9,13,20] and the mean derivative methodology [17-19]. Although there are these three methodologies to represent nonlinear dynamical systems by using artificial neural networks or any other universal approximator for functions, this article focuses only on a formal and accurate description from a mathematical perspective of the mean derivative methodology for the empirical determination of functions given by $\tan_{\Delta t}{}^k\alpha_j^i$.

From this, this paper provides one original contribution. The major is to offer a theoretical basement for the method named mean derivative, which informally appears in [18,19] and can be applied to model linear dynamic systems and non-linear ruled by autonomous ordinary differential equations. Sections 2 and 3 exclusively dedicate to this task.

The methodology presented here is an alternative method to NARMAX methodology, which can be found and described in detail in [2,7]. Nevertheless, in Appendix B of this paper, the NARMAX methodology is explained in a summarized way. Therefore, it can be stated that there are three empirical methodologies for representing dynamic systems: NARMAX methodology, the methodology that uses instantaneous derivative [9,20] and the one which utilizes mean derivative. One meaningful distinction among these three methodologies is that the NARMAX methodology and mean derivative have a fixed integration step, while the instantaneous derivative methodology has a variable step. This is a meaningful distinction because the fixed step demands a new supervised learning training whether the integration step is modified, while the method of variable step does not.

## 2. Theoretical Development.

Being the autonomous system of nonlinear ordinary differential equations,

$$\dot{y} = f(y) \tag{1.a}$$

where,

$$y = [y_1 \;\; y_2 \;\; \cdots \;\; y_n]^T \tag{1.b}$$

$$f(y) = [f_1(y) \;\; f_2(y) \;\; \cdots \;\; f_n(y)]^T \tag{1.c}$$

Consider also by definition, $y_j^i = y_j^i(t)$ for $j = 1, 2, \ldots, n$ a trajectory of the family of solutions of nonlinear differential equations system $\dot{y} = f(y)$ going past $y_j^i(t_o)$ at the initial time $t_o$, starting from a domain of interest $\left[y_j^{\min}(t_o), y_j^{\max}(t_o)\right]^n$, where $y_j^{\min}(t_o)$ and $y_j^{\max}(t_o)$ are finite. If one can say that there are initial conditions for all interior points of this domain, it means that for any discretization $i$, $y_j^i(t_o)$ there is a possible set of initial conditions in $t_o$ of the nonlinear differential equations given by (1.a) within a domain of interest in $\left[y_j^{\min}, y_j^{\max}\right]^n$ for $j = 1, 2, \ldots, n$.

It is appropriate to introduce the following vector notation concerning the possible sets of initial conditions and solutions of (1.a):

$$y_o^i = y^i(t_o) = \left[y_1^i(t_o) \;\; y_2^i(t_o) \;\; \cdots \;\; y_n^i(t_o)\right]^T \tag{2.a}$$

$$y^i = y^i(t) = \begin{bmatrix} y_1^i(t) & y_2^i(t) & \cdots & y_n^i(t) \end{bmatrix}^T \tag{2.b}$$

where $i = 1, 2, \ldots, \infty$; and $\infty$ indicates that the number of points in the discretization loop can be as big as one wants. Figure 1 is an illustrative sketch for the one-dimensional case in the variable $j$ of a family of solutions $\dot{y}_1 = f(y_1)$ that parts from the initial conditions $y_1^i(t_o)$ contained within the interest domain $\left[y_j^{\min}(t_o), y_j^{\max}(t_o)\right]^1$.

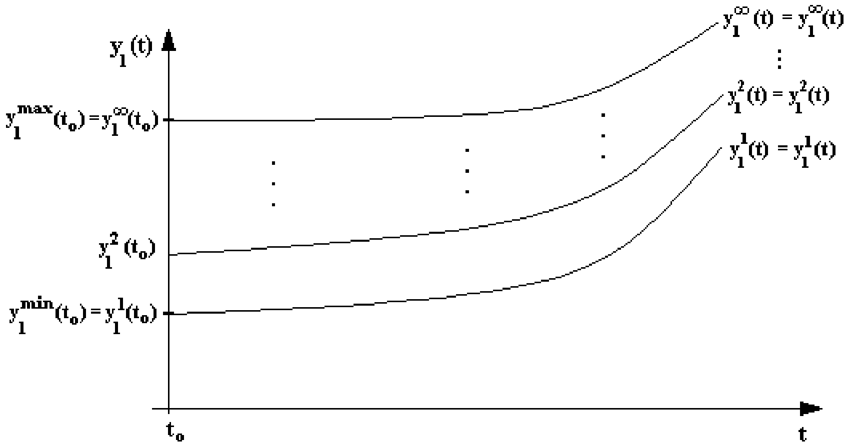


FIGURE 1. Family of curves $y_1^i = y_1^i(t)$ solutions of $\dot{y}_1 = f(y_1)$ of the interest domain $\left[y_1^{\min}(t_o), y_1^{\max}(t_o)\right]^1$ in $t_o$

It is important to demonstrate two important results (see [1]) regarding the differential equations solutions system (1.a). The first relates to the existence and oneness of solutions, and the second deals with the existence of stationary solutions of (1.a).

**Theorem 2.1 (T1).** *Assume that each of the functions $f_1(y_1, y_2, \ldots, y_n), \ldots, f_n(y_1, y_2, \ldots, y_n)$ has continuous partial derivative with respect to $y_1, \ldots, y_n$. Then, the initial value problem $\dot{y} = f(y)$, $y(t_o)$ in the interest domain $\left[y_j^{\min}, y_j^{\max}\right]^n$ to $j = 1, 2, \ldots, n$, in $t_o$, there is one and only solution $y^i = y^i(t)$ in $R^n$, from each $y^i(t_o)$. If both solutions $y = \phi(t)$ and $y = \varphi(t)$ have a common point, they should be identical.*

**Property 2.1 (P1).** *If $y = \phi(t)$ is a solution of (1.a), $y = \phi(t + c)$ is also a solution of (1.a), being **c** any real constant.*

P1 is not true if the function (1.a) explicitly depends on time [1], because in this case, assuming that $y = \phi(t + c)$ is also a solution of the non-autonomous system $\dot{y} = f(t, y)$ implies another differential equation of the type $\dot{y} = f(t + c, y)$, which leads to a contradiction.

As, in general, $\dot{y}^i = f(y^i)$ does not have analytical solution, it is common to know only for $y^i = y^i(t)$ a discrete set of points $[y^i(t + k \cdot \Delta t) \;\; y^i[t + (k+1) \cdot \Delta t] \;\; \ldots \;\; y^i[t + (k + n) \cdot \Delta t]] \equiv \left[{}^k y^i \;\; {}^{k+1}y^i \;\; \ldots \;\; {}^{k+n}y^i\right]$ for ${}^k y^i = y^i(t + k \cdot \Delta t)$ on the horizon $[t_k, t_{k+n}]$ and $\Delta t = (t_{k+n} - t_k)/n$ a constant.

**Property 2.2 (P2).** *As $\dot{y}^i = f(y^i)$ is given by (1.a), it immediately follows that, if ${}^k y^i = y^i(t + k \cdot \Delta t)$ is known, ${}^k \dot{y}^i = \dot{y}^i(t + k \cdot \Delta t)$ will also be.*

This property, although quite obvious, will be very useful when applied in the differential mean value theorems and integral on the sets of points ${}^k y^i$ and ${}^k \dot{y}^i$ in the space $R^n$ to determine important properties concerning the tangent of the secant between two consecutive points ${}^k y^i$ and ${}^{k+1}y^i$ on the curve $y^i(t)$, which is a particular solution of $\dot{y} = f(y)$ that starts from $y^i(t_o)$.

By definition (see [10]), the secant between two points ${}^{k}y^{i}$ and ${}^{k+1}y^{i}$ belonging to curve $y^{i}(t)$ is the straight segment which joins these two points. So the tangents of the secants between points ${}^{k}y_{1}^{i}$ and ${}^{k+1}y_{1}^{i}$, ${}^{k}y_{2}^{i}$ and ${}^{k+1}y_{2}^{i}$, …, ${}^{k}y_{n}^{i}$ and ${}^{k+1}y_{n}^{i}$ are defined as:

$$\tan_{\Delta t}\alpha^{i}(t+k\cdot\Delta t)=\tan_{\Delta t}{}^{k}\alpha^{i}=\left[\tan_{\Delta t}{}^{k}\alpha_{1}^{i}\ \ \tan_{\Delta t}{}^{k}\alpha_{2}^{i}\ \ \ldots\ \ \tan_{\Delta t}{}^{k}\alpha_{n}^{i}\right]^{T} \tag{3.a}$$

with,

$$\tan_{\Delta t}{}^{k}\alpha_{j}^{i}=\frac{{}^{k+1}y_{j}^{i}-{}^{k}y_{j}^{i}}{\Delta t},\ \text{to } j=1,2,\ldots,n \tag{3.b}$$

where $\alpha_{j}^{i}$ is the secant angle that joins both ${}^{k}y_{j}^{i}$ and ${}^{k+1}y_{j}^{i}$ points belonging to curve $y_{j}^{i}(t)$.

**Property 2.3 (P3).** *If ${}^{k}y^{i}$ is a discretization of solution $\dot{y}^{i}=f(y^{i})$ and $\Delta t\neq 0$, $\tan_{\Delta t}{}^{k}\alpha^{i}$ exists and it is unique.*

**Proof:** T1 guarantees the existence and oneness of $y_{j}^{i}(t)$ and of ${}^{k}y_{j}^{i}$ and ${}^{k+1}y_{j}^{i}$, given a discretization $\Delta t\neq 0$, which results in the existence and oneness of $\tan_{\Delta t}{}^{k}\alpha^{i}$.

**Property 2.4 (P4).** *Given the state vector $y^{i}(t_{o})$ at $t_{o}$ instant of the non-linear differential equations $\dot{y}=f(y)$ and if the vector $\left[{}^{k}y^{i}\ \ {}^{k+1}y^{i}\ \ \ldots\ \ {}^{k+n}y^{i}\right]$ is the discretization of a trajectory solution so that for ${}^{k}y^{i}=y^{i}(t+k\cdot\Delta t)$ it is a solution of the dynamic system, then $\tan_{\Delta t}{}^{k+l}\alpha^{i}$ for $l=0,1,\ldots,(n-1)$ exist and are unique.*

**Proof:** By successive application and induction of P3, $\tan_{\Delta t}{}^{k+l}\alpha^{i}$ for $l=0,1,\ldots,(n-1)$ exist and are unique.

Two other very important theorems that link the values of $\tan_{\Delta t}{}^{k}\alpha^{i}$ and $\tan_{\Delta t}{}^{k}\dot{\alpha}^{i}$, respectively, with the mean derivative calculated from $\left[{}^{k}y^{i}\ {}^{k+1}y^{i}\ \ldots\ {}^{k+n}y^{i}\right]$ and $\left[{}^{k}\dot{y}^{i}\ {}^{k+1}\dot{y}^{i}\ \ldots\ {}^{k+n}\dot{y}^{i}\right]$ are the differential and integral mean value theorems [10,15,21] which are set out below without proof.

**Theorem 2.2 (T2).** *(The differential mean value theorem). If the function $y_{j}^{i}(t)$ for $j=1,2,\ldots,n$ is defined and continuous function on the closed interval $[t_{k},t_{k+1}]$ and differentiable on the open interval $(t_{k},t_{k+1})$, there is at least one number $t_{k}^{*}$ with $t_{k}<t_{k}^{*}<t_{k}+\Delta t$ such that*

$$\dot{y}_{j}^{i}\left(t_{k}^{*}\right)=\frac{{}^{k+1}y_{j}^{i}-{}^{k}y_{j}^{i}}{\Delta t} \tag{4}$$

Theorem T2 states [10,21] that given a secant to the graph $y^{i}(t)$ of a differentiable curve, one can always find a graph point between the two points ${}^{k+1}y^{i}$ and ${}^{k}y^{i}$ of the intersection of the secant with the curve $y^{i}(t)$ in $t_{k}$ and $t_{k+1}$, such that the tangent straight line to the point $y^{i}(t_{k}^{*})$ is parallel to the secant. This interesting geometric property of $\dot{y}^{i}(t_{k}^{*})$ is so called *mean derivative of the function* $y^{i}(t)$ on the closed interval $[t_{k},t_{k+1}]$.

**Theorem 2.3 (T3).** *(The integral mean value theorem). If a function $y_{j}^{i}(t)$ for $j=1,2,\ldots,n$ is a continuous function on the closed interval $[t_{k},t_{k+1}]$, then there is at least one $t_{k}^{x}$ inner number to $[t_{k},t_{k+1}]$ such that*

$$y\left(t_{k}^{x}\right)\cdot\Delta t=\int_{t_{k}}^{t_{k+1}}y^{i}(t)\cdot dt \tag{5}$$

Generally, $t_{k}^{*}$ is different from $t_{k}^{x}$ and it is also important to observe that the mean values theorems say nothing on how to determine the values $t_{k}^{*}$ and $t_{k}^{x}$. The theorems simply state that $t_{k}^{*}$ and $t_{k}^{x}$ are contained in the interval $[t_{k},t_{k+1}]$.

**Property 2.5 (P5).** *Applying T3 theorem on curve $\dot{y}^{i}(t)$ is equivalent to apply T2 theorem on the curve $y^{i}(t)$ both on the same closed interval $[t_{k},t_{k+1}]$, that is, $\dot{y}^{i}(t_{k}^{x})=\dot{y}^{i}(t_{k}^{*})$.*

**Proof:** From T3 theorem applied to the curve $\dot{y}^i(t)$ it results a $\dot{y}^i(t_k^x)$ for $t_k < t_k^x < t_{k+1}$ such that $\dot{y}^i(t_k^x) \cdot \Delta t = \int_{t_k}^{t_{k+1}} \dot{y}^i(t) \cdot dt = {}^{k+1}y^i - {}^{k}y^i$ by the fundamental theorem of calculus. Thus, $\dot{y}^i(t_k^x) = \frac{{}^{k+1}y^i - {}^{k}y^i}{\Delta t} = \tan_{\Delta t} {}^{k}\alpha^i$. On the other hand, the application of T2 theorem on the curve $y^i(t)$ implies that there is a $\dot{y}^i(t_k^*)$ for $t_k < t_k^* < t_{k+1}$, so that $\dot{y}^i(t_k^*) = \frac{{}^{k+1}y^i - {}^{k}y^i}{\Delta t} = \tan_{\Delta t} {}^{k}\alpha^i$. Thus, $\dot{y}^i(t_k^x) = \dot{y}^i(t_k^*)$.

Nevertheless, this theorem does not allow one to state that $t_k^x = t_k^*$.

**Property 2.6 (P6).** *The mean derivative $\dot{y}^i(t_k^x)$ on the graph of $y^i(t)$ in the closed interval $[t_k, t_{k+1}]$ is equal to $\tan_{\Delta t} {}^{k}\alpha^i$, as an immediate consequence of the definition itself of the mean derivative.*

**Property 2.7 (P7).** *Note that if $\dot{y}^i(t_k^x) = \dot{y}^i(t_k^*)$ (P5) and $\dot{y}^i(t_k^*)$ is the mean derivative of $y^i(t)$ (T2 theorem) in the closed interval $[t_k, t_{k+1}]$, $\dot{y}^i(t_k^x)$ will also be numerically equal to the mean derivative of $y^i(t)$ for the same close interval.*

It justifies why the index $x$ was used in $t_k^x$ in the formulation of the previous theorem instead of the index $*$. Thus, following the data adopted in this work, $\ddot{y}^i(t_k^*)$ is the application of T2 theorem on the curve $\dot{y}^i(t)$, $\dddot{y}^i(t_k^*)$ is the application of T2 theorem on the curve $\ddot{y}^i(t)$ and so on. On the other hand, $\ddot{y}^i(t_k^x)$ is the application of T3 theorem on the curve $\ddot{y}^i(t)$ and so on, that is, the $*$ index is always associated to the T2 theorem application and the $x$ index is always associated to the application of T3 theorem. The application of T2 theorem on the curve $\dot{y}^i(t)$ implies that there is at least one $t_k^*$ in $[t_k, t_{k+1}]$, so that $\ddot{y}(t_k^*) = \tan_{\Delta t} {}^{k}\dot{\alpha}^i$.

Numerical and geometric equivalences between $\dot{y}^i(t_k^x)$ and $\dot{y}^i(t_k^*)$ can be interpreted as shown in Figure 2. Note that there are three possible geometric interpretations for the greatness $\dot{y}^i(t_k^x) = \dot{y}^i(t_k^*) = \tan_{\Delta t} {}^{k}\alpha^i$: the first interpretation is that it is, in fact, the mean derivative of $y^i(t)$ in the closed interval $[t_k, t_{k+1}]$, the second one is that $\Delta t \cdot \tan_{\Delta t} {}^{k}\alpha^i$ for $\Delta t = t_{k+1} - t_k$ is the area on the curve $\dot{y}^i(t)$ on the same interval $[t_k, t_{k+1}]$, and the third interpretation is that $\tan_{\Delta t} {}^{k}\alpha^i$ is, in fact, the exact derivative of at least one point within the interval $[t_k, t_{k+1}]$ of the $y^i(t)$ function, as illustrated by the two tangent straight lines

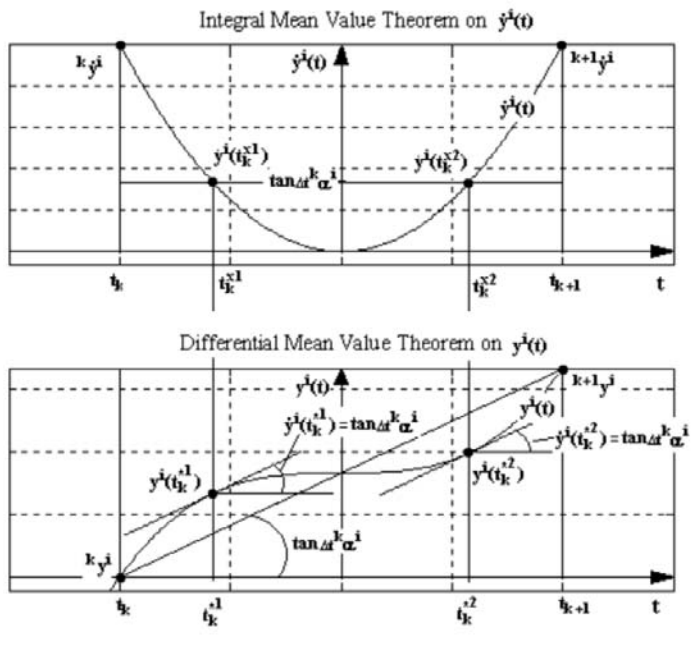


FIGURE 2. Geometrical interpretations of $\dot{y}^i(t_k^x)$ and $\dot{y}^i(t_k^*)$

on $y^i(t)$ in Figure 2. P6 suggests that if there is more than one value for $\dot{y}^i(t_k^x)$, then they should be equal to $\tan_{\Delta t}{}^k\alpha^i$.

Figure 2 illustrates this, and as it can be seen, $\dot{y}^i(t_k^{x1}) = \dot{y}^i(t_k^{x2}) = \tan_{\Delta t}{}^k\alpha^i$ even for the case where $t_k^{x1}$ is different from $t_k^{x2}$. T2 and T3 theorems can also be used to guarantee the existence of $\tan_{\Delta t}{}^k\alpha^i$, but they do not guarantee that they are unique, for that it is really required the use of T1 theorem. Figure 2 also suggests that $t_k^{x1} = t_k^{*1}$ and $t_k^{x2} = t_k^{*2}$, but this cannot be demonstrated on the basis of the theorems presented here.

**Theorem 2.4 (T4).** *The discrete general solution and exact* ${}^{k+1}y_j^i$ *for* $j = 1, 2, \ldots, n$ *of the nonlinear differential equations systems* $\dot{y}^i = f(y^i)$ *can be established through the relation* ${}^{k+1}y_j^i = \tan_{\Delta t}{}^k\alpha^i \cdot \Delta t + {}^ky_j^i$ *for a given* ${}^ky^i$ *and* $\Delta t$.

**Proof:** If $\dot{y}^i = \dfrac{dy^i}{dt} = f(y^i)$, $\int_{{}^ky^i}^{{}^{k+1}y^i} dy^i = \int_{t_k}^{t_{k+1}} f(y^i) \cdot dt$. Thus,

$$ {}^{k+1}y^i = \int_{t_k}^{t_{k+1}} f(y^i) \cdot dt + {}^ky^i \tag{6} $$

The application of the integral mean value theorem T3 on the curve $\dot{y}^i(t)$ in the closed interval $[t_k, t_{k+1}]$ implies that there is at least a number $t_k^x$ in $[t_k, t_{k+1}]$ so that

$$ \dot{y}(t_k^x) \cdot \Delta t = \int_{t_k}^{t_{k+1}} \dot{y}^i(t) \cdot dt = \int_{t_k}^{t_{k+1}} f(y^i) \cdot dt \tag{7} $$

By P6 $\dot{y}^i(t_k^x) = \tan_{\Delta t}{}^k\alpha^i$. Therefore, replacing the greatness $\dot{y}^i(t_k^x)$ in Equation (7) and this in Equation (6), results in:

$$ {}^{k+1}y^i = \tan_{\Delta t}{}^k\alpha^i \cdot \Delta t + {}^ky^i \tag{8} $$

Equation (8) is the discrete general solution and exact to the ordinary differential equations system (1.a). Note that this is a fairly simple expression. It is nothing more than Euler simple step integration structure. However, instead of using instantaneous derivative functions, it uses mean derivative functions.

**Corollary 2.1 (C1).** *The solution* ${}^{k+m}y_j^i$ *for* $j = 1, 2, \ldots, n$ *of the nonlinear differential equation system* $\dot{y}^i = f(y^i)$ *can be established by a given* ${}^ky^i$ *through the relation:*

$$ {}^{k+m}y_j^i = \sum_{l=0}^{m-1} \tan_{\Delta t}{}^{k+l}\alpha^i \cdot \Delta t + {}^ky_j^i \tag{9} $$

**Proof:** By the successive application of T4 theorem one has ${}^{k+1}y^i = \tan_{\Delta t}{}^k\alpha^i \cdot \Delta t + {}^ky^i$, ${}^{k+2}y^i = \tan_{\Delta t}{}^{k+1}\alpha^i \cdot \Delta t + {}^{k+1}y^i, \ldots, {}^{k+m}y^i = \tan_{\Delta t}{}^{k+m-1}\alpha^i \cdot \Delta t + {}^{k+m-1}y^i$. The sum of all these expressions immediately results in: ${}^{k+m}y^i = \sum_{l=0}^{m-1} \tan_{\Delta t}{}^{k+l}\alpha^i \cdot \Delta t + {}^ky^i$.

**Corollary 2.2 (C2).** *For the differential equation system* $\dot{y}^i = f(y^i)$*, the relation*

$$ \tan_{m\cdot\Delta t}{}^k\alpha_j^i = \frac{1}{m} \cdot \sum_{l=0}^{m-1} \tan_{\Delta t}{}^{k+1}\alpha_j^i $$

*for* $j = 1, 2, \ldots, n$ *is valid.*

**Proof:** From T4 theorem it immediately results in,

$$ {}^{k+m}y_j^i = \left(\tan_{m\cdot\Delta t}{}^k\alpha_j^i\right) \cdot m \cdot \Delta t + {}^ky_j^i \tag{10} $$

From Corollary C1 one has,

$$^{k+m}y_j^i = \sum_{l=0}^{m-1} \tan_{\Delta t}{}^{k+1}\alpha_j^i \cdot \Delta t + {}^k y_j^i \tag{11}$$

From the addition of equations we have,

$$\tan_{m\cdot\Delta t}{}^k\alpha_j^i = \frac{1}{m} \cdot \sum_{l=0}^{m-1} \tan_{\Delta t}{}^{k+l}\alpha_j^i \tag{12}$$

Equation (12) is graphically illustrated in Figure 3. Note that in the case in which the system (1.a) is autonomous, $y^{i_1}(t_1) = y^{i_2}(t_2)$ for $i_1 \neq i_2$ and $t_1 \neq t_2$ implies that, $\dot{y}^{i_1}(t_1) = \dot{y}^{i_2}(t_2)$. In fact, from (1.a) one obtains $\dot{y}_j^i = f(y^i)$. So,

$$\dot{y}^{i_1}(t_1) = f[y^{i_1}(t_1)] \tag{13.a}$$

$$\dot{y}^{i_2}(t_2) = f[y^{i_2}(t_2)] \tag{13.b}$$

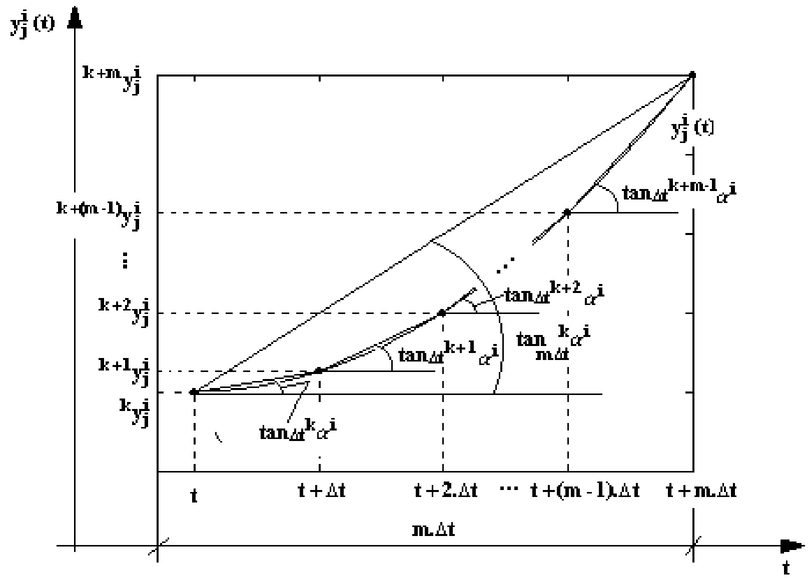


FIGURE 3. Graphical representation of the concept concerning delayed entries applied to the function $\tan_{m\cdot\Delta t}{}^k\alpha_j^i$

Hypothetically, $y^{i_1}(t_1) = y^{i_2}(t_2)$ results from (13.a) and (13.b) that $\dot{y}^{i_1}(t_1) = \dot{y}^{i_2}(t_2)$. This property states that two curves that start from two different initial conditions $y^{i_1}(t_o)$ and $y^{i_2}(t_o)$ for $i_1 \neq i_2$, from the solution family of the system $\dot{y} = f(y)$, have the same derivative if $y^{i_1}(t_1) = y^{i_2}(t_2)$ even when $t_1 \neq t_2$, that is, the system is autonomous.

P1 establishes that if $y^{i_1}(t_1) = y^{i_2}(t_2)$, $y^{i_1}(t)$ and $y^{i_2}(t)$ are in the same orbit in the phase plane, and therefore, they are just delayed solutions on the time $(t_2 - t_1)$.

Note that if $\dot{y}^{i_1}(t_1) = \dot{y}^{i_2}(t_2)$, $f[y^{i_1}(t_1)] = f[y^{i_2}(t_2)]$. In this last case, $y^{i_1}(t_1) = y^{i_2}(t_2)$ only if there is an inverse $f^{-1}(.)$ of $f(.)$ and it is unique. The question that remains now is whether the mean derivative $\tan_{\Delta t}{}^k\alpha^i$ of ${}^k y^i$ and ${}^{k+1}y^i$ are also autonomous, that is, time-invariant? To answer this question the following property is quite useful.

**Property 2.8 (P8).** *If $y^{i_1}(t)$ and $y^{i_2}(t)$ are solutions of $\dot{y} = f(y)$ which respectively arise from $y^{i_1}(t_o = 0)$ and $y^{i_2}(t_o = 0)$ and if $y^{i_1}(t_o = 0) = y^{i_2}(T)$ for $T > 0$, $y^{i_1}(\Delta t) = y^{i_2}(T + \Delta t)$ for any $\Delta t$.*

**Proof:** If $y^{i_2}(t)$ is a solution of $\dot{y} = f(y)$, $y^{i_2}(t+T)$ is also a solution by P1. Note that if $y^{i_1}(0) = y^{i_2}(T)$, $y^{i_1}(t) = y^{i_2}(t+T)$ for $t = 0$. Since $y^{i_1}(0)$ and $y^{i_2}(T)$ are solutions, for $y^{i_1}(t)$ and $y^{i_2}(t+T)$ respectively in $t = 0$, T1 theorem guarantees that they are unique for every $t$. Thus, $y^{i_1}(t) = y^{i_2}(t+T)$, and, particularly, if $t = \Delta t$, we have $y^{i_1}(\Delta t) = y^{i_2}(\Delta t + T)$. P1 warrants that if $y^{i_1}(0) = y^{i_2}(T)$ then, $y^{i_1}(t) = y^{i_2}(t+T)$ are in the same orbit of the phase plane for any $t$. In fact, the equation $y^{i_1}(t) = y^{i_2}(t+T)$ determines that the solutions $y^{i_1}(t)$ and $y^{i_2}(t+T)$ are only delayed in the time $T > 0$.

**Property 2.9 (P9).** *If $y^{i_1}(t_1) = y^{i_2}(t_2)$ for $i_1 \neq i_2$ and $t_1 \neq t_2$, then,* $\tan_{\Delta t}\alpha^{i_1}(t_1) = \tan_{\Delta t}\alpha^{i_2}(t_2)$ *for $\Delta t > 0$, that is,* $\tan_{\Delta t}{}^{k}\alpha^{i_1}$ *is time invariant.*

**Proof:** By definition, $\tan_{\Delta t}\alpha^{i_1}(t_1) = \frac{y^{i_1}(t_1+\Delta t) - y^{i_1}(t_1)}{\Delta t}$ and $\tan_{\Delta t}\alpha^{i_2}(t_2) = \frac{y^{i_2}(t_2+\Delta t) - y^{i_2}(t_2)}{\Delta t}$. Hypothetically, $y_1^{i_1}(t_1) = y_2^{i_2}(t_2)$ so $y_1^{i_1}(t_1+\Delta t) = y_2^{i_2}(t_2+\Delta t)$ and therefore, $\tan_{\Delta t}\alpha^{i_1}(t_1) = \tan_{\Delta t}\alpha^{i_2}(t_2)$.

P9 establishes that $\tan_{\Delta t}{}^{k}\alpha^{i}$ is also autonomous. This outcome is very useful because it determines that *knowing the values of* $\tan_{\Delta t}{}^{k}\alpha^{i}$ for $i = 1, 2, \ldots, \infty$ *at the initial time $t_o$* is enough for a particular interest domain $\left[y_j^{\min}, y_j^{\max}\right]^n$ for $j = 1, 2, \ldots, n$, because in the other instants $t > t_o$ *they are repeated if the solution of dynamic system does not exceed the boundaries of the n-dimensional interval of state variables* $\left[y_j^{\min}, y_j^{\max}\right]^n$ for $j = 1, 2, \ldots, n$ specified in $t_o$. Figure 4 illustrates the reasoning presented in this paragraph. One also needs to remark that when the dynamic system is propagated forward, its angle ${}^{k}\alpha(i)$ varies only in the range $-\frac{\pi}{2} < {}^{k}\alpha(i) < \frac{\pi}{2}$, and thus, it is unique. Based on this, when the dynamic system is propagated backwards, then $\frac{\pi}{2} < {}^{k}\alpha(i) < \frac{3\cdot\pi}{4}$, and therefore, the angle ${}^{k}\alpha(i)$ is also unique.

**Theorem 2.5 (T5).** *The outcome of T4 is still valid when control discretized values ${}^{k}u$ in each $[t_k, t_{k+1}]$ are used to solve the dynamic system:*

$$\dot{y}^i = f(y^i, u) \tag{14}$$

**Proof:** In fact, it is sufficient to note that in this case, the continuous function $f(y^i, u)$, with an approximated ${}^{k}u$ as a constant in each $[t_k, t_{k+1}]$ can be seen as a parameterized

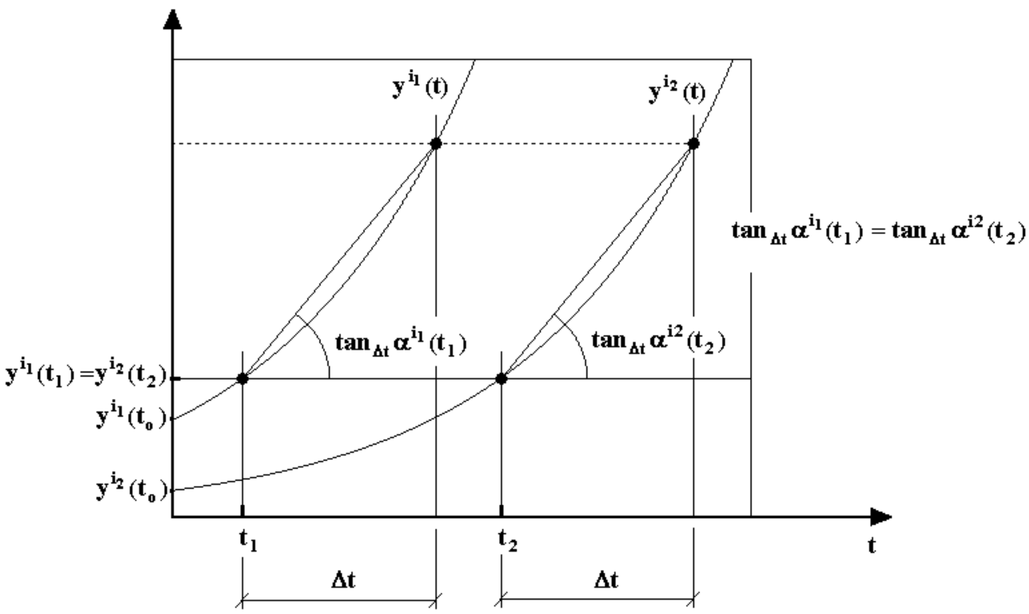


FIGURE 4. The discrete function $\tan_{\Delta t}{}^{k}\alpha^{i}$ is autonomous, that is, time invariant.

one in relation to the control variable and therefore, one can ensure that whichever is the discretization interval, the existence of mean derivative $\dot{y}^i(t_k^*) = \frac{{}^{k+1}y^i - {}^k y^i}{\Delta t} = \tan_{\Delta t}{}^k\alpha^i$ is warranted so that the result in (8) remains valid.

**Theorem 2.6 (T6).** *The definition of the secant between two distinct and any points, separated by the finite distance $\Delta t$ on the continuous curves $y_j(t)$, for $j = 1, 2, \ldots, n$ is also an exact general solution and discrete to the system of autonomous ordinary differential and nonlinear equations $\dot{y} = f[y(t)]$.*

It has been demonstrated up to this point that Euler integrator with means derivative is the general solution of nonlinear ordinary differential equations system. However, as mentioned in the introduction of this article, the reverse of this is also true. T7 theorem addresses this question.

**Theorem 2.7 (T7).** *For sequences of discrete points represented by equations ${}^{k+1}y_j^i = \tan_{\Delta t}{}^k\alpha_j^i \cdot \Delta t + {}^k y_j^i$, for $j = 1, 2, \ldots, n$, there is always an instantaneous derivative function $\dot{y} = f(y)$ associated to them.*

**Proof:** Thus, given Euler mean derivative equations:

$$ {}^{k+1}y_j^i = \tan_{\Delta t}{}^k\alpha_j^i \cdot \Delta t + {}^k y_j^i, \text{ for } j = 1, 2, \ldots, n \tag{15.a}$$

or

$$ \frac{{}^{k+1}y_j^i - {}^k y_j^i}{\Delta t} = \tan_{\Delta t}{}^k\alpha_j^i, \text{ for } j = 1, 2, \ldots, n \tag{15.b}$$

by using the differential mean value, we have $\frac{{}^{k+1}y_j^i - {}^k y_j^i}{\Delta t} = g_j'(t^*)$, for any $t^*$ in $t \le t^* < t + \Delta t$ and $j = 1, 2, \ldots, n$. By exceeding the limit $\Delta T \to 0$ in the equation involving $g_j'(t^*)$, the result is:

$$ \lim_{\Delta t \to 0} \frac{{}^{k+1}y_j^i - {}^k y_j^i}{\Delta t} = \lim_{\Delta t \to 0} g_j'(t^*) \tag{16}$$

The derivative of the left side of Equation (16) exists if a continuous and differentiable curve over the sequence of discrete points of the discrete space can be interpolated on it. This can be ensured due to the existence of universal approximators of functions (see Appendix A). Thus, Equation (16) results in $\frac{dy_j}{dt} = g_j'(t)$, for $j = 1, 2, \ldots, n$, because, in this case, when $\Delta t$ tends to zero, $t^*$ converges to $t$. Changing the $g_j'(t)$ function's name to $f_j(t)$, it follows that:

$$ \frac{dy_j}{dt} = f_j(t), \text{ for } j = 1, 2, \ldots, n \tag{17}$$

This way, Equations (15.a) and (15.b) till (17) can be summarized by Theorem 2.7.

3. **Empirical Determination of Mean Derivative Functions.** As mathematically demonstrated in the previous section, the equation ${}^{k+1}\hat{y}^i = \tan_{\Delta t}{}^k\alpha^i \cdot \Delta t + {}^k y^i$ is a discrete general solution for nonlinear ordinary differential equations system $\dot{\vec{y}} = \vec{f}(\vec{y}, \vec{u})$, where $\tan_{\Delta t}{}^k\alpha^i$ is the mean derivative function with $\Delta t$ step. It is important to highlight that the application of equation ${}^{k+1}\hat{y}^i = \tan_{\Delta t}{}^k\alpha^i \cdot \Delta t + {}^k y^i$ is only allowed if the mean derivative function is known. Theoretically, one can use any universal functions approximator for this purpose (see Appendix A). Here the symbol ˆ is applied to state that the variable associated with it is an estimation and not an exact value. Moreover, the accuracy achieved by this equation is the same as if we use a high-order integrator on the instantaneous derivative functions.

The methodology of mean derivative uses fixed integration step both in the training phase and the simulation phase, being applicable only to the first-order integrator of Euler type. There are two approaches for determining mean derivative: direct and indirect (or empirical) approaches (see [17,18]). A direct approach is outlined in Figure 5 and the indirect or empirical approach in Figure 6. As presented in Figure 5, the mean derivative are directly inserted in the universal approximator in direct approach. However, in indirect or empirical approach, as outlined in Figure 6, the mean derivative are indirectly learned from the value ${}^{k+1}\hat{y}^i$. This difference is mathematically significant because in [17], when one considers the particular case in which the universal approximator is represented by a neural network with a *feedforward* architecture, it was verified that the backpropagation remains unchanged in direct approach. Yet, in empirical approach, the backpropagation must be slightly modified.

In the methodology of the mean derivative both direct and indirect approaches can be applied to real-world data; however, it was experimentally proved that the direct approach

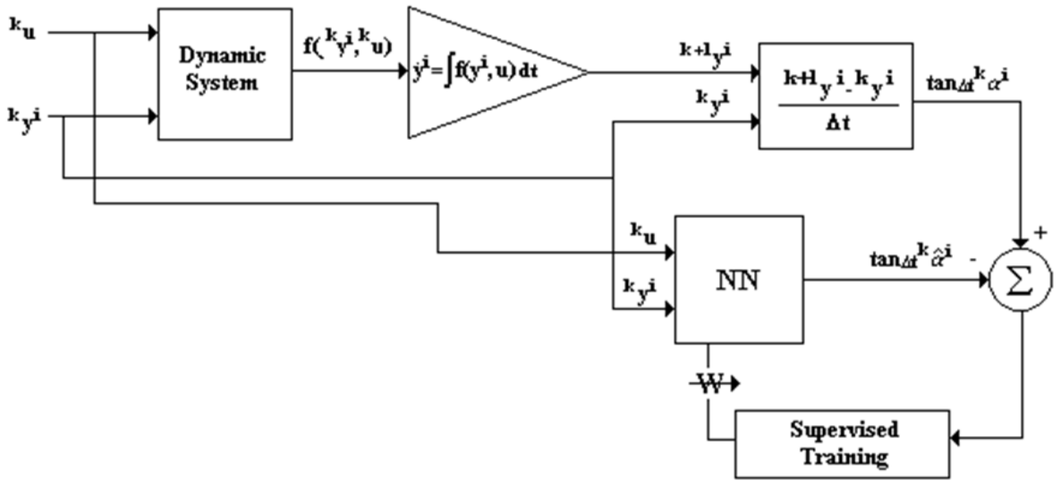


FIGURE 5. Direct approach to determining the mean derivative

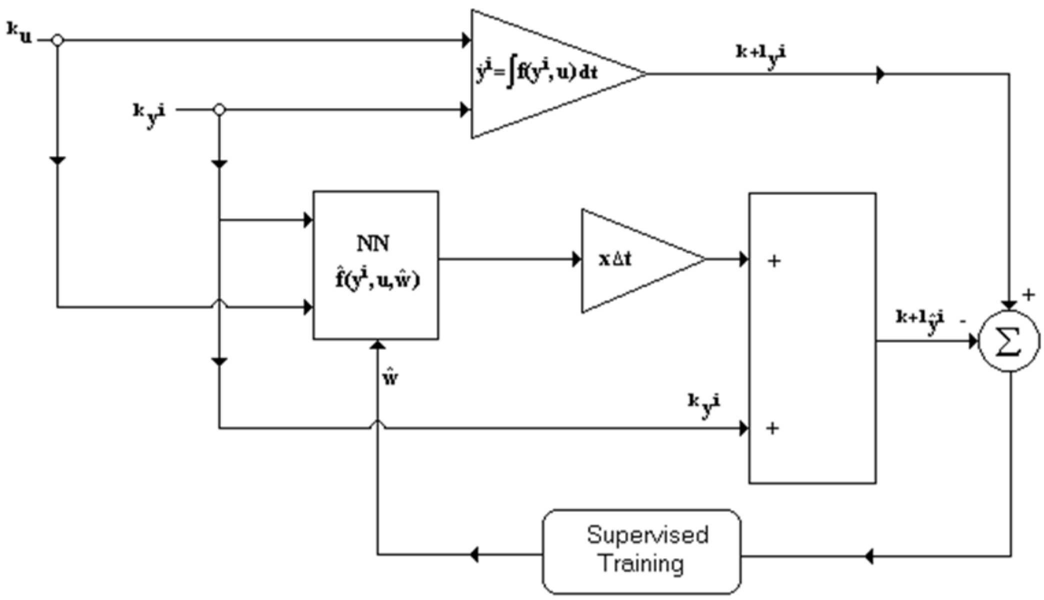


FIGURE 6. Indirect or empirical approach for determining mean derivative

is the most suitable [9]. The direct methodology of mean derivative applies to real-world problems, since the mean derivative are easily calculated from the input/output training patterns. However, if one intends to learn instantaneous derivative function (see [9,20]), the direct methodology cannot be applied to real-world problems, because instantaneous derivative functions are not easily estimated. This difference between direct and indirect approach cannot be applied to NARMAX methodology.

Hence, the following algorithm is proposed for computational determination of mean derivative via direct methodology involving both state variables and control variables:

1). Given the finite domains of interest $\left[y_j^{\min}(t_0), y_j^{\max}(t_0)\right]^{n_1}$ in $t_0$ for $j = 1, 2, \ldots, n_1$ of the state variables and also finite domains of interest $\left[u_j^{\min}(t_0), u_j^{\max}(t_0)\right]^{n_2}$ for $j = 1, 2, \ldots, n_2$ of control variables, $q$ vectors are randomly generated according to a uniform distribution within these ranges such as:

$$p_i = \left[y_1^i(t_0)\ y_2^i(t_0)\ \ldots\ y_{n_1}^i(t_0); u_1^i(t_0)\ u_2^i(t_0)\ \ldots\ u_{n_2}^i(t_0)\right]^T \tag{18.a}$$

and

$$P = [p_1 : p_2 : \ldots : p_q]_{(n_1+n_2)xq} \tag{18.b}$$

which are the input training patterns of the universal approximator at the instant $t_0$. The number of $q$ training patterns must be large enough to ensure a proper supervised learning.

2). By using a high-order integrator, we propagate all the initial conditions $p_i$ for $i = 1, 2, \ldots, q$ in order to obtain the states at the time $t_0 + \Delta t$, that is,

$$p_i^{\Delta t} = \left[y_1^i(t_0 + \Delta t)\ \ y_2^i(t_0 + \Delta t)\ \ \ldots\ \ y_{n_1}^i(t_0 + \Delta t)\right]^T \tag{19.a}$$

and

$$P^{\Delta t} = \left[p_1^{\Delta t} : p_2^{\Delta t} : \ldots : p_q^{\Delta t}\right]_{n_1xq} \tag{19.b}$$

NOTE: Alternatively in this step, one may use a computerized data acquisition system to capture the behavior of the real world dynamic systems.

3). One should establish output vectors or output training patterns $T_i$ of neural network (this is so called mean derivative direct methodology) as follows:

$$\begin{aligned} T_i &= \frac{1}{\Delta t}\left[y_1^i(t_0 + \Delta t) - y_1^i(t_0) y_2^i(t_0 + \Delta t) - y_2^i(t_0) \ldots y_{n_1}^i(t_0 + \Delta t) - y_{n_1}^i(t_0)\right]^T \\ &= \left[\tan_{\Delta t}{}^k\alpha_1^i\ \ \tan_{\Delta t}{}^k\alpha_2^i\ \ \ldots\ \ \tan_{\Delta t}{}^k\alpha_{n_1}^i\right] = \left(\tan_{\Delta t}{}^k\alpha^i\right) \end{aligned} \tag{20}$$

and

$$T = [T_1 : T_2 : \ldots : T_q]_{n_1xq} \tag{21}$$

Since function $\tan_{\Delta t}{}^k\alpha^i$ is also autonomous, it only takes one propagation over all the initial conditions of input vectors $p_i$ for $i = 1, 2, \ldots, p$, in order to model training patterns $P$ and $T$ required by the universal approximator of functions.

4). Once one has the input vectors $P$ and the output vectors $T$ required by the universal approximator, this can go through a supervised training process to learn mean derivative functions within the desired accuracy.

5). When the supervised training of the universal approximator is consolidated, it is possible to simulate the dynamics from the following discrete expression:

$${}^{k+1}\hat{y}^i = \tan_{\Delta t}{}^k\alpha^i \cdot \Delta t + {}^ky^i \tag{22}$$

For the purpose of analyzing the *local error* on Euler integration structure given by the equation ${}^{k+1}\hat{y}^i = \tan_{\Delta t}{}^k\alpha^i \cdot \Delta t + {}^ky^i$ concerning the aspects of the mean derivative functions, consider the exact value ${}^{k+1}\bar{y}^i$ and the estimated value ${}^{k+1}\hat{y}^i$ obtained by the means of Equations (23.a) and (23.b).

$${}^{k+1}\bar{y}^i = \tan_{\Delta t}{}^k\alpha^i \cdot \Delta t + {}^k y^i \tag{23.a}$$

$${}^{k+1}\hat{y}^i = \left(\tan_{\Delta t}{}^k\alpha^i + e_m\right) \cdot \Delta t + {}^k y^i \tag{23.b}$$

where $e_m$ is the mean absolute error of the output variables of the universal approximator trained to learn the mean derivative functions. By subtracting Equation (23.b) of Equation (23.a) and squaring the final result, one has:

$$\left({}^{k+1}\bar{y}^i - {}^{k+1}\hat{y}^i\right)^2 = \Delta t^2 \cdot e_m^2 \tag{24}$$

The existence of universal approximator of functions ensures that Equation (24) comes close to any desired accuracy, since $e_m^2$ can be as small as one desires for a constant and greater than zero integration step $\Delta t$. For a global error analysis of Euler type integration structure (designed with mean derivative), more studies are required, which is beyond the scope of this article. With respect to Equation (24) one also should observe that if the integration step $\Delta t$ has a smaller than one and greater than zero value, this variable receives the ability to decrease the local error learned by universal approximator. This is probably the reason why the methodology of mean derivative is easier of being computationally trained than NARMAX [18,19] methodology. However, the experience has proved that for integration steps $\Delta t > 1$ the error is amplified.

4. **Conclusions.** From what has been stated in the previous sections, one takes up the following conclusions.

1) The integration step $\Delta t$ is fixed and does not need to be infinitesimal; it can be of any value, including greater than one. However, it is important to note that if it is desirable to change the integration step to determine mean derivative functions, a new estimation is necessary, since these values will all be changed in function of the size of integration step.

2) $\tan_{\Delta t}{}^k\alpha^i = \frac{{}^{k+1}y^i - {}^k y^i}{\Delta t}$ is, by definition, the mean derivative or secant from interval $[t_k, t_{k+1}]$ of curve $y^i(t)$, and at the same time it is the discrete and general solution for the dynamic system $\dot{y}^i = f\left(y^i, u\right)$.

3) The general solution given by the mean derivative functions in Euler integrator for the nonlinear autonomous ordinary differential equations system is discrete and exact. However, the empirical determination of mean derivative functions through any universal approximator of functions is approximated, but it is always within a desired error.

4) From T1, P1, P2, P3 and P4 we have the values of $\tan_{\Delta t}{}^{k+l}\alpha^i$ to $l = 0, 1, \ldots, (L-1)$ exist and are unique, thus $\tan_{\Delta t}{}^{k+l}\alpha^i$ is a static function with the same qualitative properties of the instantaneous derivative function ${}^{k+l}\dot{y}^i = f\left({}^{k+l}y^i, {}^{k+l}u\right)$. In fact, it can be shown that $\lim\limits_{\Delta t \to 0} \tan_{\Delta t}{}^{k+l}y^i = f\left({}^{k+l}y^i, {}^{k+l}u\right)$. It is also important to realize that the instantaneous derivative function $\dot{y}^i = f\left(y^i, u\right)$ does not depend on $\Delta t$, but $\tan_{\Delta t}{}^{k+1}\alpha^i$ does. This latter property implies that the mean derivative methodology has a fixed integration step while the method of instantaneous derivatives, firstly proposed by Wang and Lin [20] may have variable integration step.

5) From Theorem T2 $\tan_{\Delta t}{}^k\alpha^i$ is really an exact derivative of at least one point inside of the interval $[t_k, t_{k+1}]$, which is also another way to ensure the existence of $\tan_{\Delta t}{}^k\alpha^i$.

6) From T3 and T4 theorems, and P6 property it follows that the recurrence relation that links ${}^{k+1}y^i$ with ${}^k y^i$ and ${}^k u$ to obtain a discretized solution for the dynamic system $\dot{y}^i = f\left(y^i, u\right)$ is, in fact, given by ${}^{k+1}y^i = \tan_{\Delta t}{}^k\alpha^i \cdot \Delta t + {}^k y^i$ which is a simple integration structure of Euler type.

7) From Theorem T5, both the instantaneous derivative functions $f\left({}^{k}y^{i}, {}^{k}u\right)$ and mean derivative functions $\tan_{\Delta t} {}^{k}\alpha^{i}$ are invariant in time; however, they are parameterized in relation to ${}^{k}u$.

8) Apparently, from a mathematical perspective, it would be extremely inaccurate to simply state geometrically or graphically that the secant definition is a general discreet solution and exact for the differential equations system $\dot{y} = f(y)$. It really needed a little bit more elaborate algebra to affirm that.

9) Interpolation or estimation of mean derivative functions up to the infinite is impossible; therefore there is a restriction imposed to finite domains outlined in Figure 1.

10) If it is desirable to use the general solution of Euler with mean derivative in control theory, then the integration step $\Delta t$ must be very small and close to zero to ensure the parameterization mentioned in T5 theorem in the dynamic $\dot{y}^{i} = f\left(y^{i,}, u\right)$ for a correct representation of temporal variation, that are subject to the values of the control variables.

11) The methodology presented in this article has been successfully tested in neural control theory [19].

12) Because of Equation (24) it seems that the use of Euler integrator designed with mean derivative for the integration step $\Delta T < 1$ is more effective from the computational point of view, than the use the NARMAX methodology.

13) The methodology of mean derivative is virtually unknown by the community of artificial neural networks.

14) Euler integrator designed with mean derivative has many applications in real-world problems; for example, it can be used to predict river flows in watersheds from past measurements or to predict the formation of sunspots.

15) The confined integration step $\Delta t$ in the interval $[0, 1]$ always decreases the local error. Nevertheless, if this $\Delta t$ value is very close to zero, for example $\Delta t = 0.0001$, the global error will increase much, though the local error is very small (very small $\Delta t$ values display much more iterations in Euler integrator with mean derivative). Thus, it is prudent to think that there is a $\Delta t$ step of integration between zero and one, not very close to zero nor too close to one, which minimizes the global error. However, the calculation of this optimum $\Delta t$ is beyond the scope of this article.

From the findings above one can guarantee the possibility of using any universal approximator of functions to represent dynamic systems through a Euler-type integration structure, for a given $\Delta t$ step. Just consider the capacity of representing functions of these universal approximators to deduce their ability to control the local error in Euler integration process with mean derivative.

Being so, the equation $\left({}^{k+1}\bar{y}^{i} - {}^{k+1}\hat{y}^{i}\right)^{2} = \Delta t^{2} \cdot e_{m}^{2}$ is close to any desired accuracy since $e_{m}^{2}$ can be as small as one desires for a fixed integration step $\Delta t > 0$. Thus, ${}^{k+1}\hat{y}^{i}$ in equation ${}^{k+1}\hat{y}^{i} = \left(\tan_{\Delta t} {}^{k}\alpha^{i} + e_{m}\right) \cdot \Delta t + {}^{k}y^{i}$ is close to any desired accuracy because the universal approximator has the ability to estimate, within a domain of interest, the mean derivative functions $\tan_{\Delta t} {}^{k}\alpha_{j}^{i}$, which is invariant in time. However, the smaller the square error desired $e_{m}^{2}$ during a supervised learning process is, the greater the computational effort to determine mean derivative functions is.

**Acknowledgments.** Initially, the authors of this article would like to thank for the excellent technical review and good written suggestions made by the reviewers of this paper. We would also like to thank the Technological Institute of Aeronautics (ITA), Casimiro Montenegro Filho Foundation (FCMF) and the 2RP Net Enterprise staffs for sponsoring this research.

## REFERENCES


[1] M. Braun, *Differential Equations and Their Applications*, 3rd Edition, Springer-Verlag, New York, 1983.

[2] S. Chen, S. A. Billings and P. M. Grant, Non-linear system identification using neural networks, *Int. J. Control*, vol.51, no.6, pp.1191-1214, 1990.

[3] N. E. Cotter, The Stone-Weierstrass and its application to neural networks, *IEEE Trans. Neural Networks*, vol.1, no.4, pp.290-295, 1990.

[4] G. Cybenko, *Continuous Valued Networks with Two Hidden Layers Are Sufficient*, Technical Report, Department of Computer Science, Tufts University, 1988.

[5] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd Edition, Prentice-Hall, Inc., New Jersey, 1999.

[6] K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks*, vol.2, no.5, pp.359-366, 1989.

[7] J. K. Hunt, D. Sbarbaro, R. Zbikowski and P. Gawthrop, Neural networks for control system – A survey, *Automatica*, vol.28, no.6, pp.1083-1112, 1992.

[8] J. S. R. Jang, C.-T. Sun and E. Mizutani, *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice-Hall, Inc., 1997.

[9] R. P. Melo and P. M. Tasinaffo, Uma metodologia de modelagem empírica utilizando o integrador neural de múltiplos passos do tipo Adams-Bashforth, *Sociedade Brasileira de Automática (SBA)*, vol.21, no.5, pp.487-509, 2010.

[10] M. A. Munem and D. J. Foulis, *Calculus with Analytic Geometry. Volumes I and II*, Worth Publishers, Inc., New York, 1978.

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

[12] K. S. Narendra and A. U. Levin, Control of nonlinear dynamical systems using neural networks: Controllability and stabilization, *IEEE Trans. Neural Networks*, vol.4, no.2, pp.192-206, 1993.

[13] A. Rios Neto, Dynamic systems numerical integrators in neural control schemes, *V Congresso Brasileiro de Redes Neurais*, Rio de Janeiro-RJ, Brazil, pp.85-88, 2001.

[14] S. Russell and P. Norvig, *Artificial Intelligence*, 2nd Edition, Prentice-Hall, Inc., New Jersey, 2003.

[15] I. S. Sokolnikoff and R. M. Redheffer, *Mathematics of Physics and Modern Engineering*, 2nd Edition, McGraw-Hill Kogakusha, LTD., 1966.

[16] J. T. Spooner, M. Maggiore, R. Ordónez and K. M. Passino, *Stable Adaptive Control and Estimation for Nonlinear Systems: Neural and Fuzzy Approximator Techniques*, Wiley-Interscience, New York, 2002.

[17] P. M. Tasinaffo, *Estruturas de Integração Neural Feedforward Testadas em Problemas de Controle Preditivo*, Ph.D. Thesis, Instituto Nacional de Pesquisas Espaciais -INPE, São José dos Campos, Brazil, 2003.

[18] P. M. Tasinaffo and A. Rios Neto, Mean derivative based neural Euler integrator for nonlinear dynamic systems modeling, *Learning and Nonlinear Models*, vol.3, no.2, pp.98-109, 2005.

[19] P. M. Tasinaffo and A. Rios Neto, Predictive control with mean derivative based neural Euler integrator dynamic model, *Sociedade Brasileira de Automática (SBA)*, vol.18, no.1, pp.94-105, 2006.

[20] Y.-J. Wang and C.-T. Lin, Runge-Kutta neural network for identification of dynamical systems in high accuracy, *IEEE Trans. Neural Networks*, vol.9, no.2, pp.294-307, 1998.

[21] E. Wilson, *Advanced Calculus*, Dover Publications, New York, 1958.

[22] J. M. Zurada, *Introduction to Artificial Neural System*, West Pub. Co., St. Paul, MN, USA, 1992.

[23] L. R. Medsker and L. C. Jain, *Recurrent Neuralbreak Networks: Design and Applications*, CRC Press, Boca Raton, FL, USA, 2001.

[24] J. F. Kolen and S. C. Kremer, *A Field Guide to Dynamical Recurrent Networks*, Wiley, New York, NY, USA, 2001.

[25] P. Shi, Y. Zhang, M. Chadli and R. Agarwal, Mixed H-infinity and passive filtering for discrete fuzzy neural networks with stochastic jumps and time delays, *IEEE Trans. Neural Networks and Learning Systems*, 2015.

[26] R. S. Guimarães, V. S. Júnior and P. M. Tasinaffo, Multipolar-valued fuzzy sets to deal with the cognitive ambiguities, *International Journal of Innovative Computing, Information and Control*, vol.11, no.6, pp.1965-1985, 2015.

**Appendix** A. **Universal Approximator for Functions.** A priori, for a proper mathematical study of the function approximation theory some definitions are needed (see [16]), as follows.

**Definition A.1.** *A function $f : D \to R^n$ is uniformly continuous on $D \subseteq R^n$ point if for each $\varepsilon > 0$ there is a $\delta(\varepsilon)$ so that for each $x, y \in D$ satisfying $|x - y| < \delta(\varepsilon)$, $|f(x) - f(y)| < \varepsilon$. The function $f : D \to R$ is continuous on $D$ if it is continuous for each point in $D$.*

**Definition A.2.** *A real number $b$ is a superior quota of a set $X \neq \phi$ of real numbers if and only if $x \leq b$ for every $x \in X$. Besides that, if any number smaller than $b$ is superior quota of $X$, one may say that $b$ is a supreme of $(sup)X$.*

Being so, from this definition of supreme, it is possible to state the fundamental principle of the existence of a supreme in real numbers domain, that is:
**Supreme Principle.** *Every real function of the type $y = f(x)$, superiorly limited, has a supreme.*

One can write the functions approximator as $F(x, w)$, where $w \in R^p$ is the parameters vector used to define the mapping of the approximator. Suppose that $\Omega^p \subset R^p$ denotes the subset of all values that the approximator parameters can obtain. Thus, it is assumed that:

$$G = \{F(x, w) : w \in \Omega^p, p \geq 0\} \tag{25}$$

is the class of functions in the form $F(x, w)$, $w \subset \Omega^p$ for any $p \geq 0$. In this case, when one refers to the functions of $G$ class, one omits how great $p$ is. That said, a *uniform approximator* is defined as follows.

**Definition A.3.** *A function $f : D \to R$ is uniformly continuous on $D \subseteq R^n$ if for each $\varepsilon > 0$, there is a $\delta(\varepsilon)$ (dependent $\varepsilon$ only) so that for every $x, y \in D$ satisfying $|x - y| < \delta(\varepsilon)$ then $|f(x) - f(x)| < \varepsilon$.*

Regarding Definition A.3, note that the supreme operation on the function $F(x) - f(x)$ establishes that even the maximum difference in the greatness of these two functions, which inevitably occurs to some $x$, should still be less than $\varepsilon$. Thus, it gives now the definition of a universal approximator.

**Definition A.4.** *A mathematical structure defining a class of functions $G_1$ is considered a universal approximator for the class functions $G_2$, if each $f \in G_2$ can be uniformly approximated by $G_1$.*

Thus, according to Jang et al. [8], the establishment of a broad class of universal approximators is mere existence theorems and not constructive methods. The starting theorem to establish the existence of universal functions approximators is the Stone-Weierstrass theorem. This theorem provides a useful way to determine whether certain approximators are really universal for a class of continuous functions and defined on a compact set. The Stone-Weierstrass theorem is stated below (see [3,8,16]).

**Theorem A.1 (T8).** *(Stone-Weierstrass). A continuous function $f : D \to R$ can be uniformly approximated on $D \subseteq R^n$ by the class functions $G$ if,*

*(1) The constant function $g(x) = 1$, $x \in D$ belongs to $G$,*

*(2) If $g_1$, $g_2$ belong to $G$, $a \cdot g_1 + b \cdot g_2$ will belong to $G$ for every $a, b \in R$,*

*(3) If $g_1$, $g_2$ belong to $G$, $g_1 \cdot g_2$ will belong to $G$, and*

*(4) If $x_1 \neq x_2$ are two distinct points in $D$, then there will be a function in $g \in G$ so that $g(x_1) \neq g(x_2)$.*

The Stone-Weierstrass theorem, which derives the real classical analysis can be used to demonstrate that some neural network architectures have a universal approximation capacity [3]. The Stone-Weierstrass theorem states conditions that warrant that neural networks with MLP (Multilayer Perceptron) architecture or RBF (Radial-Basis Functions), or Mamdani-type Fuzzy Inference System (see [8,16]) can approach any continuous functions in the sense given by Definitions A.3 and A.4. So before listing some important theorems taken from [16] concerning the theme of universal approximators, some important functions used in neural networks theory are set up.

A *sigmoid* function is every *continuous* function in the form of "s". An example of such a function can be the hyperbolic tangent function (or tansig function) with horizontal asymptotes in $-1$ and $+1$ given by the equation $\varphi(net) = \frac{2}{1+\exp(-\lambda \cdot net)} - 1$. Another example of sigmoid function can be the logistic function (or logsig function) with horizontal asymptotes in 0 and $-1$ and given by the equation $\varphi(net) = \frac{1}{1+\exp(-\lambda \cdot net)}$. Besides these, there are other functions with "s" shape, but they are discontinuous. Two of them are defined below.

Signal Function (Signum Function): a discontinuous function given by the equation $\varphi(net) = sgn(net) = \begin{cases} 1 & se\, net > 0 \\ -1 & se\, net < 0 \end{cases}$ or $\varphi(net) = \begin{cases} 1 & se\, net > 0 \\ 0 & se\, net = 0 \\ -1 & se\, net < 0 \end{cases}$. Note that this equation is the limit of $\lambda \to \infty$ in the hyperbolic tangent function.

Heaviside Function or Threshold Function: a discontinuous function given by equation $\varphi(net) = \begin{cases} 1 & \text{if } net \geq 0 \\ 0 & \text{if } net < 0 \end{cases}$. This equation was originally used in McCulloch and Pitts neuron in 1943. Note that this equation is the limit of $\lambda \to \infty$ in the Logistic function.

Note that the signal function and threshold are not sigmoid functions, since they are not continuous. Given the definitions of these functions, one may list the following universal functions approximators [16] that can be demonstrated from Theorem A.1.

**Theorem A.2 (T9).** *MLP networks of two layers, an inner layer of neurons defined by the sigmoid functions or threshold function and a linear output are universal approximators for* $f \in G_{cb}(n, D)$, $D = [a, b]$.

**Theorem A.3 (T10).** *RBF networks defined by radial basis class functions* $G_{fbr} = \{g(x) = \sum_{i=1}^{p} a_i \exp(-\gamma_i |x - c_i|^2)\}$ *with* $a_i, \gamma_i \in R$ *and* $c_i \in R^n$ *for* $i = 1, 2, \ldots, p$ *are universal approximators for* $f \in G_{cb}(n, D)$.

**Theorem A.4 (T11).** *Fuzzy systems with triangular or Gaussian pertinence functions in the input and a defuzzification that uses the average of the centers are universal approximators for* $f \in G_{cb}(n, D)$, $D = [a, b]$.

**Appendix** B. **A Brief Summary of NARMAX Methodology.** A universal approximator of functions can be used to represent a non-linear dynamic system of the type given by Equation (26) from a discrete set of input/output training patterns,

$$\dot{y} = f(y, u) \tag{26}$$

In this equation, the function $f(.)$ is invariant in time (see [2]). In order to do this, it is usually assumed the possibility of approximation of the dynamic system given by Equation (26) by a discrete model of NARMAX type, given by:

$$y(t + \Delta t) = f(y(t), y(t - \Delta t), \ldots, y(t - n_x \Delta t); u(t), u(t - \Delta t), \ldots, u(t - n_u \Delta t)) \tag{27}$$

where $n_x$, $n_u$, $\Delta t$ are constants that must be adjusted depending on the problem being addressed and the desired accuracy. This possibility is considered for the use of any universal approximator of functions to act as a discrete model as given by Equation (27). For example, for the particular case of an artificial neural network, the size of the neural network (number of layers and number of neurons per layer) can be adjusted to reach the desired accuracy for a given choice of $n_x$, $n_u$, $\Delta t$.

At this point it is convenient to remember a similar situation that occurs when numerical integrators and dynamical systems are used as in Equation (26), they are treated by discrete approximations as in Equation (27). Neural networks of multi-layer perceptron-type with one inner layer are sufficient to accomplish this representation task for virtually any function found in many engineering applications [6]. It is important to observe that the great difference between NARMAX methodology and the methodology that uses a Euler type integrator with mean derivative is that in the first one, the universal approximator outputs are the future values $y(t + \Delta t)$, while in the second, the outputs are the mean derivative functions at the instant $t$.