A Note on Smooth Approximation Capabilities of Fuzzy Systems

Manuel Landajo, María J. Río, and Rigoberto Pérez

Abstract—Modeling and prediction in some systems requires the simultaneous approximation of mappings and their derivatives to a certain finite order. In this paper, universal approximation capabilities of fuzzy systems are extended to this situation, by showing the denseness of some general classes of fuzzy models in appropriate function spaces where distance between functions is defined in terms of their derivatives. Requirements are generally very mild, and are often limited to imposing sufficient differentiability conditions on the classes of fuzzy systems to be used. The cases covered by this paper include additive fuzzy systems with Gaussian membership functions and general additive fuzzy models with realistic membership functions. Some potential application fields are considered, including examples from economics and time series prediction.

Index Terms—Derivatives, fuzzy systems, universal approximation.

I. INTRODUCTION

Together with the advantages of fuzzy systems observed in practice, such as their ability to process soft information in a systematic way and the synergies derived from the combined use of fuzzy and neural/statistical methods in order to implement learning processes, the flexibility of fuzzy systems has recently received additional support, in the form of a set of rigorous results giving a mathematical basis to the universal approximation capabilities of fuzzy systems in several function spaces.

There currently exists a considerable amount of results (e.g., [14], [24], [26], [27], [2], [18], and [16]) showing, under very general conditions, the universal approximation properties of fuzzy systems in several function spaces of practical relevance. Among others, uniform approximation of continuous real functions on compact sets and in \( L^p \) spaces are covered. These results provide a sound and objective basis to the practical use of fuzzy systems as a flexible modeling tool and—conveniently enough—their requirements are rather mild, so that many alternatives are possible in order to construct adequate fuzzy systems.

However, in a certain number of practical situations arising in economics, finance, robotics, physics, and other fields, the required approximation capabilities are especially strong, simultaneously requiring the approximation of mappings and their derivatives to a certain finite order.

Nonparametric estimation of elasticities in demand systems [7], [1], learning of smooth movements in robots [11] and the study of certain dynamic characteristics of chaotic systems useful in financial and physics time series prediction [6] are some examples. The main goal in all these cases is to be able to obtain an approximation \( \mathcal{A} \) to the relevant function \( F \), and having the guarantee that we can select \( \mathcal{A} \) in such a way that the derivatives of \( \mathcal{A} \) to a certain finite order \( r \) approximate those of \( F \) with the required precision \( \varepsilon \).

Theoretical results already exist for neural networks and other kinds of flexible estimation tools. Hence, Hornik et al. [10] and Hornik [9] have shown that, under very general conditions, perceptron networks possess the required abilities. In econometrics, similar analyzes have been conducted using kernel regression [21], splines and series estimators such as the Flexible Fourier Form (FFF) [1]. [7].

For the aforementioned reasons, it is important to know whether fuzzy rule-based systems possess such smooth approximation capabilities. As pointed out by Kosko [13], artificial intelligence (AI) models, which are constructed using step functions and are consequently able to approximate continuous and bounded measurable functions, are obviously of no use in solving our problem due to the fact that their derivatives (where they exist) are always null. As remarked in [13], multivalence has many mathematical advantages, providing fuzzy systems with much richer analytical properties than those of AI models, as well as a smooth interpolation character. To a great extent, this makes fuzzy models more akin to numerical estimators than to symbolic systems, while preserving at the same time the rule-based structures of artificial intelligence models.

Intuitively, in order to ensure smooth approximation capabilities, fuzzy systems should be constructed by using adequately differentiable membership functions and fuzzy logic operators. However, this intuition is not immediate and must not be taken for granted. In this paper, several general families of functions including some of the most popular fuzzy systems are considered, and for these it is shown that—with an essentially arbitrary choice of (sufficiently differentiable) basis functions—the resulting fuzzy systems have the ability to approximate continuous functions and their derivatives uniformly on arbitrary compact sets of \( \mathbb{R}^n \) to the desired degree.

The structure of the paper is as follows. In Section II, basic notation and concepts are presented. Section III contains theoretical results, which cover both the cases of additive fuzzy systems with Gaussian membership functions and general additive fuzzy systems with realistic membership functions. These results are immediately extended to other classes such as Takagi–Sugeno fuzzy systems (with adequately differentiable membership functions) and augmented linear models.
Section IV some practical applications to different fields are presented. Section V summarizes the results and suggests some further research directions. Mathematical proofs are gathered in the Appendix.

II. Basic Notions

Let $U$ be an open subset of the $N$-dimensional Euclidean space $R^N$ and $C(U)$ the set of continuous functions $U \to R$. Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)^T$ be an $N$-tuple of nonnegative integers (a multiindex). For $X \in R^N$, let $X^\alpha \equiv x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \ldots \cdot x_N^{\alpha_N}$, and denote by $D^\alpha$ the following partial derivative:

$$\frac{\partial^\alpha}{\partial X^\alpha} \equiv \frac{\partial^\alpha}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_N^{\alpha_N}}$$

where the order $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_N$.

We denote by $C^r(U)$ the set of all the $r$-times continuously differentiable functions, i.e., $C^r(U) = \{ f \in C(U) | D^\alpha f \in C(U) \}$ for every $|\alpha| \leq r$, where $r \in IN \cup \{0\}$. We identify $C^0(U) = C(U)$. $C^0_0(U)$ will be the subset of $C^0(U)$ composed by all the functions with compact support contained in $U$.

Let $\{K_n\}$ be a nondecreasing sequence of compact subsets covering $U$, i.e., such that $K_n \subset K_{n+1}$ and $\bigcup_n K_n = U$. For each $K_n$ we may consider the seminorm of $r$-uniform convergence

$$\rho_{K_n}(f) = \max_{0 \leq |\alpha| \leq r} \sup_{X \in K_n} |D^\alpha f(X)|.$$

The topology induced on $C^r(U)$ by this family of seminorms is equivalent to that of the following distance:

$$d(f, g) = \sum_{n=1}^{\infty} 2^{-n} \frac{\rho_{K_n}(f - g)}{1 + \rho_{K_n}(f - g)}$$

where $f, g \in C^r(U)$. Frechet space $E^r(U)$ is the set $C^r(U)$ equipped with such a topology.

A sequence of functions $\{f_i\}$ converges in $E^r(U)$ toward a function $f$ if the functions $f_i$s and their derivatives to order $r$ converge, uniformly on every compact $K$ of $U$, toward the function $f$ and its corresponding derivatives. Since convergence in $E^r(U)$ is rather strong, denseness in $E^r(U)$ is a highly desirable property for fuzzy systems. This is precisely the problem we are going to consider here. Our framework is similar to that considered by Hornik et al. [10] and Hornik [9] for the case of peceptron networks.

III. Main Results

A. Gaussian Algebras

We start by demonstrating the approximation capabilities in $E^r(R^N)$of the familiar additive fuzzy systems with Gaussian membership functions:

**Theorem 1:** Let $A(G)$ be the following class of functions:

$$A(G) \equiv \left\{ A(X) = \sum_{j=1}^{m} \beta_j \prod_{i=1}^{N} a_{ij} \exp \left( \frac{-1}{2} \frac{(x_i - \mu_{ij})^2}{\sigma_{ij}} \right) \right\}$$

with $X = (x_1, \ldots, x_i, \ldots, x_N) \in R^N$, $m, N \in IN$; $\beta_j, \mu_{ij} \in \mathbb{R}$; $\sigma_{ij}, a_{ij} \in \mathbb{R}_{++}$. Then, for every $r$ (positive integer), class $A(G)$ is dense in $E^r(R^N)$.

**Proof:** See Appendix.

The Gaussian case is evidently a leading choice, and one that permits us to make full use of the computational advantages of models which satisfy the hypotheses of the Stone–Weierstrass (SW) Theorem [3]. The Gaussian case is, in fact, a particular choice within a larger class of fuzzy systems that could be derived from the SW Theorem. Proposition A in the Appendix may be used to construct classes of fuzzy systems dense in $E^r(R^N)$, with other bell-shaped curves than Gaussians [15].

B. General (Realistic) Additive Fuzzy Systems

Since for most classes of fuzzy systems conditions of the SW Theorem do not hold, a more general framework must be considered in order to permit additive fuzzy systems with realistic membership functions to find a place therein. In this section, we derive results of density in $E^r(U)$ spaces for realistic additive fuzzy systems and propose some classes of membership functions suitable for smooth approximation which are both simple and realistic. Due to the fact that fuzzy systems with a weighted mean structure are only defined on the subset of $R^N$ where weights are positive, some pathologies may emerge. As a consequence, we have to be careful with definitions. A convenient way to express the results is the following:

**Theorem 2:** Let $T: R \to [0, 1]$ be an arbitrary nonconstant function belonging to $C^r_0(R)$ ($r$ nonnegative integer). Let $A(T)$ be the class defined as follows:

$$A(T) \equiv \left\{ A(A(X)) = \sum_{j=1}^{\infty} \prod_{i=1}^{N} a_{ij} T \left( \frac{x_i - \mu_{ij}}{\sigma_{ij}} \right) \right\}$$

with $X = (x_1, \ldots, x_i, \ldots, x_N) \in R^N$, $N \in IN$; $\beta_j, \mu_{ij} \in \mathbb{R}$; $\sigma_{ij}, a_{ij} \in \mathbb{R}_{++}$ and for every $X \in R^N$ it holds

$$\sum_{j=1}^{\infty} \prod_{i=1}^{N} a_{ij} T \left( \frac{x_i - \mu_{ij}}{\sigma_{ij}} \right) > 0.$$

Then $A(T)$ is dense in $E^r(R^N)$.

**Proof:** See Appendix.

The above definition of $A(T)$ is simpler than it appears. Due to the compactness of the support of $T$, on every compact set $K$
in $\mathbb{R}^N$, the approximators can be reduced to the habitual expression:

$$A(X) = \frac{\sum_{j=1}^{m} \beta_j \prod_{i=1}^{N} a_{ij} T \left( \frac{x_i - \mu_{ij}}{\sigma_{ij}} \right)}{\sum_{j=1}^{m} \prod_{i=1}^{N} a_{ij} T \left( \frac{x_i - \mu_{ij}}{\sigma_{ij}} \right)}.$$ 

### C. Other Classes Dense in $E^n(\mathbb{R}^N)$

Every class of functions belonging to $E^n$ and containing a subset dense in $E^n$ is also dense in such a space. This immediately applies to the following cases:

**Corollary:** Let $A$ be one of the previous classes dense in $E^n(\mathbb{R}^N)$. We have the following:

1. (Takagi–Sugeno type models): The extensions of models in $A$ obtained by substituting consequents $\beta_j$ with functions $B_j(X) = \beta_j + c_j \varphi_j(X)$, where $\varphi_j(X) \in C^n(\mathbb{R}^N)$ and $\beta_j, c_j \in \mathbb{R}$, are dense in $E^n(\mathbb{R}^N)$.

2. (Augmented linear models): The extensions of models in $A$ obtained as $B(X) = a_0 + \alpha \cdot \varphi(X) + A(X)$, where $A(X) \in A$ and $\varphi(X) \in C^n(\mathbb{R}^N)$, are dense in $E^n(\mathbb{R}^N)$. In particular, we may take $B(X) = a_0 + X \cdot \alpha + A(X)$.

3. Finite linear combinations of classes dense in $E^n(\mathbb{R}^N)$ are also dense in $E^n(\mathbb{R}^N)$.

4. (Hyperellipsoidal fuzzy systems): The modifications of models in $A$ obtained by substituting the tensor-product structures (such as the product of Gaussians) by “hyperellipsoidal” membership functions are also dense in $E^n(\mathbb{R}^N)$.

### D. Realistic Membership Functions

It is easy to find examples of basis functions suitable for constructing membership functions satisfying the requirements of simplicity, locality and smoothness we have imposed. The following are some possibilities:

$$T(z) = \begin{cases} 
(1 - z^2)^{\rho + 1} & \text{if } |z| \leq 1 \\
0 & \text{if } |z| > 1
\end{cases} \quad (1)$$

$$T(z) = \begin{cases} 
\exp(-1/(1 - z^2)) & \text{if } |z| \leq 1 \\
0 & \text{if } |z| > 1
\end{cases} \quad (2)$$

Both of these examples are bell-shaped curves of compact support (see Fig. 1). They depend only on two parameters because $z = (x-\mu)/\sigma$. The function of (1) has continuous derivatives to order $\rho$, and (2) is a $C^\infty$ function (i.e., with continuous derivatives of every order).

**Remark 1:** All the above results refer to additive fuzzy systems with weighted average structure, but may be equivalently formulated for weighted sum structures.

**Remark 2:** Some important classes of fuzzy systems are excluded from the previous results, simply because the classes of membership functions and/or fuzzy logic operators used to construct them are not sufficiently smooth. Approximation abilities of these fuzzy systems are per force weaker. Thus, one of the simplest cases, additive fuzzy systems with triangular symmetric membership functions and min conjunction operator possess approximation capabilities similar to those of piecewise linear approximators, and the same applies to Takagi–Sugeno fuzzy systems constructed on the same membership functions and conjunction.

### IV. SOME APPLICATIONS

We briefly review some simple examples of potential applications of the smooth approximation capabilities of fuzzy systems to different fields. As a general rule, for a sufficiently precise simultaneous approximation of functions and their derivatives relatively large amounts of (statistical/expert) information are required. So, when the fuzzy systems are constructed from numeric information relatively higher sample sizes are in order. Similarly, for fuzzy models constructed by using exclusively subjective information, experts must be able to implement the rule base with particular care, in order to achieve a system with the required level of precision. In practice, we may have mild sample sizes together with moderately rich and precise expert information on the system to be modeled. In these cases it may be of interest to take advantage of the flexibility of learning mechanisms in fuzzy systems by combining both sources of information to obtain models with a higher performance.

In the examples appearing in this section only statistical information will be used. We have employed (weighted average) additive fuzzy systems with Gaussian membership functions and a moderate number of rules (ranging from 6 to 10). Sample sizes between 1000–5000 appear to be sufficient for effective learning. In all cases the sample is divided into an estimation set and a test set. Performance of the models is evaluated on the test set. In particular, we report the following out-of-sample diagnostics: forecast $R^2$ for both $f$ and its derivatives, mean absolute error (MAE) and root mean square error (RMSE). Other indicators we have used to implement the applications (Theil’s...
index, several information criteria, diagnostics of prediction errors) are omitted for the sake of brevity.

The model building and learning procedures we have used here are entirely standard: in all the examples a fuzzy system is initially generated by using unsupervised competitive learning methods (AVQ), and then refined by a supervised learning procedure (a variant of the standard BP algorithm). 2

Example 1) Production Economics: Elasticity Estimation: Many problems arising in economics require the estimation of certain functions, such as production and cost functions, which summarize some relevant characteristics of the modeled system. These estimates are then used for different purposes, such as the analysis of economic/technological relations or economic forecasting. They contain information on certain characteristics of the system related to the sensitivity of some variables to variations of other magnitudes. These characteristics are measured in terms of partial derivatives (called “marginal” magnitudes) and of relative sensitivities (so-called “elasticities”).

Unfortunately, economic theory frequently does not provide us with a completely specified model for an economic relation, only giving a set of qualitative relations, and perhaps some restrictions to the admissible classes of models for the problem. The risks associated to the misspecification of economic models have led econometricians to develop or adopt a number of flexible (nonparametric and seminonparametric) procedures to deal with these problems. Flexible regression tools range from series estimators to neural networks [1], [7], [8]. A crucial point of every flexible modeling technique for it to be useful in economic applications is its interpretability, in the sense that the models—while retaining the necessary flexibility—must be readable enough to enable the imposition of economic or technological restrictions (e.g., monotonicity and convexity) on the selected models. Hence, a compromise between the flexibility of a computational black-box and the clarity of a linear regression is needed.

The rule-based structure of fuzzy systems may have useful properties in this sense because of their very nature: Experts may impose a priori restrictions by adding fuzzy rules, or may perform an a posteriori analysis of the rules generated by automatic learning mechanisms, thus being able to discard nonsense or spurious rules generated by the optimization procedure. In addition, many econometric functionals are in fact related to the modeling of technological (production) processes, a field where fuzzy systems are particularly well equipped to incorporate different sources of information, both statistical data and expert knowledge. Thus, fuzzy systems may be, at least in certain contexts, a potentially more flexible and powerful tool for economic modeling than other flexible techniques, such as neural networks, where the imposition of restrictions related to monotonicity or curvature required to obtain meaningful economic models is not a trivial task.

<table>
<thead>
<tr>
<th>No. of Rules</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Observations</td>
<td>100</td>
</tr>
<tr>
<td>( RMSE(f) )</td>
<td>0.3319</td>
</tr>
<tr>
<td>( MAE(f) )</td>
<td>0.1577</td>
</tr>
<tr>
<td>( R^2(f) )</td>
<td>0.9951</td>
</tr>
<tr>
<td>( R^2(D_2f) )</td>
<td>0.7821</td>
</tr>
<tr>
<td>( R^2*(\varepsilon_{12}) )</td>
<td>0.9660</td>
</tr>
</tbody>
</table>

Fig. 2. Cobb–Douglas production function.

Fig. 3. Fuzzy model for Cobb–Douglas production function.
We are going to consider the approximation of the main characteristics of a production process represented by a Cobb–Douglas production function \( y = f(X) = 2x_1^{0.3}x_2^{0.7} \) (0 < \( x_1, x_2 \leq 10 \)), where \( X = (x_1, x_2) \) is the input vector and \( y \) is the output of the production process (see Table I). The first goal is the estimation or (local) approximation of \( f \). Once we have estimated the production function we use this to obtain its first derivative with respect to \( x_2 \).

\[
D_2f(X) = \frac{\partial f(X)}{\partial x_2} \quad \text{(marginal productivity of } x_2) 
\]

and the corresponding partial elasticity

\[
\varepsilon_2(f)(X) = D_2f(X) \frac{x_2}{f(X)}. 
\]

A random sample of size 2000, extracted from a uniform density defined on the considered subdomain of \( f \), was used for training, and an independent sample of 100 observations was taken as the test set.

The estimated approximation is almost indistinguishable from the Cobb–Douglas production function (see Figs. 2 and 3). Results for the desired derivative and elasticity are also satisfactory (Figs. 4–7). As shown by Kim and Mendel [12], in practical situations an expert may be able to increase the accuracy of these estimates, especially on the border of the considered domain, and in areas where only a few statistical observations of the modeled system exist. The analysis may be greatly extended and refined: instead of directly estimate production functions we may decide to estimate cost functions, or other relevant characteristics such as the so-called cross-elasticities (see [22]). Some ideas of Yager [25] on the use of fuzzy rule-based systems for the modeling of economic equilibria in demand systems are much in the spirit of this section.

**Example 2) Time Series Prediction: The Logistic Map:** The distinctive feature of chaotic systems is their sensitivity to small variations of initial conditions. This makes the estimation of Lyapunov exponents a useful tool to identify chaotic behavior. A natural goal of learning is the estimation of the transfer function \( f \) relating the actual state of the system to past states, i.e., such that \( X_t = f(X_{t-1}, X_{t-2}, \ldots) \).
TABLE II
MODEL FOR LOGISTIC SERIES AND DIAGNOSTICS ON THE TEST SET

<table>
<thead>
<tr>
<th>No. of Rules</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Observations</td>
<td>500</td>
</tr>
<tr>
<td>RMSE(f)</td>
<td>0.0009</td>
</tr>
<tr>
<td>MAE(f)</td>
<td>0.0008</td>
</tr>
<tr>
<td>$R^2(f)$</td>
<td>1 (approx.)</td>
</tr>
<tr>
<td>$R^2(Df)$</td>
<td>0.9992</td>
</tr>
</tbody>
</table>

An accurate approximation of $f$ may be used for prediction purposes, and its derivatives may be employed in the estimation of Lyapunov exponents, thus being of use to detect chaos and to distinguish stochastic from chaotic (deterministic) series [6]. Neural networks have found wide application in this field and Wang [23] has made a pioneer application of fuzzy modeling to a Mackey–Glass process.

Our second example is a version of the so-called simple logistic map, $Y_t = f(X_t) = 4Y_{t-1}(1 - Y_{t-1})$, ($Y_0 = 0.2, X_t = Y_{t-1}; t = 1, 2, \ldots T$). As noted in Martin and Sawyer [17], the (global) Lyapunov exponent of this model is $\lambda = 1$, showing the presence of chaos.

When we know transfer function $f$, by using a sufficiently large sample of the series, $\{Y_t, t = 1, 2, \ldots T\}$, we are able to estimate $\lambda$ by means of the local Lyapunov exponents (see [17]):

$$\lambda(T) = T^{-1} \sum_{t=1}^{T} \log_2 |Df(Y_{t-1})|.$$  

A fuzzy rule-based system of six rules was constructed for this time series. A total of 5500 consecutive observations of the logistic map were generated. The last 500 were separated for testing one-step-ahead predictions of the fuzzy model (see Fig. 8). Results are shown in Table II.

TABLE III
ESTIMATES OF LYAPUNOV EXPONENT

<table>
<thead>
<tr>
<th></th>
<th>Training set $(T = 5,000)$</th>
<th>Test set $(T = 500)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real $\lambda$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Exact $\lambda(T)$</td>
<td>1.0001</td>
<td>0.9943</td>
</tr>
<tr>
<td>Fuzzy Model ($\hat{\lambda}(T)$)</td>
<td>1.0050</td>
<td>0.9962</td>
</tr>
</tbody>
</table>

The first derivative function of the fuzzy model $A(Y_{t-1})$ was used to estimate Lyapunov exponents of the logistic map, according to this formula:

$$\hat{\lambda}(T) = T^{-1} \sum_{t=1}^{T} \log_2 |D[A(Y_{t-1})]|$$

where $\{Y_t\}$ is the time series and $A(Y_{t-1})$ is the fuzzy model used for one-step-ahead prediction. Table III shows the results both when calculations are made on training and test sets.

V. CONCLUDING REMARKS

The results of this paper give a formal foundation to already existing intuitions and to the empirically observed “smooth approximation” characteristics of fuzzy systems. Our results permit $r$-uniform approximation on compact sets under rather mild conditions and by using mainstream classes of fuzzy systems.

Fuzzy rule-based models may be used for tasks requiring simultaneous approximation of mappings and their derivatives. We have only considered a small variety of potential application fields. In different specific areas many others may surely be found.

As to the future research directions suggested, there are at least two immediate objects of interest.

The first is the extension of the analyses of this paper to general nonadditive fuzzy systems with nonsmooth fuzzy logic operators and membership functions.

The second goal could be the formal analysis, by statistical inference methods, of the estimation processes involved in the automatic rule generation mechanisms or, in more specific terms, the possibility of the consistent (nonparametric) simultaneous estimation of functions and their derivatives by using information contained in finite random samples extracted from the modeled real systems. An important result obtained by Gallant and White [8] for the case of perceptron networks indicates that this can be achieved by standard procedures, without requiring information on derivatives to be explicitly incorporated to the learning mechanism. The practical examples considered in this paper suggest that similar results may be obtained for fuzzy systems.

APPENDIX

First, we shall state the following general-purpose result. In essence, it is a generalized Stone–Weierstrass Theorem:

**Proposition A:** Let $U \subseteq \mathbb{R}^N$ be open, and let $A$ be a part of the algebra $E^r(U)$ of the $C^r$ functions ($r$ positive integer) on $U$. 

A necessary and sufficient condition for the algebra generated by \( A \) to be dense in \( E^r(U) \) is the following:

1. For every \( x \in U \) an \( f \in A \) exists such that \( f(x) \neq 0 \).
2. For every distinct \( x, y \in U \) an \( f \in A \) exists such that \( f(x) \neq f(y) \).
3. For every \( x^* \in U \) and every nonnull vector \( h \in \mathbb{R}^N \) an \( f \in A \) exists such that \( Df(x^*) \cdot h = \theta_{h_{x^*}} \cdot f \neq 0 \) (nonnull directional derivatives).

**Proof:** This is a particular case of a result by Dieudonné [5, Ch. 17, Section 14, Exercise 5]: Differential manifold \( M \) is our \( U \) (equipped with its canonical structure of differential manifold), and \( T(M) = U \times \mathbb{R}^N \) (the tangent bundle associated to \( M \)). Conditions 1) and 2) are essentially the habitual requirements of the Stone–Weierstrass theorem. Condition 3) of Dieudonné’s Proposition particularizes into our c) for \( M = U \), and so can be reduced to a condition on directional derivatives.

**Proof of Theorem 1:** This is derived from Proposition A. We start by setting \( U = \mathbb{R}^N \), and by the same expedient as used by Wang and Mendel [24], we obtain that, for every nonnegative integer order \( r \), \( A(G) \) is a subalgebra of \( C^r(\mathbb{R}^N) \) which separates points on \( \mathbb{R}^N \) and vanishes at no point.

Condition 3) is the most laborious to test. We must show that, for every point \( X^* \in \mathbb{R}^N \) and every nonnull vector \( h = (h_1, h_2, \ldots, h_N) \in \mathbb{R}^N \), an \( f \in A(G) \) exists such that the directional derivative of \( f \) at point \( X^* \) following the direction of tangency \( h \) is nonnull.

Let us take an arbitrary point \( X^* \in \mathbb{R}^N \) and a nonnull vector \( h = (h_1, h_2, \ldots, h_N) \in \mathbb{R}^N \). Without loss of generality we may suppose that for \( i = 1, 2, \ldots, k \), \( h_i \leq 0 \) and for \( i = k + 1, k + 2, \ldots, N \), \( h_i \leq 0 \). As \( h \) is nonnull, necessarily at least one \( h_i \neq 0 \). If we are able to find \( f \in A(G) \) such that the following conditions are simultaneously satisfied:

for \( i = 1, 2, \ldots, k \), \( \frac{\partial f(X^*)}{\partial x_i} > 0 \);

for \( i = k + 1, k + 2, \ldots, N \), \( \frac{\partial f(X^*)}{\partial x_i} < 0 \),

then we shall have

\[
Df(X^*) \cdot h = \sum_{i=1}^{N} \frac{\partial f(X^*)}{\partial x_i} \cdot h_i > 0.
\]

As a consequence, our problem may be reduced to finding an \( f \in A(G) \) with positive partial derivatives at \( X^* \) with respect to its first \( k \) arguments and negative derivatives with respect to the remaining ones. To find such a function \( f \) we make use of the following procedure:

1. As \( A(G) \) separates points on \( R \) we may always choose a nonconstant function \( B : R \to R \) such that \( B \in A(G) \), and it can be immediately shown that \( B \) may be chosen (strictly) positive. Hence, there will be two points \( x_1, x_2 \in \mathbb{R} \) such that \( B(x_1) \neq B(x_2) \). From the Mean Value Theorem [4] there must exist a point \( x^* \) in \( (x_1, x_2) \) such that the derivative of \( B \) is nonnull at \( x^* \).

2. We may observe that if we define \( C(x) = B(-x + x_1 + x_2) \), both functions \( B \) and \( C \) are positive, but their respective derivatives at \( x^* \) and \( x'' = -x^* + x_1 + x_2 \) have opposite signs. And both \( B \) and \( C \) belong to \( A(G) \). Without loss of generality we shall suppose that \( DB(x') > 0 \) and \( DC(x'') < 0 \).

3. Now we construct the functions \( A_1, A_2, \ldots, A_N \) as follows:

   for \( i = 1, 2, \ldots, k \), define \( A_i(x_i) = B(x_i) \);

   for \( i = k + 1, k + 2, \ldots, N \), define \( A_i(x_i) = C(x_i) \).

The first \( k \) functions are positive and have positive derivatives at \( x' \), and the rest are also positive functions, but have negative derivatives at \( x'' \).

4. Let us define \( A(X) = \prod_{i=1}^{N} A_i(x_i) \), with \( X = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N \). It is immediate that \( A \) is a positive function and belongs to \( A(G) \). It also holds that, for every \( X \in \mathbb{R}^N \),

\[
\text{sign} \left( \frac{\partial A(X)}{\partial x_i} \right) = \text{sign} \left( \frac{dA(x_i)}{dx_i} \right).
\]

Now let us take the point \( X'' = (x''_1, x''_2, \ldots, x''_N) \in \mathbb{R}^N \) whose coordinates are as follows:

for \( i = 1, 2, \ldots, k \), \( x''_i = x'_i \);

for \( i = k + 1, k + 2, \ldots, N \), \( x''_i = x''_i \).

Hence, the following clearly holds at the point \( X'' \):

\[
DA(X'') \cdot h = \sum_{i=1}^{N} \frac{\partial A(X'')}{\partial x_i} \cdot h_i > 0.
\]

5. The final step proceeds by constructing the function \( f(X) = A(X + (X'' - X)) \), which is simply a translate of \( A(X) \). Evidently, \( f(X) \) belongs to \( A(G) \). We may observe that \( f(X^*) = A(X''') \), and \( DF(X^*) = DA(X'') \). Hence, the desired result is obtained:

\[
Df(X^*) \cdot h = \sum_{i=1}^{N} \frac{\partial f(X^*)}{\partial x_i} \cdot h_i > 0.
\]

As a consequence, Proposition A holds, and class \( A(G) \) is dense in \( E^r(\mathbb{R}^N) \). Q.E.D.

**Proof of Theorem 2:** We simply sketch the main details of the proof. A complete version may be requested from the authors. The proof essentially relies on establishing a connection between additive fuzzy systems and regularization techniques. It proceeds as follows:

1. Take arbitrary compact \( K \subset \mathbb{R}^N \), order \( r \), \( f \in C^r(\mathbb{R}^N) \) and \( \varepsilon > 0 \).

2. It follows from classical results (see [19, Theorem 6.4.15]) that, for every compact \( K \) and every \( f \in C^r(\mathbb{R}^N) \), a function \( g \in C^0(\mathbb{R}^N) \) exists such that \( f(X) = g(X) \) on a compact set \( K' \supset K \) and \( \sup g \in K' \ supset K \).

3. Under the assumptions of Theorem 2, the basis function \( T \) may be used as a kernel in order to define the following convolution:

\[
g \ast T_{\sigma}(X) = \int_{\mathbb{R}^N} g(\mu) \prod_{i=1}^{N} \left( \frac{1}{T} \left( \frac{x_i - \mu_i}{\sigma} \right) \right) d\mu_1 d\mu_2 \ldots d\mu_N
\]

for every \( X \in \mathbb{R}^N \).

4. Since \( T \in C^0(\mathbb{R}) \), classical properties of convolutions [19] guarantee the possibility to \( r \)-uniformly approximate
every \( g \in C^\infty_c(R^N) \) with arbitrary precision \( \varepsilon > 0 \) by the above expression if \( \sigma \) is properly chosen.

5) By using a construction analogous to Mao et al. [16], which involves Riemann sums both in numerator and denominator, the above convolution may be discretized:

\[
B(X) = \frac{\sum_{j=1}^{m} g(\mu_j) \prod_{i=1}^{N} \frac{1}{\sigma} T\left(\frac{x_i - \mu_{ij}}{\sigma}\right)}{\sum_{j=1}^{m} \prod_{i=1}^{N} \frac{1}{\sigma} T\left(\frac{x_i - \mu_{ij}}{\sigma}\right)} .
\]

6) By some standard lemmas, we obtain \( \varepsilon \)-uniform convergence on \( K \) for the above expression.

7) The final step consists of “extending” the support of the constructed fuzzy system to \( R^N \). This may be readily done without affecting approximation accuracy on \( K \), by using a kind of “safety-net” mechanism. As a result we obtain the following expression:

\[
A(X) = \frac{\sum_{j=1}^{m} g(\mu_j) \prod_{i=1}^{N} \frac{1}{\sigma} T\left(\frac{x_i - \mu_{ij}}{\sigma}\right) + \sum_{j=m+1}^{\infty} \prod_{i=1}^{N} \frac{1}{\sigma} T\left(\frac{x_i - \mu_{ij}}{\sigma}\right)}{\sum_{j=1}^{m} \prod_{i=1}^{N} \frac{1}{\sigma} T\left(\frac{x_i - \mu_{ij}}{\sigma}\right) + \sum_{j=m+1}^{\infty} \prod_{i=1}^{N} \frac{1}{\sigma} T\left(\frac{x_i - \mu_{ij}}{\sigma}\right)}
\]

where evidently \( A(X) \in \mathcal{A}(T) \).

8) As a consequence, we may use structures as the above \( A(X) \) to uniformly approximate derivatives to order \( r \) for every \( f \in C^r(R^N) \), to arbitrary precision \( \varepsilon \) on every compact \( K \) of \( R^N \). By a straightforward calculation it may be deduced that this is equivalent to denseness of \( \mathcal{A}(T) \) in \( E^r(R^N) \). Q.E.D.

REFERENCES


Manuel Landajo was born in Avilés, Spain, in 1969. He received the M.S. degree in 1992 and the Ph.D. degree in 1999, both in economics, from the University of Oviedo, Spain. He is an Associate Lecturer of Statistics and Econometrics at the Faculty of Economics, University of Oviedo, since 1993. His research activities focus on the development of flexible techniques for economic modeling and forecasting, with special interest on neural networks and fuzzy systems, and their integration with conventional econometric and time series modeling techniques.

Maria J. Río received the M.S. degree in mathematics, University of Santiago de Compostela, Spain, in 1978 and is currently a Ph.D. candidate with the Department of Applied Economics, University of Oviedo. Since 1978 she has held several teaching posts at the University of Oviedo, where she is currently Full Lecturer of Statistics at the Faculty of Economics. Her research fields include applications of information theory to industrial concentration measurement, economic forecasting, and applications of fuzzy sets in economics.
Rigoberto Pérez received the M.S. degree in mathematics, University of Santiago de Compostela, Spain, in 1978 and the Ph.D. degree in economics, University of Oviedo, Spain, in 1985.

He is currently Professor of Statistics and Econometrics at the University of Oviedo. Simultaneously, he acts as an advisor for several public and private regional institutions. He also leads several research projects of regional econometric modeling (a Spanish branch of the LINK project) as well as initiatives for innovation and quality in university teaching. The author of several articles on information theory and statistical inference on information measures, his current interests are economic modeling and forecasting techniques, including neurofuzzy methods.