From a Gaussian Mixture Model to Additive Fuzzy Systems

Ming-Tao Gan, *Member, IEEE*, Madasu Hanmandlu, *Member, IEEE*, and Ai Hui Tan

***Abstract*—This work explores how a kind of probabilistic system, namely the Gaussian mixture model (GMM), can be translated to an additive fuzzy system. We will prove the mathematical equivalence between the conditional mean of a GMM, and the defuzzified output of a generalized fuzzy model (GFM). The relationship between a GMM and a GFM, and the conditions for GMM to GFM translation will be made explicit in the form of theorems. The work will then extend to special cases of the GFM, specifically the Mamdani–Larsen and Takagi–Sugeno fuzzy models. The possibility of reverse translation, that is, from a GFM to a GMM will also be discussed. Finally, we will consider the generality of a GMM, specifically how it can approximate other distribution functions.**

***Index Terms*—Additive fuzzy systems, conditional means, functional equivalence, Gaussian mixture models, generalized fuzzy models, Mamdani–Larsen models, probabilistic systems, Takagi–Sugeno models.**

I. Introduction

THE mathematical theory of probability was formulated around 1660 [1]. From that time until the mid-twentieth century, uncertainty was almost solely modeled in terms of probability theory. The foundations of probability theory (PT) was continually developed and refined over a period of three centuries [2].

In the 1950s, however, alternative theories on uncertainty began to challenge the monopoly [3] of probability theory. Most notable of these are fuzzy set theory (FST) [4], [5] and possibility theory [6]. Zadeh's paper on fuzzy sets in 1965 [7] was the spark of a lively debate between proponents of PT ("probabilists") and proponents of FST ("fuzzyists"), which have continued to this day. Initial criticisms, especially by probabilists toward FST, were sometimes very harsh. Fuzzyists responded in kind, thus creating a virtual competition between fuzzy and probabilistic approaches.

At the core of the debate is the relationship between PT and FST. A survey of literature yields a variety of claims. Some are indeed valid (i.e., verified through formal mathematical proofs). Others are more of a reaction to criticisms. Still others are due to "dogmatic or overenthusiastic point of views" [8]. What follows is a sampling of claims from both sides of the debate, and where appropriate, responses from the opposition.

Manuscript received December 18, 2002; revised December 1, 2003 and August 31, 2004.

M.-T. Gan and A. H. Tan are with the Faculty of Engineering, Multimedia University, Jalan Multimedia, 63100 Cyberjaya, Selangor, Malaysia (e-mail: mtgan@mmu.edu.my; htai@mmu.edu.my).

M. Hanmandlu is with the Department of Electrical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India (e-mail: mhmandlu@ee.iitd.ernet.in).

Digital Object Identifier 10.1109/TFUZZ.2004.841728

"FST is just PT in disguise." Criticizing fuzzy sets as being nothing but misunderstood, fake or ill-conceived probabilities were popular by "pure" probabilists in the days of FST's infancy. Instances of such literature are by Stallings [9], Cheeseman [10], and Laviolette *et al.* [11], with replies by Dubois and Prade [12], Klir [13] and Zadeh [14]. Cheeseman [10], for example, was correct in interpreting the membership function of a fuzzy set as a likelihood function. However, he asserted, that on such grounds, fuzzy sets are nothing new. Dubois and Prade concurred that it is possible to interpret FST in terms of PT. However there are ways of approaching FST that has nothing to do with PT. In [15], they reviewed two interpretative settings for fuzzy sets in which the notion of probability is no longer needed. One can also consult Mabuchi [16], [17] for a systematic analysis of fuzzy sets and their possible interpretations.

"PT is a special case of FST." Predictably, some fuzzyists reacted to the above claim with this counter-claim. Kosko [18] concluded in this way when he compared fuzzy relative cardinality with conditional probability. Viertl [19] proposed a hierarchy in which probabilistic methods represent but an intermediate stage in the evolution of uncertainty models, while fuzzy models represent the highest stage. Another common assumption by fuzzy logic practitioners is to equate probability degrees with membership grades (see [20] and [21], for instance). Interestingly, this claim, also dubbed the "superset hypothesis" [22], was debunked by fuzzyists themselves. For example, Dubois and Prade [12] stressed that membership function can, *but need not be*, related to probability. As in the counter-argument for the previous claim, FST can be related to PT, or they can be viewed as standing apart from PT. Mathematical objects that behave like fuzzy sets do exist in PT, but this does not mean that fuzziness is reducible to randomness [15].

"PT is superior to FST." Radical claims along this vein were put forward by probabilists such as Cheeseman [23], and Laviolette and Seaman [22]. Some, like Lindley [24], went even further and claimed that "the only satisfactory description of uncertainty is probability." In [22], Laviolette and Seaman criticized how FST handles ambiguity and vagueness as concepts of uncertainty distinct from probability. Utilizing three examples, they alleged that FST handles uncertainty less efficiently than PT. In reply, Klir [13] cited the Cambridge debate [25], where arguments were presented in favor of vagueness as a distinct uncertainty from probability. Dubois and Prade [12] showed how uncertainty can be expressed by ordinal models that are not related to probability.

It is a relief to note that the "PT versus FST" debate has mellowed somewhat as of to-date. More likely, probabilists and fuzzyists are coming to realize that "probability theory and fuzzy logic are complementary rather than competitive" [14]. Like Dubois and Prade [12], we subscribe to a "disjoint set assumption with multiple optional bridges." We agree with them that "instead of considering probability and fuzzy sets are rivals, it sounds more promising to build bridges and take advantage of the enlarged framework for modeling uncertainty and vagueness they conjointly bring us to" [15]. Several domains where FST and PT are conjointly used are fuzzy random variables [26], statistics with fuzzy events [27], modeling of linguistic probabilities with fuzzy sets [28], and more recently, perception-based PT [14], which is a three-stage generalization of PT, i.e., fuzzification, granularization and natural-language generalization.

While it is now clear that PT is *not* FST (and vice versa), we wish to highlight that it is possible to interpret *some* aspects of FST within the framework of PT, and *vice versa*. The history of such relationships is already quite long, and we list here several existing bridges between probability and fuzzy sets ([8], [15])

i) fuzzy sets can be cast in random set theory (see [29] and [30]);
ii) fuzzy sets can be interpreted as likelihood functions (see [31] and [32]);
iii) fuzzy sets can be used in statistical inference (see [33]);
iv) fuzzy set connectives can have probabilistic interpretations (see [15]).

In subsequent sections, we will attempt a further contribution to the bridge from probability to fuzzy logic by explicitly demonstrating how a kind of probabilistic system can be translated into a kind of fuzzy system. The probabilistic system considered will be one modeled by Gaussian mixtures. Fuzzy systems are limited to the additive kind, with multiplicative conjunction and implication operators.

Specifically, we will prove that the conditional mean of a GMM [34] is mathematically equivalent to the defuzzified output of a GFM [35]. We will begin with single-input–single-output (SISO) systems and then extend our proof to multiple-input–single-output (MISO) systems. We will culminate our work by proposing a number of theorems that explicitly describe the relationship between a GMM and the fuzzy systems it translates to.

II. Additive Fuzzy Models

Additive fuzzy models found in the literature can be generally divided into 3 broad types, i.e., the Mamdani–Larsen (ML) Model [36], [37], the Takagi–Sugeno (TS) Model [38], and the recently proposed GFM [35]. While having the same IF-part form, the fuzzy models are distinguished by how the THEN-part of their fuzzy rules is defined

$$\begin{aligned}\textbf{ML}\text{: } R^k &: \text{IF } x_1 \text{ is } \mathbf{A}_1^k \wedge x_2 \text{ is } \mathbf{A}_2^k \wedge \cdots \\ &\wedge x_D \text{ is } \mathbf{A}_D^k \text{ THEN } y \text{ is } \mathbf{B}^k(b_k, v_k) \quad (1)\\ \textbf{TS}\text{: } R^k &: \text{IF } x_1 \text{ is } \mathbf{A}_1^k \wedge x_2 \text{ is } \mathbf{A}_2^k \wedge \cdots \\ &\wedge x_D \text{ is } \mathbf{A}_D^k \text{ THEN } y \text{ is } f^k(\mathbf{x}) \quad (2)\\ \textbf{GFM}\text{: } R^k &: \text{IF } x_1 \text{ is } \mathbf{A}_1^k \wedge x_2 \text{ is } \mathbf{A}_2^k \wedge \cdots \\ &\wedge x_D \text{ is } \mathbf{A}_D^k \text{ THEN } y \text{ is } \mathbf{B}^k(f^k(\mathbf{x}), v_k). \quad (3)\end{aligned}$$

Each rule is premised on the vector $\mathbf{x} = [x_1, x_2, \ldots, x_D]$, and maps fuzzy subsets in the input space $\mathbf{A}^k \subset R^D$ to a fuzzy subset in the output space, $\mathbf{B}^k \subset R$. $\mathbf{A}_d^k$ is a fuzzy set subscribed by the input variable x_d in the kth rule, and ^ is a fuzzy conjunction operator.

$\mathbf{B}^k(b_k, v_k)$, found in the THEN-part of ML's fuzzy rule, is a fuzzy set with centroid b_k and index of fuzziness v_k. If $\phi^k(y)$ is the membership function of $\mathbf{B}^k$, then v_k and b_k are computed by

$$v_k = \int_y \phi^k(y)\, dy \tag{4}$$

$$b_k = \frac{\int_y y\phi^k(y)\, dy}{\int_y \phi^k(y)\, dy}. \tag{5}$$

Equations (4) and (5) are formulas for the area and centroid of $\phi^k(y)$, respectively.

Instead of a fuzzy set on the output space, the TS model has a varying singleton defined by $f^k(\mathbf{x})$. This can be linear or nonlinear, and is a function describing the input–output relationship in a *localized* input–output space. A linear form of $f^k(\mathbf{x})$ is

$$f^k(\mathbf{x}) = b_{k0} + b_{k1}x_1 + \cdots + b_{kD}x_D. \tag{6}$$

Combining the properties of both the ML and the TS models, Azeem *et al.* [35] proposed the GFM with a consequent fuzzy set of the form $\mathbf{B}^k(f^k(\mathbf{x}), v_k)$. $\mathbf{B}^k$ now has a varying centroid, $f^k(\mathbf{x})$ (in contrast with ML's static centroid b_k), while still maintaining its "fuzziness," quantified by v_k (in contrast with the TS's singleton-type consequent).

The firing strength of the kth rule, obtained by taking the fuzzy conjunction (denoted by $\wedge$) of the membership functions of a rule's IF-part, is

$$\mu^k(\mathbf{x}) = \mu_1^k(x_1) \wedge \mu_2^k(x_2) \wedge \cdots \wedge \mu_D^k(x_D) \tag{7}$$

where $\mu_d^k(x_d)$ is the membership function of fuzzy set $\mathbf{A}_d^k$. The firing strength of the kth rule is also represented as a fuzzy set $\mathbf{A}^k \subset R^D$ in the input space. Hence, a fuzzy rule can be written more compactly as

$$R^k : \text{IF } \mathbf{x} \text{ is } \mathbf{A}^k \text{ THEN } y \text{ is } \mathbf{B}^k. \tag{8}$$

Using the fuzzy implication operator (denoted by $\rightarrow$) for mapping fuzzy subsets from the input space $\mathbf{A}^k \subset R^D$ to the fuzzy subset in the output space $\mathbf{B}^k \subset R$, the resultant fuzzy set $\mathbf{B}^{*k}$ has a membership function

$$\phi^{*k}(y) = \mu^k(\mathbf{x}) \rightarrow \phi^k(y). \tag{9}$$

The fuzzy disjunction operator (denoted by $\vee$) is used to join the mapped regions for all K rules in the output space. The aggregated fuzzy set $\mathbf{B}^o$ in the output region is obtained from

$$\mathbf{B}^o = \mathbf{B}^{*1} \vee \mathbf{B}^{*2} \vee \cdots \vee \mathbf{B}^{*K}. \tag{10}$$

Applying the weighted average gravity method for defuzzification, the defuzzified output value y^o is thus given by

$$y^o = \frac{\int_y y\phi^o(y)\,dy}{\int_y \phi^o(y)}\,dy \tag{11}$$

where $\phi^o(y)$ is the resultant membership function of $\mathbf{B}^o \subset R$ in the output space.

Different classes of each of the fuzzy models discussed can be derived on the basis of choices present in the conjunction operator, the implication operator and the disjunction operator. In this work, we shall focus only on the class employing *multiplicative* conjunction, *multiplicative* implication and *additive* disjunction. The defuzzified output formulas corresponding to this class of the ML model, TS model, and GFM are, respectively, obtained from (11) as

$$y^o = \sum_{k=1}^{K} \frac{\mu^k(\mathbf{x}) \cdot v_k}{\sum_{k'=1}^{K} \mu^{k'}(\mathbf{x}) \cdot v_{k'}} \cdot b_k \tag{12}$$

$$y^o = \sum_{k=1}^{K} \frac{\mu^k(\mathbf{x})}{\sum_{k'=1}^{K} \mu^{k'}(\mathbf{x})} f^k(\mathbf{x}) \tag{13}$$

$$y^o = \sum_{k=1}^{K} \frac{\mu^k(\mathbf{x}) \cdot v_k}{\sum_{k'=1}^{K} \mu^{k'}(\mathbf{x}) \cdot v_{k'}} f^k(\mathbf{x}). \tag{14}$$

III. GMM

Mixture distributions are statistical distributions expressed as superposition of compound distributions [34]. The most widely used finite mixture distributions are those involving normal components.

The Gaussian mixture probability density function (pdf) for a vector random variable $\mathbf{x}$, of dimension D, has the following form [34]:

$$G(\mathbf{x}; \mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{c=1}^{C} p_c N^D(\mathbf{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \tag{15}$$

where $\mathbf{p} = (p_1, p_2, \ldots, p_{C-1})$ are the $C-1$ independent mixing proportions of the mixture and are such that

$$0 < p_c < 1 \quad \text{and} \quad p_C = 1 - \sum_{c=1}^{C-1} p_c. \tag{16}$$

$\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_C)$ is a (D by CD) matrix, and $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_C)$ is a (D by C) matrix, in which $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ are the mean vector and covariance matrix, respectively. $N^D(\mathbf{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$ is the c-th component multivariate (D-dimensional) normal density, given by

$$\begin{aligned} N^D(\mathbf{x}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) &= (2\pi)^{-D/2} |\boldsymbol{\Sigma}_c|^{-1/2} \\ &\quad \times \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_c)' \boldsymbol{\Sigma}_c^{-1} (\mathbf{x} - \boldsymbol{\mu}_c)\right\}. \end{aligned} \tag{17}$$

IV. FROM A GMM TO A GFM—SISO CASE

Consider a dataset with one independent variable, x, and one dependent variable, y. Let the dataset be drawn from C separate normal distributions with different means $(\mu_1, \mu_2, \ldots, \mu_C)$ and covariances $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_C)$. Hence, we can approximate the joint pdf of (x, y) by a Gaussian mixture with C components

$$G(x, y) = \sum_{c=1}^{C} p_c N^2 \left(\begin{bmatrix} x \\ y \end{bmatrix}; \begin{bmatrix} \mu_{cx} \\ \mu_{cy} \end{bmatrix}, \boldsymbol{\Sigma}_c \right). \tag{18}$$

With $G(x, y)$, it is possible to estimate the expected value y_e, when x is given

$$\begin{aligned} y_e = E[y \,|\, x] &= \int_{-\infty}^{+\infty} yG(y \,|\, x)\,dy = \frac{\int_{-\infty}^{+\infty} yG(x, y)\,dy}{G(x)} \\ &= \frac{\int_{-\infty}^{+\infty} yG(x, y)\,dy}{\int_{-\infty}^{+\infty} G(x, y)\,dy}. \end{aligned} \tag{19}$$

The expansion of *one term* of $G(x, y)$ is given by

$$\begin{aligned} G_c(x, y) &= p_c N^2 \left(\begin{bmatrix} x \\ y \end{bmatrix}; \begin{bmatrix} \mu_{cx} \\ \mu_{cy} \end{bmatrix}, \boldsymbol{\Sigma}_c \right) \\ &= \frac{p_c}{2\pi\sqrt{|\boldsymbol{\Sigma}_c|}} \exp\left\{-\frac{1}{2} [x - \mu_{cx} \quad y - \mu_{cy}] \boldsymbol{\Sigma}_c^{-1} \begin{bmatrix} x - \mu_{cx} \\ y - \mu_{cy} \end{bmatrix}\right\} \end{aligned} \tag{20}$$

where $\boldsymbol{\Sigma}_c$ is the covariance matrix given by

$$\boldsymbol{\Sigma}_c = \begin{bmatrix} \sigma_{cxx} & \sigma_{cxy} \\ \sigma_{cyx} & \sigma_{cyy} \end{bmatrix}.$$

Note that $\sigma_{cxy} = \sigma_{cyx}$ (symmetric matrix), and $|\boldsymbol{\Sigma}_c|$ and $\boldsymbol{\Sigma}_c^{-1}$ are the determinant and inverse of $\boldsymbol{\Sigma}_c$, respectively.

The marginal pdf of $x, G_c(x)$, is given by

$$\begin{aligned} G_c(x) &= \int_{-\infty}^{+\infty} G_c(x, y)\,dy \\ &= p_c \int_{-\infty}^{+\infty} \frac{1}{2\pi\sqrt{|\boldsymbol{\Sigma}_c|}} \\ &\quad \times \exp\left\{-\frac{1}{2} [x - \mu_{cx} \quad y - \mu_{cy}] \boldsymbol{\Sigma}_c^{-1} \begin{bmatrix} x - \mu_{cx} \\ y - \mu_{cy} \end{bmatrix}\right\} dy \\ &= p_c \frac{1}{\sqrt{2\pi\sigma_{cxx}}} \exp\left\{-\frac{1}{2} \frac{(x - \mu_{cx})^2}{\sigma_{cxx}}\right\} \\ &= p_c N^1(x; \mu_{cx}, \sigma_{cxx}). \end{aligned} \tag{21}$$

Since $G(x,y) = G_1(x,y) + G_2(x,y) + \cdots + G_C(x,y)$, hence

$$\begin{aligned} G(x) &= \int_{-\infty}^{+\infty} G(x,y)\,dy \\ &= \int_{-\infty}^{+\infty} G_1(x,y)\,dy + \int_{-\infty}^{+\infty} G_2(x,y)\,dy + \cdots \\ &\quad + \int_{-\infty}^{+\infty} G_C(x,y)\,dy \\ &= G_1(x) + G_2(x) + \cdots + G_C(x) \\ &= p_1 N^1(x;\mu_{1x},\sigma_{1xx}) + p_2 N^1(x;\mu_{2x},\sigma_{2xx}) + \cdots \\ &\quad + p_C N^1(x;\mu_{Cx},\sigma_{Cxx}). \end{aligned} \tag{22}$$

The conditional pdf and hence the conditional expectation of y then becomes (23) and (24), as shown at the bottom of the page. Since the forms of all C integrals in (24) are similar, we will derive the result for only one term, and extend it to the others (see Appendix A). Substituting (A.2) and (22) into (24), we have

$$E[y\,|\,x] = \frac{\sum_{c=1}^{C} p_c N^1(x;\mu_{cx},\sigma_{cxx}) \left(\mu_{cy} + (x-\mu_{cx})\frac{\sigma_{cxy}}{\sigma_{cxx}}\right)}{\sum_{c'=1}^{C} p_{c'} N^1(x;\mu_{c'x},\sigma_{c'xx})}. \tag{25}$$

This is an explicit formula to compute the statistical expectation of the output of a SISO system, when it is known (or assumed) that the input and output are jointly distributed according to a Gaussian mixture.

Note that (25) is similar in form to (14), and both are equivalent if the following conditions are true.

Conditions I:

i) The number of rules K in the GFM rule base equals the number of components C in the GMM, i.e.,

$$\boxed{K = C}$$

for $k = c = 1, 2, \ldots, C$.

ii) The weight of each rule, v_k, is given by the corresponding prior of the mixture model p_c, i.e.,

$$\boxed{v_k = p_c}.$$

iii) The regression function in the THEN-part of the GFM rules is linear, i.e.,

$$\boxed{f^k(x) = \mu_{cy} + (x-\mu_{cx})\frac{\sigma_{cxy}}{\sigma_{cxx}}}.$$

iv) The membership function in the IF-part of the GFM rules, $\mu^k(x)$, is a Gaussian function with mean μ_{cx} and variance σ_{cxx}, i.e.,

$$\boxed{\mu^k(x) = N^1(x;\mu_{cx},\sigma_{cxx})}.$$

v) The fuzzy system is additive, with multiplicative implication.

These are the conditions under which there is mathematical equivalence between the output of a GFM and the expected output of a GMM. What this means is that for any SISO system whose input-output joint probability distribution is known to be a Gaussian mixture, there exists, under **Conditions I**, a fuzzy system which equivalently models its expected output.

The GMM to GFM translation can be stated concisely in **Theorem 1**. We denote the input as x and the output as y.

Theorem 1: If the input–output relationship of a SISO system follows a Gaussian Mixture probability distribution as defined in (18), then the system can be modeled by a GFM under **Conditions I**.

Can **Theorem 1** be generalized to MISO systems? The next section attempts to derive the mathematical equivalence between a multi-input GMM and a multi-input GFM.

V. From a GMM to a GFM—MISO Case

A MISO system is representative of all possible systems because

i) by virtue of having multiple inputs, it is more general than a SISO system;
ii) a MIMO system can be decomposed into several MISO systems.

Let the input vector to the MISO system be $\mathbf{x}$, and the output be y. The first J variables form the input vector while the $(J+1)$th variable is the output. We can rewrite the general form of the Gaussian mixture for the multiple-input case as

$$G(\mathbf{x},y) = \sum_{c=1}^{C} p_c N^{J+1}\left(\begin{bmatrix}\mathbf{x}\\ y\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_{c\mathbf{x}}\\ \mu_{cy}\end{bmatrix}, \boldsymbol{\Sigma}_c\right) \tag{26}$$

where

$$\boldsymbol{\Sigma}_c = \left[\begin{array}{c|c} \{\sigma_{cij}\}_{J\times J} & \{\sigma_{cj(J+1)}\}_{J\times 1} \\ \hline \{\sigma_{c(J+1)j}\}_{1\times J} & \sigma_{c(J+1)(J+1)} \end{array}\right],$$

for $i,j = 1,2,\ldots,J$ $\quad \boldsymbol{\mu}_{c\mathbf{x}} = [\mu_{c1} \quad \cdots \quad \mu_{cJ}]'$

$$G(y\,|\,x) = \frac{G(x,y)}{G(x)} = \frac{G_1(x,y) + G_2(x,y) + \cdots + G_C(x,y)}{G_1(x) + G_2(x) + \cdots + G_C(x)} \tag{23}$$

and

$$E[y\,|\,x] = \int_{-\infty}^{+\infty} yG(y\,|\,x)\,dy = \frac{\int_{-\infty}^{+\infty} yG_1(x,y)\,dy + \int_{-\infty}^{+\infty} yG_2(x,y)\,dy + \cdots + \int_{-\infty}^{+\infty} yG_C(x,y)\,dy}{G(x)} \tag{24}$$

and $\mu_{cy} = \mu_{c(J+1)}$. Note that $\mathbf{\Sigma}_c$ is a symmetric matrix, where $\sigma_{cjk} = \sigma_{ckj}$.

The marginal pdf of $\mathbf{x}$ for the cth term of $G(\mathbf{x}, y)$ is given by

$$\begin{aligned}G_c(\mathbf{x}) &= \int_{-\infty}^{+\infty} G_c(\mathbf{x}, y)\, dy \\ &= \int_{-\infty}^{+\infty} p_c N^{J+1}\left(\begin{bmatrix}\mathbf{x}\\ y\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_{c\mathbf{x}}\\ \mu_{cy}\end{bmatrix}, \mathbf{\Sigma}_c\right) dy \\ &= \int_{-\infty}^{+\infty} \frac{p_c}{(2\pi)^{\frac{J+1}{2}}\sqrt{|\mathbf{\Sigma}_c|}} \\ &\quad \times \exp\left\{-\frac{1}{2}\begin{bmatrix}\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}\\ y-\mu_{cy}\end{bmatrix}' \mathbf{\Sigma}_c^{-1}\begin{bmatrix}\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}\\ y-\mu_{cy}\end{bmatrix}\right\} dy \\ &= \frac{p_c}{(2\pi)^{\frac{J}{2}}\sqrt{|\boldsymbol{\sigma}_{c\mathbf{xx}}|}} \exp\left\{-\frac{1}{2}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_{c\mathbf{xx}}^{-1}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]\right\} \\ &= p_c N^J(\mathbf{x}; \boldsymbol{\mu}_{c\mathbf{x}}, \boldsymbol{\sigma}_{c\mathbf{xx}}) \end{aligned} \tag{27}$$

where $\boldsymbol{\sigma}_{c\mathbf{xx}}$ is the *minor* of the matrix $\mathbf{\Sigma}_c$, after eliminating the $(J+1)$th row and $(J+1)$th column (see Appendix B). Hence, the marginal pdf of $\mathbf{x}$ is

$$\begin{aligned}G(\mathbf{x}) &= \int_{-\infty}^{+\infty} G(\mathbf{x}, y)\, dy \\ &= \sum_{c=1}^{C}\int_{-\infty}^{+\infty} p_c N^{J+1}\left(\begin{bmatrix}\mathbf{x}\\ y\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_{c\mathbf{x}}\\ \mu_{cy}\end{bmatrix}, \mathbf{\Sigma}_c\right) dy \\ &= \sum_{c=1}^{C} p_c N^J(\mathbf{x}; \boldsymbol{\mu}_{c\mathbf{x}}, \boldsymbol{\sigma}_{c\mathbf{xx}}). \end{aligned} \tag{28}$$

The conditional pdf in (23) becomes

$$G(y \,|\, \mathbf{x}) = \frac{\sum_{c=1}^{C} p_c N^{J+1}\left(\begin{bmatrix}\mathbf{x}\\ y\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_{c\mathbf{x}}\\ \mu_{cy}\end{bmatrix}, \mathbf{\Sigma}_c\right)}{\sum_{c'=1}^{C} p_{c'} N^J(\mathbf{x}; \boldsymbol{\mu}_{c'\mathbf{x}}, \boldsymbol{\sigma}_{c'\mathbf{xx}}).} \tag{29}$$

Now, we compute the conditional expectation of y: (24)

$$\begin{aligned}E[y \,|\, \mathbf{x}] &= \int_{-\infty}^{+\infty} y \frac{\sum_{c=1}^{C} p_c N^{J+1}\left(\begin{bmatrix}\mathbf{x}\\ y\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_{c\mathbf{x}}\\ \mu_{cy}\end{bmatrix}, \mathbf{\Sigma}_c\right)}{\sum_{c'=1}^{C} p_{c'} N^J(\mathbf{x}; \boldsymbol{\mu}_{c'\mathbf{x}}, \boldsymbol{\sigma}_{c'\mathbf{xx}})} dy \\ &= \frac{\sum_{c=1}^{C} \int_{-\infty}^{+\infty} y p_c N^{J+1}\left(\begin{bmatrix}\mathbf{x}\\ y\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_{c\mathbf{x}}\\ \mu_{cy}\end{bmatrix}, \mathbf{\Sigma}_c\right) dy}{G(\mathbf{x}).} \end{aligned} \tag{30}$$

The form of the integrals for all C terms in (30) is similar, so we will derive the result for the cth term and apply it to the others (see Appendix B). Substituting (B.6) into (30), we obtain

$$E[y \,|\, \mathbf{x}] = \frac{\sum_{c=1}^{C} p_c N^J(\mathbf{x}; \boldsymbol{\mu}_{c\mathbf{x}}, \boldsymbol{\sigma}_{c\mathbf{xx}})\left(\mu_{cy} - \frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sigma_c^{yy}}\right)}{\sum_{c'=1}^{C} p_{c'} N^J(\mathbf{x}; \boldsymbol{\mu}_{c'\mathbf{x}}, \boldsymbol{\sigma}_{c'\mathbf{xx}})} \tag{31}$$

where $\boldsymbol{\sigma}_{c\mathbf{xx}}$ is a partition of the covariance matrix, $\mathbf{\Sigma}_c$, while $\boldsymbol{\sigma}_c^{\mathbf{x}y}$ and σ_c^{yy} are partitions of its inverse.

Equation (31) does not match any of the defuzzification formulas in Section II. However, if we introduce the assumption that the input variables are *mutually independent* (not unreasonable statistically), then the covariance matrix becomes

$$\mathbf{\Sigma}_c = \left[\begin{array}{c|c}\mathrm{diag}\{\sigma_{cjj}\}_{J\times J} & \{\sigma_{cj(J+1)}\}_{J\times 1} \\ \hline \{\sigma_{c(J+1)j}\}_{1\times J} & \sigma_{c(J+1)(J+1)}\end{array}\right], \quad \text{for } j = 1, 2, \ldots, J. \tag{32}$$

Note that since the input variables are independent of each other, their *covariances* are zero. However, the $(J+1)$th row and the $(J+1)$th column are generally nonzero, since we cannot assume that the input variables are independent with respect to the output variable.

With (32), (31) now becomes

$$E[y \,|\, \mathbf{x}] = \frac{\sum_{c=1}^{C} p_c \prod_{j=1}^{J} N^J(x_j; \mu_{cj}, \sigma_{cjj})\left(\mu_{cy} - \frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sigma_c^{yy}}\right)}{\sum_{c'=1}^{C} p_{c'} \prod_{j=1}^{J} N^J(x_j; \mu_{c'j}, \sigma_{c'jj})}. \tag{33}$$

Comparing (33) with (14), we see that they are equivalent under the following conditions.

Conditions II:

i) The number of rules, K, in the GFM rule base equals the number of components, C, in the GMM, i.e.,

$$\boxed{K = C}$$

for $k = c = 1, 2, \ldots, C$.

ii) The weight of each rule, v_k, is given by the corresponding prior of the mixture model p_c, i.e.,

$$\boxed{v_k = p_c}.$$

iii) The regression function in the THEN-part of the GFM rules is linear, and is a function of all input variables, given by

$$\boxed{f^k(\mathbf{x}) = \mu_{cy} - \frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sigma_c^{yy}}}.$$

iv) The membership function for each of the IF-part variables is a Gaussian function, i.e.,

$$\boxed{\mu^k(x_j) = N^1(x_j; \mu_{cj}, \sigma_{cjj})}, \quad \text{for } j = 1, 2, \ldots, J.$$

v) The fuzzy system is additive, with multiplicative conjunction and implication.

Parts i), ii), and iv) of **Conditions II** are exactly the same as those of **Conditions I**. Part iii) also bears some similarity, except that $f^k(\mathbf{x})$ is now a linear function of J input variables, instead of a single input variable. Part v) constrains the conjunctive and implicative fuzzy operators to the multiplicative kind, and the disjunctive fuzzy operator to the additive kind.

To correspond with the form of the system output in (33), the rule base of the GFM in (3) can be rewritten as

$$\begin{aligned} R^k: \text{IF } x_1 \text{ is } & \mathbf{A}_1^k(N^1(x_1,\mu_{k1},\sigma_{k11})) \\ & \wedge x_2 \text{ is } \mathbf{A}_2^k(N^1(x_2,\mu_{k2},\sigma_{k22})) \wedge \cdots \\ & \wedge x_J \text{ is } \mathbf{A}_J^k(N^1(x_J,\mu_{kJ},\sigma_{kJJ})) \\ & \text{THEN } y \text{ is } \mathbf{B}^k\left(\mu_{ky} - \frac{[\mathbf{x}-\boldsymbol{\mu}_{k\mathbf{x}}]'\boldsymbol{\sigma}_k^{\mathbf{x}y}}{\sigma_k^{yy}}, v_k\right) \end{aligned} \quad (34)$$

for $k = 1, 2, \ldots, C$.

It is clear that our conclusions for SISO systems also hold for MISO systems. Provided that **Conditions II** are satisfied, the output of a multiple-input GFM and the expected output of a multiple-input GMM are mathematically equivalent. Thus, for any MISO system with statistically independent inputs, and whose input–output joint probability distribution is known to be a Gaussian Mixture, there exists a GFM that equivalently models its expected output under **Conditions II**.

The GMM to GFM translation for MISO systems is formally expressed by **Theorem 2**.

Theorem 2: Suppose we have a MISO system with J inputs, where the inputs are mutually independent. If the input–output relationship obeys a Gaussian mixture probability distribution as defined in (26), then the system can be modeled by a GFM under **Conditions II**.

Theorems 1 and **2** are bridges between a type of probabilistic system and a type of fuzzy system. In essence, they allow us to interpret a probabilistic system modeled by Gaussian Mixtures from within a fuzzy logic framework. A note of caution is in order, however. Just because a probabilistic system *can* be interpreted as a fuzzy system does not mean it *should* be. As discussed in Section I, a fuzzy system is generally *not* a probabilistic system (and *vice versa*). It will serve us well not to repeat the same mistaken conclusions of past debates.

VI. Special Cases

Azeem *et al.* [35] showed that the GFM could be reduced to either the ML model or the TS model. It may follow that the corresponding Gaussian mixture pdf can also be reduced to accommodate these special cases. We will now attempt to derive the pdfs that can be translated into the ML and TS fuzzy models, and deduce their statistical implications.

A. From a GMM to a ML Fuzzy Model

Recall the defuzzification formula of the ML model given in (12). Equation (12) is equal to the defuzzification formula of the GFM if the regression function in (14) is reduced to a constant, i.e., $f^k(\mathbf{x}) = b_k$. We know from the previous derivations that if a GFM was translated from a GMM, then

$$f^k(\mathbf{x}) = \mu_{ky} - \frac{[\mathbf{x}-\boldsymbol{\mu}_{k\mathbf{x}}]'\boldsymbol{\sigma}_k^{\mathbf{x}y}}{\sigma_k^{yy}}. \quad (35)$$

As we can see, $f^k(\mathbf{x})$ in (35) is generally *not* a constant. However, it can be made a constant by eliminating all its variables, i.e., $\mathbf{x}$. This can be achieved by setting $\boldsymbol{\sigma}_k^{\mathbf{x}y}$ to be a zero vector. *$\boldsymbol{\sigma}_k^{\mathbf{x}y}$ becomes zero when $\boldsymbol{\sigma}_{k\mathbf{x}y}$ in $\boldsymbol{\Sigma}_k$ becomes zero.*

Proof: Recall the expressions for $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}^{-1}$ in (B.1) and (B.2). We have already imposed the constraint that $\boldsymbol{\sigma}_{\mathbf{xx}} = [\sigma_{cij}] = 0$ when $i \neq j$. To force $\boldsymbol{\sigma}^{\mathbf{x}y}$ (and, therefore, $\boldsymbol{\sigma}^{y\mathbf{x}}$) to become a zero vector, we impose another constraint on $\boldsymbol{\Sigma}$, where we set $\boldsymbol{\sigma}_{\mathbf{x}y}$ and $\boldsymbol{\sigma}_{y\mathbf{x}}$ to zero vectors. Then, $\boldsymbol{\Sigma}$ becomes a diagonal matrix

$$\boldsymbol{\Sigma} = \mathrm{diag}\{\sigma_{jj}\}, \qquad j = 1, 2, \ldots, J+1. \quad (36)$$

The inverse of diagonal matrix $\boldsymbol{\Sigma}$ is also a diagonal matrix, with the entries of $\boldsymbol{\Sigma}^{-1}$ having the inverse of the entries of $\boldsymbol{\Sigma}$ [39, p. 368]. Hence

$$\boldsymbol{\Sigma}^{-1} = \mathrm{diag}\left\{\frac{1}{\sigma_{jj}}\right\}, \qquad j = 1, 2, \ldots, J+1 \quad (37)$$

which gives $\boldsymbol{\sigma}^{\mathbf{x}y}$ and $\boldsymbol{\sigma}^{y\mathbf{x}}$ as zero vectors. The proof ends here.□

In view of (36) and (37), (35) becomes $f^k(\mathbf{x}) = \mu_{ky}$. So, we obtain the condition

$$f^k(\mathbf{x}) = b_k = \mu_{ky} \quad (38)$$

which will replace part iii) of **Conditions II** for the special case of an ML model.

Note that $\boldsymbol{\sigma}_{\mathbf{x}y}$ is the vector of covariances between the input variables and the output variable. When $\boldsymbol{\sigma}_{\mathbf{x}y}$ becomes zero, this implies that the output variable is *independent* of the input variables. *This is generally not a good statistical assumption.*

We propose the following corollary describing the statistical properties of a system that translates to a ML model.

Corollary 1: If the input–output relationship of a system follows a Gaussian mixture pdf and if its inputs and output are mutually independent of each other, then the GFM in **Theorem 2** can be reduced to a ML model such that for the kth rule, the regression function in the THEN-part is a *constant*, given by $f^k(\mathbf{x}) = \mu_{ky}$, which is the Gaussian mean of the output. All parts in **Conditions II**, except part iii), remain unchanged.

B. From a GMM to a TS Fuzzy Model

Recall the defuzzification formula for the TS model given in (13). Comparing (13) with (14), we see that they are equal if all the weights of the GFM, v_k, equals *unity*. This is not possible for GMM to GFM translation, since the weights in the GFM correspond to the priors of the GMM, and the priors are constrained by (16). However, a simple algebraic manipulation of (14) will also yield (13), and yet satisfy the constraint. If we set v_k in (14) to $(1/K)$ for all k, we obtain

$$\begin{aligned} y^o &= \sum_{k=1}^{K} \frac{\mu^k(\mathbf{x}) \cdot \frac{1}{K}}{\sum_{k'=1}^{K} \mu^{k'}(\mathbf{x}) \cdot \frac{1}{K}} f^k(\mathbf{x}) \\ &= \sum_{k=1}^{K} \frac{\mu^k(\mathbf{x})}{\sum_{k'=1}^{K} \mu^{k'}(\mathbf{x})} f^k(\mathbf{x}) \end{aligned}$$

which equals (13). Note that K is the number of rules and $\sum_{k=1}^{K} p_k = \sum_{k=1}^{K} v_k = \sum_{k=1}^{K}(1/K) = 1$, which satisfies the constraint on the priors.

So the TS model, being a special case of the GFM, is further related to a special case of the Gaussian Mixture pdf with *equal priors*. This relation can be stated formally with Corollary 2.

Corollary 2: If the input–output relationship of a system follows a Gaussian mixture pdf of *equal priors*, with each prior being $(1/K)$, where K is the number of Gaussian components, then the GFM in **Theorem 2** can be reduced to a TS model with *equal weights* (or *unweighted* TS model). Barring part ii), all other parts of **Conditions II** remain unchanged.

VII. Discussions

A. Reconstructing the GMM From a Given Fuzzy System

We have seen from **Theorems 1** and **2** how a system whose inputs and output are Gaussian mixture jointly distributed can be modeled using the GFM. Now, we are interested to investigate the reverse case. *Suppose we begin with an additive fuzzy system whose rule base is known, is it possible to derive the corresponding GMM?*

Let the output of a fuzzy system be given by

$$y^o = \sum_{k=1}^{K} \frac{v_k \prod_{j=1}^{J} N^1(x_j, \mu_{kj}, \sigma_{kjj})}{\sum_{k'=1}^{K} v_{k'} \prod_{j=1}^{J} N^1(x_j, \mu_{k'j}, \sigma_{k'jj})} \times (a_{k0} + a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kJ}x_J). \quad (39)$$

Note that y^o was derived from the rule base of an additive fuzzy system with multiplicative conjunction and implication operators.

We also expand (33), the equation of the conditional mean of a GMM by zooming in on the portion of the regression function shown in (40) at the bottom of the page. To reconstruct the GMM, we need to find the values of $p_c, \mu_{c1}, \mu_{c2}, \ldots, \mu_{cJ}, \mu_{cy}, \sigma_{c11}, \sigma_{c22}, \ldots, \sigma_{cJJ}, \sigma_c^{1y}, \sigma_c^{2y}, \ldots, \sigma_c^{Jy}$, and σ_c^{yy}. Comparing (39) and (40), we can write the following set of simultaneous equations (note that $c = k$):

$$p_c = v_k \quad (41a)$$

$$\mu_{c1} = \mu_{k1}; \quad \mu_{c2} = \mu_{k2}; \ldots; \mu_{cJ} = \mu_{kJ} \quad (41b)$$

$$\sigma_{c11} = \sigma_{k11}; \quad \sigma_{c22} = \sigma_{k22}; \ldots; \sigma_{cJJ} = \sigma_{kJJ} \quad (41c)$$

$$\mu_{cy} + \frac{\sigma_c^{1y}}{\sigma_c^{yy}}\mu_{c1} + \frac{\sigma_c^{2y}}{\sigma_c^{yy}}\mu_{c2} + \cdots + \frac{\sigma_c^{Jy}}{\sigma_c^{yy}}\mu_{cJ} = a_{k0} \quad (41d)$$

$$-\frac{\sigma_c^{1y}}{\sigma_c^{yy}} = a_{k1}; \quad -\frac{\sigma_c^{2y}}{\sigma_c^{yy}} = a_{k2}; \ldots; -\frac{\sigma_c^{Jy}}{\sigma_c^{yy}} = a_{kJ}. \quad (41e)$$

Equations (41a)–(41c) can be solved directly but (41d) and (41e) require further manipulation. If we substitute (41e) into (41d), we obtain

$$\mu_{cy} - a_{k1}\mu_{c1} - a_{k2}\mu_{c2} - \cdots - a_{kJ}\mu_{cJ} = a_{k0}.$$

It is possible to solve for μ_{cy} since the values of $\mu_{c1}, \mu_{c2}, \ldots, \mu_{cJ}$ are known from (41b)

$$\mu_{cy} = a_{k0} + a_{k1}\mu_{c1} + a_{k2}\mu_{c2} + \cdots + a_{kJ}\mu_{cJ}. \quad (41f)$$

This leaves us with $\sigma_c^{1y}, \sigma_c^{2y}, \ldots, \sigma_c^{Jy}$, and σ_c^{yy}. Looking at (41e), we see that we have a set of J equations in $J+1$ unknowns. Hence, there does not exist a unique solution for these unknowns. In other words, these unknowns can take several possible values.

If we let $\sigma_c^{yy} = t$, where t is an arbitrary value, then (41e) yields

$$\sigma_c^{1y} = -a_{k1}t; \quad \sigma_c^{2y} = -a_{k2}t; \ldots; \sigma_c^{Jy} = -a_{kJ}t.$$

The solution for $\sigma_c^{1y}, \sigma_c^{2y}, \ldots, \sigma_c^{Jy}$, and σ_c^{yy} can be written in vector form

$$\begin{bmatrix}\sigma_c^{1y} & \sigma_c^{2y} & \cdot & \cdot & \sigma_c^{Jy} & \sigma_c^{yy}\end{bmatrix}' = t[-a_{k1} \quad -a_{k2} \quad \cdot \quad \cdot -a_{kJ} \quad 1]'. \quad (41g)$$

Hence, we conclude that the reverse translation, i.e., the GFM to GMM translation is ***one-to-many****. In other words, it is possible to reconstruct the GMM given the GFM, but the GMM is* ***not unique****.*

What we have discovered here is surprisingly analogous to how fuzzy sets relate to random set theory. Quoting Dubois *et al.* [8, p. 351]

> "... for a given [membership function] $\mathrm{F}(\cdot)$, there exists some random set $\Re$ (*not unique*), such that the one-point-coverage function of $\Re$, namely $\mathrm{u} \rightarrow \mathrm{P}(\mathrm{u} \in \Re)$ is precisely F. In general, without additional assumptions on the membership function F, we know ... that F corresponds to the one-point coverage function of *more than one random set.*" (emphasis ours)

B. THEN-Part Membership Functions

Both v_k and b_k appear in the THEN-part of fuzzy rules, and are defined in (4) and (5). Recall that b_k, in the context of this work, is only valid for the ML model, while v_k is valid for both ML and GFM models.

We see that both v_k and b_k are dependent on $\phi^k(y)$, i.e., the THEN-part membership function. *To establish a connection between $\phi^k(y)$ and the GMM, we introduce an extra condition.*

Condition III: The THEN-part membership function of the kth rule, $\phi^k(y)$, is equivalent to the output marginal pdf of the kth component of the GMM.

$$\begin{aligned} E[y\,|\,\mathbf{x}] &= \frac{\sum_{c=1}^{C}\cdots}{\sum_{c'=1}^{C}\cdots}\left(\mu_{cy} - \frac{[x_1-\mu_{c1} \quad x_2-\mu_{c2} \quad \cdot \quad \cdot \quad x_J-\mu_{cJ}]\left[\sigma_c^{1y} \quad \sigma_c^{2y} \quad \cdot \quad \cdot \quad \sigma_c^{Jy}\right]'}{\sigma_c^{yy}}\right) \\ &= \frac{\sum_{c=1}^{C}\cdots}{\sum_{c'=1}^{C}\cdots}\left\{\left(\mu_{cy} + \frac{\sigma_c^{1y}}{\sigma_c^{yy}}\mu_{c1} + \frac{\sigma_c^{2y}}{\sigma_c^{yy}}\mu_{c2} + \cdots + \frac{\sigma_c^{Jy}}{\sigma_c^{yy}}\mu_{cJ}\right)\right. \\ &\quad \left. -\frac{\sigma_c^{1y}}{\sigma_c^{yy}}x_1 - \frac{\sigma_c^{2y}}{\sigma_c^{yy}}x_2 - \cdots - \frac{\sigma_c^{Jy}}{\sigma_c^{yy}}x_J\right\}. \end{aligned} \quad (40)$$

What follows is a consequence of **Condition III**. The *marginal pdf of the output* is computed from the GMM as

$$\begin{aligned} G(y) &= \int_{-\infty}^{+\infty} G(\mathbf{x}, y) d\mathbf{x} \\ &= \sum_{c=1}^{C} \int_{-\infty}^{+\infty} p_c N^{J+1}\left(\begin{bmatrix}\mathbf{x} \\ y\end{bmatrix}; \begin{bmatrix}\boldsymbol{\mu}_{c\mathbf{x}} \\ \mu_{cy}\end{bmatrix}, \boldsymbol{\Sigma}_c\right) d\mathbf{x} \\ &= \sum_{c=1}^{C} p_c N^1(y; \mu_{cy}, \sigma_{cyy}). \end{aligned} \tag{42}$$

Hence, the kth component of $G(y)$ is

$$G_k(y) = p_k N^1(y; \mu_{ky}, \sigma_{kyy}). \tag{43}$$

Equating $\phi^k(y)$ with $G_k(y)$ and substituting into (4), we obtain

$$\begin{aligned} v_k &= \int_{-\infty}^{+\infty} p_k N^1(y; \mu_{ky}, \sigma_{kyy})\, dy \\ &= p_k \int_{-\infty}^{+\infty} N^1(y; \mu_{ky}, \sigma_{kyy})\, dy = p_k \end{aligned} \tag{44}$$

imposing that the area under a Gaussian curve is unity. *The result in* (44) *agrees with part ii) of* **Conditions II** *for both the GFM* (**Theorem 2**) *and the ML model* (**Corollary 1**).

Now, substituting $\phi^k(y)$ with $G_k(y)$ in (5), we obtain

$$\begin{aligned} b_k &= \frac{\int_{-\infty}^{+\infty} y p_k N^1(y; \mu_{ky}, \sigma_{kyy})\, dy}{\int_{-\infty}^{+\infty} p_k N^1(y; \mu_{ky}, \sigma_{kyy})\, dy} \\ &= \int_{-\infty}^{+\infty} y N^1(y; \mu_{ky}, \sigma_{kyy})\, dy = \mu_{ky}. \end{aligned} \tag{45}$$

The result in (45) *agrees with part iii) of* **Conditions II** *for the ML model* (**Corollary 1**).

So, we see that (4) and (5) will yield v_k and b_k similar to their respective counterparts in the conditional mean of the GMM, if $\phi^k(y)$ is taken to be the marginal pdf of the output (i.e., **Condition III** is satisfied). The results in (44) and (45) are another rung in the bridge between the GMM and additive fuzzy systems.

C. Generality of a GMM

In this section, we turn our attention to systems modeled by probability distributions other than GMM. *Are* **Theorem 1** *and* **Theorem 2** *applicable in such cases, or are they just limited to a narrow class of systems?* This question is concerned with the generality of a GMM which, in this specific context, refers to whether a GMM can be used to approximate other distribution functions. This is just a *function approximation* problem. We are trying to decompose an arbitrary function (i.e., the probability distribution of interest) into Gaussian components.

Function approximation with Gaussian mixtures is, in fact, far from impossible. Park and Sandberg [40] proved the *universal approximation property* of radial basis function (RBF) networks. An RBF network is just a *linear superposition* of RBFs, of which Gaussian functions are a type. Jorge and Ferreira [41] zoomed in on approximation by superposition of Gaussians, while imposing distinct or weaker restrictions on the functions to be approximated (for example, continuity is not required).

Putting these results together lead us to conclude that a GMM can approximate *any* probability distribution to arbitrary accuracy, provided that.

Conditions IV:

i) The number of components, C, is sufficiently large;
ii) the parameters of the model (i.e., p, Σ, μ) are chosen correctly.

Therefore, **Theorem 1** and **2** can be applied to a system with any probability distribution, provided that

i) its probability distribution is known;
ii) the GMM approximating this distribution function satisfies **Conditions IV**.

VIII. Simulation Results

We present a brief example to illustrate the use of **Theorem 2** using the gas furnace data from [42]. In this gas furnace, methane was combined with air at a feedrate of $U(t)$ in order to obtain a mixture of gases with carbon dioxide, the concentration of which is $Y(t)$. Successive pairs of 296 observations of $(U(t), Y(t))$ were read off from the continuous records at 9 s intervals.

We selected $U(t), U(t)-U(t-1), Y(t-1)$, and $Y(t-1)-Y(t-2)$ as the input variables affecting the present output $Y(t)$. The first 250 vectors of the form $[Y(t-1)-Y(t-2)\ Y(t-1)\ U(t)-U(t-1)\ U(t)\ Y(t)]'$ obtainable from the gas furnace data were used to fit a Gaussian mixture density function with four components. The remaining vectors were reserved to validate the resulting fuzzy model. Training of the five-variable GMM using the expectation maximization (EM) algorithm yielded the following values for the model parameters, shown in Table I.

Invoking **Theorem 2**, we can construct a MISO fuzzy system with four rules. We denote the IF-part variables by $\mathbf{x} = [x_1\ x_2\ x_3\ x_4]' = [Y(t-1)-Y(t-2)\quad Y(t-1)\quad U(t)-U(t-1)\quad U(t)]'$ and the THEN-part variable by $y = Y(t)$. For the first rule, $\mathbf{R}^1$

i) the weight of the rule is 0.0720;
ii) the membership functions for each of the IF-part variables are $N^1(x_1; 0.3817, 0.7862)$, $N^1(x_2; 58.9033, 0.9134)$, $N^1(x_3; 0.1350, 0.0275)$, and $N^1(x_4; -0.9967, 0.2889)$, respectively;

iii) with the invariance covariance matrix shown at the bottom of the page, the regression function of the THEN-part is

$$f^1(\mathbf{x}) = 58.9316 - \frac{\left[\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} - \begin{bmatrix} 0.3817 \\ 58.9033 \\ 0.1350 \\ -0.9967 \end{bmatrix}\right]' \begin{bmatrix} -88.6879 \\ -77.5238 \\ -64.4961 \\ -45.2893 \end{bmatrix}}{115.1374}$$
$$= 19.2935 + 0.7703x_1 + 0.6733x_2 + 0.5602x_3 + 0.3934x_4.$$

Carrying out similar procedures for $\mathbf{R}^2, \mathbf{R}^3$, and $\mathbf{R}^4$, we obtain the following GFM rule base:

$$\begin{aligned}
\mathbf{R}^1 :\ & \text{IF } x_1 \text{ is } \mathbf{A_1^1}(N^1(x_1; 0.3817, 0.7862)) \\
& \wedge x_2 \text{ is } \mathbf{A_2^1}(N^1(x_2; 58.9033, 0.9134)) \\
& \wedge x_3 \text{ is } \mathbf{A_3^1}(N^1(x_3; 0.1350, 0.0275)) \\
& \wedge x_4 \text{ is } \mathbf{A_4^1}(N^1(x_4; -0.9967, 0.2889)), \\
& \text{THEN } \mathbf{B^1}(f^1(\mathbf{x}) = 19.2935 + 0.7703x_1 \\
& +0.6733x_2 + 0.5602x_3 + 0.3934x_4, 0.0720) \\
\mathbf{R}^2 :\ & \text{IF } x_1 \text{ is } \mathbf{A_1^2}(N^1(x_1; 0.0472, 0.5393)) \\
& \wedge x_2 \text{ is } \mathbf{A_2^2}(N^1(x_2; 53.6562, 7.1428)) \\
& \wedge x_3 \text{ is } \mathbf{A_3^2}(N^1(x_3; 0.2029, 0.0507)) \\
& \wedge x_4 \text{ is } \mathbf{A_4^2}(N^1(x_4; 0.4700, 0.8796)), \\
& \text{THEN } \mathbf{B^2}(f^2(\mathbf{x}) = 5.5758 + 0.5958x_1 \\
& +0.8943x_2 + 0.7466x_3 - 0.2909x_4, 0.4696) \\
\mathbf{R}^3 :\ & \text{IF } x_1 \text{ is } \mathbf{A_1^3}(N^1(x_1; -0.1660, 0.3054)) \\
& \wedge x_2 \text{ is } \mathbf{A_2^3}(N^1(x_2; 50.1888, 4.5797)) \\
& \wedge x_3 \text{ is } \mathbf{A_3^3}(N^1(x_3; -0.2772, 0.0951)) \\
& \wedge x_4 \text{ is } \mathbf{A_4^3}(N^1(x_4; -0.0755, 1.0478)), \\
& \text{THEN } \mathbf{B^3}(f^3(\mathbf{x}) = 5.0808 + 0.6093x_1 \\
& +0.9061x_2 + 0.7672x_3 - 0.2990x_4, 0.2695) \\
\mathbf{R}^4 :\ & \text{IF } x_1 \text{ is } \mathbf{A_1^4}(N^1(x_1; 0.0022, 0.6364)) \\
& \wedge x_2 \text{ is } \mathbf{A_2^4}(N^1(x_2; 55.6396, 1.9761)) \\
& \wedge x_3 \text{ is } \mathbf{A_3^4}(N^1(x_3; -0.1218, 0.0430)) \\
& \wedge x_4 \text{ is } \mathbf{A_4^4}(N^1(x_4; -1.0615, 0.7321)), \\
& \text{THEN } \mathbf{B^4}(f^4(\mathbf{x}) = 9.3354 + 0.6210x_1 \\
& +0.8304x_2 + 1.5080x_3 - 0.3007x_4, 0.1889).
\end{aligned}$$

To quantify the performance of the resulting GFM, we use the performance measure found in [43], which is a form of the mean squared error function. The performance measure E is expressed by the formula

$$E = \frac{\sum_{n=1}^{N} e^2(n)}{N \cdot y_r} \tag{46}$$

where

N = total number of data vectors;
$e(n)$ $= y_{\text{target}}(n) - y_{\text{model}}(n)$;
y_r $= [\max(y_{\text{target}}) - \min(y_{\text{target}})]^2$.

For the training data set (the first 250 data vectors), $E = 1.8608 \times 10^{-4}$, which is a very small value and for all practical purposes, zero. The GFM has indeed been accurately fitted to the training set. Fig. 1 plots the target output (solid line) and the model output (dotted line). Due to very small values of model error at each point, the dotted line has been almost completely hidden by the solid line.

When the inputs of the validation data set (i.e., the remaining data *not* used in training) were fed to the resulting GFM, we obtained $E = 0.0032$. The performance measure is inferior compared to that of the training data set, but this is expected when the GFM is made to predict the output of validation data. Fig. 2 plots the target output and predicted output. As can be observed, even with validation data, the GFM has captured the characteristics and nuances of the "gas furnace" input–output relationship. The shape of the predicted output plot follows the target output plot reasonably well.

It is, of course, possible to increase the prediction accuracy of the GFM. This may be achieved in a variety of ways, for example, by increasing the number of training iterations of the EM algorithm, or by increasing the number of rules in the GFM rule base, or by enlarging the set of training vectors.

IX. CONCLUSION

We have derived the conditions under which there is mathematical equivalence between the expected output of a probabilistic system modeled by Gaussian Mixtures, and the defuzzified output of a GFM. **Theorem 1** explicitly describes the GMM to GFM translation of an SISO system. **Theorem 2** deals with the more general MISO case. For the ML and TS fuzzy models, they are related to special cases of the GMM, as stated by **Corollaries 1** and **2**.

A natural overflow of our findings is a new way of training a fuzzy model, as was illustrated with the simulation example. Instead of estimating the parameters of the fuzzy rules directly, we can estimate the parameters of a GMM *first* (using any popular density estimation algorithm, such as EM), and then invoke the relevant theorem to translate the GMM into its corresponding GFM.

By interpreting a fuzzy system from a probabilistic viewpoint, a myriad of statistical tools becomes available at our disposal. For example, by using the fact that the number of mixture components is equal to the number of fuzzy rules (see **Conditions II**), the problem of finding the number of rules translates to finding the number of mixture components.

$$\mathbf{\Sigma_1^{-1}} = \begin{bmatrix} 74.5855 & 58.5393 & 32.5697 & 42.1430 & -88.6879 \\ 58.5393 & 54.4624 & 44.4077 & 26.5652 & -77.5238 \\ 32.5697 & 44.4077 & 124.3423 & 11.5597 & -64.4961 \\ 42.1430 & 26.5652 & 11.5597 & 33.1631 & -45.2893 \\ -88.6879 & -77.5238 & -64.4961 & -45.2893 & 115.1374 \end{bmatrix}$$

TABLE I
ESTIMATED PARAMETERS OF A 4-COMPONENT GMM USING THE EM ALGORITHM

Component 1:	$p_1 = 0.0720$	$\boldsymbol{\mu}_1 = \begin{bmatrix} 0.3817 \\ 58.9033 \\ 0.1350 \\ -0.9967 \\ 58.9316 \end{bmatrix}$	$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 0.7862 & -0.2183 & 0.1024 & -0.3355 & 0.3840 \\ -0.2183 & 0.9134 & 0.0003 & 0.3373 & 0.5797 \\ 0.1024 & 0.0003 & 0.0275 & -0.0236 & 0.0852 \\ -0.3355 & 0.3373 & -0.0236 & 0.2889 & 0.0691 \\ 0.3840 & 0.5797 & 0.0852 & 0.0691 & 0.7697 \end{bmatrix}$
Component 2:	$p_2 = 0.4696$	$\boldsymbol{\mu}_2 = \begin{bmatrix} 0.0472 \\ 53.6562 \\ 0.2029 \\ 0.4700 \\ 53.6050 \end{bmatrix}$	$\boldsymbol{\Sigma}_2 = \begin{bmatrix} 0.5393 & -0.0660 & 0.0668 & -0.3378 & 0.4104 \\ -0.0660 & 7.1428 & 0.2970 & -1.5582 & 7.0237 \\ 0.0668 & 0.2970 & 0.0507 & -0.0981 & 0.3718 \\ -0.3378 & -1.5582 & -0.0981 & 0.8796 & -1.9239 \\ 0.4104 & 7.0237 & 0.3718 & -1.9239 & 7.3964 \end{bmatrix}$
Component 3:	$p_3 = 0.2695$	$\boldsymbol{\mu}_3 = \begin{bmatrix} -0.1660 \\ 50.1888 \\ -0.2772 \\ -0.0755 \\ 50.2667 \end{bmatrix}$	$\boldsymbol{\Sigma}_3 = \begin{bmatrix} 0.3054 & 0.0737 & 0.0843 & -0.1979 & 0.3767 \\ 0.0737 & 4.5797 & -0.2276 & -1.2808 & 4.4030 \\ 0.0843 & -0.2276 & 0.0951 & 0.1013 & -0.1122 \\ -0.1979 & -1.2808 & 0.1013 & 1.0478 & -1.5167 \\ 0.3767 & 4.4030 & -0.1122 & -1.5167 & 4.6055 \end{bmatrix}$
Component 4:	$p_4 = 0.1889$	$\boldsymbol{\mu}_4 = \begin{bmatrix} 0.0022 \\ 55.6396 \\ -0.1218 \\ -1.0615 \\ 55.6774 \end{bmatrix}$	$\boldsymbol{\Sigma}_4 = \begin{bmatrix} 0.6364 & 0.0618 & 0.1112 & -0.0435 & 0.6273 \\ 0.0618 & 1.9761 & 0.0192 & 0.5852 & 1.5324 \\ 0.1112 & 0.0192 & 0.0430 & -0.0095 & 0.1527 \\ -0.0435 & 0.5852 & -0.0095 & 0.7321 & 0.2245 \\ 0.6273 & 1.5324 & 0.1527 & 0.2245 & 1.8832 \end{bmatrix}$

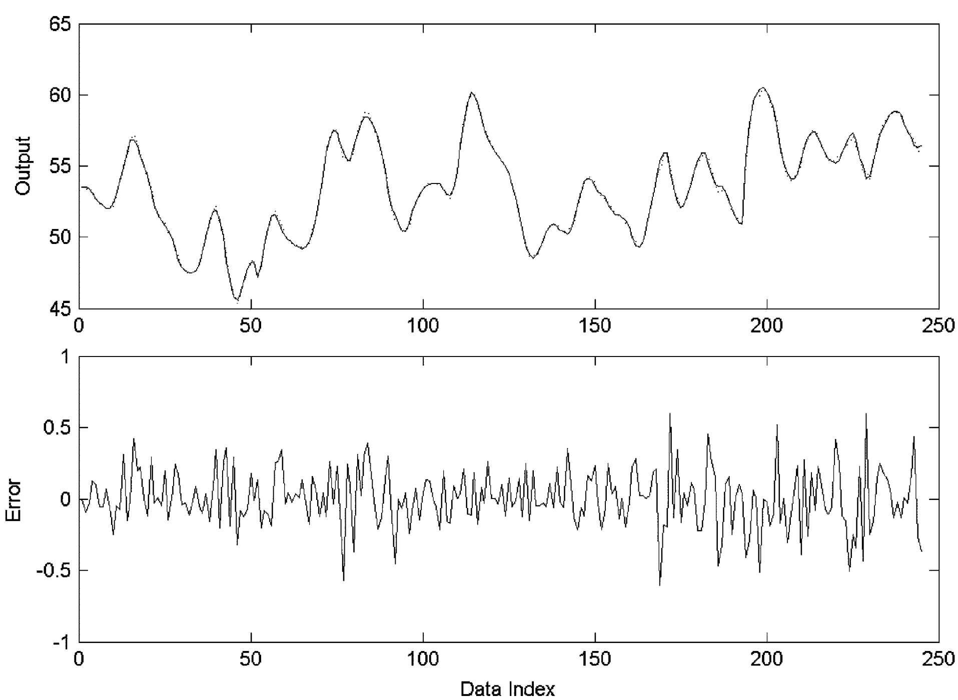

Fig. 1. Plot of target output (solid line) and model output (dotted line) of the training data set, and the corresponding model error.

Our work here primarily focused on systems modeled by Gaussian mixture probability distributions. However, it is a simple matter to show that mixture of distributions other than Gaussian (for example, triangular, trapezoidal, etc.) will result in other shapes of IF-part membership functions. Moreover, the *universal approximation property* for the superposition of Gaussians guarantees that any distribution function can be approximated as Gaussian mixtures.

Finally, we have also demonstrated the reconstruction of the GMM from the GFM, and showed that the GMM is not

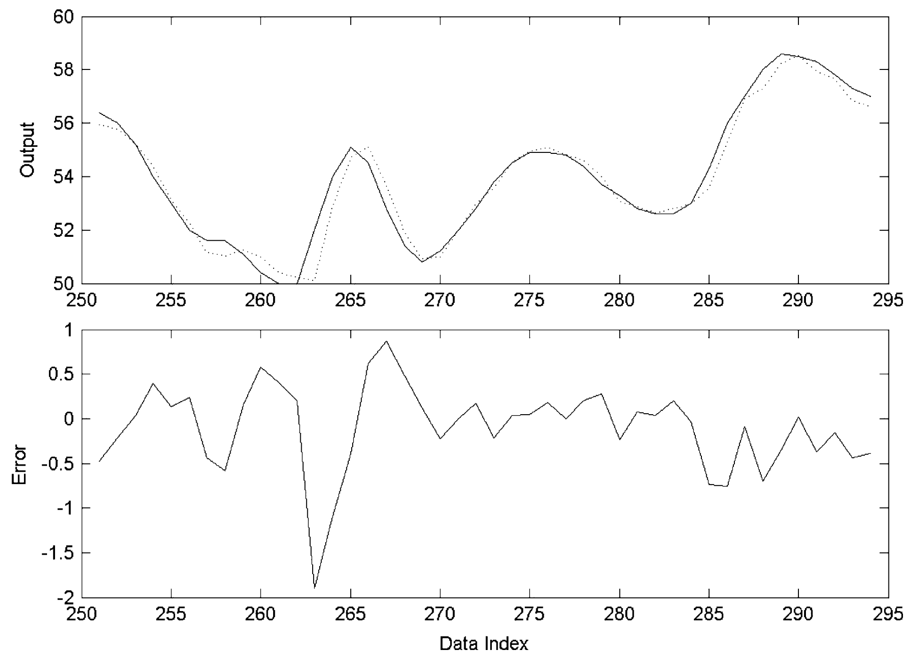

Fig. 2. Plot of target output (solid line) and model output (dotted line) of the validation data set, and the corresponding model error.

unique. Several distinct sets of GMM parameters can be derived from a single GFM. This means more than one GMM can translate into the same GFM, i.e., GFM to GMM translation is one-to-many.

Before we close, we reiterate a caution given elsewhere in this paper. Just because a probabilistic system *can* be interpreted as a fuzzy system does not mean it *should* be. In general, a fuzzy system is *not* a probabilistic system. It will serve us well not to repeat the same mistaken conclusions of past debates.

Appendix A

Expansion of one integral term in (24) yields

$$\begin{aligned}
&\int_{-\infty}^{+\infty} yG_c(x,y)\,dy \\
&= \int_{-\infty}^{+\infty} y\frac{p_c}{2\pi\sqrt{|\mathbf{\Sigma}_c|}}\exp\left\{-\frac{1}{2}[x-\mu_{cx} \quad y-\mu_{cy}]\right. \\
&\quad \times \left.\begin{bmatrix}\sigma_{cxx} & \sigma_{cxy}\\ \sigma_{cyx} & \sigma_{cyy}\end{bmatrix}^{-1}\begin{bmatrix}x-\mu_{cx}\\ y-\mu_{cy}\end{bmatrix}\right\}dy \\
&= \frac{p_c}{2\pi\sqrt{|\mathbf{\Sigma}_c|}}\int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2|\mathbf{\Sigma}_c|}[x-\mu_{cx} \quad y-\mu_{cy}]\right. \\
&\quad \times \left.\begin{bmatrix}\sigma_{cyy} & -\sigma_{cyx}\\ -\sigma_{cxy} & \sigma_{cxx}\end{bmatrix}\begin{bmatrix}x-\mu_{cx}\\ y-\mu_{cy}\end{bmatrix}\right\}dy \\
&= \frac{p_c}{2\pi\sqrt{|\mathbf{\Sigma}_c|}}\int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2|\mathbf{\Sigma}_c|}((x-\mu_{cx})^2\sigma_{cyy}\right. \\
&\quad \left. -\,2(x-\mu_{cx})(y-\mu_{cy})\sigma_{cxy}+(y-\mu_{cy})^2\sigma_{cxx})\right\}dy \\
&= \frac{p_c}{2\pi\sqrt{|\mathbf{\Sigma}_c|}}\exp\left\{-\frac{(x-\mu_{cx})^2\sigma_{cyy}}{2|\mathbf{\Sigma}_c|}\right\} \\
&\quad \times \int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2|\mathbf{\Sigma}_c|}((y-\mu_{cy})^2\sigma_{cxx}\right. \\
&\qquad \left. -2(x-\mu_{cx})(y-\mu_{cy})\sigma_{cxy})\right\}dy.
\end{aligned}$$

Completing the square for the integral on the right-hand side, we obtain

$$\begin{aligned}
&\int_{-\infty}^{+\infty} yG_c(x,y)\,dy \\
&= \frac{p_c}{2\pi\sqrt{|\mathbf{\Sigma}_c|}}\exp\left\{-\frac{1}{2|\mathbf{\Sigma}_c|}(x-\mu_{cx})^2\sigma_{cyy}\right\} \\
&\quad \times \exp\left\{\frac{1}{2|\mathbf{\Sigma}_c|}\left((x-\mu_{cx})\frac{\sigma_{cxy}}{\sqrt{\sigma_{cxx}}}\right)^2\right\} \\
&\quad \times \int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2|\mathbf{\Sigma}_c|}\left((y-\mu_{cy})\sqrt{\sigma_{cxx}}\right.\right. \\
&\quad \left.\left. -\,(x-\mu_{cx})\frac{\sigma_{cxy}}{\sqrt{\sigma_{cxx}}}\right)^2\right\}dy \\
&= \frac{p_c}{2\pi\sqrt{|\mathbf{\Sigma}_c|}}\exp\left\{-\frac{1}{2|\mathbf{\Sigma}_c|}(x-\mu_{cx})^2\left(\sigma_{cyy}-\frac{\sigma_{cxy}^2}{\sigma_{cxx}}\right)\right\} \\
&\quad \times \int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2\frac{|\mathbf{\Sigma}_c|}{\sigma_{cxx}}}\right. \\
&\quad \times \left.\left(y-\left(\mu_{cy}+(x-\mu_{cx})\frac{\sigma_{cxy}}{\sigma_{cxx}}\right)\right)^2\right\}dy.
\end{aligned} \tag{A.1}$$

Defining a Gaussian distribution $N^1(y;\mu_c,\sigma_c^2) = N^1(y;\mu_{cy} + (x-\mu_{cx})(\sigma_{cxy})/(\sigma_{cxx}), (|\mathbf{\Sigma}_c|)/(\sigma_{cxx}))$, and substituting into (A.1), we have

$$\begin{aligned}
&\int_{-\infty}^{+\infty} yG_c(x,y)\,dy \\
&= p_c \frac{\sqrt{2\pi \frac{|\mathbf{\Sigma}_c|}{\sigma_{cxx}}}}{2\pi\sqrt{|\mathbf{\Sigma}_c|}} \exp\left\{ -\frac{(x-\mu_{cx})^2}{2\frac{|\mathbf{\Sigma}_c|}{\left(\sigma_{cyy}-\frac{\sigma_{cxy}^2}{\sigma_{cxx}}\right)}} \right\} \\
&\quad\times \int_{-\infty}^{+\infty} yN^1\left(y;\mu_{cy}+(x-\mu_{cx})\frac{\sigma_{cxy}}{\sigma_{cxx}}, \frac{|\mathbf{\Sigma}_c|}{\sigma_{cxx}}\right)dy \\
&= \frac{p_c}{\sqrt{2\pi\sigma_{cxx}}} \exp\left\{-\frac{(x-\mu_{cx})^2}{2\sigma_{cxx}}\right\}\left(\mu_{cy}+(x-\mu_{cx})\frac{\sigma_{cxy}}{\sigma_{cxx}}\right) \\
&= p_c N^1(x;\mu_{cx},\sigma_{cxx})\left(\mu_{cy}+(x-\mu_{cx})\frac{\sigma_{cxy}}{\sigma_{cxx}}\right). \qquad \text{(A.2)}
\end{aligned}$$

APPENDIX B

We represent the *elements* of the covariance matrix as $\mathbf{\Sigma} = [\sigma_{ab}]$ and the elements of its inverse as $\mathbf{\Sigma}^{-1} = [\sigma^{ab}]$. Accordingly, we have

$$\mathbf{\Sigma} = \begin{bmatrix} \boldsymbol{\sigma}_{\mathbf{xx}} & \boldsymbol{\sigma}_{\mathbf{x}y} \\ \boldsymbol{\sigma}_{y\mathbf{x}} & \sigma_{yy} \end{bmatrix} = \left[\begin{array}{c|c} \{\sigma_{cij}\}_{J\times J} & \{\sigma_{cj(J+1)}\}_{J\times 1} \\ \hline \{\sigma_{c(J+1)j}\}_{1\times J} & \sigma_{c(J+1)(J+1)} \end{array}\right]. \qquad \text{(B.1)}$$

Note that since $\mathbf{\Sigma}$ is a symmetric matrix, hence $\boldsymbol{\sigma}_{\mathbf{x}y} = \boldsymbol{\sigma}'_{y\mathbf{x}}$. Therefore, the inverse is

$$\mathbf{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\sigma}^{\mathbf{xx}} & \boldsymbol{\sigma}^{\mathbf{x}y} \\ \boldsymbol{\sigma}^{y\mathbf{x}} & \sigma^{yy} \end{bmatrix}. \qquad \text{(B.2)}$$

Hence, $\int_{-\infty}^{+\infty} yG_c(\mathbf{x},y)\,dy$, or rather

$$\int_{-\infty}^{+\infty} yp_cN^{J+1}\left(\begin{bmatrix}\mathbf{x}\\ y\end{bmatrix};\begin{bmatrix}\boldsymbol{\mu}_{c\mathbf{x}}\\ \mu_{cy}\end{bmatrix},\mathbf{\Sigma}_c\right)dy$$

becomes

$$\begin{aligned}
&\frac{p_c}{(2\pi)^{\frac{J+1}{2}}\sqrt{|\mathbf{\Sigma}_c|}}\int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2}\begin{bmatrix}\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}\\ y-\mu_{cy}\end{bmatrix}' \right. \\
&\quad\left.\times\begin{bmatrix}\boldsymbol{\sigma}_{c\mathbf{xx}} & \boldsymbol{\sigma}_{c\mathbf{x}y}\\ \boldsymbol{\sigma}_{cy\mathbf{x}} & \sigma_{cyy}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}\\ y-\mu_{cy}\end{bmatrix}\right\}dy \\
&= \frac{\cdots}{\cdots}\int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2}\left([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{xx}}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]\right.\right. \\
&\quad + (y-\mu_{cy})\boldsymbol{\sigma}_c^{y\mathbf{x}}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}] + [\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}(y-\mu_{cy}) \\
&\quad \left.\left. +\ (y-\mu_{cy})\sigma_c^{yy}(y-\mu_{cy})\right)\right\}dy. \qquad \text{(B.3a)}
\end{aligned}$$

Note that like $\mathbf{\Sigma}, \mathbf{\Sigma}^{-1}$ is also a symmetric matrix. (By a theorem in [39, p. 370], the inverse of a symmetric matrix is another symmetric matrix). Hence, $\boldsymbol{\sigma}_c^{\mathbf{x}y} = (\boldsymbol{\sigma}_c^{y\mathbf{x}})'$ and $[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}(y-\mu_{cy}) = (y-\mu_{cy})\boldsymbol{\sigma}_c^{y\mathbf{x}}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]$. Then, (B.3a) becomes

$$\begin{aligned}
&= \frac{\cdots}{\cdots}\int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2}\left([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{xx}}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]\right.\right. \\
&\quad + 2(y-\mu_{cy})[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y} \\
&\quad \left.\left. +\ (y-\mu_{cy})^2\sigma_c^{yy}\right)\right\}dy \\
&= \frac{\cdots}{\cdots}\exp\left\{-\frac{1}{2}\left([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{xx}}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]\right)\right\} \\
&\quad\times\int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2}\left((y-\mu_{cy})^2\sigma_c^{yy}+2(y-\mu_{cy})\right.\right. \\
&\quad\left.\left.\times\ [\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}\right)\right\}dy. \qquad \text{(B.3b)}
\end{aligned}$$

Completing the square for the integral of (B.3b), we obtain

$$\begin{aligned}
&= \frac{\cdots}{\cdots}\exp\{\cdots\}\exp\left\{\frac{1}{2}\left(\frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sqrt{\sigma_c^{yy}}}\right)^2\right\} \\
&\quad\times\int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2}\left((y-\mu_{cy})\sqrt{\sigma_c^{yy}}+\frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sqrt{\sigma_c^{yy}}}\right)^2\right\}dy \\
&= \frac{\cdots}{\cdots}\exp\{\cdots\}\exp\{\cdots\}\int_{-\infty}^{+\infty} y\exp\left\{-\frac{1}{2\left(\frac{1}{\sigma_c^{yy}}\right)}\right. \\
&\quad\left.\times\left(y-\left(\mu_{cy}-\frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sigma_c^{yy}}\right)\right)^2\right\}dy. \qquad \text{(B.3c)}
\end{aligned}$$

We define a Gaussian distribution

$$N^1(y;\mu_c,\sigma_c^2) = N^1\left(y;\mu_{cy}-\frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sigma_c^{yy}},\frac{1}{\sigma_c^{yy}}\right)$$

and substitute into (B.3c) to obtain

$$\begin{aligned}
&= \frac{p_c\sqrt{2\pi\left(\frac{1}{\sigma_c^{yy}}\right)}}{(2\pi)^{\frac{J+1}{2}}\sqrt{|\mathbf{\Sigma}_c|}}\exp\left\{-\frac{1}{2}\left([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{xx}}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]\right.\right. \\
&\quad\left.\left. -\ \frac{([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y})^2}{\sigma_c^{yy}}\right)\right\} \\
&\quad\times\int_{-\infty}^{+\infty} yN^1\left(y;\mu_{cy}-\frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sigma_c^{yy}},\frac{1}{\sigma_c^{yy}}\right)dy \\
&= \frac{p_c}{(2\pi)^{\frac{J}{2}}\sqrt{\sigma_c^{yy}|\mathbf{\Sigma}_c|}}\exp\left\{-\frac{1}{2}\left([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{xx}}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]\right.\right. \\
&\quad\left.\left. -\ \frac{([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y})^2}{\sigma_c^{yy}}\right)\right\}\times\left(\mu_{cy}-\frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sigma_c^{yy}}\right) \\
&= \frac{\cdots}{\cdots}\exp\left\{-\frac{1}{2}\left([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{xx}}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]\right.\right. \\
&\quad\left.\left. -\ \frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}\boldsymbol{\sigma}_c^{y\mathbf{x}}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]}{\sigma_c^{yy}}\right)\right\}(\cdots) \\
&= \frac{\cdots}{\cdots}\exp\left\{-\frac{1}{2}\left([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\left[\boldsymbol{\sigma}_c^{\mathbf{xx}}-\boldsymbol{\sigma}_c^{\mathbf{x}y}\left(\sigma_c^{yy}\right)^{-1}\boldsymbol{\sigma}_c^{y\mathbf{x}}\right]\right.\right. \\
&\quad\left.\left.\times\ [\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]\right)\right\}(\cdots). \qquad \text{(B.3d)}
\end{aligned}$$

By theorem [44, Eq. (A.2.4g)]

$$\mathbf{A}^{11} = \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1}. \tag{B.4a}$$

We can also write

$$\begin{aligned}\mathbf{A}_{11} &= (\mathbf{A}^{11} - \mathbf{A}^{12}(\mathbf{A}^{22})^{-1}\mathbf{A}^{21})^{-1} \quad \text{or} \\ (\mathbf{A}_{11})^{-1} &= \mathbf{A}^{11} - \mathbf{A}^{12}(\mathbf{A}^{22})^{-1}\mathbf{A}^{21}.\end{aligned} \tag{B.4b}$$

Translating this into our notation

$$(\boldsymbol{\sigma}_{c\mathbf{xx}})^{-1} = \boldsymbol{\sigma}_c^{\mathbf{xx}} - \boldsymbol{\sigma}_c^{\mathbf{x}y}\left(\sigma_c^{yy}\right)^{-1}\boldsymbol{\sigma}_c^{y\mathbf{x}}. \tag{B.4c}$$

Substituting (B.4c) into (B.3d), the latter becomes

$$\begin{aligned}&= \frac{p_c}{(2\pi)^{\frac{J}{2}}\sqrt{\sigma_c^{yy}|\boldsymbol{\Sigma}_c|}} \exp\left\{-\frac{1}{2}\left([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_{c\mathbf{xx}}^{-1}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]\right)\right\} \\ &\qquad \times \left(\mu_{cy} - \frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sigma_c^{yy}}\right).\end{aligned} \tag{B.5}$$

$\sigma_c^{yy}|\boldsymbol{\Sigma}_c|$ is the $(J+1)(J+1)$th *cofactor* of $\boldsymbol{\Sigma}_c$, which is also the determinant of $\boldsymbol{\sigma}_{c\mathbf{xx}}$, $|\boldsymbol{\sigma}_{c\mathbf{xx}}|$. Hence, (B.5) can be finally simplified to

$$\begin{aligned}&= \frac{p_c}{(2\pi)^{\frac{J}{2}}\sqrt{|\boldsymbol{\sigma}_{c\mathbf{xx}}|}} \exp\left\{-\frac{1}{2}\left([\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_{c\mathbf{xx}}^{-1}[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]\right)\right\} \\ &\quad \times \left(\mu_{cy} - \frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sigma_c^{yy}}\right) \\ &= G_c(\mathbf{x})\left(\mu_{cy} - \frac{[\mathbf{x}-\boldsymbol{\mu}_{c\mathbf{x}}]'\boldsymbol{\sigma}_c^{\mathbf{x}y}}{\sigma_c^{yy}}\right).\end{aligned} \tag{B.6}$$

REFERENCES

[1] I. Hacking, *The Emergence of Probability*. Cambridge, U. K.: Cambridge Univ. Press, 1975.

[2] T. L. Fine, *Theories of Probability: An Examination of Foundations*. New York, NY: Academic, 1973.

[3] B. Kosko, "The probability monopoly," *IEEE Trans. Fuzzy Syst.*, vol. 2, no. 1, pp. 32–33, Feb. 1994.

[4] G. J. Klir and T. A. Folger, *Fuzzy Sets, Uncertainty and Information*. Upper Saddle River, NJ: Prentice-Hall, 1988.

[5] H. J. Zimmerman, *Fuzzy Set Theory and Its Applications*, 2nd ed. Boston, MA: Kluwer, 1991.

[6] D. Dubois and H. Prade, *Possibility Theory*. New York: Plenum, 1988.

[7] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, pp. 338–353, 1965.

[8] D. Dubois, H. T. Nguyen, and H. Prade, "Possibility theory, probability, and fuzzy sets," in *Fundamentals of Fuzzy Sets*. ser. The Handbooks of Fuzzy Sets Series, D. Dubois and H. Prade, Eds. Norwell, MA: Kluwer, 2000, pp. 343–438.

[9] W. Stallings, "Fuzzy set theory versus Bayesian statistics," *IEEE Trans. Syst., Man, Cybern.*, vol. 7, pp. 216–219, 1977.

[10] P. Cheeseman, "Probability versus fuzzy reasoning," in *Uncertainty in Artificial Intelligence 1*, L. Kanal and J. Lemmer, Eds. Amsterdam, The Netherlands: North-Holland, 1986, pp. 85–102.

[11] M. Laviolette, J. W. Seaman Jr., J. D. Barrett, and W. H. Woodhall, "Fuzziness vs. probability," *Technometrics*, vol. 37, pp. 249–259, 1995.

[12] D. Dubois and H. Prade, "Fuzzy sets—A convenient fiction for modeling vagueness and possibility," *IEEE Trans. Fuzzy Syst.*, vol. 2, no. 1, pp. 16–21, Feb. 1994.

[13] G. J. Klir, "On the alleged superiority of probabilistic representation of uncertainty," *IEEE Trans. Fuzzy Syst.*, vol. 2, no. 1, pp. 27–31, Feb. 1994.

[14] L. A. Zadeh, "Probability theory and fuzzy logic are complementary rather than competitive," *Technometrics*, vol. 37, pp. 271–276, 1995.

[15] D. Dubois and H. Prade, "Fuzzy sets and probability: Misunderstandings, bridges, and gaps," in *Proc. 2nd IEEE Int. Conf. Fuzzy Systems (FUZZ-IEEE'93)*, vol. 2, San Francisco, CA, 1993, pp. 1059–1068.

[16] S. Mabuchi, "An interpretation of membership functions and the properties of general probabilistic operators as fuzzy set operators—Part I: Case of type 1 fuzzy sets," *Fuzzy Sets Syst.*, vol. 49, pp. 271–283, 1992.

[17] ——, "An interpretation of membership functions and the properties of general probabilistic operators as fuzzy set operators—Part II: Extensions to three-valued and interval-valued fuzzy sets," *Fuzzy Sets Syst.*, vol. 92, pp. 31–50, 1997.

[18] B. Kosko, "Fuzziness vs. probability," *Int. J. Gen. Syst.*, vol. 17, pp. 211–240, 1990.

[19] R. Viertl, "Is it necessary to develop a fuzzy Bayesian inference?," in *Probability and Bayesian Statistics*, R. Viertl, Ed. New York: Plenum, 1987, pp. 471–475.

[20] R. Duda, J. Gaschnig, and P. Hart, "Model design in the prospector consultant system for mineral exploration," in *Expert Systems in the Microelectronic Age*, D. Michie, Ed. Edinburgh, U. K.: Edinburgh Univ. Press, 1979, pp. 153–167.

[21] E. J. Horvitz, D. E. Heckerman, and C. P. Langlotz, "A framework for comparing alternative formalisms for plausible reasoning," in *Proc. 5th National Conf. Artificial Intelligence (AAAI'86)*, vol. 1, Philadelphia, PA, 1986, pp. 210–214.

[22] M. Laviolette and J. W. Seaman Jr., "The efficacy of fuzzy representations of uncertainty," *IEEE. Trans. Fuzzy Syst.*, vol. 2, no. 1, pp. 4–15, Feb. 1994.

[23] P. Cheeseman, "In defence of probability," in *Proc. 9th Int. Joint Conf. Artificial Intelligence*, Los Angeles, CA, 1985, pp. 1002–1009.

[24] D. V. Lindley, "The probability approach to the treatment of uncertainty in artificial intelligence and expert systems," *Stat. Sci.*, vol. 2, no. 1, pp. 17–24, 1987.

[25] G. J. Klir, "Is there more to uncertainty than some probability theorists might have us believe?," *Int. J. Gen. Syst.*, vol. 15, no. 4, pp. 347–378, 1989.

[26] R. Kruse and K. D. Meyer, *Statistics with Vague Data*. Dordrecht, The Netherlands: D. Reidel, 1987.

[27] T. Okuda, H. Tanaka, and K. Asai, "A formulation of fuzzy decision problems with fuzzy information, using probability measures of fuzzy events," *Inf. Control*, vol. 38, pp. 135–147, 1978.

[28] T. S. Wallsten, D. V. Budescu, A. Rapoport, B. Zwick, and B. Forsyth, "Measuring the vague meanings of probability terms," *J. Experiment. Psychol.: General*, vol. 115, no. 4, pp. 348–365, 1986.

[29] P. Z. Wang, "From the fuzzy statistics to the falling random subsets," in *Advances in Fuzzy Sets, Possibility Theory and Applications*, P. P. Wang, Ed. New York: Plenum, 1983, pp. 81–96.

[30] H. T. Nguyen, "On modeling of linguistic information using random sets," *Inf. Sci.*, vol. 34, pp. 265–274, 1984.

[31] H. M. Hersh and A. Caramazza, "A fuzzy set approach to modifiers and vagueness in natural language," *J. Experiment. Psychol.: General*, vol. 105, pp. 254–276, 1976.

[32] E. Hisdal, "Naturalized logic and chain sets," *Inf. Sci. (archive)*, vol. 57–58, pp. 31–77, 1991.

[33] S. F. Thomas, "Probabilistic uncertainty and statistical inference," presented at the ORSA/TIMS Meeting, Houston, TX, 1981.

[34] B. S. Everitt and D. J. Hand, *Finite Mixture Distributions*. London, U. K.: Chapman and Hall, 1981.

[35] M. F. Azeem, M. Hanmandlu, and N. Ahmad, "Generalization of adaptive neuro-fuzzy inference systems," *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1332–1346, Nov. 2000.

[36] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic systems," *IEEE Trans. Comput.*, vol. C-26, no. 12, pp. 1182–1191, Dec. 1977.

[37] P. M. Larsen, "Industrial applications of fuzzy logic control," *Int. J. Man-Mach. Stud.*, vol. 12, pp. 3–10, 1980.

[38] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 1, pp. 116–132, Jan. 1985.

[39] E. Kreyzig, *Advanced Engineering Mathematics*, 7th ed. Singapore: Wiley, 1993.

[40] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Comput.*, vol. 3, no. 2, pp. 246–257, 1991.

[41] P. Jorge and S. G. Ferreira, "Neural networks and approximation by superposition of Gaussians," in *Proc. ICASSP'97*, vol. 4, 1997, pp. 3197–3200.

[42] E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*. San Francisco, CA: Holden Day, 1970.

[43] M. F. Azeem, M. Hanmandlu, and N. Ahmad, "Structure identification of generalized adaptive neuro-fuzzy inference systems," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 5, pp. 666–681, Oct. 2003.

[44] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. New York: Academic, 1979.

Ming-Tao Gan (M'03) received the B.Eng. degree in electronics engineering from Multimedia University (MMU), Cyberjaya, Malaysia.

Since graduating with first class honors in 2001, he has been on staff at MMU as a Teaching Assistant with the Faculty of Engineering. He is also concurrently a Postgraduate Researcher at the faculty's Center for Multimedia Communications. At the time of this writing, his research interests are inclined toward developing strategies for the identification and modeling of optimal fuzzy systems.

Madasu Hanmandlu (M'04) received the B.E. degree in electrical engineering from Osmania University, Hyderabad, India, in 1973, the M.Tech. degree in power systems from R.E.C. Warangal, Jawaharlal Nehru Technological University, India, in 1976, and the Ph.D. degree in control systems from Indian Institute of Technology, Delhi, India, in 1981.

From 1979 to 1981, he was a Senior Scientific Officer in Applied Systems Research program of the Department of Electrical Engineering, IIT Delhi, India. He joined the department as a Lecturer in 1981 and became a Professor in 1997. He was with Machine Vision Group, City University, London, U.K., in 1988 for six and half months, and the Robotics Research Group, Oxford University, Oxford, U.K., in 1993 for three months, as part of the Indo-U.K. research collaboration. He was a Visiting Professor with the Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia, from March 2001 to February 2003. He worked in the areas of power systems, control, robotics, and computer vision, before shifting to theory of fuzzy logic. His current research interests mainly include fuzzy modeling of dynamic systems and applications of fuzzy logic to image processing, document processing, bio-medical imaging, and intelligent control. Currently, he is an Associate Editor of *Pattern Recognition*. He is listed in *Reference Asia; Asia's Who's Who of Men and Women of achievement; 5000 Personalities of the World (1998), American Biographical Institute*.

Ai Hui Tan graduated with first class honors in electronics engineering from the University of Warwick, Warwick, U.K., in 1999, and received the Ph.D. degree in 2002 for research done in the Systems Modeling and Simulation Research Group at the same university.

Her research interests are in the fields of system identification, in particular the effects of nonlinearities on process models, and of control. She is currently a Senior Lecturer at Multimedia University, Cyberjaya, Malaysia.