Encoding fuzzy possibilistic diagnostics as a constrained optimisation problem

Antonio Sala

Systems Engineering and Control Dept.
Universidad Politecnica de Valencia, Cno Vera s/n,
E-46071 Valencia, Spain, Fax: 963879579, Email: asala@isa.upv.es

Abstract

This paper discusses a knowledge-base encoding methodology for diagnostic tasks, transforming such knowledge into constrained optimisation problems. The methodology is based on a re-interpretation of the consistent causal reasoning paradigm [15] as an equivalent problem of feasibility subject to equality and inequality constraints (in the binary case). Then, it is extended to the fuzzy case. Preferences under uncertain knowledge are incorporated by transforming the feasibility problem into an optimisation one, which may be interpreted in possibilistic terms. The problem is solved by efficient, widely-known, linear and quadratic programming tools, which are able to cope with large-scale problems. Examples illustrating some of the concepts and possibilities of the proposed procedure, as well as a summary comparison with other approaches are also discussed.

Key words: fault detection and diagnosis, possibilistic reasoning, approximate reasoning, optimisation

1 Introduction

Diagnosis is the process of inferring *disorders* from measurements of some *manifestations* which pinpoint departure from "normal" behaviour of a particular system (industrial, medical, ...). The result of a diagnosis process is an estimated degree of presence of each disorder. Such result may be expressed by a set of either "yes/no" assertions, or by continuously graded ones (denoting disorder "severities" from zero to, say, 100%). Different approaches to the problem appear in the literature: data-based (case-based [20]), knowledge-based, differential-equation approaches or discrete-event formulations [3,23]. The full diagnosis problem is a complex one: it should include temporal analysis [47,21] and a probabilistic setting [5], multi-scale models [30], in a framework such as hybrid recurrent Bayesian networks [19,33], as well as decision-theoretic criteria; the reader is referred to [9,4,22] for further information. Gradual presence of disorder intensities and manifestations inspired the use of fuzzy logic [1,6]. Other knowledge-based frameworks extend the basic logic reasoning schemes to handle uncertainty in the knowledge base via approximate reasoning[25,42], possibilistic and abductive reasoning [28,29,14,49], Bayesian networks [7,33], *etc.* Fuzzy-probabilistic approaches using Dempster-Shafer theory have also been proposed in diagnosis applications [44]. Optimization in the Dempster-Shafer context is proposed in, for instance, [48].

In many practical cases, diagnostic-related knowledge is available from human experts. Such experts express their knowledge in statements such as "*heavy alcohol intake* causes blood pressure *higher than normal*" : the basic building block in many logic-based approaches to fault diagnosis are logic assertions relating *disorders* and *measurable manifestations*. Such knowledge appears in literature either in direct form (*i.e.*, causal relationship from disorders –cause– to manifestations –effect– [21,28]) or in inverse form (diagnostic rules stating from a set of manifestations the most likely disorder, in a form such as "if manifestations X are present then disorder Y is present") [32,6,44,30]. In the author's opinion, the former approach is closer to the actual expert knowledge; the latter gives rise to large rulebases (accounting for many manifestation combinations), and too vague conclusions when missing measurements are present. In this paper, the first of the approaches will be pursued.

This paper is inspired in the works [15,8,28]. However, instead of following approximate reasoning and logic argumentations, the objective of the paper is formulating an interpretation of possibilistic fuzzy diagnosis in terms of constrained optimisation [13]. The idea of transforming logic inference into algebraic constraints and optimisation was also used in previous works by the author [34,35].

In this paper, the encoded diagnostic knowledge will be represented by a set of constraints and a cost index. The main contribution is realising that complex relationships between disorders and the observed manifestations can be stated in a somehow simple way if additional intermediate entities (decision variables in the optimization code) are introduced: *association variables*, *measurements* and *preference-expressing* artificial variables.

The first part of this work presents a basic knowledge encoding methodology. Knowledge will be assumed to consist on a description of "associations" between a set of measurable *manifestations*, m_j, and a particular single *disorder*, d_i. Following [15], the manifestations will be divided in three groups: those known to be certainly caused by d_i, those known to be absent when d_i alone is present, and those with uncertain presence/absence. However, instead of encoding knowledge as lists, as done in [15], knowledge will be equivalently encoded as equality and inequality constraints. Observations will also be cast in the same framework. Then, covering explanations (consistent diagnostics) [15,31] will be understood as the feasible solutions of the above set of constraints.

Extension to the fuzzy case (gradual intensity of the presence of disorders and manifestations) is a complicated issue in the most of the above references. Interestingly, in the presented methodology, it is trivial: only considering the search space for decision variables to be the full interval [0,1] is needed. Also, by conceiving additional constraint typologies, a more detailed description of disorder interaction is possible (such as “disorder 3 may appear only if disorder 5 is present and causing manifestation 2”); such a flexibility is not possible in [15,28].

The second part of this paper discusses the situations where a set of observations may have many alternative explanations. In this paper, abductive preferences under uncertain knowledge [15,29,8,31,5] will be ranked by implicitly describing a possibility distribution by means of a cost index, in the spirit of [13]. Then, obtaining the set of decision variables whith maximum possibility casts abduction as optimisation of a linear or quadratic functional subject to linear constraints, which is a deeply studied problem [43,41] for which general-purpose computer algebra software is used. Convex numerical optimization has superseded symbolic rule-based approaches in other aspects of fuzzy techniques, notably fuzzy control [45,37,36]; this paper advocates that direction in the fuzzy diagnosis field.

The main advantages of the methodology to be proposed are: (*a*) the ability to use well-known optimisation techniques with fast commercial software able to handle thousands of decision variables, (*b*) the incorporation of *both* “fuzziness” (understood as gradation in symptoms and disorders) *and* graded uncertainty (understood as relaxed inequality constraints, interpreted in possibilistic terms) – uncertainty and imprecision in the terminology in [44]–; (*c*) the ability to naturally handle multiple faults and outlier measurements (i.e., a fault in the measurement processes).

The structure of the paper is as follows: the next section discusses basic notation and definitions. Encoding of causal binary diagnostic knowledge as a set of equality and inequality constraints is discussed in Section 3. Some examples and simple enhancements are discussed in Section 4. Allowing the variables to be real-valued immediately suggests to reinterpret the equations in a fuzzy case, as discussed in Section 5. A well-known possibilistic framework is outlined in Section 6. Based on it, preferences under uncertainty in the knowledge bases are discussed in Section 7. Simple examples are given throughout the text, illustrating the main concepts. Section 8 presents additional examples. Section 9 briefly compares the proposed methodology with related ones in the literature.

2 Basic definitions

In this section, the basic definitions and notation to be used in the paper will be presented. The key elements in the diagnostic system will be disorders, manifestations (caused by the disorders) and observations (actual measurements of some of the manifestations).

Disorders. Let us consider a set of q basic single disorders $\{d_1, d_2, \ldots, d_q\}$ which may be present in a diagnosed system (with the possibility of the simultaneous presence of several disorders). For instance, in a medical application, the disorders might be: “hay fever”, “influenza”, “staphylococcus infection”, etc.

Each disorder d_i may be conceived as a structure with two fields: (1) a linguistic label (such as “hay fever”), expressing the disorder meaning for the expert which provides the knowledge, and, (2) a numerical value, to be denoted as “severity” denoting presence (severity=1), absence (severity=0), or partial degree of presence [24] with a severity ranging between zero and one. Abusing the notation, d_i will be used to denote both the disorder severity and the label, depending on the context. For convenience, the disorders will be arranged as a vector $D = (d_1, d_2, \ldots)$ in some cases.

Manifestations. The possible manifestations (symptoms) of the above disorders will be denoted by $\{m_1, \ldots, m_n\}$ where each m_i, has an associated linguistic label (such as “abnormal body temperature”, “hearth rate above 80bpm” *etc.*) and an associated “intensity of presence”, denoting its absence (zero intensity), full presence (intensity of presence equal to one), or partial degree of presence (between zero and one). As in the case of the disorders, m_i will be used to denote both the manifestation name or its intensity, depending on the context. When convenient, the intensities of the manifestations will be also arranged in a vector, denoted as M.

Observations. Each diagnostic problem to be later stated will consist on determining the most possible disorder state (or its possibility distribution) based on the factual knowledge given by the presence/absence of a subset of the manifestations. Such knowledge comes from a particular “subject” being diagnosed, via testing, questioning, measuring, etc. Data gathered from such procedures will be denoted as “observations”.

Measuring m_j will be defined as obtaining an assertion “manifestation m_j takes the value σ_j”. Ideally, the assertion would be equivalent to $m_j = \sigma_j$; however, the distinction between m_j and σ_j will be later used to easily integrate imprecise measurements (or even wrong ones, outliers). The set of measured values will be denoted as Σ.

Example 1 *Consider a medical application, such as the academic example in [49]. Several disorders may be defined, such as “d_1: hay fever”, “d_2: flu”, “d_3:food poisoning”,* etc. *Occurrence of the disorders causes appearance of some manifestations, say “m_1: nose congestion”, “m_2: fever”, “m_3: diarrhea”,* etc.

Each diagnostic case provides a set of values, say σ_j, associated to some of the manifestations m_j: a normal measured value of body temperature may be encoded *as $\sigma_2 = 0$, presence of nose congestion may be encoded as $\sigma_1 = 1$*, etc.

If a fuzzy set is defined on, say, the body temperature readings (defining the linguistic concept of "abnormal temperature" via a meaningful membership function), then the measured σ_2 may range from zero to one: for instance, a "slightly abnormal temperature", say 37.2°C, might be encoded as $\sigma_2 = 0.25$. Depending on the thermometer measurement error characteristics, any σ_2 in $[0.2, 0.3]$ might be possible.

The objective of diagnosis is estimating the actual disorder severity vector D from the *observations* and prior *knowledge* (possibly uncertain) to be described in the section below. In this work, the goal is to express knowledge as a set of *constraints* on D and M, involving as well some auxiliary variables. Intentionally, such constraints will be linear, in order to use well-known algorithms.

3 Diagnostic Methodology (binary case)

Let us first consider the binary case, where disorders, manifestations and observations are either absent or present, represented by numeric values $\{0, 1\}$.

3.1 Knowledge representation

The basic building block of diagnostic knowledge is formed by "associations" from disorders to symptoms representing causal knowledge. Following [15], given a disorder d_i, the set of manifestations known to be caused by d_i alone will be denoted as $M(d_i)^+$, and those known to be certainly absent when d_i alone is present will be denoted as $M(d_i)^-$. The *normal situation* will be associated to all manifestations m_j being equal to zero (changing the definitions of some manifestations by its logic negation, if needed).

Disorder-manifestation association. The relationship between disorder d_i and manifestation m_j will be expressed by an auxiliary "association variable", denoted as s_i^j, fulfilling the following constraints:

- If manifestation m_j is certain to be caused by d_i when d_i *alone* is present ($m_j \in M(d_i)^+$), s_i^j will fulfill, by definition, the equality constraint:

$$s_i^j = d_i \tag{1}$$

- If manifestation m_j will certainly be absent when d_i *alone* is present ($m_j \in$

$M(d_i)^-$), then [1]

$$s_i^j = 0 \tag{2}$$

- If presence or absence of manifestation m_j when d_i is present is uncertain, ***i.e.***, $m_j \in U(d_i)$, $U(d_i) = \neg(M(d_i)^+ \cup M(d_i)^-)$, the following restriction on s_i^j will be added:

$$s_i^j \leq d_i \tag{3}$$

  indicating that, when d_i is present, m_j is sometimes present sometimes absent.

Superposition. Once the relationships between s_i^j and d_i have been set up, the state of presence or absence of a manifestation m_j will be assumed to fulfill the inequalities below:

$$s_i^j \leq m_j \quad \forall\, i \tag{4}$$

$$m_j \leq \sum_i s_i^j \tag{5}$$

Indeed, the first equation indicates that the manifestation intensity must be greater or equal to the severity of the disorders for which m_j is a sure effect (as, for such disorders, the constraint $s_i^j = d_i$ must hold); the second one indicates that at least one of the disorders for which m_j may be an effect ($\neg M(d_i)^-$) must be actually present for m_j to be nonzero.

In this context, the interpretation of $s_i^j = 1$ is "d_i is present and causing m_j"; the interpretation of $s_i^j = 0$ is "either d_i is absent, or it is present but not causing m_j".

Note that the actual value of d_i, s_i^j must be inferred from the observations (known values for some m_j): the presented set of constraints involving disorders, manifestations, and association variables will be denoted as the *knowledge base* (KB).

Remark: If a and b are logic values, the inequality $a \leq b$ reproduces the classical binary implication truth table for $a \rightarrow b$. Then, the proposed knowledge base structure may be equivalently understood as a set of logical assertions ($\vee$ denotes disjunction):

$$normal \rightarrow \neg m_j,\ \forall\, j \tag{6}$$

$$s_i^j \rightarrow d_i, \qquad s_i^j \rightarrow m_j\ \forall\, i, j \tag{7}$$

$$m_j \rightarrow \bigvee_{i=1}^{q} s_i^j\ \forall\, j \tag{8}$$

$$d_i \rightarrow s_i^j \quad \text{if } m_j \in M(d_i)^+ \tag{9}$$

$$d_i \rightarrow \neg s_i^j \quad \text{if } m_j \in M(d_i)^- \tag{10}$$

[1] Note that a variable equal to zero in linear equations may be elliminated. However, equation (2) is kept for clarity and as a base for further developments.

Note that for $m_j \in M(d_i)^-$, (10) entails $s_i^j \to \neg d_i$. Jointly with (7), s_i^j must be false, otherwise both d_i and $\neg d_i$ would be entailed. As discussed in the introduction, the logic-based approach will not be pursued any further (even if, indeed, interesting), in favor of the equation-based representation which is the objective of this paper.

3.2 Incorporating observations

In each particular diagnostic case, the severities of some of the manifestations (m_j) will be assumed available, as measurements σ_j. They will lead to an additional set of constraints which will be denoted as the *Fact Base* (FB).

With a "reliable" sensor (see Section 7 discussing "unreliable" ones in detail), the observed σ_j should be equal to the actual manifestation intensity m_j, as previously commented, *i.e.*, $m_j = \sigma_j$ is added to FB, with σ_j known.

For instance, $\{m_2 = 1, m_3 = 0\}$ would denote the fact that information is available about the presence of a particular manifestation m_2 and the absence of m_3 (and no information about any other relevant manifestation).

The equation-based approach allows for some possibilities of incorporating uncertain measurements. For instance, constraints such as

$$m_2 + m_3 \geq 1 \tag{11}$$

would indicate that at least one of the manifestations m_2 or m_3 must be true, but it is not known which one. Introduction of further auxiliary variables will improve the capabilities of representing uncertain measurements (see Section 7).

3.3 Consistency-based diagnosis.

For brevity, S will denote the set of decision variables s_i^j and, as previously introduced, D will denote the disorder severity vector and M the manifestation intensities. Then, the diagnosis problem may be formulated as a *constraint satisfaction problem*:

> Find the set of feasible decision variables (D, S, M) given the constraints in KB and FB.

Each feasible solution (D, S, M) may be considered as an *explanation* of *all* the possible observations (usually, only F is of interest). If no feasible solution exists, the diagnosis will be said to conclude an *unknown disorder*.

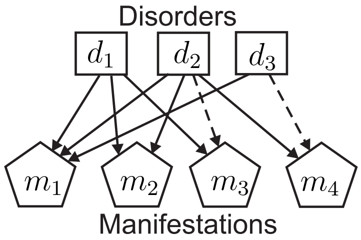

Fig. 1. Example of a basic knowledge base. (—) Solid arrow: certain association; (- - -) dash arrow: uncertain association; no arrow: no causality relation.

All presented constraints are linear in the unknown variables, so integer (more precisely, binary) *linear programming* (LP) [43] may be used as a solution tool, at least in principle. However, software for such problems is far less efficient than real-valued LP implementations (binary LP may be NP-complete, whereas real-valued LP is provably a polynomial-complexity problem). This fact motivates the use of real-valued logic values, *i.e.*, fuzzy logic, in Section 5: luckily, it has more expressivity and better computational efficiency.

4 Examples and further refinements (binary case)

In this section, some examples which illustrate the possibilities of the presented approach will be discussed. Also, additional enhancements to the knowledge representation will be outlined via further examples.

Example 2 *Let us consider the knowledge of the relationships between three disorders and four manifestations (application-dependent linguistic labels are irrelevant) given by:*

$$M(d_1)^+ = \{m_1, m_2, m_3\},\ M(d_1)^- = \{m_4\},\ U(d_1) = \emptyset$$
$$M(d_2)^+ = \{m_1, m_2, m_4\},\ M(d_2)^- = \emptyset,\ U(d_2) = \{m_3\}$$
$$M(d_3)^+ = \{m_1\},\ M(d_3)^- = \{m_2, m_3\},\ U(d_3) = \{m_4\}$$

denoting (1) "when the single disorder d_1 is present, it causes m_1, m_2 and m_3 and does not *cause m_4"; (2) "presence of d_2 causes m_1, m_2 and m_4; regarding m_3, it is sometimes present and sometimes absent (i.e., uncertain)"; (3) "presence of d_3 causes m_1 but, certainly, neither m_2 nor m_3 ever appear; however, presence of m_4 when d_3 is present is uncertain". A graphical representation appears in Figure 1.*

The causal knowledge can be equivalently expressed by the relation coefficients:

$$s_1^1 = s_1^2 = s_1^3 = d_1, \quad s_1^4 = 0 \tag{12}$$
$$s_2^1 = s_2^2 = s_2^4 = d_2, \quad s_2^3 \leq d_2 \tag{13}$$
$$s_3^1 = d_3, \quad s_3^2 = s_3^3 = 0, \quad s_3^4 \leq d_3 \tag{14}$$

jointly with the superposition equations:

$$s_1^1 \leq m_1, \quad s_2^1 \leq m_1, \quad s_3^1 \leq m_1, \quad s_1^1 + s_2^1 + s_3^1 \geq m_1 \tag{15}$$
$$s_1^2 \leq m_2, \quad s_2^2 \leq m_2, \quad s_3^2 \leq m_2, \quad s_1^2 + s_2^2 + s_3^2 \geq m_2 \tag{16}$$
$$s_1^3 \leq m_3, \quad s_2^3 \leq m_3, \quad s_3^3 \leq m_3, \quad s_1^3 + s_2^3 + s_3^3 \geq m_3 \tag{17}$$
$$s_1^4 \leq m_4, \quad s_2^4 \leq m_4, \quad s_3^4 \leq m_4, \quad s_1^4 + s_2^4 + s_3^4 \geq m_4 \tag{18}$$

Evidently, variables equal to zero and those equal to the disorders may be algebraically elliminated in an actual implementation: only the s_i^j terms related to uncertain symptoms need to be kept.

Some refinements in the above framework may be put in place in order to increase its flexibility.

Deviations to the superposition hypothesis. In some cases, the symptoms produced by the combined occurrence of several disorders may be different to the union of those expected from the individual disorders. The introduction of a dummy “compound disorder” variable is then required, as in [29].

Example 3 *Under the superposition hypothesis motivating equations* (4) *and* (5), *combined occurrence of disorder d_1 and d_2 (denoted as d_{12}) would entail a set of disorder-manifestation associations generated by:*

$$M(d_{12})^+ = M(d_1)^+ \cup M(d_2)^+, \quad M(d_{12})^- = M(d_1)^- \cap M(d_2)^-$$

*If that were not the case, a new artificial dummy variable d_{12} (and the association variables s_{12}^j) must be inserted into the knowledge base with suitable constraints describing the particular symptoms occurring (*i.e.*, encoding $M(d_{12})^+$ and $M(d_{12})^-$), plus a mutual exclusivity condition* $d_1 + d_2 + d_{12} \leq 1$.

Other refinements. In some cases, the “uncertainty” in the effects of a disorder may be expressed in terms such as “d_i causes *at least one* of the manifestations m_j, $j \in \mathcal{K}_i$”, where $\mathcal{K}_i$ is an index set of possible consequences of d_i (*cf.* $\mathcal{K}_i$ in [28]). That will amount to stating, in addition to (3), the restriction:

$$\sum_{j \in \mathcal{K}_i} s_i^j \geq d_i \tag{19}$$

The inequality in (19) should be replaced by an equality if *only one* of the manifestations in $\mathcal{K}_i$ may appear, *i.e.*, if $\mathcal{K}$ were a collection of mutually exclusive manifestations.

Example 4 *Consider a situation where a disorder d_1 has uncertain manifestations m_3 and m_4. The knowledge that "at least one of them appears with d_1" may be encoded by*

$$s_1^3 \leq d_1, \quad s_1^4 \leq d_1, \quad s_1^3 + s_1^4 \geq d_1$$

Other causality relations. Other types of disorder and manifestation "interdependence" may also be asserted via more general linear inequalities involving d_i, s_i^j and m_j, enhancing the representational capabilities. The interpretation of some of them will be illustrated below by example.

Example 5 *This example considers the translation of a series of linguistic statements into inequalities.*

- *The statement "occurrence of d_2 is only possible if d_4 or d_5 are also present" (*e.g.*, d_2 may describe a "coagulation disorder" which only occurs in "diabetic (d_4)" or "cardiopathy (d_5)" patients) is represented by the inequality:*

$$d_2 \leq d_4 + d_5 \tag{20}$$

  Note that, conveniently, only the manifestations specific to d_2 need to be defined via (1)–(3) *because* (20) *directly enforces occurrence of those ones associated to d_4 or d_5 if d_2 were present.*
- *"m_3 is* not *caused by d_2 when d_1 is absent" is represented by: $s_2^3 \leq d_1$ (in addition to $s_2^3 \leq d_2$ stating that d_2 sometimes may cause m_3).*
- *"When d_1 is present, d_2 should be also present* and *causing manifestation m_3" is represented by $s_2^3 \geq d_1$ (indeed, note that $s_2^3 \geq d_1$, jointly with the basic $s_2^3 \leq d_2$, entail $d_1 \leq d_2$.)*
- *"m_1 is not caused by d_4 when d_1 is present" is represented by: $s_4^1 \leq d_4$ jointly with $s_4^1 \leq (1 - d_1)$.*
- *"d_3 and d_5 are mutually exclusive" is represented by $d_3 + d_5 \leq 1$,*
- *"m_4 is not caused by d_2 if m_5 is absent" is represented by $s_2^4 \leq m_5$,*
- *"d_1 always causes m_2 when d_4 is present" is represented by $s_1^2 \geq d_1 + d_4 - 1$,*
- *"d_1 sometimes causes m_2 only when d_4 is present" is represented by $s_1^2 \leq d_1 + d_4 - 1$,*
- *"d_1 causes m_2 when d_4 is absent" is represented by $s_1^2 \geq d_1 - d_4$.*
- *"d_1 does not cause m_2 when d_4 is present" is represented by $s_1^2 \leq 1 - d_4$.*

Note that conditions such as those above may be an alternative in many cases to the previously considered need to introduce dummy variables for "multiple disorders".

Note that the above considered enhancements, included in the equation-based approach in this work, provide a more flexible way of expressing disorder-manifestation relationships than the plain lists $M(d)^+$ and $M(d)^-$ considered in [8,28].

A last example considers several of the presented options in the knowledge description.

Example 6 *Consider a knowledge base with 4 disorders (d_1 to d_4) and 5 manifestations (m_1 to m_5). For convenience, the disorder d_{14} will be introduced as an auxiliary variable used to express the combined occurrence of d_1 and d_4.*

The available pieces of knowledge will be now described in linguistic terms and, then, translated to inequalities in order to illustrate the methodology:

- *Knowledge on d_1: it causes m_1, it only occurs if d_3 is present.*

$$s_1^1 = d_1, \quad s_1^2 = s_1^3 = s_1^4 = s_1^5 = 0, \quad d_1 \leq d_3 \tag{21}$$

- *Knowledge on d_2: it causes m_2 and, occasionally, also m_4.*

$$s_2^1 = 0, \quad s_2^2 = d_2, \quad s_2^3 = 0, \quad s_2^4 \leq d_2, \; s_2^5 = 0 \tag{22}$$

- *Knowledge on d_3: it causes m_3 and, occasionally, also m_5 (but only when d_2 is present and causing m_4 or when d_4 is present).*

$$s_3^1 = s_3^2 = s_3^4 = 0, \quad s_3^3 = d_3, \quad s_3^5 \leq d_3, \qquad s_3^5 \leq s_2^4 + d_4 + d_{14} \tag{23}$$

- *Knowledge on d_4: it causes m_4 and m_5 and, occasionally, also m_2.*

$$s_4^1 = s_4^3 = 0, \quad s_4^2 \leq d_4, \quad s_4^4 = s_4^5 = d_4 \tag{24}$$

- *Combined occurrence of d_1 and d_4 causes m_1 and m_4 (but not m_2, which d_4 alone can sometimes cause). As d_1 causes d_3, m_5 is still possible, but not definitely sure (so the behaviour is different to that with d_4 alone).*

$$s_{14}^1 = d_{14}, \quad s_{14}^2 = s_{14}^3 = s_{14}^5 = 0, \quad s_{14}^4 = d_{14}, \quad d_{14} \leq d_3 \tag{25}$$

  The mutual exclusivity between d_1, d_4 and d_{14} is also added to the knowledge base as an inequality:

$$d_1 + d_4 + d_{14} \leq 1 \tag{26}$$

- *Manifestations m_3 and m_4 cannot occur simultaneously.*

$$m_3 + m_4 \leq 1 \tag{27}$$

The pictograph in Figure 2 approximately depicts the relationships between the different entities in the knowledge base (as well as one particular set of observations, see below).

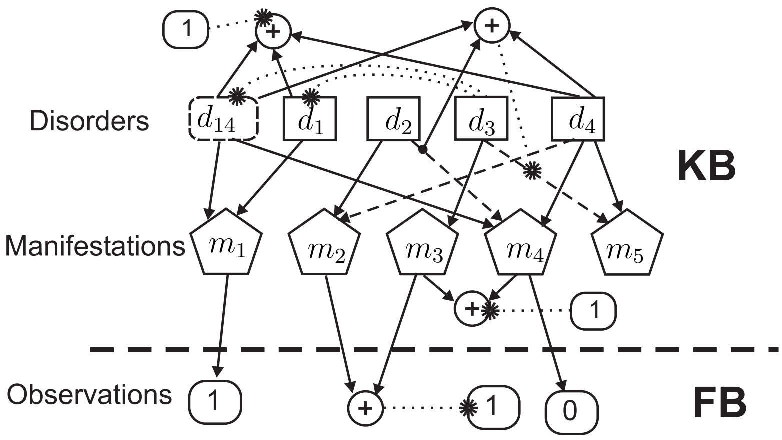

Fig. 2. A graphical description of a diagnostic problem. (—) Solid arrow: certain association; (- - -) dash arrow: uncertain association; (. . *) dotted starred-tip arrow: inequality (the star denotes the lowest term, such as a in $a \leq b$).

The knowledge base is completed with the superposition equations which are:

$$s_1^1 \leq m_1, \quad s_{14}^1 \leq m_1, \quad s_1^1 + s_{14}^1 \geq m_1 \tag{28}$$

$$s_2^2 + s_4^2 \geq m_2, \quad s_2^2 \leq m_2, \quad s_4^2 \leq m_2, \tag{29}$$

$$s_3^3 = m_3 \tag{30}$$

$$s_{14}^4 + s_2^4 + s_4^4 \geq m_4, \quad s_{14}^4 \leq m_4, \quad s_2^4 \leq m_4, \quad s_4^4 \leq m_4 \tag{31}$$

$$s_3^5 \leq m_5, \quad s_4^5 \leq m_5, \quad s_3^5 + s_4^5 \geq m_5 \tag{32}$$

Note that decision variables which were assigned a zero value in (21)–(27) *have been omitted.*

Fact base. *The equations in the fact base arising from the measurements will change for each diagnostic case. Consider, for instance, the following measured information "m_1 is present; at least one of m_2 or m_3 are present; m_4 is absent; m_5 is not measured". Then, the fact base (also depicted in the figure) will be:*

$$m_1 = 1, m_2 + m_3 \geq 1, m_4 = 0 \tag{33}$$

Inference. *Consistency based-inference amounts to finding the set of feasible diagnostics. For instance, in the example being considered, the presence of m_1 entails d_1 and, hence, d_3 (which would explain the observed "m_2 or m_3"). d_4 must be absent as, otherwise, m_4 would be present (the joint d_{14} disorder is also discarded by the absent m_4) . Regarding d_2, it may be present (but not causing m_4). Denoting* $D = (d_1, d_2, d_3, d_4)$*, the feasible diagnostics are* $(1, 0, 1, 0)$ *and* $(1, 1, 1, 0)$.

5 The fuzzy approach.

In the above knowledge-representation framework, the need for integer programming is removed if the decision variables are allowed to take any positive real value or, possibly better, values in the interval $[0, 1]$ in order to integrate the approach into a standard fuzzy logic framework. Indeed, such bounding has an elegant interpretation in the framework of fuzzy sets, when:

- The observations σ_i are the result of fuzzification of real-world measurements, indicating an "intensity of presence" of a particular manifestation, $0 \leq \sigma_i \leq 1$. For instance, a fuzzy observation may indicate wether a patient's fever is more or less "high"; obviously, it increases the expressive power of the knowledge base (considering "degrees of fever" instead of a plain binary choice "fever" vs. "not fever").
- The disorder severities f_i may be gradual (partial occurrence of a disorder; [24]).
- Expression (1) is interpreted as: "the intensities of the manifestations appearing with a disorder increase as its severity d_i does", which is indeed reasonable from a commonsense point of view, at least approximately.

With this simple modification to the search space for solutions and the allowable values of observations, the examples in the previous sections could be run with, say, a fact base given by: $m_1 = 0.7$, $m_2 + m_3 \geq 0.8$, $m_4 < 0.1$.

Remark: In the above interpretation, fuzziness refers to "gradation" in fulfilment of linguistic properties and certainly not to "uncertainty" or "possibility". Uncertainty in the knowledge base or the observations will be incorporated below by relaxing some constraints. A more complete discussion on uncertain knowledge appears in Sections 6 and 7.

5.1 Uncertainty in fuzzy diagnosis

Up until now, the only "uncertainty" in an association between disorders and manifestation was in the form (3) or in inequalities in the enhancement examples discussed in Section 4. The fuzzy approach opens up additional flexibility.

Uncertain fuzzy knowledge. Equations (1), (2) and (3) are a particular case of a more general "sector constraint":

$$a_i^j d_i \leq s_i^j \leq b_i^j d_i \tag{34}$$

where a_i^j and b_i^j are user-defined parameters. Indeed:

- (1) uses $a_i^j = b_i^j = 1$ to denote a "certain" disorder→manifestation association,

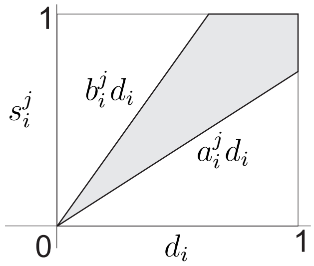

Fig. 3. Uncertain fuzzy knowledge (feasible zone appears shaded).

- (2) uses $a_i^j = b_i^j = 0$, to denote the opposite (*i.e.*, non-association)
- (3) uses $a_i^j = 0$, $b_i^j = 1$ to denote “uncertain” associations.

However, nothing precludes the possibility of using different values for a_i^j and b_i^j to suit particular intermediate requirements (Figure 3), even setting up arbitrary linear constraints such as

$$a_i^j d_i + p_i^j \leq s_i^j \leq b_i^j d_i + q_i^j$$

letting the expert providing the rules choose such constants, if he wishes so.

Example 7 *For instance, asserting “when d_5 is fully present, manifestation m_3 is sometimes fully present, sometimes only* `weakly` *present” may be interpreted as $0.3d_5 \leq s_5^3 \leq d_5$. If m_3 always appeared weakly with d_5, the piece of knowledge might be encoded as* $0.3d_5 \leq s_5^3 \leq 0.4d_5$.

Also, (19) may be generalised to:

$$\alpha d_i \leq \sum_{j \in J} s_i^j \leq \beta d_i \tag{35}$$

for some user-defined α and β, with the interpretation in the example below.

Example 8 *The assertion “d_1 may cause m_1, m_2, m_3 or m_4; in particular, it is known to cause at least one of the manifestations plus another one weakly” may be approximately translated to the constraint* $1.5d_1 \leq s_1^1 + s_1^2 + s_1^3 + s_1^4$ *(in addition to the basic $s_1^1 \leq d_1$, $s_1^2 \leq d_1, \ldots$).*

Remark: As realistic applications are complex and non-linear, a disorder with multiple associated manifestations will not, in most cases, fulfill the “exact” linear equation (1): introducing uncertainty in the rules (or in the observations, see below and Section 7) will be a must to obtain feasible solutions (see footnote 2 for a trivial example).

Uncertain observations. Some observations may be uncertain or not available. They may be represented by considering observation data Σ having an interval logic value $[\sigma_{l,j}, \sigma_{u,j}]$. In this way, uncertain measurements are translated as the constraints:

$$\sigma_{l,j} \leq m_j \leq \sigma_{u,j} \tag{36}$$

where $\sigma_{l,j}$ and $\sigma_{u,j}$ are known quantities.

Additionally, if the only knowledge is that at least one manifestation in a set, say K, is true to a degree α, then the restriction $\sum_{k\in K} m_k \geq \alpha$ may be added to express such a situation, generalising (11).

In this way, the equation-based approach presented in this work approximately incorporates some ideas related to intuitionistic fuzzy sets [2,10] or interval-valued fuzzy set theory [12,39].

5.2 Fuzzy consistency-based diagnosis

In the above presented context, the consistency-based diagnosis problem can be understood as a real-valued LP *feasibility* problem [18,43] with decision variables d_i, s_i^j and m_j. Hence, elementary LP theory shows that the set of feasible decision variables will be a convex polytope. The set of feasible diagnostics, *i.e.*, the set of feasible values of D will be denoted as $\mathcal{D}^{feas}$, representing the possible disorder coordinates consistent with the available observations.

The presented methodology in a fuzzy case is, however, overly impractical at this stage. Indeed, describing $\mathcal{D}^{feas}$ may be cumbersome (enumerating a possible large set of vertices) and, even worse, the solution set $\mathcal{D}^{feas}$ may be very large (*i.e.*, almost uninformative, for instance when few observations are available) or empty [2]. Fortunately, such difficulties will be overcome in the next sections.

6 A Possibilistic Framework: Preliminaries

In the paper [13], possibilistic constraint satisfaction problems (CSP) are presented. The authors introduce constraints which are satisfied to a degree, transforming the feasibility/infeasibility of a potential solution into a gradual notion: given a CSP with multiple solutions $\delta \in \Delta$ (where Δ denotes the set of all feasible values for the decision variables), a fuzzy relation $R : \Delta \to [0, 1]$ was suggested in order to represent preference or priority as a "consistency degree"; R was built by conjunction of individual relations (expressing preference or priority on each individual constraint), and the best CSP solutions were defined as those which satisfy the global problem to the maximal degree. In the cited work, the authors state that "R can be viewed as a possibility distribution, prescribing to what extent a value δ for the decision variables is judged suitable according to the constraints".

[2] For instance, an unfeasible set of constraints results with the trivial knowledge base: "d_1 always causes m_1 and m_2", encoded as $\{d_1 = m_1 = m_2\}$ after variable ellimination, and the fact base obtained from measurements $\{m_1 = 0.5, m_2 = 0.50001\}$.

In related literature [17], possibility distributions were denoted with π, instead of R. The possibility of an event A (subset of Δ) is computed, as usual, via:

$$\pi(A) = \sup_{\delta \in A} \pi(\delta) \tag{37}$$

In particular, for a multidimensional $\Delta = \Delta_1 \times \Delta_2$, $\delta = (\delta_1, \delta_2) \in \Delta$, the *marginal* possibility of δ_1 is defined as:

$$\pi(\delta_1) = \sup_{\delta_2 \in \Delta_2} \pi(\delta_1, \delta_2) \tag{38}$$

Hence, possibility computations are optimisation problems.

Conversely, consider a cost function $J : \Delta \mapsto \mathbb{R}^+$ (*i.e.*, verifying $J(\delta) \geq 0$ for all $\delta \in \Delta$), so that there exists $\delta_0 \in \Delta$ such that $J(\delta_0) = 0$. Then, a possibility distribution may be defined on Δ via:

$$\pi(\delta) = e^{-J(\delta)} \quad \delta \in \Delta \tag{39}$$

so the possibility of an event A is given by replacing the possibility definition (39) in (37), resulting in:

$$\pi(A) = e^{-\inf_{\delta \in A} J(\delta)} \tag{40}$$

In the next sections, an event A will be usually described by a set of constraints on the decision variables δ. In this way, numeric constrained optimisation problems may be interpreted in possibilistic terms: the cost $J(\delta)$ will be interpreted as the log-possibility of δ and, by definition, unfeasible values of decision variables will be assigned zero possibility. For details on definitions and computations with possibility distributions, the reader is referred to [17,13].

7 Possibilistic knowledge bases

Up to now, the diagnostic problem has been cast as a feasibility linear programming setup, i.e., each set of decision variables (D, M, S) is either feasible or unfeasible.

By means of the introduction of some artificial variables and a cost index, as discussed above, Expression (39) will enable grading the different candidate diagnosis as more or less "possible": optimization software will be able to query the "most possible disorder" or the (marginal) possibility of a particular event such as "disorder 7 is present and not causing manifestation 2".

In summary, the knowledge and fact bases (after some modifications or re-interpretations) will define a possibility distribution in the space of decision variables (disorders, association variables, manifestations, *etc.*). Let us discuss the details on how a meaningful cost index can be built.

Remark: The proposed methodology is not the only available choice as, in principle, optimization could be carried out with arbitrary cost indices and constraints. However, the proposals below are, in the author's opinion, a sensible compromise between, on one hand, readability and linguistic interpretability, and on the other hand, optimization performance.

Possibility of basic disorders. The basic disorders will appear in the overall cost index via a component:

$$J_D = \alpha_1 d_1 + \alpha_2 d_2 + \dots \tag{41}$$

so, each basic disorder d_i will contribute to the cost index via a linear term $\alpha_i d_i$. The log-possibilistic interpretation of such a term is that "the *a priori* possibility of a particular disorder d_i having a severity f is $e^{-\alpha_i f}$". With the above cost index, disorders are assumed *non-interactive* (its prior possibility is independent of that of other disorders) and the normal situation (all $d_i = 0$) is assigned possibility one [3] .

The term J_d may be modified including cross-product terms such as $\alpha_{ij} d_i d_j$, indicating that simultaneous occurrence of d_i and d_j (*i.e.*, both different from zero) is "less possible". This is a case where the non-interactivity assumption is relaxed. Quadratic programming would be required to solve the optimisation problem. Example 15, discussed later, incorporates such a possibility.

Knowledge constraints. All the inequalities conforming a knowledge base described in Section 3 may be cast as an equality via suitable artificial slack variables. Indeed, $a \leq b$ is equivalent to saying that there exists a positive ϵ such that $a + \epsilon = b$. This is a widely used trick in optimization.

Then, these additional artificial variables may be present in the cost index to generate a possibility distribution. This procedure will be denoted as "constraint softening" and its details will be outlined below.

[3] An alternative guide to set up suitable cost index may be a probabilistic interpretation: *a priori* possibility may be (roughly) linked to *a priori* probability. For ample discussion on the relationships between possibility and probability the reader is referred to [16].
Let us assume that the *a priori* probability of disorder d_i having severity f_i is governed by a density distribution proportional to $e^{-\lambda_i f_i}$, where λ_i is a known coefficient. Let us also assume that the various disorders are statistically independent. Then, the joint probability density would be proportional to $e^{-\sum_{i=1}^{k} \lambda_i f_i}$, so the cost index would be proportional to the log-probability. In order to set up an approximate "rule of thumb" to determine λ_i, we have:

$$P(f_i \geq a) = \frac{\int_a^1 e^{-\lambda_i \psi}\, d\psi}{\int_0^1 e^{-\lambda_i \psi}\, d\psi} = \frac{e^{-\lambda_i a} - e^{-\lambda_i}}{1 - e^{-\lambda_i}} \tag{42}$$

For instance, if "significant disorder" is understood as one with severity higher than 0.66, the assertion "*10% of the diagnostic cases exhibit disorder d_i with a significant degree*" will result in setting $\lambda_i = 2.8$ (obtained by solving for λ_i in $(e^{-0.66\lambda_i} - e^{-\lambda_i})/(1 - e^{-\lambda_i}) = 0.1$). If the frequency were 5%, $\lambda_i \approx 4.1$; if it were 1%, $\lambda_i \approx 6.8$.

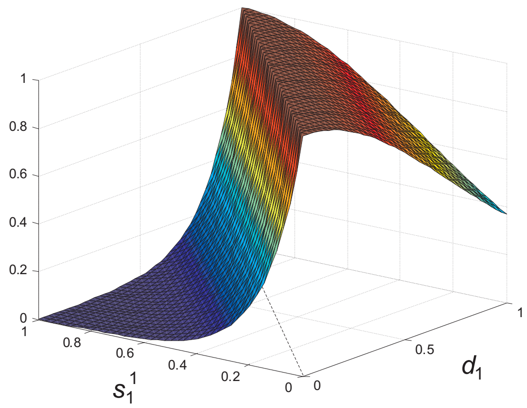

Fig. 4. Possibility distribution arising from an "asymmetrically softened" equality constraint.

Softening equality constraints. Consider an equality restriction $a = b$. A relaxed ("softened") version of such restriction may be written as:

$$a = b + \epsilon - \nu, \quad \epsilon \geq 0, \quad \nu \geq 0 \tag{43}$$

with ϵ and ν being artificial variables. If ϵ and μ are penalised in an optimisation index (typically with cost index terms such as, say, 10ϵ or $7\epsilon^2$), the possibilistic interpretation of (43) may be:" $a = b$ is fully possible and consistent with my knowledge; the more a differs from b, the less possible such situation is". The initial equality constraint would be recovered if the cost index weights on the artificial variables tend to infinity.

Example 9 *An association $s_1^1 = d_1$ may be "softened" by stating*

$$s_1^1 = d_1 + \epsilon - \nu, \quad \epsilon \geq 0, \quad \nu \geq 0$$

and building a cost index

$$J(s_1^1, d_1, \epsilon, \nu) = \nu^2 + 7\epsilon$$

In this way, the possibility of $(s_1^1 = 0.5, d_1 = 0.5)$ is 1 (the cost J is zero), that of $(s_1^1 = 1, d_1 = 0.5)$ is $e^{-7*0.5} = 0.03$, and that of $(s_1^1 = 0, d_1 = 0.5)$ is $e^{-0.5^2} = 0.78$. In linguistic terms: with the disorder d_1 occurring at medium intensity, it is "quite possible" for manifestation m_1 not to appear, it is "almost impossible" for it to appear at a higher degree than that of d_1. Figure 4 depicts the possibility distribution. Due to the conception of J, the softening has been intentionally made non-symmetric (cf., *for instance, with $J = 7\epsilon + 7\nu$ which would have resulted in a "symmetric" penalization).*

Inequalities. A "softened" inequality restriction is nothing but an equality one with no penalisation on one of the artificial variables above.

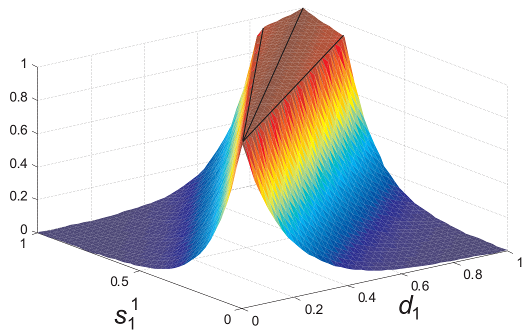

Fig. 5. Possibility distribution with 4 artificial variables: softened sector constraint.

Example 10 *Considering Example 9, if the cost index were:*

$$J(s_1^1, d_1, \epsilon, \nu) = 7\epsilon$$

all pairs (s_1^1, d_1) *with* $s_1^1 \leq d_1$ *would have possibility equal to 1 (zero cost). The possibility of* $(s_1^1 = 1, d_1 = 0.5)$ *is still* $e^{-7*0.5} = 0.03$.

Example 11 *The use of more artificial variables allows greater flexibility of the proposed representations. For instance, expressing:*

$$s_1^1 = d_1 + \epsilon_1 + \epsilon_2 - \nu_1 - \nu_2, \tag{44}$$
$$0 \leq \epsilon_1 \leq 0.2d_1, \quad 0 \leq \epsilon_2, \quad 0 \leq \nu_1 \leq 0.2d_1, \quad 0 \leq n_2 \tag{45}$$

with an associated:

$$J(s_1^1, d_1, \epsilon_1, \epsilon_2, \nu_1, \nu_2) = 0.4\epsilon_1 + 7\epsilon_2 + 0.4\nu_1 + 3\nu_2 \tag{46}$$

induces a possibility distribution. The possibility of a pair (s_1^1, d_1) *is calculated, from* (40)*, from the minimum value of the cost index over the set of feasible decision variables (*ϵ_1*,* ϵ_2*,* ν_1*,* ν_2*). As all constraints above are linear, the problem can be solved by linear programming. For instance, the possibility of* $(s_1^1 = 0.55, d_1 = 0.5)$ *is* $e^{-0.4*0.05} = 0.98$*, corresponding to the possibility of the "singleton" in the decision variable space given by* $(s_1^1 = 0.55, d_1 = 0.5, \epsilon_1 = 0.05, \epsilon_2 = \nu_1 = \nu_2 = 0)$*. By computing such possibility values in a grid of values for* s_1^1 *and* d_1*, the marginal possibility distribution in Figure 5 can be computed. The setting proposed in this example can be used to soften sector constraints* (34)*. Note that, once the constraints* (44)–(45) *and cost* (46) *are integrated with the rest of the knowledge and fact bases, all the above computations are carried out by the optimization software, transparently to the end-user.*

Manifestations and Observations. If the measured value of manifestation m_j is σ_j, the basic assertion $m_j = \sigma_j$ may be relaxed with additional variables, related to each sensor's "reliability", conforming a possibility distribution in (m_j, σ_j) associated to some cost index terms J_M.

Example 12 *Analogously to Example 11, an interval measurement (plus some possibility of outlier measurements) may be described via:*

$$m_j = \sigma_j + \epsilon_1 - \nu_1 + \epsilon_2 + \nu_2,\ 0 \le \epsilon_1 \le 0.05,\ 0 \le \nu_1 \le 0.05,\ 0 \le \epsilon_2,\ 0 \le \nu_2$$

and a cost index incorporating the term $5\epsilon_2 + 5\nu_2$. *The constraints describe an interval measurement, where the possibility of* $\sigma_j \in [m_j - 0.05, m_j + 0.05]$ *is one and the possibility of* σ_j *being out of the referred interval depends on the cost index weights. For instance, with a measured* $\sigma_j = 0.1$, *the possibility of the actual manifestation being* $m_j = 0.95$ *would be* $e^{-5*0.8} = 0.018$, *the possibility of* $m_j = 0.14$ *would be one.*

Sets of equations: the overall cost associated to the diagnostic problem. Consider a set of equations (softened constraints) with its associated artificial variables $\delta_1 = \{\epsilon_1, \nu_1\}$, $\delta_2 = \{\epsilon_2, \nu_2\}$, *etc.* defined as above, with individual cost indices $J_1(\delta_1)$, $J_2(\delta_2)$, etc. The cost index defined by:

$$J(\delta_1, \delta_2, \dots) = J_1(\delta_1) + J_2(\delta_2) + \dots \tag{47}$$

will be the choice to define a possibility distribution in the product space, interpreting the possibilistic conjunction operator in [13] as an algebraic product:

$$\pi(\delta_1, \delta_2, \dots) = e^{J(\delta_1, \delta_2, \dots)} = e^{J_1(\delta_1)} e^{J_2(\delta_2)} \dots = \pi_1(\delta_1)\pi_1(\delta_2)\dots \tag{48}$$

under the assumption of non-interactivity between the decision variables in different constraints [4] .

As a result, an overall cost index can be built with reasonable ease by adding up all components from individual constraints plus those from the *a priori* disorder possibilities in (41):

$$J = J_D + J(\delta_1, \delta_2, \dots) \tag{49}$$

[4] The choice of the algebraic product as the aggregation operator is based on heuristics: the more not-fully-possible constraints that are needed to explain an hypothesis, the lower the possibility of a diagnostic result is considered to be. It is not the only alternative: the "minimum" conjunction might be interpreted as "the possibility of an hypothetical disorder is that of the least-possible knowledge item explaining it". It would correspond to defining

$$J(\delta_1, \delta_2, \dots) = \max_i J_i(\delta_i)$$

and solving the optimization problem:

$$\begin{array}{rl} \text{minimise} & \gamma \\ \text{subject to KB, FB and} & J_i(\delta_i) \le \gamma \end{array}$$

which (if the cost and constraints are linear) is a well-known minimax problem, solvable as well by linear programming.

In this way, a possibilistic knowledge base will be formed by a set of constraints where the involved decision variables are:

(1) The disorder severities d_i
(2) The association variables s_i^j
(3) The manifestation variables m_j
(4) The artificial variables ϵ, ν needed for the "softening" of the basic constraints in the knowledge and fact bases.

where some (known) measurement data σ_j appear in some of the constraints for each particular diagnostic problem.

The cost index value reflects the log-possibility of a particular combination of the above decision variables. Section 8 provides detailed examples on the procedure.

7.1 Possibilistic diagnosis results.

Once a set of measurements σ_j is available, the output of the diagnostic system should be the minimum-cost (maximum possibility) combination of decision variables after fixing σ_j. The obtained possibility value is, hence, the "a priori" possibility of encountering such measurements.

However, in a situation where multiple diagnostic results might be reasonably possible, there would be a need to report to the end-user such alternate options.

Conveniently, marginal possibility distributions may be easily plotted: the marginal possibility (38) of $d_i = f$, (where f is a known numerical value) is computed by adding $d_i = f$ to FB and minimising J over the rest of the decision variables. Such marginal possibility distributions can be plotted on a grid of values for f in the $[0, 1]$ interval; such a plot will be interpreted as the "possibility distribution of the fault severity given the observations".

Furthermore, if a quotient definition for conditional possibility [5] of an event A given event B were used (i.e., $\pi(A|B) := \pi(A \cap B)/\pi(B)$), conditional possibilities may be easily computed.

Indeed, let B stand for the event (i.e., the set of decision variables values) given by the knowledge and fact base constraints $\{$KB,FB$\}$; and let A be a smaller event (i.e., with additional constraints) described by an additional set of constraints $\{$Q$\}$.

[5] Apart from the above considered formula, reminiscent of conditional probability, different alternative definitions for conditional possibility have been proposed in literature, see [16,11].

Then:

$$\pi(A|B) := \frac{\pi(A \cap B)}{\pi(B)} = \frac{exp\left(\min_{\text{constraints \{KB,FB,Q\} in place}} J\right)}{exp\left(\min_{\text{constraints \{KB,FB\} in place}} J\right)} \tag{50}$$

The numerator and denominator above can be obtained by solving the associated optimization problems.

In the diagnostic problem at hand, the conditional possibility of a particular event (disorder state) given the observations may be understood as an *a posteriori* possibility, in an analogue to Bayesian inference.

The details of some of the above issues will be illustrated by example in next section.

8 Additional Examples

This section presents some examples illustrating the approach in the possibilistic-binary and possibilistic-fuzzy cases.

Example 14. Let us consider the binary diagnostic problem in [49]§5, where a set of 3 diseases and 3 manifestations (binary) are defined. Denote disorders by $\{d_1$: hay fever, d_2: flu , d_3: food poisoning$\}$. Denote manifestations by:$\{m_1$: nasal congestion, m_2: high fever, m_3: diarrhea$\}$.

The knowledge base in the cited reference provides conditional possibilities for each of the $3 \times 3 \times 2 = 18$ cases (both the manifestations and its negation are considered). For instance,

$$\pi(nc : hay|hay) = 1, \quad \pi(\overline{nc : hay}|hay) = 0.3$$

denotes that the possibility of *hay fever* causing *nose congestion* is 1, but its necessity is 0.7 (the possibility of *not* causing the symptom is 0.3). In the approach in this paper, the conditional possibility provided gives information about the s_1^1 association variable; the above assertions are translated into the constraints:

$$s_1^1 = d_1 - \epsilon, \quad 0 \le \epsilon \le d_1 \tag{51}$$

jointly with an additive term in the overall cost index $-\ln(0.3)\epsilon$, *i.e.*, 1.204ϵ. Indeed, the interpretation of (51) jointly with the cost in ϵ is that the most possible situation is $s_1^1 = d_1$ (i.e., hay fever causes nose congestion), because in that case $\epsilon = 0$ gives zero cost (possibility one); otherwise, the possibility of ($d_1 = 1$, $s_1^1 = 0$, $\epsilon = 1$) is $e^{-1.204} = 0.3$.

As another example, the assertions

$$\pi(diarrhea : flu|flu) = 0.4, \quad \pi(\overline{diarrhea : flu}|flu) = 1$$

are translated into:

$$s_2^3 \leq d_2, \quad s_2^3 = \epsilon_5 \tag{52}$$

adding a term in the cost index $-\ln(0.4)\epsilon_5 = 0.92\epsilon_5$. Indeed, assuming $d_2 = 1$, the most possible situation (zero cost) is $s_2^3 = 0$ (no diarrhea produced by flu); the situation $s_2^3 = 1$ would entail $\epsilon_5 = 1$, *i.e.*, a cost of 0.92 associated to a possibility 0.4.

Following this methodology, all the possibilistic knowledge base in [49] may be expressed as:

$$\begin{array}{r}
s_1^1 = d_1 - \epsilon_1, \quad s_2^1 = \epsilon_3, \quad s_3^1 = 0 \\
s_1^2 = \epsilon_2, \quad s_2^2 = d_2 - \epsilon_4, \quad s_3^2 = \epsilon_6 \\
s_1^3 = 0, \quad s_2^3 = \epsilon_5, \quad s_3^3 = d_3 - \epsilon_7 \\
0 \leq \epsilon_1 \leq d_1, \quad 0 \leq \epsilon_2 \leq d_1, \quad 0 \leq \epsilon_3 \leq d_2, \quad 0 \leq \epsilon_4 \leq d_2 \\
0 \leq \epsilon_5 \leq d_2, \quad 0 \leq \epsilon_6 \leq d_3, \quad 0 \leq \epsilon_7 \leq d_3
\end{array}$$

jointly with the superposition equations

$$\begin{array}{r}
m_1 \leq s_1^1 + s_2^1 + s_3^1 \quad m_2 \leq s_1^2 + s_2^2 + s_3^2 \quad m_3 \leq s_1^3 + s_2^3 + s_3^3 \\
s_i^j \leq m_j, \qquad i = 1, \ldots 3, \quad j = 1, \ldots, 3
\end{array}$$

with a cost index:

$$J = 0.51d_1 + 0d_2 + 0.92d_3 + 1.2\epsilon_1 + 1.61\epsilon_2 + 0.36\epsilon_3 + 1.2\epsilon_4 + 0.92\epsilon_5 + 0.69\epsilon_6 + 2.3\epsilon_7$$

In the particular case considered in [49], observation data are knowledge about the presence of *hay fever*, $d_1 = 1$, as well as *high fever*, $m_2 = 1$, *i.e.*, FB=$\{d_1 = 1$,$m_2 = 1\}$.

Given the knowledge base, the possibility of such measured data is $e^{-0.51} = 0.6$, corresponding to the minimum value of J obtained after enforcing the constraints in KB and FB. The notation $\pi(\{KB, FB\}) = 0.6$ will be used. The most possible set of decision variables, yielding $J = 0.51$ is (only nonzero variables shown):

$$d_1, d_2, m_1, m_2, s_1^1, s_2^2 \tag{53}$$

pinpointing that, possibly, *high fever* (m_2) is caused by *flu* (d_2), as $s_2^2 = 1$, and that presence of m_1 (*nose congestion*) is caused by d_1(*hay fever*), as $s_1^1 = 1$.

To get a better description of the resulting diagnosis, the marginal possibility for *not flu* (given the current observations) may be calculated, by adding $d_2 = 0$ to the equality constraints in FB. The minimum cost is $J = 2.12$, achieved with two alternative explanations:

- "$d_1 = d_3 = \epsilon_6 = 1$, rest of variables equal to zero", interpreted as:
  "if flu were assumed not possible, simultaneous *hay fever* and *food poisoning* would be the most possible diagnosis, *food poisoning* being responsible for the present *high fever*".
- "$d_1 = \epsilon_2 = 1$, the rest of the variables zero", interpreted as "*hay fever* is causing the observed fever and no other disorder is present".

The exponential of the achieved cost may be considered as possibility $e^{-2.12} = 0.12$ of *not flu*.

The conditional *a posteriori* possibility of $d_2 = 1$ (conditioned to the available knowledge and fact bases) is 1, because the most possible explanation features $d_2 = 1$. The conditional possibility of $d_2 = 0$ is given by:

$$\pi(\{d_2 = 0\} \mid \{KB, FB\}) = \frac{\pi(\{d_2 = 0\} \cup \{KB, FB\})}{\pi(\{KB, FB\})} = \frac{0.12}{0.6} = 0.2$$

Conditional possibilities for all decision variables may be calculated in the same way. For instance, $\pi(\{m_1 = 1\} \mid \{KB, FB\}) = 1$, $\pi(\{m_1 = 0\} \mid \{KB, FB\}) = 0.3$, *etc.*. The numerical results in this case are coincident with those in [49]; however, accurate equality of results is not claimed in a general case.

Example 15. This example briefly illustrates the advantages of the possibilistic approach in situations where measurements are too few to distinguish between several disorders with similar manifestations. Also, the example will show the usefulness of the approach in inconsistent situations arising from a faulty measurement.

Consider the knowledge base in Example 2 describing the relationship between three disorders and four abnormal manifestations, refined with the possibilistic information below:

- Disorder d_1 causes m_1, m_2 and m_3, but not m_4. The *a priori* log-possibility of d_1 is $-2.8d_1$ (which can be interpreted as "d_1 occurs significatively in 10% of cases", from (42))
- d_2 causes m_1, m_2, and m_4 with its effects on m_3 being partly uncertain: $0.5d_2 \leq s_2^3 \leq d_2$. The *a priori* log-possibility of d_2 is $-4.1d_2$ ('d_2 occurs significatively in 5% of cases")
- d_3 causes m_1, but neither m_2 nor m_3; its effects on m_4 are uncertain (in most cases m_4 is present but the possibility of m_4 not appearing when d_3 is present is 0.1; translated as $s_3^4 = d_3 - \epsilon$ jointly with $-\ln(0.1)\epsilon = 2.3\epsilon$ in the cost index). The *a priori* log-possibility of d_3 is $-4.5d_3$.

The knowledge base discussed in page 8 has been now modified to the following one:

$$s_1^1 = s_1^2 = s_1^3 = d_1, \quad s_2^1 = s_2^2 = s_2^4 = d_2,\ 0.5d_2 \le s_2^3 \le d_2$$
$$s_3^1 = d_3\ s_3^2 = s_3^3 = 0, s_3^4 = d_3 - \epsilon$$

jointly with the superposition equations, uncertain measurement equations $m_1 = \sigma_1 + \epsilon_1 - \nu_1$, $m_2 = \sigma_2 + \epsilon_2 - \nu_2$, $m_3 = \sigma_3 + \epsilon_3 - \nu_3$, $m_4 = \sigma_4 + \epsilon_4 - \nu_4$ and constraining all variables to belong to the interval $[0, 1]$. The log-possibilistic cost index has been defined as

$$J_1 = 2.8d_1 + 4.1d_2 + 4.5d_3 + 2.3\epsilon + 10\sum_{i=1}^{4}(\epsilon_i + \nu_i)$$

where the rightmost term denotes that the possibility of a measurement totally contrary to the actual value of a manifestation (i.e, either ϵ_i or ν_i equal to one) is set to be $e^{-10} = 0.000045$. In this way, no observation will have zero possibility and infeasibility problems will be avoided [6] .

Furthermore, cross-product terms have been added penalising $20d_id_j$ to indicate preference for single-disorder results; cross-product penalisation has also been added in the positive ϵ and negative ν sensor error terms, to penalise combinations of sensor errors with the same sign ($2\epsilon_i\epsilon_j$ and 2 $\nu_i\nu_j$), deemed a priori less possible. The overall cost index is, then

$$J = J_1 + 20(d_1d_2 + d_1d_3 + d_2d_3) + 2(\epsilon_1\epsilon_2 + \epsilon_1\epsilon_3 + \epsilon_1\epsilon_4 + \epsilon_2\epsilon_3 + \epsilon_2\epsilon_4 + \epsilon_3\epsilon_4) +$$
$$+ 2(\nu_1\nu_2 + \nu_1\nu_3 + \nu_1\nu_4 + \nu_2\nu_3 + \nu_2\nu_4 + \nu_3\nu_4)$$

Case 1: In order to illustrate the diagnostic output when many disorders are possible, let us consider that the only available measurement is that of m_1, *i.e.*, $\sigma_1 = 0.9$. As m_1 is produced by all disorders, all of them are partially possible. The most possible disorder vector resulting from the optimisation of J is (0.9,0,0,0), with possibility $e^{-2.8*0.9} = 0.08$. The marginal possibilities for each disorder are plotted in Figure 6: they indicate that there is some possibility (0.025) towards accepting $d_2 = 0.9$ as a diagnostic, as well as some evidence (0.017) of the possibility of $d_3 = 0.9$. Under such an scarce measurement information, the result is in accordance with the prior possibilities, as intuitively expected. Conditional *a posteriori* possibilities can be obtained by scaling the plots in Figure 6 to a maximum height equal to 1: $D = (0.9, 0, 0)$ is fully possible, but the posterior possibility of $D = (0, 0.9, 0)$ is 0.31, and that of $D = (0, 0, 0.9)$ is 0.22.

[6] This is parallel to Bayesian approaches avoiding zero-probability events, in order to compute meaningful conditional probabilities in any situation.

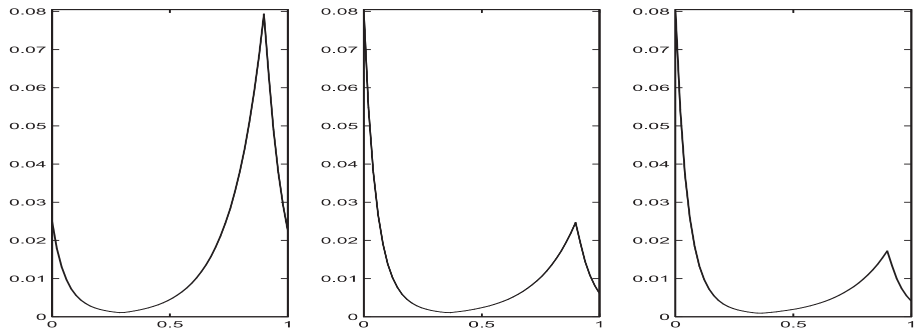

Fig. 6. Marginals in case 1. Abscissae: severity (from left to right: d_1, d_2, d_3); Ordinate: possibility.

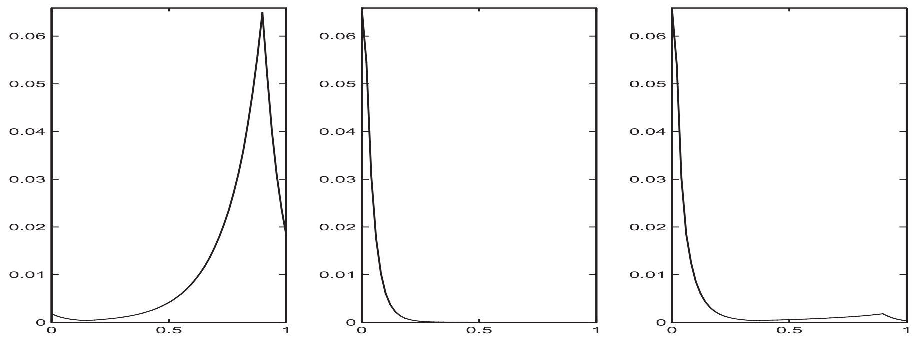

Fig. 7. Marginals in case 2.

Case 2: Consider now measuring almost full presence of m_1 and almost full absence of m_4, i.e., $\sigma_1 = 0.9$, $\sigma_4 = 0.02$. In this case, the most likely diagnosis is the same as above: $d_1 = 0.9$, $d_2 = d_3 = d_4 = 0$, with possibility 0.066. Figure 7 illustrates the obtained marginals, which show that now d_2 is basically discarded by the new measurement σ_4. There is, however, a situation with $d_3 = 0.9$ being the cause (because d_3 in some rare cases matches the sensor pattern) with possibility 0.002. The *a posteriori* possibility of $D = (0.9, 0, 0)$ is one, that of $D = (0, 0, 0.9)$ is 0.03.

Case 3: Consider now $\sigma_1 = 0.8$, $\sigma_2 = 0.77$, $\sigma_3 = 0.05$. This is an "almost impossible" situation (the peak possibility is 0.001). Indeed, no disorder in the knowledge base seems to be able to provide such a pattern: m_2 and m_3 should take the same value for all encoded disorders.

The low possibility of the present measurement combination should be used as a warning of an unexpected situation. However, inference concludes that $d_2 = 0.77$ is the most possible diagnosis, jointly with a fault in sensor 3 reading 0.34 units less than expected (note that m_2^3 is partially uncertain, but $d_2 = 0.77$ should produce at least $m_3 \geq 0.385$, which is not in agreement with the measured value). The same pattern might have been produced by less possible situations, such as $d_1 = 0.77$ and larger sensor 3 fault, as the marginal for d_1 shows in Figure 8.

Note how, transparently, sensor faults are integrated into the framework as well as *warning* that the pattern at hand is quite a "rare" one: particular attention or outlier

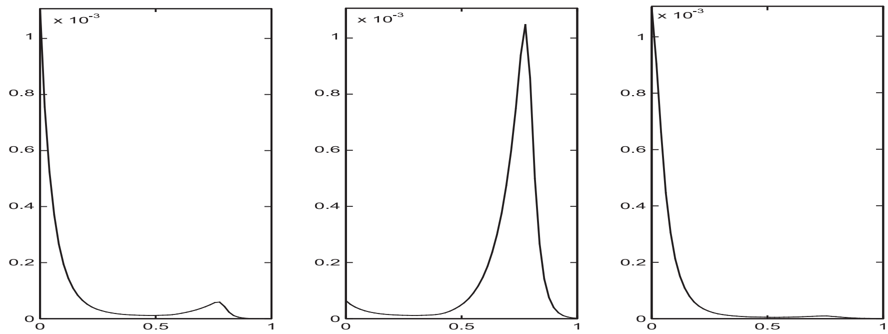

Fig. 8. Marginals in case 3.

handling may be considered [7].

A realistic example. Part of the procedures presented in this work have been applied to encode expert knowledge in a diesel engine diagnosis application, via oil analysis (improving over preliminary attempts reported in [26,27]). The task involved 10 disorders (combustion problems, water contamination, wear, injection problems ...) and 15 measurements (metallic elements, soot, glycol, oil viscosity and age...). There were around 60 decision variables and another 60 linear constraints.

Given the analysis results of an oil sample, a pure linear programming approach in Mathematica obtained the most possible diagnosis in around 0.2 seconds. A marginal possibility plot (which requires solving a linear programming problem for each of the plotted points), with a step of 0.05, took under 3 seconds on a single-core Pentium M laptop. A detailed description of such an example appears in [38]. The experts' (mechanical engineers) opinion on the developed procedure was favorable: they agreed that the proposed approach was more flexible and closer to their understanding of the problem than earlier naive attempts to compile a set of

[7] There are several alternatives for such outlier cases:

- Considering that σ_3 is conveying partially correct information about s_3 being "low", hence accepting disorder 2 (0.77) as the diagnostic.
- Considering that σ_3 is a totally unreliable measurement and, hence, presenting the diagnostic results inferred from only σ_1 and σ_2, removing $\{\sigma_3 = 0.05\}$ from the fact base (which would pinpoint d_1 as the most relevant disorder).
- Obtaining either σ_3 again (revising the measurement procedures in order to improve its reliability) or a different related manifestation which could provide additional information,
- Considering σ_3 as a reliable piece of data and, hence, concluding an unknown disorder (for instance, if the repeated measurement confirms the unexpected value).

Of course, in a practical case, the choice of the most suitable alternative will depend on the particular application and the expected failure mode of the sensor.

"if-symptoms-then-disorder" fuzzy diagnosis rules [27,39]. The previous approach seemed to be an easy task (some binary-logic rules were developed in [26]), but ended up with lots of exceptions, rules for cases where a measurement was missing, etc. Instead, the optimization approach resulted in better performance and a more elegant interpretation.

9 Comparative discussion

The equivalence between inequalities and implications, and between additions and conjunctions has been previously discussed, indicating that the binary approach may be handled by means of logic inference. Regarding the methodology for fuzzy and uncertain knowledge, other comparisons may be relevant, as discussed below.

Comparison to fuzzy relational approaches. Fuzzy logic is used in [40,46] in order to establish a "fuzzy relation" R between disorders and manifestations as the basic knowledge representation. In particular, the incorporation of both fuzziness (gradualness) and uncertainty would require defining uncertain fuzzy relations, such as those in [10]. Compositional operators are used to carry out diagnosis. However, the meaning of such fuzzy relations and the choice of composition operators is not clear right away; quoting [15], "*the literature on fuzzy relational equations for diagnosis usually lacks a concern for these representational issues*". Considering this work, the possibility distributions implicitly described via the constraints and the cost on the decision variables may be interpreted as a sort of uncertain fuzzy relation. The contribution of this work is illustrating that a particular type of such disorder-manifestation-measurement relations may be expressed via linear constraints and linear or quadratic cost indices, while keeping both a reasonably useful *expressive power* (correlations, multiple disorders, abduction, sensor voting with outlier measurements) and an *understandable* semantics. The key step is the addition of auxiliary variables.

Comparison to abductive causal reasoning. Causal reasoning is based on the definitions of the manifestations caused ($M(d_i)^+$) and not caused ($M(d_i)^-$) by a disorder, considered in 5. Then, to sort out between different alternative consistent diagnosis, some "abduction" criteria are used [8], particularly "possibilistic (*binary*) diagnosis under graded uncertainty" in [15]. This paper extends those results as it incorporates both graded uncertainty and graded severity. Furthermore, the enhancements discussed in sections 4, 5.1 and 7 allow for finer descriptions of the interrelations between disorders and manifestations than just the sets $M(d)^+$ and $M(d)^-$ originally used in [15], while keeping computational tractability.

Regarding comparison with [49], Yamada's approach also considers possibility as graded uncertainty of binary events, so the approach here may be considered more flexible and a generalisation to fuzzy events. Example 4 in the previous section showed how the translation of Yamada's knowledge bases into equations may be easily carried out by transforming possibilities lower than 1 into preference-related weights.

As a final remark, note the transparent integration of sensor error handling illustrated in example 15 may be considered advantageous with respect to other approaches in the literature.

Note also that the approach in this work interprets abduction as setting up a "prior" possibility of disorders, sensor behaviour, etc. and inferring from actual data a "posterior" possibility, somehow reminiscent of a Bayesian approach. However, there are other abduction possibilities (minimal cardinality, preference for single faults, preference for faults whose "certain" manifestation have been fully observed, *etc*. [8,28]) which have not been considered and could prove interesting.

Other approaches. The main objective in the paper is extending the abductive causal reasoning ideas above, hence, no detailed comparison has been made to other approaches such as [44], based on combinining fuzzy and probabilistic representation into a Dempster-Shafer framework. Briefly, in that paper, focal elements – in the form "if (fuzzy) manifestations then disorder" with a probability mass assignment– are combined into an overall diagnosis. Hence, the representation is the inverse to the one in this work –"if disorders then manifestations" with a possibility assignment–.

Computational issues. The main objective of the paper has been avoiding logic- or set-based approaches (even if interesting and fruitful, and possibly equivalent to many aspects of the proposed methodology) in favor of exploiting the algebraic equivalences in order to use standard optimisation software. The necessary software is widely available; actually, the knowledge and fact bases can be *literally* typed into a symbolic computer algebra program such as *Mathematica*®.

Linear and quadratic programming can solve optimisation problems with hundreds and thousands of decision variables in a reasonable time, hence the proposed methodology can escalate to complex diagnostic knowledge bases: one optimization step for a diesel engine problem with 10 faults and 15 measurements took less than 0.2 seconds (linear programming). In the author's opinion, current computers should be able to solve realistic problems with, say, a few dozen faults and measurements in a few seconds, and plot all relevant marginal plots in at most a couple of minutes. Note that the particular linear constraints in these problems may be amenable to sparse matrix representation, as only a handful of decision variables appear in each constraint.

A front-end application is being developed at this moment in order to ease editing of knowledge data for larger-scale cases. The application will provide a much more friendly user-interface experience than the present raw lists of equations and optimization software commands.

10 Conclusions

This paper has presented a methodology to encode knowledge-based diagnostic reasoning as an optimisation problem. Considering binary cases, the methodology seems to be able to incorporate the knowledge structures considered in consistency-based approaches as well as some possibilistic-based ones, at least approximately. Furthermore, this work incorporates the following interesting features: (a) the ability to seamlessly incorporate inequalities involving d_i, m_i, s_i^j which enable more expressivity than the referenced approaches, in some cases; (b) the ability to naturally handle multiple disorders, particularly sensor faults (outliers); (c) the simultaneous use of possibility (abduction) and gradualness (fuzziness), producing possibility distributions on the truth value (severity) of the disorders; (d) the ability to express the diagnostic knowledge as a linear or quadratic programming problem, for which efficient code exists. The availability of efficient code makes the methodology suitable for large-scale cases, in the author's opinion.

References

[1] C. Angeli. Online expert system for fault diagnosis in hydraulic systems. *Expert Systems*, 16:115–120, 1999.

[2] K. Atanasov. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 20:87–96, 1986.

[3] P. Bhowal, D. Sarkar, S. Mukhopadhyay, and A. Basu. Fault diagnosis in discrete time hybrid systems–A case study. *Information Sciences*, 177(5):1290–1308, 2007.

[4] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, editors. *Diagnosis and Fault-Tolerant Control*. Springer, London, 2003.

[5] Christian Borgelt and Rudolf Kruse. *Abductive inference with probabilistic networks*, pages 281–314. Kluwer, MA, USA, 2000.

[6] E.F. Carrasco and J. Rodríguez *et. al.* Diagnosis of acidification states in an anaerobic wastewater treatment plant using a fuzzy-based expert system. *Control Engineering Practice*, 12(1):59–64, 2004.

[7] E. Castillo, J.M. Gutierrez, and A.S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer, London, 1997.

[8] D. Cayrac. *Diagnostic opérational exploitant des connaissances incomplètes (in French)*. PhD thesis, Univ. Paul Sabatier, Toulouse, 1995.

[9] L.H. Chiang, E.L. Russell, and R.D. Braatz. *Fault Detection and Diagnosis in Industrial Systems*. Springer-Verlag, 2001.

[10] S.K. De, R. Biswas, and A.R. Roy. An application of intuitionistic fuzzy sets in medical diagnosis. *Fuzzy Sets and Systems*, 117(2):209–213, 2001.

[11] G. De Cooman. Possiblity theory II: conditional possibility. *International Journal of General Systems*, 25(4):325–351, 1997.

[12] G. Deschrijver. Arithmetic operators in interval-valued fuzzy set theory. *Information Sciences*, 177(14):2906–2924, 2007.

[13] D. Dubois, H. Fargier, and H. Prade. Possibility theory in constraint satisfaction problems: handling priority, preference and uncertainty. *Applied Inteligence*, 6(4):287–309, 1996.

[14] D. Dubois, M. Grabisch, O. De Mouzon, and H. Prade. A possibilistic framework for single-fault causal diagnosis under uncertainty. *Int. J. General Systems*, 30(2):167–192, 2001.

[15] D. Dubois and H. Prade. Fuzzy relation equations and causal reasoning. *Fuzzy Sets and Systems*, 45(2):119–134, 1995.

[16] D. Dubois and H. Prade. Possibility theory, probability theory and multiple-valued logics: a clarification. *Annals of Mathematics and Artificial Intelligence*, 32(1):35–66, 2001.

[17] Didier Dubois and Henri Prade. *Possibility Theory: an Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.

[18] S.I. Gass. *Linear Programming: methods and applications*. Dover, 5th edition, 2003.

[19] Z. Gharamani. An introduction to hidden markov models and bayesian networks. *Int. J. of Pattern Reconnition and Artif. Intell.*, 15(1):9–42, 2001.

[20] C.C. Hsu and C.S. Ho. A new hybrid case-based architecture for medical diagnosis. *Information Sciences*, 166(1-4):231–247, 2004.

[21] PR Innocent and RI John. Computer aided fuzzy medical diagnosis. *Information Sciences*, 162(2):81–104, 2004.

[22] R. Isermann. Model-based fault-detection and diagnosis–status and applications. *Annual Reviews in Control*, 29(1):71–85, 2005.

[23] E. Kilic. Diagnosability of fuzzy discrete event systems. *Information Sciences*, 178:858–870, 2008.

[24] J. Kitowski and M. Bargiel. Diagnosis of faulty states in complex physical systems using fuzzy relational equations. In E. Sanchez and L.A. Zadeh, editors, *Approx. Reasoning in Intelligent Systems, Decision and Control*, pages 175–194. Pergamon, 1987.

[25] R. Kruse, E. Schwecke, and J. Heinsohn, editors. *Uncertainty and vagueness in Knowledge Based Systems: Numerical Methods (Artificial Intelligence)*. Springer-Verlag, 1991.

[26] V. Macián, B. Tormos, and Mª.J. Lerma. Knowledge based systems for predictive maintenance of diesel engines. In *Proc. Euromaintenance Conf.*, volume 1, pages 49–54. Swedish Maintenance Society-ENFMS, 2000.

[27] V. Macián, B. Tormos, A. Sala, and J. Ramirez. Fuzzy logic-based expert system for diesel engine oil analysis diagnosis. *Insight (BINDT)*, 48(8):462–469, 2006.

[28] O. Mouzon, D. Dubois, and H. Prade. Using consistency and abduction based indices in possibilistic causal diagnosis. In *Proc. IEEE Int. Fuzzy Systems Conf.*, pages 729–734, 2000.

[29] O. Mouzon, D. Dubois, and H. Prade. Twofold fuzzy sets in single and multiple fault diagnosis, using information about normal values. In *Proc. IEEE Int. Fuzzy Systems Conf.*, pages 1563–1566, 2001.

[30] E. Németh, R. Lakner, KM Hangos, and IT Cameron. Prediction-based diagnosis and loss prevention using qualitative multi-scale models. *Information Sciences*, 177(8):1916–1930, 2007.

[31] Y. Peng and J.A. Reggia. *Abductive inference models for diagnostic problem-solving*. Springer, 1990.

[32] A. Roychowdhury, D.K. Pratihar, N. Bose, KP Sankaranarayanan, and N. Sudhahar. Diagnosis of the diseases—-using a GA-fuzzy approach. *Information Sciences*, 162(2):105–120, 2004.

[33] S. Russell and P. Norvig. *Artificial Intelligence: a modern approach*. Prentice-Hall, 2nd edition, 2003.

[34] A. Sala and P. Albertos. Fuzzy systems evaluation: The inference error approach. *IEEE Trans. on Syst. Man & Cybernetics*, 28B(2):268–275, 1998.

[35] A. Sala and P. Albertos. Inference error minimisation: Fuzzy modelling of ambiguous functions. *Fuzzy Sets and Systems*, 121(1):95–111, 2001.

[36] A. Sala and C. Ariño. Asymptotically necessary and sufficient conditions for stability and performance in fuzzy control: Applications of Polya's theorem. *Fuzzy Sets and Systems*, 158(24):2671–2686, 2007.

[37] A. Sala, T.M. Guerra, and R. Babuska. Perspectives of fuzzy systems and control. *Fuzzy sets and systems*, 156(3):432–444, 2005.

[38] A. Sala, J.C. Ramirez, B. Tormos, and M. Yago. An optimization approach to fuzzy diagnosis: Oil analysis application. In *Proceedings of IEEE Int. Conf. on Fuzzy Systems FUZZ-IEEE'07*, London, 2007. IEEE.

[39] A. Sala, B. Tormos, V. Macian, and E. Royo. Fuzzy diagnosis module based on interval fuzzy logic: Oil analysis application. In *Informatics in Control, Automation and Robotics II*. Springer, 2007.

[40] E. Sánchez. Inverses of fuzzy relations. Application to possibility distributions and medical diagnosis. *Fuzzy Sets and Systems*, 2:75–86, 1979.

[41] L.E. Schrage. *linear, integer and quadratic programming with LINDO*. Course Technology, 1998.

[42] G. Shafer and J. Pearl, editors. *Readings in uncertain reasoning*. Morgan Kauffman, San Mateo (CA), USA, 1990.

[43] G. Sierksma. *Linear and Integer Programming: Theory and Practice*. Marcel Dekker Pub., New York, 2nd edition, 2001.

[44] E. Straszecka. Combining uncertainty and imprecision in models of medical diagnosis. *Information Sciences*, 176(20):3026–3059, 2006.

[45] K. Tanaka and H.O. Wang. *Fuzzy Control Systems Design and Analysis: A Linear Matrix Inequality Approach*. John Wiley & Sons, Inc. New York, NY, USA, 2001.

[46] Y. Tsukamoto and T. Terano. Failure diagnosis by using fuzzy logic. In *Proc. of IEEE Control and Decision Conf.*, pages 1390–1395, 1977.

[47] J. Wainer and A. Rezende. A temporal extension to the parsimonious covering theory. *Artificial Intelligence in Medicine*, 10:235–255, 1997.

[48] Y.M. Wang, J.B. Yang, D.L. Xu, and K.S. Chin. On the combination and normalization of interval-valued belief structures. *Information Sciences*, 177(5):1230–1247, 2007.

[49] K. Yamada. Possibilistic causality consistency problem based on asymmetrically-valued causal model. *Fuzzy Sets and Systems*, 132:33–48, 2002.