

UDC 004.822

QUERY ANSWERING OVER FACT BASES IN ZADEH LOGIC

(This work is partially supported by RFBR grant № 12-01-00705)

Plesniewicz G.S., Ph.D. (Tech.Sc.), professor

(National Research University "MPEI", Krasnokazarmennaya St., 14, Moscow, 111250, Russian Federation, salve777@mail.ru)

Аннотация. Let \mathbf{L} be Zadeh logic i.e. the fuzzy propositional logic based on triangular norm $\min(x, y)$. A fact in \mathbf{L} is an expression of the form $r \leq \varphi \leq s$ where $\varphi \in \mathbf{L}$ and $0 \leq r \leq s \leq 1$. In a fuzzy interpretation I of \mathbf{L} every fact is true or false, and $I(r \leq \varphi \leq s) = 1$ if and only if the two-side inequality $r \leq I(\varphi) \leq s$ is satisfied. Thus, the set \mathbf{FL} of all facts in \mathbf{L} defines a crisp logic with fuzzy interpretations. Logical consequence " \models " in the logic \mathbf{FL} is defined as usual: for any set E of facts and any fact α , $E \models \alpha$ if there are no an interpretation I such that $I(\alpha) = 1$ and $I(\beta) = 0$ for all $\beta \in E$. But in the logic \mathbf{FL} there is also strong logical consequence \models^* : $E \models^* r \leq \varphi \leq s$ if $E \models r \leq \varphi \leq s$ and it is not true that $E \models r' \leq \varphi \leq s$ with $r' > r$ and not true $E \models r \leq \varphi \leq s'$ with $s' < s$.

A fact base is a finite set F of facts: $F = \{r_i \leq \varphi_i \leq s_i / 1 \leq i \leq n\}$. One can consider the set $K = \{\varphi_i / 1 \leq i \leq n\}$ as a fuzzy knowledge base and F as an instance of K . A query is an expression of the form $?\psi$ where $\psi \in \mathbf{L}$. The answer to the query to the fact base F is the fact $r \leq \psi \leq s$ such that $F \models^* r \leq \psi \leq s$.

The problem of query answering over fact bases in \mathbf{FL} can be solved by analytical tableaux method. The method results in an algorithm with the exponential worst-case estimate (relatively to the size of $F \cup \{\kappa\}$ where κ is a query). However, consider the situation when the knowledge base K and the query κ are fixed but fact bases F are arbitrary instances of K . Then, it is possible to answer query κ to fact bases F quickly. But preliminary we should deal with the parametric fact base associated with K .

Under a parametric fact we mean an expression of the form $a \leq \varphi \leq b$ where a and b are not numbers but parameters – variables with values in $[0, 1]$. The parametric fact base associated with K is $P = \{a_i \leq \varphi_i \leq b_i / 1 \leq i \leq n\}$ where a_i, b_i are different parameters. Thus, if we replace the parameters by specific numbers from $[0, 1]$ (with adherence to corresponding inequality) we obtain a specific fact base which is an instance of the parametric fact base P .

One can also consider query answering over parametric fact bases. Let P be a parametric fact base and $?\psi$ be a query. The answer to the query is the expression $g \leq \psi \leq h$ such that $K\lambda \models^* g\lambda \leq \psi \leq h\lambda$ for any substitution λ of numbers from $[0, 1]$ for the parameters from K . Here g and h are appropriate expressions with the parameters from K .

Using the analytical tableaux method, we show how to design an algorithm for finding the expressions g and h for a given knowledge base K . So, let a knowledge base $K = \{\varphi_i / 1 \leq i \leq n\}$ and a query $\kappa: ?\psi$ be fixed. Suppose we need to answer the query to any fact base $F = \{r_i \leq \varphi_i \leq s_i / 1 \leq i \leq n\}$. Then we (1) apply the algorithm to the parametric fact base $P = \{a_i \leq \varphi_i \leq b_i / 1 \leq i \leq n\}$ and obtain the expressions g and h ; (2) apply the substitution $\lambda = \{a_i / r_i, b_i / s_i / 1 \leq i \leq n\}$ to g and h ; thus, we obtain the answer $g\lambda \leq \psi \leq h\lambda$.

Ключевые слова: query, fact base, Zadeh logic, knowledge base, contrary condition, inequalities.

Introduction. Main definitions

Let \mathbf{L} be Zadeh logic i.e. the propositional fuzzy logic based on a triangular norm $\min\{x, y\}$. The syntax of \mathbf{L} is the same as the syntax of usual propositional logic. The semantics of \mathbf{L} is defined by interpretations $I: \mathbf{L} \rightarrow [0, 1]$ satisfying the following conditions for any formulas $\varphi, \psi \in \mathbf{L}$:

$$I(\neg \varphi) = 1 - I(\varphi), I(\varphi \wedge \psi) = \min\{I(\varphi), I(\psi)\},$$

$$I(\varphi \vee \psi) = \max\{1 - I(\varphi), 1 - I(\psi)\},$$

$$I(\varphi \rightarrow \psi) = \max\{1 - I(\varphi), I(\psi)\}.$$

We associate with each formula $\varphi \in \mathbf{L}$ and numbers r, s ($0 \leq r \leq s \leq 1$) the sentence (fact) $r \leq \varphi \leq s$ which is true or false in I , and $I(r \leq \varphi \leq s) = 1 \Leftrightarrow_{\text{df}} r \leq I(\varphi) \leq s$. (We also write $\varphi \geq r$ instead of $r \leq \varphi \leq 1$ and $\varphi \leq r$ instead of $0 \leq \varphi \leq r$.) A fact base F is a finite set of facts.

Let \mathbf{FL} denote the set of all facts for \mathbf{L} . Thus, \mathbf{FL} is a crisp logic with fuzzy interpretations. As in any logic, there is the logical consequence relation \models in \mathbf{FL} : for any $E \subseteq \mathbf{FL}$ and $\alpha \in \mathbf{FL}$

$$E \models \alpha \Leftrightarrow_{\text{df}} \text{there is no interpretation } I \text{ such that } I(\alpha) = 0 \text{ and } I(\beta) = 1 \text{ for all } \beta \in E.$$

But in \mathbf{FL} there is also strong logical consequence \models^* :

$$E \models^* r \leq \varphi \leq s \Leftrightarrow_{\text{df}} E \models r \leq \varphi \leq s \text{ and it is not true that } E \models r' \leq \varphi \leq s \text{ with } r' > r \text{ and } E \models r \leq \varphi \leq s' \text{ with } s' < s.$$

A *knowledge base* is a finite set K of formulas from \mathbf{L} : $K = \{\varphi_i / 1 \leq i \leq n\}$. Let us choose numbers $0 \leq r_i \leq s_i \leq 1$ ($1 \leq i \leq n$); then the fact base $F = \{r_i \leq \varphi_i \leq s_i / 1 \leq i \leq n\}$ is an *instance* of the knowledge base K . An expression of the form $? \psi$ is a *query* to the fact bases – instances of K if ψ is a formula of \mathbf{L} in the signature of K . The *answer* to a query $? \psi$ to a fact base F is the fact $r \leq \psi \leq s$ such that $F \models^* r \leq \psi \leq s$.

Example 1. Let us consider the knowledge base $K = \{p \wedge q, q \rightarrow r\}$ and its instance (the fact base) $F = \{0.7 \leq p \wedge q, 0.4 \leq q \rightarrow r \leq 0.6\}$. Let $\kappa = ?p \wedge \neg r$ be the query to F . Then $0.4 \leq p \wedge \neg r \leq 0.6$ is the answer to κ .

Under a *parametric fact* we mean an expression of the form $a \leq \varphi \leq b$ where a and b are not numbers but parameters – variables with values in $[0, 1]$. A *parametric fact base* P for a knowledge base $K = \{\varphi_i / 1 \leq i \leq n\}$ is the set of parametric facts with different parameters: $P = \{a_i \leq \varphi_i \leq b_i / 1 \leq i \leq n\}$ ($a_i \neq b_j$ if $i \neq j$).

Let λ be a substitution numbers for parameters: $\lambda = \{r_i / a_i, s_i / b_i / 1 \leq i \leq n\}$ ($r_i \leq s_i$). Then, applying λ to P we obtain the fact base $P\lambda = \{r_i \leq \varphi_i \leq s_i / 1 \leq i \leq n\}$. One can put a query $? \psi$ to the parametric fact base P for a knowledge base K . Then the answer to this query is the expression $g \leq \psi \leq h$ such that $K\lambda \models^* g\lambda \leq \psi \leq h\lambda$ for any substitution λ where g and h are some expressions containing parameters from K .

Example 2. Let the knowledge base K and the query κ be the same as in Example 1. The parametric fact base for the knowledge base is $P = \{a \leq p \wedge q \leq b, c \leq q \rightarrow r \leq d\}$. Then $g \leq p \wedge \neg r \leq h$ is the answer to κ where

$$g = \min\{a, 1 - d\},$$

$$h = \text{case}[\max\{b, 1 - c\} \text{ if } a + c > 1 \wedge b + d > 1,$$

$$b \text{ if } a + c \leq 1 \wedge b + d > 1,$$

$$1 - c \text{ if } a + c > 1 \wedge b + d \leq 1,$$

$$0 \text{ if } a + c \leq 1 \wedge b + d \leq 1].$$

Query answering to fact bases

Let \mathbf{M} be the set of all sentences from \mathbf{FL} of the forms $\alpha \leq c$, $\alpha < c$, $\alpha \geq c$ and $\alpha > c$. Obviously, the problem of logical consequence for logic \mathbf{FL} is reduced to the problem of inconsistency for logic \mathbf{M} since

$$F \models r \leq \alpha \leq s \Leftrightarrow F^+ \cup \{\alpha > r\} \text{ and } F^+ \cup \{r < s\} \text{ are inconsistent sets}$$

where $F^+ = \{\beta \leq s, \beta \geq r \mid (r \leq \beta \leq s) \in F\}$.

The method of analytical tableaux can be applied to solve the problem of inconsistency in \mathbf{M} [2]. In Table 1 there are the inference rules for the logic \mathbf{M} . This method also can be applied to the problem of finding answers to queries to fact bases. We show by example, how to do it.

Table 1

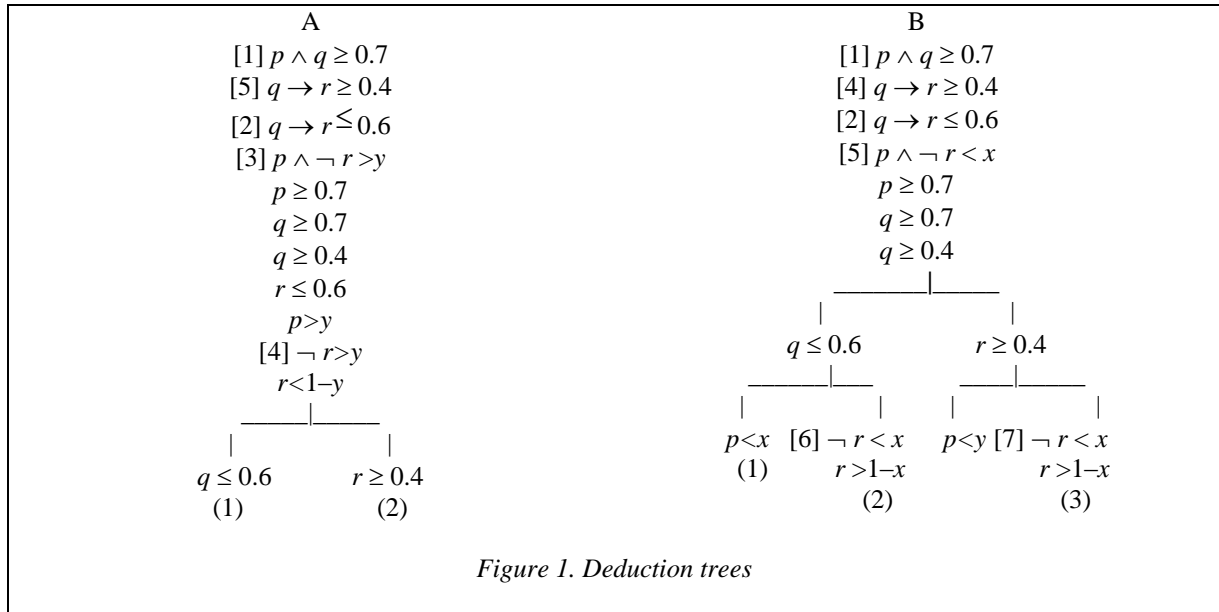
$\frac{\neg \varphi \geq c \quad \neg \varphi \leq c \quad \neg \varphi > c \quad \neg \varphi < c}{\varphi \leq 1 - c \quad \varphi \geq 1 - c \quad \varphi < 1 - c \quad \varphi > 1 - c}$
$\frac{\varphi \wedge \psi \geq c \quad \varphi \wedge \psi \leq c \quad \varphi \wedge \psi > c \quad \varphi \wedge \psi < c}{\varphi \geq c \quad \varphi \leq c \mid \psi \leq c \quad \varphi > c \quad \varphi < c \mid \psi < c}$ $\psi \geq c \quad \psi > c$
$\frac{\varphi \vee \psi \geq c \quad \varphi \vee \psi \leq c \quad \varphi \vee \psi > c \quad \varphi \vee \psi < c}{\varphi \geq c \mid \psi \geq c \quad \varphi \leq c \quad \varphi > c \mid \psi > c \quad \varphi < c}$ $\psi \leq c \quad \psi < c$
$\frac{\varphi \rightarrow \psi \geq c \quad \varphi \rightarrow \psi \leq c \quad \varphi \rightarrow \psi > c \quad \varphi \rightarrow \psi < c}{\varphi \leq 1 - c \mid \psi \geq c \quad \varphi \geq 1 - c \quad \varphi < 1 - c \mid \psi > c \quad \varphi > 1 - c}$ $\psi \leq c \quad \psi < c$

Example 3. Let the knowledge base K , the fact base F and the query κ be such as in Example 1. The fact base F can be replaced with the equivalent fact base $F^+ = \{p \wedge q \geq 0.7, q \rightarrow r \geq 0.4, q \rightarrow r \leq 0.6\}$. Then $F^+ \models x \leq p \wedge \neg r \leq y$ if and only if the following two sets are inconsistent:

$$F_1 = \{p \wedge q \geq 0.7, q \rightarrow r \geq 0.4, q \rightarrow r \leq 0.6, p \wedge \neg r \leq x\},$$

$$F_2 = \{p \wedge q \geq 0.7, q \rightarrow r \geq 0.4, q \rightarrow r \leq 0.6, p \wedge \neg r > y\}.$$

In Figure 1 the deduction trees A and B for these sets are shown. In A the branch (1) is closed since it contains contrary inequalities $q \geq 0.7$ and $q \leq 0.6$. The branch (2) will be closed if we choose y such that inequalities $r < 1-y$ and $r \geq 0.4$ becomes contrary. Also (2) will be closed if $y=1$ (then $p > 1$ and that is impossible). Hence, (2) is closed if and only if $y=1$ or $0.4 \geq 1-y$, i.e. $y \geq 0.6$. Therefore, $h = \min y = 0.6$. In B branches (1) and (2) are closed since they contain contrary inequalities $q \geq 0.7$ and $q \leq 0.6$. Clearly, branch (3) is closed if and only if $1-x = 1$, i.e. $x = 0$. Therefore, $g = 0$. Hence, $p \wedge \neg r \leq 0.6$ is the answer to κ .



Remark. The similar method has been offered in [1].

Query answering for parametric fact bases

One can find answers to queries to parametric fact bases by applying analytical tableaux method. Here is an example (how to do it).

Example 4. Let the knowledge base F , the query κ and the parametric fact base P are such as in Example 2. In Figure 2 and Figure 3 the deduction trees for the sets $P_1 = P \cup \{p \wedge \neg r > y\}$ and $P_2 = P \cup \{p \wedge \neg r < x\}$ are shown.

In the first tree, consider two inequalities $q \geq a$ and $q \leq 1-c$ which lie on branch (1). Clearly, they are contrary (inconsistent) if and only if $a > 1-c$ i.e. $a+c > 1$. We say that $(q \geq a, q \leq 1-c)$ is a *candidate contrary pair* and $a+c > 1$ is a *condition of its contrariness*.

In Table 2 there are all the candidate contrary pairs together with the contrariness conditions and with references to the branches closed by the contrary pairs. From the table we see that pairs 1 and 4 (and also pairs 2 and 3) block up all branches of the tree. Therefore, the tree in Figure 2 is closed if and only if the following condition is satisfied:

$$(a + c > 1 \wedge y \geq 1 - c) \vee (b + d > 1 \wedge y \geq b) \tag{1}$$

Table 2

1	$(q \leq 1 - c, q \geq a)$	$a + c > 1$	(1), (3)
2	$(q \leq b, q \geq 1 - d)$	$b + d > 1$	(3), (4)
3	$(p \leq b, p > y)$	$y \geq b$	(1), (2)
4	$(q \leq 1 - c, q \geq a)$	$y \geq 1 - c$	(2), (4)

So, $h = \min\{y \mid y \text{ satisfies (1)}\}$. Depending on what conditions $a+c > 1$ and $b+d > 1$ are true or false the condition (2.1) is reduced to:

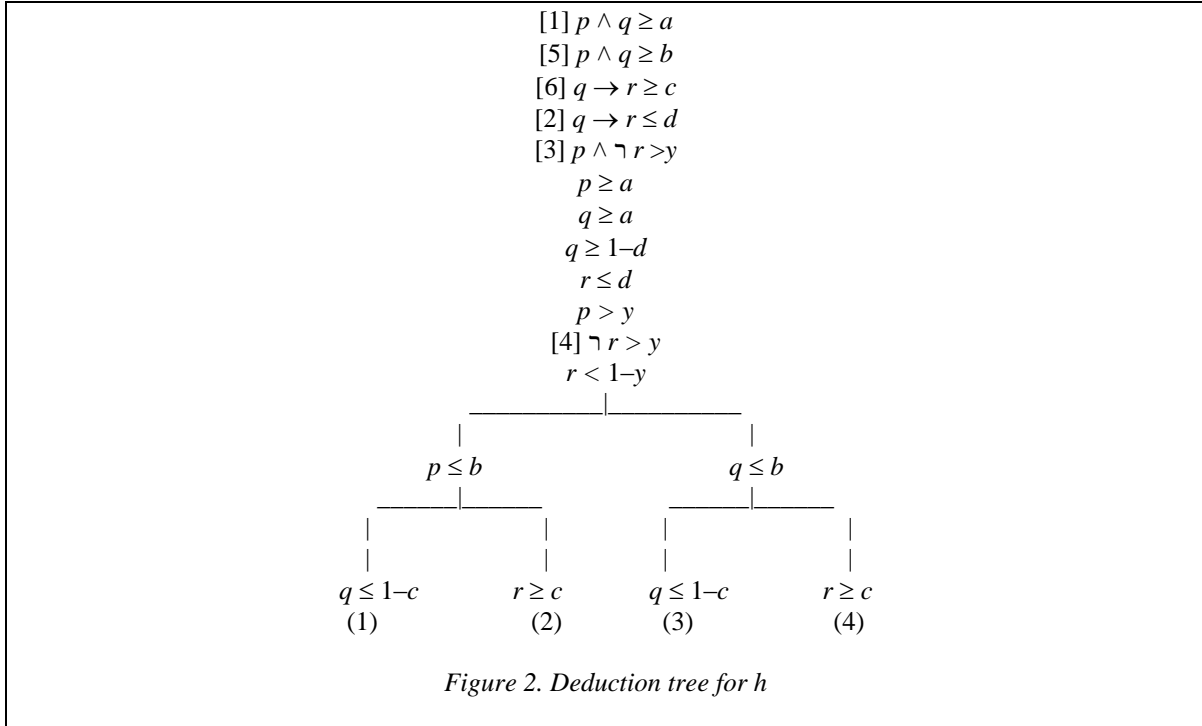
$$y \geq 1 - c \vee y \geq b \quad \text{if } a + c > 1 \text{ and } b + d > 1,$$

$$y \geq b \quad \text{if } a + c \leq 1 \text{ and } b + d > 1,$$

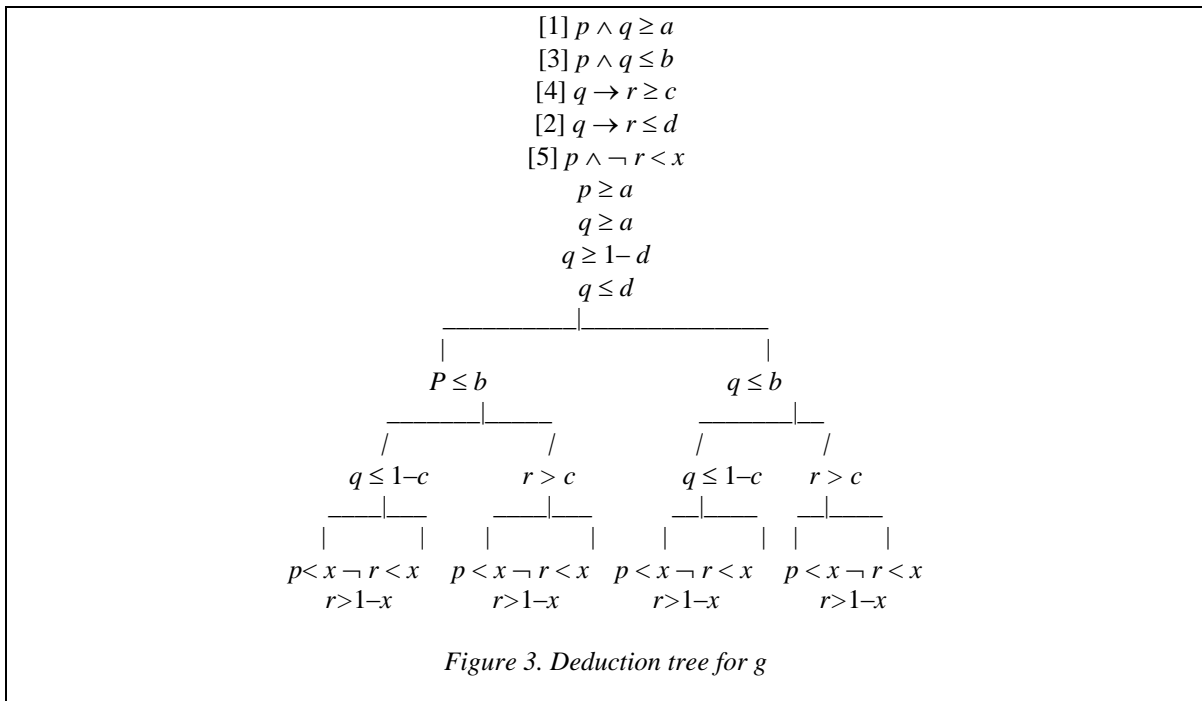
$y \geq 1 - c$ if $a + c > 1$ and $b + d \leq 1$,
 0 if $a + c \leq 1$ and $b + d \leq 1$.

From this we obtain

$h = \text{case}[\max\{b, 1 - c\}$ if $a + c > 1 \wedge b + d > 1$,
 b if $a + c \leq 1 \wedge b + d > 1$,
 $1 - c$ if $a + c > 1$ and $b + d \leq 1$,
 0 if $a + c \leq 1$ and $b + d \leq 1]$.



In a similar way, considering the tree in Figure 3, we find:
 $g = \min\{a, 1 - d\}$.



Example 5. Let the knowledge base K , the parametric fact base P , the query κ , and the expressions g and h be such as in Example 4. Let us take the substitution $\lambda = \{a / 0.7, b / 1, c / 0.4, d / 0.6\}$. Applying λ to P , g and h , we obtain

$$P\lambda = F = \{0.7 \leq p \wedge q, 0.4 \leq q \rightarrow r \leq 0.6\}, g\lambda = 0.4, h\lambda = 0.4.$$

Thus, we have the answer $0.4 \leq p \wedge \neg r \leq 0.6$ to the query $p \wedge \neg r$ to the fact base F .

Let us consider a general situation when a knowledge base K , a query κ and a parametric fact base P for K are arbitrary. Let P have parameters a_i ($i = 1, 2, \dots$). Suppose we construct the deduction trees T_1 and T_2 for the sets $P \cup \{\kappa < x\}$ and $P \cup \{\kappa > y\}$. It is easy to see that in T_1 there can be candidate pairs and contrary conditions of the following forms:

$$\begin{array}{ll} (p \geq a_j, p \leq a_k), & a_j > a_k, \\ (p \geq 1 - a_j, p \leq a_k), & a_j + a_k < 1, \\ (p \geq a_j, p \leq 1 - a_k), & a_j + a_k > 1, \\ (p \geq a_j, p < x), & x \leq a_j, \\ (p \geq 1 - a_j, p < x), & x \leq 1 - a_j, \\ (p > 1 - x, p \leq a_j), & x \leq 1 - a_j, \\ (p > 1 - x, p \leq 1 - a_j), & x \leq a_j, \\ (p > 1 - x, p > x), & x \leq 1/2. \end{array}$$

In the first tree, consider two inequalities $q \geq a$ and $q \leq 1 - c$ which lie on branch (1). Clearly, they are contrary (inconsistent) if and only if $a > 1 - c$ i.e.

Every candidate pair in T_1 blocks some branches. Let $b(\pi)$ denote the set of branches which is blocked up by pair π , and let $c(\pi)$ denote the contrary condition for pair π . Also, let $b(S) = \cup\{b(\pi) \mid \pi \in E\}$ and $c(S) = \cap\{c(\pi) \mid \pi \in S\}$ where S is a set of candidate pairs.

A set S of candidate pairs is a *covering* if $b(S)$ coincides with the set of all branches of T_1 . Thus, if S is a covering and the condition $c(S)$ is satisfied with a given substitution σ then the tree $T_1\sigma$ is closed. A covering S is (locally) *minimal* if $S \setminus \{\pi\}$ is not a covering for each $\pi \in S$.

Let S_1, S_2, \dots, S_m be all minimal coverings for T_1 . Take the condition $C = c(S_1) \vee c(S_2) \vee \dots \vee c(S_m)$. Thus, C is a disjunction of conjunctions made of inequalities of the form: $a_j > a_k, a_j + a_k < 1, a_j + a_k > 1, x \leq a_j, x \leq 1 - a_j, x \leq 1/2$.

Let R be the set of all conditions which are occurred in C and have no the variable x . Let θ be any assignment of truth values 0 or 1 to the conditions from R , i.e. $\theta: R \rightarrow \{0, 1\}$. One can consider θ as a substitution truth values for inequalities. Thus, $C\theta$ has the form $C^0_1 \vee C^0_2 \vee \dots \vee C^0_m$ where $C^0_i = c(S_i)\theta$ and C^0_i has the form $(x \leq e_{i1}) \wedge (x \leq e_{i2}) \wedge \dots \wedge (x \leq e_{iq(i)})$. Let us denote $r(C^0_i) = \{e_{i1}, e_{i2}, \dots, e_{iq(i)}\}$.

It is clear that:

$$x \text{ satisfies } C^0_i \Leftrightarrow (x \leq e_{i1}) \wedge (x \leq e_{i2}) \wedge \dots \wedge (x \leq e_{iq(i)})$$

$$\Leftrightarrow x \leq \min\{e_{i1}, e_{i2}, \dots, e_{iq(i)}\}$$

$$\Leftrightarrow x \leq \min r(C^0_i);$$

$$x \text{ satisfies } C\theta \Leftrightarrow x \text{ satisfies } C^0_1 \vee C^0_2 \vee \dots \vee C^0_m$$

$$\Leftrightarrow (x \leq \min r(C^0_1)) \vee (x \leq \min r(C^0_2)) \vee \dots \vee (x \leq \min r(C^0_m))$$

$$\Leftrightarrow x \leq \max\{\min r(C^0_1), \min r(C^0_2), \dots, \min r(C^0_m)\}$$

$$\Leftrightarrow x \leq \max\{\min r(C^0_j) \mid 1 \leq j \leq m\} \quad (2)$$

Let θ^* be the conjunction of the inequalities from R or their negations. We include in θ^* an inequality if θ assigns 1, and the contrary inequality if θ assigns 0, i.e.

$$\theta^\# = (\bigwedge\{\sigma \mid \sigma \in R, \theta(\sigma) = 1\}) \wedge (\bigwedge\{\neg\sigma \mid \sigma \in K, \theta(\sigma) = 0\}).$$

We have

$$C = \bigwedge\{\theta^\# \rightarrow C\theta \mid \theta: R \rightarrow [0, 1]\} \quad (3)$$

This can be understood considering the following example.

Example 6. Let α be a formula of propositional variables p, q and r : $\alpha = \alpha[p, q, r]$. Then

$$\alpha[p, q, r] = (p \wedge q \rightarrow \alpha[1, 1, r]) \wedge (p \wedge \neg q \rightarrow \alpha[1, 0, r]) \wedge$$

$$(\neg p \wedge q \rightarrow \alpha[0, 1, r]) \wedge (\neg p \wedge \neg q \rightarrow \alpha[0, 0, r]).$$

Indeed, for example, if $p = 0, q = 1$ then in the right part of this equality we have $\alpha[0, 1, r]$. Hence, the equality is true for $p = 0, q = 1$.

From (2) and (3) we obtain

$$g = \max\{x \mid x \text{ satisfies } C\}$$

$$= \text{case}[\max\{\min r(C^0_j) \mid 1 \leq j \leq m\} \text{ if } \theta^\# \mid \theta: R \rightarrow [0, 1]].$$

In the similar way we obtain the expression for h (see Table 3).

Table 3

$g = \text{case}[\max\{\min r(C^0_j) \mid 1 \leq j \leq m\} \text{ if } \theta^\# \mid \theta: R \rightarrow [0, 1]]$

$$h = \text{case}[\max\{\min r(C_j^0) \mid 1 \leq j \leq m\} \text{ if } \theta^\# \mid \theta: R \rightarrow [0, 1]]$$

Example 7. In the tree T_2 (Figure 3) there are the candidate pairs which are written in Table 3. From here $R = \{a+c>1, b+d>1\}$. There are exactly four substitutions

$$\theta_1 = \{1/(a+c>1), 1/(b+d>1)\}, \theta_2 = \{1/(a+c>1), 0/(b+d>1)\}, \\ \theta_3 = \{0/(a+c>1), 1/(b+d>1)\}, \theta_4 = \{0/(a+c>1), 0/(b+d>1)\}.$$

Then we have

$$\theta_1^\# = (a+c>1) \wedge (b+d>1), \theta_2^\# = (a+c>1) \wedge (b+d \leq 1), \\ \theta_3^\# = (a+c \leq 1) \wedge (b+d>1), \theta_4^\# = (a+c \leq 1) \wedge (b+d \leq 1).$$

For the condition

$$C = (a+c>1 \wedge y \geq 1-c) \vee (b+d>1 \wedge y \geq b)$$

we have

$$C\theta_1 = ((y \geq 1-c) \vee (y \geq b)), C\theta_2 = (y \geq 1-c), \\ C\theta_3 = (y \geq b), C\theta_4 = 0.$$

Hence,

$$r(C\theta_1) = \{1-c, b\}, r(C\theta_2) = \{1-c\}, r(C\theta_3) = \{b\}, r(C\theta_4) = \{\}.$$

Using the formula presented in the second row of Table 3 we obtain the expression g coinciding with what is presented in Example 2.

References

1. Straccia U. Reasoning and Experimenting within Zadeh's Fuzzy propositional Logic. *Technical Report 2000-b4-011*. Istituto di Elaborazione dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy, 2000.
2. Hahnle R. Automated Deduction in Multi-Valued Logic. *International Series of Monographs on Computer Science 10*. Oxford University Press, 1995.