DEMONSTRATIO MATHEMATICA
Vol. XXIV No 1-2 1991

Janusz A. Pomykała

SOME REMARKS ON APPROXIMATION

1. Introduction

The notion of information system, which is a starting point of the present paper, was introduced by Pawlak (see [1]) in 1981, and since then it has been intensively investigated. In particular related notions of a nondeterministic information system (see [2], [3]) and an approximation space (see [4], [5], [6]) were also examined. These notions are used to analyse computer and empirical data, being helpful in understanding indiscernibility and similarity of objects.

In sections 2, 3, 4 we recall basic notions and we give the short motivation for considering generalized approximation space. In section 5 we examine several approximation operations $\underline{\mathbb{E}}_i$, $\overline{\mathbb{E}}_i$, i=0...4, in view of elementary lattice theory. In section 6 we introduce the notion of approximation algebra and we use this notion to characterize families of definable sets with respect to the operations $\underline{\mathbb{E}}_i$, $\overline{\mathbb{E}}_i$ i=0...4.

Throughout the paper we use the standard mathematical notation, in particular P(X) stands for the family of all subsets of the set X. A family $\mathbb{E} \subseteq P(X)$ such that $\bigcup \mathbb{E} = X$ is called a cover of X. Frequently we will consider the cover $\mathbb{E}$ whose elements are nonempty, pairwise disjoint subsets of X. In such a case it is called a partition of X. Any relation τ on a set U which is reflexive and symmetric is called the tolerance relation. A set $E \subseteq U$ such that $E \times E \subseteq \tau$ and which is maximal with respect to inclusion is called a tolerance class (see [8]).

2. Information system and approximation space

We recall basic notions from papers [1]-[4].

Throughout the paper U will be an arbitrary fixed set, traditionally called universe.

An information system is a quadruple $(U, A, (V_a)_{a\in A}, f)$ where U is a set of objects, A stands for a set of attributes, V_a is a set of values of attribute a, and $f: U\times A \longrightarrow \bigcup_{a\in A} V_a$ is a function (called information function) such that $f(x,a)\in V_a$ for any $x\in U$ and $a\in A$.

For every subset $B\subseteq A$ an indiscernibility relation $Ind(B)\subseteq U^2$ is defined in the following way: for any $x, y \in U$

(1) $x\ Ind(B)\,y$ iff $f(x, a)=f(y, a)$ for every $a \in B$.

If $x\ Ind(B)\,y$ we say that x, y are indiscernible with respect to B.

Suppose that R is an equivalence relation in U. The pair (U,R) is called an approximation space. $[x]_R$ will stand for the equivalence class of the relation R determined by $x \in U$.

Traditionally the equivalence classes of R are called R-elementary sets.

For any set $X \subseteq U$ its lower (resp. upper) approximation $\underline{R}(X)$ (resp. $\overline{R}(X)$) is defined as follows:

$$\underline{R}(X)=\left\{x:\ [x]_R\subseteq X\right\},$$

(2)

$$\overline{R}(X)=\left\{x:\ [x]_R \cap X\neq\emptyset\right\}.$$

For brevity, we often write $\underline{X}$ instead of $\underline{R}(X)$ and $\overline{X}$ instead of $\overline{R}(X)$.

Let us recall that any set $X \subseteq U$ is called definable iff $\underline{R}(X)=\overline{R}(X)$. Equivalently, X is definable iff $\underline{R}(X)=X$ iff $\overline{R}(X)=X$ iff X is a union of some R-elementary sets. Thus, the family Def(U,R) of all definable sets is a complete atomic Boolean algebra with the usual set operations, having as atoms the elementary sets. The family Def(U,R) is topology for U while the family of all elementary sets is a base for Def(U,R). $\underline{R}(X)$ ($\overline{R}(X)$) is an interior (a closure) of X, respectively.

3. Nondeterministic information system

Suppose we are given the information system $(U, A, (V_a)_{a\in A}, f)$. It may happen that the information function f is not determined precisely i.e. the values of f are not settled uniquely. For instance, assume that one has to estimate the value of a light stimulus on a given measurement scale; then, an estimation is given by the interval in which we expect to find the actual value of the stimulus. Then it may be reasonable to consider a function F having as values the subsets of V.

Formally we define: The quadruple $(U, A, (V_a)_{a\in A}, F)$ where F is an arbitrary function satisfying
$F: U \times A \longrightarrow P(V)$ and $F(U\times\{a\}) \subseteq P(V_a)$ forany $a\in A$, is called a **nondeterministic information system** (see [2]).

Now let $(U, A, (V_a)_{a\in A}, F)$ be the nondeterministic information system. For any subset $B\subseteq A$ we define a similarity of objects with respect to B in the following way: for any $x, y\in U$

(3) $\quad (x, y) \in sim(B)$ iff $\forall_{b\in B}\ F(x,b) \cap F(y,b) \neq \emptyset$.

The relation sim(B) is called B-similarity relation and, if $(x, y) \in sim(B)$ then we say that x, y are B-similar (see [12]). Some other tolerances in the system $(U, A, (V_a)_{a\in A}, F)$ are worth mentioning:

$$(x, y)\in\Pi \text{ iff } \forall_{a,b\in B}\ (F(x,a)\subseteq F(y,a) \text{ or } F(y,b)\subseteq F(x,b)),$$

$$(x, y)\in\Pi' \text{ iff } \forall_{b\in B}\ F(x,b)\cap F(y,b)\neq\emptyset \text{ and } \exists_{b\in B}\ F(x,b)=F(y,b).$$

Thus it seems to be desirable to examine systems with tolerance.

4. Approximation space

As was mentioned above, the notion of approximation space has been defined by Pawlak as a pair (U,R) where R is an equivalence relation in U. Now, suppose that R is an arbitrary binary relation in U. Any set $E\subseteq U$ satisfying $E\times E\subseteq R$ and maximal with respect to inclusion will be called R-elementary.

Applying Kuratowski-Zorn lemma we shall prove the following

Lemma 1. Suppose R is a reflexive relation in U. Then the family $\mathbb{E}$ of all R-elementary sets is a cover of U.

Proof. Suppose x is an arbitrary element of U; we have $\{x\}\times\{x\}\subseteq R$ by reflexivity of R. Now consider any chain $\{E_\eta : \eta<\lambda\}$ of sets such that $E_\eta \times E_\eta \subseteq R$. We have

$$\bigcup_{\eta<\lambda} E_\eta \times \bigcup_{\eta<\lambda} E_\eta \subseteq R.$$

Indeed, if $x\in\bigcup_{\eta<\lambda} E_\eta$ and $y\in\bigcup_{\eta<\lambda} E_\eta$ then $x\in E_\delta$ for some $\delta<\lambda$ and $y\in E_\gamma$ for some $\gamma<\lambda$. Since $E_\delta\subseteq E_\gamma$ or $E_\gamma\subseteq E_\delta$ we infer that $(x, y) \in E_\gamma\times E_\gamma$ or $(x, y)\in E_\delta\times E_\delta$ hence finally $(x,y)\in R$. In other words $\bigcup_{\eta<\lambda} E_\eta$ is an upper bound of the chain $\{E_\eta:\eta<\lambda\}$. Therefore there exists a maximal set E satisfying $E\times E\subseteq R$ such that $x\in E$, in view of Kuratowski-Zorn lemma. Hence x belongs to some R-elementary set, as required.

As a consequence we obtain the wel' known

Corollary 1. Suppose τ is a tolerance in U. Then the family $\mathbb{E}(\tau)$ of all tolerance classes of τ is a cover of U.

This is our motivation to consider in what follows the space (U, $\mathbb{E}$), where $\mathbb{E}$ is a cover of U. The pair (U, $\mathbb{E}$) will be called generalized approximation space.

5. Approximation operations

Suppose (U, $\mathbb{E}$) is a generalized approximation space. Let us recall that the indiscernibility neighbourhood of an element $x\in U$ is the set

$$O_x^{\mathbb{E}} = \bigcup \{\, E_t : x\in E_t \}.$$

For any element $x\in U$, the set

$$I_x^{\mathbb{E}} = \{\, y\in U : \forall E_t \ (x\in E_t \Leftrightarrow y\in E_t) \,\}$$

will be called the kernel of x in view of its analogy to the notion used in the theory of tolerance relations. If no confusion is possible we shall write O_x and I_x instead of O_x^E and I_x^E, respectively. Let J be the family of all the kernels of $(U, \mathbb{E})$:

$$J(\mathbb{E}) = J = \{\ I_x:\ x \in U\ \}.$$

It is easy to verify that J is a partition; the equivalence relation determined by J will be denoted by I. If xIy then we say that x, y are $\mathbb{E}$-inseparable.

Let -X stand for U-X. We say that two operations G,G':P(U)->P(U) are conjugated iff for any X U, the following condition is satisfied:

$$G(X) = -G'(-X).$$

Now we apply the introduced notions to define some special pairs of conjugated approximation operations in the space $(U, \mathbb{E})$. When $\mathbb{E}$ is a partition of U, all those operations will coincide with the well known lower and upper approximation operations of Pawlak. The motivation to consider pairs of conjugated operations comes from two sources: first, the operations G and G' may be used to define operators of necessity and possibility in a respective modal logic and second, in cases when G, G' are topological operations then in order to define the same topology on U, they have to be conjugated.

Let X be a subset of U. We will define the operations $\underline{\mathbb{E}}_i$, as follows: (see [6])

(4)

$$\underline{\mathbb{E}}_1(X) = \{x:\ O_x \subseteq X\}$$

$$\overline{\mathbb{E}}_1(X) = \bigcup \{E_t : E_t \cap X \neq \emptyset\}$$

$$\underline{\mathbb{E}}_2(X) = \bigcup \{O_x : O_x \subseteq X\}$$

$$\overline{\mathbb{E}}_2(X) = \{z : \forall y\ (z \in O_y \Rightarrow O_y \cap X \neq \emptyset\},$$

$$\underline{\mathbb{E}}_3(X) = \bigcup \{E_t : E_t \subseteq X\}$$

$$\overline{\mathbb{E}}_3(X) = \{y : \forall E_t (y \in E_t \Rightarrow E_t \cap X \neq \emptyset\}$$

$$\underline{\mathbb{E}}_4(X) = \bigcup \{I_y : I_y \subseteq X\}$$

$$\overline{\mathbb{E}}_4(X) = \bigcup \{I_y : I_y \cap X \neq \emptyset\}.$$

First, let us observe that for i=2, 3, 4 the operations $\underline{\mathbb{E}}_i$, $\overline{\mathbb{E}}_i$ are idempotent i.e. for any $n\in\omega$, $X\subseteq U$ and for i=2, 3, 4 the following conditions are valid:

(a) $(\underline{\mathbb{E}}_i^n(X)) = \underline{\mathbb{E}}_i(X)$,

(b) $((\overline{\mathbb{E}}_i)^n(X)) = \overline{\mathbb{E}}_i(X)$.

The situation is more complicated when we iterate operation $\underline{\mathbb{E}}_1$ or $\overline{\mathbb{E}}_1$. For any $X\subseteq U$ we have the inclusions

$$\overline{\mathbb{E}}_1(X)\subseteq(\overline{\mathbb{E}}_1)^2(X)\subseteq(\overline{\mathbb{E}}_1)^3(X)\ldots$$

but it may happen that the elements of this sequence are pairwise distinct. So, we introduce one more approximation operation $\overline{\mathbb{E}}_0$ in the following manner:

$$\overline{\mathbb{E}}_0(X) = \bigcup_{i<\omega}(\overline{\mathbb{E}}_1)^i(X). \tag{5}$$

We shall call $\overline{\mathbb{E}}_0$ the transitive closure operation (by analogy to the terminology used in the theory of tolerance relations). The set $\overline{\mathbb{E}}_0(\{x\})$ denoted by C_x will be called the component of x in U,

$$C_x \overset{\text{def}}{=} \overline{\mathbb{E}}_0(\{x\}).$$

It is easy to check that $\overline{\mathbb{E}}_0(X)=\bigcup\{C_y\colon C_y\cap X \neq \emptyset\}$ and the conjugated operation $\underline{\mathbb{E}}_0$ satisfies

$$\underline{\mathbb{E}}_0(X) = \{x\in X\colon C_x\subseteq X\},$$

since the family $\mathbb{C}=\{C_x : x\in U\}$ of all components in U, is a partition of U.

Let us also observe that the following inclusions hold:

$$\underline{\mathbb{E}}_0 \le \underline{\mathbb{E}}_1\le \underline{\mathbb{E}}_2\le \underline{\mathbb{E}}_3\le \underline{\mathbb{E}}_4\le \mathrm{Id} \le \overline{\mathbb{E}}_4\le \overline{\mathbb{E}}_3\le \overline{\mathbb{E}}_2\le \overline{\mathbb{E}}_1\le \overline{\mathbb{E}}_0.$$

Now, to express the algebraic properties of the above operations, we recall some notions from lattice theory:

Let G be a mapping of P(U) into itself. We shall say that G is a lower (upper) operation on U iff for any $X\subseteq U$, $G(X)\subseteq X$ ($G(X)\supseteq X$), respectively. (The upper operation is also called extensive) (see [7]).

The mapping G is said to be monotonic iff (if $X \subseteq Y$ then $G(X) \subseteq G(Y)$ for any $X, Y \subseteq U$). Any monotonic and lower or monotonic and upper operation will be called an approximation operation. The most important examples of operations satisfying this definition are the lower $\underline{R}$ and upper $\overline{R}$ approximation operations of Pawlak (see [2]). The mapping G is said to be **idempotent** iff for every $X \subseteq U$, $G(X) = G(G(X))$. If G is an upper, monotonic and idempotent mapping then G is called a **closure mapping** and the pair (U, G) is called a **closure space** (see [7]).

To summarize this section we recall that a closure operator H on the set U is an **algebraic** (resp. **topological**) closure operator if for every $X \subseteq U$

$$H(X) = \bigcup \{ H(X') : X' \subseteq X \text{ and } X' \text{ is finite} \}$$

(resp. for every $X, Y \subseteq U$ $H(X \cup Y) = H(X) \cup H(Y)$).

Theorem 1. Assume $(U, \mathbb{E})$ is a generalized approximation space and $\underline{E}_i, \overline{E}_i$, $i=0,\ldots,4$ are the approximation operations defined by (4). Then it holds:

(a) $\overline{E}_0, \overline{E}_4$ are topological algebraic closure operations;

(b) $\overline{E}_2, \overline{E}_3$ are closure operations;

(c) $\overline{E}_1$ is monotonic, extensive and it satisfies the condition

$$\overline{E}_1(X \cup Y) = \overline{E}_1(X) \cup \overline{E}_1(Y), \text{ for any } X, Y \subseteq U;$$

(d) $\overline{E}_1$ is a topological closure operation iff $\{O_x : x \in U\}$ is a partition of U.

Proof. It is easy to prove (a), (b) and (c). For a little bit more dificult (d) see [6].

6. Approximation algebra

In applications it is often considered a family of all definable subsets of the universe U. To formulate definitions of these families in a unified way, we introduce the following approximation algebra:

An algebra $(P(U), \{\underline{G}_i, \overline{G}_i : i \in I\})$ is called **an approximation algebra** on U if, for any $X, Y \subseteq U$ and $i \in I$, it satisfies:

1) $\overline{G}_i : P(U) \longrightarrow P(U)$,

2) $X \subseteq \overline{G}_i$,

3) $X \subseteq Y$ implies $\overline{G}_i(X) \subseteq \overline{G}_i(Y)$,

4) $\underline{G}_i(X) = -\overline{G}_i(-X)$.

A subset X of U is called a definable subset with respect to $\{\overline{G}_i : i \in I_0 \subseteq I\}$ if for every $i \in I_0$ it holds $\overline{G}_i(X) = X$. In other words X is a fixed point of all $\overline{G}_i$, $i \in I_0$. Similary, X is definable with respect to $\{\underline{G}_i : i \in I_0\}$ if $\forall_{i \in I_0} \underline{G}_i(X) = X$. The family of all definable sets with respect to $\{\underline{G}_i : i \in I_0\}$ will be denoted by Def(U, $\{\underline{G}_i : i \in I_0\}$), or in short by $\underline{\mathrm{Def}}(I_0)$. $\overline{\mathrm{Def}}(I_0)$ = Def(U, $\{\overline{G}_i : i \in I_0\}$) denotes the family of all definable sets with respect to $\{\overline{G}_i : i \in I_0\}$.

Lemma 2. The family $\overline{\mathrm{Def}}(I_0)$ is closed on intersections i.e. if $\forall_{t \in S}\, X_t \in \overline{\mathrm{Def}}(I_0)$ then $\bigcap_{t \in S} X_t \in \overline{\mathrm{Def}}(I_0)$.

Proof. Assume $i \in I_0$. It holds

$$\overline{G}_i\Big(\bigcap_{t \in S} X_t\Big) \subseteq \bigcap_{t \in S} \overline{G}_i(X_t) = \bigcap_{t \in S} X_t$$

in view of the monotonicity of $\overline{G}_i$ and the hypothesis. On the other hand $\overline{G}_i(\bigcap_{t \in S} X_t) \supseteq \bigcap_{t \in S} X_t$, since $\overline{G}_i$ is extensive. Finally, $\overline{G}_i(\bigcap_{t \in S} X_t) = \bigcap_{t \in S} X_t$.

Corollary 2. $\overline{\mathrm{Def}}(I_0)$ is a complete lattice with respect to set inclusion, and

$\inf\{X_t : t \in S\} = \bigcap_{t \in S} X_t$, $\sup\{X_t : t \in S\} = \bigcap \{X \in \overline{\mathrm{Def}}(I_0) : X \supseteq X_t \ \forall t \in S\}$

Proof. It is a consequence of Lemma 9, p.184 in [7], or Theorem 4.2 p.14 in [9].

Lemma 3. The family $\underline{\mathrm{Def}}(I_0)$ is closed on arbitrary unions.

Proof. If $X_t \in \underline{\mathrm{Def}}(I_0)$ for every $t \in S$, then

$$\bigcup_{t\in S} X_t = \bigcup_{t\in S} \underline{G}_i(X) \leq \underline{G}_i(\bigcup_{t\in S} X_t) \leq \bigcup_{t\in S} X_t, \text{ for } i\in I_0.$$

Hence

$$\underline{G}_i(\bigcup_{t\in S} X_t) = \bigcup_{t\in S} X_t, \text{ for } i\in I_0, \text{ i.e. } \bigcup_{t\in S} X_t \in \underline{Def}(I_0).$$

Corollary 3. $\underline{Def}(I_0)$ is a complete lattice with respect to set inclusion and $\sup\{X_t : t\in S\} = \bigcup_{t\in S} X_t$

$\inf\{X_t : t\in S\} = \bigcup\{X\in\underline{Def}(I_0) : X \subseteq \bigcap_{t\in S} X_t\}$.

Applying these lemmas to the approximation algebra $(P(U), \{\underline{E}_i, \overline{E}_i, i\in\{0\ldots4\}\})$ we obtain:

Corollary 4. Assume that $\underline{E}_i, \overline{E}_i$, $i=0,\ldots,4$, are approximation operations in the space (U, E). Then it holds:

(a) $Def(U, \underline{E}_1) = Def(U, \overline{E}_1) = Def(U, \underline{E}_0) = Def(U, \overline{E}_0)$ and $Def(U, \underline{E}_4) = Def(U, \overline{E}_4)$ are fields of sets;

(b) $Def(U, \overline{E}_2)$ and $Def(U, \overline{E}_3)$ are complete lattices with respect to set inclusion and $\inf Y = \bigcap Y$, $\sup Y = \overline{E}_2(\bigcup Y)$ $(\sup Y = \overline{E}_3(\bigcup Y))$, for $Y \subseteq Def(U,\overline{E}_2)$, $(Y \subseteq Def(U,\overline{E}_3))$, respectively;

(c) $Def(U,\underline{E}_2)$, $Def(U,\underline{E}_3)$ are complete lattices with respect to set inclusion and $\sup Y = \bigcup Y$, $\inf Y = \underline{E}_2(\bigcap Y)$, $(\inf Y = \underline{E}_3(\bigcap Y))$, for $Y \subseteq Def(U,\underline{E}_2)$, $(Y \subseteq Def(U,\underline{E}_3))$ respectively.

REFERENCES

[1] **Z. Pawlak:** Information systems, theoretical foundations, Information Systems 6, (1981) 205-218.

[2] **Z. Pawlak:** Information systems, theoretical foundations (in Polish) Warsaw (1983).

[3] **E. Orłowska, Z. Pawlak:** Logical foundations of knowledge representation, ICS PAS Rep. 537 Warsaw (1984).

[4] **Z. Pawlak:** Rough sets, Int.J.Inf.Comp.Sci. 11(5), (1982), 341-356.

[5] **W. Żakowski:** Approximations in the space (U,Π), Demonstratio Math. 14 (1983) 761-769.

[6] **J.A. Pomykała:** Approximation operations in approximation space, Bull.Acad.Polon.Sci.Ser.Sci.Math. 35, (1987), 654-662.

[7] **G. Gratzer:** General Lattice Theory. Academic Press Inc, 1978.

[8] **E.O. Zeeman:** Topology of the Brain.(in M.K.Fort,Topology on 3-Manifolds, 1962).

[9] **H.P. Sankappanavar, S. Burris:** A Course in Universal Algebra, Springer, New York, 1981.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF AGRICULTURE AND PEDAGOGICS, 08-110 SIEDLCE, POLAND

Received November 8, 1989.