

Rough Sets and Matroids

Victor W. Marek¹ and Andrzej Skowron²

¹ Department of Computer Science,
University of Kentucky, Lexington, KY 40506
marek@cs.uky.edu

² Institute of Mathematics, The University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
skowron@mimuw.edu.pl

Abstract. We prove the recent result of Liu and Zhu [2] and discuss some consequences of that and related facts for the development of rough set theory.

Key words: rough set, matroid

1 Introduction

The goal of this note is to provide a proof of the recent statement by Liu and Zhu [2] and look at some properties of rough sets related to Liu and Zhu realization [10] that rough sets [4, 6, 9, 8, 7] relate to one of classical structures of combinatorics and computer science, namely matroid. The importance of that result is that it allows to tie various rough set methods to *greedy algorithms* that succeed when underlying combinatorial structure is defined by matroid [1, 3]. This allows for developments of algorithms for finding properties of maximal and minimal sets in various classes of sets (see also Propositions 2 and 3 below.)

2 Preliminaries

Below we introduce basic notions used in this paper. Generally, we assume that the reader is familiar with the notion of rough sets of [4, 6, 9].

2.1 Rough Sets

Any pair (U, \sim) , where U is a finite set and \sim is an equivalence relation in U is called an *approximation space*. We denote by $[x]$ the set $\{y \in U : x \sim y\}$. We call sets of the form $[x]$, *monads* (or *elementary granules* [9]). Monads of an equivalence relation \sim form a partition of the set U . Given a set $X \subseteq U$, the sets \underline{X} and \overline{X} are defined as $\bigcup\{[x] : [x] \subseteq X\}$, and $\bigcup\{[x] : [x] \cap X \neq \emptyset\}$, respectively. The sets \underline{X} and \overline{X} are called the *lower approximation* of X (relative to \sim) and the *upper approximation* of X (relative to \sim), respectively. The set $BN(X) = \overline{X} \setminus \underline{X}$ is called the boundary region of X (relative to \sim). If $BN(X) = \emptyset$ then X is *crisp* (relative to \sim), otherwise X is *rough* (relative to \sim). Pawlak, in [4], established the basic properties of these operations. We assume that the reader is familiar with these properties.

2.2 Matroids

Matroids are one of basic structures studied by combinatorists [3, 1]. Matroids occur in many areas of Mathematics and Computer Science as a common generalization of concepts such a collection of independent sets in a linear space and of the cycle-free sets in a graph. Matroids are closely related to issues in combinatorial optimization because of relationship between greedy algorithms and matroids.

Formally, a matroid over a set U is a nonempty family \mathcal{M} of subsets of U satisfying the following conditions:

1. $\emptyset \in \mathcal{M}$.
2. Whenever $X \in \mathcal{M}$, and $Y \subseteq X$, then $Y \in \mathcal{M}$.
3. (Steinitz Exchange Principle) Whenever $X, Y \in \mathcal{M}$ and $|X| < |Y|$ then for some $y \in Y \setminus X$, $X \cup \{y\} \in \mathcal{M}$.

By a parameterized matroid over an index set I we mean a family of matroids $\langle \mathcal{M}_i \rangle_{i \in I}$. In our case the set I will be the powerset of U , $\mathcal{P}(U)$.

3 Matroids Generated by Approximation Spaces

In this section we give a proof of the result of Liu and Zhu [2] on the parametric matroid associated with an approximation space defined by an equivalence relation \sim over a finite set U .

Definition 1. *Given an approximation space (U, \sim) we define for a set $Y \subseteq U$ the family of sets \mathcal{M}_Y as*

$$\{A \subseteq U : \underline{A} \subseteq Y\}. \quad (1)$$

Then, we prove

Theorem 1 (Liu and Zhu). *Let (U, \sim) be an approximation space. Then for every subset $Y \subseteq U$, \mathcal{M}_Y is a matroid.*

Proof: The family \mathcal{M}_Y is closed under subsets because whenever $B \subseteq A$ and $A \in \mathcal{M}_Y$, then, by the definition, $\underline{A} \subseteq Y$. Since $B \subseteq A$, we have $\underline{B} \subseteq \underline{A}$, thus $\underline{B} \subseteq Y$, i.e. $B \in \mathcal{M}_Y$. Therefore the first two conditions on matroid hold for \mathcal{M}_Y .

We will now show the exchange property for \mathcal{M}_Y . To this end, let A, B be two sets, $A, B \in \mathcal{M}_Y$, $|A| < |B|$. We need to find $x \in B \setminus A$ so that $A \cup \{x\} \in \mathcal{M}_Y$.

Our argument consists of two cases.

Case 1. Some $x \in B \setminus A$ has the property that $[x] = \{x\}$. That is, for $y \neq x$, $y \not\sim x$. We claim that for that x , $A \cup \{x\} \in \mathcal{M}_Y$.

Since $[x] = \{x\}$ and $x \notin A$, we have

$$\underline{A \cup \{x\}} = \underline{A} \cup \{x\}$$

Now, $\underline{A} \subseteq Y$ (because $A \in \mathcal{M}_Y$), and also $x \in Y$ because $\underline{B} \subseteq Y$ and $\{x\} = [x] \subseteq \underline{B} \subseteq Y$. Thus $\underline{A} \cup \{x\} \subseteq Y$, and so $A \cup \{x\} \in \mathcal{M}_Y$.

Case 2. No $x \in B \setminus A$ has the property that $[x] = \{x\}$. We will now assume that for no $x \in B \setminus A$, $A \cup \{x\} \in \mathcal{M}_Y$ and show that this leads to the contradiction.

Let us look at an arbitrary $x \in B \setminus A$. Under our assumption ($A \cup \{x\} \notin \mathcal{M}_Y$), it must be the case that $\overline{A \cup \{x\}}$ is strictly bigger than \underline{A} (because if $\overline{A \cup \{x\}} = \underline{A}$ then as $\underline{A} \subseteq Y$, then, since $\underline{A} \subseteq Y$, $\overline{A \cup \{x\}} \subseteq Y$, so $A \cup \{x\} \in \mathcal{M}_Y$, a contradiction.)

But what is $\overline{A \cup \{x\}}$?

There are two possibilities:

1. $\overline{A \cup \{x\}} = \underline{A}$, or
2. $\overline{A \cup \{x\}} = \underline{A} \cup [x]$.

Since the first possibility has already been eliminated, it must be the case that $\overline{A \cup \{x\}} = \underline{A} \cup [x]$. But this means that for all y such that $y \neq x$, $y \sim x$, the element y must belong to A .

Moreover, since x was an arbitrary element of $B \setminus A$, it must be the case that whenever $x \in B \setminus A$, $y \neq x$, $y \sim x$ then $y \in A$.

Next, we ask if it is possible that for some $x, y \in B \setminus A$, $x \neq y$, $x \sim y$. We claim that this is impossible. Indeed let us assume that for some $x, y \in B \setminus A$, $x \neq y$, $x \sim y$, then $[x] = [y]$ and $[x] \setminus \{x\} \subseteq A$ and $[y] \setminus \{y\} \subseteq A$. Then $y \in [x] \setminus \{x\}$, i.e. $\{y\} \subseteq [x] \setminus \{x\}$. Therefore

$$[y] = ([y] \setminus \{y\}) \cup \{y\} \subseteq ([x] \setminus \{x\}) \cup ([x] \setminus \{x\}) \subseteq A$$

contradicting the fact that $y \notin A$.

Now, for every $x \in B \setminus A$ let us select an element y_x so that:

1. $y_x \sim x$
2. $y_x \in A$.

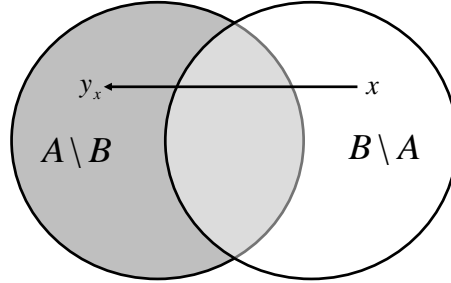


Fig. 1. Mapping $B \setminus A$ into $A \setminus B$

Figure 1 illustrates the fact that $B \setminus A$ can be injected into $A \setminus B$.

We observe that there is such mapping $x \mapsto y_x$ because we are in Case 2. Also, the mapping $x \mapsto y_x$ is an injection, i.e., $x_1 \neq x_2$ implies $y_{x_1} \neq y_{x_2}$. But, of course y_x belongs to $A \setminus B$ (because $y_x \sim x$ and $y_x \notin B$). Therefore we now have an injection of $B \setminus A$ into $A \setminus B$. But then, $|B \setminus A| \leq |A \setminus B|$. This contradicts the fact that $|A| < |B|$ and completes the proof. \square

The matroid \mathcal{M}_Y is called the *matroid defined by Y and the approximation space (U, \sim)* .

4 Properties of Parameterized Matroids of Approximation Spaces, and Their Characterization

For a given approximation space (U, \sim) we consider a *parameterized matroid associated with the approximation space (U, \sim)* assuming that $\mathcal{M}^\sim = \{\mathcal{M}_Y : Y \in \mathcal{P}(U)\}$, where \mathcal{M}_Y is a matroid defined by Y and the approximation space (U, \sim) , and $\mathcal{P}(U)$ is the powerset of U . Instead \mathcal{M}^\sim we also write \mathcal{M} , for short.

We now show the following fact.

Proposition 1. *Let \mathcal{M} be a parameterized matroid associated with the approximation space (U, \sim) . Then*

$$\underline{Z} = \underline{X} \text{ iff } \mathcal{M}_Z = \mathcal{M}_X \text{ for any } X, Z \in \mathcal{P}(U). \quad (2)$$

Proof: First assume that $\underline{Z} = \underline{X}$. Then if $A \in \mathcal{M}_Z$, then $\underline{A} \subseteq Z$ and thus $\underline{A} \subseteq \underline{Z}$. Therefore $\underline{A} \subseteq \underline{X}$, thus $\underline{A} \subseteq X$, so $A \in \mathcal{M}_X$. Therefore $\mathcal{M}_Z \subseteq \mathcal{M}_X$. But if $A \in \mathcal{M}_X$, then $\underline{A} \subseteq X$. Then $\underline{A} \subseteq \underline{X} = \underline{Z} \subseteq Z$. This completes implication \Rightarrow . Conversely, let us assume that $\mathcal{M}_Z = \mathcal{M}_X$. We want to show that $\underline{Z} = \underline{X}$. If $\underline{Z} \neq \underline{X}$ then there is Y such that $Y = \underline{Y}$, $Y \neq \emptyset$ and $Y \subseteq \underline{Z}$, $Y \cap X = \emptyset$, or $Y \subseteq \underline{X}$, $Y \cap Z = \emptyset$.

We consider the first case, the other is similar. For that set Y , $Y \in \mathcal{M}_Z$ since $Y \subseteq \underline{Z} \subseteq Z$ and $Y = \underline{Y}$. But $Y \cap X = \emptyset$ so $\underline{Y} \cap X = \emptyset$, a contradiction. \square

Definition 2. *Let (U, \sim) be an approximation space. Let $X \subseteq U$ be a crisp set, i.e., $X = \underline{X}$. We define \mathcal{D}_X as the collection of all monads M such that $M \cap X = \emptyset$.*

We observe that the elements of \mathcal{D}_X are pairwise disjoint and nonempty. Moreover, the union of all monads from \mathcal{D}_X is equal to the lower approximation of $U \setminus X$, i.e., $U \setminus X = \bigcup \mathcal{D}_X$. Let us assume that $X \neq U$. Then the family \mathcal{D}_X possesses *selectors*, i.e., sets S such that $S \subseteq \bigcup \mathcal{D}_X$ and for all $D \in \mathcal{D}_X$, $|S \cap D| = 1$. We can now present the description of bases of matroids in \mathcal{M} .

Proposition 2 (Truszczyński). *Let (U, \sim) be an approximation space, and let \mathcal{M} be its parameterized matroid. Then for every set $X \neq U$ such that $X = \underline{X}$, the bases for \mathcal{M}_X are precisely the sets of the form $U \setminus S$, where S is a selector for $\mathcal{D}_{\underline{X}}$.*

Proof: A base B of \mathcal{M}_X is an inclusion-maximal set in \mathcal{M}_X . This means that for any $x \notin B$, $B \cup \{x\}$ does not belong to \mathcal{M}_X , that is $\overline{B \cup \{x\}}$ is strictly larger than \underline{B} . But $\underline{B} = \underline{X}$. Thus $B \cup \{x\}$ contains at least one more monad M . This means that all the remaining elements of the monad M are already in B . But as this monad M was arbitrary among those not included in \underline{X} , we have that B is of the form $U \setminus S$ where S is a selector for the family $\{M \in U / \sim : M \cap \underline{X} = \emptyset\}$. The converse implication is obvious. \square

One consequence of Proposition 2 is that one can use a greedy algorithm to compute a maximal weight set roughly equivalent to a given set X .

To make this claim precise, let us say that subsets X and Y of U are *roughly equivalent* if and only if $\underline{X} = \underline{Y}$ and $\overline{X} = \overline{Y}$ [6]. The following property characterizes roughly equivalent sets $X, Y \subseteq U$: $[x] \subseteq X$ if and only if $[x] \subseteq Y$, for all $x \in U$.

A *weight function* on the set U is any function $wt : U \rightarrow \mathbf{R}^+$, where \mathbf{R}^+ is the set of all positive reals. The weight of a set $Z \subseteq U$ is equal to $\sum_{z \in Z} wt(z)$.

Our task now is, given $X \subseteq U$ to find a roughly equivalent to X set Y of maximum weight. Each basis of \mathcal{M}_X is roughly equivalent to X and by Proposition 2 all we need to do is to find a selector for \mathcal{D}_X of minimal weight. But such selector can be found by choosing in each element $[z]$ of \mathcal{D}_X a single element of least possible weight (we observe that such element does not need to be unique.)

Another class of sets associated with rough sets is that of *representative sets*³.

A set $X \subseteq U$ is representative if $\overline{X} = U$, that is for every $x \in U$, there is $y \in X$ so that $x \sim y$. The parameterized matroid \mathcal{M}^\sim associated with the approximation space (U, \sim) determines a class of special representative sets. Specifically, let $X \subseteq U$ be a crisp set in an approximation space (U, \sim) , i.e., $X = \underline{X}$. Then we characterize the minimal representative sets including X that belong to the matroid \mathcal{M}_X as follows.

Proposition 3. *Let $X \subset U$ be a crisp set in an approximation space (U, \sim) , i.e., $X = \underline{X}$. Then the minimal representative sets including X belonging to the matroid \mathcal{M}_X are precisely the sets of the form $X \cup S$ where S is a selector for \mathcal{D}_X .*

Given Proposition 3, a greedy algorithm can be used to find the minimal representative set of minimal weight.

We now list a number of properties of the parameterized matroid \mathcal{M}^\sim .

- Proposition 4.**
1. For every $X \in \mathcal{P}(U)$, $X \in \mathcal{M}_X$.
 2. For every $X \in \mathcal{P}(U)$, $\mathcal{M}_X = \mathcal{M}_{\underline{X}}$.
 3. For all $X, Y \in \mathcal{P}(U)$, $X \subseteq Y$ implies $\mathcal{M}_X \subseteq \mathcal{M}_Y$.
 4. For every $X \in \mathcal{P}(U)$, \underline{X} is the \subseteq -least set in $\mathcal{M}_X \setminus \bigcup \{\mathcal{M}_Y : \mathcal{M}_Y \subset \mathcal{M}_X\}$.
 5. The family $\{\underline{X} : X \in \mathcal{P}(U)\}$ forms a Boolean algebra.
 6. For all $X, Y \in \mathcal{P}(U)$ if $Y \in \mathcal{M}_X$ then $Y \setminus \underline{X} = \emptyset$.
 7. For all $X, Y, Z \in \mathcal{P}(U)$ if $Y = \underline{X} \cup Z$ and $\overline{Z} = \emptyset$ then $Y \in \mathcal{M}_X$.
 8. If $X \subseteq Z \subseteq Y$ and $\mathcal{M}_X = \mathcal{M}_Y$ then $\mathcal{M}_X = \mathcal{M}_Z$.

Points (1)–(8) are almost obvious, except possibly (4).

But the same points provide a key to the answer to the following question:

Given a parameterized matroid

$$\mathcal{N} = \{\mathcal{N}_X : X \in \mathcal{P}(X)\},$$

³ Note that in [6, 9] such sets are called externally or totally undefinable relative to a given approximation space. Such sets were also used by Pawlak in investigating the notion of rough truth [5].

when there exists an approximation space (U, \approx) so that

$$\mathcal{N} = \mathcal{M}^\approx.$$

Specifically, we will formulate seven abstract conditions, corresponding to points (1)–(7) above and show that under these conditions, indeed the parameterized matroid is determined by an approximation space (U, \approx) that is determined by a parameterized matroid \mathcal{N} .

So let $\mathcal{N} = \langle \mathcal{N}_X \rangle_{X \in \mathcal{P}(U)}$ be a parameterized matroid. We formulate *conditions* (A)–(F) that \mathcal{N} needs to satisfy.

- (A) For all $X \in \mathcal{P}(U)$, $X \in \mathcal{N}_X$.
- (B) For all $X, Y \in \mathcal{P}(U)$, $X \subseteq Y$ implies $\mathcal{N}_X \subseteq \mathcal{N}_Y$.
- (C) For all $X \in \mathcal{P}(U)$, the family

$$\mathcal{N}_X \setminus \bigcup \{ \mathcal{N}_Y : \mathcal{N}_Y \subset \mathcal{N}_X \},$$

possesses a \subseteq -least element, further referred as $[X]$.

- (D) The family $\{[X] : X \in \mathcal{P}(U)\}$ forms a Boolean Algebra, further referred as $\mathcal{B}_\mathcal{N}$, or simply \mathcal{B} .
- (E) For all $X \in \mathcal{P}(U)$, $\mathcal{N}_X = \mathcal{N}_{[X]}$.
- (F) For all $X, Y \in \mathcal{P}(U)$, if $Y \in \mathcal{N}_X$ then $[Y \setminus [X]] = \emptyset$.
- (G) For all $X, Y, Z \in \mathcal{P}(U)$, if $Y = [X] \cup Z$ and $[Z] = \emptyset$ than $Y \in \mathcal{N}_X$.

Once \mathcal{N} is a parameterized matroid satisfying conditions (A)–(F), we define a relation \approx in U by setting:

$x \approx y$ if and only if there is an atom A of \mathcal{B} such that $x \in A$ and $y \in A$.

It is easy to see that (under conditions (A)–(G), in particular condition (D)), we have

Proposition 5. $x \approx y$ if and only if

$$\text{for every } X \subseteq U, x \in [X] \text{ if and only if } y \in [X].$$

One can also observe the following fact:

Proposition 6. Let $\mathcal{N} = \langle \mathcal{N}_X \rangle_{X \in \mathcal{P}(U)}$ be a parameterized matroid satisfying conditions (A)–(G). Then for any $Y \subseteq U$ we have $[Y] = \underline{Y}$, where \underline{Y} is the lower approximation of Y in the approximation space (U, \approx) and $[Y] \in \mathcal{B}$.

Proof: Let us assume $x \in [Y]$. Then from Proposition 5 we have $y \in Y$ for $y \approx x$. Since $[Y] \subseteq Y$, we obtain $[x]_\approx \subseteq Y$, i.e., $x \in \underline{Y}$.

Now let us assume $x \in \underline{Y}$, i.e., $[x]_\approx \subseteq Y$. Suppose that $x \notin [Y]$. Then by Proposition 5 we have $[x]_\approx \subseteq U \setminus [Y]$. Hence, $[x]_\approx \subseteq Y \setminus [Y]$, a contradiction with (F) (where we take $X = Y$). \square

We now show the main result of this section.

Proposition 7. Let $\mathcal{N} = \langle \mathcal{N}_X \rangle_{X \in \mathcal{P}(U)}$ be a parameterized matroid. Then \mathcal{N} is a parameterized matroid defined by some approximation space, i.e., $\mathcal{N} = \mathcal{M}^\sim$ for some approximation space (U, \sim) if and only if \mathcal{N} satisfies conditions (A)–(G) above.

Proof: By Proposition 4, if \mathcal{N} is a parameterized matroid for an approximation space, then \mathcal{N} satisfies conditions (A)-(G).

Conversely, if \mathcal{N} satisfies conditions (A)-(G), then we show that $\mathcal{M}^{\approx} = \mathcal{N}$. That is we show that for every $X \subseteq U$, $\mathcal{N}_X = \mathcal{M}_X^{\approx}$.

First, assume $Y \in \mathcal{M}_X^{\approx}$. The set $Y \setminus [Y]$ is sparse w.r.t. \approx , *i.e.*, $\underline{Y \setminus [Y]} = \emptyset$, where the lower approximation is relative to the approximation space (U, \approx) . By proposition 6 we obtain $[Y \setminus [Y]] = \emptyset$. By condition (G), $[X] \cup (Y \setminus [Y]) \in \mathcal{N}_X$. But $\underline{Y} \subseteq X$. Hence, by Proposition 6 $[Y] \subseteq X$. Therefore $[Y] \subseteq [X]$. But then, $Y = [Y] \cup (Y \setminus [Y]) \subseteq [X] \cup (Y \setminus [Y]) \in \mathcal{N}_X$, as desired.

Conversely, let $Y \in \mathcal{N}_X$. By Proposition 6 we need only to show that $[Y] \subseteq X$. But $Y \in \mathcal{N}_X$ means (see (F)) that $[Y] \subset [X]$ or $[Y] = [X]$. In either case, as $[X] \subseteq X$, we have $[Y] \subseteq X$, that is $\underline{Y} \subseteq X$, by Proposition 6. Hence $Y \in \mathcal{M}_X^{\approx}$. This completes the argument. \square

5 Conclusions

In the paper we have presented some relationships between rough set theory and metroid theory. We plan to explore possibilities of application of heuristics based on combinatorial optimization developed in matroid theory to algorithmic problems in rough set theory.

Acknowledgements

The authors acknowledge valuable discussions with Professor Mirosław Truszczyński. Andrzej Skowron was supported by the Polish National Science Centre grant 2012/05/B/ST6/ 03215 as well as by the Polish National Centre for Research and Development (NCBiR) under the grant SYNAT No. SP/I/1/77065/10 in frame of the strategic scientific research and experimental development program: “Interdisciplinary System for Interactive Scientific and Scientific-Technical Information” and the grant No. O ROB/0010/ 03/001 in frame of the Defence and Security Programmes and Projects: “Modern engineering tools for decision support for commanders of the State Fire Service of Poland during Fire & Rescue operations in the buildings”.

References

1. J. Edwards. Matroids and the greedy algorithm. *Mathematical Programming* 1 (1971) 127–136.
2. Y. Liu, W. Zhu. Parametric matroid of rough set. arXiv: 1209.4975 (2012) 1–15.
3. J. Oxley. *Matroid Theory*. Oxford University Press, Oxford, UK, 2006.
4. Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences* 11 (1982) 341–356.
5. Z. Pawlak. Rough logic. *Bulletin of the Polish Academy of Sciences, Technical Sciences* 35(5-6) (1987) 253–258.
6. Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning about Data, System Theory, Knowledge Engineering and Problem Solving*, vol. 9. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1991.

7. Z. Pawlak, A. Skowron. Rough sets and Boolean reasoning. *Information Sciences* 177(1) (2007) 41–73.
8. Z. Pawlak, A. Skowron. Rough sets: Some extensions. *Information Sciences* 177(1) (2007) 28–40.
9. Z. Pawlak, A. Skowron. Rudiments of rough sets. *Information Sciences* 177(1) (2007) 3–27.
10. J. Tanga, K. Shea, F. Minb, W. Zhu. A matroidal approach to rough set theory. *Theoretical Computer Science* 471 (2013) 1–11.