

Dualization on Partially Ordered Sets: Preliminary Results

Lhouari Nourine¹ and Jean Marc Petit²(✉)

¹ Clermont-Université, Université Blaise Pascal, LIMOS, CNRS,
Aubière, France

² Université de Lyon, CNRS, LIRIS, INSA, Lyon, France
`jmpetit@liris.cnrs.fr`

Abstract. The dualization problem on arbitrary posets is a crucial step in many applications in logics, databases, artificial intelligence and pattern mining.

The objective of this paper is to study *reductions* of the dualization problem on arbitrary posets to the dualization problem on boolean lattices, for which output quasi-polynomial time algorithms exist. We introduce *convex embedding* and *poset reflection* as key notions to characterize such reductions. As a consequence, we identify posets, which are not boolean lattices, for which the dualization problem remains quasi-polynomial and propose a classification of posets with respect to dualization.

As far as we know, this is the first contribution to explicit non-trivial reductions for studying the hardness of dualization problems on arbitrary posets.

1 Introduction

The dualization problem in arbitrary finite¹ partially ordered sets (poset for short), is a crucial step in many applications in logics, databases, artificial intelligence and pattern mining and has been intensively studied for years [1–3]. The dualization problem can be very difficult and the decision problem associated to dualization is still open, even for lattices. Only a few results exist, mainly dualization is quasi-polynomial whenever (P, \leq) is isomorphic to a powerset ordered under set inclusion (or boolean lattice) [2], which corresponds to the enumeration of minimal transversal of hypergraph.

The dualization problem on hypergraphs has been studied by many researchers, among which we quote [1, 2] while only a few results exist on dualization on other posets [4]. Recently Kante *et al.* [5] have shown that the enumeration of minimal transversal of an hypergraph is *equivalent* to the enumeration of minimal dominating sets of a cobipartite graph. Interestingly, this result brings the dualization problem to the large graph theory community.

¹ It also works for infinite partially ordered sets that are well ordered, i.e. all antichains are finite.

Some theoretical frameworks for pattern mining have studied dualization, for instance [6–8]. In [8], we introduced the idea of weak representations as sets for pattern mining problems and showed how the frequent rigid sequences mining problem benefits from such representation. The embryo of a reduction for the dualization problem on arbitrary posets was present but was clearly implicit in [8].

To measure the complexity of enumeration algorithms, we always refer to the complexity in the size of the input and the size of the output, see [9] for details.

The objective of this paper is to study *reductions* of the dualization problem on arbitrary posets to the dualization problem on boolean lattices. On posets, the dualization problem can be stated as follows:

DualizeOnPoset

Input: A representation of a poset (P, \leq) , B^+ an antichain of P .

Output: B^- such that (B^+, B^-) are dual sets² in P .

Let us consider the particular instance of this problem:

DualizeOnSet

Input: A finite set E , B^+ an antichain of $\mathcal{P}(E)$ (the powerset of E).

Output: B^- such that (B^+, B^-) are dual sets in $\mathcal{P}(E)$.

As already mentioned, the complexity of **DualizeOnSet** is known to be quasi-polynomial while the complexity of **DualizeOnPoset** is still open in most posets (for example, lattice) [4]. In this setting, we are interested in studying the *reduction* from **DualizeOnPoset** to **DualizeOnSet**, i.e. under which conditions **DualizeOnSet** is at least as hard as **DualizeOnPoset**. Notice that reductions for the hardness of enumeration problems are not well established as for decision problems. In this paper, we consider only polynomial time reduction as explained in Fig. 1 which is inspired from classical polynomial reduction of decision problems.

Contribution on Dualization: We introduce *convex embedding* and *poset reflection* as key notions to characterize such reductions. As a consequence, we identify posets, which are not boolean lattices, for which the dualization problem remains quasi-polynomial and propose a classification of posets with respect to dualization.

2 Preliminaries

We briefly recall definitions on partial orders, embeddings and borders [8, 10].

A partial order is a binary relation \leq over a set P which is reflexive, anti-symmetric, and transitive. Let x, y be elements of P , if $x \leq y$ or $y \leq x$, then x and y are comparable, otherwise they are incomparable. A partial order under which every pair of elements is comparable is called a *chain*. A subset of a poset in which no two distinct elements are comparable is called an *antichain*. We say that y *covers* x if whenever $x \leq z \leq y$ then $z = x$ or $z = y$; we denote by \prec the covering relation. For $S \subseteq P$, $\downarrow S$ (resp. $\uparrow S$) is the downward (resp. upward)

² Dual sets are also known as blocker and anti-blocker or positive and negative borders.

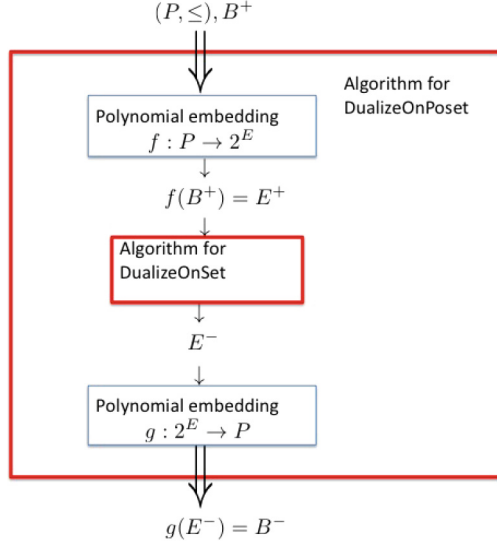


Fig. 1. Reduction from DualizeOnPoset to DualizeOnSet

closed set of S under the relation \leq (i.e. $\downarrow S$ is an ideal and $\uparrow S$ a filter of (P, \leq)). In case of ambiguity, $\downarrow S$ (resp. $\uparrow S$) will be denoted by $\downarrow^{\leq} S$ (resp. $\uparrow^{\leq} S$). A subset $S \subseteq P$ is *convex* in P if for all $x, y, z \in P$, $x, y \in S$ and $x \leq z \leq y$ implies $z \in S$. We denote by $\text{Max}_{\leq}(S)$ (resp. $\text{Min}_{\leq}(S)$) denotes the maximal (resp. minimal) elements of S with respect to \leq . When \leq is clear from context, (P, \leq) (resp. $\text{Max}_{\leq}(S)$ and $\text{Min}_{\leq}(S)$) will be denoted by P (resp. $\text{Max}(S)$, $\text{Min}(S)$).

Let (P, \leq_P) and (Q, \leq_Q) be posets and $f : P \rightarrow Q$ a mapping (total function). f is an *embedding* if for all $x, y \in P$, $x \leq_P y$ iff $f(x) \leq_Q f(y)$. The mapping f is an *isomorphism* if f is a bijective embedding. In this case P and Q are said to be *isomorphic*. f is a *convex embedding* if f is an *embedding* and $f(P)$ is convex in (Q, \leq_Q) . Whenever f is injective but not surjective, there exists another mapping $g : f(P) \rightarrow P$ such that $g \circ f = \text{Id}$, the identity function. A *reflection* of a poset (P, \leq) is a poset (P, \leq') on the same ground set P such that for all $x, y \in P$, $x \leq' y \Rightarrow x \leq y$.

Two antichains $(\mathcal{B}^+, \mathcal{B}^-)$ of P are said to be *dual* if $\downarrow \mathcal{B}^+ \cup \uparrow \mathcal{B}^- = P$ and $\downarrow \mathcal{B}^+ \cap \uparrow \mathcal{B}^- = \emptyset$. The relationship between these dual sets is known as the dualization, i.e. given \mathcal{B}^+ , compute \mathcal{B}^- (or inversely). In the sequel, $(\mathcal{B}^+, \mathcal{B}^-)$ will be referred to as a “border”.

Let $f : P \rightarrow Q$ be a mapping and $(\mathcal{B}^+, \mathcal{B}^-)$ a border in P . The border $(\mathcal{E}^+, \mathcal{E}^-)$ in Q is an *extension* of $(\mathcal{B}^+, \mathcal{B}^-)$ with respect to f , if $f(\mathcal{B}^+) \subseteq \mathcal{E}^+$ and $f(\mathcal{B}^-) \subseteq \mathcal{E}^-$. The extension $(\mathcal{E}^+, \mathcal{E}^-)$ is said to be a *polynomial extension* of $(\mathcal{B}^+, \mathcal{B}^-)$ if $|\mathcal{E}^+| + |\mathcal{E}^-|$ is polynomial in $|\mathcal{B}^+| + |\mathcal{B}^-|$.

The intuition of the reduction of enumeration problems used in this paper is based on finding a mapping between posets such that borders are polynomial preserved, i.e. every border has a polynomial extension.

In the rest of this paper, we assume that a poset is given by an implicit representation \mathcal{L} and we shall denote by (\mathcal{L}^*, \preceq) the poset defined by \mathcal{L} . Clearly, the size of \mathcal{L}^* may be exponential in the size of the representation \mathcal{L} . For instance, the free monoid Σ^* ordered by sub word is implicitly defined by the alphabet Σ , and a lattice is implicitly defined by its poset of (join and meet) irreducible elements [11].

3 Classification of Posets with Respect to Dualization

In this section we describe two properties of posets that lead us to have polynomial time reductions to **DualizeOnSet**. First we show that a convex embedding from a poset (\mathcal{L}^*, \preceq) to $\mathcal{P}(E)$ for some set E is sufficient to re-use algorithms of **DualizeOnSet**. Second, we show that the convex embedding is not a necessary condition and introduce the reflection of a poset (\mathcal{L}^*, \preceq) to obtain a new poset $(\mathcal{L}^*, \preceq')$ in which there is a convex embedding. Indeed, a reflection of a poset (\mathcal{L}^*, \preceq) corresponds to an embedding which preserves incomparabilities (or antichains), even if some comparabilities could be lost. The previous two embeddings introduce extra-elements to the dualization. Whenever these extra-elements are bounded by a polynomial, the dualization can be polynomial reduced to **DualizeOnSet**. To do so, we ask the following questions:

Given a poset of patterns (\mathcal{L}^, \preceq) ,*

- *Does there exist a convex embedding of (\mathcal{L}^*, \preceq) into $(\mathcal{P}(E), \subseteq)$ for some finite set E ? If not,*
- *Does there exist a reflection $(\mathcal{L}^*, \preceq')$ of (\mathcal{L}^*, \preceq) such that there exists a convex embedding of $(\mathcal{L}^*, \preceq')$ into $(\mathcal{P}(E), \subseteq)$ for some finite set E ?*

These two questions and their associated computational costs allow to come up with new classes of posets with respect to dualization. Figure 2 gives an illustration with a diagram where posets and borders are placed side by side.

3.1 Convex Embedding

First, let us recall that any poset has an embedding into a boolean lattice.

Proposition 1 [10]. For any poset (\mathcal{L}^*, \preceq) , there exists an embedding from (\mathcal{L}^*, \preceq) to $(\mathcal{P}(E), \subseteq)$, for some finite set E .

It follows that any poset has a set representation but obviously the dualization on (\mathcal{L}^*, \preceq) may be much more complex than the dualization on $(\mathcal{P}(E), \subseteq)$ [8]. We define the **RAS** class as follows:

Definition 1. $(\mathcal{L}^*, \preceq) \in \mathbf{RAS}$ iff (\mathcal{L}^*, \preceq) and $(\mathcal{P}(E), \subseteq)$ are isomorphic, for some finite set E .

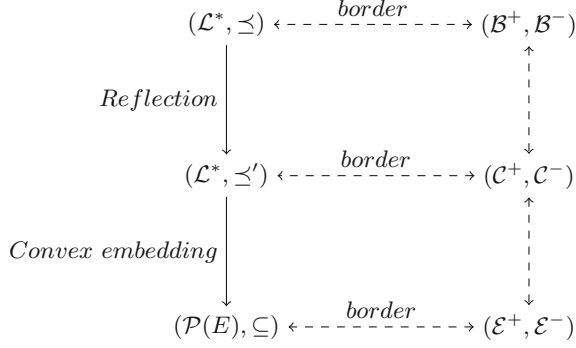


Fig. 2. Reflection and convex embedding

This class of posets gathers together many patterns such as frequent itemsets (FIM) [12], functional dependencies (FD) [13], inclusion dependencies (IND) [14]. This class is known as the *representation as sets* class of pattern mining problems defined in [6].

Nevertheless, requirements to be in \mathcal{RAS} are restrictive, since the poset must be isomorphic to a boolean lattice, and then its size has to be equal to 2^n where $n = |E|$. Now we will relax the bijective constraint of \mathcal{RAS} but we keep the convexity property on the set representation. Hence, we extend \mathcal{RAS} to a new class, called \mathcal{XRAS} , for *convex* \mathcal{RAS} .

Definition 2. $(\mathcal{L}^*, \preceq) \in \mathcal{XRAS}$ iff there exists a convex embedding from (\mathcal{L}^*, \preceq) to $(\mathcal{P}(E), \subseteq)$, for some finite set E .

The idea is still to require an isomorphism but just between the poset of patterns and some subset of $\mathcal{P}(E)$, instead of the entire set $\mathcal{P}(E)$ (see Fig. 3). Note also that f is injective since f is an embedding. The following proposition points out a simple yet important characterization of \mathcal{XRAS} problems.

Proposition 2. $(\mathcal{L}^*, \preceq) \in \mathcal{XRAS}$ iff (\mathcal{L}^*, \preceq) is isomorphic to $(\mathcal{P}(E) \setminus (\downarrow B_0^+ \cup \uparrow B_0^-), \subseteq)$ for some antichains $B_0^+ \subseteq \mathcal{P}(E)$ and $B_0^- \subseteq \mathcal{P}(E)$.

Proof. Let f be a convex embedding from (\mathcal{L}^*, \preceq) to $(\mathcal{P}(E), \subseteq)$ and $\mathcal{F} = f(\mathcal{L}^*)$. Let us consider $\mathcal{F}^+ = \text{Min}(\mathcal{F})$ and $\mathcal{F}^- = \text{Max}(\mathcal{F})$ two antichains of \mathcal{F} . Moreover, let $B_0^+ = \text{Max}(\{X \in \mathcal{P}(E) \mid X \subset Y, Y \in \mathcal{F}^+\})$ and $B_0^- = \text{Min}(\mathcal{P}(E) \setminus (\mathcal{F} \cup \downarrow B_0^+))$. Let $X \in \mathcal{P}(E)$. Then either $X \in \downarrow B_0^+$, or $X \in \uparrow B_0^-$ or $X \in \mathcal{P}(E) \setminus (\downarrow B_0^+ \cup \uparrow B_0^-)$. In the latter case, there exists $Y_1 \in \mathcal{F}^+, Y_2 \in \mathcal{F}^-$ such that $Y_1 \subseteq X \subseteq Y_2$. Since \mathcal{F} is convex, $X \in \mathcal{F}$ and the result follows.

The other direction holds since $\mathcal{P}(E) \setminus (\downarrow B_0^+ \cup \uparrow B_0^-)$ is a convex set of $\mathcal{P}(E)$. \square

Figure 3 gives an illustration of the Proposition 2.

Note that the sets B_0^+ and B_0^- can be exponential in the size of E . The next definition introduces *efficient* problems of \mathcal{XRAS} , called \mathcal{EXRAS} .

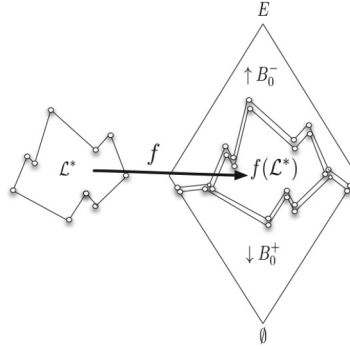


Fig. 3. The class \mathcal{XRAS}

Definition 3. $(\mathcal{L}^*, \preceq) \in \mathcal{EXRAS}$ if $(\mathcal{L}^*, \preceq) \in \mathcal{XRAS}$ and $|B_0^+ \cup B_0^-|$ is polynomial in $|E|$.

The following proposition points out that a polynomial extension of any border of (\mathcal{L}^*, \preceq) exists if $(\mathcal{L}^*, \preceq) \in \mathcal{EXRAS}$.

Proposition 3. Let $(\mathcal{L}^*, \preceq) \in \mathcal{EXRAS}$ and $f : \mathcal{L}^* \rightarrow \mathcal{P}(E)$ a convex embedding, for some finite set E .

Then for any border $(\mathcal{B}^+, \mathcal{B}^-)$ of (\mathcal{L}^*, \preceq) , $(\text{Max}(B_0^+ \cup f(\mathcal{B}^+)), \text{Min}(B_0^- \cup f(\mathcal{B}^-)))$ is a polynomial extension of $(\mathcal{B}^+, \mathcal{B}^-)$ in $(\mathcal{P}(E), \subseteq)$.

Proof. It suffices to notice that $(\mathcal{B}^+, \mathcal{B}^-)$ is a border of (\mathcal{L}^*, \preceq) iff $(f(\mathcal{B}^+), f(\mathcal{B}^-))$ is a border of $(\mathcal{P}(E) \setminus (\downarrow B_0^+ \cup \uparrow B_0^-), \subseteq)$ since (\mathcal{L}^*, \preceq) is isomorphic to $(\mathcal{P}(E) \setminus (\downarrow B_0^+ \cup \uparrow B_0^-), \subseteq)$. \square

3.2 Polynomial Reflection of Posets

We now consider posets that are not in \mathcal{XRAS} . Our idea is to transform the initial poset to a new poset over the same ground set, in order to get a convex embedding. As a consequence, two natural question arise:

- (1) For a given poset (\mathcal{L}^*, \preceq) , does there exist a “polynomial reflection” $(\mathcal{L}^*, \preceq')$ such that $(\mathcal{L}^*, \preceq')$ belongs to \mathcal{EXRAS} ?
- (2) How to quantify the “lost comparabilities” induced by a reflection?

In the sequel, we study poset reflection to give answers to the previous questions. Since the reflection of a poset induces the lost of some comparabilities in the original poset, we have to recover them efficiently.

Before that, we consider different examples of posets over sequences.

Examples with Different Posets of Sequences. Let us consider sequences with or without wildcard (denoted \star), see e.g. [15].

Let Σ be an alphabet. A sequence is an element of Σ^* and a rigid sequence an element of $(\Sigma \cup \{\star\})^*$ of the form $P = P[1] \cdots P[m]$ such that $P[1] \neq \star$ and $P[m] \neq \star$. Let Σ_R^* be the set of rigid sequences and Σ^* the set of sequences. We denote by Σ^n the set of all sequences in Σ^* of size at most n .

Different partial orders over Σ_R^* and Σ^* exist. Let us first consider sub-word (resp. factor and prefix) posets over Σ^* , denoted (Σ^*, \leq_s) (resp. (Σ^*, \leq_f) and (Σ^*, \leq_p)). Let $P[1..m], Q[1..n] \in \Sigma^*$. We have:

- $P \leq_s Q$ if there exists integers $i_1 < \dots < i_m$ in $[1..n]$ such that $P[j] = Q[i_j]$ for all $j \in [1..m]$.
- $P \leq_f Q$ if $P \leq_s Q$ and $i_j = i_{j+1} - 1$ for all $j \in [1..m-1]$
- $P \leq_p Q$ if $P \leq_f Q$ and $i_m = m$.

These different posets are illustrated in Fig. 4 on a simple example.

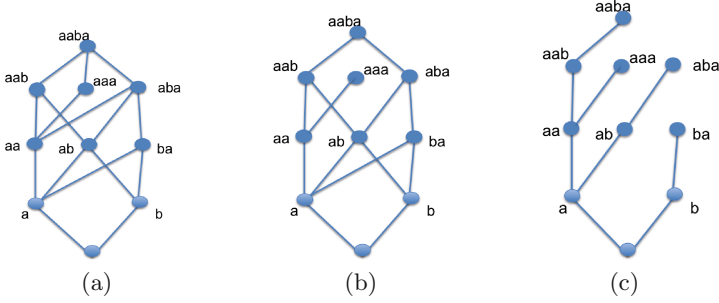


Fig. 4. (a) The sub word poset induced by the word $w=aaba$ on $\Sigma = \{a, b\}$; (b) the factor poset which is a reflection of (a); and (c) the prefix poset which is a reflection of posets (b) and (a).

In Fig. 4, let us consider the set $\{aa, ab\}$ for the different posets. Its dual set is equal to $\{ba, aab, aaa\}$ for (a) and (b) and $\{b, aab, aaa, aba\}$ for (c).

Second, we consider two posets for rigid sequences, one similar to the factor poset and another one to the prefix poset, denoted $(\Sigma_R^*, \sqsubseteq)$ and $(\Sigma_R^*, \sqsubseteq_1)$. Let $P[1..m], Q[1..n] \in \Sigma_R^*$. We have:

- $P \sqsubseteq Q$ if there exists $p \in [1..n]$ such that for every $i \in [1..m]$, either $P[i] = Q[p+i-1]$ or $P[i] = \star$
- $P \sqsubseteq_1 Q$ if for every $i \in [1..m]$, either $P[i] = Q[i]$ or $P[i] = \star$

These two posets are illustrated in Fig. 5, where the poset (b) suggests the existence of hidden hypercubes in the poset (a).

Now, we claim that the posets (Σ^*, \leq_s) , (Σ^*, \leq_f) , (Σ^*, \leq_p) and $(\Sigma_R^*, \sqsubseteq)$ do not have any convex embedding. Consider again the example given in Fig. 4 and the two following sets: $A = \{a, b, ab, ba\}$ and $A' = \{a, aa, aaa\}$.

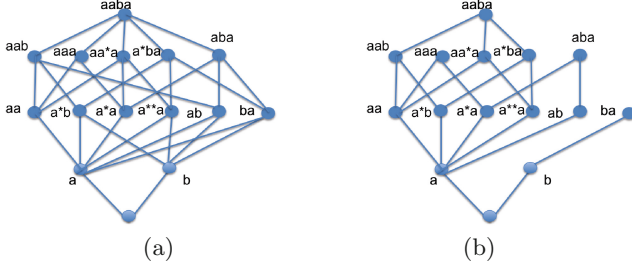


Fig. 5. (a) The factor poset of rigid sequences induced by the word $w=aaba$ on $\Sigma = \{a, b\}$; (b) the prefix poset of rigid sequences which is a reflection of posets (a).

- A is convex in (Σ^*, \leq_s) (Fig. 4(a)) but its image by any embedding cannot be convex in $(\mathcal{P}(E), \subseteq)$ since $(\mathcal{P}(E), \subseteq)$ is a lattice. The same reasoning applies for (Σ^*, \leq_f) and (Σ_R^*, \subseteq) .
- A' is convex in (Σ^*, \leq_p) but its image by any embedding cannot be convex in $(\mathcal{P}(E), \subseteq)$ since $\mathcal{P}(E)$ cannot contain a convex set which is a chain of length 3.

However, Fig. 5 also shows a reflection that leads to a convex embedding. Consider $(\Sigma_R^n, \subseteq_1)$. Let $f : \Sigma_R^n \setminus \{\epsilon\} \rightarrow \mathcal{P}(E)$ be an embedding, for some finite set E [8, 15]: f associates to each letter of a sequence a couple (indice, letter). For instance, let ab and ba be two patterns. Then $f(ab) = \{(1, a), (2, b)\}$ and $f(ba) = \{(1, b), (2, a)\}$. It is easy to verify that $f(\Sigma_R^n \setminus \{\epsilon\})$ is convex in $(\mathcal{P}(E), \subseteq)$ [8]. Let us again consider $A = \{a, b, ab, ba\}$: we only have $a \subseteq_1 ab, b \subseteq_1 ba$, i.e. two comparabilities ($a \not\subseteq_1 ba, b \not\subseteq_1 ab$) are lost, allowing to reach the convexity constraint.

These examples point out that we have to study poset reflection to be able to obtain some convex embedding.

3.3 Reaching Convexity by Poset Reflection

As shown in previous examples involving sequences, whenever possible, we have now to quantify the lost comparabilities induced by a poset reflection.

For a given element of a poset, we define its successors and its predecessors induced by a poset reflection.

Definition 4. Let (P, \leq') a reflection of a poset (P, \leq) and $x \in P$. The lost predecessors of x in the reflection of (P, \leq) to (P, \leq') , denoted by $\text{LostPred}(x)$, are defined by:

$\text{LostPred}(x) = \text{Max}_{\leq'} \{y \in P \mid y \leq x, y \not\leq' x\}$. Similarly, the lost successors are defined by: $\text{LostSucc}(x) = \text{Min}_{\leq'} \{y \in P \mid x \leq y, x \not\leq' y\}$.

By extension, we note $\text{LostPred}(X) = \bigcup_{x \in X} \text{LostPred}(x)$ (resp. $\text{LostSucc}(X)$) for $X \subseteq P$.

Example 1. Let us consider the reflection $(\Sigma_R^n, \sqsubseteq_1)$ of $(\Sigma_R^n, \sqsubseteq)$ [8]. Let $S \in \Sigma_R^n$. We have $\text{LostPred}(S) = \{S[i..|S|] \mid 1 \leq i \leq |S|, S[i] \neq \star\}$ and $\text{LostSucc}(S) = \{x \underbrace{\star \dots \star}_i S \mid 0 \leq i \leq n - |S| - 1, x \in \Sigma\}$.

For instance with $n = 5$, $\text{LostPred}(a \star ba) = \{a \star ba, ba, a\}$ and $\text{LostSucc}(ba) = \{aba, bba, a \star ba, b \star ba, a \star \star ba, b \star \star ba\}$.

As shown in the following lemma, we can recover the initial poset from any reduced poset with LostPred and LostSucc .

Lemma 1. *Let $x \in P$ and (P, \leq') a reflection of (P, \leq) . Then:*

1. $\downarrow \leq x = \downarrow \leq' \text{LostPred}(x)$ and
2. $\uparrow \leq x = \uparrow \leq' \text{LostSucc}(x)$.

Proof. (1) Let $y \in \downarrow \leq x$. We have either $y \leq' x$ or $y \not\leq' x$. If $y \leq' x$, then $y \in \downarrow \leq' \text{LostPred}(x)$ since $x \in \text{LostPred}(x)$. If $y \not\leq' x$, then there exists $z \in P$ such that $y \leq' z$, $z \in \text{LostPred}(x)$, i.e. $y \in \downarrow \leq' \text{LostPred}(x)$. The same reasoning applies for (2). \square

Some remarks have to be made: First, for any poset, there always exists a reflection that has a convex embedding into a boolean lattice. It suffices to take a reflection which is an antichain, i.e. that deletes all comparabilities. In this case, the number of lost comparabilities can be exponential in the size of the description of the poset. Second, we would like to be able to recover lost comparabilities in polynomial time. This is formalized with the notion of *poly-reflection* as follows.

Definition 5. (\mathcal{L}^*, \leq') is a poly-reflection of (\mathcal{L}^*, \leq) if (\mathcal{L}^*, \leq') is a reflection of (\mathcal{L}^*, \leq) and for all $x \in \mathcal{L}^*$, $\text{LostPred}(x)$ and $\text{LostSucc}(x)$ are computable in polynomial time in the size of the description \mathcal{L} .

Example 2. Continuing the previous example, for all $S \in \Sigma_R^n$, $\text{LostPred}(S)$ is polynomial in n and for all $s \sqsubseteq S$, there exists $s' \in \text{LostPred}(S)$ such that $s \sqsubseteq_1 s'$. Therefore, $(\Sigma_R^n, \sqsubseteq_1)$ is a poly-reflection of $(\Sigma_R^n, \sqsubseteq)$.

Now we show the relationship between borders in a poset and its reflection. For a given border on the initial poset, we define its extension in the reduced poset to take into account lost comparabilities.

Definition 6. Let (\mathcal{L}^*, \leq') be a poly-reflection of (\mathcal{L}^*, \leq) and $(\mathcal{B}^+, \mathcal{B}^-)$ a border of (\mathcal{L}^*, \leq) . The extension of $(\mathcal{B}^+, \mathcal{B}^-)$ in (\mathcal{L}^*, \leq') , denoted by $(\text{ext}(\mathcal{B}^+), \text{ext}(\mathcal{B}^-))$, is defined by:

$$\begin{aligned} \text{ext}(\mathcal{B}^+) &= \text{Max}_{\leq'} \{\text{LostPred}(x) \mid x \in \mathcal{B}^+\} \\ \text{ext}(\mathcal{B}^-) &= \text{Min}_{\leq'} \{\text{LostSucc}(x) \mid x \in \mathcal{B}^-\}. \end{aligned}$$

The “preservation” of borders can now be formally stated.

Proposition 4. Let (\mathcal{L}^*, \leq') be a poly-reflection of (\mathcal{L}^*, \leq) and $(\mathcal{B}^+, \mathcal{B}^-)$ a border of (\mathcal{L}^*, \leq) . Then $(\text{ext}(\mathcal{B}^+), \text{ext}(\mathcal{B}^-))$ is a polynomial extension of $(\mathcal{B}^+, \mathcal{B}^-)$.

Proof. We have to show:

1. $(\mathbf{ext}(\mathcal{B}^+), \mathbf{ext}(\mathcal{B}^-))$ is a border of $(\mathcal{L}^*, \preceq')$ with $\mathcal{B}^+ \subseteq \mathbf{ext}(\mathcal{B}^+)$ and $\mathcal{B}^- \subseteq \mathbf{ext}(\mathcal{B}^-)$,
 2. $|\mathbf{ext}(\mathcal{B}^+)| + |\mathbf{ext}(\mathcal{B}^-)|$ is polynomial in $|\mathcal{B}^+| + |\mathcal{B}^-|$.
- (1) Any reflection preserves all incomparabilities and $x \in \mathbf{LostPred}(x)$ (resp $x \in \mathbf{LostSucc}(x)$) for all $x \in \mathcal{L}^*$. Then $\mathcal{B}^+ \subseteq \mathbf{ext}(\mathcal{B}^+)$ and $\mathcal{B}^- \subseteq \mathbf{ext}(\mathcal{B}^-)$. By definition, $\mathbf{ext}(\mathcal{B}^+)$ and $\mathbf{ext}(\mathcal{B}^-)$ are antichains in $(\mathcal{L}^*, \preceq')$. By Lemma 1, the result follows.
- (2) $|\mathbf{ext}(\mathcal{B}^+)| + |\mathbf{ext}(\mathcal{B}^-)|$ is polynomial in $|\mathcal{B}^+| + |\mathcal{B}^-|$ since $(\mathcal{L}^*, \preceq')$ is a poly-reflection of (\mathcal{L}^*, \preceq) since computing $\mathbf{LostPred}(x)$ and $\mathbf{LostSucc}(x)$ can be done in polynomial time in the size of the description of \mathcal{L} . \square

The notion of poly-reflection allows to define the last class of posets, called \mathcal{EWRAS} , meaning *Efficient weak representation as sets*. \mathcal{EWRAS} is the more general class ensuring the existence of quasi-polynomial algorithms. It combines both poly-reflection of posets and \mathcal{EXRAS} .

Definition 7. $(\mathcal{L}^*, \preceq) \in \mathcal{EWRAS}$ iff there exists a poly-reflection $(\mathcal{L}^*, \preceq')$ of (\mathcal{L}^*, \preceq) such that $(\mathcal{L}^*, \preceq') \in \mathcal{EXRAS}$.

Then, this definition means that if some comparabilities can be forgotten – up to a polynomial cost to recover them – to get a new poset satisfying the condition of \mathcal{EXRAS} , then the dualization problem on the initial poset can be reduced to **DualizationOnSet**.

Example 3. Continuing previous examples, we have $(\Sigma_R^*, \sqsubseteq_1)$ is a poly-reflection of $(\Sigma_R^*, \sqsubseteq)$ and $(\Sigma_R^*, \sqsubseteq_1)$ belongs to \mathcal{EXRAS} . Then, $(\Sigma_R^*, \sqsubseteq)$ belongs to \mathcal{EWRAS} .

The main result concerning the \mathcal{EWRAS} class is now given:

Theorem 1. Let $(\mathcal{L}^*, \preceq) \in \mathcal{EWRAS}$. Assume that $(\mathcal{L}^*, \preceq')$ is a poly-reflection of (\mathcal{L}^*, \preceq) such that $(\mathcal{L}^*, \preceq')$ belongs to \mathcal{EXRAS} , i.e. $(\mathcal{L}^*, \preceq')$ isomorphic to $\mathcal{P}(E) \setminus (B_0^+ \cup B_0^-)$. Then, for any border $(\mathcal{B}^+, \mathcal{B}^-)$ of (\mathcal{L}^*, \preceq) :

- $(B_0^+ \cup f(\mathbf{ext}(\mathcal{B}^+)), B_0^- \cup f(\mathbf{ext}(\mathcal{B}^-)))$ is a border in $(\mathcal{P}(E), \subseteq)$,
- $(B_0^+ \cup f(\mathbf{ext}(\mathcal{B}^+)), B_0^- \cup f(\mathbf{ext}(\mathcal{B}^-)))$ is a polynomial extension of $(\mathcal{B}^+, \mathcal{B}^-)$,

The following corollary gives the relationship between all the classes introduced so far.

Corollary 1. Let $(\mathcal{L}^*, \preceq')$ be a poly-reflection of (\mathcal{L}^*, \preceq) such that $(\mathcal{L}^*, \preceq')$ belongs to \mathcal{XRAS} , i.e. $(\mathcal{L}^*, \preceq')$ isomorphic to $\mathcal{P}(E) \setminus (B_0^+ \cup B_0^-)$. We have:

1. $(\mathcal{L}^*, \preceq) \in \mathcal{RAS}$ if $(\mathcal{L}^*, \preceq) = (\mathcal{L}^*, \preceq')$ and $B_0^+ = B_0^- = \emptyset$.
2. $(\mathcal{L}^*, \preceq) \in \mathcal{XRAS}$ if $(\mathcal{L}^*, \preceq) = (\mathcal{L}^*, \preceq')$.
3. $(\mathcal{L}^*, \preceq) \in \mathcal{EXRAS}$ if $(\mathcal{L}^*, \preceq) = (\mathcal{L}^*, \preceq')$ and the size of $B_0^+ \cup B_0^-$ is polynomial in the size of the description of (\mathcal{L}^*, \preceq) .
4. $(\mathcal{L}^*, \preceq) \in \mathcal{EWRAS}$ if the size of $B_0^+ \cup B_0^-$ is polynomial in the size of the description of (\mathcal{L}^*, \preceq) .

3.4 DualizeOnSeq is Equivalent to DualizeOnSet

Recall that we consider rigid sequences only. The dualization problem can be stated as follows:

DualizeOnSeq

Input: Σ a totally ordered alphabet with a minimal element \star , n a positive integer, B^+ a positive border of Σ_R^n .

output: B^- such that (B^+, B^-) is a border of Σ_R^n .

We have already showed that **DualizeOnSeq** is at least as hard as **DualizeOnSet** in [8]. In the sequel, we point out that **DualizeOnSet** is at least as hard as **DualizeOnSeq**, and therefore the two problems are polynomially equivalent. Indeed, we show that **DualizeOnSet** is a particular case of **DualizeOnSeq**.

Let $\Sigma = \{1, 2, \dots, n, \star\}$ be an ordered alphabet (i.e. $\star < 1 < 2 \dots < n$ and $S \in \Sigma^n$). The sequence S is said to be an *ordered sequence* if for any $i, j \in [1..n]$ such that $i < j$, $S[i] \neq \star$ and $S[j] \neq \star$ we have $S[j] - S[i] = j - i$. We denote $\Sigma_O^n \subseteq \Sigma^n$ the set of all ordered sequences of size at most n . For example, the sequence $2 \star 5$ is an ordered sequence but $2 \star 5$ is not.

The following lemma characterizes ordered sequences.

Lemma 2. *Let $\Sigma = \{1, 2, \dots, n, \star\}$ be an ordered alphabet and $S \in \Sigma^n$. Then $S \in \Sigma_O^n$ iff S does not contain a subsequence of the form $i \underbrace{\star \dots \star}_k j$ with either $i < j$ and $k \neq j - i - 1$, or $i \geq j$ with $k \in [0..n - 2]$.*

Consider the set $B_0^- = \{i \star^k j \mid i \geq j, k \in [0..n - 2]\} \cup \{i \star^k j \mid i < j, k \in [0..n - 2], k \neq j - i - 1\}$. For example for $\Sigma = \{1, 2, 3, \star\}$ we have $B_0^- = \{11, 1 \star 1, 22, 2 \star 2, 33, 3 \star 3, 21, 2 \star 1, 31, 3 \star 1, 32, 3 \star 2\} \cup \{13\}$.

Lemma 3. $\Sigma^n \setminus \uparrow B_0^- = \Sigma_O^n$.

Let $V = \{1, \dots, n\}$ be a set. We define the mapping $f : \mathcal{P}(V) \rightarrow \Sigma^n$ such that for any $E \in \mathcal{P}(V)$, $f(E) = S$ with $S[i] = i$ if $i \in E$ and $S[i] = \star$ otherwise. Without loss of generality, we delete the symbols \star that are prefix or suffix of $f(E)$. Note that $f(E)[i] = \star$ means that $i \notin E$. For example $f(\{2, 5\}) = 2 \star \star 5$ and $f(\{\})$ is the empty sequence.

Proposition 5. *Let $V = \{1, \dots, n\}$ be a set. Then the mapping f is a convex embedding of $\mathcal{P}(V)$ into Σ^n . Moreover $f(\mathcal{P}(V)) = \Sigma_O^n$.*

Proof. Let P, Q be two sequences that are images of two sets $A \subset B \subseteq V$, i.e. $f(A) = P$ and $f(B) = Q$. Clearly $f(A) \sqsubset f(B)$.

Now suppose there is a sequence S such that $P \sqsubset S \sqsubset Q$.

For every $i \neq j \in [1..n]$ we have either $S[i] \neq S[j]$ or $S[i] = S[j] = \star$, by definition of the embedding f . Then the set $C = \{x \in V \mid S[i] = x, i \in [1..n]\}$ is clearly defined. Moreover $f(C) = S$ and $A \subset C \subset B$, since for any $x \in \Sigma$, $x \not\sqsubset \star$, but $\star \sqsubset x$.

We have $f(\mathcal{P}(V)) = \Sigma_O^n$ by construction. \square

Theorem 2. *DualizeOnSeq and DualizeOnSet are polynomially equivalent.*

Proof. We have shown that **DualizeOnSeq** is polynomially reducible to **DualizeOnSet** [8]. In Proposition 5, we have shown the existence of a convex embedding from $\mathcal{P}(V)$ into Σ^n . Moreover we have shown that $B_0^+ = \emptyset$ and according to Lemma 2 the size of B_0^- is bounded by $O(n^3)$. Thus **DualizeOnSet** is polynomially reducible to **DualizeOnSeq**. \square

We have shown that the dualization on rigid sequences with wildcard is equivalent to the dualization on set, i.e. enumerating minimal transversals of a given hypergraph.

References

1. Eiter, T., Gottlob, G.: Identifying the minimal transversals of a hypergraph and related problems. *SIAM J. Comput.* **24**(6), 1278–1304 (1995)
2. Fredman, M.L., Khachiyan, L.: On the complexity of dualization of monotone disjunctive normal forms. *J. Algorithms* **21**(3), 618–628 (1996)
3. Eiter, T., Gottlob, G., Makino, K.: New results on monotone dualization and generating hypergraph transversals. *SIAM J. Comput.* **32**, 514–537 (2003)
4. Elbassioni, K.M.: Algorithms for dualization over products of partially ordered sets. *SIAM J. Discrete Math.* **23**(1), 487–510 (2009)
5. Kanté, M.M., Limouzy, V., Mary, A., Nourine, L.: On the enumeration of minimal dominating sets and related notions. Revised version submitted (2013)
6. Mannila, H., Toivonen, H.: Levelwise search and borders of theories in knowledge discovery. *Data Min. Knowl. Discov.* **1**(3), 241–258 (1997)
7. Gunopulos, D., Khardon, R., Mannila, H., Saluja, S., Toivonen, H., Sharm, R.S.: Discovering all most specific sentences. *ACM Trans. Database Syst.* **28**(2), 140–174 (2003)
8. Nourine, L., Petit, J.M.: Extending set-based dualization: application to pattern mining. In: Press, I. (ed.) *ECAI 2012*, August 2012
9. Elbassioni, K.: Incremental algorithms for enumerating extremal solutions of monotone systems of submodular inequality and their applications. Ph.D. thesis, Rutgers, The state university of New Jersey (2002)
10. Davey, B.A., Priestley, H.A.: *Introduction to Lattices and Order*. Cambridge Press, New York (1990)
11. Ganter, B., Wille, R.: *Formal Concept Analysis*. Springer, Heidelberg (1999)
12. Agrawal, R., Imielinski, T., Swami, A.: Mining associations between sets of items in massive databases. In: *ACM SIGMOD 1993*, Washington D.C., pp. 207–216 (1993)
13. Mannila, H., Rih, K.J.: Algorithms for inferring functional dependencies from relations. *Data Knowl. Eng.* **12**(1), 83–99 (1994)
14. De Marchi, F., Petit, J.M.: Zigzag: a new algorithm for mining large inclusion dependencies in databases. In: *ICDM 2003*, USA, pp. 27–34, November 2003
15. Arimura, H., Uno, T.: Polynomial-delay and polynomial-space algorithms for mining closed sequences, graphs, and pictures in accessible set systems. In: *SDM*, pp. 1087–1098 (2009)

Information Search, Integration and Personalization
9th International Workshop, ISIP 2014, Kuala Lumpur,
Malaysia, October 9-10, 2014, Revised Selected Papers
Kotzinos, D.; Choong, Y.W.; Spyratos, N.; Tanaka, Y.
(Eds.)
2016, XII, 113 p. 26 illus., Softcover
ISBN: 978-3-319-38900-4