Some Notes on Managing Closure Operators

Sebastian Rudolph

Karlsruhe Institute of Technology, Germany
rudolph@kit.edu

Abstract. It is widely known that closure operators on finite sets can be represented by sets of implications (also known as inclusion dependencies) as well as by formal contexts. In this paper we survey known results and present new findings concerning time and space requirements of diverse tasks for managing closure operators, given in contextual, implicational, or black-box representation. These tasks include closure computation, size minimization, finer-coarser-comparison, modification by "adding" closed sets or implications, and conversion from one representation into another.

1 Introduction

Closure operators and closure systems are a basic notion in algebra and occur in various computer science scenarios such as logic programming or functional dependencies in databases. One central task when dealing with closure operators is to represent them in a succinct way while still allowing for their efficient computational usage. Formal concept analysis (FCA) naturally provides two complementary ways of representing closure operators: by means of *formal contexts* on one side and *implication sets* on the other. Although being complementary, these two representation modes share a lot of properties:

- Both allow for tractable closure computation.
- Both kinds of data structures do not uniquely represent the corresponding closure operator, but in either case, there is a well-known minimal "normal form" which is unique up to isomorphism: row-reduced formal contexts and canonical bases.
- In both cases, this normal form can be computed with polynomial effort.

For a given closure operator, the space needed to represent it in one or the other way may differ significantly: there are closure operators whose minimal implicational representation is exponentially larger than their minimal contextual one and vice versa (see Section 3).
Thus, it seems worthwhile to modify algorithms which store and manipulate closure operators (as many FCA algorithms do) such that they can switch between the two representations depending on which is more memory-efficient. To

this end, algorithms performing basic operations on closure operators need to be available for both representations. Moreover, conversion methods from one representation to the other are needed and their computational complexity needs to be analyzed. Thereby, it is not only interesting to determine the required resources related to the size of the input, but also to the size of the output. This is to account for the fact, that (in order to be "fair") an algorithm creating a larger output should be allowed to take more time and use more memory.
Next to surveying well-known complexity results for tasks related to closure operators in different representations, this paper's noteworthy original contributions are the following:

- We clarify the complexities for comparing closure operators in different representations in terms of whether one is a refinement of the other. Interestingly, some of the investigated comparison tasks are tractable (i.e. time-polynomial), others are not.
- We show how to compute an implication set which realizes the closure operator of a given context and has polynomial size compared to the size of the context. This is achieved by virtue of auxiliary attributes. Note that this contrasts results according to which without such auxiliary attributes, a worst-case exponential blow-up is unavoidable.
- Exploiting this polynomial representation, we propose an alternative algorithm for computing the Duquenne-Guigues base of a given context.

The paper is organized as follows. After recalling some basics about closure operators and formal concept analysis in Section 2, we note that no representation is generally superior to the other in terms of the size needed to store it in Section 3. Finally, Section 4 provides algorithms and complexity results before we conclude in Section 5.

2 Preliminaries

We start providing a condensed overview over the notions used in this paper.

2.1 Closure Operators

Definition 1. *Let M be an arbitrary set. A function $\varphi : 2^M \to 2^M$ is called a* closure operator *on M if it is*

1. extensive*, i.e., $A \subseteq \varphi(A)$ for all $A \subseteq M$,*
2. monotone*, i.e., $A \subseteq B$ implies $\varphi(A) \subseteq \varphi(B)$ for all $A, B \subseteq M$, and*
3. idempotent*, i.e., $\varphi(\varphi(A)) = \varphi(A)$ for all $A \subseteq M$.*

A set $A \subseteq M$ is called closed *(or* φ-closed *in case of ambiguity), if $\varphi(A) = A$. The set of all closed sets $\{A \mid A = \varphi(A) \subseteq M\}$ is called* closure system.

It is easy to show that for an arbitrary closure system $\mathcal{S}$, the corresponding closure operator φ can be reconstructed by

$$\varphi(A) = \bigcap_{B \in \mathcal{S},\ A \subseteq B} B.$$

Hence, there is a one-to-one correspondence between a closure operator and the according closure system.

Definition 2. *Given two closure operators φ and ψ on M, φ is called* finer than *ψ (written $\varphi \preceq \psi$, alternatively we also say ψ is* coarser than *φ) if every φ-closed set is also ψ-closed. We call φ and ψ* equivalent *(written $\varphi \equiv \psi$), if $\varphi(A) = \psi(A)$ for all $A \subseteq M$.*

It is well-known that the set of all closure operators together with the "finer than" relation constitutes a complete lattice.

2.2 Contexts

Following the normal line of argumentation of FCA [8], we use formal contexts as data structure to encode closure operators.

Definition 3. *A* formal context *$\mathbb{K}$ is a triple (G, M, I) with an arbitrary set G called* objects*, an arbitrary set M called* attributes*, and a relation $I \subseteq G \times M$ called* incidence relation*. The* size *of $\mathbb{K}$ (written: $\#\mathbb{K}$) is defined as $|G| \cdot |M|$, i.e. the number of bits to store I.*

This basic data structure can then be used to define operations on sets of objects or attributes, respectively.

Definition 4. *Let $\mathbb{K} = (G, M, I)$ be a formal context. We define a function $(\cdot)^I : 2^G \to 2^M$ with $\widetilde{G}^I := \{m \mid gIm \text{ for all } g \in \widetilde{G}\}$ for $\widetilde{G} \subseteq G$. Furthermore, we use the same notation to define the function $(\cdot)^I : 2^M \to 2^G$ where $\widetilde{M}^I := \{g \mid gIm \text{ for all } m \in \widetilde{M}\}$ for $\widetilde{M} \subseteq M$. For convenience, we sometimes write g^I instead of $\{g\}^I$ and m^I instead of $\{m\}^I$.*

Applied to an object set, this function yields all attributes common to these objects; by applying it to an attribute set we get the set of all objects having those attributes. The following facts are consequences of the above definitions:

- $(\cdot)^{II}$ is a closure operator on G as well as on M.

- For $A \subseteq G$, A^I is a $(\cdot)^{II}$-closed set and dually
- for $B \subseteq M$, B^I is a $(\cdot)^{II}$-closed set.

In the following, we will focus only on the closure operator on attribute sets and exploit the fact that this closure operator is independent from the concrete object set G; it suffices to know the set of the context's object intents. Thus, we will directly use intent sets, that is: families $\mathcal{F}$ of subsets of M to represent formal contexts.

Definition 5. *Given a family $\mathcal{F} \subseteq 2^M$, we let $\mathbb{K}(\mathcal{F})$ denote the formal context (G, M, I) with $G = \mathcal{F}$ and, for an $A \in \mathcal{F}$, we let AIm exactly if $m \in A$. Given $B \subseteq M$, we use the notation $B^{\mathcal{F}}$ to denote the attribute closure B^{II} in $\mathbb{K}(\mathcal{F})$ and let $\#\mathcal{F} = \#\mathbb{K}(\mathcal{F}) = |\mathcal{F}| \cdot |M|$.*

For the sake of simplicity we will from now on to refer to $\mathcal{F}$ as contexts (on M). We recall the first basic complexity result:

Proposition 1. *For any context $\mathcal{F}$ on a set M and any set $A \subseteq M$, the closure $A^{\mathcal{F}}$ can be computed in $O(\#\mathcal{F}) = O(|\mathcal{F}| \cdot |M|)$ time and $O(|M|)$ space.*

Given an arbitrary context $\mathcal{F}$ representing some closure operator φ on some set M, the question whether there exists another $\mathcal{F}'$ representing φ and satisfying $\#\mathcal{F}' < \#\mathcal{F}$ – and if so, how to compute it – is straightforwardly solved by noting that this coincides with the question if $\mathbb{K}(\mathcal{F})$ is row-reduced and how to row-reduce it. Hence we obtain:

Proposition 2. *Given a context $\mathcal{F}$ on M, a size-minimal context $\mathcal{F}'$ with $(\cdot)^{\mathcal{F}} \equiv (\cdot)^{\mathcal{F}'}$ can be computed in $O(|\mathcal{F}| \cdot \#\mathcal{F}) = O(|\mathcal{F}|^2 \cdot |M|)$ time and $O(|M|)$ space.*

Algorithm 1 displays the according method cast in our representation via set families.
We close this section by noting that for a given closure operator φ, the minimal $\mathcal{F}$ with $\varphi \equiv (\cdot)^{\mathcal{F}}$ is uniquely determined. We will denote it by $\mathcal{F}(\varphi)$.

2.3 Implications

Given a set of attributes, *implications* on that set are logical expressions that can be used to describe certain attribute correspondences which are valid for all objects in a formal context.

Definition 6. *Let M be an arbitrary set. An* implication *on M is a pair (A, B) with $A, B \subseteq M$. To support intuition we write $A \to B$ instead of (A, B). We say an implication $A \to B$* holds *for an attribute set C (also: C* respects *$A \to B$),* *if $A \nsubseteq C$ or $B \subseteq C$. Moreover, an implication* $\mathfrak{i}$ holds *(or: is* valid*) in a formal context $\mathbb{K} = (G, M, I)$ if it holds for all sets g^I with $g \in G$. We then write $\mathbb{K} \models \mathfrak{i}$. The size of an implication set $\mathfrak{I}$ (written: $\#\mathfrak{I}$) is defined as $|\mathfrak{I}| \cdot |M|$. Given a set $A \subseteq M$ and a set $\mathfrak{I}$ of implications on M, we write $A^{\mathfrak{I}}$ for the smallest set that contains A and respects all implications from $\mathfrak{I}$. (Since those two requirements are preserved under intersection, the existence of a smallest such set is assured).*

It is obvious that for any set $\mathfrak{I}$ of implications on M, the operation $(\cdot)^{\mathfrak{I}}$ is a closure operator on M. Furthermore, it can be easily shown that an implication $A \to B$ is valid in a formal context $\mathbb{K} = (G, M, I)$ exactly if $B \subseteq A^{II}$.
The following result is an often noted and straightforward consequence from [16].

Proposition 3. *For any attribute set $B \subseteq M$ and set $\mathfrak{I}$ of implications, $B^{\mathfrak{I}}$ can be computed in $O(\#\mathfrak{I}) = O(|\mathfrak{I}| \cdot |M|)$ time and $O(|M|)$ space.*

Like in the case of the contextual encoding, also here it is natural to ask for a size-minimal set of implications that corresponds to a certain closure operator.
Although there is in general no unique minimal implication set for a given closure operator φ, the so-called Duquenne-Guigues base or stem base [10] is often used as a (minimal) canonical representation. We follow this practice and denote it by $\mathfrak{I}(\varphi)$.
Algorithm 2 (cf. [1, 22, 19]) provides a well-known way to turn an arbitrary implication set into an equivalent Duquenne-Guigues base. Thus we can note the following complexity result.

Proposition 4 (Day 1992). *Given a set $\mathfrak{I}$ of implications on M, a size-minimal $\mathfrak{I}'$ with $(\cdot)^{\mathfrak{I}} \equiv (\cdot)^{\mathfrak{I}'}$ can be computed in $O(|\mathfrak{I}| \cdot \#\mathfrak{I}) = O(|\mathfrak{I}|^2 \cdot |M|)$ time and $O(|\mathfrak{I}| \cdot |M|)$ space.*

A closer look at the algorithm reveals that the $O(|\mathfrak{I}| \cdot |M|)$ space bound comes about by the necessity of a 2-pass processing of the implication set. Note that both passes can be performed *in situ* (i.e., by overwriting the input with the output) which would require only $O(|M|)$ additional memory.

3 Size Comparisons

Given these two encodings which are very alike with respect to the complexities of computing closures and minimization, a question which arises naturally is whether one encoding is superior to the other in terms of memory required to store it. First of all, note that for a given M, the size of both representations is bounded by $2^{|M|} \cdot |M|$, i.e. at most exponential in the size of M.

Algorithm 1 `minimizeContext`

Input: context $\mathcal{F}$ on M
Output: size-minimal context $\mathcal{F}'$ such that $(\cdot)^{\mathcal{F}} \equiv (\cdot)^{\mathcal{F}'}$

$\mathcal{F} := \mathcal{F}'$
for each $A \in \mathcal{F}'$ **do**
  if $A = A^{\mathcal{F}' \setminus \{A\}}$ **then**
    $\mathcal{F}' := \mathcal{F}' \setminus \{A\}$
  end if
end for
output $\mathcal{F}'$

Algorithm 2 `minimizeImpSet`

Input: implication set $\mathfrak{I}$ on M
Output: size-minimal implication set $\mathfrak{I}'$ such that $(\cdot)^{\mathfrak{I}} \equiv (\cdot)^{\mathfrak{I}'}$

$\widetilde{\mathfrak{I}} := \emptyset$
for each $A \to B \in \mathfrak{I}$ **do**
  $\widetilde{\mathfrak{I}} := \widetilde{\mathfrak{I}} \cup \{A \to (A \cup B)^{\mathfrak{I}}\}$
end for
$\mathfrak{I}' := \emptyset$
for each $A \to B \in \widetilde{\mathfrak{I}}$ **do**
  delete $A \to B$ from $\widetilde{\mathfrak{I}}$
  $C := A^{\widetilde{\mathfrak{I}} \cup \mathfrak{I}'}$
  if $C \neq B$ **then**
    $\mathfrak{I}' := \mathfrak{I}' \cup \{C \to B\}$
  end if
end for
output $\mathfrak{I}'$

The following proposition shows that for some φ, $\#\mathcal{F}(\varphi)$ is exponentially larger than $\#\mathfrak{I}(\varphi)$.

Proposition 5. *There exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$ of closure operators such that $\#\mathcal{F}(\varphi_n) \in \Theta(2^n)$ whereas $\#\mathfrak{I}(\varphi_n) \in \Theta(n^2)$.*

Proof. We define φ_n as the closure operator on the set $M_n = \{1, \dots, 2n\}$ that corresponds to the implication set $\mathfrak{I}_b$ containing the implication $\{2i-1, 2i\} \to M_n$ for every $i \in \{1, \dots, n\}$. Then, we obtain $\#\mathfrak{I}(\varphi_n) = 2n^2$. On the other hand, $\mathcal{F}(\varphi_n) = \{\{2k - a_k | 1 \leq k \leq n\} \mid \langle a_1, \dots, a_n\rangle \in \{0,1\}^n\}$ (as schematically displayed in Fig. 1) whence we obtain $\#\mathcal{F}(\varphi_n) = 2^n \cdot 2n$. □

	1	2	...	$2n-3$	$2n-2$	$2n-1$	$2n$
g_1	×		...	×		×	
g_2	×		...	×			×
g_3	×		...		×	×	
g_4	×		...		×		×
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
g_{2^n-1}		×	...		×	×	
g_{2^n}		×	...		×		×

Fig. 1. Example for a context that is exponential in the size of its stem base

On the other hand, as a consequence of a result on the number of pseudo-intents [13, 17], we know that the converse is true as well: for some φ, $\#\Im(\varphi)$ is exponentially larger than $\#\mathcal{F}(\varphi)$.

Proposition 6 (Kuznetsov 2004, Mannila & Räihä 1992). *There exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$ of closure operators such that $\#\mathcal{F}(\varphi_n) \in \Theta(n^2)$ but $\#\Im(\varphi_n) \in \Theta(2^n)$.*

This result seems to imply that in general, one cannot avoid the exponential blowup if a contextually represented closure operator is to be represented by means of implications.
However, as the following definition and theorem show, this only holds if M is supposed to be fixed. If we allow for a bit more freedom in terms of the used attribute set, this blowup can be avoided.

Definition 7. *Given a context $\mathcal{F}$ on a set M, let M^+ denote the set M extended by a one new attribute m_F for each $F \in \mathcal{F}$. Then we define $\Im_{\mathcal{F}}$ as implication set on M^+ containing for every $m \in M$ the two implications $\{m\} \to \{m_F \mid F \in \mathcal{F}, m \notin F\}$ and $\{m_F \mid F \in \mathcal{F}, m \notin F\} \to \{m\}$.*

Theorem 7. *Given a context $\mathcal{F}$ on a set M, the following hold*

1. $\#\Im_{\mathcal{F}} = 2 \cdot |M| \cdot |M^+| = 2 \cdot |M| \cdot (|M| + |\mathcal{F}|) \leq (\#\mathcal{F})^2$.
2. $(\cdot)^{\mathcal{F}} \equiv (\cdot)^{\Im_{\mathcal{F}}}|_M$*, that is,* $A^{\mathcal{F}} = A^{\Im_{\mathcal{F}}} \cap M$ *for all* $A \subseteq M$.

Proof. The first claim is obvious.
For the second claim, we first show that for an arbitrary set $A \subseteq M$ holds $A^{\Im_{\mathcal{F}}} = B \cup C$ with $B := \{m_F \mid F \in \mathcal{F}, A \nsubseteq F\}$ and $C := \{m \mid \{m_F \mid F \in \mathcal{F}, m \notin F\} \subseteq B\}$.
To show $A^{\Im_{\mathcal{F}}} \subseteq B \cup C$ we note that $A \subseteq B \cup C$ and that $B \cup C$ is $\Im_{\mathcal{F}}$-closed: $B \cup C$ satisfies all implications of the type $\{m_F \mid F \in \mathcal{F}, m \notin F\} \to \{m\}$ by definition of C. To check implications of the second type, $\{m\} \to \{m_F \mid F \in \mathcal{F}, m \notin F\}$, we note that

$$\begin{aligned}
C &:= \{m \mid \{m_F \mid F \in \mathcal{F}, m \notin F\} \subseteq B\} \\
&= \{m \mid \{m_F \mid F \in \mathcal{F}, m \notin F\} \subseteq \{m_F \mid F \in \mathcal{F}, A \nsubseteq F\}\} \\
&= \{m \mid \forall F \in \mathcal{F} : m \notin F \to A \nsubseteq F\}
\end{aligned}$$

Now, picking an $m \in C$, we find that every m_F for which $m \notin F$ must also satisfy $A \nsubseteq F$ and therefore $m_F \in B$ so we find all implications of the second type satisfied.
Further, we show $B \cup C \subseteq A^{\Im_{\mathcal{F}}}$, by proving $B \subseteq A^{\Im_{\mathcal{F}}}$ and $C \subseteq A^{\Im_{\mathcal{F}}}$ separately.
We obtain $B = \{m_F \mid F \in \mathcal{F}, A \nsubseteq F\} \subseteq A^{\Im_{\mathcal{F}}}$ due to the following: given an $F \in \mathcal{F}$ with $A \nsubseteq F$, we find an $m \in A$ with $m \notin F$ and thus an implication $m \to \{m_F, \ldots\}$ contained in $\Im_{\mathcal{F}}$, therefore $A^{\Im_{\mathcal{F}}}$ must contain m_F.

We then also obtain $C := \{m \mid \{m_F \mid F \in \mathcal{F}, m \notin F\} \subseteq B\} \subseteq A^{\mathfrak{I}_\mathcal{F}}$ by the following argument: picking an $m \in C$, we find the implication $\{m_F \mid F \in \mathcal{F}, m \notin F\} \to \{m\}$ contained in $\mathfrak{I}_\mathcal{F}$. On the other hand, we already know $B \subseteq A^{\mathfrak{I}_\mathcal{F}}$ and $B \supseteq \{m_F \mid F \in \mathcal{F}, m \notin F\}$, hence $m \in A^{\mathfrak{I}_\mathcal{F}}$.
Finally, we obtain $A^{\mathfrak{I}_\mathcal{F}}|_M = A^{\mathfrak{I}_\mathcal{F}} \cap M = C = \{m \mid \forall F \in \mathcal{F} : m \notin F \to A \nsubseteq F\} = \{m \mid \forall F \in \mathcal{F} : A \subseteq F \to m \in F\} = \bigcap_{F \in \mathcal{F}, A \subseteq F} F = A^\mathcal{F}$ for any $A \subseteq M$. □

Thus, we obtain a polynomially size-bounded implicational representation of a context. In our view this is a remarkable – although not too intricate – insight as it seems to challenge the practical relevance of computationally hard problems w.r.t. pseudo-intents (recognizing, enumerating, counting), on which theoretical FCA research has been focusing lately [14, 19, 15, 21, 20, 2].

4 Algorithms for Managing Closure Operators

4.1 Finer or Coarser?

Depending on how closure operators are represented, there are several ways of checking if one is finer than the other. As the general case, we consider the situation where both closure operators are given in a "black-box" manner, i.e. as opaque functions that we can call and that come with runtime guarantees.

Theorem 8. *Let φ and ψ be closure operators on a set M for which computing of closures can be performed in time t_φ and t_ψ, respectively and space s_φ and s_ψ, respectively. Moreover, let $cl_\varphi = |\{\varphi(A) \mid A \subseteq M\}|$.*
Then, $\varphi \leq \psi$ can be decided in $O(cl_\varphi \cdot (|M| \cdot t_\varphi + t_\psi))$ and in $O(2^{|M|} \cdot (t_\varphi + t_\psi))$.
The space complexity is bounded by $O(s_\varphi + s_\psi)$.

Proof. For the $O(cl_\varphi \cdot (|M| \cdot t_\varphi + t_\psi))$ time bound, we employ Ganter's NextClosure algorithm [6, 7] for enumerating the closed sets of φ. We note that (i) it only makes use of the closure operator in a black-box manner (that is, it does not depend on a its specific representation) by calling it as a function and (ii) it has polynomial delay, more precisely the time between two closed sets being output is $O(|M| \cdot t_\varphi)$. For each delivered closed set, we have to additionally check if it is ψ-closed, hence the overall time needed per φ-closed set is $O(|M| \cdot t_\varphi + t_\psi)$.
The $O(2^{|M|} \cdot (t_\varphi + t_\psi))$ bound can be obtained by naïvely checking all subsets of M for φ-closedness and ψ-closedness.
In both cases, no intermediary information needs to be stored between testing successive sets which explains the space complexity. □

Note that all known black-box algorithms require exponentially many closure computations w.r.t. to $|M|$. This raises the question whether this bound can be

Algorithm 3 `finerThanContext`

Input: closure operator φ on set M,
context $\mathcal{F}$

Output: YES if $\varphi \leq (\cdot)^{\mathcal{F}}$, NO otherwise

for each $A \in \mathcal{F}$ **do**
  if $A \neq \varphi(A)$ **then**
    output NO
    exit
  end if
end for
output YES

Algorithm 4 `coarserThanImpSet`

Input: closure operator φ on set M,
implication set $\mathfrak{I}$

Output: YES if $(\cdot)^{\mathfrak{I}} \leq \varphi$, NO otherwise

for each $A \to B \in \mathfrak{I}$ **do**
  if $B \nsubseteq \varphi(A)$ **then**
    output NO
    exit
  end if
end for
output YES

improved if one or both of the to-be-compared closure operators are available in a specific representation. The subsequent theorem captures the cases where polynomially many calls suffice.

Theorem 9. *Let φ be a closure operator on a set M for which computing of closures can be performed in t_φ time and s_φ space. Then, the following hold:*

- *for a context $\mathcal{F}$ on M, the problem $\varphi \leq (\cdot)^{\mathcal{F}}$ can be decided in $|\mathcal{F}| \cdot t_\varphi$ time and s_φ space and*
- *for an implication set $\mathfrak{I}$ on M, the problem $(\cdot)^{\mathfrak{I}} \leq \varphi$ can be decided in $|\mathfrak{I}| \cdot t_\varphi$ time and s_φ space.*

Proof. Algorithm 3 provides a solution for the first case. It verifies that every element (in other words: every object intent) of $\mathcal{F}$ is φ-closed, this suffices to guarantee that all $\mathcal{F}$-closed sets are φ-closed since every $\mathcal{F}$-closed set is an intersection of elements of $\mathcal{F}$ and φ-closed sets are closed under intersections (since this holds for every closure operator).
Algorithm 4 provides a solution for the second case. To ensure that every φ-closed set is also $(\cdot)^{\mathfrak{I}}$-closed, it suffices to show that every φ-closed set respects all implications from $\mathfrak{I}$. If every φ-closed set respects an implication $A \to B \in \mathfrak{I}$ can in turn be verified by checking if $B \subseteq \varphi(A)$. □

The results established in the above theorem give rise to precise polynomial complexity bounds for three of the four possible comparisons of closure operators which are contextually or implicationally represented.

Corollary 10. *Given contexts $\mathcal{F}, \mathcal{F}'$ and implication sets $\mathfrak{I}, \mathfrak{I}'$ on some set M, it is possible to check*

- $(\cdot)^{\mathcal{F}} \leq (\cdot)^{\mathcal{F}'}$ *in time $O(|\mathcal{F}| \cdot |\mathcal{F}'| \cdot |M|)$,*

– $(\cdot)^{\mathfrak{I}} \leq (\cdot)^{\mathfrak{I}'}$ *in time* $O(|\mathfrak{I}| \cdot |\mathfrak{I}'| \cdot |M|)$*, and*
– $(\cdot)^{\mathcal{F}} \leq (\cdot)^{\mathfrak{I}}$ *in time* $O(|\mathcal{F}| \cdot |\mathfrak{I}| \cdot |M|)$.

Surprisingly, the ensuing question – whether it is possible to establish a polynomial time complexity bound for the missing comparison case – has to be denied assuming $P \neq NP$, as the following theorem shows.[1]

Theorem 11. *Given a context* $\mathcal{F}$ *and an implication set* $\mathfrak{I}$ *on some set* M*, deciding if* $(\cdot)^{\mathfrak{I}} \leq (\cdot)^{\mathcal{F}}$ *is coNP-complete.*

Proof. To show coNP membership, we note that $(\cdot)^{\mathfrak{I}} \nleq (\cdot)^{\mathcal{F}}$ if and only if there is a set A and which is $(\cdot)^{\mathfrak{I}}$-closed but not $(\cdot)^{\mathcal{F}}$-closed. Clearly, we can guess such a set and check the above properties in polynomial time.
We show coNP hardness, by a polynomial reduction of the problem to 3SAT [11]. Given a set $C = \{C_1, \ldots, C_k\}$ of 3-clauses (i.e. $|C_i| = 3$) over a set of literals $L = \{p_1, \neg p_1, \ldots p_\ell, \neg p_\ell\}$, we let $M = L$ and define

$$\mathfrak{I} := \{\{p_i, \neg p_i\} \to M \mid p_i \in L\}$$

as well as

$$\mathcal{F} := \{M \setminus (C_i \cup \{m\}) \mid C_i \in C, m \in M\}.$$

We now show that there is a set A with $A^{\mathfrak{I}} = A$ but $A^{\mathcal{F}} \neq A$ exactly if there is a valuation on $\{p_1, \ldots, p_\ell\}$ for which C is satisfied.
For the "if" direction assume $val : \{p_1, \ldots, p_\ell\} \to \{true, false\}$ to be that valuation and define $A := \{p_i \mid val(p_i) = true\} \cup \{\neg p_i \mid val(p_i) = false\}$. Obviously, A is $(\cdot)^{\mathfrak{I}}$-closed. On the other hand, since by definition A must contain one element from each $C_i \in C$, we have that $F \nsubseteq A$ for all $F \in \mathcal{F}$ and hence $A^{\mathcal{F}} = M \neq A$.
For the "only if" direction, assume $A^{\mathfrak{I}} = A$ but $A^{\mathcal{F}} \neq A$. By construction of $\mathcal{F}$, the latter can only be the case if A contains one element of each $C_i \in C$. Thus, the valuation $val : \{p_1, \ldots, p_\ell\} \to \{true, false\}$ with

$$val(p_i) = \begin{cases} true & \text{if } p_i \in A \\ false & \text{otherwise} \end{cases}$$

witnesses the satisfiability of C. □

Drawing from the above black-box case, a straightforward deterministic algorithm for testing $(\cdot)^{\mathfrak{I}} \leq (\cdot)^{\mathcal{F}}$ would need to subsequently generate all closed sets of $\mathfrak{I}$ (e.g. by Ganter's algorithm [6, 7]) and check closedness w.r.t. $\mathcal{F}$. This algorithm would, however, require exponential time w.r.t. $|M|$ in the worst case.

[1] As indicated by a reviewer, this result in a slightly different formulation is already known in other communities, cf. [9].

4.2 Adding a Closed Set

We now consider the task of making a closure operator φ minimally "finer" by requiring that a given set A be a closed set.

Definition 8. *Given a closure operator φ on M and some $A \subseteq M$, the A-*refinement *of φ (written $\varphi{\downarrow}A$) is defined as the coarsest closure operator ψ with $\psi \leq \varphi$ and $\psi(A) = A$.*

It is straightforward to show that B is a $\varphi{\downarrow}A$-closed set exactly if it is φ-closed or the intersection of A and a φ-closed set. Clearly, if a closure operator is represented as formal context, refinements can be computed by simply adding a row, i.e. for any context $\mathcal{F}$ on M and set $A \subseteq M$ we have for $\mathcal{F}' := \mathcal{F} \cup \{A\}$ that $(\cdot)^{\mathcal{F}}{\downarrow}A \equiv (\cdot)^{\mathcal{F}'}$. Of course, $\mathcal{F}'$ will in general not be size-minimal even if $\mathcal{F}$ is.

Fact 12. *Given a context $\mathcal{F}$ on M and some $A \in M$, an $\mathcal{F}'$ with $(\cdot)^{\mathcal{F}'} \equiv (\cdot)^{\mathcal{F}}{\downarrow}A$ can be computed in $O(|M|)$ time and constant space. Moreover, we have $|\mathcal{F}'| \leq |\mathcal{F}| + 1$.*

If the closure operator is represented in terms of implications, a little more work is needed for this task.

Proposition 13. *Given an implication set $\mathfrak{I}$ on M and some $A \in M$, an $\mathfrak{I}'$ with $(\cdot)^{\mathfrak{I}'} \equiv (\cdot)^{\mathfrak{I}}{\downarrow}A$ can be computed in $O(|\mathfrak{I}| \cdot |M|^2)$ time. Moreover, we have $|\mathfrak{I}'| \leq |\mathfrak{I}| \cdot |M|$.*

Proof. Algorithm 5 ensures the claimed complexity behavior. We now show that it is correct by proving that a subset of M is $(\cdot)^{\mathfrak{I}'}$-closed iff it is $(\cdot)^{\mathfrak{I}}$-closed or the intersection of A and some $(\cdot)^{\mathfrak{I}}$-closed set.
For the "if" direction, note that all implications from $\mathfrak{I}'$ are entailed by $\mathfrak{I}$, therefore all $(\cdot)^{\mathfrak{I}}$-closed sets are $(\cdot)^{\mathfrak{I}'}$-closed. Further we note that A is obviously $(\cdot)^{\mathfrak{I}'}$-closed. As intersections preserve closedness the above implies that all intersections of A and an $\mathfrak{I}$-closed sets must be $(\cdot)^{\mathfrak{I}'}$-closed.
For the "only if" direction, let S be a $(\cdot)^{\mathfrak{I}'}$-closed set. If we assume $S \subseteq A$ we find that $S^{\mathfrak{I}'} = S^{\mathfrak{I}} \cap A$. If $S \nsubseteq A$, there exists an $m \in S \setminus A$. But then we find $S^{\mathfrak{I}'} = S^{\mathfrak{I}}$. □

4.3 Adding an Implication

The task dual to the one from the preceding section is to make a given closure operator coarser by requiring that all closed sets of the coarsened version respect a given implication. In other words, all closed sets not respecting the implication are removed.

Definition 9. *Given a closure operator φ on M and some implication $\mathfrak{i} = A \to B$ with $A, B \subseteq M$, the* $\mathfrak{i}$-coarsening *of φ (written $\varphi{\uparrow}\mathfrak{i}$) is defined as the finest closure operator ψ with $\varphi \leq \psi$ and $B \subseteq \psi(A)$.*

Clearly, if a closure operator is represented as implication set, coarsenings can be computed by simply adding the implication to the set. Note that $\mathfrak{I}' := \mathfrak{I} \cup \{\mathfrak{i}\}$ will in general not be size-minimal.

Fact 14. *Given an implication set $\mathfrak{I}$ on M and some implication $\mathfrak{i}$ on M, an $\mathfrak{I}'$ with $(\cdot)^{\mathfrak{I}'} \equiv (\cdot)^{\mathfrak{I}}{\uparrow}\mathfrak{i}$ can be computed in $O(|M|)$ time and constant space. Moreover, we have $|\mathcal{F}'| \leq |\mathcal{F}| + 1$.*

If the closure operator is represented by a context, a little more work is needed for this task.

Proposition 15. *Given a context $\mathcal{F}$ on M and some implication $\mathfrak{i}$ on M, an $\mathcal{F}'$ with $(\cdot)^{\mathcal{F}'} \equiv (\cdot)^{\mathcal{F}}{\uparrow}\mathfrak{i}$ can be computed in $O(|\mathcal{F}|^2 \cdot |M|)$ time. Moreover, we have $|\mathcal{F}'| \leq |\mathcal{F}|^2$.*

Proof. It is easy to check that Algorithm 6 satisfies the given complexity bounds. We show its correctness by verifying that a set is $(\cdot)^{\mathcal{F}'}$-closed exactly if it is $(\cdot)^{\mathcal{F}}$-closed and respects $A \to B$.
For the "if" direction, let S be an $(\cdot)^{\mathcal{F}}$-closed set that respects $A \to B$. This means that either $B \subseteq S$ or $A \nsubseteq S$. In the first case, note that every $F \in \mathcal{F}$ with $S \subseteq F$ respects $A \to B$ and thus each such F is contained in $\mathcal{F}'$ as well. Since S is the intersection of all these F it must itself be $(\cdot)^{\mathcal{F}'}$-closed. In the second case, there must be some $F \in \mathcal{F}$ with $S \subseteq F$ with $A \nsubseteq F$. Thus we obtain

$$\begin{aligned} S &= \textstyle\bigcap_{S \subseteq F' \in \mathcal{F}} F' \\ &= (\textstyle\bigcap_{S \subseteq F' \in \mathcal{F}, S \text{ respects } A \to B} F') \cap (\textstyle\bigcap_{S \subseteq F' \in \mathcal{F}, S \text{ violates } A \to B} F') \cap F \\ &= (\textstyle\bigcap_{S \subseteq F' \in \mathcal{F}, S \text{ respects } A \to B} F') \cap (\textstyle\bigcap_{S \subseteq F' \in \mathcal{F}, S \text{ violates } A \to B} F' \cap F) \end{aligned}$$

and see that S is an intersection of $(\cdot)^{\mathcal{F}'}$-closed sets and hence itself $(\cdot)^{\mathcal{F}'}$-closed.
For the "only if" direction, consider an arbitrary $(\cdot)^{\mathcal{F}'}$-closed set S. It can be easily checked that all $F \in \mathcal{F}'$ respect $A \to B$, hence also S does. Moreover, by definition, every $F \in \mathcal{F}'$ is an intersection of elements of $\mathcal{F}$ and thus $(\cdot)^{\mathcal{F}}$-closed. □

4.4 Conversion of Representations

We conclude this paper by considering the problem of extracting minimal implicational or contextual representations from black-box closure operators. While

Algorithm 5 `addClosedSet`

Input: implication set $\mathfrak{I}$ on M, set $A \subseteq M$
Output: implication set $\mathfrak{I}'$ with $(\cdot)^{\mathfrak{I}'} \equiv (\cdot)^{\mathfrak{I}}\downarrow A$

$\mathfrak{I}' := \emptyset$
for each $B \to C \in \mathfrak{I}$ **do**
  if $B \to C$ is respected by A **then**
    $\mathfrak{I}' := \mathfrak{I}' \cup \{B \to C\}$
  else
    $\mathfrak{I}' := \mathfrak{I}' \cup \{B \to C \cap A\}$
    for each $m \in M \setminus A$ **do**
      $\mathfrak{I}' := \mathfrak{I}' \cup \{B \cup \{m\} \to C\}$
    end for
  end if
end for
output $\mathfrak{I}'$

Algorithm 6 `addImplication`

Input: context $\mathcal{F}$ on M,
implication $\mathfrak{i} = A \to B$ on M
Output: context $\mathcal{F}'$ with $(\cdot)^{\mathcal{F}'} \equiv (\cdot)^{\mathcal{F}}\uparrow\mathfrak{i}$

$\mathcal{F}' := \emptyset$
for each $C \in \mathcal{F}$ **do**
  if C respects $A \to B$ **then**
    $\mathcal{F}' := \mathcal{F}' \cup \{C\}$
  else
    for each $D \in \mathcal{F}$ with $A \not\subseteq D$ **do**
      $\mathcal{F}' := \mathcal{F}' \cup \{C \cap D\}$
    end for
  end if
end for
output $\mathcal{F}'$

the problem of finding a minimal implication base has been considered extensively, the dual task has hardly been considered so far. In both cases, however, no output-polynomial algorithm could be established.

We start by considering the dual task: given a black-box closure operator φ, how can we compute $\mathcal{F}(\varphi)$? Algorithm 7 displays a semi-naïve approach which essentially computes the row-reduced version of the context containing all φ-closed sets, but size-minimizes the context on the way by progressing in reverse lectic order. This yields us with an algorithm requiring $O(2^{|M|} \cdot (t_\varphi + \#\mathcal{F}(\varphi))$ time but only $|M|$ space. If φ is represented by an implication set $\mathfrak{I}$, this amounts to a time complexity of $O(2^{|M|} \cdot |M| \cdot (|\mathfrak{I}| + |\mathcal{F}(\varphi)|)$

Unfortunately, this still means that the algorithm is worst-case time-exponential w.r.t. $|M|$, even if $\#\mathcal{F}(\varphi)$ is "small" (i.e., polynomially bounded w.r.t. $|M|$). As a straightforward example, consider the closure operator φ_{id} with $\varphi_{\text{id}}(A) = A$ for all $A \subseteq M$, for which $\mathcal{F}(\varphi_{\text{id}}) = \{M \setminus \{m\} \mid m \in M\}$. In fact, the question whether a better, tractable behavior can be obtained at all has to be refuted: it follows rather directly from Thm 5.2 of [5], that no output-polynomial algorithm for this task can exist.[2]

Finally reviewing the more popular task of determining the stem base of a given formal context, the following can be shown by an inspection of Ganter's algorithm for enumerating all pseudo-closed sets of a closure operator [6, 7].

[2] More precisely, the authors of [5] provide a representation of propositional Horn theories that admits for polynomial computation of the associated closure operator but does not allow for polynomial delay enumeration of "'characteristic models", that is intents of the corresponding reduced context.

Algorithm 7 `extractContext`

Input: closure operator φ on set M
Output: context $\mathcal{F}(\varphi)$

$\mathcal{F} = \emptyset$
for each $F \subseteq M$, enumerated in inverse lectic order **do**
  if $F = \varphi F$ **then**
    if $F \neq F^{\mathcal{F}}$ **then**
      $\mathcal{F} = \mathcal{F} \cup \{F\}$
    end if
  end if
end for
output $\mathcal{F}$

Algorithm 8 `contextToImpSet`

Input: context $\mathcal{F}$ on set M
Output: implication set $\Im((\cdot)^{\mathcal{F}})$

compute $\Im_{\mathcal{F}}$
$\Im = \Im_{\mathcal{F}}$
for each $m_F \in M^+ \setminus M$ **do**
  $\Im' = \emptyset$
  for each $A \to B \in \Im$ **do**
    $\Im' := \Im' \cup \{A \to B \setminus \{m_F\}\}$
    if $m_F \in B \setminus A$ **then**
      for each $C \to D \in \Im$ with $m_F \in C$ **do**
        $\Im' := \Im' \cup \{A \cup C \setminus \{m_F\} \to D\}$
      end for
    end if
  end for
  $\Im :=$ `minimizeImpSet`$(\Im')$
end for
output $\Im$

Proposition 16 (essentially Ganter 1984). *Let φ be a closure operator on a set M for which computing of closures can be performed in time t_φ and space s_φ. Then $\Im(\varphi)$ can be computed in time $O(2^{|M|}\cdot(t_\varphi+\#\Im(\varphi))) = O(2^{|M|}\cdot(t_\varphi+|M|\cdot|\Im(\varphi)|))$ and space $O(s_\varphi)$.*

For the case of φ being explicitly represented by a context, this implies that converting a contextual representation into an implicational one can be done in time $O(2^{|M|} \cdot (\#\mathcal{F} + \#\Im((\cdot)^{\mathcal{F}}))) = O(2^{|M|} \cdot |M| \cdot (|\mathcal{F}| + |\Im((\cdot)^{\mathcal{F}})|))$.

The results from Section 3 give rise to a quite different approach of computing $\Im((\cdot)^{\mathcal{F}})$ from a given $\mathcal{F}$. Starting from the polynomial-size implicational representation $\Im^{\mathcal{F}}$ of a context $\mathcal{F}$, one can one-by-one remove the auxiliary attributes m_F by a resolution procedure, while minimizing the intermediate implicational representations via `minimizeImpSet`. This method is formally specified in Algorithm 8. While the correctness of the algorithm is a rather immediate, establishing complexity results is the subject of ongoing work. Whether the algorithm turns out to be output-polynomial must, however, be doubted given that this would imply the existence of an output-polynomial algorithm for finding the transversal hypergraph of a given hypergraph (as first observed in [12] and put in FCA terms in [3]), which has been an open problem for over 20 years now (see [4] for a comprehensive overview). Moreover, it has been shown that no polynomial-delay algorithms for enumerating the stembase in lectic [2] or inverse lectic [18] order can exist unless $P = NP$.

5 Conclusion

We have investigated runtime and memory requirements for diverse tasks related to closure operators. The overview displayed in Table 1 reveals a certain duality between the two representations forms – context or implication set – and ascertains that none can be generally preferred to the other.

Table 1. Time complexities for different representations and tasks.

	context $\mathcal{F}$	implication set $\mathfrak{I}$
closure	$O(\|\mathcal{F}\| \cdot \|M\|)$	$O(\|\mathfrak{I}\| \cdot \|M\|)$
turn to minimal $\mathcal{F}'$	$O(\|\mathcal{F}\|^2 \cdot \|M\|)$	$O(2^{\|M\|} \cdot \|M\| \cdot (\|\mathfrak{I}\| + \|\mathcal{F}((\cdot)^{\mathfrak{I}})\|)$
turn to minimal $\mathfrak{I}'$	$O(2^{\|M\|} \cdot \|M\| \cdot (\|\mathcal{F}\| + \|\mathfrak{I}((\cdot)^{\mathcal{F}})\|))$	$O(\|\mathfrak{I}\|^2 \cdot \|M\|)$
check if $\mathfrak{I}'$ finer	$O(\|\mathcal{F}\| + \|\mathfrak{I}'\|) \cdot \|M\| \cdot 2^{\|M\|}$	$O(\|\mathfrak{I}\| \cdot \|\mathfrak{I}'\| \cdot \|M\|)$
check if $\mathcal{F}'$ finer	$O(\|\mathcal{F}\| \cdot \|\mathcal{F}'\| \cdot \|M\|)$	$O(\|\mathcal{F}'\| \cdot \|\mathfrak{I}\| \cdot \|M\|)$
check if $\mathfrak{I}'$ coarser	$O(\|\mathcal{F}\| \cdot \|\mathfrak{I}'\| \cdot \|M\|)$	$O(\|\mathfrak{I}\| \cdot \|\mathfrak{I}'\| \cdot \|M\|)$
check if $\mathcal{F}'$ coarser	$O(\|\mathcal{F}\| \cdot \|\mathcal{F}'\| \cdot \|M\|)$	$O(\|\mathcal{F}'\| + \|\mathfrak{I}\|) \cdot \|M\| \cdot 2^{\|M\|}$
extract from φ	$O(\|M\| \cdot \|\mathcal{F}(\varphi)\|$ $\cdot(t_\varphi + \|M\| \cdot \|\mathfrak{I}(\varphi)\|))$	$O(\|M\| \cdot \|\mathcal{F}(\varphi)\|$ $\cdot(t_\varphi + \|M\| \cdot \|\mathfrak{I}(\varphi)\|))$
add implication	$O(\|\mathcal{F}\|^2 \cdot \|M\|)$	$O(M)$
add closed set	$O(M)$	$O(\|\mathfrak{I}\| \cdot \|M\|^2)$

There are many open questions left. On the theoretical side, central open problems are if there are algorithms transforming contextual into implicational representations and vice versa in output polynomial time. Note that a negative answer to this question would also disprove the existence of polynomial-delay algorithms.
On the practical side, coming back to our initial motivation, it should be experimentally investigated if variants of standard FCA algorithms can be improved by adding the option of working with alternative closure operator representations.
Moreover, the proposed alternative algorithm for computing the Duquenne-Guigues base should be evaluated against Ganter's algorithm on typical datasets from practical use cases, in order to assess its practical use.

Acknowledgements

The author is deeply indebted to the reviewers who gave extremely valuable hints to existing related work. This work has been supported by the project ExpresST funded by the German Research Foundation (DFG).

References

1. Day, A.: The lattice theory of functional dependencies and normal decompositions. International Journal of Algebra and Computation 2(4), 409–431 (1992)
2. Distel, F.: Hardness of enumerating pseudo-intents in the lectic order. In: Proc. ICFCA. LNCS, vol. 5986, pp. 124–137. Springer (2010)
3. Distel, F., Sertkaya, B.: On the complexity of enumerating pseudo-intents. Discrete Applied Mathematics 159(6), 450–466 (2011)
4. Eiter, T., Gottlob, G.: Hypergraph transversal computation and related problems in logic and ai. In: Proc. JELIA. pp. 549–564 (2002)
5. Eiter, T., Ibaraki, T., Makino, K.: Computing intersections of Horn theories for reasoning with models. Tech. Rep. IFIG research report 9803, Universität Gießen (1998), http://bibd.uni-giessen.de/ghtm/1998/uni/r980014.htm
6. Ganter, B.: Two basic algorithms in concept analysis. Tech. Rep. 831, FB4, TH Darmstadt (1984)
7. Ganter, B.: Two basic algorithms in concept analysis. In: Proc ICFCA. LNCS, vol. 5986, pp. 312–340. Springer (2010)
8. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer-Verlag (1997)
9. Gottlob, G., Libkin, L.: Investigations on armstrong relations, dependency inference, and excluded functional dependencies. Acta Cybernetica 9(4), 385–402 (1990)
10. Guigues, J.L., Duquenne, V.: Familles minimales d'implications informatives resultant d'un tableau de données binaires. Math. Sci Humaines 95, 5–18 (1986)
11. Karp, R.M.: Reducibility Among Combinatorial Problems. In: Miller, R.E., Thatcher, J.W. (eds.) Complexity of Computer Computations, pp. 85–103. Plenum Press (1972)
12. Kavvadias, D., Papadimitriou, C., Siedri, M.: On Horn envelopes and hypergraph traversals. In: Proc. ISAAC. LNCS, vol. 762. Springer (1993)
13. Kuznetsov, S.O.: On the intractability of computing the duquenne-guigues base. Journal of Universal Computer Science 10(8), 927–933 (2004)
14. Kuznetsov, S.O., Obiedkov, S.A.: Counting pseudo-intents and #p-completeness. In: ICFCA. LNCS, vol. 3874, pp. 306–308. Springer (2006)
15. Kuznetsov, S.O., Obiedkov, S.A.: Some decision and counting problems of the duquenne-guigues basis of implications. Discrete Applied Mathematics 156(11), 1994–2003 (2008)
16. Maier, D.: The Theory of Relational Databases. Computer Science Press (1983)
17. Mannila, H., Räihä, K.J.: Design of Relational Databases. Addison-Wesley (1992)
18. Mikhail A. Babin, S.O.K.: Recognizing pseudo-intents is conp-complete. In: Marzena Kryszkiewicz, S.O. (ed.) Proc. 7th Int. Conf. on Concept Lattices and Their Applications (CLA 2010). CEUR Workshop Proceedings, vol. 672. CEUR-WS.org (2010)
19. Rudolph, S.: Some notes on pseudo-closed sets. In: Proc. ICFCA. LNCS, vol. 4390, pp. 151–165. Springer (2007)
20. Sertkaya, B.: Some computational problems related to pseudo-intents. In: Proc. ICFCA. LNCS, vol. 5548, pp. 130–145. Springer (2009)
21. Sertkaya, B.: Towards the complexity of recognizing pseudo-intents. In: Proc. ICCS. LNCS, vol. 5662, pp. 284–292. Springer (2009)
22. Wild, M.: Implicational bases for finite closure systems. In: Lex, W. (ed.) Arbeitstagung Begriffsanalyse und Künstliche Intelligenz. pp. 147–169. Springer (1991)