On the lattice of antichains of finite intervals

Paolo Boldi* Sebastiano Vigna†

Abstract

Motivated by applications to information retrieval, we study the lattice of antichains of finite intervals of a locally finite, totally ordered set. Intervals are ordered by reverse inclusion; the order between antichains is induced by the lower set they generate. We discuss in general properties of such *antichain completions*; in particular, their connection with Alexandrov completions. We prove the existence of a unique, irredundant $\wedge$-representation by $\wedge$-irreducible elements, which makes it possible to write the relative pseudo-complement in closed form. We also discuss in detail properties of additional interesting operators used in information retrieval. Finally, we give a formula for the rank of an element and for the height of the lattice.

1 Introduction

Modern information-retrieval systems, such as web search engines, rely on different models to compute the answer to a query. The simplest one is the *Boolean model*, in which operators are just conjunction, disjunction and negation, and the answer is just "relevant" or "not relevant". A richer model is given by *minimal-interval semantics* [4], which uses *antichains* (w.r.t. inclusion) of *intervals* (i.e., textual passages in the document) to represent the answer to a query; this is a very natural framework in which operators such as ordered conjunction, proximity restriction, etc., can be defined and combined freely. Each interval in the text of a document is a *witness* of the satisfiability of the query, that is, it represents a region of the document that satisfies the query. Words in the document are numbered (starting from 0), so regions of text are identified with integer intervals, that is, sets of consecutive natural numbers. For example, a query formed by the conjunction of two terms is satisfied by the minimal intervals representing the regions of the document that contain both terms. These intervals can be used not only to estimate the relevance of the document to the query [3], but also to provide the user with *snippets*—fragments of texts witnessing where (and why) the document satisfies the query.

Clarke, Cormack and Burkowski defined minimal-interval semantics in their seminal work [4], but they missed the connection with lattice theory, proving every single property (e.g., distributivity) from scratch: in this paper, we firstly aim at providing a principled introduction of their framework in terms of *antichain completions* and their relation with Alexandrov completions. On the other hand, we extend the study of these structures by allowing infinite antichains, and characterizing in an elementary manner operators, such as the relative pseudo-complement, some of which have an immediate interpretation in information retrieval.

We conclude by discussing interesting algebraic properties connecting the operators with one another, and by providing a closed form for the rank of an element.

In the rest of the paper we use the Hoare–Ramshaw notation [9] for intervals in a partially ordered set: $[a\,..\,b] = \{x \mid a \leq x \leq b\}$ denotes a closed interval from a to b, $(\leftarrow\,..\,b] = \{x \mid x \leq b\}$ denotes an interval containing all elements less than or equal to b, and $[a\,..\,\rightarrow) = \{x \mid$

*Dipartimento di Informatica, Università degli Studi di Milano, via Comelico 39, I-20135 Milano, Italy. paolo.boldi@unimi.it.

†Contact author. Dipartimento di Informatica, Università degli Studi di Milano, via Comelico 39, I-20135 Milano, Italy. sebastiano.vigna@unimi.it.

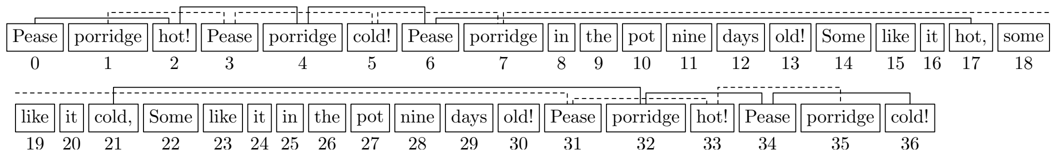


Figure 1: A document; the intervals corresponding to the semantics of the query "*porridge* AND *pease* AND (*hot* OR *cold*)" are shown. For easier reading, every other interval is dashed.

$a \leq x\}$ denotes an interval containing all elements greater than or equal to a. In case we need to treat indifferently intervals of the form $[a\,..\,b]$ or $(\leftarrow ..\, b]$, we will use the shortcut notation $[-\,..\,b]$, and analogously for $[a\,..\,-]$. We also write $[x]$ as a shortcut for $[x\,..\,x] = \{x\}$.

2 A motivating example

In this section we provide an introduction to minimal-interval semantics by examples. Consider the text document represented in Figure 1. Queries associated with a single keyword have a natural semantics in terms of antichains of intervals—the list of positions where the keyword occurs as singleton intervals. For example, since the word "hot" appears only three times (in positions 2, 17 and 33), the semantics of "*hot*" will be

$$\{\,[2\,..\,2], [17\,..\,17], [33\,..\,33]\,\}.$$

If we start combining terms disjunctively, we get simply the union of their positions. For instance, "*hot* OR *cold*" gives

$$\{\,[2\,..\,2], [5\,..\,5], [17\,..\,17], [21\,..\,21], [33\,..\,33], [36\,..\,36]\,\}.$$

If we consider the conjunction of two terms, we will start getting non-singleton intervals: the semantics of "*pease* AND *porridge*" is computed by picking all possible pairs of positions of *pease* and *porridge* and keeping the minimal intervals among those spanned by such pairs:

$$\{\,[0\,..\,1], [1\,..\,3], [3\,..\,4], [4\,..\,6], [6\,..\,7], [7\,..\,31], [31\,..\,32], [32\,..\,34], [34\,..\,35]\,\}.$$

The more complex query "(*pease* AND *porridge*) OR *hot*" is interesting because we have to take the intervals just computed, put them together with the positions of *hot*, and remove the non-minimal intervals:

$$\{\,[0\,..\,1], [2\,..\,2], [3\,..\,4], [4\,..\,6], [6\,..\,7], [17\,..\,17], [31\,..\,32], [33\,..\,33], [34\,..\,35]\,\}.$$

One can see, for example, that the presence of *hot* in position 2 has eliminated the interval $[1\,..\,3]$.

Let's try something more complex: "*pease* AND *porridge* AND (*hot* OR *cold*)". We have again to pick one interval from each of the three sets associated to "*pease*", "*porridge*" and "*hot* OR *cold*", and keep the minimal intervals among those spanned by such triples (see Figure 1):

$$\{\,[0\,..\,2], [1\,..\,3], [2\,..\,4], [3\,..\,5], [4\,..\,6], [5\,..\,7], [6\,..\,17], [7\,..\,31],$$
$$[21\,..\,32], [31\,..\,33], [32\,..\,34], [33\,..\,35], [34\,..\,36]\,\}.$$

From this rich semantic information, a number of different outputs can be computed. A simple snippet extraction algorithm would compute greedily the first k smallest nonoverlapping intervals of the antichain, which would yield, for $k = 3$, the intervals $[0\,..\,2]$, $[3\,..\,5]$, $[31\,..\,33]$, that is, "*Pease porridge hot*!", "*Pease porridge cold*!", and, again, "*Pease porridge hot*!".

A ranking scheme such as that proposed by Clarke and Cormack [3] would use the number and the length of these intervals to assign a score to the document with respect to the query. In a simplified setting, we can assume that each interval yields a score that is the inverse of its length. The resulting score for the query above would be

$$\frac{1}{|[0\,..\,2]|}+\frac{1}{|[1\,..\,3]|}+\cdots+\frac{1}{|[6\,..\,17]|}+\frac{1}{|[7\,..\,31]|}+\cdots\frac{1}{|[34\,..\,36]|}$$
$$=\frac{1}{3}+\frac{1}{3}+\cdots+\frac{1}{12}+\frac{1}{25}+\cdots\frac{1}{3}=\frac{177}{50}=3.54.$$

Clearly, documents with a large number of intervals are more relevant, and short intervals increase the document score more than long intervals, as short intervals are more informative. The score associated to "*hot*" would be just 3 (i.e., the number of occurrences). One can also incorporate positional information, making, for example, intervals appearing earlier in the document more important [1].

3 Preliminaries of lattice theory

In this section we are going to recall and briefly summarize some results from lattice and order theory that we will need in the rest of the paper. The reader can find more details, for example, in [8]. Note that in the previous sections we used the word "minimal" always meaning "by inclusion", but in the following sections its meaning will depend on the underlying order.

Let $P=(P,\leq)$ be a partially ordered set (poset). If P has a minimum element, or bottom, 0 (a maximum element, or top, 1, respectively) an *atom* (*coatom*, respectively) of P is an element $x\in P$ such that $0<y\leq x$ ($x\leq y<1$, respectively) implies $y=x$. We say that P is *atomic* (*coatomic*, respectively) if, for every $x\neq 0$ ($x\neq 1$, respectively), there exists an atom (coatom, respectively) a such that $a\leq x$ ($x\leq a$, respectively); it is *strongly (co)atomic* if, for every two $x,y\in P$ such that $x<y$, the poset induced by $[x\,..\,y]$ is (co)atomic. We say that P is *atomistic* (*coatomistic*, respectively) if every $x\neq 1$ ($x\neq 0$, respectively) is the $\vee$ ($\wedge$, respectively) of a set of atoms (coatoms, respectively). Note that (co)atomistic implies (co)atomic. For a given poset $P=(P,\leq)$, we let P^{op} denote the dual of P, that is, $(P,\leq)^{\mathrm{op}}=(P,\geq)$.

A $\vee$-*semilattice* ($\wedge$-*semilattice*) is a poset that has all binary joins (meets). A *lattice* is a $\vee$-semilattice that is also a $\wedge$-semilattice; it is *bounded* if it has a top and a bottom.

We say that a lattice $L=(L,\leq)$ is a *Heyting algebra* [11] if it is bounded and, for every x, the map $x\wedge\cdot$ has a right adjoint. In other words, for every $x,z\in L$ there exists a greatest y such that $x\wedge y\leq z$. Such a y is denoted by $x\to z$ and called the *pseudo-complement of x relative to z*.

Dually, it is a *Brouwerian algebra* [12][1] if it is bounded and, for every x, the map $x\vee\cdot$ has a left adjoint. In other words, for every $x,z\in L$ there exists a smallest y such that $z\leq x\vee y$. Such a y is denoted by $x-z$ and called the *pseudo-difference between x and y*.

For the remainder of this section, $L=(L,\leq)$ will be a *complete* lattice, that is, a poset with $\vee$ and $\wedge$ for arbitrary sets. We say that L is a *completely distributive lattice* if, for every collection $\mathscr{L}\subseteq 2^L$ of subsets of L, we have

$$\bigwedge\Bigl(\bigvee\mathscr{L}\Bigr)=\bigvee\Bigl(\bigwedge\mathscr{L}^\sharp\Bigr)$$

where $\mathscr{L}^\sharp$ is the family of all subsets of L that have a nonempty intersection with all sets of $\mathscr{L}$. Every completely distributive lattice is at the same time a Heyting and a Brouwerian algebra.

An element $x\in L$ is called:

[1] There is some confusion in the literature about the usage of the adjective "Brouwerian": some authors call "Brouwerian lattice" what other authors call a "Heyting algebra".

- $\vee$-*irreducible* ($\wedge$-irreducible, respectively) if and only if $x = a \vee b$ ($x = a \wedge b$, respectively) implies $x = a$ or $x = b$, for all $a, b \in L$;
- $\vee$-*prime* ($\wedge$-prime, respectively) if and only if $x \leq a \vee b$ ($x \geq a \wedge b$, respectively) implies $x \leq a$ or $x \leq b$ ($x \geq a$ or $x \geq b$, respectively), for all $a, b \in L$;
- *completely* $\vee$-*irreducible* ($\wedge$-irreducible, respectively) if and only if, for every $X \subseteq L$, $x = \bigvee X$ ($x = \bigwedge X$, respectively) implies $x \in X$;
- *completely* $\vee$-*prime* ($\wedge$-prime, respectively) if and only if, for every $X \subseteq L$, $x \leq \bigvee X$ ($x \geq \bigwedge X$, respectively) implies $x \leq a$ ($x \geq a$, respectively) for some $a \in X$.

In general, (complete) primality implies (complete) irreducibility. In (completely) distributive lattices, the reverse holds, too.

We say that L is *superalgebraic* if every element is the join of a set of completely $\vee$-prime elements. This notion is self-dual, that is, in a superalgebraic lattice every element is also a meet of a set of completely $\wedge$-prime elements. If a lattice is completely distributive, of course, one can replace primality with irreducibility in the definition above.

4 The Alexandrov completion

The *Alexandrov completion* of a poset $P = (P, \sqsubseteq)$ is defined as $\mathscr{L}(P) = (\mathscr{L}(P), \subseteq)$ where $\mathscr{L}(P)$ is the family of lower sets[2] of P.

The nature of $\mathscr{L}(P)$ is well known. Erné [8] studied the interplay between Alexandrov completions of posets, Alexandrov spaces (topological spaces where arbitrary unions of closed sets is still closed) satisfying the T_0 separation axiom (for every pair of points there is an open set containing exactly one of them), and completely distributive lattices. In general, Alexandrov completions are superalgebraic [8, Proposition 2.2.B]. However, we can prove more if P satisfies the *ascending chain condition* (ACC), that is, if it does not contain infinite chains of the form $x_0 \sqsubset x_1 \sqsubset x_2 \sqsubset \cdots$:

Theorem 1 ([8, Lemma 5.12]) If P satisfies the ACC then $\mathscr{L}(P)$ is a strongly coatomic, superalgebraic, completely distributive lattice.[3]

The smallest lower set containing $X \subseteq P$ will be denoted by $\downarrow X$ (analogously for the smallest upper set $\uparrow X$). Note that the map from P to $\mathscr{L}(P)$ sending an element x to $\downarrow\{x\}$, the *principal ideal associated with* x, is an *order embedding* (an injective function preserving and reflecting the order relations) and preserves all meets existing in P.

(Completely) irreducible elements. Since Alexandrov completions are superalgebraic and completely distributive, every element is the $\vee$ of a set of completely $\vee$-irreducible elements, and the $\wedge$ of a set of completely $\wedge$-irreducible elements.

The completely $\vee$-irreducible elements in the Alexandrov completion $\mathscr{L}(P)$ are in one-to-one correspondence with the elements of P, via the map sending $x \in P$ to $\downarrow\{x\}$. Since $\mathscr{L}(P)$ is strongly coatomic, there is no difference between $\vee$-irreducible and completely $\vee$-irreducible elements [5]. Moreover, each element has an *irredundant*[4] $\vee$-representation by (completely) $\vee$-irreducible elements [5, Proposition 6.4].

Also the completely $\wedge$-irreducible elements in $\mathscr{L}(P)$ are in one-to-one correspondence with the elements of P, this time via the map sending $x \in P$ to $P \setminus \uparrow\{x\}$. If P is coatomistic, the

[2] A set $X \subseteq P$ is a *lower set* if $x \sqsubseteq y \in X$ implies $x \in X$. *Upper sets* are defined dually.

[3] In fact, much more is true: every strongly coatomic, superalgebraic, completely distributive lattice is isomorphic to the Alexandrov completion of a poset satisfying the ACC. Moreover, such lattices are exactly the lattices of closed sets of *sober* T_0 Alexandrov spaces [8]. By a natural choice of morphisms, this correspondence can be made into a categorical equivalence.

[4] A $\vee$-representation X of x is *irredundant* if no proper subset of X can be used to represent x, that is, $x = \bigvee X \neq \bigvee Y$ for every $Y \subset X$. Analogously, one can define irredundant $\wedge$-representations.

set $P \setminus \uparrow\{x\}$ can be rewritten as the union of the principal ideals $\downarrow\{c\}$ for all the coatoms c not above x. We remark, however, that in this case there is no automatic equivalence between $\wedge$-irreducible and completely $\wedge$-irreducible elements, as $\mathscr{L}(P)$ is not strongly atomic in general. For the same reason, some elements might not have an irredundant $\wedge$-representation by completely $\wedge$-irreducible elements.

5 The antichain completion

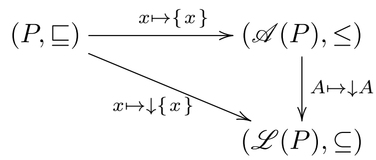
An *antichain* of a poset $P = (P, \sqsubseteq)$ is a subset $A \subseteq P$ whose elements are pairwise $\sqsubseteq$-incomparable. The set of antichains of P is denoted by $\mathscr{A}(P)$. It is possible to define a partial order on the antichains of P by letting, for all $A, B \in \mathscr{A}(P)$,

$$A \leq B \qquad \text{iff} \qquad \downarrow A \subseteq \downarrow B.$$

We can equivalently describe (perhaps in a more direct way) the order of $\mathscr{A}(P)$ as follows: given antichains A and B,

$$A \leq B \qquad \text{iff} \qquad \forall x \in A\ \exists y \in B \quad x \sqsubseteq y. \tag{1}$$

We call the partial order $\mathscr{A}(P) = (\mathscr{A}(P), \leq)$ the *antichain completion of* P. The map $\downarrow(-)$ defines an order embedding from $\mathscr{A}(P)$ to $\mathscr{L}(P)$, through which the embedding of P into $\mathscr{L}(P)$ factors as $x \mapsto \{x\}$:



Note that, as any order embedding, $\downarrow(-)$ reflects all joins and meets: many proofs about $\mathscr{A}(P)$ can be carried out by exploiting judiciously this property. In the main object of study of this paper, the embedding $\downarrow(-)$ will be an isomorphism (see Theorem 4 below), but this is not true in general.

In fact, $\mathscr{A}(P)$ may not even be a lattice: if we consider, for example, a poset P with elements $a_0 \sqsubset a_1 \sqsubset a_2 \sqsubset \cdots \sqsubset a_k \sqsubset \cdots$ and two additional, incomparable elements b and c that are greater than all of the a_i's, it is easy to see that the meet $\{b\} \sqcap \{c\}$ does not exist in $\mathscr{A}(P)$. However, $\mathscr{A}(P)$ is always a $\vee$-semilattice endowed with unique $\vee$-representations by $\vee$-irreducible elements:[5]

Theorem 2 Let $P = (P, \sqsubseteq)$ be a poset. The antichain completion $\mathscr{A}(P)$ is a $\vee$-semilattice, where the join of $A, B \in \mathscr{A}(P)$ is given by the maximal elements of $A \cup B$. Moreover, the $\vee$-irreducible elements of $\mathscr{A}(P)$ are the singleton antichains, and given $A \in \mathscr{A}(P)$

$$A = \bigvee_{x \in A} \{x\}$$

is the unique irredundant $\vee$-representation of A by $\vee$-irreducible elements.

Proof. The first statement is proved by noting that the maximal elements of $\downarrow A \cup \downarrow B$ are exactly the maximal elements of $A \cup B$; since the lower set of the maximal elements of $A \cup B$ is $\downarrow A \cup \downarrow B$, we obtain the result by the fact that order embeddings reflect joins. The second statement is a trivial consequence. For the third statement, note that A is equal to the maximal elements of $\bigcup_{x \in A} \downarrow\{x\}$, and that irredundant $\vee$-representations by $\vee$-irreducible elements (i.e., singleton antichains) are in bijection with antichains of elements of P. ∎

[5] The reader should note that we defined $\vee$-irreducible elements for *lattices*, but in fact the definition is sensible in any $\vee$-semilattice.

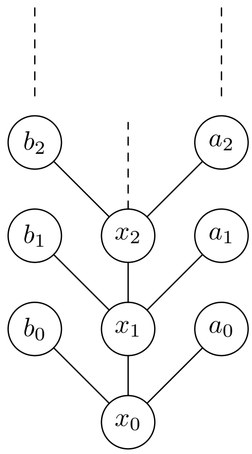


Figure 2: An example in which condition (2) of Theorem 2 is not satisfied, even assuming that existence of meets: $\downarrow\{a_i \mid i \in \mathbf{N}\} \cap \downarrow\{b_i \mid i \in \mathbf{N}\} = \{x_i \mid i \in \mathbf{N}\}$, but the latter set is not a lower set generated by an antichain.

The poset $a_0 \sqsubset a_1 \sqsubset a_2 \sqsubset \cdots \sqsubset a_k \sqsubset \cdots$ shows that an antichain completion is not necessarily $\vee$-complete: the join of the singleton antichains $\{a_i\}$, $i \in \mathbf{N}$, does not exist. On the other hand, the poset $a_0 \sqsubset a_1 \sqsubset a_2 \sqsubset \cdots \sqsubset a_k \sqsubset \cdots \sqsubset a$ shows that singleton antichains are not necessarily *completely* $\vee$-irreducible, as $\{a\} = \bigsqcup_{i \in \mathbf{N}} \{a_i\}$.

The antichain completion as a lattice. There is a simple necessary and sufficient condition that will turn the antichain completion into a lattice:

Theorem 3 Let $P = (P, \sqsubseteq)$ be a poset. Then, $\mathscr{A}(P)$ is a lattice iff for every $A, B \in \mathscr{A}(P)$ there is a $C \in \mathscr{A}(P)$ such that

$$\downarrow A \cap \downarrow B = \downarrow C, \tag{2}$$

and then $A \wedge B = C$; moreover, under this condition $\mathscr{A}(P)$ is also distributive. In particular, if P is a $\sqcap$-semilattice and $\mathscr{A}(P)$ is a lattice the meet of $A, B \in \mathscr{A}(P)$ is given by the maximal elements of $\{a \sqcap b \mid a \in A, b \in B\}$.

Proof. Condition (2) is sufficient for $\mathscr{A}(P)$ to be a lattice: indeed, it is equivalent to the fact that the image of $\mathscr{A}(P)$ with respect to $\downarrow(-)$ is a $\cap$-semilattice. Since we know already from Theorem 2 that it is a $\cup$-semilattice, we conclude that it is a lattice, so the same is true of $\mathscr{A}(P)$. Since the image of $\mathscr{A}(P)$ is a set lattice, it is distributive, and once again the same is true of $\mathscr{A}(P)$.

Let us show that condition (2) is also necessary: suppose by contradiction that $\mathscr{A}(P)$ is a lattice, but there are $A, B \in \mathscr{A}(P)$ such that the lower set $X = \downarrow A \cap \downarrow B$ is not generated by an antichain. In this case, X contains an element x that is not bounded by any maximal element of X. Still there is a C such that $C = A \wedge B$. Then necessarily $C \subseteq \downarrow C \subset X$. Since both $\downarrow C$ and X are lower sets, this implies that there is an $x \in X$ that is strictly greater or incomparable with any element of C: thus, $C < C \vee \{x\} \leq A \wedge B$, a contradiction.

For the second statement, recall that $\downarrow X = \downarrow Y$ implies that X and Y have the same maximal elements. Since it is immediate to show that

$$\downarrow\{a \sqcap b \mid a \in A, b \in B\} = \downarrow A \cap \downarrow B = \downarrow C$$

and C is an antichain, the thesis follows. ▮

Note that Condition (2) is weaker than the ACC (consider, for example, the case in which P is an ascending chain), and it is not tautological, as the example in Figure 2 shows.

Moreover, even in the best case of Theorem 3 (i.e., P a $\sqcap$-semilattice) $\mathscr{A}(P)$ may not be complete. Consider the poset $c_1 \sqsubset c_2 \sqsubset c_3 \sqsubset \cdots \sqsubset c_k \sqsubset \cdots \sqsubset c_{-k} \sqsubset \cdots \sqsubset c_{-3} \sqsubset c_{-2} \sqsubset c_{-1}$. Clearly, $\bigsqcap_{i<0}\{c_i\}$ and $\bigsqcup_{i>0}\{c_i\}$ do not exist in $\mathscr{A}(P)$.

Under the ACC, however, one can prove much more:

Theorem 4 ([7, Corollary 10.5]) If P satisfies the ACC, the order embedding $\downarrow(-) : \mathscr{A}(P) \to \mathscr{L}(P)$ is an isomorphism.

Thus, in case P satisfies the ACC the antichain completion of P enjoys all strong properties of the Alexandrov completion of P. Alternatively, the previous theorem shows that in the ACC case the antichain completion is a handy representation of the Alexandrov completion.

Remark. Theorems 2 and 3 remain true for the set of *finite* antichains; moreover, in the case of finite antichains the hypothesis of P being a $\sqcap$-semilattice becomes sufficient for obtaining a lattice, because the set $\{a \sqcap b \mid a \in A, b \in B\}$ is finite and its maximal elements provide the finite antichain C that makes condition (2) true.[6] The case of finite antichains is interesting from a computational point of view, as discussed in [2]. Moreover, every finite-$\vee$-preserving map from the poset of finite antichains to a finitely complete $\vee$-semilattice factors uniquely through the embedding $\downarrow(-)$: in other words, finite antichains are (isomorphic to) the free finitely complete $\vee$-semilattice over their base poset [14].

6 Atoms and coatoms

Let us first characterize the atoms of our completions:

Proposition 1 Given a poset $P = (P, \sqsubseteq)$, the atoms of the Alexandrov completion $\mathscr{L}(P)$ are exactly the lower sets of the form $\downarrow\{x\}$, where x is a minimal element of P; the atoms of the antichain completion $\mathscr{A}(P)$ are exactly the antichains of the form $\{x\}$, where x is a minimal element of P.

Proof. Consider an atom $A \subseteq P$ of $\mathscr{L}(P)$. If $A \neq \varnothing$ contains more than one element, pick $x \in A$ so that it is not the minimum of A; note that $A \setminus \uparrow\{x\}$ is still a lower set. Then, $\varnothing \subset A \setminus \uparrow\{x\} \subset A$, contradicting the fact that A is an atom. In the opposite direction, we just note that for a minimal $x \in P$ the set $\downarrow\{x\}$ contains just x, and thus covers $\varnothing$. The proof for $\mathscr{A}(P)$ is analogous. ∎

Since in a poset with a least element but no atoms every non-bottom element is the start of an infinite descending chain, we have the following:

Corollary 1 Consider a poset $P = (P, \sqsubseteq)$. If P has no minimal elements then each non-bottom element of $\mathscr{A}(P)$ or $\mathscr{L}(P)$ is the start of an infinite descending chain.

We now turn our attention to coatoms. Here we consider a scenario where P satisfies the ACC (hence we state it directly for $\mathscr{A}(P)$) and is coatomistic.

Proposition 2 Let $P = (P, \sqsubseteq)$ be a coatomistic poset satisfying the ACC, let 1_P be its top element, and let $\operatorname{coat}(P)$ be the set of its coatoms. Then, the top element of $\mathscr{A}(P)$ is $\{1_P\}$, and the only coatom is $\operatorname{coat}(P)$. Moreover, the elements A of $\mathscr{A}(P)$ from which an infinite ascending chain starts are exactly those for which $\operatorname{coat}(P) \setminus A$ is infinite, that is, such that there are infinite coatoms of P not in A.

[6] The hypothesis is not necessary, though. Consider the case of two incomparable elements a and b both smaller than two incomparable elements x and y: $x \sqcap y$ does not exist, but the set of finite antichains forms nonetheless a lattice.

Proof. Note that $\mathrm{coat}(P)$ is by definition an antichain. Since every element of P except for 1_P is a meet of coatoms, $\mathrm{coat}(P)$ is greater than every other element of $\mathscr{A}(P)$ except for the top $\{1_P\}$, and thus the only coatom.

Now, let $N_A = \mathrm{coat}(P) \setminus A$ be the set of coatoms not in A. If N_A is infinite, say $N_A = \{c_0, c_1, c_2, \dots\}$, the sequence

$$A < A \vee \{c_0\} < A \vee \{c_0, c_1\} < \dots$$

is an infinite ascending chain (the elements of the sequence are all distinct because the c_i's are always maximal).

Suppose by contradiction that N_A is finite and that

$$A = A_0 < A_1 < A_2 < \dots$$

is an infinite ascending chain starting from A. Note that the set of non-coatoms in A must be finite, as they must be the meet of elements from N_A.

We can assume without loss of generality that at each step of the chain one of the following two events happens:

- a new element is added to A;
- a non-coatom in A is substituted by a smaller element (and because of this, possibly some other elements are dropped from A).

Note that the first event can happen only a finite number of times, as such a new element must be the meet of coatoms in N_A. Moreover, the second event can happen only a finite number of times for a given element, or the ACC would be violated. This contradicts the existence of the chain. ∎

7 The Clarke–Cormack–Burkowski lattice

Let $O = (O, \sqsubseteq)$ be a totally ordered set. An *interval* of O is any subset $I \subseteq O$ such that, for all $x, y, z \in O$, if $x \sqsubseteq z \sqsubseteq y$ and $x, y \in I$ then $z \in I$. The set of all *finite* intervals of O is denoted by $\mathscr{I}_O$. If $\ell \sqsubseteq r$, the interval $[\ell\,..\,r]$ is nonempty, ℓ is its least element and r its greatest element; ℓ (r, respectively) will be called the *left* (*right*, resp.) extreme of the interval.

The following proposition shows some properties of the set of finite intervals ordered by *reverse* inclusion $(\mathscr{I}_O, \supseteq)$ that are relevant to its antichain completion:

Proposition 3 Let $O = (O, \sqsubseteq)$ be a totally ordered set. Then, the poset $(\mathscr{I}_O, \supseteq)$ enjoys the following properties:

1. it has a minimum iff O is finite (and the minimum is O); if O is infinite, $(\mathscr{I}_O, \supseteq)$ has no minimal element;
2. it satisfies the ACC;
3. it is a $\vee$-semilattice, and the join of I and J is $I \cap J$;
4. it is a $\wedge$-semilattice (and thus a lattice) iff O is locally finite[7], and the meet of $[\ell\,..\,r]$ and $[\ell'\,..\,r']$ is
$$[\ell \sqcap \ell'\,..\,r \sqcup r'];$$
5. it is coatomistic;

[7] A poset $(O, \sqsubseteq)$ is locally finite iff for every $x, y \in O$ the interval $[x\,..\,y]$ is finite.

6. it is a subposet of the Cartesian product of two totally ordered sets, hence its dimension is at most 2; it is exactly 2 iff $|O| > 1$.

Proof. The first item is trivial. For (2) an infinite ascending chain would be a sequence of finite intervals of the form $I_0 \supset I_1 \supset I_2 \supset \cdots$.
(3) Trivial.
(4) If O is not locally finite, there exists an interval $[\ell \mathrel{..} r]$ of infinite cardinality, and there exists no common $\supseteq$-lower bound to $[\ell]$ and $[r]$ in $\mathscr{I}_O$. So $(\mathscr{I}_O, \supseteq)$ cannot be a $\wedge$-semilattice. On the other hand, if O is locally finite the interval in the statement is clearly the smallest finite interval containing both $[\ell \mathrel{..} r]$ and $[\ell' \mathrel{..} r']$.
(5) $\varnothing$ is the largest element of $\mathscr{I}_O$. The elements just below it (the coatoms) are exactly the singleton intervals, and each finite interval $[\ell \mathrel{..} r]$ is the meet of $[\ell]$ and $[r]$.
(6) Consider the poset $P_O = (O, \sqsubseteq) \times (O, \sqsupseteq)$; the injection $\iota : \mathscr{I}_O \to P_O$ sending $[\ell \mathrel{..} r]$ to the pair (ℓ, r) respects the order. In fact, $\iota([\ell \mathrel{..} r]) \sqsubseteq \iota([\ell' \mathrel{..} r'])$ if and only if $\ell \sqsubseteq \ell'$ and $r \sqsupseteq r'$, which happens precisely when $[\ell \mathrel{..} r] \supseteq [\ell' \mathrel{..} r']$. Since P_O is a subposet of the Cartesian product of two total orders, its dimension is at most 2 [13]. ∎

Justified by the previous proposition, we assume from now on that $O = (O, \sqsubseteq)$ is a fixed locally finite, totally ordered set, so $(\mathscr{I}_O, \supseteq)$ is a $\wedge$-semilattice, and we can use Theorem 3 to compute meets easily. Locally finite, totally ordered sets correspond, up to isomorphisms, to the subsets of $(\mathbf{Z}, \leq)$.

The fact that O is locally finite implies that it is strongly atomic and coatomic. In particular, for every $x \in O$, either x is the greatest element of O or there exists a single element $x' \in O$ such that $x \sqsubset x'$ and $x \sqsubseteq y \sqsubseteq x'$ implies $y \in \{x, x'\}$: the element x' is called the *successor* of x and, when it exists, it is denoted by $x + 1$. Similarly, either x is the least element of O or there exists a single element $x'' \in O$ such that $x'' \sqsubset x$ and $x'' \sqsubseteq y \sqsubseteq x$ implies $y \in \{x, x''\}$: the element x'' is called the *predecessor* of x and, when it exists, it is denoted by $x - 1$.

Note that with a slight abuse of notation we will write $[\ell + 1 \mathrel{..} \rightarrow)$ even when ℓ has no successor to mean the empty set. Analogously for $(\leftarrow \mathrel{..} r - 1]$ when r has no predecessor. This convention makes it possible to avoid the introduction of open or semi-open intervals.

We are now ready to state our main definition:

Definition 1 Given a locally finite, totally ordered set O, the *Clarke–Cormack–Burkowski lattice on* O, denoted by $\mathscr{C}_O$, is the antichain completion of $(\mathscr{I}_O, \supseteq)$.

The above definition says that $\mathscr{C}_O$ is the set of antichains of finite intervals with respect to inclusion, with partial order given by

$$A \leq B \qquad \text{iff} \qquad \forall I \in A \; \exists J \in B \quad J \subseteq I,$$

which is just an explicit restatement of equation (1). In other words, $A \leq B$ if for every interval I in A there is some better interval (witness) J in B, where “better” means that the new interval J is contained in I. This corresponds to the intuition that smaller intervals are more *precise*, and thus convey more information.

In fact, the lattice $\mathscr{C}_O$ extends the definition provided by Clarke, Cormack and Burkowski [4] in two ways: first, our base set can contain also infinite (ascending, descending or bidirectional) chains; second, we allow for the empty interval. The first generalization makes it possible to avoid specifying the document length, allowing for infinite virtual documents, too. The second one gains us a top element that is useful in modeling negation.[8]

We remark that in our setting each antichain can be totally ordered by left (or, equivalently, right) extreme. This ordering (which we will call *natural*) is locally finite, so it makes sense to talk about the predecessor or successor of an interval in an antichain. An interval that has no successor (predecessor, respectively) in an antichain will be called the *last* (*first*,

[8]It is interesting to note that in our formulation $\mathscr{C}_\varnothing$ is the Boolean lattice (containing only “false” and “true”): if there is no spatial information, minimal-interval semantics reduces to the Boolean case.

```
procedure enumerate(A, ℓ, r) begin
   emit(A);
   for i = ℓ to n − 1 do
      for j = max{i, r} to n − 1 do
         enumerate(A ∪ { [i . . j] } , i + 1, j + 1)
      od
   od
end.
```

Algorithm 1: The CAT algorithm enumerating all antichains (except for 1). The base call is enumerate($\varnothing, 0, 0$).

respectively) interval of the antichain. With $\bigcup A$ we denote the union of the intervals in A: in other words, the set of elements of O covered by some interval in A.

We usually write $\mathscr{E}_n$ instead of $\mathscr{E}_{\{0,1,\ldots,n-1\}}$. It is known that $|\mathscr{E}_n| = C_{n+1} + 1$, where C_n is the n-th Catalan number [16, item 183].[9] There is a simple, recursive CAT (Constant Amortized Time) algorithm (Algorithm 1) that lists all antichains except for $\{\varnothing\}$: the algorithm greedily adds to a base antichain A a new interval, assuming that the first interval in natural order that can be added to A has left margin ℓ and right margin r. The correctness proof is an easy induction.

The following theorem instantiates the results of Section 4 to $\mathscr{E}_O$, using Proposition 3. The first two items show that $\mathscr{E}_n$ is exactly the lattice defined in [4].

Theorem 5 $\mathscr{E}_O$ is a strongly coatomic, superalgebraic, completely distributive lattice. Moreover, for any $A, B \in \mathscr{E}_O$, we have that:

1. $A \vee B$ is the set of all $\subseteq$-minimal elements of $A \cup B$;
2. $A \wedge B$ is the set of all $\subseteq$-minimal elements of the set
$$\{\, [\min\{\ell, \ell'\} \mathinner{.\,.} \max\{r, r'\}] \mid [\ell \mathinner{.\,.} r] \in A, [\ell' \mathinner{.\,.} r'] \in B \,\};$$
3. the least element 0 of $\mathscr{E}_O$ is $\varnothing$;
4. the greatest element 1 of $\mathscr{E}_O$ is $\{\varnothing\}$; no other element of $\mathscr{E}_O$ contains $\varnothing$;
5. $\mathscr{E}_O$ has exactly one coatom, $\{\,[x] \mid x \in O\,\}$, denoted by 1^-;
6. if O is finite, then $\mathscr{E}_O$ has exactly one atom, $\{O\}$; no other element of $\mathscr{E}_O$ contains O;
7. if O is infinite, 0 is not covered by any element (i.e., $\mathscr{E}_O$ has no atoms), and for each $A \neq 0$ there is an infinite descending chain starting at A;
8. if O is infinite, if $A \neq 1$ there is an infinite ascending chain starting at A iff there are infinite singleton intervals not in A.

8 Normal forms

$\vee$-representations. Theorem 2 has already provided the unique $\vee$-representation of an antichain A by $\vee$-irreducible elements, which we restate in our case:

[9] There is an off-by-one with respect to [16], due to the fact the author considers nonempty intervals only.

Theorem 6 Let $A \in \mathscr{E}_O$. Then,

$$A = \bigvee_{I \in A} \{ I \}.$$

and this is the only irredundant $\vee$-representation of A by $\vee$-irreducible elements of $\mathscr{E}_O$.

$\wedge$-representations. A $\wedge$-representation in terms of completely $\wedge$-irreducible elements is also easy to write, as explained in Section 4, but we immediately meet a computational issue: if O is infinite, a *finite* antichain cannot be represented by a finite set of completely $\wedge$-irreducible elements, as the $\wedge$ of any such set contains infinite singleton intervals. Moreover, when O is infinite $\mathscr{E}_O$ is not strongly atomic, and thus there are elements without an irredundant $\wedge$-representation by completely $\wedge$-irreducible elements [5, Proposition 6.3].

As we did for $\vee$-irreducible elements, we thus turn to the study of $\wedge$-irreducible elements and of the associated $\wedge$-representations, which, as we will see, are also unique and irredundant; moreover, such representations will be finite for finite antichains, and they will make it possible to describe the relative pseudo-complement in closed form.

Definition 2 We denote with $\mathscr{I}_O^\infty$ the set of all (finite and infinite) intervals of O. If $I \in \mathscr{I}_O^\infty$, we let $\tilde{} I$ denote the set $\{ [x] \mid x \notin I \}$, that is, the set of all singletons that are not contained in I.

The following proposition shows that the antichains of the form $\tilde{} I$ are precisely the $\wedge$-irreducible elements:

Proposition 4 An element $A \neq 1$ of $\mathscr{E}_O$ is $\wedge$-irreducible iff $A = \tilde{} I$ for some $I \in \mathscr{I}_O^\infty$.

Proof. Suppose that $A \neq 1$ and assume that there is $[\ell \,..\, r] \in A$ with $\ell \sqsubset r$. Since $\{ [\ell \,..\, r] \} = \{ [\ell] \} \wedge \{ [r] \}$, we have, letting $A' = A \setminus \{ [\ell \,..\, r] \}$,

$$A = A' \vee \{ [\ell \,..\, r] \} = A' \vee (\{ [\ell] \} \wedge \{ [r] \}) = (A' \vee \{ [\ell] \}) \wedge (A' \vee \{ [r] \}).$$

Now we notice that $A' \vee \{ [\ell] \}$ cannot contain $[\ell \,..\, r]$, because it contains $[\ell]$, and the same is true of $A' \vee \{ [r] \}$, so A is not $\wedge$-irreducible. Then necessarily $|I| = 1$ for all $I \in A$. Now, suppose that $(\bigcup A)^c$ is not an interval; then we have

$$A = (A \cup \{ [x] \}) \wedge (A \cup \{ [y] \}),$$

where x and y are taken so that they do not belong to the same convex component of $(\bigcup A)^c$, and once again A is not $\wedge$-irreducible.

Conversely, suppose by contradiction that $I \in \mathscr{I}_O^\infty$ and $A = \tilde{} I = A_1 \wedge A_2$. Then, for every $x \notin I$ we must have $[x] \in A_1 \cap A_2$. Thus, the other intervals in A_1 and A_2 must be subintervals of I (because A_1 and A_2 are antichains), and there must be at least one such interval both in A_1 and in A_2. The meet of those two intervals is entirely contained in I, which implies that such interval, or a smaller one, is in A, contradicting its definition. ∎

Figure 3 shows the lattice $\mathscr{E}_4$, with the $\wedge$-irreducible elements highlighted. Note that in the infinite case the antichains $\tilde{} I$ with I infinite are exactly the $\wedge$-irreducible elements that are *not* completely $\wedge$-irreducible.

We remark that $(\mathscr{I}_O^\infty, \supseteq)$ is a complete lattice where the meet of a family of intervals is the smallest interval containing all members of the family, and the join is just the intersection. Moreover, for any two intervals $I, J \in \mathscr{I}_O^\infty$ we have $\tilde{} I \leq \tilde{} J$ if and only if $I \supseteq J$. In other words,

Proposition 5 The map from $(\mathscr{I}_O^\infty, \supseteq)$ to the poset of $\wedge$-irreducible elements of $\mathscr{E}_O$ defined by the tilda operator is an isomorphism.

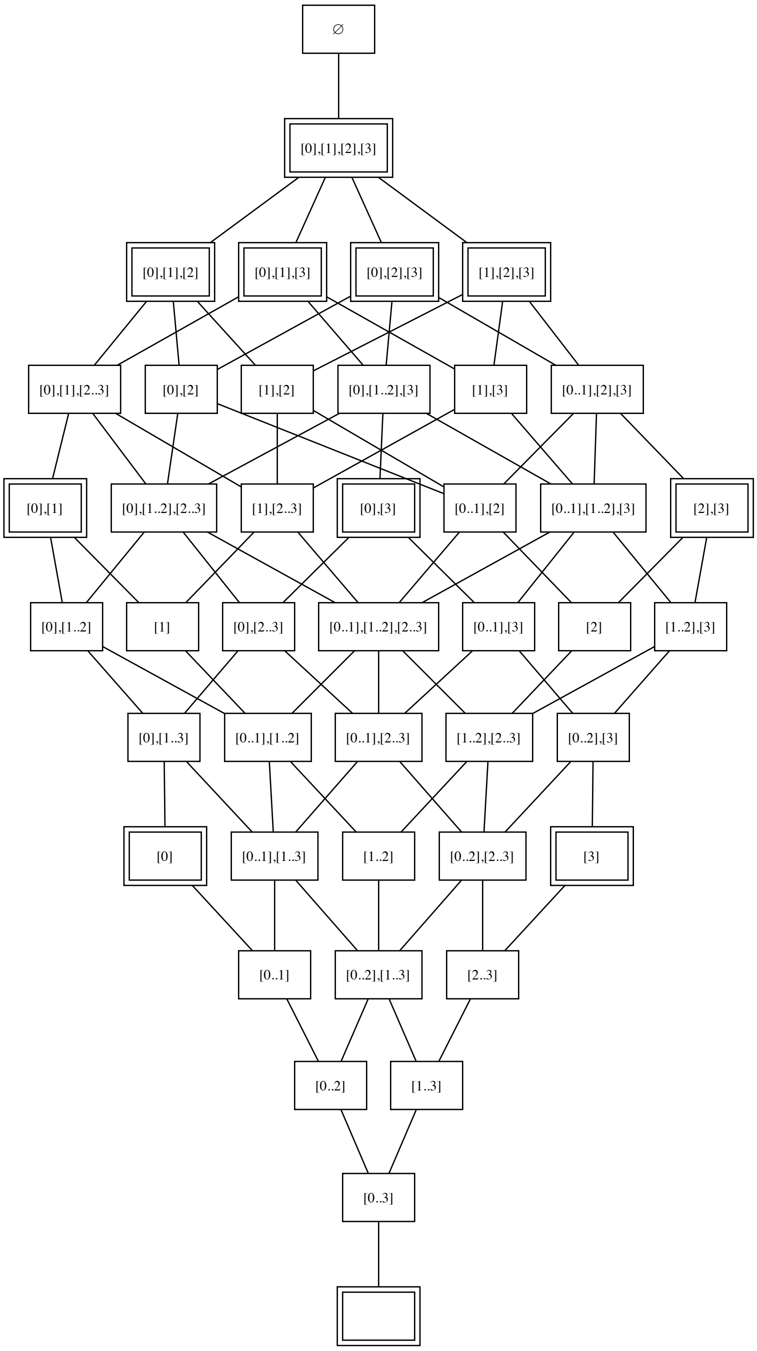


Figure 3: The lattice $\mathscr{E}_4$. The $\wedge$-irreducible elements have a double border. Each horizontal layer corresponds to a level set (the set of elements of a given rank; see Section 12).

The fact that $\wedge$-irreducible elements correspond naturally to intervals of $(\mathscr{I}_O^\infty, \supseteq)$ leads us to the following definition:

Definition 3 An interval $I \in (\mathscr{I}_O^\infty, \supseteq)$ is *critical* for an antichain $A \in \mathscr{E}_O$ iff it is minimal among the intervals of $(\mathscr{I}_O^\infty, \supseteq)$ that do not contain any interval of A. The set of critical intervals for A is denoted by $\mathscr{C}_A$.

Said otherwise, an interval is critical iff it is $\supseteq$-minimal in $\mathscr{I}_O^\infty \setminus \downarrow A$, where the $\downarrow$ operator is computed in $(\mathscr{I}_O^\infty, \supseteq)$ (as an antichain of $(\mathscr{I}_O, \supseteq)$ is also an antichain of $(\mathscr{I}_O^\infty, \supseteq)$).

The condition that a critical interval I does not contain any interval of A is trivially equivalent to $A \leq {}^\sim I$. By Proposition 5,

Lemma 1 Let $I \in \mathscr{I}_O^\infty$ and $A \in \mathscr{E}_O$. Then, I is critical for A iff $A \leq {}^\sim I$ and ${}^\sim I$ is minimal with this property.

Thus, the critical intervals of A are exactly those associated with the minimal $\wedge$-irreducible elements that dominate A. We now give an explicit characterization, which also shows that $\mathscr{C}_A$ is finite if A is finite:

Theorem 7 An interval $I \in \mathscr{I}_O^\infty$ is critical for an antichain A iff one of the following happens:

- $I = [\ell + 1 \,..\, r' - 1] \neq \varnothing$ and there are two intervals $[\ell \,..\, r], [\ell' \,..\, r'] \in A$ such that the latter is the successor of the former in A;
- $I = (\leftarrow \,..\, r - 1] \neq \varnothing$ and $[\ell \,..\, r]$ is the first element of A for some ℓ;
- $I = [\ell + 1 \,..\, \rightarrow) \neq \varnothing$ and $[\ell \,..\, r]$ is the last element of A for some r;
- $I = O$ and $A = 0$;
- $I = \varnothing$ and $A = 1^-$.

Proof. Assume first that $I \in \mathscr{I}_O^\infty$ is critical for A.

- If $I = (\leftarrow \,..\, x]$, then necessarily $x \sqsubset r$, where r is the right extreme of the first interval of A, otherwise the first interval would be included in I. Thus, by $\supseteq$-minimality $x = r - 1$; the case $I = [x \,..\, \rightarrow)$ is analogous.
- Suppose that $I = [x \,..\, y]$; since no interval of A can be included in I, every interval starts before x or ends after y, and we can assume that there is at least one interval starting before x and at least one interval ending after y (otherwise we fall in the previous case). Let $[\ell \,..\, r]$ be an interval of A starting before x, but with $\ell \sqsubset x$ as large as possible; since it cannot be the last one, its successor $[\ell' \,..\, r']$ will end after y (i.e., $y \sqsubset r'$). But then by $\supseteq$-minimality $\ell = x - 1$ and $y = r' - 1$.

The other implication follows by a trivial case-by-case analysis. ∎

Our goal now is to show that the mapping $A \mapsto \mathscr{C}_A$ is actually an isomorphism

$$\mathscr{A}(\mathscr{I}_O, \supseteq) \cong \mathscr{A}(\mathscr{I}_O^\infty, \subseteq)^{\mathrm{op}}.$$

More explicitly,

$$\mathscr{E}_O = \mathscr{A}(\mathscr{I}_O, \supseteq) \cong \mathscr{A}\big((\mathscr{I}_O^\infty, \supseteq)^{\mathrm{op}}\big)^{\mathrm{op}} = \mathscr{A}(\mathscr{I}_O^\infty, \subseteq)^{\mathrm{op}}.$$

In particular, this isomorphism will imply that elements of $\mathscr{E}_O$ have unique, irredundant $\wedge$-representations by antichains of $\wedge$-irreducible elements. Moreover, Theorem 2 and 3 will provide the rules to perform computations in $\mathscr{A}(\mathscr{I}_O^\infty, \subseteq)^{\mathrm{op}}$, and thus on $\wedge$-representations.

We also remark that the antichain completion $\mathscr{A}(\mathscr{I}_O^\infty, \subseteq)^{\mathrm{op}}$ is *not* an Alexandrov completion, because $(\mathscr{I}_O^\infty, \subseteq)$ does not satisfy the ACC in general. This consideration also explains why the two dualizations do not cancel out. The reader should contrast this theorem with the fact that $\mathscr{A}(P)$ is isomorphic to the antichain completion of its (completely) $\vee$-irreducible elements.

The order in $\mathscr{A}(\mathscr{I}_O^\infty, \subseteq)^{\mathrm{op}}$ can be written in elementary form by unwinding (1): given antichains S, T of intervals in $\mathscr{I}_O^\infty$, we have the following chain of equivalences:

$$S \leq T \text{ in } \mathscr{A}(\mathscr{I}_O^\infty, \subseteq)^{\mathrm{op}} \iff T \leq S \text{ in } \mathscr{A}(\mathscr{I}_O^\infty, \subseteq) \iff \forall I \in T\ \exists J \in S \quad I \subseteq J.$$

As a first step towards proving the isomorphism, we now characterize the meets of incomparable $\wedge$-irreducible elements in $\mathscr{E}_O$ in terms of the associated antichain in $(\mathscr{I}_O^\infty, \supseteq)$:

Proposition 6 Consider an antichain S of $(\mathscr{I}_O^\infty, \supseteq)$ and the meet (in $\mathscr{E}_O$) $M = \bigwedge\{\,{}^\sim I \mid I \in S\,\}$. Then, M contains the maximal intervals of $(\mathscr{I}_O, \supseteq)$ that are not contained in any interval of S.

Proof. Note that ${\downarrow}M = \bigcap_{I \in S} {\downarrow}{}^\sim I$. Since ${\downarrow}{}^\sim I$ contains exactly all intervals that are not contained in I, $\bigcap_{I \in S} {\downarrow}{}^\sim I$ contains exactly all intervals that are not contained in any interval of S, and since $(\mathscr{I}_O, \supseteq)$ satisfies the ACC M contains exactly the maximal elements among such intervals. ▮

Said otherwise, an interval belongs to M iff it is $\supseteq$-maximal in $\mathscr{I}_O \setminus {\uparrow}S$. An immediate corollary shows a special property of singleton intervals:

Corollary 2 Consider a nonempty antichain S of $(\mathscr{I}_O^\infty, \supseteq)$ and the meet $M = \bigwedge\{\,{}^\sim I \mid I \in S\,\}$. Then, $[x] \in M$ iff $x \notin \bigcup S$.

The next theorem describes explicitly the meets of incomparable $\wedge$-irreducible elements, using the following notation:

$$[\![\ell, r]\!] = {}^\sim[\ell + 1 \mathinner{..} \rightarrow) \wedge {}^\sim(\leftarrow \mathinner{..} r - 1] = \begin{cases} \{\,[\ell \mathinner{..} r]\,\} & \text{if } \ell \sqsubset r, \\ \{\,[x] \mid r \sqsubseteq x \sqsubseteq \ell\,\} & \text{otherwise.} \end{cases}$$

Theorem 8 Consider an antichain S of $(\mathscr{I}_O^\infty, \supseteq)$ and the meet $M = \bigwedge\{\,{}^\sim I \mid I \in S\,\}$. Then, M contains exactly the following pairwise disjoint sets of intervals:

- $[\![\ell' - 1, r + 1]\!]$, if S contains two consecutive intervals $[- \mathinner{..} r]$, $[\ell' \mathinner{..} -]$.
- ${}^\sim[\ell \mathinner{..} \rightarrow)$ if S has a first interval, and it is of the form $[\ell \mathinner{..} -]$.
- ${}^\sim(\leftarrow \mathinner{..} r]$ if S has a last interval, and it is of the form $[- \mathinner{..} r]$.
- ${}^\sim\varnothing$ if $\varnothing \in S$.
- $\{\,\varnothing\,\}$ if $S = \varnothing$.

Proof. We will denote with M^* the set of intervals specified by the statement of the theorem, and with M the meet $\bigwedge\{\,{}^\sim I \mid I \in S\,\}$. Proposition 6 and a simple case-by-case analysis shows that $M^* \subseteq M$. Moreover, Corollary 2 shows that the singleton intervals of M and M^* are the same.

We are left to prove that no other non-singleton interval can belong to M (hence $M^* = M$). Assume by contradiction that there is an interval $I = [a \mathinner{..} b] \in M \setminus M^*$, $a \sqsubset b$, which must be incomparable with all intervals in M^* (because M is an antichain, and $M^* \subseteq M$). Moreover, it cannot be contained in any interval of S.

Since I is not a singleton, by Corollary 2 it must be contained in $\bigcup S$ (otherwise there would be an $x \in I$ with $[x] \in M$, which is impossible). Consider the last interval $J = [\ell \mathinner{..} r]$ of S with $\ell \sqsubseteq a$. J must necessarily overlap with I, but then $r \sqsubset b$, as I cannot be contained in any interval in S. Since $I \subseteq \bigcup S$, the interval J must have a successor in S, say $[\ell' \mathinner{.\,.} r']$, and of course $a \sqsubset \ell'$, so I contains $[\ell' - 1 \mathinner{.\,.} r+1] \in M^*$, a contradiction.

A straightforward check shows that all the unions appearing in the description of M in the statement of the theorem are disjoint. ▮

We now fulfill our promise:

Theorem 9 Given $A \in \mathscr{E}_O$ and $S \in \mathscr{A}(\mathscr{I}_O^\infty, \subseteq)^{\mathrm{op}}$, the maps

$$A \overset{f}{\mapsto} \mathscr{C}_A$$
$$S \overset{g}{\mapsto} \bigwedge_{I \in S} {}^{\sim}I$$

define an isomorphism between $\mathscr{E}_O$ and $\mathscr{A}(\mathscr{I}_O^\infty, \subseteq)^{\mathrm{op}}$.

Proof. We start by proving monotonicity. Note that if $S, T \in \mathscr{A}(\mathscr{I}_O^\infty, \subseteq)^{\mathrm{op}}$, if we have $S \leq T$ then for each $I \in T$ there is a $J \in S$ such that $I \subseteq J$, which means ${}^{\sim}J \leq {}^{\sim}I$. Thus, trivially $S \leq T$ implies $g(S) \leq g(T)$.

Suppose now we have $A, B \in \mathscr{E}_O$ with $A \leq B$, and let $I \in f(B)$. Assume by contradiction that there is a $J \in A$ such that $J \subseteq I$. Since $A \leq B$ there must be a $K \in B$ such that $K \subseteq J \subseteq I$, contradicting the fact that, by definition, I does not contain any interval of B. Since I does not contain any interval of A, it must be contained by definition in a critical interval of A, that is, an interval in $f(A)$. We conclude that $f(A) \leq f(B)$.

We now prove that $g \circ f = \mathbf{1}$. By Definition 3 and Proposition 6, we have to show that

$$g(f(A)) = \max_{\supseteq}(\mathscr{I}_O \setminus \uparrow \min_{\supseteq}(\mathscr{I}_O^\infty \setminus \downarrow A)) = A,$$

where $\max_\supseteq$ ($\min_\supseteq$) computes the set of maximal (minimal) elements by reverse inclusion.[10]

We first show that $\uparrow \min_\supseteq(\mathscr{I}_O^\infty \setminus \downarrow A) = \mathscr{I}_O^\infty \setminus \downarrow A$. Let us start observing that given any family $X \subseteq \mathscr{I}_O^\infty$ whose elements all contain a common interval, the union $\bigcup_{J \in X} J$ is also an interval. As a consequence, for every $I \in \mathscr{I}_O^\infty \setminus \downarrow A$ we have that $\bigcup\{ J \in \mathscr{I}_O^\infty \setminus \downarrow A \mid I \subseteq J \}$ is still an interval, so it is a $\supseteq$-minimal element of $\mathscr{I}_O^\infty \setminus \downarrow A$. In other words, $\mathscr{I}_O^\infty \setminus \downarrow A$ has the property that each element dominates a $\supseteq$-minimal element, so taking the upper set of $\min_\supseteq(\mathscr{I}_O^\infty \setminus \downarrow A)$ gives back $\mathscr{I}_O^\infty \setminus \downarrow A$. Since $\mathscr{I}_O$ satisfies the ACC

$$\max_{\supseteq}(\mathscr{I}_O \setminus \uparrow \min_{\supseteq}(\mathscr{I}_O^\infty \setminus \downarrow A)) = \max_{\supseteq}(\mathscr{I}_O \setminus (\mathscr{I}_O^\infty \setminus \downarrow A)) = \max_{\supseteq}(\mathscr{I}_O \cap \downarrow A) = A.$$

Finally, we show that $f \circ g = \mathbf{1}$. This is equivalent to

$$f(g(S)) = \min_{\supseteq}(\mathscr{I}_O^\infty \setminus \downarrow \max_{\supseteq}(\mathscr{I}_O \setminus \uparrow S)) = S.$$

We first show that $\downarrow \max_\supseteq(\mathscr{I}_O \setminus \uparrow S) = \mathscr{I}_O^\infty \setminus \uparrow S$. Note that $\mathscr{I}_O \setminus \uparrow S$ is a lower set of finite intervals and $\mathscr{I}_O$ satisfies the ACC, so taking in $\mathscr{I}_O$ the lower set of $\max_\supseteq(\mathscr{I}_O \setminus \uparrow S)$ would give back $\mathscr{I}_O \setminus \uparrow S$. Building the lower set in $\mathscr{I}_O^\infty$, instead, we include also the *infinite* intervals of $\mathscr{I}_O^\infty$ that contain intervals of $\mathscr{I}_O \setminus \uparrow S$, and this gives exactly $\mathscr{I}_O^\infty \setminus \uparrow S$. Thus,

$$\min_{\supseteq}(\mathscr{I}_O^\infty \setminus \downarrow \max_{\supseteq}(\mathscr{I}_O \setminus \uparrow S)) = \min_{\supseteq}(\mathscr{I}_O^\infty \setminus (\mathscr{I}_O^\infty \setminus \uparrow S)) = \min_{\supseteq}(\uparrow S) = S. \text{ ▮}$$

Thanks to the isomorphism we have just presented, we have an analogous of Theorem 6 for the case of $\wedge$-representations:

[10] Recall that when we write $\downarrow A$, we are computing the lower set in $(\mathscr{I}_O^\infty, \supseteq)$ using the trivial injection $(\mathscr{I}_O, \supseteq) \to (\mathscr{I}_O^\infty, \supseteq)$, which extends to antichains.

Corollary 3 Let $A \in \mathscr{E}_O$. Then,

$$A = \bigwedge_{I \in \mathscr{C}_A} {}^\sim I$$

and this is the only irredundant $\wedge$-representation of A by $\wedge$-irreducible elements of $\mathscr{E}_O$.

Proof. By Theorem 9, $g(f(A)) = A$, so $\{{}^\sim I \mid I \in \mathscr{C}_A\}$ is an irredundant $\wedge$-representation of A. Suppose that

$$A = \bigwedge_{I \in S} {}^\sim I,$$

and that the representation is irredundant. Then, S must be an antichain, so $A = g(S)$, but then $\mathscr{C}_A = f(A) = f(g(S)) = S$. ∎

9 Relative pseudo-complement and pseudo-difference

Using the well-known characterization of the relative pseudo-complement in terms of $\wedge$-irreducible elements [15], we have that given $A, B \in \mathscr{E}_O$, if $B = \bigwedge_i C_i$ is the $\wedge$-representation of B, then

$$A \to B = \bigwedge\{C_i \mid A \not\leq C_i\}. \tag{3}$$

The right-hand side, albeit explicit, does not suggest an easy way to compute $A \to B$, but we will provide a closed form in the next section. As an aside, this characterization implies that the negation induced by the relative pseudo-complement is trivial: for every $A \neq 0$ we have $\neg A = A \to 0 = 0$, whereas $\neg 0 = 1$.

By duality, we have that given $A, B \in \mathscr{E}_O$, if $A = \bigvee_i C_i$ is the $\vee$-representation of A, then

$$A - B = \bigvee\{C_i \mid C_i \not\leq B\}.$$

We can further simplify this expression by noting that if $B = \bigvee_j D_j$ is the $\vee$-representation of B, then $C_i \not\leq B$ iff there is no D_j such that $C_i \leq D_j$:

$$A - B = \bigvee\{C_i \mid \nexists D_j \text{ such that } C_i \leq D_j\}.$$

Finally, since we know that such representations are simply given by sets of singleton antichains we can just write

$$A - B = \{I \in A \mid \nexists J \in B \text{ such that } J \subseteq I\}, \tag{4}$$

which characterizes pseudo-difference in $\mathscr{E}_O$ in a completely elementary way. Note that the Brouwerian complement of A, $1 - A$, is always equal to 1, except for $1 - 1 = 0$.

It is customary to define a few derived operators, such as the *symmetric pseudo-difference*:

$$A \,\Delta\, B = (A - B) \vee (B - A).$$

It is well known that in any Brouwerian algebra

$$(A \vee B) - (A \wedge B) = A \,\Delta\, B.$$

In our lattice, it is easy to verify that moreover

$$(A \vee B) - (A \,\Delta\, B) = A \cap B, \tag{5}$$

where $\cap$ is the standard set intersection. In other words, intersection of antichains can be reconstructed using only lattice operations. In particular, this means that the $\wedge$-semilattice of antichains ordered by set inclusion can be be reconstructed, too, since $A \subseteq B$ iff $A \cap B = A$.

Another straightforward but important property we will use is that

$$A - (B \vee C) = (A - B) \cap (A - C). \tag{6}$$

We remark that the observations above hold true in any Brouwerian algebra that has unique $\vee$-representations by $\vee$-irreducible elements. The only difference, in the general case, is that $(A \vee B) - A \Delta B$ will be equal to the join of the $\vee$-irreducible elements in the intersection of the representations of A and B.

9.1 An elementary characterization of the relative pseudo-complement

The simplicity of the elementary characterization of the pseudo-difference (4) is due the fact that we are implicitly representing the elements of $\mathscr{E}_O$ by $\vee$-irreducible elements. Characterizing the relative pseudo-complement would be analogously easy if we were using a representation based on $\wedge$-irreducible elements.

It is possible, however, to obtain an elementary characterization of the relative pseudo-complement $A \to B$ based on $\vee$-irreducible elements (i.e., on an antichain of $\mathscr{I}_O$). To do that, first convert B to a representation by $\wedge$-irreducible elements using Theorem 7; then, select those elements that are not greater than A, as in (3). Finally, convert the set of selected elements to a representation by $\vee$-irreducible elements using Theorem 8. In this section we will make this process explicit when B is finite.

Note that for $A \in \mathscr{E}_O$, $I \in \mathscr{I}_O^\infty$ we have $A \not\leq {\sim}I$ iff $\{I\} \leq A$. Then, if $A, B \in \mathscr{E}_O$ with $A \neq 1$, $B \neq 0, 1$ finite[11], say, $B = \{[\ell_i \mathinner{..} r_i] \mid 0 \leq i < n\}$, $A \to B$ is the meet of the following $\wedge$-irreducible elements:

- ${\sim}(\leftarrow \mathinner{..} r_0 - 1]$, if $\{(\leftarrow \mathinner{..} r_0 - 1]\} \leq A$;
- ${\sim}[\ell_{i-1} + 1 \mathinner{..} r_i - 1]$, for $0 < i < n$, if $\{[\ell_{i-1} + 1 \mathinner{..} r_i - 1]\} \leq A$;
- ${\sim}[\ell_{n-1} + 1 \mathinner{..} \to)$, if $\{[\ell_{n-1} + 1 \mathinner{..} \to)\} \leq A$.

Given this description, by a tedious but straightforward application of Theorem 8 we can provide a description in terms of $\vee$-irreducible elements:

Theorem 10 Let $A, B \in \mathscr{E}_O$ with $A \neq 1$, $B \neq 0, 1$ finite, and $B = \{[\ell_i \mathinner{..} r_i] \mid 0 \leq i < n\}$. Define T as the set of all i with $0 < i < n$ such that $\{[\ell_{i-1} + 1 \mathinner{..} r_i - 1]\} \leq A$, and, if $T \neq \varnothing$, $t_{\min} = \min T$, $t_{\max} = \max T$, $T^+ = T \setminus \{t_{\min}\}$. For every $i \in T^+$, let $P(i)$ denote the predecessor of i in T (i.e., the largest element of T smaller than i). Then, if $T \neq \varnothing$,

$$A \to B = U^- \cup \bigcup_{i \in T^+} [\![\ell_{i-1}, r_{P(i)}]\!] \cup U^+$$

where

$$U^- = \begin{cases} [\![\ell_{t_{\min}-1}, r_0]\!] & \text{if } \{(\leftarrow \mathinner{..} r_0 - 1]\} \leq A, \\ {\sim}[\ell_{t_{\min}-1} + 1 \mathinner{..} \to) & \text{otherwise,} \end{cases}$$

$$U^+ = \begin{cases} [\![\ell_{n-1}, r_{t_{\max}}]\!] & \text{if } \{[\ell_{n-1} + 1 \mathinner{..} \to)\} \leq A, \\ {\sim}(\leftarrow \mathinner{..} r_{t_{\max}} - 1] & \text{otherwise.} \end{cases}$$

If $T = \varnothing$,

$$A \to B = \begin{cases} [\![\ell_{n-1}, r_0]\!] & \text{if } \{[\ell_{n-1} + 1 \mathinner{..} \to)\}, \{(\leftarrow \mathinner{..} r_0 - 1]\} \leq A, \\ {\sim}[\ell_{n-1} + 1 \mathinner{..} \to) & \text{otherwise, if } \{[\ell_{n-1} + 1 \mathinner{..} \to)\} \leq A, \\ {\sim}(\leftarrow \mathinner{..} r_0 - 1] & \text{otherwise, if } \{(\leftarrow \mathinner{..} r_0 - 1]\} \leq A, \\ 1 & \text{otherwise.} \end{cases}$$

In the case of an infinite antichain B, if $T \neq \varnothing$ has no minimum (maximum), one needs to eliminate U^- (U^+, respectively) from the result.

[11] It is immediate to verify that $1 \to A = A$, $A \to 1 = 0 \to 0 = 1$, and that $A \to 0 = 0$ for $A \neq 0$.

10 Containment operators

Along the lines of Brouwerian difference, it is possible to define four binary operators[12] that have specific applications in information retrieval:

$$A \ntrianglerighteq B = \{ I \in A \mid \nexists J \in B \text{ such that } J \subseteq I \} = A \setminus {\downarrow} B \tag{7}$$
$$A \trianglerighteq B = \{ I \in A \mid \exists J \in B \text{ such that } J \subseteq I \} = A \cap {\downarrow} B \tag{8}$$
$$A \ntrianglelefteq B = \{ I \in A \mid \nexists J \in B \text{ such that } J \supseteq I \} = A \setminus {\uparrow} B \tag{9}$$
$$A \trianglelefteq B = \{ I \in A \mid \exists J \in B \text{ such that } J \supseteq I \} = A \cap {\uparrow} B \tag{10}$$

These operators are named "not containing", "containing", "not contained in" and "contained in", respectively, in [4]. They are an essential tool in finding part of a text satisfying further positional constraints: for example, if K denotes the antichain of intervals of text where a certain set of keywords appears, and T denotes the antichain of intervals specifying which parts of the text are titles, $K \trianglelefteq T$ contains the intervals where the keywords appear inside a title.

Note that:

$$\begin{aligned} A \ntrianglerighteq B &= A - B \\ A \trianglerighteq B &= A - (A - B). \end{aligned}$$

The first equality is due to (4), whereas the second follows from observing that the pseudo-difference with a subset is just complementation. Other operators definable by pseudo-difference are the strict versions of $\ntrianglerighteq$ and $\trianglerighteq$:

$$\begin{aligned} A - (B - A) = A \ntriangleright B &= \{ I \in A \mid \nexists J \in B \text{ such that } J \subset I \} \\ A - (A - (B - A)) = A \triangleright B &= \{ I \in A \mid \exists J \in B \text{ such that } J \subset I \} \end{aligned}$$

These operators are important because $A \ntriangleright B$ contains exactly the intervals in A that do not disappear by minimization in $A \vee B$. In a formula,

$$A \vee B = (A \ntriangleright B) \vee (B \ntriangleright A) = (A \ntriangleright B) \cup (B \ntriangleright A). \tag{11}$$

The two operators $\trianglelefteq$ and $\ntrianglelefteq$ do not seem to admit an easy description in terms of lattice operators, albeit we can always, of course, resort to unique $\vee$-representations by $\vee$-irreducible elements: the definitions (7)-(10) can be indeed rewritten in any lattice in which such a representation exists, reading the $\in$ symbol as "is an element of the $\vee$-representation" and replacing containment with $\leq$.

Since these operators are defined by taking a subset of elements (of the $\vee$-representation) of the first operand that depends only on the second operand, any chain of applications of these operands can be permuted without affecting the results. In other words, as noted in [4],

$$(A \ominus B) \oplus C = (A \oplus C) \ominus B \quad \ominus, \oplus \in \{ \trianglerighteq, \trianglelefteq, \ntrianglerighteq, \ntrianglelefteq \}.$$

We note that an immediate consequence of $\ntrianglerighteq$ being a lower adjoint is that

$$(A \vee B) \ntrianglerighteq C = (A \ntrianglerighteq C) \vee (B \ntrianglerighteq C).$$

This property is a form of distributivity. Some pseudo-distributivity properties are listed in the following:

Theorem 11 Let $A, B, C \in \mathscr{E}_O$. Then,

$$\begin{aligned} A \ntrianglerighteq (B \wedge C) &= (A \ntrianglerighteq B) \vee (A \ntrianglerighteq C) \\ A \ntrianglerighteq (B \vee C) &= (A \ntrianglerighteq B) \cap (A \ntrianglerighteq C) \\ A \trianglerighteq (B \wedge C) &= (A \trianglerighteq B) \cap (A \trianglerighteq C) \end{aligned}$$

[12] A warning: although the notation is reminiscent of binary *relations*, the reader should keep in mind that we are defining binary *operations*.

Proof. The first equality is well known [12]; the second was observed in (6). For the third one,

$$\begin{aligned} A \trianglerighteq (B \wedge C) &= A - (A - (B \wedge C)) \\ &= A - ((A - B) \vee (A - C)) \\ &= (A - (A - B)) \cap (A - (A - C)) \\ &= (A \trianglerighteq B) \cap (A \trianglerighteq C). \ \blacksquare \end{aligned}$$

Finally, other distributivity properties involving the containment operators hold:

Theorem 12
1. $\trianglelefteq$ and $\ntrianglerighteq$ are left-distributive over $\vee$; that is, for every $A, B, C \in \mathscr{E}_O$ and $\oplus \in \{\trianglelefteq, \ntrianglerighteq\}$ we have
$$(A \vee B) \oplus C = (A \oplus C) \vee (B \oplus C).$$
2. $\trianglerighteq$ is right-distributive over $\vee$; that is, for every $A, B, C \in \mathscr{E}_O$ we have
$$A \trianglerighteq (B \vee C) = (A \trianglerighteq B) \vee (A \trianglerighteq C).$$
3. No other distributive property of any of the operators $\{\trianglelefteq, \ntrianglerighteq, \trianglerighteq, \ntrianglelefteq\}$ over any of $\{\vee, \wedge\}$ holds.

Proof. For (1), as we observed above, the case $\oplus = \ntrianglerighteq$ depends on $\ntrianglerighteq$ being a lower adjoint.

Let us prove the case $\oplus = \trianglelefteq$. The left-hand side is formed by the set of minimal[13] intervals in $A \cup B$ that are contained in some interval of C. Such intervals are either in $A \trianglelefteq C$ or $B \trianglelefteq C$, and they are obviously minimal in $(A \trianglelefteq C) \cup (B \trianglelefteq C) \subseteq A \cup B$. Thus,

$$(A \vee B) \trianglelefteq C \subseteq (A \trianglelefteq C) \vee (B \trianglelefteq C).$$

The right-hand side is made by minimal intervals in $(A \trianglelefteq C) \cup (B \trianglelefteq C)$. Let I be such an interval, and assume without loss of generality that it belongs to A. We have necessarily that $I \in A \vee B$, for otherwise there should be an interval $J \in B$ such that $J \subset I$. Such an interval would be contained *a fortiori* in some interval of C, and this contradicts the minimality of I.

For (2),

$$\begin{aligned} A \trianglerighteq (B \vee C) &= A - (A - (B \vee C)) \\ &= A - ((A - B) \cap (A - C)) \\ &= A - ((A - B) \wedge (A - C)) \\ &= (A - (A - B)) \vee (A - (A - C)) \\ &= (A \trianglerighteq B) \vee (A \trianglerighteq C), \end{aligned}$$

where we used Theorem 11, and the fact that intervals in $(A - B) \wedge (A - C)$ but not in $(A - B) \cap (A - C)$ are spans of two distinct intervals of A, so they cannot be included in an interval of A.

For (3), there is one counterexample for each instance in Table 1 and 2. $\blacksquare$

[13] From this section, "minimal" will always mean minimal by inclusion, that is, $\supseteq$-maximal.

$\oplus$	$A \oplus (B \vee C)$	$(A \vee B) \oplus C$
$\not\trianglerighteq$	$A = C = \{\,[a]\,\}$, $B = \{\,[a\,..\,b]\,\}$	—
$\trianglerighteq$	—	$A = C = \{\,[a\,..\,b]\,\}$, $B = \{\,[a]\,\}$
$\not\trianglelefteq$	$A = B = \{\,[a]\,\}$, $C = \{\,[b]\,\}$	$A = \{\,[a\,..\,b]\,\}$, $B = C = \{\,[a]\,\}$
$\trianglelefteq$	$A = B = \{\,[a\,..\,b]\,\}$, $C = \{\,[a]\,\}$	—

Table 1: Counterexamples to missing distributivity laws for containment operators over $\vee$.

$\oplus$	$A \oplus (B \wedge C)$	$(A \wedge B) \oplus C$
$\not\trianglerighteq$	$A = \{\,[a], [b]\,\}$, $B = \{\,[a]\,\}$, $C = \{\,[b]\,\}$	$A = \{\,[a]\,\}$, $B = \{\,[b]\,\}$, $C = \{\,[c]\,\}$, $a \sqsubset c \sqsubset b$
$\trianglerighteq$	$A = \{\,[a], [b]\,\}$, $B = \{\,[a]\,\}$, $C = \{\,[b]\,\}$	$A = \{\,[a]\,\}$, $B = \{\,[b]\,\}$, $C = \{\,[c]\,\}$, $a \sqsubset c \sqsubset b$
$\not\trianglelefteq$	$A = \{\,[a\,..\,b]\,\}$, $B = \{\,[a]\,\}$, $C = \{\,[b]\,\}$	$A = \{\,[a]\,\}$, $B = \{\,[b]\,\}$, $C = \{\,[a], [b]\,\}$
$\trianglelefteq$	$A = \{\,[a\,..\,b]\,\}$, $B = \{\,[a]\,\}$, $C = \{\,[b]\,\}$	$A = \{\,[a]\,\}$, $B = \{\,[b]\,\}$, $C = \{\,[a], [b]\,\}$

Table 2: Counterexamples to distributivity laws for containment operators over $\wedge$.

11 Ordered operators

For completeness, we introduce two final operators that are fundamental in information retrieval. For $A, B \neq 1$ the *ordered non-overlapping meet operator*

$$A < B = \bigvee \big\{\,\{\,[\min I\,..\,\max J]\,\}\mid I \in A, J \in B, \max I \sqsubset \min J\,\big\}$$

returns the minimal intervals spanned by an interval of A followed (without overlaps) by an interval of B. Such an operator is useful when looking, for example, for a sequence of terms in a given order in a window of size k. It suffices to compute the antichain generated by the operator, and check whether it contains some interval of length at most k.

The *block operator*

$$A \,\square\, B = \{\,[\min I\,..\,\max J] \mid I \in A, J \in B, \max I + 1 = \min J\,\}$$

returns the intervals formed by two consecutive intervals from A and B (such intervals form an antichain). The block operator is used to implement phrase search. We define 1 to be the identity for both operators.

Both operators are associative, and moreover the ordered meet distributes with respect to joins: more precisely,

Theorem 13 We have

$$(A \vee B) < C = (A < C) \vee (B < C)$$
$$A < (B \vee C) = (A < B) \vee (A < C)$$

Proof. If any of the antichain is 1, the result is trivial. We prove $(A \vee B) < C = (A < C) \vee (B < C)$. The other case is similar.

Consider an interval I in $(A \vee B) < C$. It is formed by the span of an interval that we can assume without loss of generality to belong to A, and of an interval in C. We know that I is minimal among all spans between elements of $A \vee B$ and C. If I does not belong to $A < C$, this means that there is another smaller span, which contradicts minimality in $(A \vee B) < C$. If I belongs to $A < C$ but not to $(A < C) \vee (B < C)$, it means that there's a smaller span in $(B < C)$, again contradicting the minimality of I. So $(A \vee B) < C \subseteq (A < C) \vee (B < C)$

Consider now an interval J in $(A < C) \vee (B < C)$, and assume without loss of generality that it comes from $A < C$. If J does not belong to $(A \vee B) < C$, it means that there is a smaller, minimal interval in $(A \vee B) < C$. But such an interval is *a fortiori* minimal in $(A < C)$ or $(B < C)$, which contradicts the minimality of J. ∎

The example $A = \{[0\,..\,1]\}$, $B = \{[0]\}$, $C = \{[2]\}$ shows that the block operator does not satisfy the first distributivity law of the theorem; the example $A = \{[0]\}$, $B = \{[1\,..\,2]\}$, $C = \{[2]\}$ that it does not satisfy the second one.

12 Ranking and height

Every finite distributive lattice L is *ranked*: there is a rank function $\mathfrak{r} : L \to \mathbf{N}$ such that $\mathfrak{r}(x) = \mathfrak{r}(y) + 1$ whenever x covers y, and $\mathfrak{r}(0) = 0$. For a finite distributive lattice, the rank of an element is the number of $\vee$-irreducible elements it dominates. This characterization makes the explicit computation of the rank very simple, if O is finite; from now on, we let $O = \{0, 1, \ldots, n-1\}$.

Theorem 14 Let $[\ell\,..\,r] \in \mathscr{I}_O$; then,

$$\mathfrak{r}(\{[\ell\,..\,r]\}) = (1 + \ell)(n - r).$$

Moreover, let $k > 0$ and $X = \{[\ell_0\,..\,r_0], \ldots, [\ell_k\,..\,r_k]\} \in \mathscr{E}_O$ with $\ell_0 \sqsubset \ell_1 \sqsubset \ldots \sqsubset \ell_k$. Then,

$$\mathfrak{r}(X) = (1 + \ell_0)(n - r_0) + \sum_{i=1}^{k} (\ell_i - \ell_{i-1})(n - r_i).$$

Proof. Since $\vee$-irreducible elements are singleton antichains, to rank $\{[\ell\,..\,r]\}$ we must compute the number of intervals that contain $[\ell\,..\,r]$, which are exactly those of the form $[\ell'\,..\,r']$ with $\ell' \sqsubseteq \ell$ and $r \sqsubseteq r'$, resulting in the first formula above. For the second formula, we recall that in every distributive lattice

$$\mathfrak{r}(A) + \mathfrak{r}(B) = \mathfrak{r}(A \vee B) + \mathfrak{r}(A \wedge B);$$

taking $A = \{[\ell_0\,..\,r_0], \ldots, [\ell_{k-1}\,..\,r_{k-1}]\}$, $B = \{[\ell_k\,..\,r_k]\}$ and observing that $A \wedge B = \{[\ell_{k-1}\,..\,r_k]\}$, we have that

$$\mathfrak{r}(X) = \mathfrak{r}(\{[\ell_0\,..\,r_0], \ldots, [\ell_{k-1}\,..\,r_{k-1}]\}) + \mathfrak{r}([\ell_k\,..\,r_k]) - \mathfrak{r}([\ell_{k-1}\,..\,r_k]).$$

Applying the same idea k times, we have

$$\begin{aligned}\mathfrak{r}(X) = \sum_{i=0}^{k} \mathfrak{r}(\{[\ell_i\,..\,r_i]\}) - \sum_{i=1}^{k} \mathfrak{r}(\{[\ell_{i-1}\,..\,r_i]\}) \\ = \sum_{i=0}^{k} (1 + \ell_i)(n - r_i) - \sum_{i=1}^{k} (1 + \ell_{i-1})(n - r_i),\end{aligned}$$

whence the result. ∎

Corollary 4 The rank of the top element of $\mathscr{E}_n$ is $1 + n(n+1)/2$. Hence $\mathscr{E}_n$ has height $2 + n(n+1)/2$.

Proof. The rank of $1^- = \{[0], [1], \ldots, [n-1]\}$ is

$$\mathfrak{r}(1^-) = n + \sum_{i=1}^{n-1} (n - i) = n(n+1)/2;$$

the statement follows immediately. ∎

13 Conclusions and open problems

We have presented a detailed analysis of the Clarke–Cormack–Burkowski lattice on a locally finite, totally ordered set. Besides analyzing its basic properties, we have provided elementary characterizations for all operations, including relative pseudo-complement and pseudo-difference, and a closed formula for the rank of an element.

It is easy to check that join and meet of two elements are computable in linear time using simple greedy algorithms. More efficient algorithms for the computation of n-ary joins and meets, as well as algorithms for the containment and ordered operators, are described in [2]. Finally, in the description of $A \to B$ in the statement of Theorem 10 all intervals appear exactly once, and moreover they are enumerated in their natural order. This implies that in the finite case it can be turned into an $O(|A| + |B| + |A \to B|)$ algorithm to compute the relative pseudo-complement: first, an easy $O(|A| + |B|)$ greedy algorithm computes T. Then, depending on the case conditions (which can be all evaluated in $O(|A|)$) a simple loop emits the output (which cannot take more time than $O(|A \to B|)$). Note that in this particular case we have to state explicitly that the algorithm is linear *both in the input and in the output size*, as the output might have roughly the same size as the base set even for a constant-sized input (e.g., $\{[n-1]\} \to \{[n-2]\} = \{[0], [1], \ldots, [n-2]\}$).[14] All these algorithms have been implemented in LaMa4J (Lattice Manipulation for Java), a free Java library supporting computation in lattices.[15]

An interesting open question is that relative to the *width* of the lattice $\mathscr{E}_n$, that is, the cardinality of a maximum antichain. The sequence of the first few widths of $\mathscr{E}_n$ for $n = 0, 1, \ldots$ is

$$1, 2, 3, 7, 17, 44, 118, 338, 1003, 3039, 9466, 30009, \ldots,$$

which does not appear in the "On-Line Encyclopedia of Integer Sequences" [10]. At these sizes, the width coincide with the cardinality of the maximum *level sets* (a level set is the set of elements of given rank). The fact that this happens at all sizes is known at the *Sperner property* [6]. It would be interesting to find a closed form for the sequence above, and prove or disprove the Sperner property for $\mathscr{E}_n$. Finally, it would be interesting to relax the totality assumption on the base set O, to include, for instance, tree-structured documents.

14 Acknowledgements

We are very grateful to the anonymous reviewer and to the editor: their suggestions improved enormously the quality of the presentation.

References

[1] Paolo Boldi and Sebastiano Vigna. MG4J at TREC 2005. In Ellen M. Voorhees and Lori P. Buckland, editors, *The Fourteenth Text REtrieval Conference (TREC 2005) Proceedings*, number SP 500-266 in Special Publications. NIST, 2005. `http://mg4j.di.unimi.it/`.

[2] Paolo Boldi and Sebastiano Vigna. Efficient optimally lazy algorithms for minimal-interval semantics. *Theoretical Computer Science*, 648:8–25, 2016.

[3] Charles L. A. Clarke and Gordon V. Cormack. Shortest-substring retrieval and ranking. *ACM Trans. Inf. Syst*, 18(1):44–78, 2000.

[4] Charles L. A. Clarke, Gordon V. Cormack, and Forbes J. Burkowski. An algebra for structured text search and a framework for its implementation. *Comput. J.*, 38(1):43–56, 1995.

[14] If we admit a symbolic representations for infinite sets of the form $\tilde{}[\ell \,..\, \to)$ and $\tilde{}(\leftarrow \,..\, r]$, the time bound becomes $O(|A| + |B|)$ even when O is infinite, as long as the input antichains are finitely representable.

[15] `http://lama4j.di.unimi.it/`

[5] Peter Crawley and Robert P. Dilworth. *Algebraic Theory of Lattices.* Prentice Hall, 1973.

[6] Konrad Engel. *Sperner Theory.* Cambridge Solid State Science Series. Cambridge University Press, 1997.

[7] Marcel Erné. *Einführung in die Ordnungstheorie.* Bibliographisches Institut, 1982.

[8] Marcel Erné. The ABC of order and topology. In *Category Theory at Work*, volume 18 of *Res. Expo. Math.*, pages 57–83. Heldermann Verlag, 1991.

[9] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete Mathematics.* Addison–Wesley, second edition, 1994.

[10] OEIS Foundation Inc. The on-line encyclopedia of integer sequences, 2017.

[11] Peter T. Johnstone. *Stone Spaces.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1986.

[12] J.C.C. McKinsey and Alfred Tarski. On closed elements in closure algebras. *Annals of Mathematics*, 47(1):122–162, 1946.

[13] Oynstein Ore. *Theory of Graphs*, volume XXXVIII of *American Mathematical Society Colloquium Publications.* American Mathematical Society, 1962.

[14] Paulo Ribenboim. Ordering the set of antichains of an ordered set. *Collect. Math.*, 46(1–2):159–170, 1995.

[15] Dorothy P. Smith. Meet-irreducible elements in implicative lattices. *Proc. Amer. Math. Soc.*, 34(1):57–62, 1972.

[16] Richard P. Stanley. *Catalan Numbers.* Cambridge University Press, 2015.