Stepwise Construction of the Dedekind-MacNeille Completion

Bernhard Ganter[1] and Sergei O. Kuznetsov[2]

[1] Technische Universität Dresden, Institut für Algebra, D-01062 Dresden
[2] Department of Theoretical Foundations of Informatics, All-Russia Institute for Scientific and Technical Information (VINITI), ul. Usievicha 20 a, 125219 Moscow, Russia

Abstract. Lattices are mathematical structures which are frequently used for the representation of data. Several authors have considered the problem of incremental construction of lattices. We show that with a rather general approach, this problem becomes well-structured. We give simple algorithms with satisfactory complexity bounds.

For a subset $A \subseteq P$ of an ordered set $(P, \leq)$ let $A^\uparrow$ denote the set of all upper bounds. That is,

$$A^\uparrow := \{p \in P \mid a \leq p \text{ for all } a \in A\}.$$

The set $A^\downarrow$ of lower bound is defined dually. A **cut** of $(P, \leq)$ is a pair (A, B) with $A, B \subseteq P$, $A^\uparrow = B$, and $A = B^\downarrow$. It is well known that these cuts, ordered by

$$(A_1, B_1) \leq (A_2, B_2) :\iff A_1 \subseteq A_2 \quad (\iff B_2 \subseteq B_1)$$

form a complete lattice, the **Dedekind-MacNeille completion** (or short **completion**) of $(P, \leq)$. It is the smallest complete lattice containing a subset order-isomorphic with $(P, \leq)$.

The size of the completion may be exponential in $|P|$. The completion can be computed in steps: first complete a small part of $(P, \leq)$, then add another element, complete again, et cetera. Each such step increases the size of the completion only moderately and is moreover easy to perform. We shall demonstrate this by describing an elementary algorithm that, given a (finite) ordered set $(P, \leq)$ and its completion $(L, \leq)$, constructs the completion of any one-element extension of $(P, \leq)$ in

$$O(|L| \cdot |P| \cdot \omega(P))$$

steps, where $\omega(P)$ denotes the width of $(P, \leq)$. The special case that $(P, \leq)$ is itself a complete lattice and thus isomorphic to its completion, has been considered as the problem of **minimal insertion** of an element into a lattice, see e.g. Valtchev [4]. We obtain that the complexity of inserting an element into a lattice $(L, \leq)$ and then forming its completion is bounded by

$$O(|L|^2 \cdot \omega(L)).$$

The elementary considerations on the incidence matrix of $(P, \leq)$, which we use in the proof, do not utilize any of the order properties. Our result therefore generalizes to arbitrary incidence matrices. In the language of Formal Concept Analysis this may be interpreted as *inserting a preconcept into a concept lattice.*

1 Computing the completion

Let us define a **precut** of an ordered set to be a pair (S, T), where S is an order filter and T is an order ideal such that $S \subseteq T^{\downarrow}$, $T \subseteq S^{\uparrow}$. We consider the following construction problem:

INSTANCE: A finite ordered set $(P, \leq)$, its completion, and a precut (S, T) of $(P, \leq)$.
OUTPUT: The completion of $(P \cup \{x\}, \leq)$, where $x \notin P$ is some new element with $p \leq x \iff p \in S$ and $x \leq p \iff p \in T$ for all $p \in P$.[1]

$(P, \leq)$ may be given by its incidence matrix (of size $O(|P|^2)$). The completion may be represented as a list of cuts, that is, of pairs of subsets of P.

With a simple case analysis we show how the cuts of $(P \cup \{x\}, \leq)$ can be obtained from those of $(P, \leq)$.

Proposition 1. *Each cut of $(P \cup \{x\}, \leq)$, except $(S \cup \{x\}, T \cup \{x\})$, is of the form*

$$(C, D), \quad (C \cup \{x\}, D \cap T), \quad \text{or} \quad (C \cap S, D \cup \{x\})$$

for some cut (C, D) of $(P, \leq)$. If (C, D) is a cut of $(P, \leq)$ then

1. *$(C \cup \{x\}, D \cap T)$ is a cut of $(P \cup \{x\}, \leq)$ iff $S \subset C = (D \cap T)^{\downarrow}$,*
2. *$(C \cap S, D \cup \{x\})$ is a cut of $(P \cup \{x\}, \leq)$ iff $T \subset D = (C \cap S)^{\uparrow}$,*
3. *(C, D) is a cut of $(P \cup \{x\}, \leq)$ iff $C \not\subseteq S$ and $D \not\subseteq T$.*

For a **proof** of this result and of the following see the next section.

Proposition 2. *The number of cuts of $(P \cup \{x\}, \leq)$ does not exceed twice the number of cuts of $(P, \leq)$, plus two.*

A natural embedding of the completion of $(P, \leq)$ into that of $(P \cup \{x\}, \leq)$ is given by the next proposition:

Proposition 3. *For each cut (C, D) of $(P, \leq)$ exactly one of*

$$(C, D), \quad (C \cup \{x\}, D), \quad (C, D \cup \{x\}), \quad (C \cup \{x\}, D \cup \{x\})$$

is a cut of $(P \cup \{x\}, \leq)$.

These cuts can be considered to be the "old" cuts, up to a modification. "New" cuts are obtained only from cuts (C, D) that satisfy 3) and simultaneously 1) or 2). An algorithm can now be given:

[1] For elements of P different from x, the order remains as it was.

Algorithm to construct the completion of $(P \cup \{x\}, \leq)$. Let L denote the set of all cuts of $(P, \leq)$.

- Output $(S \cup \{x\}, T \cup \{x\})$.
- For each $(C, D) \in L$ do:
  1. If $C \subseteq S$ and $D \not\subseteq T$ then output $(C, D \cup \{x\})$.
  2. If $C \not\subseteq S$ and $D \subseteq T$ then output $(C \cup \{x\}, D)$.
  3. If $C \not\subseteq S$ and $D \not\subseteq T$ then
     (a) output (C, D),
     (b) if $C = (D \cap T)^{\downarrow}$ then output $(C \cup \{x\}, D \cap T)$,
     (c) if $D = (C \cap S)^{\uparrow}$ then output $(C \cap S, D \cup \{x\})$.
- End.

It follows from the above propositions that this algorithm outputs every cut of $(P \cup \{x\}, \leq)$ exactly once. Each step of the algorithm involves operations for subsets of P. The most time consuming one is the computation of $(D \cap T)^{\downarrow}$ and of $(C \cap S)^{\uparrow}$. Note that $(D \cap T)^{\downarrow} = (\min(D \cap T))^{\downarrow}$, where $\min(D \cap T)$ is the set of the minimal elements of $D \cap T$ and can be computed in $O(|P| \cdot \omega(P))$ steps. Since $|\min(D \cap T)| \leq \omega(P)$ and, moreover,

$$(\min(D \cap T))^{\downarrow} = \bigcap\{p^{\downarrow} \mid p \in \min(D \cap T)\},$$

we conclude that $(D \cap T)^{\downarrow}$ can be obtained with an effort of $O(|P| \cdot \omega(P))$. The dual argument for $(C \cap S)^{\uparrow}$ leads to the same result. So if L is the set of cuts of $(P, \leq)$, then the algorithm can be completed in $O(|L| \cdot |P| \cdot \omega(P))$ steps.

Let us mention that computing an incidence matrix of the completion can be done in $O(|L|^2)$ steps, once the completion has been computed, see Proposition 6.

2 Inserting a preconcept

A triple (G, M, I) is called a **formal context** if G and M are sets and $I \subseteq G \times M$ is a relation between G and M. For each subset $A \subseteq G$ let

$$A^I := \{m \in M \mid (g, m) \in I \text{ for all } g \in A\}.$$

Dually, we define for $B \subseteq M$

$$B^I := \{g \in G \mid (g, m) \in I \text{ for all } m \in B\}.$$

A **formal concept** of (G, M, I) is a pair (A, B) with $A \subseteq G$, $B \subseteq M$, $A^I = B$, and $A = B^I$. The formal concepts, ordered by

$$(A_1, B_1) \leq (A_2, B_2) :\iff A_1 \subseteq A_2 \quad (\iff B_2 \subseteq B_1),$$

form a complete lattice, the **concept lattice** of (G, M, I).

Most of the arguments given below become rather obvious if one visualizes a formal context as a $G \times M$ - cross table, where the crosses indicate the incidence relation I. The concepts (we sometimes omit the word "formal") then correspond to maximal rectangles in such a table. Note that if $A = B^I$ for some set $B \subseteq M$, then (A, A^I) automatically is a concept of (G, M, I).

A pair (A, B) with $A \subseteq G$, $B \subseteq M$, $A \subseteq B^I$, and $B \subseteq A^I$ is called a **preconcept** of (G, M, I). In order to change a preconcept into a concept, one may extend each of the sets G and M by one element with the appropriate incidences. So as a straightforward generalization of the above, we consider the following construction problem:

INSTANCE: A finite context (G, M, I), its concept lattice, and a preconcept (S, T) of (G, M, I).
OUTPUT: The concept lattice of $(G \cup \{x\}, M \cup \{x\}, I^+)$, where $x \notin G \cup M$ is a new element and

$$I^+ := I \cup ((S \cup \{x\}) \times (\{x\} \cup T)).$$

The special case of section 1 is obtained by letting

$$G = M := P \quad \text{and} \quad (g, m) \in I :\Longleftrightarrow g \leq m.$$

Proposition 4. *Each formal concept of $(G \cup \{x\}, M \cup \{x\}, I^+)$, with the exception of $(S \cup \{x\}, T \cup \{x\})$, is of the form*

$$(C, D), \quad (C \cup \{x\}, D \cap T), \quad \text{or} \quad (C \cap S, D \cup \{x\})$$

for some formal concept (C, D) of (G, M, I). With the obvious modifications, the conditions given in Proposition 1 hold.

Proof. Each formal concept (A, B) of $(G \cup \{x\}, M \cup \{x\}, I^+)$ belongs to one of the following cases:

1. $x \in A, x \in B$. Then $A = S \cup \{x\}$, $B = T \cup \{x\}$.
2. $x \in A, x \notin B$. Then $B \subseteq T$ and $B^I = A \setminus \{x\}$. Therefore $(C, D) := (A \setminus \{x\}, (A \setminus \{x\})^I)$ is a formal concept of (G, M, I) satisfying

$$S \subset C = (D \cap T)^I. \tag{1}$$

Conversely if (C, D) is a formal concept of (G, M, I) satisfying (1), then

$$(A, B) := (C \cup \{x\}, D \cap T)$$

is a formal concept of $(G \cup \{x\}, M \cup \{x\}, I^+)$.

3. $x \notin A, x \in B$, dual to 2. Then $(C, D) := ((B \setminus \{x\})^I, B \setminus \{x\})$ is a concept of $(G \cup \{x\}, M \cup \{x\}, I^+)$ with

$$T \subset D = (C \cap S)^I. \tag{2}$$

Conversely each formal concept (C, D) with (2) yields a formal concept

$$(A, B) := (C \cap S, D \cup \{x\})$$

of $(G \cup \{x\}, M \cup \{x\}, I^+)$.

4. $x \notin A, x \notin B$. Then $(C, D) := (A, B)$ is a formal concept also of (G, M, I), satisfying
$$C \not\subseteq S, \quad D \not\subseteq T. \tag{3}$$
Conversely is each pair with (3) also a concept of $(G \cup \{x\}, M \cup \{x\}, I^+)$.

If both $(C \cup \{x\}, D \cap T)$ and $(C \cap S, D \cup \{x\})$ happen to be concepts, then $S \subseteq C$ and $T \subseteq D$, which implies $C \cup \{x\} = T^I$, $D \cup \{x\} = S^I$. Thus apart from perhaps one exceptional case these two possibilities exclude each other. From each concept of (G, M, I), we therefore obtain at most two concepts of $(G \cup \{x\}, M \cup \{x\}, I^+)$, except in a single exceptional case, which may lead to three solutions. On the other hand, each concept of $(G \cup \{x\}, M \cup \{x\}, I^+)$, except $(S \cup \{x\}, T \cup \{x\})$, is obtained in this manner. This proves Proposition 2.

To see that Proposition 3 holds in the general case, note that each formal concept (C, D) of (G, M, I) belongs to one of the following cases:

1. $C = S$, $D = T$. Then $(C \cup \{x\}, D \cup \{x\})$ is a concept of $(G \cup \{x\}, M \cup \{x\}, I^+)$.
2. $C \subseteq S$, $T \subset D$. Then $D = C^I$ and condition (2) (from the proof of Proposition 4) is fulfilled. Thus $(C, D \cup \{x\})$ is a concept of $(G \cup \{x\}, M \cup \{x\}, I^+)$.
3. $S \subset C$, $D \subseteq T$. Then $C = D^I$ and condition (1) is satisfied. Therefore $(C \cup \{x\}, D)$ is a concept of $(G \cup \{x\}, M \cup \{x\}, I^+)$.
4. $C \not\subseteq S$, $D \not\subseteq T$. Then (C, D) is a concept of $(G \cup \{x\}, M \cup \{x\}, I^+)$.

It is clear that each of the possible outcomes determines (C, D), and that therefore the possibilities are mutually exclusive.

It is a routine matter to check that these formal concepts are ordered in the same way than those of (G, M, I). The construction thus yields a canonical order embedding of the small concept lattice into that of the enlarged context.

Since all details have carried over to the more general case, we may resume:

Proposition 5. *The algorithm given in section 1, when applied to the concept lattice L of (G, M, I), computes the concept lattice of $(G \cup \{x\}, M \cup \{x\}, I^+)$.*

The abovementioned complexity considerations apply as well, but it is helpful to introduce a parameter for contexts that corresponds to the width. The incidence relation induces a quasiorder relation on G by
$$g_1 \leq g_2 :\iff \{g_2\}^I \subseteq \{g_2\}^I.$$
Let $\omega(G)$ be the width of this quasiorder, and let $\omega(M)$ denote the width of the corresponding quasiorder on M. Let
$$\tau(G, M, I) := (\omega(G) + \omega(M)) \cdot (|G| + |M|).$$
Of course, $\tau(G, M, I) \leq (|G| + |M|)^2$. Provided the induced quasiorders on G and M are given as incidence matrices (these can be obtained in $O(|G| \cdot |M| \cdot (|G| + |M|))$ steps), we have a better bound on the complexity of the derivation operators: the set A^I can be computed from A with complexity $O(\tau(G, M, I))$.

Computing A^I was the most time consuming step in the algorithm on section 1. Thus computing the new concept lattice can be performed with

$$O(|L| \cdot \tau(G, M, I))$$

bit operations.

Each concept of $(G \cup \{x\}, M \cup \{x\}, I^+)$, except $(S \cup \{x\}, T \cup \{x\})$, is generated by exactly one of the steps 1, 2, 3a, 3b, 3c of the algorithm, and precisely 3b) and 3c) lead to "new" concepts (other than $(S \cup \{x\}, T \cup \{x\})$. When performing the algorithm, we may note down how the concepts were obtained. These data can be used later to construct an incidence matrix of the new lattice:

Proposition 6. *The order relation of the new lattice can be computed in additional $O(|L|^2)$ steps.*

Proof. $(S \cup \{x\}, T \cup \{x\})$ is the largest concept containing x in its extent and the smallest concepts containing x in its intent. In other words, $(S \cup \{x\}, T \cup \{x\})$ is greater than all concepts generated in steps 2) and 3b) and smaller than all concepts generated by steps 1) and 3c). It is incomparable to the other elements. So we may exclude this concept from further considerations.

The order relation between the "old" concepts, i.e. between those generated in steps 1), 2), and 3a), is the same as before.

For the remaining case, we consider w.l.o.g. a concept $(C \cup \{x\}, D \cap T)$, which was generated in step 3b) from a concept (C, D) of (G, M, I). Now $(C \cup \{x\}, D \cap T) \leq (E, F)$ if and only if (E, F) has been generated in steps 2) or 3b) from some concept $(E \setminus \{x\}, (E \setminus \{x\})^I) \geq (C, D)$ of (G, M, I). If $x \in E$, then similarly $(E, F) \leq (C \cup \{x\}, D \cap T)$ is true if and only if (E, F) has been generated in steps 2) or 3b) from some concept $(E \setminus \{x\}, (E \setminus \{x\})^I) \leq (C, D)$ of (G, M, I).

Suppose $x \notin E$. If (E, F) was obtained in steps 1) or 3a) of the algorithm, than (E, E^I) is a concept of (G, M, I) and $(E, F) \leq (C \cup \{x\}, D \cap T)$ is equivalent to $(E, E^I) \leq (C, D)$. If (E, F) was obtained in step 3c), then $S^I \subseteq F$, which implies $D \cap T \subseteq S^I \subseteq F$. So in this case $(E, F) \leq (C \cup \{x\}, D \cap T)$ always holds.

Summarizing these facts, we obtain all comparabilities of a concept $(C \cup \{x\}, D \cap T)$ of $(G \cup \{x\}, M \cup \{x\}, I^+)$ which was derived from a concept (C, D) of (G, M, I) in step 3b): Concepts greater than $(C \cup \{x\}, D \cap T)$ are those obtained in steps 2 or 3b) from concepts greater than (C, D), concepts smaller than $(C \cup \{x\}, D \cap T)$ are those obtained in steps 1), 2), 3a) or 3b) from those smaller than (C, D) and all those obtained in step 3c). Thus the comparabilities of $(C \cup \{x\}, D \cap T)$ can be obtained from those of (C, D) using only a bounded number of elementary operations in each case. Filling the corresponding row of the incidence matrix is of complexity $O(|L|)$. The argument for concepts obtained by 3c) is analogous.

The generalized algorithm may be applied to the context $(P, P, \not\geq)$, obtained from an arbitrary ordered set $(P, \leq)$. The concept lattice is the lattice of maximal antichains of $(P, \leq)$ (see Wille [5]). Our result therefore relates to that of Jard, Jourdan and Rampon [2].

3 A non-incremental procedure may be more convenient

In practice, a strategy suggests itself that may be more time-consuming, but is nevertheless simpler than the algorithm presented in section 1. Rather than pursuing an incremental algorithm, it may be easier to compute the lattice "from scratch" (i.e. from the formal context, or, in the special case, from the ordered set $(P, \leq)$) each time. For this task there is an algorithm that is remarkably simple (it can be programmed in a few lines) and at the same time is not dramatically slower than the incremental approach: it computes the concept lattice L of a formal context (G, M, I) in $O(|L| \cdot |G|^2 \cdot |M|)$ steps. Using the parameter introduced above, we can improve this to $O(|L| \cdot |G| \cdot \tau(G, M, I))$. This algorithm generates the formal concepts inductively and does not require a list of concepts to be stored.

Let us exemplify the advantage of this by a simple calculation: A formal context (G, M, I) with $|G| = |M| = 50$ may have as may as 2^{50} formal concepts in the extreme. But even if the lattice is "small" and has only, say, 10^{10} elements, it would require almost a hundred Gigabytes of storage space. Generating such a lattice with the inductive algorithm appears to be time-consuming, but not out of reach; the storage space required would be less than one Kilobyte. Moreover, this algorithm admits modifications that allow to search specific parts of the lattice.

For details and proofs we refer to the literature (see [1]), but the algorithm itself is so simple that it can be recalled here. For simplicity assume $G := \{1, \ldots, n\}$, and define for subsets $A, B \subseteq G$

$$A <_i B :\iff i \in B \text{ is minimal in } (A \setminus B) \cup (B \setminus A).$$

Then the definition

$$A < B :\iff A <_i B \text{ for some } i$$

yields a strict linear order on the set of all subsets of G (a lexicographic or, for short, **lectic** order).

If (A, B) is a formal concept of (G, M, I) then A is called its **extent**. Since $B = A^I$, the extents uniquely determine the concepts. To generate all concepts, it therefore suffices to generate these. This can be done in lectic order, starting with $\emptyset^{II}$. The step that constructs from a given set A the "next" extent is of the form

$$A \oplus i := ((A \cap \{1, \ldots, i-1\}) \cup \{i\})^{II}.$$

The following theorem describes how the element i must be chosen:

Theorem 1 (see [1]). *Let (G, M, I) be a formal context with $G := \{1, \ldots, n\}$. For given $A \subset G$, the smallest extent that is larger than A (with respect to the lectic order) is given by*

$$A^+ := A \oplus i,$$

where i is maximal with respect to $A <_i A \oplus i$.

It is easy to see that computing $A \oplus i$ requires at most $O(|G| \cdot |M|)$ steps, using the induced quasiorders only $O(\tau(G, M, I))$ steps. The "next" extent therefore is found at an expense of $O(|G|^2 \cdot |M|)$, or even $O(|G| \cdot \tau(G, M, I))$.

If a lattice diagram is to be generated, the inductive approach may even be faster than the incremental one. For a given extent $A \neq G$, the extents of the upper covers are precisely the minimal sets of the form

$$(A \cup \{i\})^{II}, \quad i \notin A.$$

Computing these requires $O(|G|^2 \cdot |M|)$ steps. Localizing such an upper cover in a linear list of extents, using a binary search algorithm, can be done with $O(\log |L|)$ comparisons of subsets of G. The complexity thus is $O(|G| \cdot |M|)$, since $|L| \leq 2^{|M|}$.

Every finite lattice $(L, \leq)$ is isomorphic to some concept lattice. A natural choice is the formal context $(J(L), M(L), \leq)$, where $J(L)$ and $M(L)$ denote the sets of join- and meet-irreducible elements of $(L, \leq)$, respectively. If we denote the cardinalities of these sets by

$$j(L) := |J(L)|, \qquad m(L) := |M(L)|,$$

we can resume:

Corollary 1. *The covering relation of a finite lattice $(L, \leq)$ can be computed in $O(j(L)^2 \cdot m(L) \cdot |L|)$ steps, provided the sets $J(L)$ and $M(L)$ of join- and of meet-irreducible elements are given.*

This is considerably better e.g. than the bound given by Skorsky[3]. Again, the bound can be refined using the width of the induced orders on G and on M.

References

1. Bernhard Ganter, Rudolf Wille: *Formale Begriffsanalyse - Mathematische Grundlagen.* Springer Verlag 1996.
2. C. Jard, G.-V. Jourdan and J.-X. Rampon: *Computing On-Line the Lattice of Maximal Antichains of Posets.* Order 11 (1994).
3. Martin Skorsky: *Endliche Verbände – Diagramme und Eigenschaften.* Shaker, 1992.
4. Petko Valtchev: *An Algorithm for Minimal Insertion in a Type Lattice.* Second International KRUSE Symposium, Vancouver, 1997.
5. Rudolf Wille: *Finite distributive lattices as concept lattices.* Atti Inc. Logica Mathematica, 2 (1985).