



HAL
open science

Difference hierarchies and duality with an application to formal languages

Célia Borlido, Mai Gehrke, Andreas Krebs, Howard Straubing

► **To cite this version:**

Célia Borlido, Mai Gehrke, Andreas Krebs, Howard Straubing. Difference hierarchies and duality with an application to formal languages. *Topology and its Applications*, 2019. hal-02413264

HAL Id: hal-02413264

<https://hal.science/hal-02413264>

Submitted on 16 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Difference hierarchies and duality with an application to formal languages[☆]

Célia Borlido^{a,*}, Mai Gehrke^a, Andreas Krebs^b, Howard Straubing^c

^a*Laboratoire J. A. Dieudonné, CNRS, Université Côte d'Azur, France*

^b*Wilhelm-Schickard-Institut Universität Tübingen, Germany*

^c*Boston College, United States*

Abstract

The notion of a difference hierarchy, first introduced by Hausdorff, plays an important role in many areas of mathematics, logic and theoretical computer science such as descriptive set theory, complexity theory, and the theory of regular languages and automata. Lattice theoretically, the difference hierarchy over a distributive lattice stratifies the Boolean algebra generated by it according to the minimum length of difference chains required to describe the Boolean elements. While each Boolean element is given by a finite difference chain, there is no canonical such writing in general. We show that, relative to the filter completion, or equivalently, the lattice of closed upsets of the dual Priestley space, each Boolean element over the lattice has a canonical minimum length decomposition into a Hausdorff difference chain. As a corollary, each Boolean element over a co-Heyting algebra has a canonical difference chain (and an order dual result holds for Heyting algebras). With a further generalization of this result involving a directed family of closure operators on a Boolean algebra, we give an elementary proof of the fact that if a regular language is given by a Boolean combination of universal sentences using arbitrary numerical predicates then it is also given by a Boolean combination of universal sentences using only regular numerical predicates.

[☆]This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No. 670624). For further partial funding, see acknowledgements.

*Corresponding author

Email addresses: cborlido@unice.fr (Célia Borlido), mgehrke@unice.fr (Mai Gehrke), mail@krebs-net.de (Andreas Krebs), howard.straubing@bc.edu (Howard Straubing)

Keywords: Difference hierarchies, Stone-Priestley duality, Logic on Words
2000 MSC: 06D50, 68F05

In honour of Aleš Pultr on the occasion of his 80th birthday

1. Introduction

Hausdorff introduced the notion of a difference hierarchy in his work on set theory [17]. Subsequently, the notion has played an important role in descriptive set theory as well as in complexity theory. More recently, it has seen a number of applications in the theory of regular languages and automata [15, 4]. From a lattice theoretic point of view, the difference hierarchy over a distributive lattice D stratifies the universal Boolean envelope of D . This is the (unique up to isomorphism) Boolean algebra B containing D as a sublattice and generated (as a Boolean algebra) by D . We follow the tradition in lattice theory calling B the Booleanization of D . However, we warn any frame theorist reading the paper that this is not the same thing as the Booleanization of a frame [1], which generalizes the construction of the complete Boolean algebra of regular opens of a topological space.

In the difference hierarchy over D , the stratification of the Booleanization B of D is made according to the minimum length of difference chains required to describe an element $b \in B$:

$$b = a_1 - (a_2 - (\dots (a_{n-1} - a_n)\dots)) \quad (1)$$

where $a_1 \geq a_2 \geq \dots \geq a_{n-1} \geq a_n$ are elements of D . One difficulty in the study of difference hierarchies is that in general elements $b \in B$ do not have canonical associated difference chains.

Priestley duality [25] for distributive lattices uses the Stone space of the Booleanization equipped with a partial order to represent the lattice as the closed and open (henceforth called clopen) upsets of the associated Priestley space. Priestley duality provides an elucidating tool for the study of difference hierarchies. For one, the minimum length of difference chains for an element $b \in B$ has a nice description relative to the Priestley dual space X of D as the length of the longest chain of points $x_1 < x_2 < \dots < x_n$ in X so that x_i belongs to the clopen corresponding to b if and only if i is odd. Further, if we allow difference chains of closed upsets of the Priestley

29 space, rather than clopen upsets, then every element $b \in B$ has a canonical
30 difference chain which is of minimum length. If in addition D is a co-Heyting
31 algebra, then the elements of this canonical difference chain are clopen and
32 thus every $b \in B$ has a canonical difference chain in D . We present this ma-
33 terial, which is closely related to work by Leo Esakia on skeletal subalgebras
34 of closure algebras [9], in Section 3. We have chosen to present this first part
35 in the language of point-set topology to provide a treatment which is more
36 easily accessible for researchers not used to point-free topology.

37 The main purpose of Section 5 is to obtain sufficient conditions so that,
38 given a Boolean algebra B , a Boolean subalgebra B' , and a sublattice D of B ,
39 the intersection of B' and D generates the intersection of B' with the Boolean
40 algebra generated by D . For this purpose we consider a situation where B
41 is equipped with a family of closure operators whose meet-subsemilattices of
42 closed elements form a directed family the union of which is a sublattice of B .
43 Using Stone-Priestley duality in the point-free form of canonical extensions,
44 we generalize the results of Section 3 and use them to derive the main theorem
45 of the section.

46 In turn, the results of Section 5 are used in Section 8 for an application in
47 logic on words. More precisely, we give an elementary proof of the equality

$$\mathcal{B}\Pi_1[\mathcal{N}] \cap \text{Reg} = \mathcal{B}\Pi_1[\text{Reg}]. \quad (2)$$

48 The idea is that this equality may be translated into an instance of the main
49 theorem of Section 5 where $B = \mathcal{P}(A^*)$ is the Boolean algebra of all formal
50 languages over the alphabet A , $B' = \text{Reg}$ is the Boolean subalgebra of regular
51 languages, and $D = \Pi_1[\mathcal{N}]$ is the sublattice of languages given by universal
52 sentences using arbitrary numerical predicates. Universally quantifying over
53 a finite number of variables yields a closure operator, and the directed union
54 of the closed elements is exactly $D = \Pi_1[\mathcal{N}]$. The equation (2) was first
55 proved by Macial, Péladeau and Thérien [20]. For more details, see [29].

56 Before each of the main Sections 3, 5, and 8, we include the background
57 needed. The aim is in this way to make the paper accessible both to lattice
58 and frame theorists and to researchers working with formal languages and
59 automata. Thus, in Section 2 we introduce the basics on lattices and duality,
60 Section 4 is an introduction to canonical extensions, and Section 6 contains
61 the preliminaries on recognition and logic on words. Section 7 provides a
62 lattice theoretical perspective on logic on words, as well as its connections
63 with model theory, which is the point of view adopted in Section 8. Although,

64 strictly speaking, the results of this section can not be considered as new, to
 65 the best our knowledge, the presentation of the material in this form is an
 66 original contribution of the paper. Finally, in Section 9 we discuss possible
 67 generalizations of the present work, and in particular, we give some examples
 68 of open problems that one could try to handle with the techniques developed
 69 here.

70 2. Preliminaries on lattices and duality

71 We consider all distributive lattices to be bounded and we view the classes
 72 of distributive lattices and of Boolean algebras as categories in which the
 73 morphisms are the algebraic homomorphisms, that is, the maps that preserve
 74 all the basic operations (including the bounds). For readers needing more
 75 detailed preliminaries on lattices and duality, we refer to [7].

76 *Priestley duality.* The *Priestley dual space* of a distributive lattice D con-
 77 sists of the set $S(D)$ of homomorphisms from D to the two-element lattice $\mathbf{2}$
 78 (or equivalently, of the *prime filters* of D) ordered point-wise and equipped
 79 with the topology generated by the sets $\hat{a} = \{x \in S(D) \mid x(a) = 1\}$ and their
 80 complements, for $a \in D$. One can show that the resulting ordered topological
 81 space, (X, \leq, π) , is compact and *totally order disconnected*. That is, if $x \not\leq y$
 82 in X then there is a clopen upset V of X with $x \in V$ and $y \notin V$. Totally
 83 order disconnected compact spaces are called *Priestley spaces* and the ap-
 84 propriate structure preserving maps are the continuous and order preserving
 85 maps. In the other direction, given a Priestley space (X, \leq, π) the collection
 86 $\mathbf{UpClopen}(X, \leq, \pi)$ of subsets of X that are clopen upsets forms a lattice of
 87 sets (this may also be seen as the Priestley morphisms into the Priestley space
 88 based on the two-element chain). Morphisms correspond contravariantly and
 89 the correspondence is given by pre-composition. These functors account for
 90 the dual equivalence of the category of distributive lattices and the category
 91 of Priestley spaces. On objects, this means that $D \cong \mathbf{UpClopen}(S(D))$ (via
 92 the map $a \mapsto \hat{a}$) for any distributive lattice D and $X \cong S(\mathbf{UpClopen}(X))$ (via
 93 the map $x \mapsto \chi_x$ where χ_x is the characteristic function of the point x re-
 94 stricted to $\mathbf{UpClopen}(X)$) for any Priestley space X . In addition, the double
 95 dual of a morphism, on either side of the duality, is naturally isomorphic to
 96 the original. For more details see [7, Chapter 11].

97 *Booleanization.* The *Booleanization*, D^- , of a distributive lattice D is a
 98 Boolean algebra with a lattice embedding $D \hookrightarrow D^-$ so that any lattice

99 homomorphism $h: D \rightarrow B$ into a Boolean algebra, uniquely extends to a
 100 homomorphism $h^-: D^- \rightarrow B$ making the following diagram commutative:

$$\begin{array}{ccc}
 D & \xrightarrow{\quad} & D^- \\
 & \searrow h & \vdots h^- \\
 & & B
 \end{array}$$

101

102 It is well-known that for *any* embedding of D into *any* Boolean algebra, the
 103 Boolean algebra generated by the image is isomorphic to D^- . The existence
 104 of such a map is a consequence of Stone-Priestley duality, but showing that
 105 this map is an embedding requires a non-constructive principle. For more on
 106 Booleanizations of lattices in a constructive manner, please see [21, 23, 16, 6].

107 In the setting of Priestley duality, we have seen that D is isomorphic to
 108 the lattice of clopen upsets of its dual space X . It thus follows that D^- is
 109 isomorphic to the Boolean subalgebra of $\mathcal{P}(X)$ generated by $\text{UpClopen}(X)$.
 110 One can show that this is the Boolean algebra of all clopen subsets of X .
 111 That is, $D^- \cong \text{Clopen}(X)$.

112 *Adjunctions and closure operators.* Let P and Q be posets. We say that
 113 maps $f: P \rightleftarrows Q : g$ form an *adjoint pair* provided

$$\forall p \in P, q \in Q \quad (f(p) \leq q \iff p \leq g(q)).$$

114 Note that in this case, f and g uniquely determine each other since

$$f(p) = \bigwedge \{q \in Q \mid p \leq g(q)\} \quad \text{and} \quad g(q) = \bigvee \{p \in P \mid f(p) \leq q\}.$$

115 We call f the *lower adjoint* of g and g the *upper adjoint* of f . One can
 116 show that lower adjoints preserve all existing suprema, while upper adjoints
 117 preserve all existing infima. In the case that the posets P and Q are complete
 118 lattices this gives a simple criterion for the existence of adjoints.

119 **Proposition 1.** *A map between complete lattices has a lower adjoint if and*
 120 *only if it preserves arbitrary meets, and it has an upper adjoint if and only*
 121 *if it preserves arbitrary joins.*

122 Adjoint pairs are intimately related to *closure operators*. Recall that, for
 123 a poset P , a function $c: P \rightarrow P$ is a *closure operator* provided

$$\forall p_1, p_2 \in P \quad (p_1 \leq c(p_2) \iff c(p_1) \leq c(p_2)).$$

124 **Proposition 2.** *If $(f: P \rightleftarrows Q : g)$ is an adjoint pair, then gf is a closure*
 125 *operator on P . Conversely, every closure operator $c: P \rightarrow P$ may be obtained*
 126 *in this way, e.g. $(c': P \rightleftarrows \text{Im}(c) : \iota)$, where c' is the co-restriction of c to its*
 127 *image and ι the inclusion map, is such an adjoint pair.*

128 Finally, since upper adjoints preserve all existing infima, one can show
 129 that if $c: P \rightarrow P$ is a closure operator and P admits a meet-semilattice
 130 structure, then $\text{Im}(c)$ is a meet-subsemilattice of P . For more details, see [7,
 131 Chapter 7].

132 *Heyting and co-Heyting algebras.* Heyting algebras are the algebras for in-
 133 tuitionistic propositional logic in the same sense that Boolean algebras are
 134 the algebras for classical propositional logic. The order dual notion is called
 135 a co-Heyting algebra. We will focus on co-Heyting algebras here as this is
 136 more convenient for the sequel, but any result about one notion has a corre-
 137 sponding order dual result about the other notion.

138 **Definition 3.** A *co-Heyting algebra* is a distributive lattice equipped with an
 139 additional binary operation, $/$, which is the lower adjoint of the operation \vee
 140 in the sense that we have

$$\forall a, b, c \in D \quad (a/b \leq c \iff a \leq b \vee c).$$

141 Notice that this property implies that the co-Heyting operation on a dis-
 142 tributive lattice, if it exists, is unique and is given by

$$a/b = \bigwedge \{c \mid a \leq b \vee c\}.$$

143 Thus one may think of a co-Heyting algebra as a special kind of distributive
 144 lattice. Indeed, the following proposition, which is the order dual of the
 145 corresponding fact for Heyting algebras, see [10, Proposition 3], gives such a
 146 characterization.

147 **Proposition 4.** *A distributive lattice D is a co-Heyting algebra if and only*
 148 *if the inclusion of D in its Booleanization has a lower adjoint.*

149 **Definition 5.** The lower adjoint mentioned in Proposition 4 gives, for each
 150 $b \in D^-$, the least over-approximation of b in D . We denote it by

$$D^- \longrightarrow D, \quad b \mapsto [b] = \bigwedge \{c \in D \mid b \leq c\}.$$

151 and call it the *ceiling function* (of D).

152 It will be useful to understand which Priestley spaces correspond to co-
 153 Heyting algebras. The order dual characterization for Heyting algebras is
 154 due to Esakia [8] independently of Priestley's work. We include a proof of
 155 the result for co-Heyting algebras to illustrate the correspondence between
 156 the algebraic and topological formulations.

157 **Theorem 6** ([8]). *Let D be a distributive lattice and X its Priestley dual.*
 158 *Then D admits a co-Heyting structure if and only if for each $V \subseteq X$ clopen,*
 159 *$\uparrow V$ is again clopen. When this is the case, the map $\lceil \rceil : D^- \rightarrow D$ is naturally*
 160 *isomorphic to the map $V \mapsto \uparrow V$ on clopen subsets of X .*

161 *Proof.* This is a simple consequence of Proposition 4. Note that by total
 162 order disconnectedness of X , for any closed (and thus compact) $K \subseteq X$, we
 163 have

$$\uparrow K = \bigcap \{W \subseteq X \mid K \subseteq W \text{ and } W \text{ clopen upset}\}. \quad (3)$$

164 Therefore, there is a least clopen upset (i.e., element of D) above V if and
 165 only if $\uparrow V$ is clopen. \square

166 *Closed upsets in Priestley spaces.* Note that closed subspaces of Priestley
 167 spaces are again Priestley spaces. In fact, the closed subspaces of a Priestley
 168 space X correspond to the lattice quotients of its dual D , cf. [7, Section 11.32].

169 The following well-known fact about upsets of closed sets will be used
 170 extensively in the sequel. We include a proof for the sake of completeness.
 171 For a subset $S \subseteq P$ of a poset P , we use $\min(S)$ to denote the set of minimal
 172 elements of S .

173 **Proposition 7.** *Let X be a Priestley space and $K \subseteq X$ a closed subset.*
 174 *Then, $\uparrow K = \uparrow \min(K)$ and this is a closed subset of X .*

175 *Proof.* As seen in (3), $\uparrow K$ is a closed subset of X whenever $K \subseteq X$ is. Now
 176 consider X as the dual space of a distributive lattice D . Then the points
 177 of X are the prime filters of D . Let x be any element of K and let C be a
 178 maximal chain of prime filters contained in K with $x \in C$. Since C is a chain,
 179 it is easy to show that $x_0 = \bigcap_{x \in C} x$ is again a prime filter. Also, if $W = \hat{a}$
 180 is any clopen upset of X with $K \subseteq W$, then $a \in y$ for all $y \in K$ and thus
 181 $a \in x_0$. It follows that $x_0 \in W$ for all clopen upsets W of X with $K \subseteq W$
 182 and thus $x_0 \in \uparrow K$. Now by maximality of C it follows that $x_0 \in \min(K)$ and
 183 $x_0 \leq x$. Thus $\uparrow K = \uparrow \min(K)$. \square

184 **3. The difference hierarchy and closed upsets**

185 Let D be a distributive lattice and D^- its Booleanization. Since D^- is
 186 generated by D as a Boolean algebra and because of the disjunctive normal
 187 form of Boolean expressions, every element of D^- may be written as a finite
 188 join of elements of the form $a - b$ with $a, b \in D$. A fact, that is well known
 189 but somewhat harder to see is that every element of D^- is of the form

$$a_1 - (a_2 - (\dots - (a_{n-1} - a_n)\dots)), \quad (4)$$

190 for some $a_1, \dots, a_n \in D$. The usual proof is by algebraic computation and
 191 is not particularly enlightening. It is also a consequence of our results here.
 192 We begin with a technical observation.

193 **Proposition 8.** *Let B be a Boolean algebra and let $a_1 \geq \dots \geq a_{2m}$ be a*
 194 *decreasing sequence of elements of B . Then, the following equality holds:*

$$a_1 - (a_2 - (\dots - (a_{2m-1} - a_{2m})\dots)) = (a_1 - a_2) \vee (a_3 - (\dots - (a_{2m-1} - a_{2m})\dots)) \quad (5)$$

195 *where the join is disjoint, and by induction we obtain*
 196

$$a_1 - (a_2 - (\dots - (a_{2m-1} - a_{2m})\dots)) = \bigvee_{n=1}^m (a_{2n-1} - a_{2n})$$

197 *where the joinands are pairwise disjoint.*

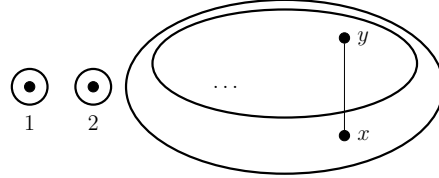
198 *Proof.* Let us denote $b = a_3 - (a_4 - (\dots - (a_{2m-1} - a_{2m})\dots))$. A simple algebraic
 199 computation yields $a_1 - (a_2 - b) = (a_1 - a_2) \vee (a_1 \wedge b)$. Since a_1, \dots, a_{2m}
 200 is a decreasing chain and $b \leq a_3$, it follows that $a_1 \wedge b = b$, which in turn
 201 yields (5). \square

202 One problem with difference chain decompositions of Boolean elements
 203 over a distributive lattice, which makes them difficult to understand and
 204 work with, is that, in general, there is no ‘most efficient’ such decomposition.
 205 We give an example of a Boolean element over a distributive lattice that
 206 illustrates this problem.

207 **Example 9.** Consider $X = \mathbb{N} \cup \{x, y\}$ equipped with the topology of the
 208 one-point compactification of the discrete topology on $\mathbb{N} \cup \{x\}$. That is, the
 209 frame of opens of X is:

$$\pi = \mathcal{P}(\mathbb{N} \cup \{x\}) \cup \{C \cup \{y\} \mid C \subseteq \mathbb{N} \cup \{x\} \text{ is cofinite}\}.$$

210 The order relation on X is as depicted. That is, the only non-trivial order
 211 relation in X is $x \leq y$. It is not hard to verify that X is a Priestley space.



212

213 The clopen upsets of X are the finite subsets of \mathbb{N} and the cofinite subsets
 214 of X containing y and they form the lattice D dual to X . Note that $V = \{x\}$
 215 is clopen in X and thus $V \in D^-$. On the other hand, any clopen upset W
 216 of X containing V must be cofinite. We can write

$$V = W - W'$$

217 where $W' = W - \{x\}$ is also a clopen upset of X . There is no smallest choice
 218 for W as $\uparrow V = \{x, y\}$ is not open and thus not in D .

219 However, if we look for difference chains for V relative to *the lattice of*
 220 *closed subsets of X* , then we have a least choice of difference chain, namely
 221 $V = K_1 - K_2$ where $K_1 = \uparrow V$ and $K_2 = K_1 - V$.

222 We show that there is an algorithm for deriving, for each element of the
 223 Booleanization of a distributive lattice, a difference chain of *closed upsets* (cf.
 224 Theorem 14), and that, in the case of a co-Heyting algebra, this provides
 225 a difference chain of the form (4) for each element of its Booleanization
 226 (cf. Corollary 17). We show that the difference chain thus obtained is of
 227 minimum length and is element-wise contained in any other such sequence
 228 (cf. Proposition 15). For this reason we will call it the *canonical difference*
 229 *chain of closed upsets* for the Boolean element in question. Recall that a
 230 subset $S \subseteq P$ of a poset is said to be *convex* provided $x \leq y \leq z$ with
 231 $x, z \in S$ implies $y \in S$.

232 **Definition 10.** If P is a poset, $S \subseteq P$, and $p \in P$, then we say that $p_1 <$
 233 $p_2 < \dots < p_n$ in P is an *alternating sequence of length n for p* (with respect
 234 to S) provided

235 (a) $p_i \in S$ for each $i \in \{1, \dots, n\}$ which is odd;

236 (b) $p_i \notin S$ for each $i \in \{1, \dots, n\}$ which is even;

237 (c) $p_n = p$.

238 Further, we say that $p \in P$ has *degree* n (with respect to S), written
239 $\deg_S(p) = n$, provided n is the largest natural number k for which there
240 is an alternating sequence of length k for p . In particular, if there is no alter-
241 nating sequence for p with respect to S (i.e. if $p \in P - \uparrow S$) then $\deg_S(p) = 0$.

242 Notice that an element of finite degree is of odd degree if and only if it belongs
243 to S . Also, if S is convex, then every element of S has degree 1, while every
244 element of $\uparrow S - S$ has degree 2. In general, there will be non-empty subsets
245 of posets with respect to which no element has finite degree. However, that
246 is not the case for clopen subsets of Priestley spaces.

247 **Proposition 11.** *Let X be a Priestley space and V a clopen subset of X .
248 Then every element of X has finite degree with respect to V .*

249 *Proof.* The elements of the Booleanization of a distributive lattice D may be
250 written as a finite disjunctions of differences of elements from D . Thus, if V
251 is a clopen subset of a Priestley space X , then there is an m so that we may
252 write

$$V = \bigcup_{i=1}^m (U_i - W_i),$$

253 where $U_i, W_i \subseteq X$ are clopen upsets of X . In particular, since each $U_i - W_i$
254 is convex and, by the Pigeonhole Principle, there is no alternating sequence
255 with respect to V of length strictly greater than $2m$. \square

256 **Lemma 12.** *Let X be a Priestley space and V a clopen subset of X . Let*

$$K_1 = \uparrow V, \quad K_2 = \uparrow(\uparrow V - V).$$

257 *Then, for each $i \in \{1, 2\}$, K_i is closed and*

$$K_i = \{x \in X \mid \deg_V(x) \geq i\} = \uparrow\{x \in X \mid \deg_V(x) = i\}.$$

258 *Proof.* By Proposition 11, every element of X has a finite degree. Also if $x \leq$
259 y , then it is clear that $\deg_V(x) \leq \deg_V(y)$. Furthermore, by Proposition 7,
260 we have that $\uparrow K = \uparrow \min(K)$ for any closed set K . Now since both V and

261 $\uparrow V - V$ are closed, it suffices to show that the elements of $\min(V)$ have
 262 degree 1, and the elements of $\min(\uparrow V - V)$ have degree 2. It is clear that
 263 $\deg_V(x) = 1$ for any $x \in \min(V)$. Now suppose $x \in \min(\uparrow V - V)$. Since
 264 $x \in \uparrow V$, there is $x' \in V$ with $x' \leq x$. Since $x \notin V$, this is an alternating
 265 sequence of length 2 for x . On the other hand, if $x_1 < x_2 < \dots < x_n = x$ is
 266 an alternating sequence for x , then $x_2 \in \uparrow V - V$ and thus $x \notin \min(\uparrow V - V)$
 267 unless $n = 2$ and $x_2 = x$. Thus $\deg_V(x) = 2$ for any $x \in \min(\uparrow V - V)$. \square

268 **Corollary 13.** *Let X be a Priestley space, V a clopen subset of X , and*
 269 *$G_1 \supseteq G_2 \supseteq \dots \supseteq G_{2p}$ a sequence of closed upsets in X satisfying*

$$V = G_1 - (G_2 - (\dots - (G_{2p-1} - G_{2p})\dots)). \quad (6)$$

270 *If K_1 and K_2 are as defined in Lemma 12, then*

$$K_1 \subseteq G_1, \quad K_2 \subseteq G_2, \quad \text{and} \quad G_1 - G_2 \subseteq K_1 - K_2.$$

271 *Proof.* By (6), we have $V \subseteq G_1$. Also, since G_1 is an upset we have $K_1 =$
 272 $\uparrow V \subseteq G_1$. Now, since $G_1 - G_2 \subseteq V$ we have $G_1 - V \subseteq G_2$ and as G_2
 273 is an upset, it follows that $\uparrow(G_1 - V) \subseteq G_2$. Also, $K_1 \subseteq G_1$ implies $K_2 =$
 274 $\uparrow(K_1 - V) \subseteq \uparrow(G_1 - V)$ and thus, $K_2 \subseteq G_2$. In particular, we have $G_1 - G_2 \subseteq$
 275 $V - K_2 \subseteq K_1 - K_2$. \square

276 An iteration of Lemma 12 leads to the main result of this section.

277 **Theorem 14.** *Let X be a Priestley space and V a clopen subset of X . Define*
 278 *a sequence of subsets of $\uparrow V$ as follows:*

$$K_1 = \uparrow V, \quad K_{2i} = \uparrow(K_{2i-1} - V), \quad \text{and} \quad K_{2i+1} = \uparrow(K_{2i} \cap V),$$

279 *for $i \geq 1$. Then, $K_1 \supseteq K_2 \supseteq \dots$ is a decreasing sequence of closed upsets of*
 280 *X and, for every $n \geq 1$, we have*

$$K_n = \{x \in X \mid \deg_V(x) \geq n\} = \uparrow\{x \in X \mid \deg_V(x) = n\}. \quad (7)$$

281 *In particular,*

$$V = \bigcup_{i=1}^m (K_{2i-1} - K_{2i}) = K_1 - (K_2 - (\dots (K_{2m-1} - K_{2m})\dots)), \quad (8)$$

282 *where $2m - 1 = \max\{\deg_V(x) \mid x \in V\}$.*

283 *Proof.* Note that if (7) holds, then (8) holds since $K_{2i-1} - K_{2i}$ will consist
 284 precisely of the elements of V of degree $2i - 1$ and since each element of V
 285 has an odd degree less than or equal to the maximum degree achieved in V .

286 For the first statement and for (7), the proof proceeds by induction on
 287 the parameter i used in (8). The case $i = 1$ is exactly Lemma 12. For
 288 the inductive step, suppose the statements hold for $n \leq 2i$ and notice that
 289 $K_{2i+1} = \uparrow(K_{2i} \cap V)$ and $K_{2i+2} = \uparrow(K_{2i+1} - V)$ are in fact the sets K_1 and
 290 K_2 of Lemma 12 when we apply it to the Priestley space $X' = K_{2i}$ and its
 291 clopen subset $V' = K_{2i} \cap V$. Thus, to complete the proof, it suffices to notice
 292 that, for every $x \in X'$, we have $\deg_V(x) = \deg_{V'}(x) + 2i$. \square

293 Using Corollary 13 we can now prove the minimality of the chain in (8).

294 **Proposition 15.** *Let X be a Priestley space and $V \subseteq X$ be a clopen subset*
 295 *of X . Let $G_1 \supseteq G_2 \supseteq \dots \supseteq G_{2p}$ be a decreasing sequence of closed upsets of*
 296 *X satisfying*

$$V = \bigcup_{i=1}^p (G_{2i-1} - G_{2i}) = G_1 - (G_2 - (\dots (G_{2p-1} - G_{2p}) \dots)). \quad (9)$$

297 *Then, taking $(K_i)_{i \geq 1}$ as defined in Theorem 14, we have $p \geq m$ and, for*
 298 *every $n \in \{1, \dots, p\}$, the following inclusions hold:*

$$K_{2n-1} \subseteq G_{2n-1}, K_{2n} \subseteq G_{2n}, \text{ and } \bigcup_{i=1}^n (G_{2i-1} - G_{2i}) \subseteq \bigcup_{i=1}^n (K_{2i-1} - K_{2i}). \quad (10)$$

299 *Proof.* We proceed by induction on n . The case $n = 1$ is the content of
 300 Corollary 13. Now suppose that (10) holds for a certain $n \in \{1, \dots, p\}$. As
 301 in the proof of Theorem 14, we consider the new Priestley space $X' = K_{2n}$ and
 302 its clopen subset $V' = X' \cap V$. Setting $G'_i = X' \cap G_i$ for each $i \in \{1, \dots, p\}$,
 303 we obtain a decreasing sequence of closed upsets of X' that form a difference
 304 chain for V' . On the other hand, by the induction hypothesis we have

$$\bigcup_{i=1}^n (G'_{2i-1} - G'_{2i}) = \left(\bigcup_{i=1}^n (G_{2i-1} - G_{2i}) \right) \cap K_{2n} \subseteq \left(\bigcup_{i=1}^n (K_{2i-1} - K_{2i}) \right) \cap K_{2n} = \emptyset$$

305 so that the first $2n$ sets do not contribute to the writing of V' as a difference.
 306 It follows that the sequence $G'_{2n+1} \supseteq G'_{2n+2} \supseteq \dots \supseteq G'_{2p}$ is a difference chain

307 of closed upsets of X' for V' . Now applying Corollary 13 to this sequence,
 308 we see that

$$K_{2n+1} = \uparrow(K_{2n} \cap V) = \uparrow V' \subseteq G'_{2n+1} \subseteq G_{2n+1}$$

309 and

$$K_{2n+2} = \uparrow(K_{2n+1} - V) = \uparrow(\uparrow V' - V') \subseteq G'_{2n+2} \subseteq G_{2n+2}.$$

310 We also obtain that $(G_{2n+1} - G_{2n+2}) \cap K_{2n} = G'_{2n+1} - G'_{2n+2} \subseteq K_{2n+1} - K_{2n+2}$.
 311 On the other hand, by Theorem 14, we have

$$(G_{2n+1} - G_{2n+2}) - K_{2n} \subseteq V - K_{2n} = \bigcup_{i=1}^n (K_{2i-1} - K_{2i}).$$

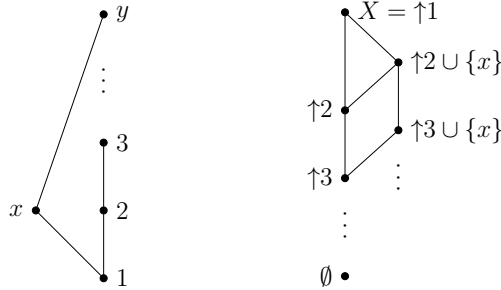
312 We thus conclude that

$$\bigcup_{i=1}^{n+1} (G_{2i-1} - G_{2i}) \subseteq \bigcup_{i=1}^{n+1} (K_{2i-1} - K_{2i})$$

313 as required for the inductive step. □

314 Note that, for $V = \bigcup_{i=1}^m (K_{2i-1} - K_{2i})$ as in Theorem 14, each of the
 315 unions $\bigcup_{i=1}^j (K_{2i-1} - K_{2i})$ is an open subset of X , while $\bigcup_{i=j}^m (K_{2i-1} - K_{2i})$
 316 is closed ($j = 1, \dots, m$). In the next example, we illustrate Theorem 14 in a
 317 case where $m = 2$, $(K_1 - K_2)$ is not closed, and $(K_3 - K_4)$ is not open.

318 **Example 16.** Let $X = \mathbb{N} \cup \{x, y\}$ be the same topological space as in Ex-
 319 ample 9, that is, the one-point compactification by y of the discrete space
 320 $\mathbb{N} \cup \{x\}$, but now ordered by $1 < 2 < \dots < y$ and $1 < x < y$. The order on
 321 X and the dual lattice are depicted below:



322

323 Consider the clopen subset $V = \mathbb{N} \cup \{y\}$ of X . Then, we have:

$$K_1 = X, \quad K_2 = \{x, y\}, \quad K_3 = \{y\}, \quad \text{and} \quad K_4 = \emptyset.$$

324 So $K_1 - K_2 = \mathbb{N}$ is open but not closed and vice versa for $K_3 - K_4 = \{y\}$.

325 As a consequence of Theorem 6, Theorem 14 and Proposition 15 we have
326 the following corollary.

327 **Corollary 17.** *Let D be a co-Heyting algebra and $b \in D^-$. Define the*
328 *following sequence (recall Definition 5):*

$$a_1 = \lceil b \rceil, \quad a_{2i} = \lceil a_{2i-1} - b \rceil, \quad \text{and} \quad a_{2i+1} = \lceil a_{2i} \wedge b \rceil, \quad \text{for } i \geq 1.$$

329 *Then, the sequence $\{a_i\}_{i \geq 1}$ is decreasing, and there exists $m \geq 1$ such that*
330 *$a_{2m+1} = 0$. For the least such m we have*

$$b = a_1 - (a_2 - (\dots (a_{2m-1} - a_{2m}) \dots)) \quad (11)$$

331 *and, for every other writing*

$$b = c_1 - (c_2 - (\dots (c_{2p-1} - c_{2p}) \dots))$$

332 *as a difference chain with $c_1 \geq \dots \geq c_{2p}$ in D , we have $p \geq m$, $c_i \geq a_i$ for $i \in$*
333 *$\{1, \dots, 2p\}$, and for each $n \leq p$ we have $\bigvee_{i=1}^n (c_{2i-1} - c_{2i}) \leq \bigvee_{i=1}^n (a_{2i-1} - a_{2i})$.*

334 We have a order dual algorithm for getting difference chains for Boolean
335 elements over Heyting algebras. To obtain these sequences we use the floor
336 function $\lfloor _ \rfloor : D^- \rightarrow D$ (its existence is the order dual of Proposition 4).

337 **Corollary 18.** *Let D be a Heyting algebra and $b \in D^-$. Define a sequence*
338 *of elements in D as follows:*

$$a_1 = \lfloor \neg b \rfloor, \quad a_{2i} = \lfloor a_{2i-1} \vee b \rfloor, \quad a_{2i+1} = \lfloor a_{2i} \vee \neg b \rfloor, \quad \text{for } i \geq 1.$$

339 *Then, the sequence $\{a_i\}_{i \geq 0}$ is increasing, and there exists $m \geq 1$ such that*
340 *$a_{2m+1} = 1$. For the least such m we have*

$$b = a_{2m} - (a_{2m-1} - (\dots (a_2 - a_1) \dots))$$

341 and, for every other writing

$$b = c_{2p} - (c_{2p-1} - (\dots (c_2 - c_1)\dots))$$

342 as a difference chain with $c_1 \leq \dots \leq c_{2p}$ in D , we have $p \geq m$, $c_i \leq a_i$ for $i \in$
 343 $\{1, \dots, 2p\}$, and for each $n \leq p$ we have $\bigvee_{i=1}^n (c_{2i-1} - c_{2i}) \geq \bigvee_{i=1}^n (a_{2i-1} - a_{2i})$.

344 *Proof.* Apply Corollary 17 to the co-Heyting algebra $D' = \{\neg a \mid a \in D\}$ to
 345 get, for any $b \in B = D^- = (D')^-$, a difference chain of elements in D' :

$$d_1 = \lceil b \rceil, \quad d_{2i} = \lceil d_{2i-1} - b \rceil, \quad \text{and} \quad d_{2i+1} = \lceil d_{2i} \wedge b \rceil,$$

346 where the ceiling function is the one of D' . Since $a - b = \neg b - (\neg a)$, by
 347 Proposition 8, we may write $b \in B$ as

$$b = \neg d_{2m} - (\neg d_{2m-1} - (\dots (\neg d_2 - \neg d_1)\dots)),$$

348 and we have

$$\neg d_1 = \lfloor \neg b \rfloor, \quad \neg d_{2i} = \lfloor \neg d_{2i-1} \vee b \rfloor, \quad \text{and} \quad \neg d_{2i+1} = \lfloor \neg d_{2i} \vee \neg b \rfloor,$$

349 since $\neg \lceil u \rceil = \lfloor \neg u \rfloor$ for all $u \in B$, where the ceiling function is the one for
 350 D' while the floor function is the one for D . Each a_i in the statement of the
 351 corollary is the $\neg d_i$ of the proof. \square

352 Since finite distributive lattices are co-Heyting algebras, Corollary 17 ap-
 353 plies. Combined with the fact that every distributive lattice is the direct
 354 limit of its finite sublattices and that the Booleanization is the direct limit of
 355 the Booleanizations of these finite sublattices, we have a proof of the original
 356 observation by Hausdorff.

357 **Corollary 19.** *Every Boolean element over any distributive lattice may be*
 358 *written as a difference chain of elements of the lattice.*

359 The results of this section were proved using Priestley duality. This makes
 360 them non-constructive. However, we could have proved them in a point-free
 361 setting (with very similar proofs). We trust that anyone interested in the
 362 constructive aspect can see for themselves that this is the case. In Section 5
 363 we will continue in the point-free setting of so-called canonical extensions as
 364 this makes the more involved proofs there simpler.

365 **4. Preliminaries on canonical extensions**

366 Here we provide the required information on canonical extensions. For
 367 further details, please see [10] and [12].

368 *Canonical extensions.* Let D be a distributive lattice and X its dual Priestley
 369 space. Then, Priestley duality implies that the Stone map

$$D \longrightarrow \mathbf{Up}(X, \leq), \quad a \mapsto \hat{a}$$

370 is an embedding of D into the complete lattice $\mathbf{Up}(X, \leq)$ of upsets of the poset
 371 underlying X . An embedding into a complete lattice is called a *completion*,
 372 and canonical extension, first introduced by Jónsson and Tarski [18], comes
 373 about from the fact that the above completion can be uniquely characterized
 374 in abstract terms among all the completions of D . Indeed, it is the unique
 375 completion $e : D \hookrightarrow C$ (up to isomorphism) satisfying the following two
 376 properties:

377 **(dense)** Each element of C is a join of meets and a meet of joins of elements
 378 in the image of D ;

379 **(compact)** For $S, T \subseteq D$ with $\bigwedge e[S] \leq \bigvee e[T]$ in C , there are finite subsets
 380 $S' \subseteq S$ and $T' \subseteq T$ with $\bigwedge e[S'] \leq \bigvee e[T']$.

381 Thus, instead of working with the dual space of a distributive lattice D , we
 382 will work with its canonical extension, denoted D^δ . It comes with an em-
 383 bedding $D \hookrightarrow D^\delta$, which is compact and dense in the above sense. As stated
 384 above this implies (modulo a non-constructive axiom) that D^δ is isomorphic
 385 to $\mathbf{Up}(X, \leq)$, where X is the Priestley space of D , and that the embedding
 386 e is naturally isomorphic to the map $a \mapsto \hat{a}$. In what follows, to lighten the
 387 notation, we will assume (WLOG) that the embedding e is an inclusion so
 388 that D sits as a sublattice in D^δ .

389 *Filter and ideal elements.* Since D sits in D^δ as the clopen upsets sit in
 390 $\mathbf{Up}(X, \leq)$, the join-closure of D in D^δ corresponds to the lattice of open
 391 upsets of X . One can show that these are in one-to-one correspondence
 392 with the ideals of D . Thus we denote the join-closure of D in D^δ by $I(D^\delta)$
 393 and call the elements of $I(D^\delta)$ *ideal elements* of D^δ . Similarly the meet-
 394 closure of D in D^δ corresponds to the lattice of closed upsets of X and one
 395 can show that these are in one-to-one correspondence with the filters of D .

396 Accordingly we denote the meet-closure of D in D^δ by $F(D^\delta)$ and call its
 397 elements *filter elements* of D^δ . Note that, relative to the concepts of filter
 398 and ideal elements, the density property of D^δ states that every element of
 399 D^δ is a join of filter elements and a meet of ideal elements.

400 Another abstract characterization of $F(D^\delta)$ is that it is the free down-
 401 directed meet completion of D . As such it is uniquely determined by the
 402 following two properties [13, Proposition 2.1]:

403 **(filter dense)** Each element of $F(D^\delta)$ is a down-directed meet of elements
 404 from D ;

405 **(filter compact)** For $S \subseteq F(D^\delta)$ down-directed and $a \in D$, if $\bigwedge S \leq a$, then
 406 there is $s \in S$ with $s \leq a$.

407 Notice, that in the particular case of a Boolean algebra B , the order on
 408 the dual space is trivial and thus B^δ is isomorphic to the full powerset of the
 409 dual space X of B . Also, the ideal elements of B^δ correspond to all the opens
 410 of X while the filter elements of B^δ correspond to all the closed subsets of X .

411 *Ceiling functions at the level of canonical extensions.* We saw in Proposi-
 412 tion 4, that if D is a co-Heyting algebra, then the inclusion of D in its
 413 Booleanization, D^- , has a lower adjoint $\lceil \rceil : D^- \rightarrow D$. Here we will show,
 414 that *on the level of canonical extensions* any embedding has a lower adjoint
 415 with nice properties for filter elements.

416 Consider a situation where we have a Boolean algebra B and a sublattice
 417 D of B . Then the embedding of D in B extends to a complete embedding
 418 $D^\delta \hookrightarrow B^\delta$ [12, Theorem 3.2] which restricts to embeddings for the filter
 419 elements as well as for the ideal elements [12, Theorem 2.19]. Since the
 420 embedding is complete, it has both an upper and a lower adjoint, see Propo-
 421 sition 1. We are interested in the lower adjoint, which we will study via the
 422 corresponding closure operator $\bar{(\)} : B^\delta \rightarrow D^\delta \hookrightarrow B^\delta$, cf. Proposition 2. Thus
 423 we have, for $v \in B^\delta$,

$$\bar{v} = \bigwedge \{u \in D^\delta \mid v \leq u\} = \bigwedge \{y \in I(D^\delta) \mid v \leq y\},$$

424 where the second equality follows by the density property of D^δ . We are
 425 particularly interested in the restriction of this closure operator to the filter
 426 elements.

427 **Proposition 20.** *Let B be a Boolean algebra and D a sublattice of B , and*
 428 *let $\overline{(\)}: B^\delta \rightarrow B^\delta$ be the closure operator associated with $D \leq B$ as above.*
 429 *Then the following properties hold:*

430 (a) *for each $u \in B^\delta$, \bar{u} is the least element of D^δ which lies above u ;*

431 (b) *the map $\overline{(\)}: B^\delta \rightarrow B^\delta$ sends filter elements to filter elements;*

432 (c) *the map $\overline{(\)}: F(B^\delta) \rightarrow F(B^\delta)$ preserves down-directed meets.*

433 *Proof.* Part (a) follows by the definition of adjoints. For (b), let $v \in F(B^\delta)$
 434 and let $y \in I(D^\delta)$ with $v \leq y$. Since $y = \bigvee\{a \in D \mid a \leq y\}$, by compactness,
 435 there is $a_y \in D$ with $v \leq a_y \leq y$. Thus, we have

$$\bar{v} = \bigwedge\{y \in I(D^\delta) \mid v \leq y\} = \bigwedge\{a_y \in D \mid v \leq y\} \in F(D^\delta) \subseteq F(B^\delta).$$

436 For (c), let S be a down-directed subset of $F(B^\delta)$ with $v = \bigwedge S$. By (b),
 437 $\bar{v} \in F(D^\delta)$, and thus $\bar{v} = \bigwedge\{a \in D \mid v \leq a\}$. Let $a \in D$ with $v \leq a$, then
 438 $\bigwedge S \leq a$ and by the filter compactness property of $F(B^\delta)$, there is $w_a \in S$
 439 with $w_a \leq a$. Therefore we have

$$\bigwedge\{\bar{w} \mid w \in S\} \leq \bigwedge\{\bar{w}_a \mid v \leq a \in D\} \leq \bigwedge\{a \mid v \leq a \in D\} = \bar{v}.$$

440 On the other hand, since $v \leq w$ for each $w \in S$, by monotonicity of the
 441 closure operator, we also have $\bar{v} \leq \bigwedge\{\bar{w} \mid w \in S\}$ and thus the closure
 442 operator, restricted to $F(B^\delta)$, preserves down-directed meets. \square

443 **Remark 21.** Notice that if B is the Booleanization of D , and X is the
 444 Priestley space of D , then $B^\delta \cong \mathcal{P}(X)$, $D^\delta \cong \mathbf{Up}(X, \leq)$, and the closure
 445 operator is the map $S \mapsto \uparrow S$. Furthermore, Proposition 20(b) tells us that
 446 if $K \subseteq X$ is closed then so is $\uparrow K$. That is, it is the canonical extension
 447 formulation of the second assertion in Proposition 7. We did not prove
 448 Proposition 20(c) in topological terms, but we could have. It says that if
 449 $\{W_i\}_{i \in I}$ is a down-directed family of closed subsets of a Priestley space, then

$$\bigcap_{i \in I} \uparrow W_i = \uparrow \left(\bigcap_{i \in I} W_i \right).$$

450 A statement that is not true in general for down-directed families of subsets
 451 of a poset.

452 **5. The difference hierarchy and families of closure operators**

453 Our main aim in this section is Theorem 27 and, more specifically, the
 454 Corollary 30, which will provide a simple proof of an important result in
 455 logic on words in Section 8. We will be working in the following general
 456 setting. We have a Boolean algebra B and D a sublattice of B . Recall
 457 that the Boolean subalgebra of B generated by D is (up to isomorphism)
 458 the Booleanization D^- of D . Accordingly, we work with D^- as being this
 459 generated subalgebra of B .

460 We start by formulating Theorem 14 in terms of canonical extensions and
 461 closure operators. As in the previous section, we let $\overline{(\)}: B^\delta \rightarrow D^\delta \subseteq B^\delta$ be
 462 the closure operator which is the lower adjoint of the (complete) embedding
 463 $D^\delta \hookrightarrow B^\delta$. Given $b \in B$, we define the sequence $\{k_n\}_{n \geq 1}$ in D^δ as follows:

$$k_1 = \bar{b}, \quad k_{2n} = \overline{k_{2n-1} - b}, \quad \text{and} \quad k_{2n+1} = \overline{k_{2n} \wedge b}, \quad \text{for } n \geq 1. \quad (12)$$

464 Notice that $\{k_n\}_{n \geq 1}$ is a decreasing sequence of filter elements of D^δ . More-
 465 over, in the case where $B = D^-$, by Remark 21, this is exactly the canonical
 466 extension incarnation of the sequence $\{K_n\}_{n \in \mathbb{N}}$ for $V = \widehat{b}$ in Theorem 14. We
 467 thus have the following:

468 **Theorem 22.** *Let D be a distributive lattice, $B = D^-$ and, for $b \in D^-$,
 469 consider the sequence $\{k_n\}_{n \geq 1}$ of D^δ as defined in (12). Then, there exists
 470 $m \geq 1$ such that $k_{2m+1} = 0$, and*

$$b = k_1 - (k_2 - (\dots (k_{2m-1} - k_{2m}) \dots)) = \bigvee_{l=1}^m (k_{2l-1} - k_{2l}).$$

471 We will need the following slight generalization of Theorem 22.

472 **Corollary 23.** *Let B be a Boolean algebra and D a sublattice of B . If $b \in D^-$
 473 and $\{k_n\}_{n \geq 1}$ is the sequence defined in (12), then there exists $m \geq 1$ such
 474 that $k_{2m+1} = 0$ and*

$$b = k_1 - (k_2 - (\dots (k_{2m-1} - k_{2m}) \dots)) = \bigvee_{l=1}^m (k_{2l-1} - k_{2l}).$$

475 *Proof.* The embedding $e: D \hookrightarrow B$ factors through D^- so that $e = e_2 e_1$ where
 476 $e_1: D \hookrightarrow D^-$ and $e_2: D^- \hookrightarrow B$. These maps all lift to complete embeddings

477 with $e^\delta = e_2^\delta e_1^\delta$ [12, Theorem 3.2 and Theorem 2.33], and these all have lower
 478 adjoints, $f, f_2,$ and $f_1,$ respectively. It follows that $f = f_1 f_2$. In a picture:

$$\left(\begin{array}{ccc} D^\delta & \xrightarrow{e_1^\delta} & (D^-)^\delta \\ \xleftarrow{f_1} & & \xleftarrow{f_2} \\ & & B^\delta \end{array} \right) = \left(\begin{array}{ccc} D^\delta & \xrightarrow{e^\delta} & B^\delta \\ \xleftarrow{f} & & \end{array} \right)$$

479

480 Now the closure operator associated with D^- is the map $c_1 = e_1^\delta f_1$, whereas
 481 the closure operator associated with B is the map $c = e^\delta f$. The difference
 482 between Theorem 22 and Corollary 23 is that in the former c_1 is used to
 483 produce the sequence $\{k_n\}_{n \geq 1}$, and in the latter c is used. Thus the corollary
 484 follows if we can show that for all $x \in (D^-)^\delta$ we have $c(x) = c_1(x)$, or
 485 including the action of the inclusion e_2^δ , that $c(e_2^\delta(x)) = e_2^\delta(c_1(x))$. This is
 486 verified by the following calculation.

$$c(e_2^\delta(x)) = (e^\delta f e_2^\delta)(x) = (e_2^\delta e_1^\delta f_1 f_2 e_2^\delta)(x) = e_2^\delta(c_1([f_2 e_2^\delta](x))) = e_2^\delta(c_1(x))$$

487 since $f_2 e_2^\delta = \text{id}_{(D^-)^\delta}$. □

488 Motivated by the application to logic on words presented in Section 8, we
 489 will now work in the following more general setting:

490 **Definition 24.** Let B be a Boolean algebra and I a directed partially or-
 491 dered set. A *directed family of closure operators* on B indexed by I is a family
 492 of closure operators $\{\overline{(\)}^i : B \rightarrow B\}_{i \in I}$ satisfying the following conditions:

- 493 (a) The meet-subsemilattices $S_i := \{b \in B \mid \overline{b}^i = b\}$ for $i \in I$ form an
 494 I -directed family of subsets of B . That is, $S_i \subseteq S_j$ whenever $i \leq j$.
- 495 (b) $D := \bigcup_{i \in I} S_i$ is a sublattice of B . That is, if $a, b \in S_i$ then there is j
 496 with $a \vee b \in S_j$.

497 We start by showing that we have the following relationship between the
 498 closure operators $\overline{(\)}^i$ and the one given by D .

499 **Proposition 25.** Let B be a Boolean algebra and $\{\overline{(\)}^i : B \rightarrow B\}_{i \in I}$ a
 500 directed family of closure operators on B . Then, for each $x \in B$, we have:

$$\overline{x} = \bigwedge_{i \in I} \overline{x}^i$$

501 where the meet is taken in B^δ .

502 *Proof.* For each $x \in B$, $x \leq \bar{x}^i \in S_i \subseteq D \subseteq D^\delta$. Also \bar{x} is the least element
503 of D^δ above x . Thus $\bar{x} \leq \bigwedge_{i \in I} \bar{x}^i$. On the other hand, by Proposition 20(b),
504 since $x \in B \subseteq F(B^\delta)$, we have $\bar{x} \in F(D^\delta)$. That is, $\bar{x} = \bigwedge \{a \in D \mid x \leq a\}$.
505 Now let $a \in D$ with $x \leq a$. Then, since $D = \bigcup_{i \in I} S_i$, there is $j \in I$ with
506 $a \in S_j$. Now using the fact that $x \leq a$ and the monotonicity of $(\bar{\quad})^j$ we obtain

$$\bigwedge_{i \in I} \bar{x}^i \leq \bar{x}^j \leq \bar{a}^j = a.$$

507 We thus have

$$\bigwedge_{i \in I} \bar{x}^i \leq \bigwedge \{a \in D \mid x \leq a\} = \bar{x}. \quad \square$$

508 Now, for each $i \in I$, we define the $\{c_{n,i}\}_{n \geq 1}$ as follows:

$$c_{1,i} = \bar{b}^i, \quad c_{2n,i} = \overline{c_{2n-1,i} - \bar{b}^i}, \quad \text{and} \quad c_{2n+1,i} = \overline{c_{2n,i} \wedge \bar{b}^i} \quad (13)$$

509 **Lemma 26.** *The following properties hold for the sequences as defined above:*

510 (a) $i \leq j$ implies $k_n \leq c_{n,j} \leq c_{n,i}$ for all $n \in \mathbb{N}$ and $i, j \in I$;

511 (b) $k_n = \bigwedge_{i \in I} c_{n,i}$ for all $n \in \mathbb{N}$.

512 *Proof.* Define $k_0 = c_{0,i} = 1$ for all $i \in I$. Also, define $b_n = b$ for n odd and
513 $b_n = \neg b$ for n even then we have, for all $n \geq 1$, $k_{n+1} = \overline{k_n \wedge b_n}$ and similarly
514 for the $c_{n,i}$. Proceeding by induction on n , we suppose (a) holds for $n \in \mathbb{N}$
515 and that $i \leq j$. Note that since $S_i \subseteq S_j \subseteq D$, we have $\bar{x} \leq \bar{x}^j \leq \bar{x}^i$ for all
516 $x \in B$. Also, by the induction hypothesis $k_n \leq c_{n,j} \leq c_{n,i}$, and thus we have

$$\overline{k_n \wedge b_n} \leq \overline{c_{n,j} \wedge b_n} \leq \overline{c_{n,j} \wedge b_n^j} \leq \overline{c_{n,j} \wedge b_n^i} \leq \overline{c_{n,i} \wedge b_n^i}.$$

517 That is, $k_{n+1} \leq c_{n+1,j} \leq c_{n+1,i}$ as required.

518 For (b), again the case $n = 0$ is clear by definition and we suppose
519 $k_n = \bigwedge_{i \in I} c_{n,i}$. Then we have

$$k_{n+1} = \overline{k_n \wedge b_n} = \overline{\left(\bigwedge_{i \in I} c_{n,i} \right) \wedge b_n} = \overline{\bigwedge_{i \in I} (c_{n,i} \wedge b_n)}.$$

520 Now applying Proposition 20(c) and then Proposition 25, we obtain

$$k_{n+1} = \bigwedge_{i \in I} \overline{c_{n,i} \wedge b_n} = \bigwedge_{i \in I} \bigwedge_{j \in I} \overline{c_{n,i} \wedge b_n^j}.$$

521 Now given $i, j \in I$, since I is directed, there is $k \in I$ with $i, j \leq k$. By
 522 Lemma 26(a) we have $c_{n,k} \leq c_{n,i}$. Combining this with the fact that $S_j \subseteq S_k$
 523 we obtain

$$\overline{c_{n,k} \wedge b_n^k} \leq \overline{c_{n,i} \wedge b_n^k} \leq \overline{c_{n,i} \wedge b_n^j}$$

524 and thus

$$k_{n+1} = \bigwedge_{(i,j) \in I^2} \overline{c_{n,i} \wedge b_n^j} = \bigwedge_{k \in I} \overline{c_{n,k} \wedge b_n^k} = \bigwedge_{k \in I} c_{n+1,k}. \quad \square$$

525 We are now ready to state and prove our main theorem.

526 **Theorem 27.** *Let B be a Boolean algebra and $\{\overline{(\)}^i : B \rightarrow B\}_{i \in I}$ a directed*
 527 *family of closure operators on B . For each $b \in B$ let $\{k_n\}_{n \geq 1}$ be the sequence*
 528 *of filter elements as defined in (12) and $\{c_{n,i}\}_{n \geq 1, i \in I}$ be the sequence defined*
 529 *in (13). If $b \in D^- \subseteq B$, then, there is $m \in \mathbb{N}$ and $i \in I$ so that, for each*
 530 *$j \in I$ with $i \leq j$ we have*

$$\begin{aligned} b &= k_1 - (k_2 - \dots - (k_{2m-1} - k_{2m}) \dots) = \bigvee_{l=1}^m (k_{2l-1} - k_{2l}) \\ &= c_{1,j} - (c_{2,j} - \dots - (c_{2m-1,j} - c_{2m,j}) \dots) = \bigvee_{l=1}^m (c_{2l-1,j} - c_{2l,j}). \end{aligned}$$

531 *Proof.* Note that for $b \in D^-$ the fact that the first line of the conclusion
 532 holds is simply the content of Corollary 23. The fact that the second line
 533 holds follows inductively from Lemma 28 below. \square

534 **Lemma 28.** *Let $b, b' \in B$ and $v \in B^\delta$ be such that $v \wedge k_{2l+1} = 0$ and*
 535 *$b' \leq k_{2l+2}$. Suppose $b = v \vee (k_{2l+1} - k_{2l+2}) \vee b'$. Then there is an $i \in I$ so*
 536 *that, for each $j \in I$ with $i \leq j$ we have $b = v \vee (c_{2l+1,j} - c_{2l+2,j}) \vee b'$.*

537 *Proof.* Since both v and $k_{2l+1} - k_{2l+2}$ are below $\neg k_{2l+2}$ we have $b \leq \neg k_{2l+2} \vee b'$,
 538 or equivalently, $b \wedge k_{2l+2} \leq b'$. Now by Lemma 26(b) we have $b \wedge \bigwedge_{i \in I} c_{2l+2,i} \leq$
 539 b' and by compactness there is an $i_1 \in I$ so that for all $j \in I$ with $i_1 \leq j$ we

540 have $b \wedge c_{2l+2,j} \leq b'$, or equivalently, $b \leq \neg c_{2l+2,j} \vee b'$. Now, for each $j \in I$
 541 with $i_1 \leq j$

$$\begin{aligned}
 b &= (\neg k_{2l+1} \wedge b) \vee (k_{2l+1} \wedge b) = v \vee (k_{2l+1} \wedge b) \\
 &\leq v \vee (k_{2l+1} \wedge (\neg c_{2l+2,j} \vee b')) = v \vee (k_{2l+1} \wedge \neg c_{2l+2,j}) \vee (k_{2l+1} \wedge b') \\
 &\leq v \vee (k_{2l+1} - c_{2l+2,j}) \vee b' \\
 &\leq v \vee (k_{2l+1} - k_{2l+2}) \vee b' = b.
 \end{aligned}$$

542 Consequently, for each $j \in I$ with $i_1 \leq j$ we have $b = v \vee (k_{2l+1} - c_{2l+2,j}) \vee b'$.
 543 Now, since $c_{2l+2,j} = \overline{c_{2l+1,j} - b^j} \geq c_{2l+1,j} - b$, and thus, $b \geq c_{2l+1,j} - c_{2l+2,j}$,
 544 using also the inequality $c_{2l+1,j} \geq k_{2l+1}$ given by Lemma 26(a), we may
 545 deduce

$$\begin{aligned}
 b &= v \vee b \vee b' \geq v \vee (c_{2l+1,j} - c_{2l+2,j}) \vee b' \\
 &\geq v \vee (k_{2l+1,j} - c_{2l+2,j}) \vee b' = b.
 \end{aligned}$$

546 It then follows that for all $j \in I$ with $j \geq i_1$ we have

$$b = v \vee (c_{2l+1,j} - c_{2l+2,j}) \vee b'. \quad \square$$

547 **Remark 29.** Notice that Corollary 19, stating that Boolean elements over
 548 a lattice are difference chains of elements of the lattice, can also be seen
 549 as a consequence of Theorem 27. Let D be any distributive lattice and
 550 B its Booleanization. For each finite sublattice D' of D , the embedding
 551 $D' \hookrightarrow D \hookrightarrow B$ has an upper adjoint $g' : B \rightarrow D'$ given by $g'(b) = \bigwedge \{a \in$
 552 $D' \mid b \leq a\} = \min\{a \in D' \mid b \leq a\}$. Thus, Theorem 27 applies and we
 553 get Corollary 19. In fact, in this way, we obtain more information as we see
 554 that the minimum length chain in D is equal to the minimum length chain
 555 in $F(D^\delta)$, or equivalently, in the lattice of closed upsets of the dual space of
 556 D . In turn, this is the same as the maximum length of difference chains in
 557 the dual with respect to the clopen corresponding to the given element.

558 In Section 8 we will give an application of the following consequence of
 559 Theorem 27, which needs its full generality.

560 **Corollary 30.** Let B be a Boolean algebra and $\{\overline{(\)}^i : B \rightarrow B\}_{i \in I}$ a directed
561 family of closure operators on B . Let $B' \leq B$ be a Boolean subalgebra closed
562 under each of the closure operators $\overline{(\)}^i$ for $i \in I$. Then,

$$(D \cap B')^- = D^- \cap B',$$

563 where we view the Booleanization of any sublattice of B as the Boolean sub-
564 algebra of B that it generates.

565 *Proof.* Since $D \cap B'$ is contained in both of the Boolean algebras D^- (also
566 viewed as a subalgebra of B) and B' , the Booleanization of $D \cap B'$ is contained
567 in their intersection.

568 For the converse, let $b \in D^- \cap B'$. By Theorem 27, there exists an index j
569 so that b can be written as a difference chain

$$b = c_{1,j} - (c_{2,j} - (\dots - (c_{2m-1,j} - c_{2m,j}) \dots)),$$

570 where $c_{1,j} = \bar{b}^j$, $c_{2n,j} = \overline{c_{2n-1,j} - b^j}$ and $c_{2n+1,j} = \overline{c_{2n,j} \wedge b^j}$, for $n \geq 1$.
571 But then, by hypothesis that B' is closed under $\overline{(\)}^j$ and a straightforward
572 induction argument, it follows that $c_{1,j} \geq \dots \geq c_{2m,j}$ is a chain in $\overline{B'}^j \subseteq$
573 $D \cap B'$. Thus, b belongs to $(D \cap B')^-$. \square

574 **Remark 31.** We remark that the closure of B' under the operators $\overline{(\)}^i$ for
575 $i \in I$ implies that the closure operator $\overline{(\)}^i : B' \rightarrow B'$ on B' , whose image
576 is $S'_i = B' \cap S_i$, is such that $\{S'_i\}_{i \in I}$ is an I -directed family of subsets of B' .
577 Moreover, $D' = \bigcup_{i \in I} S'_i$ is precisely the distributive lattice $D \cap B'$.

578 We give an example to show that the conclusion of Corollary 30 is by no
579 means true in general.

580 **Example 32.** Let $B = \mathcal{P}(\{a, b, c\})$ be the eight-element Boolean algebra.
581 Further, let D be the sublattice generated by $\{a\}$ and $\{a, b\}$ and let B' be
582 the Boolean subalgebra generated by $\{b\}$. Then B is, up to isomorphism, the
583 Booleanization of D , and thus $D^- \cap B' = B'$, whereas $D \cap B' = (D \cap B')^-$
584 is the two-element Boolean subalgebra of B .

585 In order to formulate the application to the theory of formal languages,
586 we will need some concepts from logic on words.

587 **6. Preliminaries on formal languages and logic on words**

588 *Formal languages.* An *alphabet* is a finite set A , a *word over A* is an ele-
 589 ment of the free A -generated monoid A^* , and a *language* is a set of words
 590 over some alphabet. For a word $w \in A^*$, we use $|w|$ to denote the *length*
 591 of w , that is, if $w = a_1 \dots a_n$ with each $a_i \in A$, then we have $|w| = n$.
 592 Given a homomorphism $f : A^* \rightarrow M$ into a finite monoid M , we say that
 593 a language $L \subseteq A^*$ is *recognized by f* provided there is a subset $P \subseteq M$
 594 such that $L = f^{-1}(P)$, or equivalently, if $L = f^{-1}(f[L])$. The language L
 595 is *recognized by a finite monoid M* provided there is a homomorphism into
 596 M recognizing L . Finally, a language is said to be *regular* if it is recognized
 597 by some finite monoid. Notice that the set of all regular languages forms a
 598 Boolean algebra. Indeed, if a language is recognized by a given finite monoid
 599 then so is its complement, and if L_1 and L_2 are recognized, respectively, by
 600 M_1 and M_2 , then $L_1 \cap L_2$ is recognized by the Cartesian product $M_1 \times M_2$.
 601 (The more commonly encountered definitions of ‘regular language’ refer to
 602 finite automata or regular expressions; one can show that these are indeed
 603 the same as the regular languages defined here.)

604 The following well-known technical result [28] will be needed in Section 8.

605 **Lemma 33.** *Let $f : A^* \rightarrow B^*$ be a homomorphism. Then the forward image*
 606 *under f of a regular language over A is a regular language over B .*

607 We are interested in languages defined by first-order formulas of *logic on*
 608 *words* which we briefly introduce now. For further details please see [29,
 609 Chapter II].

610 *Syntax of first-order logic on words.* Fix an alphabet A . We denote *first-*
 611 *order variables* by x, y, z, x_1, x_2, \dots . *First-order formulas* are inductively
 612 built as follows. For each letter $a \in A$, we consider a *letter predicate*, also
 613 denoted by a , which is unary. Thus, for any variable x , $a(x)$ is an (atomic)
 614 formula. A *k -ary numerical predicate* is a function $R : \mathbb{N} \rightarrow \mathcal{P}(\mathbb{N}^k)$ satisfying
 615 $R(n) \subseteq \{1, \dots, n\}^k$ for every $n \in \mathbb{N}$. That is, R is an element of the Boolean
 616 algebra $\prod_{n \in \mathbb{N}} \mathcal{P}(\{1, \dots, n\}^k)$. When we fix a set \mathcal{R} of numerical predicates,
 617 we will assume it forms a Boolean subalgebra of $\prod_{n \in \mathbb{N}} \mathcal{P}(\{1, \dots, n\}^k)$. Each k -
 618 ary numerical predicate R and any sequence x_1, \dots, x_k of first-order variables
 619 define an (atomic) formula $R(x_1, \dots, x_k)$. Finally, Boolean combinations
 620 of formulas are formulas, and if φ is a formula and x_1, \dots, x_k are distinct
 621 variables, then $\forall x_1, \dots, x_k \varphi$ is a formula. In order to simplify the notation,

622 we usually also consider the quantifier \exists : the formula $\exists x_1, \dots, x_k \varphi$ is an
623 abbreviation for $\neg \forall x_1, \dots, x_k \neg \varphi$. As usual in logic, we say that a variable
624 x occurs *freely* in a formula provided it is not in the scope of a quantifier
625 that quantifies over x , and a formula is said to be a *sentence* provided it
626 has no free variables. *Quantifier-free* formulas are those that are Boolean
627 combinations of atomic formulas.

628 *Semantics of first-order logic on words.* Let us fix an alphabet A and a set
629 of numerical predicates \mathcal{R} . To each non-empty word $w = a_1 \dots a_n \in A^*$
630 with $a_i \in A$, we associate the relational structure $\mathcal{M}_w = (\downarrow n, A \cup \mathcal{R})$, where
631 $\downarrow n = \{1, \dots, n\}$, $a^w = \{i \in \downarrow n \mid a_i = a\}$, for each $a \in A$, and $R^w = R(n)$,
632 for each $R \in \mathcal{R}$. Models of first-order sentences are words, while models of
633 formulas with free variables are the so-called *structures*. For a set of distinct
634 variables $\mathbf{x} = \{x_1, \dots, x_k\}$, an \mathbf{x} -*structure* is an element of $A^* \times (\downarrow |w|)^{\mathbf{x}}$. We
635 identify maps from \mathbf{x} to $\downarrow |w|$ with k -tuples $\mathbf{i} = (i_1, \dots, i_k) \in (\downarrow |w|)^k$. Given
636 a word $w \in A^*$ and a vector $\mathbf{i} = (i_1, \dots, i_k) \in (\downarrow |w|)^{\mathbf{x}}$, we denote by $w_{\mathbf{x}=\mathbf{i}}$
637 the \mathbf{x} -structure based on w equipped with the map given by \mathbf{i} . Moreover,
638 if $\mathbf{x} = \{x_1, \dots, x_k\}$ and $\mathbf{y} = \{y_1, \dots, y_\ell\}$ are disjoint sets of variables, $\mathbf{i} =$
639 $(i_1, \dots, i_k) \in (\downarrow |w|)^{\mathbf{x}}$ and $\mathbf{j} = (j_1, \dots, j_\ell) \in (\downarrow |w|)^{\mathbf{y}}$, then $w_{\mathbf{x}=\mathbf{i}, \mathbf{y}=\mathbf{j}}$ denotes
640 the \mathbf{z} -structure $w_{\mathbf{z}=\mathbf{k}}$, where $\mathbf{z} = \mathbf{x} \cup \mathbf{y}$ and $\mathbf{k} = (i_1, \dots, i_k, j_1, \dots, j_\ell)$.

641 We denoted the set of all \mathbf{x} -structures by $A^* \otimes \mathbf{x}$. Let $\varphi(\mathbf{x})$ be a formula
642 all of whose free variables are in \mathbf{x} . A *model* of $\varphi(\mathbf{x})$ is an \mathbf{x} -structure that
643 satisfies $\varphi(\mathbf{x})$, using the standard interpretation of quantifiers in formulas.
644 We denote by $L_{\varphi(\mathbf{x})}$ the set of all models of the formula. We will say that
645 the formula *defines* this set of \mathbf{x} -structures.

646 **Example 34.** Suppose $A = \{a, b\}$. The sentence $\varphi = \exists x, y (x < y \wedge a(x) \wedge$
647 $b(y))$ is read: “there are positions x and y such that x comes before y and
648 there is an a at position x and a b at position y ”. Thus, φ defines the regular
649 language L_φ given by the regular expression $A^*aA^*bA^*$.

650 It is worth remarking how this informal intuitive interpretation matches
651 up with our definitions of the syntax and semantics of first-order formulas.
652 The subformula $x < y$ is the binary numerical predicate, which, formally
653 speaking, maps each $n \in \mathbb{N}$ to the set $\{(i, j) \mid 1 \leq i < j \leq n\} \subseteq (\downarrow n)^2$. Since
654 φ has no free variables, L_φ is a set of \emptyset -structures—that is, simply a set of
655 words over A .

656 We give an example involving formulas that contain free variables.

657 **Example 35.** Again, let $A = \{a, b\}$. Let us define a numerical predicate,
 658 which we denote informally by $x = \frac{y}{2}$. Formally, this maps each $n \in \mathbb{N}$ to the
 659 set $\{(i, 2i) \mid 1 \leq i \leq n/2\}$. The quantifier-free formula $\psi_1(x, y)$ given by

$$\left(x = \frac{y}{2}\right) \wedge b(y)$$

660 defines a set of $\{x, y\}$ -structures. It is convenient to think of each such
 661 structure as a word over A in which each of the two variables labels a position
 662 in the word. Thus, for example, $ab(a, x)ba(b, y)aa$ represents the structure
 663 $abababaa_{x=3, y=6}$, which is a model of $\psi_1(x, y)$.

664 Let ψ_2 be the formula $\exists x \psi_1(x, y)$. This formula has a single free variable
 665 y , says that y labels an even-numbered position, which contains the letter b .
 666 So, for example, $ababa(b, y)aa \in L_{\psi_2(y)}$.

667 Further, let ψ_3 denote the formula $\exists x, y \psi_1(x, y)$. Its set of models L_{ψ_3}
 668 consists of all words that contain the letter b in an even-numbered position.
 669 This is again a regular language, given by the regular expression $(A^2)^* AbA^*$.

670 *Regular languages of structures.* We now give a more formal account of a
 671 device we used informally in the examples above. We fix a set of distinct
 672 variables $\mathbf{x} = \{x_1, \dots, x_k\}$. Then, $2^{\mathbf{x}}$ is isomorphic to the powerset $\mathcal{P}(\mathbf{x})$.
 673 There is a natural embedding of the set of all \mathbf{x} -structures into the free
 674 monoid $(A \times 2^{\mathbf{x}})^*$. Indeed, to an \mathbf{x} -structure $w_{\mathbf{x}=\mathbf{i}}$, where $\mathbf{i} = (i_1, \dots, i_k)$,
 675 we may assign the word $(a_1, S_1) \dots (a_n, S_n)$, where $w = a_1 \dots a_n$ with each
 676 $a_i \in A$ and, for $\ell \in \downarrow n$, $S_\ell = \{x_j \in \mathbf{x} \mid i_j = \ell\}$. It is not hard to see
 677 that this mapping defines an injection $A^* \otimes \mathbf{x} \hookrightarrow (A \times 2^{\mathbf{x}})^*$. Moreover,
 678 an element $(a_1, S_1) \dots (a_n, S_n)$ of $(A \times 2^{\mathbf{x}})^*$ represents an \mathbf{x} -structure under
 679 this embedding precisely when the non-empty sets among S_1, \dots, S_n form a
 680 partition of \mathbf{x} . From hereon, we view $A^* \otimes \mathbf{x}$ as a subset of $(A \times 2^{\mathbf{x}})^*$ without
 681 further mention.

682 Since we view \mathbf{x} -structures as words over $A \times 2^{\mathbf{x}}$, we can talk about *regular*
 683 languages of structures. Moreover, it is easy to see that the set $A^* \otimes \mathbf{x}$ of
 684 all \mathbf{x} -structures is itself a regular language: To see this, let $x \in \mathbf{x}$ be any
 685 variable. Let $N = \{0, 1, m\}$ be the three-element monoid in which 0 is
 686 absorbent, 1 is the identity, and $m^2 = 0$, and let $f : (A \times 2^{\mathbf{x}})^* \rightarrow N$ be
 687 the unique homomorphism satisfying $f(a, S) = m$ if $x \in S$ and $f(a, S) =$
 688 1 otherwise. This homomorphism recognizes, via $\{m\}$, the set of words
 689 $(a_1, S_1) \dots (a_n, S_n)$ such that the variable x occurs exactly once among the
 690 S_i . The set of structures $A^* \otimes \mathbf{x}$ is the intersection of these languages over

691 all $x \in \mathbf{x}$, and thus is a regular language itself, since the family of regular
 692 languages is closed under finite intersection.

693 As a result, a language L of structures is a regular language if and only
 694 if it consists of all the \mathbf{x} -structures in some regular language L' over $A \times 2^{\mathbf{x}}$,
 695 because we then have $L = L' \cap (A^* \otimes \mathbf{x})$. We will make use of this observation
 696 in the next section.

697 **Example 36.** We consider again the formulas of Example 35. The set L
 698 of structures defined by $\psi_1(x, y)$ is a non-regular language: Consider the
 699 homomorphism $\alpha : (A \times 2^{\{x, y\}})^* \rightarrow \{c, d\}^*$ defined by mapping every letter of
 700 the form (e, \emptyset) to c and all other letters to d . Then $\alpha[L] = \{c^k d c^k d c^\ell\}_{k, \ell \geq 0}$. If
 701 L were regular, then Lemma 33 implies that so is $\alpha[L]$. However, well-known
 702 elementary techniques of the theory of automata show that this is not the
 703 case. (See, e.g. [29, Chapter I], or a standard textbook like Sipser [26].)

704 On the other hand, the set of structures defined by $\psi_2(y)$ is regular: It is
 705 just the language given by the regular expression $(A^2)^* A(b, \{y\}) A^*$.

706 *Fragments of first-order logic.* Formulas will always be considered up to se-
 707 mantic equivalence, even if not explicitly said. We denote by $\mathbf{FO}[\mathcal{N}]$ the set
 708 of all first-order sentences with arbitrary numerical predicates (up to seman-
 709 tic equivalence). For formulas whose free variables are in \mathbf{x} , we will write
 710 $\mathbf{FO}_{\mathbf{x}}[\mathcal{N}]$. And for a set \mathcal{R} of numerical predicates, $\mathbf{FO}[\mathcal{R}]$ denotes the set
 711 of first-order sentences using numerical predicates from \mathcal{R} . Notice that, as
 712 a Boolean algebra, $\mathbf{FO}[\mathcal{N}]$ is naturally equipped with a partial order, which
 713 in turn may be characterized in terms of semantic containment: $\varphi \leq \psi$ if
 714 and only if $L_\varphi \subseteq L_\psi$. For this reason, we will identify formulas and the
 715 corresponding languages of models switching freely between φ and L_φ . In
 716 particular, we see $\mathbf{FO}[\mathcal{N}]$ as a Boolean subalgebra of $\mathcal{P}(A^*)$ and $\mathbf{FO}_{\mathbf{x}}[\mathcal{N}]$ as
 717 a Boolean subalgebra of $\mathcal{P}(A^* \otimes \mathbf{x})$.

718 *Quantifier alternation.* We can measure the complexity of first-order formu-
 719 las by the minimum number of alternations of quantifiers that is needed to
 720 express them in prenex-normal formula, that is, in the form

$$\psi = Q_1 \mathbf{x}_1 \dots Q_m \mathbf{x}_m \varphi(\mathbf{x}_1, \dots, \mathbf{x}_m), \quad (14)$$

721 where φ is a quantifier-free formula, $Q_1, \dots, Q_m \in \{\forall, \exists\}$ and $Q_\ell = \forall$ if and
 722 only if $Q_{\ell+1} = \exists$ for each $\ell = 1, \dots, m - 1$. It is a well-known fact that
 723 for every first-order formula there is a semantically equivalent one in prenex-
 724 normal form. For $m \geq 1$ and a set of numerical predicates \mathcal{R} , $\Pi_m[\mathcal{R}]$ consists

725 of all the sentences of $\mathbf{FO}[\mathcal{R}]$ that are semantically equivalent to a sentence
726 of the form (14) where $Q_1 = \forall$. This is the *alternation hierarchy* over \mathcal{R} .
727 In particular, we have $\Pi_m[\mathcal{R}] \subseteq \Pi_\ell[\mathcal{R}]$ whenever $m \leq \ell$. Similarly, $\Sigma_m[\mathcal{R}]$
728 denotes the set of all sentences that are semantically equivalent to a sentence
729 of the form (14) with $Q_1 = \exists$, and $\Sigma_m[\mathcal{R}] \subseteq \Sigma_\ell[\mathcal{R}]$ whenever $m \leq \ell$. It is
730 not hard to see that both $\Pi_m[\mathcal{R}]$ and $\Sigma_m[\mathcal{R}]$ are closed under disjunction
731 and conjunction but not under negation in general. In other words, $\Pi_m[\mathcal{R}]$
732 and $\Sigma_m[\mathcal{R}]$ are lattices, but not Boolean algebras. We denote by $\mathcal{B}\Pi_m[\mathcal{R}]$
733 and by $\mathcal{B}\Sigma_m[\mathcal{R}]$ the Boolean algebras generated by $\Pi_m[\mathcal{R}]$ and by $\Sigma_m[\mathcal{R}]$,
734 respectively, that is,

$$\mathcal{B}\Pi_m[\mathcal{R}] = (\Pi_m[\mathcal{R}])^- \quad \text{and} \quad \mathcal{B}\Sigma_m[\mathcal{R}] = (\Sigma_m[\mathcal{R}])^-.$$

735 Clearly, we have $\mathcal{B}\Pi_m[\mathcal{R}] = \mathcal{B}\Sigma_m[\mathcal{R}] \subseteq \Pi_{m+1}[\mathcal{R}], \Sigma_{m+1}[\mathcal{R}]$.

736 When \mathcal{R} consists of just the order relation $<$, then the languages defined
737 by first-order sentences are all regular. In this case, the alternation hierarchy
738 has been extensively studied. An outstanding open problem is to determine
739 effectively whether a given first-order definable regular language belongs to
740 the m^{th} level of the hierarchy. The only cases for which this is known is when
741 $m \leq 2$ (see [24]).

742 In this paper we are only concerned with the first level of the hierarchy
743 over different bases of numerical predicates. For notational convenience, we
744 will work with the fragment $\Pi_1[\mathcal{R}]$, although everything we prove for $\Pi_1[\mathcal{R}]$
745 admits a dual statement for $\Sigma_1[\mathcal{R}]$. Every formula of $\Pi_1[\mathcal{R}]$ is of the form
746 $\psi = \forall \mathbf{x} \varphi(\mathbf{x})$, for some quantifier-free formula $\varphi(\mathbf{x})$. Inside $\Pi_1[\mathcal{R}]$, we classify
747 formulas according to the size of \mathbf{x} : we let $\Pi_1^k[\mathcal{R}]$ consist of all equivalence
748 classes of such formulas for which there is a representative ψ for which \mathbf{x}
749 has k variables. We remark that, $\Pi_1^k[\mathcal{R}]$ is closed under conjunction, since
750 the formulas $\forall \mathbf{x} \varphi(\mathbf{x}) \wedge \forall \mathbf{x} \psi(\mathbf{x})$ and $\forall \mathbf{x} (\varphi(\mathbf{x}) \wedge \psi(\mathbf{x}))$ are semantically
751 equivalent. However, it is easy to see that, in general, $\Pi_1^k[\mathcal{R}]$ fails to be
752 closed under disjunction, as we now show.

753 **Example 37.** Let $\varphi(x) = a(x)$ and $\psi(x) = b(x)$. Then, $\forall x \varphi(x)$ defines the
754 language a^* , while $\forall x \psi(x)$ defines the language b^* and thus these are both
755 in $\Pi_1^1[\mathcal{N}]$. The disjunction $\forall x \varphi(x) \vee \forall x \psi(x)$ defines the language $a^* \cup b^*$,
756 while $\forall x (\varphi(x) \vee \psi(x))$ defines the language $\{a, b\}^*$. Indeed, one can show
757 that $\forall x \varphi(x) \vee \forall x \psi(x)$ is not in $\Pi_1^1[\mathcal{N}]$ while it is in $\Pi_1^2[\mathcal{N}]$ as witnessed by
758 the sentence $\forall x_1, x_2 (\varphi(x_1) \vee \psi(x_2))$.

759 We will use \mathbf{Reg} to denote the set of first-order sentences φ for which L_φ
760 is a regular language. We will also use \mathbf{Reg} to denote the set of numerical
761 predicates for which the associated language of structures is regular. In [2]
762 it is shown that

$$\mathbf{FO}[\mathcal{N}] \cap \mathbf{Reg} = \mathbf{FO}[\mathbf{Reg}].$$

763 In other words, if a first-order sentence defines a regular language L , then L
764 can be defined using only regular numerical predicates. This is proved as a
765 consequence of results in circuit complexity.

766 It is conjectured that this equality holds at each level of the alternation
767 hierarchy, in other words that, for $m \geq 0$, we have

$$\mathcal{B}\Pi_m[\mathcal{N}] \cap \mathbf{Reg} = \mathcal{B}\Pi_m[\mathbf{Reg}]. \quad (15)$$

768 For $m > 1$, this question is open. In Section 8, we will use the results of
769 Section 5 to provide a proof of the case $m = 1$. This was first proved in [20],
770 and a different proof appears in [30]. The proof in [20] relies on some hard
771 results in semigroup theory and ideas from circuit complexity, and the one
772 in [30] on Ramsey theory coupled with the algebra of finite semigroups. In
773 contrast, our argument is entirely different and quite elementary.

774 **Example 38.** In an earlier example, we saw that the sentence

$$\exists x, y \left(x = \frac{y}{2} \wedge b(y) \right)$$

775 defines the regular language $(A^2)^*AbA^*$ over $A = \{a, b\}$. On the other hand
776 $x = \frac{y}{2}$ is a non-regular numerical predicate. The dual equality of (15) for
777 existential fragments and $m = 0$ implies that the same language can be
778 defined by a sentence of $\Sigma_1[\mathbf{Reg}]$. Such a sentence is given by

$$\exists y ((y \equiv 0 \pmod{2}) \wedge b(y)),$$

779 in which the numerical predicate is regular.

780 7. A lattices-and-duality perspective on logic on words

781 *Universal quantifiers as adjoints.* Again, we fix a finite alphabet A and a set
782 of variables \mathbf{x} . We consider the projection map given by

$$\pi: A^* \otimes \mathbf{x} \rightarrow A^*, \quad w_{\mathbf{x}=\mathbf{i}} \mapsto w.$$

783 This gives rise, via the duality between sets and complete and atomic Boolean
784 algebras, to the complete embedding of Boolean algebras

$$\pi^{-1} = () \otimes \mathbf{x}: \mathcal{P}(A^*) \hookrightarrow \mathcal{P}(A^* \otimes \mathbf{x}), \quad L \mapsto \pi^{-1}(L) = L \otimes \mathbf{x}$$

785 This embedding, being a complete homomorphism between complete lattices,
786 has an upper adjoint which we may call \forall (and a lower adjoint \exists). These are
787 given by

$$\begin{aligned} \forall: \mathcal{P}(A^* \otimes \mathbf{x}) &\rightarrow \mathcal{P}(A^*) \\ K \mapsto \forall K &= \max\{L \in \mathcal{P}(A^*) \mid L \otimes \mathbf{x} \subseteq K\} \\ &= \{w \in A^* \mid \forall \mathbf{i} \in \{1, \dots, |\mathbf{x}|\}^{\mathbf{x}}, w_{\mathbf{x}=\mathbf{i}} \in K\} \\ &= (\pi[K^c])^c \end{aligned}$$

788 and similarly

$$\begin{aligned} \exists: \mathcal{P}(A^* \otimes \mathbf{x}) &\rightarrow \mathcal{P}(A^*) \\ K \mapsto \exists K &= \min\{L \in \mathcal{P}(A^*) \mid K \subseteq L \otimes \mathbf{x}\} \\ &= \{w \in A^* \mid \exists \mathbf{i} \in \{1, \dots, |\mathbf{x}|\}^{\mathbf{x}}, w_{\mathbf{x}=\mathbf{i}} \in K\} \\ &= \pi[K] \end{aligned}$$

789 As is well-known in categorical logic, \forall and \exists are the semantic incarnations
790 of the classical universal and existential quantifiers, respectively. Explicitly,
791 for the universal quantifier, when $K = L_{\varphi(\mathbf{x})}$ is definable by a formula $\varphi(\mathbf{x})$,
792 we have

$$\forall L_{\varphi(\mathbf{x})} = \{w \in A^* \mid \forall \mathbf{i} \in \{1, \dots, |\mathbf{x}|\}^{\mathbf{x}}, w_{\mathbf{x}=\mathbf{i}} \models \varphi(\mathbf{x})\} = L_{\forall \mathbf{x} \varphi(\mathbf{x})}. \quad (16)$$

793 *Recognition, model theoretic types, and duality.* Let $\text{LOG}_{\mathbf{x}}$ denote the Boolean
794 algebra of formulas with free variables in \mathbf{x} relative to some logic (up to se-
795 mantic equivalence), e.g. $\text{LOG}_{\mathbf{x}} = \mathbf{FO}_{\mathbf{x}}[\mathcal{N}]$. Further, let X be the dual space

796 of this Boolean algebra and let $\mathbf{x}\text{-Str}$ denote the set (or class) of intended
 797 models for this logic, e.g. $\mathbf{x}\text{-Str} = A^* \otimes \mathbf{x}$. Then we have a mapping

$$\begin{aligned} \mathbf{typ} : \mathbf{x}\text{-Str} &\longrightarrow X \\ (\mathcal{M}, \mathbf{i}) &\mapsto \{\varphi(\mathbf{x}) \mid \mathcal{M} \models \varphi(\mathbf{i})\} \end{aligned} \tag{17}$$

798 which sends a model to the ultrafilter of formulas that it satisfies. In model
 799 theory, the image of this map is known as the *type* of the given model. In
 800 language theory, this kind of map is used to study the logic itself since it is
 801 encoded via recognition. That is, this map *topologically recognizes* the logic
 802 in the sense that the Boolean algebra $\text{LOG}_{\mathbf{x}}$ is isomorphic to the Boolean
 803 subalgebra of those $L \subseteq \mathcal{P}(\mathbf{x}\text{-Str})$ such that

$$L = \mathbf{typ}^{-1}(V) \text{ for some clopen } V \subseteq X.$$

804 In classical model theory, Gödel's Completeness Theorem tells us that every
 805 type $x \in X$ is realized (by some model). However, for this to be the case
 806 here, we would need to consider not just the finite models of our logic but the
 807 so-called pseudo-finite models. Then the map in (17) becomes surjective and
 808 topological methods may be applied. This is closely related to recognition
 809 by profinite monoids as studied in language theory, where these ideas are
 810 combined with those of monoid recognition as described above. See [14, 19]
 811 for a study of the connections between model theoretic type theory and
 812 recognition in language theory and [11] for a study of the connections between
 813 Stone duality and recognition in language theory.

814 Here, we will be able to work *just with the finite models*. This is because
 815 the logic fragment we want to consider here consists of the quantifier-free
 816 formulas in $\mathbf{FO}_{\mathbf{x}}[\mathcal{N}]$ and, as we will see in Corollary 40, these form a complete
 817 and atomic Boolean algebra and thus fall within the discrete duality between
 818 sets and complete and atomic Boolean algebras.

819 *Quantifier-free formulas.* Consider a set of distinct variables $\mathbf{x} = \{x_1, \dots, x_k\}$.
 820 We will give an algebraic characterization of the languages of the form $L_{\varphi(\mathbf{x})}$
 821 for $\varphi(\mathbf{x})$ a quantifier-free formula whose free variables are in \mathbf{x} .

822 We first provide a characterization of these languages via discrete duality,
 823 bringing out the connection between recognition and the notion of types from
 824 model theory as described above. For this purpose, we say that $L \subseteq X$ is *set*
 825 *theoretically recognized* by $f: X \rightarrow Y$ provided there is a subset $P \subseteq Y$ with
 826 $L = f^{-1}(P)$.

827 We will need the following notation: For $w = a_1 \dots a_n$ with each $a_i \in A$
828 and $\mathbf{i} = (i_1, \dots, i_k)$, we write $w(\mathbf{i})$ for the tuple $(a_{i_1}, \dots, a_{i_k})$. For a vector
829 of letters $\mathbf{a} = (a_1, \dots, a_k) \in A^{\mathbf{x}}$, we denote by $\mathbf{a}(\mathbf{x})$ the conjunction $a_1(x_1) \wedge$
830 $\dots \wedge a_k(x_k)$.

831 **Lemma 39.** *Let $K \subseteq A^* \otimes \mathbf{x}$. Then, K is given by a quantifier-free first-*
832 *order formula over \mathbf{x} if and only if is it set theoretically recognized by the*
833 *map*

$$c_A : A^* \otimes \mathbf{x} \rightarrow \mathbb{N}^{k+1} \times A^k, \quad w_{\mathbf{x}=\mathbf{i}} \mapsto (|w|, \mathbf{i}, w(\mathbf{i})).$$

834 *Proof.* First suppose that $K = c_A^{-1}(P)$ for some $P \subseteq \mathbb{N}^{k+1} \times A^k$. For each
835 $\mathbf{a} \in A^k$ and $n \in \mathbb{N}$, let

$$R^{\mathbf{a}}(n) = \{\mathbf{i} \in (\downarrow n)^k \mid (n, \mathbf{i}, \mathbf{a}) \in P\}.$$

836 Then $R^{\mathbf{a}}$ is a (k -ary) numerical predicate for each $\mathbf{a} \in A^k$ and it is not
837 difficult to see that $c_A^{-1}(P) = L_{\varphi(\mathbf{x})}$ for

$$\varphi(\mathbf{x}) = \bigvee_{\mathbf{a} \in A^k} (\mathbf{a}(\mathbf{x}) \wedge R^{\mathbf{a}}(\mathbf{x})).$$

838 On the other hand, for $a \in A$ and $i \in \{1, \dots, k\}$

$$L_{a(x_i)} = c_A^{-1}(\mathbb{N}^{k+1} \times \{\mathbf{a} \in A^k \mid a_i = a\})$$

839 and for $R \subseteq \mathbb{N}^{m+1}$ an m -ary numerical predicate, and (not necessarily dis-
840 tinct) variables $y_1, \dots, y_m \in \mathbf{x}$, we have

$$L_{R(y_1, \dots, y_m)} = c_A^{-1}(R' \times A^k)$$

841 where $(i_1, \dots, i_k) \in R'(n)$ if and only if $(j_1, \dots, j_m) \in R(n)$ where $j_s = i_t$ if
842 and only if $y_s = x_t$. □

843 Now, we obtain an *algebraic* characterization of the quantifier free formu-
844 las. That is, a characterization in the form of recognition by a monoid rather
845 than just by a set. for this purpose, let $\varepsilon \notin A$ be a new symbol and denote
846 $A_\varepsilon = A \cup \{\varepsilon\}$. We consider the homomorphism $\Theta_{\mathbf{x}} : (A \times 2^{\mathbf{x}})^* \rightarrow (A_\varepsilon \times 2^{\mathbf{x}})^*$
847 given by

$$\Theta_{\mathbf{x}}(a, S) = \begin{cases} (a, S), & \text{if } S \neq \emptyset; \\ (\varepsilon, S), & \text{if } S = \emptyset. \end{cases}$$

848 Notice that, given \mathbf{x} -structures $v_{\mathbf{x}=\mathbf{i}}$ and $w_{\mathbf{x}=\mathbf{j}}$, we have

$$\Theta_{\mathbf{x}}(v_{\mathbf{x}=\mathbf{i}}) = \Theta_{\mathbf{x}}(w_{\mathbf{x}=\mathbf{j}}) \iff |v| = |w|, \mathbf{i} = \mathbf{j}, \text{ and } v(\mathbf{i}) = w(\mathbf{j}). \quad (18)$$

849 Using this observation, it is straightforward to show:

850 **Corollary 40.** *Let $L \subseteq A^* \otimes \mathbf{x}$ be a language. Then, the following are*
851 *equivalent:*

852 (a) *L is definable by a quantifier-free formula;*

853 (b) *$L = \Theta_{\mathbf{x}}^{-1}(\Theta_{\mathbf{x}}[L])$;*

854 (c) *there is a subset $P \subseteq A_{\varepsilon}^* \otimes \mathbf{x}$ such that $L = \Theta_{\mathbf{x}}^{-1}(P)$.*

855 *In particular, the set of all quantifier-free formulas of $\mathbf{FO}_{\mathbf{x}}[\mathcal{N}]$ forms a com-*
856 *plete and atomic Boolean algebra.*

857 8. An application to Logic on Words

858 In this section we combine Corollary 30 and Remark 31 to prove the
859 equality

$$\mathcal{B}\Pi_1[\mathcal{N}] \cap \text{Reg} = \mathcal{B}\Pi_1[\text{Reg}]. \quad (19)$$

860 The idea is the following. Combining the fact that universal quantification
861 may be seen as an adjoint and our algebraic characterization of quantifier-
862 free formulas we obtain a directed family of adjunctions on $\mathcal{P}(A^*)$ with joint
863 image equal to $\Pi_1[\mathcal{N}]$ allowing us to fit into the setting of Theorem 27. Finally
864 we show that these adjunctions restrict correctly to the regular fragment
865 thus allowing us to apply Corollary 30 and Remark 31, thereby concluding
866 that (19) holds.

867 Let \mathbf{x} be a set of k variables. Universal quantification (as an adjoint) and
868 recognition of quantifier-free formulas are based on the following two maps,
869 respectively

$$A^* \xleftarrow{\pi} A^* \otimes \mathbf{x} \xrightarrow{\Theta_{\mathbf{x}}} A_{\varepsilon}^* \otimes \mathbf{x}$$

Dually this gives rise to

$$\begin{array}{ccc}
 & \forall & (\Theta_{\mathbf{x}}[(\)^c])^c \\
 & \curvearrowright & \curvearrowright \\
 \mathcal{P}(A^*) & \xrightarrow{\pi^{-1}} & \mathcal{P}(A^* \otimes \mathbf{x}) & \xleftarrow{\Theta_{\mathbf{x}}^{-1}} & \mathcal{P}(A_{\varepsilon}^* \otimes \mathbf{x}) \\
 & \curvearrowleft & & \curvearrowleft & \\
 & \exists & & \Theta_{\mathbf{x}}[\] &
 \end{array}$$

870 In particular, we have a (correct) composition of adjunctions as follows

$$\begin{array}{ccc}
 & \forall & \\
 & \curvearrowright & \\
 \mathcal{P}(A^*) & \xrightarrow{\pi^{-1}} & \mathcal{P}(A^* \otimes \mathbf{x}) & \xleftarrow{\Theta_{\mathbf{x}}^{-1}} & \mathcal{P}(A_{\varepsilon}^* \otimes \mathbf{x}) \\
 & & & \curvearrowleft & \\
 & & & \Theta_{\mathbf{x}}[\] &
 \end{array}$$

871 That is,

$$f_k = \Theta_{\mathbf{x}}[\pi^{-1}(\)]: \mathcal{P}(A^*) \xleftrightarrow{\quad} \mathcal{P}(A_{\varepsilon}^* \otimes \mathbf{x}): \forall(\Theta_{\mathbf{x}}^{-1}(\)) = g_k$$

872 is an adjunction and, combining quantification as adjunction with the de-
 873 scription of quantifier-free formulas in Corollary 40(c), we have $L \subseteq A^*$ is in
 874 $\Pi_1^k[\mathcal{N}]$ if and only if $L = \forall(\Theta_{\mathbf{x}}^{-1}(P)) = g_k(P)$ for some $P \subseteq A_{\varepsilon}^* \otimes \mathbf{x}$. That is,
 875 (f_k, g_k) is an adjunction with associated closure operator

$$[L]_k := g_k f_k(L) = \forall \Theta_{\mathbf{x}}^{-1}(\Theta_{\mathbf{x}}[L \otimes \mathbf{x}]),$$

876 and $\text{Im}([\]_k) = \text{Im}(g_k) = \Pi_1^k[\mathcal{N}]$. Notice that, since $[\]_k$ is a closure
 877 operator on $\mathcal{P}(A^*)$, we have

$$L \subseteq [K]_k \iff [L]_k \subseteq [K]_k,$$

878 for every $K, L \subseteq A^*$. Therefore, $[L]_k$ may be seen as the best over-approximation
 879 of L by a language definable in $\Pi_1^k[\mathcal{N}]$. Accordingly, we are in the situation
 880 of Theorem 27 with

$$B = \mathcal{P}(A^*) \text{ and } D = \bigcup_{k \in \mathbb{N}} \text{Im}(g_k) = \bigcup_{k \in \mathbb{N}} \Pi_1^k[\mathcal{N}] = \Pi_1[\mathcal{N}].$$

881 We now aim to apply Corollary 30 with $B' = \mathbf{Reg}$, the Boolean algebra of all
 882 regular languages over A . This is possible given the following fact.

883 **Lemma 41.** *For each $k \in \mathbb{N}$, if $L \subseteq A^*$ is regular, then so is $\lceil L \rceil_k$.*

884 *Proof.* Fix $k \in \mathbb{N}$ and suppose $L \subseteq A^*$ is regular. We proceed through the
885 four maps whose composition defines $\lceil \cdot \rceil_k$.

886 **Claim 1.** $L \otimes \mathbf{x}$ is regular.

887 Note that if $\mu: A^* \rightarrow M$ is a finite monoid recognizing L , then $\pi^*: (A \times$
888 $2^{\mathbf{x}})^* \rightarrow A^*$ composed with μ , where π^* is the homomorphism extending the
889 projection of $A \times 2^{\mathbf{x}}$ onto A , recognizes $L \otimes \mathbf{x}$ once we restrict to structures.
890 As we remarked earlier, the set of all \mathbf{x} -structures is itself regular so it follows
891 from closure under intersection that $L \otimes \mathbf{x}$ is itself regular.

892 **Claim 2.** $\Theta_{\mathbf{x}}[L \otimes \mathbf{x}]$ is regular.

893 This follows from the previous claim and Lemma 33.

894 **Claim 3.** $\Theta_{\mathbf{x}}^{-1}(\Theta_{\mathbf{x}}[L \otimes \mathbf{x}])$ is regular.

895 This is immediate as the inverse image with respect to a homomorphism
896 between free monoids of a regular language is always regular: If $\Theta_{\mathbf{x}}[L \otimes \mathbf{x}]$ is
897 recognized by $f': (A_{\varepsilon} \times 2^{\mathbf{x}})^* \rightarrow M'$ then the composition $f' \circ \Theta_{\mathbf{x}}: (A \times 2^{\mathbf{x}})^* \rightarrow$
898 M' recognizes $\Theta_{\mathbf{x}}^{-1}(\Theta_{\mathbf{x}}[L \otimes \mathbf{x}])$.

899 **Claim 4.** $\forall(\Theta_{\mathbf{x}}^{-1}(\Theta_{\mathbf{x}}[L \otimes \mathbf{x}]))$ is regular.

900 As observed in Section 7, the upper adjoint \forall is given by $K \mapsto (\pi[K^c])^c$
901 where $\pi: A^* \otimes \mathbf{x} \rightarrow A^*$ is the restriction of $\pi^*: (A \times 2^{\mathbf{x}})^* \rightarrow A^*$ to structures.
902 It follows that $\pi[K^c] = \pi^*[K^c \cap (A^* \otimes \mathbf{x})]$. Now, since $K = \Theta_{\mathbf{x}}^{-1}(\Theta_{\mathbf{x}}[L \otimes \mathbf{x}])$ is
903 regular, K^c is also regular and $K^c \cap (A^* \otimes \mathbf{x})$ is regular. Further, it follows by
904 Lemma 33 that $\pi^*[(\Theta_{\mathbf{x}}^{-1}(\Theta_{\mathbf{x}}[L \otimes \mathbf{x}]))^c \cap A \otimes \mathbf{x}]$ is regular. Finally, we conclude
905 that its complement $\forall(\Theta_{\mathbf{x}}^{-1}(\Theta_{\mathbf{x}}[L \otimes \mathbf{x}])) = (\pi^*[(\Theta_{\mathbf{x}}^{-1}(\Theta_{\mathbf{x}}[L \otimes \mathbf{x}]))^c \cap A \otimes \mathbf{x}])^c$
906 is regular as required. \square

907 As a consequence, Corollary 30 applies and we obtain:

908 **Corollary 42.** *Considering each of the following Booleanizations as subal-*
909 *gebras of $\mathcal{P}(A^*)$, we have*

$$(\Pi_1[\mathcal{M}] \cap \text{Reg})^- = (\Pi_1[\mathcal{M}])^- \cap \text{Reg}.$$

910 Finally, applying Remark 31 in this particular case, we see that

$$\Pi_1[\mathcal{M}] \cap \text{Reg} = \bigcup_{k \in \mathbb{N}} g_k f_k[\text{Reg}].$$

911 The languages in $g_k f_k[\text{Reg}]$ are exactly the languages $\lceil L \rceil_k$ for L regular.
912 By the proof of Lemma 41, we have that $\lceil L \rceil_k = \forall(\Theta_{\mathbf{x}}^{-1}(P))$ where $P =$

913 $\Theta_{\mathbf{x}}[L \otimes \mathbf{x}] \subseteq (A_\varepsilon \times 2^{\mathbf{x}})^*$, and, by Claim 2 in particular, we have that P is
 914 regular. That is, $[L]_k = L_{\forall \mathbf{x} \varphi(\mathbf{x})}$ where the atomic formula $\varphi(\mathbf{x})$ is regular,
 915 or equivalently, $[L]_k \in \Pi_1[\mathbf{Reg}]$.

916 On the other hand if $\varphi(\mathbf{x})$ is an atomic formula that is regular, then the
 917 arguments in Claims 3 and 4 show that $L_{\forall \mathbf{x} \varphi(\mathbf{x})}$ is regular. Thus

$$g_k f_k[\mathbf{Reg}] = \Pi_1^k[\mathbf{Reg}]$$

918 and we have

$$(\Pi_1[\mathbf{Reg}])^- = \left(\bigcup_{k \in \mathbb{N}} \Pi_1^k[\mathbf{Reg}] \right)^- = (\Pi_1[\mathcal{M}] \cap \mathbf{Reg})^- = (\Pi_1[\mathcal{M}])^- \cap \mathbf{Reg}.$$

919 We thus obtain the desired result:

920 **Theorem 43.** *The following equality holds:*

$$\mathcal{B}\Pi_1[\mathbf{Reg}] = \mathcal{B}\Pi_1[\mathcal{M}] \cap \mathbf{Reg}.$$

921 9. Final remarks

922 Can the techniques of this paper be pushed further? A few possible
 923 directions suggest themselves. As we mentioned earlier, Theorem 43 is not
 924 new, but the techniques of our proof are completely different from what was
 925 used before.

926 In [30], a more general result is proved, concerning *modular quantifiers*:
 927 If \mathbf{x} is a set of k variables and $0 \leq j < m$, $t > 0$, then we allow quantified
 928 formulas of the form

$$\exists^{(j,m,t)}_{\mathbf{x}} \varphi(\mathbf{x}).$$

929 Such formulas are interpreted as follows: If $w \in A^*$, then $w \models \exists^{(j,m,t)}_{\mathbf{x}} \varphi(\mathbf{x})$
 930 if the number of k -tuples \mathbf{i} such that

$$(w, \mathbf{i}) \models \varphi(\mathbf{x})$$

931 is both congruent to j modulo m and at least t . Observe that with $m = t = 1$,
 932 this is just ordinary existential quantification. If $\varphi(\mathbf{x})$ is quantifier-free, then
 933 $\exists^{(j,m,t)}_{\mathbf{x}} \varphi(\mathbf{x})$ is called a *generalized* Σ_1 -sentence. The main result of [30]
 934 is that the analogue of Theorem 43 holds for Boolean combinations of such

935 generalized Σ_1 -sentences, as well as for Boolean combinations of ordinary
 936 Σ_1 -sentences. It would be interesting to know whether the approach of the
 937 present paper can be used to give a different proof of this result.

938 What about rising higher in the alternation hierarchy? The identity

$$\mathbf{FO}[\mathcal{N}] \cap \mathbf{Reg} = \mathbf{FO}[\mathbf{Reg}],$$

939 proved in [2], is equivalent to well-known lower bounds results in circuit
 940 complexity, and the only proof known depends on the circuit complexity
 941 results. Yet this equality appears to be saying something fundamental, and
 942 rather simple, about automata and logic: If you can define a regular language
 943 with a sentence that uses unrestricted numerical predicates, then you can
 944 define it using just regular numerical predicates. Can the methods used here
 945 provide a different proof of this fact? One might first try to show that

$$\mathcal{B}\Pi_2[\mathcal{N}] \cap \mathbf{Reg} \subseteq \mathbf{FO}[\mathbf{Reg}].$$

946 It should be cautioned that things are not so neat when we get to Π_2 formulas,
 947 and that this would require the development of new techniques.

948 Combining the two directions of generalization - from ordinary to modular
 949 quantifiers, and from Π_1 formulas to formulas with more levels of quantifier
 950 alternation - one might conjecture

$$\mathbf{FOMOD}[\mathcal{N}] \cap \mathbf{Reg} = \mathbf{FOMOD}[\mathbf{Reg}].$$

951 Here, $\mathbf{FOMOD}[\mathcal{R}]$ represents all the formulas one can build, starting from
 952 numerical predicates in \mathcal{R} , using both modular quantification and Boolean
 953 operations. This conjecture is equivalent to the long-open question in circuit
 954 complexity of whether the complexity class ACC^0 is strictly contained in
 955 NC^1 (see [2] and [29]).

Acknowledgments

In addition to the funding received from the ERC, during the revision process, the first-named author was partially supported by the Center for Mathematics of the University of Coimbra (UID/MAT/00324/2019) and by the Center for Mathematics of the University of Porto (UID/MAT/00144/2019), which are funded by the Portuguese Government through FCT/MCTES and cofunded by the European Regional Development Fund through the Partnership Agreement PT2020.

Finally, we would like to thank the anonymous referee for the careful reading of the paper and for the useful suggestions that helped to improve the presentation of our work. We also want to thank Charles Paperman for fruitful discussions with the two last-named authors.

References

- [1] B. Banaschewski and A. Pultr. Booleanization. *Cahiers de Topologie et Géométrie Différentielle Catégoriques* 37 (1), 41-60, 1996.
- [2] D. Barrington, K. Compton, H. Straubing, and D. Thérien. Regular languages in NC^1 . *Journal of Computer and System Sciences* 44(3):478-499, 1992.
- [3] C. Borlido and M. Gehrke. A note on powers of Boolean spaces with internal semigroups. *ArXiv e-prints*, page arXiv:1811.12339, November 2018.
- [4] O. Carton, D. Perrin, and J.-É. Pin. A survey on difference hierarchies of regular languages. *Logical Methods in Computer Science* 14(1), 2017.
- [5] E. Casanovas and R. Farré. Omitting Types in Incomplete Theories. *The Journal of Symbolic Logic* 61(1), 236–245, 1996.
- [6] C. C. Chen. Free Boolean extensions of distributive lattices. *Nanta Math.*, 1:1–14, 1966/1967.
- [7] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order, 2nd edition*. Cambridge University Press, 2002.
- [8] L. L. Esakia. Topological Kripke models. *Dokl. Akad. Nauk SSSR*, 214:298–301, 1974.
- [9] L. L. Esakia. *Heyting algebras: Duality theory*. “Metsniereba”, Tbilisi, 1985. (in Russian).
- [10] M. Gehrke. Canonical extensions, Esakia spaces, and universal models. In *Leo Esakia on duality in modal and intuitionistic logics*, volume 4 of *Outst. Contrib. Log.*, pages 9–41. Springer, Dordrecht, 2014.

- [11] M. Gehrke, S. Grigorieff, and J.-É. Pin. Duality and equational theory of regular languages. In *Automata, languages and programming. Part II*, volume 5126 of *Lecture Notes in Comput. Sci.*, pages 246–257. Springer, Berlin, 2008.
- [12] M. Gehrke and B. Jónsson. Bounded distributive lattices expansions. *Mathematica Scandinavica*, 94(2):13–45, 2004.
- [13] M. Gehrke and H. A. Priestley. Canonical extensions and completions of posets and lattices. *Rep. Math. Logic*, 43:133–152, 2008.
- [14] S. van Gool and B. Steinberg. Pro-aperiodic monoids and model theory. *Israel Journal of Mathematics*, to appear.
- [15] C. Glaßer, H. Schmitz, and V. Selivanov. Efficient algorithms for membership in Boolean hierarchies of regular languages. *Theoret. Comput. Sci.*, 646:86–108, 2016.
- [16] G. Grätzer and E. T. Schmidt. On the generalized Boolean algebra generated by a distributive lattice. *Indag. Math.*, 20:547–553, 1959.
- [17] F. Hausdorff. *Set theory*. Chelsea Publishing Company, New York, 1957. Translated by John R. Aumann, et al.
- [18] B. Jónsson and A. Tarski. Boolean algebras with operators I. *Amer. J. Math.*, 73:891–939, 1951.
- [19] L. Libkin. *Elements of finite model theory*. Springer, Berlin, 2004.
- [20] A. Maciel, P. Péladéau, and D. Thérien. Programs over semigroups of dot-depth one. *Theoret. Comput. Sci.*, 245(1):135–148, 2000. Semigroups and algebraic engineering (Fukushima, 1997).
- [21] H. M. MacNeille. Extension of a distributive lattice to a Boolean ring. *Bull. Amer. Math. Soc.*, 45(6):452–455, 1939.
- [22] A. Nerode. Some Stone spaces and recursion theory. *Duke Math. J.*, 26:397–406, 1959.
- [23] W. Peremans. Embedding of a distributive lattice into a Boolean algebra. *Indag. Math.*, 60:73–81, 1957.

- [24] T. Place and M. Zeitoun. Going higher in first-order quantifier alternation hierarchies on words. *Journal of the ACM*, 2018.
- [25] H. A. Priestley. Representation of distributive lattices by means of ordered stone spaces. *Bull. London Math. Soc.*, 2:186–190, 1970.
- [26] M. Sipser. *Introduction to the theory of computation, 2nd edition* Thomson, Boston, MA, 2006.
- [27] M. H. Stone. Applications of the theory of Boolean rings to general topology. *Trans. Amer. Math. Soc.*, 41(3):375–481, 1937.
- [28] H. Straubing. Recognizable sets and power sets of finite semigroups. *Semigroup Forum*, 18(4):331–340, 1979.
- [29] H. Straubing. *Finite automata, formal logic, and circuit complexity*. Progress in Theoretical Computer Science. Birkhäuser Boston, Inc., Boston, MA, 1994.
- [30] H. Straubing. Languages defined with modular counting quantifiers. *Information and Computation* 166(2):112–132, 2001.