

Representing Interval Orders by Weighted Bases: some Complexity Results

Pierre Marquis
CRIL, Université d'Artois
marquis@cril.fr

Meltem Öztürk
CRIL, Université d'Artois
ozturk@cril.fr

ABSTRACT

This paper is centered on the notion of interval order as a model for preferences. We introduce a family of representation languages for such orders, parameterized by a scale and an aggregation function. We show how interval orders can be represented by elements of those languages, called weighted bases. We identify the complexity of the main decision problems.

1. INTRODUCTION

Dealing with preferences is an important issue in many fields, like economics, decision theory, artificial intelligence and data bases.

In general, a preference relation is defined as a complete preorder (i.e., a reflexive and transitive relation) over the set of alternatives. However, such a model does not prove adequate to all situations, and other models (generalizing the complete preorder one) have been pointed out. In particular, a well-known problem with the complete preorder structure for preference is that the associated indifference relation is necessarily transitive, and it is known that such a property may be violated in the presence of thresholds as shown in the famous example of Luce [13] with a sweetened cup of coffee.

Among other relational structures like quasi-orders, tolerance orders, split semiorders, etc. (for more details, see e.g. [9]), interval orders have been introduced for handling such scenarios. Indeed, in contrast to the associated strict preference relation, the indifference relation induced by an interval order is not necessarily transitive. In their classical numerical representation an interval (with a uniform length in the case of semiorders) is associated to each alternative and each alternative is said to be preferred to another one if and only if its associated interval is completely to the right of the other's interval. This is what we call an interval representation of an interval order in the following.

Unfortunately, when the set of alternatives is huge, neither the explicit representation of the interval order (i.e.,

its representation by pairs) nor its interval representations are feasible; especially when the set of alternatives has a combinatorial structure, i.e., it is a cartesian product of a finite domain for each one of a set of variables. Hence, it is important to define more compact representation languages for interval orders and to evaluate them. For this purpose, the following three criteria are of great value:

- **Simplicity and modularity:** a representation language is expected to have a simple (yet formal) semantics expressing the connection between the representation and the corresponding (explicit) preference relation. Modularity is the ability to specify the preference relation within the representation language in a piecewise way.
- **Complexity issues:** they indicate the computational effort which must be spent to achieve a number of treatments of interest on the preference relation represented in the chosen language. Some basic queries consist in determining whether a given alternative is preferred or not to another and determining whether a given alternative is undominated. The focus is typically laid on worst case scenarios.
- **Expressiveness and spatial efficiency:** expressiveness gives the aptitude of a representation language to encode a family of preference relations (total order, preorder, partial order, etc.), while spatial efficiency gives its aptitude to do it using few space (it refines expressiveness)

Since the complexity of any algorithm is a function relating its input size with the amount of resources (time or space) needed to achieve the computation, the complexity results for treatments based on a given representation language must be interpreted in light of its spatial efficiency.

The first criterion (simplicity and modularity) is fundamental in the perspective of preference elicitation and other human-computer interactions about the preference relation. It shows the perspective of preference representation more general than preference compilation, which mainly consists in turning a given preference representation into another one, so as to optimize at least one of the last two criteria.

While much effort has been devoted to the representation issue for utility functions or preorders (complete or partial) for the last years (see among others [7, 1, 2, 3, 10, 12, 4]), the compact representation of interval orders has not been addressed so far (as far as we know).

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires a fee and/or special permission from the publisher, ACM.

VLDB '07, September 23-28, 2007, Vienna, Austria.
Copyright 2007 VLDB Endowment, ACM 978-1-59593-649-3/07/09.

In this paper, we contribute to fill this gap by showing how interval orders can be compactly encoded by weighted bases. The paper is organized as follows. Some formal preliminaries are given in Section 2. How interval orders can be represented in a simple, modular way by weighted bases is shown in Section 3. The expressiveness and spatial efficiency issues are addressed in Section 4. Complexity results are provided in Section 5. Section 6 concludes the paper and gives some perspectives for further research.

2. FORMAL PRELIMINARIES

2.1 Propositional logic

We consider a propositional language $PROP_{PS}$ generated in the usual way from a finite set PS of propositional atoms and the connectives $\wedge, \vee, \Rightarrow, \Leftrightarrow$, and the Boolean constants *true* (verum), *false* (falsum). A world over PS is a total function (i.e., a mapping) ω from PS to $\{0, 1\}$. The set of all worlds is denoted by 2^{PS} . The notion of satisfaction is defined in the standard truth functional way. When a world ω satisfies a formula ϕ , we write $\omega \models \phi$ and say that ω is a model of ϕ . $Mod(\phi)$ denotes the set of all models of ϕ . Inference is defined as model containment, à la Tarski. \models denotes logical entailment and \equiv logical equivalence. For every formula Σ from $PROP_{PS}$, $Var(\Sigma)$ denotes the set of all atoms from PS occurring in Σ .

2.2 Binary relations, scales and intervals

Let A be a set and $R \subseteq A \times A$ be a relation over A :

- The relation I_R over A defined by $\forall a_1, a_2 \in A, (a_1, a_2) \in I_R$ if and only if $(a_1, a_2) \in R$ and $(a_2, a_1) \in R$ is called the symmetric part of R .
- The relation P_R over A defined by $\forall a_1, a_2 \in A, (a_1, a_2) \in P_R$ if and only if $(a_1, a_2) \in R$ and $(a_2, a_1) \notin R$ is called the asymmetric part of R .

By construction, I_R is a symmetric relation and P_R an asymmetric one. If R is interpreted as a relation such that $(a_1, a_2) \in R$ if and only if a_1 is at least as preferred as a_2 , then I_R is the associated indifference relation ($(a_1, a_2) \in I_R$ if and only if a_1 and a_2 are equally preferred) and P_R is the associated strict preference relation ($(a_1, a_2) \in P_R$ if and only if a_1 is strictly preferred to a_2).

A scale S is a totally ordered set with a least element \perp and a greatest element \top ($\top \neq \perp$). \leq denotes the corresponding order. Let $=$ denote the identity relation over S . We denote by $<, >, \geq$ the binary relations over S given respectively by $s_1 < s_2$ if and only if $s_1 \leq s_2$ and not $(s_1 = s_2)$, $s_1 > s_2$ if and only if not $(s_1 \leq s_2)$ and $s_1 \geq s_2$ if and only if not $(s_1 < s_2)$, whatever the elements s_1 and s_2 of S .

The set of all intervals over S is Int_S given by

$$Int_S = \{[s_1, s_2], [s_1, s_2), (s_1, s_2], (s_1, s_2) \mid s_1, s_2 \in S\},$$

Observe that such intervals can be equal to the empty set (especially when $s_1 > s_2$).

(Int_S, \subseteq) is a lattice, \emptyset is the least element of Int_S and $S = [\perp, \top]$ is the greatest one. For any non-empty interval i of Int_S of the form $i = [s_1, s_2], [s_1, s_2), (s_1, s_2],$ or (s_1, s_2) , we note $l(i) = s_1$ and $u(i) = s_2$. $lc(i)$ (resp. $rc(i)$) means that i is left-closed (resp. right-closed). Let $>_{Int_S}, \sim_{Int_S}$ and \geq_{Int_S} be the relations from $Int_S \times Int_S$ s.t. for every pair of non-empty intervals i_1 and i_2 , we have:

- $i_1 >_{Int_S} i_2$ if and only if
 - $l(i_1) \geq u(i_2)$ if not $rc(i_2)$,
 - $l(i_1) > u(i_2)$ otherwise;
- $i_1 \sim_{Int_S} i_2$ if and only if $i_1 \cap i_2 \neq \emptyset$;
- $i_1 \geq_{Int_S} i_2$ if and only if $i_1 >_{Int_S} i_2$ or $i_1 \sim_{Int_S} i_2$.

Let S be a scale. Given an element s_1 of S , each relational operator $r \in \{=, \leq, \geq, <, >\}$ characterizes the interval $r(s_1) = \{s_2 \in S \mid s_2 r s_1\}$ of Int_S . Each set $r(s_1)$ consists of the set of all elements from S which are located between two elements of S among s_1, \top, \perp , one of them being s_1 :

- $=(s_1) = [s_1, s_1] = \{(s_1, s_1)\}$;
- $\leq(s_1) = [\perp, s_1]$; $\geq(s_1) = [s_1, \top]$;
- $<(s_1) = [\perp, s_1)$; $>(s_1) = (s_1, \top]$.

3. REPRESENTING INTERVAL ORDERS

3.1 Interval orders

Let us first give a formal definition of interval orders:

DEFINITION 1. An interval order $R \subseteq A \times A$ is a relation which is:

- Complete: $\forall a_1, a_2 \in A, (a_1, a_2) \in R$ or $(a_2, a_1) \in R$.
- Ferrers: $\forall a_1, a_2, a_3, a_4 \in A$, if $(a_1, a_2) \in R$ and $(a_3, a_4) \in R$, then $(a_1, a_4) \in R$ or $(a_3, a_2) \in R$.

Interval orders are not necessarily transitive relations (and this is where they depart from preorders). Nevertheless, P_R associated to an interval order R is transitive and Ferrers.

EXAMPLE 1. Let $A = \{a, b, c, d\}$ and let $R = \{(a, a), (a, b), (a, c), (a, d), (b, a), (b, b), (b, c), (b, d), (c, b), (c, c), (c, d), (d, c), (d, d)\}$. R is an interval order over A (but not a preorder over A since e.g., we have $(c, b) \in R$ and $(b, a) \in R$ but $(c, a) \notin R$).

The following representation theorem due to Fishburn [8] shows that any interval order over a countable set can be characterized by intervals over \mathbb{R} :

PROPOSITION 1. [8] Let A be a countable set. A binary relation $R \subseteq A \times A$ is an interval order if and only if there exist a mapping g from A to \mathbb{R} and a mapping q from \mathbb{R} to \mathbb{R}^+ such that for any $a_1, a_2 \in A$, we have

$$(a_1, a_2) \in R \text{ if and only if } g(a_1) + q(g(a_1)) \geq g(a_2).$$

Indeed, the mappings g and q come down to associating to any element a of A an interval $[l_a = g(a), u_a = g(a) + q(g(a))]$ over \mathbb{R} such that for any $a_1, a_2 \in A$, we have:

- $(a_1, a_2) \in P_R$ if and only if $l_{a_1} > u_{a_2}$,
- $(a_1, a_2) \in I_R$ if and only if $l_{a_1} \leq u_{a_2}$ and $l_{a_2} \leq u_{a_1}$.

EXAMPLE 2. Let g and q such that: $g(a) = 6, g(b) = 4, g(c) = 2, g(d) = 0$, and $q(x) = 3$ for every $x \in \mathbb{R}$. (g, q) represents the relation R given in the previous example.

When A is finite, Fishburn's theorem (Proposition 1) can be specialized so that closed intervals of \mathbb{N} included in $[0, 2 \times Card(A) - 1]$ are enough.

3.2 Weighted bases

We are now ready to define the notion of weighted base, at the syntax level first and then from a semantical point of view:

DEFINITION 2. Let PS be a finite set of atomic propositions and S be a scale. A weighted base B over PS and S is a finite multiset B of $(\phi, r_1, s_1, r_2, s_2)$ where ϕ is a formula of $PROP_{PS}$, r_1 (resp. r_2) is a relational operator from $\{=, >, \geq\}$ (resp. $\{=, <, \leq\}$) and s_1, s_2 are elements of S .

EXAMPLE 3. Let $PS = \{x, y, z\}$ and $S = [\perp = 0, \top = 100] \subseteq \mathbb{R}$ naturally ordered. $B = \{(\neg x, \geq, 30, \leq, 80), (\neg z, \geq, 20, \leq, 60), (x \wedge \neg y, \geq, 20, \leq, 40), (x \wedge z, \geq, 50, \leq, 100)\}$ is a weighted base over PS and S .

Intuitively, $(\phi, r_1, s_1, r_2, s_2)$ must be viewed as pieces of preferential evidence, corresponding to imprecise evaluations of the alternatives under consideration w.r.t. different criteria or w.r.t. different sources.

REMARK 1. For sake of notation simplicity, when one of the bound s_1 or s_2 of the interval associated to a formula is also the bound of the scale S , a triple (ϕ, r, s) may be used instead of $(\phi, r_1, s_1, r_2, s_2)$. For example the last 5-uple of the previous example may be written as $(x \wedge z, \geq, 50)$ since 100 is one of the bounds of $S = [0, 100]$.

Putting the criteria or sources together calls for aggregation functions:

DEFINITION 3. Let f be any mapping from $Int_S \times Int_S$ to Int_S which is associative, commutative, has a neutral element $n_f \in Int_S$. f is called an S -aggregation function. It is extended to a mapping from any finite subset of Int_S to Int_S , also referred to as f , and defined inductively by:

- $f(\emptyset) = n_f$, $f(\{i\}) = i$,
- $f(\{i_1, i_2, \dots, i_n\}) = f(i_1, f(\{i_2, \dots, i_n\}))$.

We are now ready to define the notion of satisfaction of a weighted base by an interpretation. $I(\omega)$ is the interval where the actual (qualitative) utility of world ω lies.

DEFINITION 4. Let PS be a finite set of atomic propositions and let S be a scale. Let I be an interpretation over PS and S , i.e., a mapping from 2^{PS} to Int_S . Let f be an S -aggregation function. I f -satisfies a weighted base B over PS and S , noted $I \models_f B$ if and only if $\forall \omega \in 2^{PS}$, we have

$$I(\omega) \subseteq f(\{r_1(s_1) \cap r_2(s_2) \mid (\phi, r_1, s_1, r_2, s_2) \in B \text{ and } \omega \models \phi\}).$$

The neutral element of the S -aggregation function will be used as the interpretation of a word which satisfies none of the formulae of the given weighted base.

Among the possible S -aggregation functions f for any scale S are \cap , whose neutral element n_\cap is $S = [\perp, \top]$ and $+$ which is such that for any $i_1, i_2 \in Int_S$, $i_1 + i_2$ is the smallest interval (w.r.t. \subseteq) containing both i_1 and i_2 . Clearly enough, the neutral element n_+ is the empty interval \emptyset . It is obvious to show that \cap is the infimum function for the lattice (Int_S, \subseteq) , and that $+$ is the supremum function for this lattice.

Simple S -aggregation functions are min and max : when none of i_1 and i_2 is the empty interval, $l(i_1 \min i_2)$ is the

minimum of $l(i_1)$ and $l(i_2)$, $u(i_1 \min i_2)$ is the minimum of $u(i_1)$ and $u(i_2)$. Otherwise, $i_1 \min i_2$ is the empty interval (i.e., \emptyset is an absorbing element). We have $n_{min} = [\top, \top]$.

When none of i_1 and i_2 is the empty interval, $l(i_1 \max i_2)$ is the maximum of $l(i_1)$ and $l(i_2)$, $u(i_1 \max i_2)$ is the maximum of $u(i_1)$ and $u(i_2)$. Otherwise, $i_1 \max i_2$ is the empty interval. We have $n_{max} = [\perp, \perp]$.

Another simple S -aggregation function (containing 0) is sum : when none of i_1 and i_2 is the empty interval, $l(i_1 \text{ sum } i_2)$ is the sum of $l(i_1)$ and $l(i_2)$, $u(i_1 \text{ sum } i_2)$ is the sum of $u(i_1)$ and $u(i_2)$, $lc(i_1 \text{ sum } i_2)$ holds iff $lc(i_1)$ and $lc(i_2)$ holds and $rc(i_1 \text{ sum } i_2)$ holds if and only if $rc(i_1)$ and $rc(i_2)$ holds; otherwise, $i_1 \text{ sum } i_2$ is the empty interval. We have $n_{sum} = [0, 0]$. $b\text{-sum}$ (bounded sum) is a variant of sum . It can be used instead of sum as an S -aggregation function when S is a numerical scale containing 0, but not closed under addition at the bounds; for instance, consider the scale $S = [0, 100]$, S is not closed under addition since the sum of $u(i_1)$ and $u(i_2)$ with $i_1, i_2 \in Int_S$ may be greater than 100. In any case where the sum of two values from S would exceed \top by above (resp. \perp by below), their bounded sum is equal to \top (resp. \perp) (for example in the case of $S = [0, 100]$, we have $[0, 60] \text{ b-sum } [30, 55] = [30, 100]$).

For instance, the weighted base $B = \{(sea \wedge summer, \geq, 20), (sea \wedge summer, <=, 30), (\neg sea \wedge summer, \geq, 10), (\neg sea \wedge summer, <=, 25), (sea, \geq, -10), (sea, <=, -5)\}$ can be used to represent preferences concerning the period (during summer or not) and the location (near the sea or not) of future vacations; 5-uples can be concerned here with two criteria: pleasure and cost (for the last two). sum , min and max are (among others) natural ways to aggregate intervals associated to 5-uples in this case.

$B = \{(electric - guitar, \geq, 20), (electric - guitar, <=, 1000), (electric - guitar, \geq, 200), (electric - guitar, <=, 500), (electric - guitar, \geq, 300), (electric - guitar, <=, 700)\}$ collects evidence given by three friends when asking them whether we should buy an electric guitar. Here \cap is a natural aggregation function if the sources are equally reliable (it leads to uncertainty reduction).

The presence of several criteria or several sources explains why a multiset representation of weighted bases is more convenient than a set-based one: it can be the case that two criteria (or sources) evaluate alternatives in the same way; while this has no impact when the aggregation function under consideration is idempotent (like \cap and $+$), this is not the case in the remaining case (just consider sum).

3.3 Canonical interpretations

In general, a weighted base can be associated to many interpretations that f -satisfy it. However, one of them plays a specific role: the f -canonical interpretation.

DEFINITION 5. Let PS be a finite set of atomic propositions and let S be a scale. Let f be any S -aggregation function. The f -canonical interpretation I_B^f associated to the weighted base B over PS and S and f is the interpretation over PS and S given by: for any $\omega \in 2^{PS}$,

$$I_B^f(\omega) = f(\{r_1(s_1) \cap r_2(s_2) \mid (\phi, r_1, s_1, r_2, s_2) \in B \text{ and } \omega \models \phi\}).$$

EXAMPLE 4. Let $PS = \{x, y, z\}$ ordered in this way. Let S be the scale $[0, 100] \subseteq \mathbb{R}$ naturally ordered. $B = \{(\neg x, \geq, 30, \leq, 80), (\neg z, \geq, 20, \leq, 60), (x \wedge \neg y, \geq, 20, \leq, 40), (x \wedge z, \geq, 50, \leq, 100)\}$ is a weighted base over PS and S . Table 1 gives interval representations of I_B^\cap , I_B^+ , I_B^{\max} , I_B^{\min} and $I_B^{b\text{-sum}}$.

| 2^{PS} | I_B^\cap | I_B^+ | I_B^{\max} | I_B^{\min} | I_B^{b-sum} |
|----------|-------------|-----------|--------------|--------------|---------------|
| 000 | [30, 60] | [20, 80] | [30, 80] | [20, 60] | [50, 100] |
| 001 | [30, 80] | [30, 80] | [30, 80] | [30, 80] | [30, 80] |
| 010 | [30, 60] | [20, 80] | [30, 80] | [20, 60] | [50, 100] |
| 011 | [30, 80] | [30, 80] | [30, 80] | [30, 80] | [30, 80] |
| 100 | [20, 40] | [20, 60] | [20, 60] | [20, 40] | [40, 100] |
| 101 | \emptyset | [20, 100] | [50, 100] | [20, 40] | [70, 100] |
| 110 | [20, 60] | [20, 60] | [20, 60] | [20, 60] | [20, 60] |
| 111 | [50, 100] | [50, 100] | [50, 100] | [50, 100] | [50, 100] |

Table 1: Canonical interpretations.

Now, a notion of f -equivalence between weighted bases over PS and S can be easily defined: two bases B_1 and B_2 are said to be f -equivalent when the set of all interpretations f -satisfying B_1 is equal to the set of all interpretations f -satisfying B_2 . It is easy to prove that two weighted bases are f -equivalent whenever they are associated to the same f -canonical interpretation:

PROPOSITION 2. *Let PS be a finite set of atomic propositions and let S be a scale. Let B_1, B_2 be two weighted bases over PS and S . Let f be any S -aggregation function. B_1 and B_2 are f -equivalent if and only if $I_{B_1}^f = I_{B_2}^f$.*

In order to make precise the connection between I_B^f and the corresponding interval order, we still need two notions, namely the notion of f -consistent weighted base and the notion of f -possible world.

Obviously enough, the trivial interpretation I_\emptyset s.t. $\forall \omega \in 2^{PS}, I(\omega) = \emptyset$ f -satisfies any weighted base B . For this reason, the notion of f -consistency of a weighted base cannot be defined as the existence of an interpretation f -satisfying it but has to be a bit more elaborate:

DEFINITION 6. *Let PS be a finite set of atomic propositions and let S be a scale. Let f be any S -aggregation function. A weighted base B over PS and S is said to be f -consistent if and only if there exists an interpretation I over PS and S such that $I \models_f B$ and $I \neq I_\emptyset$.*

EXAMPLE 5. *The base B given in Example 4 is f -consistent for f among $\cap, +, \max, \min, \text{sum}, b\text{-sum}$. The base $\{(x, >, 3), (x, =, 2), (\neg x, <, 1), (\neg x, >, 3)\}$ is not \cap -consistent but is $+$ -consistent.*

It is easy to prove that every weighted base B is $+$ -consistent if and only if B is not the empty multiset and that every weighted base B is \min -consistent, \max -consistent, sum -consistent and $b\text{-sum}$ -consistent except when it contains a 5-uple $(\phi, r_1, s_1, r_2, s_2)$ such that $r_1(s_1) \cap r_2(s_2) = \emptyset$ (i.e., the interval specified by r_1, s_1, r_2, s_2 is empty).

Now, the preimage of \emptyset by I_B^f is a subset of 2^{PS} of worlds which are considered as impossible given B :

DEFINITION 7. *Let PS be a finite set of atomic propositions and let S be a scale. Let B be a weighted base over PS and S . Let f be any S -aggregation function. Let ω be a world from 2^{PS} . ω is said to be f -impossible given B if and only if $I_B^f(\omega) = \emptyset$. ω is said to be f -possible when it is not f -impossible.*

$Possible_f(B)$ denotes the subset of 2^{PS} consisting of all worlds which are f -possible given B . Clearly enough, we have $Possible_f(B) = \emptyset$ if and only if B is not f -consistent if and only if $I_B^f = I_\emptyset$.

EXAMPLE 6. *Let us consider again the base B used in the running example. We have $Possible_\cap(B) = 2^{PS} \setminus \{101\}$.*

f -possible worlds correspond to feasible alternatives and f -consistency ensures that at least some alternative is feasible.

We are now ready to define the preference relations induced by a weighted base and an aggregation function:

DEFINITION 8. *Let PS be a finite set of atomic propositions and let S be a scale. Let f be any S -aggregation function. Let B be a f -consistent weighted base over PS and S . We define the binary relations \succ_B^f, \sim_B^f and $\succeq_B^f \subseteq Possible_f(B) \times Possible_f(B)$ by: let $\omega_1, \omega_2 \in Possible_f(B)$,*

- $\omega_1 \succ_B^f \omega_2$ if and only if $I_B^f(\omega_1) >_{Int_S} I_B^f(\omega_2)$;
- $\omega_1 \sim_B^f \omega_2$ if and only if $I_B^f(\omega_1) \sim_{Int_S} I_B^f(\omega_2)$;
- $\omega_1 \succeq_B^f \omega_2$ if and only if $\omega_1 \succ_B^f \omega_2$ or $\omega_1 \sim_B^f \omega_2$.

The next result is a soundness one: it shows that every weighted base and S -aggregation function actually specifies an interval order over S :

PROPOSITION 3. *Let PS be a finite set of atomic propositions and let S be a scale. Let f be any S -aggregation function. Let B be a f -consistent weighted base over PS and S . \succeq_B^f is an interval order on $Possible_f(B)$. \sim_B^f (resp. \succ_B^f) is its symmetric (resp. asymmetric) part.*

Finally, notice that two f -equivalent weighted bases specify the same interval order but the converse is false. Therefore, we define another, less demanding, notion of equivalence called (f, g) -order-equivalence: if f is an S_1 -aggregation function and g is an S_2 -aggregation function, then a weighted base B_1 over PS and S_1 is said to be (f, g) -order-equivalent to B_2 over PS and S_2 if and only if $\succeq_{B_1}^f = \succeq_{B_2}^g$.

3.4 Simplicity and modularity

Basically, in our approach, 5-uples $(\phi, r_1, s_1, r_2, s_2)$ of B are viewed as pieces of preferential evidence and are aggregated using a suited function f ; the corresponding f -canonical interpretation maps each interpretation (representing an alternative) to an interval; the associated interval order follows in an obvious way once alternatives associated to the empty interval have been removed.

In our opinion, this semantics is simple; of course, simplicity has no formal definition but let us stress that the semantics of weighted propositional formulae (or penalty-based representations) as considered in [11, 5, 12, 4] and used for representing cardinal preferences is very close to the semantics we point out. Indeed, in penalty logic, weighted bases P are sets of pairs of the form (ϕ, α) where $\phi \in PROP_{PS}$ and $\alpha \in \mathbb{R} \cup \{+\infty\}$; α is the penalty to be paid whenever ϕ is not satisfied; associated to each weighted base P is a propositional formula K used to discriminate impossible worlds: the possible worlds are restricted to the models of K . Disutilities α are aggregated in an additive way so

that the disutility of any world $\omega \in Mod(K)$ is given by $d(\omega) = \sum_{(\phi, \alpha) \in P \mid \omega \models \phi} \alpha$.

The main differences between penalty-based representations and our framework are: in penalty-based representations, the purpose is to represent complete preorders (and not interval orders); each piece of preferential evidence allows to map worlds to values from a numerical scale S (disutilities) and not to intervals over any S ; *sum* is the aggregation function (while we consider other functions in our framework); finally, a further propositional formula K is used to discriminate impossible worlds in the penalty-based approach (while we consider as impossible those worlds associated to the empty interval).

The modularity of the representation in our framework is achieved, at the syntax level, by the multiset representation of weighted base (5-uples can be considered in a separate way) and at the semantical level, by the compositionality of the construction of the f -canonical interpretation.

4. EXPRESSIVENESS AND SPATIAL EFFICIENCY

After the soundness result given by Proposition 3, let us first show that a completeness result can be achieved for some specific choices of scales S , whatever the chosen S -aggregation function f . Of specific interest is the scale $S = [0, 2 \times Card(A) - 1] \subseteq \mathbb{N}$.¹ From Fishburn's result, it comes that any interval order over a finite set of alternatives A can be represented by a weighted base in our setting:

PROPOSITION 4. *Let A be a finite set of alternatives and \succeq an interval order on A . For every S -aggregation function f with $S = [0, 2 \times Card(A) - 1] \subseteq \mathbb{N}$, there exists a set of propositional atoms PS and a weighted base B over PS and S such that there is a bijection b between A and $Possible_f(B)$ and for every $a_1, a_2 \in A$, we have $a_1 \succeq a_2$ if and only if $b(a_1) \succeq_B^f b(a_2)$.*

Obviously enough, this proposition can be trivially extended to every case a scale isomorphic to $[0, 2 \times Card(A) - 1] \subseteq \mathbb{N}$ (naturally ordered) is considered. This proposition also shows that the choice of the aggregation function has no impact on the expressiveness issue, as soon as such a scale is considered.

Let us now show that a language of weighted bases for representing interval orders is strictly more expressive than the language of penalty-based representations for cardinal preferences:

PROPOSITION 5. *Every complete preorder on A can be represented in the language of weighted bases using $S = \mathbb{R} \cup \{-\infty, +\infty\}$ as the scale and *sum* as the S -aggregation function. Contrastingly, it is not the case that every interval order can be represented in the language of penalty logic.*

Let us finally turn to the spatial efficiency issue and briefly compare our representation languages using weighted bases

¹Alternatively, we could consider the scale $S = \mathbb{R} \cup \{-\infty, +\infty\}$ where \leq extends the standard ordering on real numbers so that $\perp = -\infty$ is the least element of S and $\top = +\infty$ is the greatest element of S . Our preference to integers instead of reals is motivated by representational and computational issues.

to both the explicit representation and the interval representations or interval orders.

The explicit representation of interval orders given by Definition 1 is just as compact as its representation using intervals over \mathbb{R} . Indeed, from Fishburn's theorem (see Proposition 1), we immediately get that for any pair $(a_1, a_2) \in R$ if and only if $l_{a_1} > u_{a_2}$ or $(l_{a_1} \leq u_{a_2} \text{ and } l_{a_2} \leq u_{a_1})$. Hence from any set S_I of pairs (a, i) representing an interval order R , deriving the corresponding explicit representation of R requires $\mathcal{O}(|S_I|^2)$ time in the worst case (hence a polynomial amount of space). Conversely, the explicit representation of an interval order being given, an interval representation of this order can be computed in polynomial time. For instance Doignon [6] pointed out a polynomial time algorithm for computing the (minimal) interval representation of an interval order.

Contrastingly, the ability offered by propositional logic to represent sets of worlds in a compact way is enough for ensuring that the languages of weighted bases are strictly more compact than both the explicit representation and the interval representations. Indeed, the problem with the explicit representation and the interval representations is that they both require an explicit representation of the alternatives. Assume that $A = 2^{PS}$ for a given set PS of n symbols: the explicit representation of any interval order R over A contains at least 2^n elements; this is also the case for any interval representation of A . However, some interval orders over A can be represented in a much more compact way using a weighted base. For instance, if $PS = \{x_1, \dots, x_n\}$, $B = \{(x_1 \vee \dots \vee x_n, =, \top)\}$ can be used for representing a complete preorder where all the models of $x_1 \vee \dots \vee x_n$ are indifferent but strictly preferred to the remaining world (where all atoms are set to 0) (it is enough to consider *max* as the aggregation function).

5. COMPLEXITY RESULTS

In this section, we investigate the complexity of exploiting weighted bases as interval orders representations. We mainly assume that f is a polytime function. Under this assumption (which holds for any of \cap , $+$, *min*, *max*, *sum* and *b-sum*), it is obvious that for any world $\omega \in 2^{PS}$ and any weighted base B , the value of $I_B^f(\omega)$ can be computed in time polynomial in the size of B plus the size of ω . We also assume that the ordering \leq on any scale S can be decided in polynomial time and that we can decide in polynomial time whether a given interval from Int_S is empty (this clearly depends on S since e.g. $(2, 3)$ is empty when $S = \mathbb{N}$ but not empty when $S = \mathbb{R}$).

The first question to be addressed is whether B is f -consistent.

DEFINITION 9. *f -CONSISTENCY is the decision problem given by*

- **Input:** *A weighted base B over PS and S .*
- **Question:** *Is B f -consistent?*

Since the set of alternatives characterized by a weighted base is not always equal to the set of all worlds, another important question is to determine when a world is among the alternatives:

DEFINITION 10. f -POSSIBILITY is the decision problem given by

- **Input:** A weighted base B over PS and S and a world $\omega \in 2^{PS}$.
- **Question:** Is ω f -possible given B ?

Since many different bases share the same f -canonical interpretation in the general case, it is important to identify the complexity of deciding whether two given bases are f -equivalent; indeed, if two bases are f -equivalent then they represent the same interval order:

DEFINITION 11. f -EQUIVALENCE is the decision problem given by

- **Input:** Two weighted bases B_1 and B_2 over PS and S .
- **Question:** Are B_1 and B_2 f -equivalent?

Similarly, it is important to identify the complexity of deciding whether a first base is (f, g) -order-equivalent to a second base (since, as we have seen, it can be the case that two bases that are not f -equivalent nevertheless represent the same interval order); we assume here that both f and g are polytime aggregation functions (possibly over different scales S_1 and S_2):

DEFINITION 12. (f, g) -ORDER-EQUIVALENCE is the decision problem given by

- **Input:** A weighted base B_1 over PS and S_1 , and a weighted base B_2 over PS and S_2 .
- **Question:** Is B_1 (f, g) -order-equivalent to B_2 ?

The following three decision problems are in some sense natural problems when dealing with compactly represented interval orders (and, as such, have been considered in other papers from the literature dealing with compact representation of preferences, see [12]); let R be an element of $\{\succeq, \succ, \sim\}$:

DEFINITION 13. f -COMPARISON(R) is the decision problem given by

- **Input:** A f -consistent weighted base B over PS and S and two worlds $\omega_1, \omega_2 \in \text{Possible}_f(B)$.
- **Question:** Is $\omega_1 R_B^f \omega_2$ true?

DEFINITION 14. f -NON-DOMINANCE is the decision problem given by:

- **Input:** A f -consistent weighted base B over PS and S and a world $\omega \in \text{Possible}_f(B)$.
- **Question:** Is ω undominated w.r.t. \succ_B^f in the set of all worlds f -possible given B ?

DEFINITION 15. f -CAND-OPT-SAT is the decision problem given by:

- **Input:** A f -consistent weighted base B over PS and S and a formula $\phi \in \text{PROP}_{PS}$.

- **Question:** Does there exist a undominated world ω w.r.t. \succ_B^f in the set of all worlds f -possible given B s.t. $\omega \models \phi$?

We have derived the following results:

PROPOSITION 6.

- f -CONSISTENCY is in NP; it is NP-complete for $f = \cap$ and in P for $f = +, f = \min, f = \max, f = \text{sum}, f = b - \text{sum}$.
- f -POSSIBILITY is in P.
- f -EQUIVALENCE is coNP-complete.
- (f, g) -ORDER-EQUIVALENCE is coNP-complete.
- f -COMPARISON(R) is in P.
- f -NON-DOMINANCE is coNP-complete.
- f -CAND-OPT-SAT is in Σ_2^P . It is both NP-hard and coNP-hard (hence it is not in $\text{NP} \cup \text{coNP}$ unless the polynomial hierarchy collapses).
 - If $\text{card}(S) \geq \text{card}(B)$, f -CAND-OPT-SAT is Θ_2^P -hard.
 - Let P_B^f be the set of potential intervals given B and f defined as the subset of Int_S consisting of all intervals $f(\{r_1(s_1) \cap r_2(s_2) \mid (\phi, r_1, s_1, r_2, s_2) \in B\})$. If a superset P of P_B^f can be computed in polynomial time in the size of B , then f -CAND-OPT-SAT is in Θ_2^P .
 - If B can be turned in polynomial time into a (f, sum) -order-equivalent base B' over PS and S' s.t. every 5-uple $(\phi, r_1, s_1, r_2, s_2)$ satisfies $r_1 = \geq$ and $r_2 = \leq$, and S' is a closed interval of \mathbb{N} containing 0, then sum -CAND-OPT-SAT is in Δ_2^P .

We present in the following the two hardest proofs of these results.

PROOF. • f -CAND-OPT-SAT: Membership to Σ_2^P comes from the following non-deterministic polytime algorithm using an NP-oracle: (1) guess $\omega \in 2^{V^{ar}(B)}$; (2) check that ω is f -possible; (3) check that ω is not dominated w.r.t. \succ_B^f using one call to an NP-oracle; (4) check that $\omega \models \phi$.

NP-hardness comes from the following reduction from the satisfiability problem SAT: let Σ be any CNF formula from PROP_{PS} ; to Σ let us associate in polynomial time the base $B = \{\{\text{true}, =, \top\}\}$ and the formula $\phi = \Sigma$. By construction, Σ is satisfiable if and only if there exists a undominated world ω w.r.t. \succ_B^f in the set of all worlds f -possible given B such that $\omega \models \phi$. Indeed, any model ω of Σ does the job.

coNP-hardness comes from a straightforward polynomial reduction from f -NON-DOMINANCE: ω is not dominated w.r.t. \succ_B^f if and only if there exists a undominated world ω w.r.t. \succ_B^f in the set of all worlds f -possible given B such that $\omega \models \phi$ with $\phi = \omega$.

– When $\text{card}(S) \geq \text{card}(B)$, Θ_2^p -hardness comes from a polynomial reduction from PARITY-SAT [17]: given a sequence ϕ_1, \dots, ϕ_n of formulae from $PROP_{PS}$ such that for all $i \in \{1, \dots, n-1\}$, if ϕ_i is unsatisfiable then ϕ_{i+1} is unsatisfiable, is the maximum index i such that ϕ_i is satisfiable an odd number? $n = 2m$ is assumed even without loss of generality. The reduction is as follows: to $\phi_1, \dots, \phi_n \in PROP_{PS}$, let us associate in polynomial time the base $B = \{(\phi_i \wedge \text{new}_i \wedge \bigwedge_{j=1 \dots n, j \neq i} \neg \text{new}_j, =, s_i) \mid i \in \{1, \dots, n\}\}$ and the formula $\phi = \bigvee_{j=0}^{m-1} \phi_{2j+1}$. By construction, the maximum index i such that ϕ_i is satisfiable is odd if and only if there exists a model ω of ϕ which is not dominated w.r.t. \succ_B^f .

– Let us now show the membership of f -CAND-OPT-SAT to Θ_2^p in a restricted case. Roughly, the approach consists in determining an interval i_{max} of Int_S which is maximal w.r.t. $>_{Int_S}$ and such that there exists a world ω satisfying $I_B^f(\omega) = i_{max}$; this is done using binary search and an NP-oracle; then it is enough to check using one call to an NP-oracle that there exists a model ω of ϕ s.t. $I_B^f(\omega) \sim_{Int_S} i_{max}$.

The difficulty here lies in binary searching since $>_{Int_S}$ is not necessarily complete; especially, it can be the case that i_{max} is not unique. In order to overcome it, we refrain from considering $>_{Int_S}$ directly, but a complete, strict ordering \succ closely related to it. Formally, let \succ be the binary relation over the set of all non-empty intervals from Int_S defined by $i_1 \succ i_2$ if and only if $l(i_1) > l(i_2)$ or $(l(i_1) = l(i_2) \text{ and } ((lc(i_1) \text{ and not } lc(i_2)) \text{ or } u(i_1) > u(i_2) \text{ or } (u(i_1) = u(i_2) \text{ and } rc(i_1) \text{ and not } rc(i_2))))))$. For any non-empty subset E of Int_S consisting of non-empty intervals, $\text{max}(E, \succ)$ is a singleton $\{i_{max}^E\}$. It is obvious that $\forall i_1, i_2 \in E$, if $i_1 >_{Int_S} i_2$ then $i_1 \succ i_2$. We also have the following lemma:

LEMMA 1. *Let E be any non-empty subset of Int_S consisting of non-empty intervals. We have:*

$$\text{max}(E, >_{Int_S}) = \{i \in E \mid i \sim_{Int_S} i_{max}^E\}.$$

Based on this lemma, one can design a polynomial time algorithm SOLVE- f -CAND-OPT-SAT-1 for deciding f -CAND-OPT-SAT (under the requirements given in the proposition), using a logarithmic number of calls to an NP-oracle. It mainly consists in binary searching $\{i_{max}^E\}$ in the set of potential intervals (or the given superset of it); here, $E = \{i \in P \setminus \{\emptyset\} \mid \exists \omega \in 2^{Var(B) \cup Var(\phi)} \text{ s.t. } I_B^f(\omega) = i\}$.

- (1) compute $P = P_B^f$ (or a superset of it) in time polynomial in the size of B and remove from it the empty interval (if present); let O be the resulting (finite and non-empty) set;
- (2) sort the elements of O w.r.t. \succ ;
- (3) if O is a singleton $\{i_{max}\}$, then go to (5);
- (4) consider the interval i of rank $\lfloor \text{card}(O)/2 \rfloor$ in O ; if there exists $\omega \in 2^{Var(B) \cup Var(\phi)}$ such that

$I_B^f(\omega) \in O$ and $I_B^f(\omega) \succ i$ then remove from O every interval j such that $I_B^f(\omega) \succ j$ and resume from (3) else remove from O every interval j such that $j \succ i$ and resume from (3);

(5) if there exists $\omega \in 2^{Var(B) \cup Var(\phi)}$ such that $\omega \models \phi$ and $I_B^f(\omega) \sim_{Int_S} i_{max}$ then RETURN(true) else RETURN(false).

The evaluation of the condition of each conditional in SOLVE- f -CAND-OPT-SAT-1 requires one call to an NP-oracle. The total number of calls to such an oracle is thus upper bounded by $\lceil \log_2 \text{Card}(P) \rceil + 1$; the result then follows immediately since $\text{Card}(P)$ is upper bounded by $p(|B|)$ where p is a polynomial such that $|P| = p(|B|)$.

– Let us now finally show the membership of f -CAND-OPT-SAT to Δ_2^p in a second restricted case. Roughly, the algorithm SOLVE- f -CAND-OPT-SAT-2 used to prove it consists in first computing B' , then in binary searching $m = l(i_{max}^E)$ with $E = \{i \in Int_{S'} \setminus \{\emptyset\} \mid \exists \omega \in 2^{Var(B') \cup Var(\phi)} \text{ s.t. } I_{B'}^g(\omega) = i\}$, in the interval $[0, M = \sum_{(\phi, \geq, s_1, \leq, s_2) \in B'} s_2] \subseteq S'$. Indeed, for any interval $i \in Int_{S'}$, we have that $i \sim_{Int_{S'}} i_{max}^E$ if and only if $m \in i$.

(1) $l = 0$ and $u = M$;

(2) while $l \neq u$ do

$m = \lfloor (l + u)/2 \rfloor$;

if there exists $\omega \in 2^{Var(B') \cup Var(\phi)}$ s.t. $I_{B'}^g(\omega) \neq \emptyset$ and $l(I_{B'}^g(\omega)) > m$

then $l = m + 1$

else if there exists $\omega \in 2^{Var(B') \cup Var(\phi)}$ s.t.

$m \in I_{B'}^g(\omega)$

then $l = u = m$

else $u = m$;

(3) if there exists $\omega \in 2^{Var(B') \cup Var(\phi)}$ such that $\omega \models \phi$ and $m \in I_{B'}^g(\omega)$ then RETURN(true) else RETURN(false).

The evaluation of the condition of each conditional in SOLVE- f -CAND-OPT-SAT-2 requires one call to an NP-oracle. The total number of calls to such an oracle is thus upper bounded by $2 \times \lceil \log_2 M + 1 \rceil + 1$. Finally, the number of bits in the binary representation of any s_2 in B' is upper bounded by $p(|B|)$ where p is a polynomial such that $|B'| = p(|B|)$; similarly, the cardinal of B' is upper bounded by $p(|B|)$, and as a consequence the number of bits in the binary representation of M is upper bounded by $2 \times p(|B|)$; it comes that the value of M is upper bounded by $2^{2 \times p(|B|)}$; thus, the total number of calls to an NP-oracle in SOLVE- f -CAND-OPT-SAT-2 is upper bounded by $2 \times \lceil \log_2 (2^{2 \times p(|B|)}) + 1 \rceil + 1 = 2 \times \lceil 2 \times p(|B|) + 1 \rceil + 1$, hence by a polynomial in the input size, and the result follows.

□

To conclude with the complexity results, observe that the set $\{\{s_1, s_2\}, [s_1, s_2], (s_1, s_2], (s_1, s_2) \mid ((\phi_1, r_1, s_1) \in B \text{ or } s_1 \in \{\perp, \top\}) \text{ and } ((\phi_2, r_2, s_2) \in B \text{ or } s_2 \in \{\perp, \top\})\}$ is a superset of both P_B^{\cap} , P_B^+ , P_B^{min} and P_B^{max} and it can be

computed in time polynomial in the size of B ; this shows that \cap -CAND-OPT-SAT, $+$ -CAND-OPT-SAT, min -CAND-OPT-SAT and max -CAND-OPT-SAT are in Θ_2^P .

In light of our results, it turns out that several key decision problems when dealing with preferences (f -COMPARISON(R), f -POSSIBILITY) remain tractable when interval orders are represented by weighted bases, while some other key decision problems (f -NON-DOMINANCE, f -EQUIVALENCE, (f, g) -ORDER-EQUIVALENCE, f -CONSISTENCY) become “mildly” har-

der than the corresponding problems based on explicit or interval representations (which are tractable). This appears as the price to be paid for the gain in spatial efficiency offered by our representation languages. We do not know whether f -CAND-OPT-SAT is at the first level of the polynomial hierarchy or at the second level in the general case but our results show it at the first level of the polynomial hierarchy for many interesting cases.

Finally, our results show that the additional expressive power offered by our approach does not lead to a complexity shift compared to the penalty-based approach to complete preorders representation. Indeed, the complexity results for f -POSSIBILITY, f -COMPARISON(R) and f -NON-DOMINANCE coincide with the complexity results for the corresponding problems reported in [12]. In contrast to our framework, the complexity of CAND-OPT-SAT in the penalty-based approach is in Δ_2^P while this is not ensured in our approach; the difficulty comes from the fact that not only the scale is not necessarily numerical in our setting, but it is not part of the input of the decision problem (hence there is no way to compute a notion of “mean” value, which is required for binary searching).

6. CONCLUSION

In this paper, we have shown how interval orders can be encoded as weighted bases, subject, in our knowledge, that has not been studied before. Our presentation is simple and general in the sense that aggregation functions are not fixed but just have to satisfy some basic properties. Among other things, we have shown that all interval orders can be represented using some weighted bases and we have identified the complexity of a number of decision problems pertaining to the exploitation of compactly represented interval orders.

This work calls for a number of perspectives. One of them consists in designing compact representations for other preference relations, including partially ordered intervals and PQI interval orders. PQI interval orders are preference structures with three relations P (strict preference relation), Q (weak preference relation) and I (indifference). They have been introduced and characterized by Tsoukiàs and Vincke [15, 16]. PQI interval orders have an interval representation: P holds when one interval is completely to the right of the other one, Q holds when two intervals have a non-empty intersection without inclusion and finally I holds in case of inclusion. Concerning their numerical representation Ngo The and Tsoukiàs ([14]) have given two algorithms, the first one in $\mathcal{O}(n^2)$ determines a general representation of these structures and the second one in $\mathcal{O}(n)$ minimizes the first one. Our weighted base representation appears appropriate for these structures.

A second perspective consists in investigating further the expressiveness and spatial efficiency issues. Indeed, the choice of S and f has a clear impact on both issues in the general

case. It turns out that some spatial efficiency results from [5] for complete preorders can be easily extended to the case of interval orders (for instance, when a numerical scale is considered, the inability of compactly encoding some exponentially long P_R -chains using min or max can be exploited to show the corresponding languages less succinct than those for which sum is used). It would prove valuable to determine the expressiveness landscape and the spatial efficiency landscape for various choices of the parameters S and f .

7. REFERENCES

- [1] F. Bacchus and A.J. Grove. Utility independence in a qualitative decision theory. In *Proc. of KR'96*, pages 542–552, 1996.
- [2] S. Bistarelli, H. Fargier, U. Montanari, F. Rossi, Th. Schiex, and G. Verfaillie. Semiring-based CSPs and valued CSPs: Framework, properties and comparison. *Constraints*, 4(3):199–240, 1999.
- [3] C. Boutilier, R. Brafman, C. Domshlak, H. Hoos, and D. Poole. CP-nets: a tool for representing and reasoning with conditional ceteris paribus statements. *Journal of Artificial Intelligence Research*, 21:135–191, 2004.
- [4] Y. Chevaleyre, U. Endriss, and J. Lang. Expressive power of weighted propositional formulas for cardinal preference modeling. In *Proc. of KR'06*, pages 145–152, 2006.
- [5] S. Coste-Marquis, J. Lang, P. Liberatore, and P. Marquis. Expressive power and succinctness of propositional languages for preference representation. In *Proc. of KR'04*, pages 203–212, 2004.
- [6] J.P. Doignon. Sur les représentations minimales des semiordres et des ordres d'intervalles. *Mathématiques et Sciences Humaines*, 101:49–59, 1988.
- [7] J. Doyle and M.P. Wellman. Preferential semantics for goals. In *Proc. of AAAI'91*, pages 698–703, 1991.
- [8] P.C. Fishburn. *Interval orders and interval graphs*. John Wiley and Sons, 1985.
- [9] P.C. Fishburn. Generalisations of semiorders: a review note. *Journal of Mathematical Psychology*, 41:357–366, 1997.
- [10] Ch. Gonzales and P. Perny. GAI networks for utility elicitation. In *Proc. of KR'04*, pages 224–234, 2004.
- [11] C. Lafage and J. Lang. Logical representation of preferences for group decision making. In *Proc. of KR'00*, pages 457–468, 2000.
- [12] J. Lang. Logical preference representation and combinatorial vote. *Annals of Mathematics and Artificial Intelligence*, 42(1-3):37–71, 2004.
- [13] R.D. Luce. Semi-orders and a theory of utility discrimination. *Econometrica*, 24, 1956.
- [14] A. Ngo The and A. Tsoukiàs. Numerical representation of PQI interval orders. *Discrete Applied Mathematics*, 147:125–146, 2005.
- [15] A. Tsoukiàs and Ph. Vincke. Double threshold orders: A new axiomatization. *Journal of Multi-criteria Decision Analysis*, 7:285–301, 1998.
- [16] A. Tsoukiàs and Ph. Vincke. A characterization of PQI interval orders. *Discrete Applied Mathematics*, 127(2):387–397, 2003.
- [17] K.W. Wagner. Bounded query classes. *SIAM Journal on Computing*, 19(5):833–846, 1990.