Weighted abstract dialectical frameworks through the lens of approximation fixpoint theory

Bart Bogaerts
KU Leuven, Department of Computer Science, Celestijnenlaan 200A, 3001 Heverlee, Belgium
Vrije Universiteit Brussel (VUB), Department of Computer Science, Pleinlaan 2, 1050 Brussels, Belgium
bart.bogaerts@vub.be

Weighted abstract dialectical frameworks (wADFs) were recently introduced, extending abstract dialectical frameworks to incorporate degrees of acceptance. In this paper, we propose a different view on wADFs: we develop semantics for wADFs based on approximation fixpoint theory, an abstract algebraic theory designed to capture semantics of various non-monotonic reasoning formalisms. Our formalism deviates from the original definition on some basic assumptions, the most fundamental is that we assume an ordering on acceptance degrees. We discuss the impact of the differences, the relationship between the two versions of the formalism, and the advantages each of the approaches offers. We furthermore study complexity of various semantics.

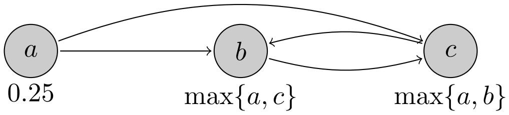
1 Introduction

Abstract argumentation frameworks (AFs) [18] are simple and abstract systems to deal with contentious information and draw conclusions from it. An AF is a directed graph where the nodes are arguments and the edges encode a notion of attack between arguments. In spite of their conceptual simplicity, there exist many different semantics of AFs.

Abstract dialectical frameworks (ADFs) [6, 8] constitute a generalization of AFs in which not only attack, but also support, joint attack and joint support can be expressed.

Recently, ADFs were further generalized into *weighted abstract dialectical frameworks* [10, 9]. These frameworks allow for a fine-grained distinction between degrees of acceptance of arguments. In a partial interpretation of an ADF, each argument is either *accepted*, *rejected* or *unknown*;[1] in wADFs, on the other hand, one associates with each argument a value from an arbitrary set of acceptance values. For instance, arguments can take values from the unit interval, where 1 means acceptance, 0 rejection and numbers in between mean it partially is accepted with a certain strength.

Example 1.1. Consider three arguments a, b and c and assume acceptance values in the unit interval. Partial interpretations assign to each argument either a value in the unit interval, or $\mathbf{u}$ (unknown). Assume that the strength of the acceptance value of a is 0.25, independently of b and c. Furthermore, assume a and b support c and that a and c support b where the strength of the acceptance for b is $\max\{a, c\}$ and the strength of the acceptance for c is $\max\{a, b\}$. This situation can be represented as a weighted ADF (depicted below) and various semantics are defined for it (some of them are compactenumd below).



- The *models* of this wADF are all assignments I such that $I(a) = 0.25$ and $I(b) = I(c) \in [0.25, 1]$.
- The *grounded (partial) interpretation* of this wADF assigns 0.25 to a and $\mathbf{u}$ to b and c.
- *Stable models* are defined with respect to a subset $W \subseteq V$. In this case, a model I is stable if $I(b) \in W$. ▲

The previous example highlights two issues with the semantics of wADFs. First of all, intuitively, the grounded interpretation is supposed to collect all information that is beyond any doubt. In this example, however, that is not the case. It is beyond any doubt that a is assigned 0.25 and thus, since the value of b is defined as the maximum of the values of a and c, it is also beyond any doubt that b takes at least acceptance value 0.25. The reason why this is not discovered is that Brewka et al. [9] use partial interpretations as approximations of interpretations, but these are not refined enough to represent this kind of information. Brewka et al. already noticed this: in their conclusion, they propose (as a topic for future work) to research so-called *generalized partial interpretations* to tackle exactly this problem.

A second issue in Example 1.1 is that stable interpretations are unnatural. In AFs and ADFs, stable interpretations are (as in logic programming and other nonmonotonic logics) used to express some (constructive) form of minimality, where no arguments are accepted without reason. Intuitively, if acceptance values are in the unit interval and higher numbers mean "more accepted", we would expect the only stable interpretation to be the one that assigns 0.25 to all three arguments. In the definitions of Brewka et al. [9], this is not the case. There, stability is determined by a set of values for which, intuitively, no justification is needed. No minimality properties of such stable models are known.

[1] Instead of unknown, sometimes also the terminology *undefined* or *undecided* is used.

What these issues illustrate are two open questions regarding wADFs that we aim to answer in the current paper:

1. **What are good approximations of interpretations and how can we extend the semantic operator to such approximations?** In the original work on wADFs, partial interpretations are proposed; these assign to each argument either a value or unknown. An alternative they proposal are "generalized partial interpretations": sets of interpretations. We develop a middle-ground between the two.
2. **How can we, systematically, generalize the asymmetry between true and false in ADFs to wADFs, taking an order on the acceptance values into account?** And thus, how can we obtain a generalization of stable semantics (and of other semantics) in which smaller acceptance values are preferred over larger acceptance values?

We answer these two questions by direct applications of *approximation fixpoint theory* (AFT), an algebraical unifying study of semantics of nonmonotonic logics. Given a lattice operator and a so-called approximating operator, Denecker, Marek and Truszczyński (henceforth DMT) [12] defined several types of fixpoints. They showed that all of the main semantics of logic programming are induced by AFT, using Fitting's three-valued immediate consequence operator [20] as an approximator of the two-valued immediate consequence operator of van Emden and Kowalski [33]. They identified approximating operators for default logic [30] and autoepistemic logic [28] and showed that AFT induces all main and some new semantics in these fields [13]. AFT has been applied to various other research domains, including active integrity constraints [3], and extensions of logic programming [1, 2, 11]. Most importantly for this paper, Strass [32] showed that many semantics for Dung's argumentation frameworks and abstract dialectical frameworks can be retrieved from AFT. Strass' study revealed some anomalies with the definitions of ADF semantics of Brewka and Woltran [6], resulting in a revision [8].

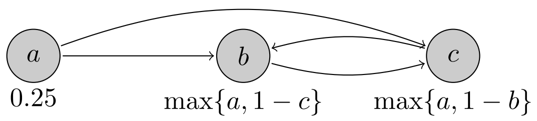
We show that direct applications of AFT yield an adequate answer to both of the above questions. As approximations, AFT uses intervals. In our setting, this means that each argument is assigned a lower bound and an upper bound on its acceptability. A semantic operator for wADFs can be defined by analogy with the operator for ADFs; this has already been done by Brewka et al. [9]. Once this operator is defined, there exists an automatic way to obtain an approximator [14] that acts on interval-based approximations instead of on interpretations. Uncoincidentally, this so-called ultimate approximator was also used to define the (revisited) semantics for ADFs [8]. Using this approximator and the characterizations of Strass [32], we obtain generalizations of all major semantics of ADFs to wADFs, including an improved definition of stable semantics and some semantics that were not generalized to the weighted case yet, namely partial stable and well-founded semantics.

Example 1.2 (Example 1.1 continued)**.** In this example, by additionally imposing the standard order on the unit interval, our semantics for wADFs are as follows[2]:

- the *models* of this wADF are as before all assignments I such that $I(a) = 0.25$ and $I(b) = I(c) \in [0.25, 1]$,
- the *grounded (partial) interpretation* of this wADF assigns 0.25 to a and $[0.25, 1]$ to b and c,
- the unique *stable model* assigns 0.25 to a, b and to c.
- the *well-founded interpretation* equals the unique stable model. ▲

We end this introduction with another example.

Example 1.3. Again consider arguments, a, b and c with acceptance values in the unit interval. Assume a is given acceptance value 0.25, independently of b and c, and that b and c are contradicting arguments, but both are supported by a. Formally, the acceptance strength of b is $\max\{a, 1-c\}$ and the strength of the acceptance for c is $\max\{a, 1-b\}$.



With our definitions, this framework has an infinite number of stable models, namely exactly all the models

$$a = 0.25, b = k, c = 1-k \mid k \in [0.25, 0.75].$$

The well-founded interpretation equals the grounded interpretation and assigns

$$a = [0.25, 0.25], b = [0.25, 0.75], c = [0.25, 0.75],$$

i.e., it finds the exact value for a and determines that b and c have acceptance values in the interval $[0.25, 0.75]$. In comparison, in the approach of Brewka et al., the grounded interpretation assigns $\mathbf{u}$ to b and c and stable models (with respect to W) are those in which b or c is assigned a value in W. ▲

The rest of this paper is structured as follows. In Section 2 we recall some background on AFT and on how AFT is used to characterize semantics of ADFs. Next, we present our AFT-based version of wADFs in Section 3. In Section 4, we recall Brewka et al.'s [9] definition of wADFs. We compare our version of wADFs to the original one and discuss strengths and weaknesses in Section 5. We study complexity of our semantics in Section 6 and conclude in Section 7.

Since we use two different versions of the formalism of wADFs, we will from now on refer to AFT-wADFs for the ones developed in this paper and BSWW-wADFs for the ones developed by Brewka et al. [9].

2 Preliminaries

We now recall the basics of AFT and its application to ADFs, following the preliminaries of [5].

Lattices and Operators A *complete lattice* $\langle L, \leq\rangle$ is a set L equipped with a partial order $\leq$, such that every set $S \subseteq L$ has both a least upper bound and a greatest lower bound, denoted $\mathrm{lub}(S)$ and $\mathrm{glb}(S)$ respectively. A complete lattice has a least element $\bot$ and a greatest element $\top$. As custom, we also use the notations $\bigwedge S = \mathrm{glb}(S)$, $x \wedge y = \mathrm{glb}(\{x, y\})$, $\bigvee S = \mathrm{lub}(S)$ and $x \vee y = \mathrm{lub}(\{x, y\})$.

[2] We define more semantics than what is compactenumd here.

An operator $O : L \to L$ is *monotone* if $x \leq y$ implies that $O(x) \leq O(y)$. An element $x \in L$ is a *fixpoint* of O if $O(x) = x$. Every monotone operator O in a complete lattice has a least fixpoint, denoted $\mathrm{lfp}(O)$.

Approximation Fixpoint Theory Given a lattice L, AFT makes uses of the bilattice L^2. We define *projections* for pairs as usual: $(x, y)_1 = x$ and $(x, y)_2 = y$. Pairs $(x, y) \in L^2$ are used to approximate all elements in the interval $[x, y] = \{z \mid x \leq z \wedge z \leq y\}$. We call $(x, y) \in L^2$ *consistent* if $x \leq y$, that is, if $[x, y]$ is non-empty. We use L^c to denote the set of consistent elements. Elements $(x, x) \in L^c$ are called *exact*. We sometimes abuse notation and use the tuple (x, y) and the interval $[x, y]$ interchangeably. The *precision order* on L^2 is defined as $(x, y) \leq_p (u, v)$ if $x \leq u$ and $v \leq y$. If (u, v) is consistent, this means that (x, y) approximates all elements approximated by (u, v). If L is a complete lattice, then so is $\langle L^2, \leq_p\rangle$.

An operator $A : L^2 \to L^2$ is an *approximator* of O if it is $\leq_p$-monotone, and has the property that for all x, $O(x) \in [x', y']$, where $(x', y') = A(x, x)$. Approximators are internal in L^c (i.e., map L^c into L^c). As usual, we restrict our attention to *symmetric* approximators: approximators A such that for all x and y, $A(x, y)_1 = A(y, x)_2$. DMT [14] showed that the consistent fixpoints of interest (supported, stable, well-founded) are uniquely determined by an approximator's restriction to L^c, hence, sometimes we only define approximators on L^c.

AFT studies fixpoints of O using fixpoints of A.

- The *A-Kripke-Kleene fixpoint* is the $\leq_p$-least fixpoint of A; it approximates all fixpoints of O.
- A *partial A-stable fixpoint* is a pair (x, y) such that $x = \mathrm{lfp}(A(\cdot, y)_1)$, where $A(\cdot, y)_1 : L \to L : x \mapsto A(x, y)_1$.
- The *A-well-founded fixpoint* is the least precise partial A-stable fixpoint.
- An *A-stable fixpoint* of O is a fixpoint x of O such that (x, x) is a partial A-stable fixpoint.

In general, a lattice operator $O : L \to L$ has a family of approximators of different precision. For two approximators A and B of O, we say that $A \leq_p B$ if $A(x, y) \leq_p B(x, y)$ for all $(x, y) \in L^c$. In this case, all A-stable fixpoints are B-stable fixpoints, and the B-well-founded fixpoint is more precise than the A-well-founded fixpoint. DMT [14] showed that there exists a most precise approximator, U_O, called the ultimate approximator of O. This operator is defined by $U_O : L^c \to L^c : (x, y) \mapsto (\bigwedge O([x, y]), \bigvee O([x, y]))$, where $O([x, y]) = \{O(z) \mid z \in [x, y]\}$.

Abstract Dialectical Frameworks Consider two truth values true ($\mathbf{t}$) and false ($\mathbf{f}$). The set of truth values is $\mathbb{B} = \{\mathbf{t}, \mathbf{f}\}$. A *vocabulary* S is a set of so-called *arguments*. An *interpretation* I of S is a function $S \to \mathbb{B}$, where $I(s) = \mathbf{t}$ means that s is accepted and $I(s) = \mathbf{f}$ means s is rejected. The set of all interpretations of S is denoted $int(S)$.

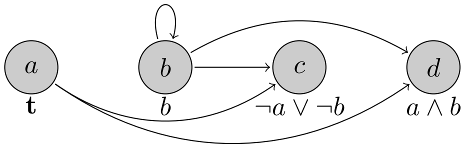
An *abstract dialectical framework* [6, 8] is a tuple $\Xi = (S, C)$, where

- S is a vocabulary, i.e. a set of arguments
- $C = \{C_s^{in}\}_{s \in S}$ is a collection of functions $C_s^{in} : int(S) \to \mathbb{B}$.

The function C_s^{in} specifies for an argument s whether or not it should be accepted, given the knowledge which arguments are accepted. These functions are often specified as propositional formulas over the vocabulary S.

Our definition is a simplified (but equivalent) version of the definition most often found in the literature, given by Brewka et al. [8]. It is sometimes referred to as the *logical* representation of ADFs [10].

Example 2.1. Let S be the set of arguments $\{a, b, c, d\}$. Furthermore, assume $C_a^{in}(I) = \mathbf{t}$ for all I, $C_b^{in}(I) = \mathbf{t}$ iff $I(b) = \mathbf{t}$, $C_c^{in}(I) = \mathbf{t}$ iff $I(a)$ and $I(b)$ are not both $\mathbf{t}$ and $C_d^{in}(I) = \mathbf{t}$ iff both $I(a)$ and $I(b)$ are $\mathbf{t}$. The ADF $\Xi = (S, C)$ is compactly depicted as the graph below, where nodes are arguments, labeled with their acceptance function and an edge from a_1 to a_2 denotes that the value of a_2 depends on the value assigned to a_1.



The following observations provide an intuitive reading of Ξ:

- a is a valid argument since it has trivial support;
- b supports itself: the only "reason" to believe b is b itself;
- a and b jointly attack c: c is rejected if a and b are both accepted; otherwise, c is accepted;
- a and b jointly support d: d only is accepted if a and b are both accepted.

Intuitively, we should accept a. Whether or not to accept b depends on which semantics for ADFs is used. Argument c can only be accepted in case we reject b, d should be accepted if we accept b. ▲

If X and Y are two interpretations of S, we say that $X \leq Y$ if $Y(s) = \mathbf{t}$ whenever $X(s) = \mathbf{t}$, i.e., if $\{s \in S \mid X(s) = \mathbf{t}\} \subseteq \{s \in S \mid Y(s) = \mathbf{t}\}$. Often, interpretations X are identified with the set $\{s \in S \mid X(s) = \mathbf{t}\}$ and hence are referred to as "sets of arguments". With an ADF Ξ, we associate an operator G_Ξ on the lattice $\langle int(S), \leq\rangle$ as follows [32]:

$$G_\Xi(X) : s \mapsto C_s^{in}(X).$$

That is, G_Ξ maps an interpretation X to the interpretation in which acceptance of arguments is determined by the value of their acceptance function in X. I.e., G_Ξ revises acceptability of arguments based on the value of the arguments they depend on in the current interpretation.

The semantical objects of interest for ADFs are either interpretations X, or pairs (X, Y) of interpretations with $X \leq Y$. We call the latter *partial interpretations*. The meaning of a partial interpretation (X, Y) is that all arguments in $\{s \in S \mid X(s) = \mathbf{t}\}$ are accepted, all arguments in $\{s \in S \mid Y(s) = \mathbf{f}\}$ are rejected and all others are unknown. This means that X is a lower bound for the acceptance and Y an upper bound. The precision order on partial interpretations is defined as usual: $(X, Y) \leq_p (X', Y')$ if $X \leq X'$

and $Y' \leq Y$. It is easy to see that the set of partial interpretations is exactly the set $int(S)^c$, i.e., the set of consistent elements of the square lattice of $int(S)$. As such, approximators of G_Ξ are defined on the set of partial interpretations. By U_Ξ we denote the ultimate approximator of G_Ξ.

Strass [32] showed that many semantics for ADFs can be characterized with AFT and new ones (that generalize corresponding AF semantics) are obtained by direct applications of AFT. Below we discuss some of them; for a complete overview, we refer to Strass [32].

- The *grounded partial interpretation* of Ξ is the $\leq_p$-least fixpoint of U_Ξ (i.e., the ultimate Kripke-Kleene fixpoint of G_Ξ).
- A partial interpretation (X, Y) is *admissible* with respect to Ξ if $(X, Y) \leq_p U_\Xi(X, Y)$.
- A partial interpretation (X, Y) is *complete* with respect to Ξ if $(X, Y) = U_\Xi(X, Y)$ (i.e., if it is a fixpoint of the ultimate approximator of G_Ξ).
- An interpretation X is a *model* of Ξ if (X, X) is complete with respect to Ξ (i.e., iff X is a fixpoint of G_Ξ).
- A partial interpretation (X, Y) is *preferred* with respect to Ξ if it is $\leq_p$-maximal among all admissible partial interpretations.
- A partial interpretation (X, Y) is *stable* with respect to Ξ if it is a stable fixpoint of U_Ξ (i.e., if it is an ultimate stable fixpoint of G_Ξ),
- An interpretation X is a *stable model* of Ξ if (X, X) is stable.
- The *AFT-well-founded partial interpretation*[3] is the well-founded fixpoint of U_Ξ.

Example 2.2 (Example 2.1 continued)**.** Here, the grounded partial interpretation is $(\{a\}, \{a, b, c, d\})$, i.e., the interpretation that assigns $\mathbf{t}$ to a and $\mathbf{u}$ to all other elements. The AFT-well-founded interpretation is more precise (as always) and here even exact; it is $(\{a, c\}, \{a, c\})$; $\{a, c\}$ is the only stable model. ▲

3 Weighted Abstract Dialectical Frameworks: An AFT Perspective

We now show how ADFs can be generalized to a multi-valued setting, building on the AFT characterization of ADF semantics. Therefore, we assume that a set ν and a partial order $\leq_a$ are given such that $\langle \nu, \leq_a \rangle$ forms a complete lattice. Intuitively, ν represents a set of acceptance values and $\leq_a$ represents a relation on those values such that $u \leq_a v$ means that the acceptance value v is closer to acceptance (more accepted) than u. Examples of such sets are

- The set $\nu = \{\mathbf{f}, \mathbf{t}\}$ with $\mathbf{f} \leq_a \mathbf{t}$; as we see later, taking this set for ν yields exactly standard ADFs.
- The unit interval, where each number represents a *degree of acceptance* with the normal order.
- The set $\{Rej, Cond_{1,2}, Cond_1, Cond_2, Acc\}$, where Rej stands for "reject", $Cond_c$ for "conditional accept, based on all the conditions in c" (these might be external conditions to be checked in order to accept an argument) and Acc stands for "accept". A sensible order on this set is depicted below.

$$\begin{tikzcd} & & Cond_1 \arrow[dr, "\leq_a"] & \\ Rej \arrow[r, "\leq_a"] & Cond_{1,2} \arrow[ur, "\leq_a"] \arrow[dr, "\leq_a"'] & & Acc \\ & & Cond_2 \arrow[ur, "\leq_a"'] & \end{tikzcd}$$

Now, given a set of values ν and a vocabulary S, a ν-interpretation of S is a function $S \to \nu$, i.e., an assignment of a value to each argument in the vocabulary. The set of all ν-interpretations of S is denoted $int(\nu, S)$. We extend the order $\leq_a$ pointwise to interpretations: $X \leq_a Y$ if for all arguments $s \in S$, $X(s) \leq_a Y(s)$. Now, we have all the ingredients to define weighted ADFs.

Definition 3.1. A *weighted abstract dialectical framework* (AFT-wADF) over ν is a tuple $\Xi = (S, C)$, where

- S is a vocabulary, i.e. a set of arguments
- $C = \{C_s^{in}\}_{s \in S}$ is a collection of functions $C_s^{in} : int(\nu, S) \to \nu$.

The intuitions are the same as in the case of standard abstract dialectical frameworks: the functions C_s^{in} determine the acceptance value of an argument, given an interpretation (an acceptance value) for all the arguments it depends on.

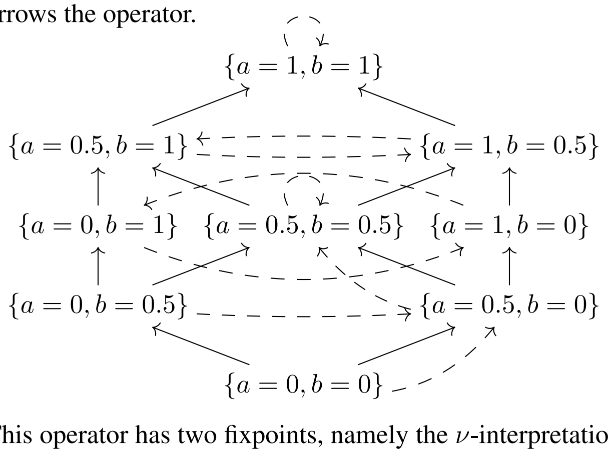
Example 3.2. Assume $S = \{a, b\}$, $\nu = \{0, 0.5, 1\}$ and $\leq_a$ is the standard order. The value 0 represents rejection of an argument, 1 represents acceptance and 0.5 represents indifference. Let $\Xi = (S, C)$ be the AFT-wADF such that $C_a^{in}(X) = \max\{0.5, X(b)\}$ and $C_b^{in}(X) = X(a)$. This represents a situation where b and a support each other, and a is supported (with strength 0.5) for some other reason. ▲

As with standard ADFs [32], we associate with an AFT-wADF over ν a semantic operator W_Ξ^ν on the lattice $\langle int(\nu, S), \leq_a \rangle$ as follows:

$$W_\Xi^\nu(X) : s \mapsto C_s^{in}(X).$$

I.e., the operator maps an interpretation X to the interpretation that assigns to each argument s, the value of its acceptance function in X.

Example 3.3 (Example 3.2 continued)**.** The operator W_Ξ^ν associated with Ξ is depicted below, where full arrows represent the $\leq_a$ relation between interpretations and dashed arrows the operator.



This operator has two fixpoints, namely the ν-interpretation that assigns 0.5 to both a and b and the one that assigns 1 to both. ▲

[3] This is a different fixpoint than the one that was called well-founded by Brewka and Woltran [6], hence the prefix.

An ν^c-*interpretation*[4] is a tuple (X, Y) of two ν-interpretations with $X \leq_a Y$. The name ν^c stems from the fact that these interpretations are *consistent* approximations of ν-interpretations, i.e., elements of $int(\nu, S)^c$. Alternatively, a ν^c-interpretation can be seen as assigning an interval $[X(a), Y(a)]$ (in the $\leq_a$ order) to each argument a. We use these two views on ν^c-interpretations interchangeably. The precision order $\leq_p$ on ν^c-interpretation is defined in the normal way: $(X, Y) \leq_p (X', Y')$ if $X \leq_a X'$ and $Y' \leq_a Y$. In the view as intervals this means that for each a, the interval $[X'(a), Y'(a)]$ is a subset of $[X(a), Y(a)]$. We say that (X, Y) *approximates* an interpretation Z if $X \leq_a Z \leq_a Y$. By U^{ν}_{Ξ} we denote the ultimate approximator of W^{ν}_{Ξ}. The semantics for AFT-wADFs are then straightforward extensions of their counterparts for traditional ADFs:

Definition 3.4.

- The *grounded ν^c-interpretation* of Ξ is the least fixpoint of U^{ν}_{Ξ} (i.e., the ultimate Kripke-Kleene fixpoint of W^{ν}_{Ξ}).
- A ν^c-interpretation (X, Y) is *admissible* with respect to Ξ if $(X, Y) \leq_p U^{\nu}_{\Xi}(X, Y)$.
- A ν^c-interpretation (X, Y) is *complete* with respect to Ξ if $(X, Y) = U^{\nu}_{\Xi}(X, Y)$ (i.e., if it is a fixpoint of the ultimate approximator of W^{ν}_{Ξ}).
- An interpretation X is a *model* of Ξ if (X, X) is complete with respect to Ξ (i.e., iff X is a fixpoint of W^{ν}_{Ξ}).
- A ν^c-interpretation (X, Y) is *preferred* with respect to Ξ if it is maximal w.r.t. $\leq_p$ among all admissible ν^c-interpretations.
- A partial interpretation (X, Y) is *stable* with respect to Ξ if it is a stable fixpoint of U^{ν}_{Ξ} (i.e., if it is an ultimate stable fixpoint of W^{ν}_{Ξ}),
- An interpretation X is a *stable model* of Ξ if (X, X) is stable.
- The *AFT-well-founded ν^c-interpretation* is the well-founded fixpoint of U^{ν}_{Ξ}.

Example 3.5 (Example 3.3 continued)**.** In this example, the grounded ν^c-interpretation assigns $[0.5, 1]$ to both a and b: it can derive that both take at least acceptance 0.5, but cannot derive anything more specific. Ξ has two models, namely the ν-interpretation $X_{0.5}$ that assigns 0.5 to both a and b and X_1 that assigns 1 to both a and b; $(X_{0.5}, X_{0.5})$ and (X_1, X_1) are also its preferred ν^c-interpretations. Other semantics, such as stable and AFT-well-founded semantics are more critical with respect to acceptance: they prefer assigning smaller acceptance values. In this case, they exclude the model X_1 where a and b are jointly self-supporting. The unique stable model of this theory is $X_{0.5}$; the AFT-well-founded ν^c-interpretation is exact and equal to $(X_{0.5}, X_{0.5})$. ▲

Many relations between the various semantics follow directly from AFT. The following proposition lists a few.

Proposition 3.6. *Assume $\Xi =$ is an AFT-wADF over ν. The following claims hold.*

1. *The grounded ν^c interpretation of Ξ approximates all models of Ξ.*
2. *The AFT-well-founded ν^c approximates all stable interpretations of Ξ.*
3. *Stable models of Ξ are $\leq_a$-minimal models of Ξ.*

The relationship between AFT-wADFs and standard ADFs follows immediately from the definitions.

Theorem 3.7. *Let Ξ be an AFT-wADF over $\mathbb{B}$ where $\mathbf{f} \leq_a \mathbf{t}$. In this case, each semantics defined in Definition 3.1 coincides with the equally named ADF semantics.*

4 Preliminaries: BSWW-wADFs

In this section, we summarize the semantics of wADFs as defined by Brewka et al. [9], a corrected version of the original semantics of [10]. Let ν be a set of values (without any specific ordering) with $\mathbf{u} \notin \nu$. By $\nu_{\mathbf{u}}$ we denote the set $\nu \cup \{\mathbf{u}\}$; here $\mathbf{u}$ represents an undefined value. A *partial ν-interpretation* of a vocabulary S is a function $I : S \to \nu_{\mathbf{u}}$, i.e., this is a $\nu_{\mathbf{u}}$-interpretation. A BSWW-wADF is a tuple $\Xi = (S, C, \nu, \leq_i)$, where

- S is a vocabulary, i.e. a set of arguments,
- $C = \{C^{in}_s\}_{s \in S}$ is a collection of functions $C^{in}_s : int(\nu, S) \to \nu$, and
- $\leq_i$ is a complete partial order[5] (CPO) on $\nu_{\mathbf{u}}$ with least element $\mathbf{u}$.

The order $\leq_i$ is pointwise extended to partial interpretations. A partial ν-interpretation I is said to be a *completion* of a partial ν-interpretation J if I is more informative then J. The set of all completions of J is denoted $[J]_c = \{I \in \mathbb{Z}(\nu, S) \mid I \geq_i J\}$. It is important to note that even if J is an exact interpretation (does not assign $\mathbf{u}$ to any arguments), the set of completions of J can still contain other interpretations than J itself. While this might seem unnatural, it is essential from a technical perspective: in the original on BSWW-wADFs, this was not the case and led to a bug in the semantics (the grounded interpretation was ill-defined). The modified definition of the completion, that we presented above, leads to well-defined semantics.

A BSWW-wADF $(S, C, \nu, \leq_i)$ induces an operator Γ_Ξ that maps a partial interpretation J to

$$\Gamma_\Xi(J) : s \mapsto \mathrm{glb}_{\leq_i}\{C^{in}_s(I) \mid I \in [J]_c\}.$$

As such, in $\Gamma_\Xi(J)$, each argument s is interpreted as the consensus over all interpretations more precise than J.

Most semantics of BSWW-wADFs are entirely defined by this operator, analogous to Definition 3.4.

Definition 4.1. Let $\Xi = (S, C, \nu, \leq_i)$ be a BSWW-wADF.

- The *grounded $\nu_{\mathbf{u}}$-interpretation* of Ξ is lfp Γ_Ξ
- A $\nu_{\mathbf{u}}$-interpretation (X, Y) is *admissible* with respect to Ξ if $(X, Y) \leq_i \Gamma_\Xi(X, Y)$.
- A $\nu_{\mathbf{u}}$-interpretation (X, Y) is *complete* with respect to Ξ if $(X, Y) = \Gamma_\Xi(X, Y)$.
- A ν-interpretation X is a *model* of Ξ if (X, X) is complete with respect to Ξ (i.e., iff X is a fixpoint of W^{ν}_{Ξ}).
- A $\nu_{\mathbf{u}}$-interpretation (X, Y) is *preferred* with respect to Ξ if it is maximal w.r.t. $\leq_i$ among all admissible $\nu_{\mathbf{u}}$-interpretations.

[4] We refrain from using the terminology *partial* interpretation here to avoid confusion (in later sections) with what was called a partial interpretation by BSWW.

[5] This means that every subset of $\nu_{\mathbf{u}}$ has a greatest lower bound and every ascending chain has a least upper bound.

BSWW also give a definition of stable semantics, not directly from Ξ but using an auxiliary set of values. We refer to [9] for the formal definition.

5 wADFs: A Comparison

We now discuss how our semantics relates to the original definition. We discuss two aspects: first, we discuss the assumptions both frameworks make; next, we research how they relate when the assumptions of both are satisfied.

5.1 Assumptions – Similarities and Differences

The two definitions start from:

- A set S of arguments,
- A set ν of acceptance values,
- A collection of acceptance functions C_s^{in}.

An *interpretation* is in both frameworks an assignment of acceptance degrees to arguments. The essence of a wADF is an operator that maps interpretations to interpretations. (this operator is not explicated in the BSWW case, but can easily be obtained by restricting Γ_Ξ to interpretations).

For AFT-wADFs, additionally, an acceptance order $\leq_a$ on ν is imposed. For BSWW-wADFs, additionally, an information order $\leq_i$ on ν is imposed.

The approximations in the two frameworks differ. For AFT-wADFs, we use ν^c-interpretations: assignments of a lower and upper bound on the acceptability of each argument. On such intervals, a natural precision ordering $\leq_p$ is imposed (an interval is more precise than another interval if the former is a subset of the latter). BSWW-wADFs use partial ν-interpretations, which are $\nu_{\mathbf{u}}$-interpretations. The order $\leq_i$ is extended to such interpretations.

For both frameworks, an operator on approximations is defined. The definition of these operators is remarkably similar. For ν^c-interpretations, the ultimate approximator is used. A direct definition of this operator is: the operator U_Ξ that maps a ν^c-interpretation X to the ν^c-interpretation

$$U_\Xi(X) : s \mapsto \mathrm{glb}_{\leq_p}\{C_s^{in}(Y) \mid Y \in int(\nu, S) \wedge Y \geq_p X\}$$

For partial ν-interpretations, the approximator is the operator Γ_Ξ that maps a partial interpretation X to the partial interpretation

$$\Gamma_\Xi(X) : s \mapsto \mathrm{glb}_{\leq_i}\{C_s^{in}(Y) \mid Y \in int(\nu, S) \wedge Y \geq_i X\}$$

I.e., these definitions are identical; only the space of approximations differs (and the orders defined on it).

Both frameworks define their semantics based on these operators. The semantics are defined similarly, with three notable exceptions: stable semantics (which in the BSWW case is defined using an auxiliary set of variables, whereas in our case directly from the operator) and partial stable and well-founded semantics (which are new).

The most important differences stem from the orders used: an order on acceptability $\leq_a$ in our case, versus an information order $\leq_i$ in the BSWW case. The information order is very similar to the precision order $\leq_p$, derived from $\leq_a$. Yet, we believe its inclusion to be a bad design choice. The main reason for this is that it obscures the line between interpretations and approximations. To illustrate this, notice that in many formalisms approximations of interpretations are present. For instance, in standard ADFs, or in logic programming, these approximations are partial interpretations. In constraint programming, approximations of interpretations sometimes assign upper and lower bounds to integer values (depending on the type of propagation mechanism used). What all of these formalisms have in common is that approximations are comparable by some form of precision or information ordering and that interpretations are maximal with respect to this ordering. A consequence is that it can never be the case that one interpretation is more informative (precise) than another one. This is not the case in BSWW-wADFs, leading to a lack of an informal explanation of what an interpretation is. Also, on the formal side, this has some unpleasant effects. First of all, as mentioned by Brewka et al. [9], the bug in the original wADF paper was due to this fact: the definitions in [10] only work for flat information orderings (i.e., orderings such that all interpretations are $\leq_i$-maximal). Second, the following example shows that with the definitions of BSWW, not all models are preferred interpretations, which might come as a surprise to the reader familiar with AFs and ADFs.

Example 5.1. Consider $\nu = \{1, 2\}; S = \{a\}$ and an information ordering given by $\mathbf{u} \leq_i 1 \leq_i 2$. Furthermore, consider the BSWW-wADF such that $C_a^{in}(X) = X(a)$ for each ν-interpretation X. In this case, the interpretation X_1 that assign 1 to a and X_2 that assigns 2 to a are both models. However, since $X_1 \leq_i X_2$, X_1 is not preferred. ▲

Third, intuitively, the function C_s^{in} specifies with which strength an argument s is acceptable in a given interpretation. The operator Γ_Ξ revises an interpretation based on this acceptance. One might expect that such revision respects the acceptance functions in the sense that in an interpretation X, it always holds that

$$\Gamma_\Xi(X)(s) = C_s^{in}(X), \tag{1}$$

i.e. that the strength of the argument s in the revised interpretation equals what the acceptance function dictates. This equation indeed holds for the original definition [10], but was sacrificed when making the technical changes required to eliminate the bug [9]. In the revised report, the intuition that (1) should hold is given in the definition of a model, where it is stated that intuitively in models "the value is exactly what is required by the acceptance functions". This informal statement is only correct if $\Gamma_\Xi(X)(s) = C_s^{in}(X)$. All in all, we feel that the information order $\leq_i$ does more harm than good, both on an informal and a formal level.

5.2 Overlap

Despite the disadvantages $\leq_i$ presents, it can be used to "emulate" an interval-based precision ordering, as given in the AFT approach. Assume that an AFT-wADF $\Xi = (S, C)$ over $\langle \nu, \leq_a \rangle$ is given. The idea now is: given a set of values ν, ν^c consists of all intervals in ν. Let us now define $\nu' = \nu^c \setminus \{\bot_{\leq_p}\}$, i.e., we exclude the least precise approximation. In this case $\nu'_{\mathbf{u}}$ is isomorphic to ν_c. In BSWW-wADFs, due to the order $\leq_i$, part of the approximation space (everything except for the least precise approximation) is captured within the set of values. and hence the approximations in both frameworks coincide. We show a correspondence between AFT-wADFs over ν and BSWW-wADFS over ν'. Therefore, let Ξ' for the rest of this section denote the BSWW-wADF $\Xi' = (S, C', \nu', \leq_p)$, where

$$C_s'^{in}(X) = \bigwedge_{\leq_p}\{C_s^{in}(I) \mid I \in int(\nu, S) \text{ and } I \geq_p X\}$$

With these definitions, there is a strong overlap between semantics, as defined in the current paper and in the original work by BSWW. The most fundamental relationship is the one between the defined operators (that are in both cases used to define the semantics).

Theorem 5.2. *Let Ξ and Ξ' be as above. The operators U_Ξ^ν and Γ_Ξ coincide.*

From this result, it follows that many semantics coincide, though (as explained below) not all semantics do.

Corollary 5.3. *Let Ξ and Ξ' be as above. For the following semantics, Definitions 3.4 and 4.1 coincide: grounded, admissible, complete, and preferred.*

For the other semantics the formalisms do not neccesarily agree: the well-founded semantics is only defined for AFT-wADFs; the difference in stable semantics is already illustrated in Examples 1.1 and 1.2, where we argued that stable models in the AFT sense are more natural, provided that an acceptance order is available. One last difference in semantics between the two formalism is found in the notion of models. In the AFT approach, models are only allowed to take values in ν. I.e., each argument should be assigned an exact (maximally precise) value; not an approximation. In the BSWW approach on the other hand in a model, each argument is assigned a value in ν or an approximation thereof (but not the least precise approximation $\mathbf{u}$ which in this case corresponds to the interval $(\bot_{\leq_a}, \top_{\leq_a})$).

6 Complexity

We now study complexity of tasks related to wADFs. We assume that $\langle \nu, \leq_a \rangle$ is fixed (hence not part of the input of the problem). Furthermore, we assume that we can in polynomial time determine if $v_1 \leq_a v_2$ for $v_1, v_2 \in \nu$ and that we can, in polynomial time compute $C_s^{in}(X)$ for an interpretation X and acceptance function C_s^{in}. Some (but not all) of our results furthermore require that ν be finite.

The size of the input of all considered problems is the cardinality of the finite set S of arguments. Hardness results follow (by Theorem 3.7) from standard ADFs [31]. Complexity results for admissible, complete, and preferred semantics can (using Corollary 5.3) be carried over from BSWW-wADFs [10]. Hence, we focus on the other semantics.

Proposition 6.1. *Verifying if a given ν-interpretation X is a model of Ξ is in P. If ν is finite, checking if there exists a model of Ξ is in NP.*

While the fact that checking if X is a model is in P might seem like an obvious property, we inform the reader that this property does not hold for BSWW-wADFs.

Proposition 6.2. *If ν is finite, verifying if a given ν^c-interpretation (X, Y) is stable with respect to Ξ is in Δ_2^P.*

Proposition 6.3. *If ν is finite, the problem of checking whether Ξ has a stable model is in Σ_2^P.*

Proposition 6.4. *Assume $a \in S$ and $v_1 \leq_a v_2 \in \nu$. Let (X, Y) denote the grounded ν^c-interpretation of Ξ. The problem of checking whether a is assigned a value at least as precise as (v_1, v_2) in (X, Y) (i.e., whether $(X(a), Y(a)) \geq_p (v_1, v_2)$) is in Δ_2^P.*

The same holds when we take (X, Y) to be the AFT-well-founded ν^c-interpretation of Ξ.

7 Related Work and Conclusion

There is a large body of work on weighted argumentation [6, 7, 19, 21, 22], probabilistic argumentation [17, 27, 25, 23, 15, 29, 16, 24], and social argumentation [26]. In some of these papers weights on nodes are considered; in others weights on attacks. Some of these formalism take the number of attacking nodes or votes into account to determine these weights. Our current work relates to that body of work in a similar way as how the original work on wADFs relates to it, hence we refer to Brewka et al. [10] for an overview.

In this paper we applied approximation fixpoint theory to wADFs. Our main contributions are two-fold. The first contribution is that by applying AFT, we shed light on the relationship of wADFs with other non-monotonic formalisms and we obtain several new semantics for free (partial stable semantics, well-founded semantics). Furthermore, we can immediately use many of the theoretic results from AFT, such as stratification results [34] or knowledge compilation techniques [4]. Most of the semantics are defined similar to BSWW semantics, with the notable exception being stable semantics, where our version (contrary to the previous one) yields acceptance-minimal models. What AFT gives us here is the guarantee that all developed semantics are based on the same fundamental principles as the semantics of ADFs, and in fact of many other nonmonotonic formalisms. For instance, we automatically obtain a characterization of the AFT-well-founded interpretation as the least precise partial stable interpretation. In order to be convinced of the correctness of AFT-style semantics, one only needs to verify is that **i)** the defined acceptance order makes sense, **ii)** intervals are a desired type of approximations, and **iii)** the semantic operator is defined correctly. For the last point, the semantic operator we use was already present in the work of Brewka et al. [10].

While we believe interval-based approximations (used in our approach) are often natural, we can imagine situations where different approximations prove more useful; this is where our second contribution kicks in: in Section 5 we highlight a troublesome design decision in wADFs, namely the lack of distinction between exact and approximate values. This has far-reaching consequences, both on an informal and on a formal level. We believe such distinction to be important and hence, the current work can be seen as a plea towards including such a distinction in wADFs, independently of whether the AFT approach, using interval-based approximations, is followed or not.

References

[1] Antic, C.; Eiter, T.; and Fink, M. 2013. Hex semantics via approximation fixpoint theory. In *Proceedings of LPNMR*, 102–115.

[2] Bi, Y.; You, J.-H.; and Feng, Z. 2014. A generalization of approximation fixpoint theory and application. In *Proceedings of RR*, 45–59.

[3] Bogaerts, B., and Cruz-Filipe, L. 2018. Fixpoint semantics for active integrity constraints. *AIJ* 255:43–70.

[4] Bogaerts, B., and Van den Broeck, G. 2015. Knowledge compilation of logic programs using approximation fixpoint theory. *TPLP* 15(4–5):464–480.

[5] Bogaerts, B.; Vennekens, J.; and Denecker, M. 2015. Grounded fixpoints and their applications in knowledge representation. *AIJ* 224:51–71.

[6] Brewka, G., and Woltran, S. 2010. Abstract dialectical frameworks. In *Proceedings of KR*, 102–111.

[7] Brewka, G., and Woltran, S. 2014. GRAPPA: A semantical framework for graph-based argument processing. In *Proceedings of ECAI*, 153–158.

[8] Brewka, G.; Strass, H.; Ellmauthaler, S.; Wallner, J. P.; and Woltran, S. 2013. Abstract dialectical frameworks revisited. In *Proceedings of IJCAI*, 803–809.

[9] Brewka, G.; Pührer, J.; Strass, H.; Wallner, J. P.; and Woltran, S. 2018a. Weighted abstract dialectical frameworks: Extended and revised report. *CoRR* abs/1806.07717.

[10] Brewka, G.; Strass, H.; Wallner, J. P.; and Woltran, S. 2018b. Weighted abstract dialectical frameworks. In *Proceedings of AAAI*.

[11] Charalambidis, A.; Rondogiannis, P.; and Symeonidou, I. 2018. Approximation fixpoint theory and the well-founded semantics of higher-order logic programs. *CoRR* abs/1804.08335. to appear in TPLP.

[12] Denecker, M.; Marek, V.; and Truszczyński, M. 2000. Approximations, stable operators, well-founded fixpoints and applications in nonmonotonic reasoning. In *Logic-Based Artificial Intelligence, Springer*, volume 597, 127–144.

[13] Denecker, M.; Marek, V.; and Truszczyński, M. 2003. Uniform semantic treatment of default and autoepistemic logics. *AIJ* 143(1):79–122.

[14] Denecker, M.; Marek, V.; and Truszczyński, M. 2004. Ultimate approximation and its application in nonmonotonic knowledge representation systems. *Information and Computation* 192(1):84–121.

[15] Dondio, P. 2014. Toward a computational analysis of probabilistic argumentation frameworks. *Cybernetics and Systems* 45(3):254–278.

[16] Dondio, P. 2017. Propagating degrees of truth on an argumentation framework: an abstract account of fuzzy argumentation. In *Proceedings of SAC*, 995–1002.

[17] Dung, P. M., and Thang, P. M. 2010. Towards (probabilistic) argumentation for jury-based dispute resolution. In *Proceedings of COMMA*, 171–182.

[18] Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *AIJ* 77(2):321 – 357.

[19] Dunne, P. E.; Hunter, A.; McBurney, P.; Parsons, S.; and Wooldridge, M. 2011. Weighted argument systems: Basic definitions, algorithms, and complexity results. *AIJ* 175(2):457–486.

[20] Fitting, M. 2002. Fixpoint semantics for logic programming — A survey. *Theoretical Computer Science* 278(1-2):25–51.

[21] Gabbay, D. M. 2012. Equational approach to argumentation networks. *Argument & Computation* 3(2-3):87–142.

[22] Grossi, D., and Modgil, S. 2015. On the graded acceptability of arguments. In *Proceedings of IJCAI*, 868–874.

[23] Hunter, A., and Thimm, M. 2014. Probabilistic argumentation with incomplete information. In *Proceedings of ECAI*, 1033–1034.

[24] Hunter, A.; Polberg, S.; and Thimm, M. 2018. Epistemic graphs for representing and reasoning with positive and negative influences of arguments. *CoRR* abs/1802.07489.

[25] Hunter, A. 2013. A probabilistic approach to modelling uncertain logical arguments. *Int. J. Approx. Reasoning* 54(1):47–81.

[26] Leite, J., and Martins, J. 2011. Social abstract argumentation. In *Proceedings of IJCAI*, 2287–2292.

[27] Li, H.; Oren, N.; and Norman, T. J. 2011. Probabilistic argumentation frameworks. In *TAFA, Revised Selected Papers*, 1–16.

[28] Moore, R. C. 1985. Semantical considerations on nonmonotonic logic. *AIJ* 25(1):75–94.

[29] Polberg, S., and Doder, D. 2014. Probabilistic abstract dialectical frameworks. In *Proceedings of JELIA*, 591–599.

[30] Reiter, R. 1980. A logic for default reasoning. *AIJ* 13(1-2):81–132.

[31] Strass, H., and Wallner, J. P. 2015. Analyzing the computational complexity of abstract dialectical frameworks via approximation fixpoint theory. *AIJ* 226:34–74.

[32] Strass, H. 2013. Approximating operators and semantics for abstract dialectical frameworks. *AIJ* 205:39–70.

[33] van Emden, M. H., and Kowalski, R. A. 1976. The semantics of predicate logic as a programming language. *J. ACM* 23(4):733–742.

[34] Vennekens, J.; Gilis, D.; and Denecker, M. 2006. Splitting an operator: Algebraic modularity results for logics with fixpoint semantics. *ACM Trans. Comput. Log.* 7(4):765–797.