Initial Models and Serialisability in Abstract Dialectical Frameworks

Lars Bengel, **Matthias Thimm**
Artificial Intelligence Group, University of Hagen, Germany
{lars.bengel, matthias.thimm}@fernuni-hagen.de

Abstract

We introduce *initial models* for abstract dialectical frameworks (ADFs) as a notion of minimal justifiable valuations and based on that, generalise the concept of *serialisability* of argumentation semantics to ADFs. In particular, we show that the characteristic operator-based semantics for ADFs can be characterised through *serialisation sequences*, which are, essentially, decompositions of a model into a series of initial models, representing a more fine-grained view into why a model is acceptable wrt. the semantics. We also analyse the computational complexity of tasks related to initial models.

1 Introduction

Ever since its introduction by Dung [1995], *abstract argumentation* has been an important and growing field in artificial intelligence in general and within the area of knowledge representation and reasoning in particular. Specifically, in the context of explainable artificial intelligence, argumentation represents a promising approach with numerous recent developments [Leofante *et al.*, 2024]. An *abstract argumentation framework* (AF) is a directed graph, where the nodes represent abstract arguments and the edges represent attacks among them. Various extensions of AFs have been proposed in the literature [Brewka *et al.*, 2014], but *abstract dialectical frameworks* (ADF) have emerged as a particularly powerful generalisation of AFs [Brewka *et al.*, 2013]. In an ADF, the relations between the arguments are represented by *acceptance conditions*, i. e., (propositional) logical formulae that specify the conditions under which arguments may be accepted. This enables the representation of more complex relationships between arguments than simple attacks, such as collective attacks or support relations.

Formal semantics for both AFs and ADFs are given through functions that determine admissible sets of arguments, called extensions [Baroni *et al.*, 2018], or, in the case of ADFs, three-valued models [Brewka *et al.*, 2017]. In particular, the classical admissibility-based semantics of Dung have been generalised to ADFs [Brewka *et al.*, 2013].

As the name suggests, ADFs are inspired by dialectics [Brewka *et al.*, 2013]. An important element of dialectics is procedurality, i. e., the fact that arguments are put forward and are then followed by counter-arguments [Hage, 2000; Rescher, 1977]. While this aspect is modelled well on the syntactic level in ADFs, on the semantical side this aspect is somewhat lost, just like in the case of AFs [Verheij, 1996]. Consider, for instance, the ADF in Figure 1, where acceptance conditions of arguments are placed right above them (we will provide formal definitions in Section 2). We have that a and b can only be accepted if the other is rejected, meaning they form a sort of atomic conflict that must be resolved. Only after resolving this conflict, for example by accepting a and rejecting b, can we turn to the remaining arguments and evaluate them properly. Now, rejecting b directly implies that c must also be rejected and in turn that we shall accept d afterwards. On the other hand, if we accept b and reject a in the initial conflict, that implies that we accept c and subsequently reject d. If we only consider the resulting admissible model that assigns the respective truth values, we disregard this information about the reasoning process of the argumentation performed to arrive at the conclusion.

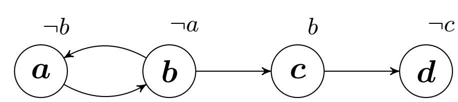

Figure 1: An ADF, the arguments a and b are in a conflict, while b supports c and the argument d may only be accepted if c is rejected.

An approach to address the above-described problem is the notion of *serialisability* for AFs [Thimm, 2022]. It provides a non-deterministic construction scheme for extensions, where initial sets of the AF are selected iteratively. An initial set [Xu and Cayrol, 2018] is thereby defined as a non-empty, minimal admissible set, essentially representing a solution to an atomic conflict of the AF. An extension can then be represented by *serialisation sequences*, i. e., sequences of initial sets that represent an order in which the corresponding extension can be build. The serialisation sequences thus lend themselves well to build explanations for the extension that they represent inside the AF [Bengel, 2022]. Even more so, they also induce a very expressive semantical equivalence notion for AFs [Bengel *et al.*, 2024].

In this work, we characterise initial models for ADFs and generalise the notion of serialisability to ADFs. In particular, we define initial models as the minimal, non-empty admissible models with respect to characteristic operator of [Brewka *et al.*, 2013]. Based on that, we show that initial models coincide with the initial sets for those ADFs that correspond directly to AFs. Furthermore, we generalise the notion of serialisation sequences from AFs to ADFs and characterise admissibility as well as the preferred, complete, grounded and two-valued semantics of [Brewka *et al.*, 2013] in terms of serialisation sequences for ADFs in general. Due to the higher expressiveness of the three-valued models of ADFs (compared to extensions), this yields fine-grained serialisation sequences that make explicit the process of both accepted and rejected arguments within acceptable models. That means, with our approach we incorporate the procedural aspect of dialectic argumentation directly into the semantics of ADFs. Finally, we also analyse the computational complexity of tasks related to initial models, which surprisingly turns out to be harder than the respective tasks for admissible models.

To summarise, the main contributions of this work are:

1. We define the notion of *initial models* for ADFs and show that they are a generalisation of initial sets in AFs (Section 3).
2. We define *serialisability* and *serialisation sequences* for ADFs, and analyse their properties (Section 4).
3. We characterise most admissibility-based semantics for ADFs via serialisation sequences (Section 5).
4. We provide complexity results for various tasks related to initial models in ADFs (Section 6).

In Section 2 we introduce the necessary background and Section 7 concludes the paper. Omitted proofs for all technical results can be found in an online appendix[1].

2 Preliminaries

In the following, we introduce the necessary background on abstract argumentation frameworks [Dung, 1995] and abstract dialectical frameworks [Brewka *et al.*, 2013].

2.1 Abstract Argumentation Frameworks

An *abstract argumentation framework* (AF) is a pair $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ where $\mathcal{A}$ is a finite set of arguments and $\mathcal{R}$ is a relation $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. For two arguments $a, b \in \mathcal{A}$, the relation $a\mathcal{R}b$ means that argument a *attacks* b. For a set $S \subseteq \mathcal{A}$ we may write $S\mathcal{R}a$ (resp. $a\mathcal{R}S$) iff there is some $c \in S$ with $c\mathcal{R}a$ (resp. $a\mathcal{R}c$). We say that a set $S \subseteq \mathcal{A}$ is *conflict-free* iff for all $a, b \in S$ it is not the case that $a\mathcal{R}b$. A set S *defends* an argument $b \in \mathcal{A}$ iff for all a with $a\mathcal{R}b$ there is $c \in S$ with $c\mathcal{R}a$. Furthermore, a set S is called *admissible* (ad) iff it is conflict-free and S defends all $a \in S$. Let $\mathsf{ad}(\mathcal{F})$ denote the set of admissible sets of $\mathcal{F}$.

Non-empty minimal admissible sets have been coined *initial sets* by Xu and Cayrol (2018).

Definition 1. For $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, a set $S \subseteq \mathcal{A}$ with $S \neq \emptyset$ is called an *initial set* if S is admissible and there is no admissible $S' \subsetneq S$ with $S' \neq \emptyset$.

With $\mathsf{is}(\mathcal{F})$ we denote the set of initial sets of the AF $\mathcal{F}$.

[1]`https://doi.org/10.5281/zenodo.15357088`

2.2 Abstract Dialectical Frameworks

Let At be a set of propositional atoms and $\mathcal{L}_{\mathsf{At}}$ the propositional language over At closed under the usual connectives $\neg$, $\vee$, and $\wedge$. An *interpretation* υ is a function that assigns to each atom in At a truth value $\mathtt{true}$ or $\mathtt{false}$. Truth evaluation wrt. an interpretation is extended to arbitrary formulas in the usual way. If an interpretation υ satisfies a formula $\phi \in \mathcal{L}_{\mathsf{At}}$, i. e., it assigns $\mathtt{true}$ to ϕ, we may also call υ a *model* of ϕ. If a formula ϕ has at least one model it is *satisfiable*. Furthermore, we denote with $\textsc{Taut}_{\mathsf{At}}$ the set of tautologies and with $\textsc{Unsat}_{\mathsf{At}}$ the set of unsatisfiable formulae over At.

Definition 2. An *abstract dialectical framework* (ADF) is a pair $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ where $\mathcal{A}$ is a set of arguments and $\mathcal{C}$ is a set of propositional formulae $\{\phi_a\}_{a \in \mathcal{A}}$ over $\mathcal{A}$, called *acceptance conditions*.[2]

Reasoning in ADFs is performed via three-valued propositional interpretations that satisfy all acceptance conditions called *models*.

Definition 3. Let $\mathcal{D}$ be an ADF and $\upsilon : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ is three-valued interpretation. For a propositional formula ϕ over $\mathcal{A}$, we define $\upsilon(\phi)$ inductively as follows

$$\upsilon(\neg\psi) = \begin{cases} \mathsf{t} & \text{if } \upsilon(\psi) = \mathsf{f} \\ \mathsf{f} & \text{if } \upsilon(\psi) = \mathsf{t} \\ \mathsf{u} & \text{if } \upsilon(\psi) = \mathsf{u} \end{cases}$$

$$\upsilon(\psi_1 \wedge \psi_2) = \begin{cases} \mathsf{t} & \text{if } \upsilon(\psi_1) = \upsilon(\psi_2) = \mathsf{t} \\ \mathsf{f} & \text{if } \upsilon(\psi_1) = \mathsf{f} \text{ or } \upsilon(\psi_2) = \mathsf{f} \\ \mathsf{u} & \text{otherwise} \end{cases}$$

$$\upsilon(\psi_1 \vee \psi_2) = \begin{cases} \mathsf{t} & \text{if } \upsilon(\psi_1) = \mathsf{t} \text{ or } \upsilon(\psi_2) = \mathsf{t} \\ \mathsf{f} & \text{if } \upsilon(\psi_1) = \upsilon(\psi_2) = \mathsf{f} \\ \mathsf{u} & \text{otherwise} \end{cases}$$

An interpretation υ is called a *(three-valued) model* of $\mathcal{D}$, iff $\upsilon(a) = \mathsf{u}$ or $\upsilon(a) = \upsilon(\phi_a)$ for all $a \in \mathcal{A}$.

We denote with $\mathsf{val}_3(\mathcal{D})$ the set of *three-valued models* of $\mathcal{D}$ and with υ_{u} the three-valued model that assigns u to all $a \in \mathcal{A}$. We also define the set of all *two-valued models* of $\mathcal{D}$ as $\mathsf{val}_2(\mathcal{D}) = \{\upsilon \in \mathsf{val}_3(\mathcal{D}) \mid \forall a \in \mathcal{A} : \upsilon(a) \neq \mathsf{u}\}$.

Furthermore, we also consider the inverse function $\upsilon^{-1} : \{\mathsf{t}, \mathsf{f}, \mathsf{u}\} \to 2^{\mathcal{A}}$, meaning we denote with $\upsilon^{-1}(\mathsf{x})$ the set arguments assigned the truth value x by the interpretation υ for all $\mathsf{x} \in \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$. We will also write $a \mapsto \mathsf{x}$ instead of $\upsilon(a) = \mathsf{x}$ for readability. $\boldsymbol{a \mapsto \mathsf{x}}$ will be written in bold if $\mathsf{x} \neq \mathsf{u}$ to highlight $\leq_i$-maximal assignments (see below).

We consider the following partial order $\leq_i$ according to information content: $\mathsf{u} <_i \mathsf{t}$, $\mathsf{u} <_i \mathsf{f}$ and no other pair in $<_i$. For a pair of three-valued interpretations $\upsilon_1, \upsilon_2 : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ we define

$$\upsilon_1 \leq_i \upsilon_2 \quad \text{iff } \upsilon_1(a) \leq_i \upsilon_2(a) \text{ for all } a \in \mathcal{A}.$$

The *consensus operator* $\sqcap$ combines two interpretations υ_1 and υ_2 into a new interpretation υ_3, that coincides with υ_1

[2]Note that the original definition of ADFs [Brewka *et al.*, 2013] includes the *links* between arguments explicitly, but here we assume them implicitly given by the acceptance conditions.

and υ_2 wherever both interpretations coincide and assigns u otherwise. We define $\upsilon_3 = \upsilon_1 \sqcap \upsilon_2$ as

$$\upsilon_3(a) = \begin{cases} \mathsf{t} & \text{if } \upsilon_1(a) = \upsilon_2(a) = \mathsf{t} \\ \mathsf{f} & \text{if } \upsilon_1(a) = \upsilon_2(a) = \mathsf{f} \\ \mathsf{u} & \text{otherwise} \end{cases} \tag{1}$$

for all $a \in \mathcal{A}$.

For some three-valued interpretation υ, we define the set of *completions* $[\upsilon]_2$ as

$$[\upsilon]_2 = \{\upsilon' \mid \upsilon \leq_i \upsilon', (\upsilon')^{-1}(\mathsf{u}) = \emptyset\} \tag{2}$$

The set $[\upsilon]_2$ is then essentially the set of all two-valued interpretations that contain more information than υ.

For the semantic evaluation of an ADF $\mathcal{D}$ we then define the *characteristic operator* $\Gamma_{\mathcal{D}}$ which computes for a model υ the consensus of all its completions for every argument.

Definition 4. Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF and $\upsilon : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ be a three-valued interpretation. We define the *characteristic operator* $\Gamma_{\mathcal{D}}$ as

$$\Gamma_{\mathcal{D}}(\upsilon)(a) = \bigsqcap \{\upsilon'(\phi_a) \mid \upsilon' \in [\upsilon]_2\}$$

for all $a \in \mathcal{A}$.

Analogous to the characteristic function of AFs, we utilise the characteristic operator to define the classical admissibility-based semantics for ADFs.

Definition 5. Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF. We define $\upsilon_0 = \upsilon_{\mathsf{u}}$ and $\upsilon_i = \Gamma_{\mathcal{D}}(\upsilon_{i-1})$ for all $i \in \mathbb{N}$. Let $k \in \mathbb{N}$ be the smallest number with $\upsilon_k = \upsilon_{k-1}$. Then υ_k is called the *grounded model* of $\mathcal{D}$.

Definition 6. Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF. An interpretation $\upsilon : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ is called

- an *admissible model* of $\mathcal{D}$, iff $\upsilon \leq_i \Gamma_{\mathcal{D}}(\upsilon)$,
- a *complete model* of $\mathcal{D}$, iff $\upsilon = \Gamma_{\mathcal{D}}(\upsilon)$,
- a *preferred model* of $\mathcal{D}$, iff υ is complete and there is no complete model υ' with $\upsilon \leq_i \upsilon'$.

The stable semantics has been revised in [Brewka *et al.*, 2013] compared to its original proposal in [Brewka and Woltran, 2010] and is defined as follows.

Definition 7. Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF. A two-valued model υ of $\mathcal{D}$ is a *stable model* of $\mathcal{D}$, iff $\upsilon^{-1}(\mathsf{t})$ equals $\upsilon_{gr}^{-1}(\mathsf{t})$ of the grounded model υ_{gr} of the reduced ADF $\mathcal{D}_{\downarrow\upsilon} = (\upsilon^{-1}(\mathsf{t}), \mathcal{C}')$ such that for all $a \in \upsilon^{-1}(\mathsf{t})$ we set $\phi'_a = \phi_a^{[b/\mathsf{f}\,:\,\upsilon(b)=\mathsf{f}]}$.

With $\sigma(\mathcal{D})$ we denote the set of σ-models of $\mathcal{D}$ for $\sigma \in \{\mathsf{ad}, \mathsf{gr}, \mathsf{co}, \mathsf{pr}, \mathsf{val}_2, \mathsf{st}\}$.

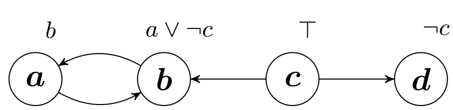

Figure 2: The ADF $\mathcal{D}_1$ from Example 1.

Example 1. Consider the ADF $\mathcal{D}_1$ in Figure 2. The ADF has the grounded model

$$\upsilon_1 = \{a \mapsto \mathsf{u}, b \mapsto \mathsf{u}, \boldsymbol{c} \mapsto \mathbf{t}, \boldsymbol{d} \mapsto \mathbf{f}\},$$

and two preferred models

$$\upsilon_2 = \{\boldsymbol{a} \mapsto \mathbf{f}, \boldsymbol{b} \mapsto \mathbf{f}, \boldsymbol{c} \mapsto \mathbf{t}, \boldsymbol{d} \mapsto \mathbf{f}\},$$
$$\upsilon_3 = \{\boldsymbol{a} \mapsto \mathbf{t}, \boldsymbol{b} \mapsto \mathbf{t}, \boldsymbol{c} \mapsto \mathbf{t}, \boldsymbol{d} \mapsto \mathbf{f}\}.$$

All three models are complete and the latter two are also two-valued models. However, only υ_2 is a stable model.

For a given argumentation framework $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, we define the corresponding ADF $\mathcal{D}_{\mathcal{F}}$ as follows.

Definition 8. Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF. Then, we define the ADF $\mathcal{D}_{\mathcal{F}} = (\mathcal{A}, \mathcal{C})$, where $\mathcal{A}$ is the same set of arguments and $\mathcal{C}$ is defined as $\{\phi_a = \bigwedge_{b \in a^-} \neg b\}_{a \in \mathcal{A}}$.

Furthermore, for some ADF $\mathcal{D}_{\mathcal{F}} = (\mathcal{A}, \mathcal{C})$ and a given interpretation υ of $\mathcal{D}_{\mathcal{F}}$, we define the unique extension E_υ corresponding to υ as

$$E_\upsilon = \upsilon^{-1}(\mathsf{t}). \tag{3}$$

3 Characterising Initial Models in ADFs

To characterise initial models and serialisability for ADFs we build on the characteristic operator-based approach to admissibility in ADFs as reviewed in Section 2.2. It should be noted however that there exist other approaches to admissibility in ADFs, for instance the decisive outing formulation of [Polberg *et al.*, 2013]. The concept of serialisability for semantics of AFs is closely related to the characteristic function [Thimm, 2022], i. e., it can be understood as a generalisation of the construction scheme of the grounded extension via applying the characteristic function iteratively. In other words, it generalises from iteratively adding unattacked arguments to iteratively adding initial sets as the atomic units. For that reason, we consider the characteristic operator-based approach to be the most suitable for defining serialisability.

Before characterising initial models for ADFs, consider again the translation from a model υ of an ADF $\mathcal{D}_{\mathcal{F}}$ to an extension E_υ of the corresponding AF $\mathcal{F}$ as defined in Equation (3). Notably, there are cases where multiple models υ of $\mathcal{D}_{\mathcal{F}}$ correspond to the same extension E_υ of $\mathcal{F}$, as shown by the following example.

Example 2. Consider the ADF $\mathcal{D}_2$ in Figure 3. $\mathcal{D}_2$ has the following three admissible models:

$$\upsilon_1 = \{a \mapsto \mathsf{u}, b \mapsto \mathsf{u}\},$$
$$\upsilon_2 = \{\boldsymbol{a} \mapsto \mathbf{t}, b \mapsto \mathsf{u}\},$$
$$\upsilon_3 = \{\boldsymbol{a} \mapsto \mathbf{t}, \boldsymbol{b} \mapsto \mathbf{f}\}.$$

Both υ_2 and υ_3 correspond to the preferred extension $\{a\}$ of the corresponding AF $\mathcal{F} = (\{a, b\}, \{(a, b)\})$. However, υ_2 is not a complete model of $\mathcal{D}_2$ while υ_3 is complete and $\leq_i$-maximal in $\mathcal{D}_{\mathcal{F}}$ and thus a preferred model.

To obtain a unique correspondence between extensions and models we define the reverse direction as follows. For some

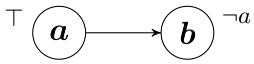

Figure 3: The ADF $\mathcal{D}_2$ from Example 2.

conflict-free set of arguments $S \subseteq \mathcal{A}$ of $\mathcal{F}$ we define the corresponding interpretation υ_S of $\mathcal{D}_\mathcal{F}$ via

$$\upsilon_S(a) = \begin{cases} \mathsf{t} & \text{if } a \in S \\ \mathsf{f} & \text{if } a\mathcal{R}S \\ \mathsf{u} & \text{otherwise} \end{cases} \quad (4)$$

Note that, according to Equation (4) the truth value f is assigned only if a *attacks* S. With that, we deviate, for good reason, from previous literature where f is assigned to an argument a if $S\mathcal{R}a$, cf. [Brewka *et al.*, 2017]. Note first that our translation is well-defined and indeed constructs an admissible model υ_S iff S is an admissible set of $\mathcal{F}$.

Proposition 1. *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF and $\mathcal{D}_\mathcal{F}$ is the corresponding ADF. $S \subseteq \mathcal{A}$ is an admissible set of $\mathcal{F}$, if and only if υ_S is an admissible model of $\mathcal{D}_\mathcal{F}$.*

Now, the reason for introducing the new translation from an admissible set to a three-valued model is rooted in the fundamental idea of initial sets, i. e., the fact that they represent a *minimal* resolution to an atomic conflict [Thimm, 2022]. Equation (4) ensures exactly that, i. e., the corresponding model υ_S will be $\leq_i$-minimal, as we will show in the proof of Theorem 1 below.

Example 3. We continue Example 2. According to Equation (4), only the extension $\{a\}$ corresponds to the unique admissible model

$$\upsilon_2 = \{\boldsymbol{a} \mapsto \mathbf{t}, b \mapsto \mathsf{u}\}$$

which is a $\leq_i$-minimal admissible model of $\mathcal{D}_2$ (excluding the trivial model υ_u).

With that, we now define the *initial models* of an ADF as those models that are admissible and minimal wrt. the information ordering $\leq_i$, excluding the model υ_u that assigns u to all arguments.

Definition 9. Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF. An interpretation $\upsilon : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ is called an *initial model* of $\mathcal{D}$, iff υ is admissible with $\upsilon \neq \upsilon_\mathsf{u}$ and there is no admissible model $\upsilon' \neq \upsilon_\mathsf{u}$ with $\upsilon' <_i \upsilon$.

We denote with $\mathsf{is}(\mathcal{D})$ the initial models of $\mathcal{D}$. It follows then from Proposition 1 and the fact that Equation (4) indeed induces a $\leq_i$-minimal model that the initial sets of any argumentation framework $\mathcal{F}$ correspond exactly to the initial models of the corresponding ADF $\mathcal{D}_\mathcal{F}$.

Theorem 1. *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF and $\mathcal{D}_\mathcal{F}$ is the corresponding ADF. Then $S \subseteq \mathcal{A}$ is an initial set of $\mathcal{F}$ if and only if υ_S is an initial model of $\mathcal{D}_\mathcal{F}$.*

Example 4. Consider the ADF $\mathcal{D}_3$ in Figure 4. The initial models of $\mathcal{D}_3$ are

$$\upsilon_1 = \{\boldsymbol{a} \mapsto \mathbf{t}, b \mapsto \mathsf{u}, c \mapsto \mathsf{u}, d \mapsto \mathsf{u}, e \mapsto \mathsf{u}\},$$
$$\upsilon_2 = \{a \mapsto \mathsf{u}, b \mapsto \mathsf{u}, \boldsymbol{c} \mapsto \mathbf{t}, \boldsymbol{d} \mapsto \mathbf{f}, e \mapsto \mathsf{u}\},$$
$$\upsilon_3 = \{a \mapsto \mathsf{u}, b \mapsto \mathsf{u}, \boldsymbol{c} \mapsto \mathbf{f}, \boldsymbol{d} \mapsto \mathbf{t}, e \mapsto \mathsf{u}\},$$
$$\upsilon_4 = \{a \mapsto \mathsf{u}, \boldsymbol{b} \mapsto \mathbf{f}, c \mapsto \mathsf{u}, d \mapsto \mathsf{u}, \boldsymbol{e} \mapsto \mathbf{t}\}.$$

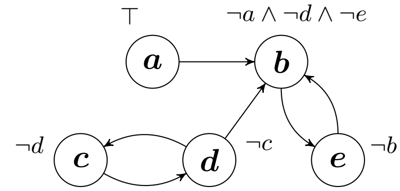

Figure 4: The ADF $\mathcal{D}_3$ from Example 4.

They correspond directly to the initial sets $\{a\}$, $\{c\}$, $\{d\}$ and $\{e\}$ of the associated AF, respectively.

As we can observe in the following example, if we do not restrict the acceptance conditions to simple attack relations, the initial models of an ADF can be quite different than the initial sets in AFs. In particular, we can have initial models in which only f and u are assigned, i. e., no argument is accepted, either because of an unsatisfiable acceptance condition or because of cyclic support relations. In fact, the latter case always leads to two initial models (assuming no further influences): one where the arguments in the support-cycle are assigned t and one where they are assigned f.

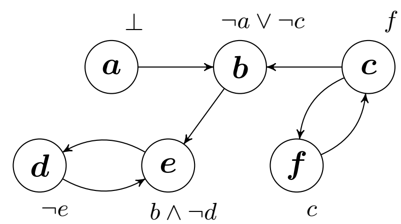

Figure 5: The ADF $\mathcal{D}_4$ from Example 5.

Example 5. Consider the ADF $\mathcal{D}_4$ in Figure 5. The initial models of $\mathcal{D}_4$ are:

$$\upsilon_1 = \{a \mapsto \mathsf{u}, b \mapsto \mathsf{u}, \boldsymbol{c} \mapsto \mathbf{t}, d \mapsto \mathsf{u}, e \mapsto \mathsf{u}, \boldsymbol{f} \mapsto \mathbf{t}\},$$
$$\upsilon_2 = \{a \mapsto \mathsf{u}, b \mapsto \mathsf{u}, \boldsymbol{c} \mapsto \mathbf{f}, d \mapsto \mathsf{u}, e \mapsto \mathsf{u}, \boldsymbol{f} \mapsto \mathbf{f}\},$$
$$\upsilon_3 = \{\boldsymbol{a} \mapsto \mathbf{f}, b \mapsto \mathsf{u}, c \mapsto \mathsf{u}, d \mapsto \mathsf{u}, e \mapsto \mathsf{u}, f \mapsto \mathsf{u}\},$$
$$\upsilon_4 = \{a \mapsto \mathsf{u}, b \mapsto \mathsf{u}, c \mapsto \mathsf{u}, \boldsymbol{d} \mapsto \mathbf{t}, \boldsymbol{e} \mapsto \mathbf{f}, f \mapsto \mathsf{u}\}.$$

To better investigate the relation between different models of an ADF, we introduce a notion of conflict between models. For two models υ_1, υ_2 of an ADF $\mathcal{D} = (\mathcal{A}, \mathcal{C})$, we say that they are *conflicting* iff there exists an argument $a \in \mathcal{A}$ such that $\upsilon_1(a) \neq \upsilon_2(a)$ with $\upsilon_1(a) \neq \mathsf{u}$ and $\upsilon_2(a) \neq \mathsf{u}$. With this notion of conflict we can now distinguish between three types of initial models for ADFs, similar to how it is done for AFs [Thimm, 2022].

Definition 10. For $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ and some initial model υ of $\mathcal{D}$, we say that

1. υ is *unattacked* iff for all $a \in \mathcal{A}$ if $\upsilon(a) \neq \mathsf{u}$, then $\phi_a \in \textsc{Taut}_\mathcal{A} \cup \textsc{Unsat}_\mathcal{A}$,
2. υ is *unchallenged* iff υ is not unattacked and there is no initial model υ' that is conflicting with υ,
3. υ is *challenged* iff there is some model υ' that is conflicting with υ and υ' is an initial model.

We denote with $\mathsf{is}^{\not\leftarrow}(\mathcal{D})$, $\mathsf{is}^{\not\leftrightarrow}(\mathcal{D})$, and $\mathsf{is}^{\leftrightarrow}(\mathcal{D})$ the set of unattacked, unchallenged, and challenged initial models, respectively. Intuitively, unattacked initial models are not involved in any conflict while (un)challenged initial models are and the latter even conflicts with some other initial model.

Example 6. We continue Example 4 with the ADF $\mathcal{D}_3$ in Figure 4. We have that υ_1 is an unattacked initial model. The models υ_2 and υ_3 are challenged initial models, both conflicting with each other. Finally, υ_4 is an unchallenged initial model of $\mathcal{D}_3$.

Example 7. We continue Example 5 with the ADF $\mathcal{D}_4$ in Figure 5. The models υ_1 and υ_2 are challenged initial models, conflicting both on c and f. υ_3 is unattacked initial and υ_4 is the only unchallenged initial model of $\mathcal{D}_4$.

Analogously to the initial sets in AFs, we have that an unattacked initial model is always a singleton. There is however an important difference, while for AFs this singleton set obviously contains only one accepted argument, in the unattacked initial models for ADFs we can have that the single argument is evaluated to either t or f.

Corollary 1. *Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF and υ is an unattacked initial model of $\mathcal{D}$. Then, we have that $|\{a \in \mathcal{A} \mid \upsilon(a) \neq \mathsf{u}\}| = 1$.*

4 Serialisability in ADFs

We now characterise admissibility for ADFs in terms of serialisation sequences, i. e. sequences of initial models. Before we consider the notion of a serialisation sequence we generalise the notion of the *reduct* in the sense of [Baumann *et al.*, 2020b] to ADFs. For some model υ, we define the υ-reduct $\mathcal{D}^\upsilon$ of $\mathcal{D}$ as the ADF where all arguments that are evaluated to t or f by υ are removed and their occurrence in the acceptance condition of some other argument is replaced by $\top$ or $\bot$ respectively. The intuition being that the reduct of an ADF $\mathcal{D}$ wrt. some model υ represents the part of the ADF that is unresolved by υ.

Definition 11. Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF and $\upsilon : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ is a three-valued interpretation. Then we define the *υ-reduct* of $\mathcal{D}$ as the ADF $\mathcal{D}^\upsilon = (\mathcal{A}', \{\phi'_a\}_{a\in\mathcal{A}'})$ where

$$\begin{aligned} \mathcal{A}' = &\quad \mathcal{A} \setminus \{a \in \mathcal{A} \mid \upsilon(a) \neq \mathsf{u}\}, \\ \mathcal{C}' = &\quad \{\phi'_a\}_{a\in\mathcal{A}'} \end{aligned}$$

with $\mathsf{x} \in \{\mathsf{t}, \mathsf{f}\}$ and $\phi'_a = \phi_a^{[b/\mathsf{x}\,:\,\upsilon(b)=\mathsf{x}]}$.

Example 8. Consider again the ADF $\mathcal{D}_3$ in Figure 4 and the model

$$\upsilon = \{\boldsymbol{a} \mapsto \mathbf{t}, b \mapsto \mathsf{u}, \boldsymbol{c} \mapsto \mathbf{f}, \boldsymbol{d} \mapsto \mathbf{t}, e \mapsto \mathsf{u}\}.$$

Then, the reduct wrt. υ is the ADF $\mathcal{D}_3^\upsilon = (\{b, e\}, \mathcal{C}')$ with

$$\phi'_b = \bot \land \neg e \quad \text{and} \quad \phi'_e = \neg b.$$

Notably, while $\mathcal{D}_3$ correspond to an AF, the reduct $\mathcal{D}_3^\upsilon$ does not, because $\phi'_b \in \textsc{Unsat}_{\mathcal{A}'}$.

It should be noted that this new definition of a reduct wrt. a three-valued model υ is different from the reduct used in the definition of the stable models in Definition 7. That definition is used for two-valued models, as the ADF where only arguments evaluated to t are retained and in the acceptance conditions all arguments that are evaluated to f are replaced by their truth value. The intention behind that approach is to verify whether a two-valued model υ is the grounded model of this reduct. In contrast to that, in our approach, where the intent is to obtain the unresolved part of the framework to iteratively compute a model for the original ADF. Meaning, using the reduct from Definition 11 with a two-valued model $\upsilon \in \mathsf{val}_2(\mathcal{D})$ will always yield the empty ADF $\mathcal{D}^\upsilon = (\emptyset, \emptyset)$.

Dual to the consensus operator $\sqcap$, we define the *union operator* $\sqcup$, which combines two *non-conflicting* interpretations υ_1 and υ_2 into an interpretation υ_3 by unifying their value assignments. So in case there is no a with $\{\upsilon_1(a), \upsilon_2(a)\} = \{\mathsf{t}, \mathsf{f}\}$, we define $\upsilon_3 = \upsilon_1 \sqcup \upsilon_2$ as

$$\upsilon_3(a) = \begin{cases} \mathsf{t} & \text{if } \upsilon_1(a) = \mathsf{t} \text{ or } \upsilon_2(a) = \mathsf{t} \\ \mathsf{f} & \text{if } \upsilon_1(a) = \mathsf{f} \text{ or } \upsilon_2(a) = \mathsf{f} \\ \mathsf{u} & \text{otherwise} \end{cases} \tag{5}$$

for all $a \in \mathcal{A}$. If υ_1 and υ_2 are in conflict (so there is a with $\{\upsilon_1(a), \upsilon_2(a)\} = \{\mathsf{t}, \mathsf{f}\}$) we leave $\upsilon_1 \sqcup \upsilon_2$ undefined.

We now define the concept of the *serialisation sequence* for ADFs as a series of initial models of the respective reducts.

Definition 12. A *serialisation sequence* for $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ is a sequence $\mathcal{Y} = (\upsilon_1, \dots \upsilon_n)$ with $\upsilon_1 \in \mathsf{is}(\mathcal{D})$ and for each $2 \leq i \leq n$ we have that $\upsilon_i \in \mathsf{is}(\mathcal{D}^{\upsilon_1 \sqcup \dots \sqcup \upsilon_{i-1}})$.

As an intermediary result we formulate a generalisation of Dung's fundamental lemma for AFs [Dung, 1995], stating that given non-conflicting admissible models υ_1, υ_2 of an ADF $\mathcal{D}$, the union $\upsilon_1 \sqcup \upsilon_2$ is itself an admissible model of $\mathcal{D}$.

Lemma 1. *Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF and υ_1, υ_2 are admissible models of $\mathcal{D}$. If υ_1 and υ_2 are not in conflict with each other, then $\upsilon_1 \sqcup \upsilon_2$ is an admissible model of $\mathcal{D}$.*

If υ_2 is an admissible model of the reduct $\mathcal{D}^{\upsilon_1}$ then υ_1 and υ_2 are non-conflicting and we can infer in that case that $\upsilon_1 \sqcup \upsilon_2$ is an admissible model of $\mathcal{D}$. The following corollary formalises this statement and is a generalisation of the modularisation property for AFs from [Baumann *et al.*, 2020a].

Corollary 2. *Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF and υ_1, υ_2 are three-valued models. If $\upsilon_1 \in \mathit{ad}(\mathcal{D})$ and $\upsilon_2 \in \mathit{ad}(\mathcal{D}^{\upsilon_1})$ then $\upsilon_1 \sqcup \upsilon_2 \in \mathit{ad}(\mathcal{D})$.*

Note that serialisability is generally more expressive than (iterative) modularisation, since the former considers initial sets (resp. models), independently of the semantics [Bengel and Thimm, 2022]. Based on the above results, we can then show that the union of all initial models υ_i in some serialisation sequence $\mathcal{Y} = (\upsilon_1, \dots \upsilon_n)$ corresponds directly to an admissible model. In particular, we can characterise the admissible models for ADFs in this way.

Theorem 2. *A serialisation sequence $\mathcal{Y} = (\upsilon_1, \dots \upsilon_n)$ induces an admissible model $\upsilon = \upsilon_1 \sqcup \dots \sqcup \upsilon_n$ and for every admissible model there is at least one such sequence.*

Note that, every serialisation sequence corresponds to exactly one admissible model, but an admissible model may have multiple corresponding serialisation sequences, each representing one particular order in which the admissible model can be constructed.

Interestingly, the serialisation sequences for ADFs provide an even more fine-grained sequence than we can obtain for AFs. Recall that ADFs enable us to have unattacked initial models that assign f to one argument and u to all others. In addition to that, the reduct for ADFs is more cautious and may contain arguments with an unsatisfiable acceptance condition. Thus, a serialisation sequence of an ADF not only provides a sequence of accepted arguments within some model (like in the case of AFs), but rather a sequence of both accepted and rejected arguments. This ties in nicely with the procedural aspect of dialectical argumentation and means that we bring this aspect to the semantical level of ADFs with our approach.

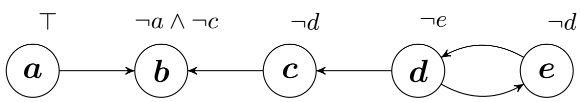

Figure 6: The ADF $\mathcal{D}_5$ from Example 9.

Example 9. Consider the ADF $\mathcal{D}_5$ in Figure 6. We have for instance the initial model

$$\upsilon_1 = \{\boldsymbol{a} \mapsto \mathbf{t}, b \mapsto \mathsf{u}, c \mapsto \mathsf{u}, d \mapsto \mathsf{u}, e \mapsto \mathsf{u}\}.$$

Now, in the reduct $\mathcal{D}_5^{\upsilon_1}$, we have $\phi'_b \in \textsc{Unsat}_{\mathcal{A}'}$ and thus the unattacked initial model

$$\upsilon_2 = \{\boldsymbol{b} \mapsto \mathbf{f}, c \mapsto \mathsf{u}, d \mapsto \mathsf{u}, e \mapsto \mathsf{u}\}.$$

Note that, in the corresponding AF $\mathcal{F}$ b would be removed in the reduct $\mathcal{F}^{\{a\}}$ since $b \in a^+$ and thus the information that b can be rejected after accepting a would not be explicitly encoded in the serialisation sequence for the AF.

In the reduct $\mathcal{D}_5^{\upsilon_1 \sqcup \upsilon_2}$ we have two initial models

$$\upsilon_3 = \{c \mapsto \mathsf{u}, \boldsymbol{d} \mapsto \mathbf{f}, \boldsymbol{e} \mapsto \mathbf{t}\},$$
$$\upsilon_4 = \{c \mapsto \mathsf{u}, \boldsymbol{d} \mapsto \mathbf{t}, \boldsymbol{e} \mapsto \mathbf{f}\}.$$

Then, we have that in the reduct $\mathcal{D}_5^{\upsilon_1 \sqcup \upsilon_2 \sqcup \upsilon_3}$ only the argument c is left with $\phi'_c = \top$ and thus the initial model $\upsilon_5 = \{\boldsymbol{c} \mapsto \mathbf{t}\}$.

It follows then that $(\upsilon_1, \upsilon_2, \upsilon_3, \upsilon_5)$ is an admissible serialisation sequence corresponding to the admissible model

$$\upsilon = \{\boldsymbol{a} \mapsto \mathbf{t}, \boldsymbol{b} \mapsto \mathbf{f}, \boldsymbol{c} \mapsto \mathbf{t}, \boldsymbol{d} \mapsto \mathbf{f}, \boldsymbol{e} \mapsto \mathbf{t}\}.$$

Intuitively, the sequence then tells us the following. We can accept a unconditionally and after that rejecting b follows logically, but not necessarily. On the other hand, when accepting e we must necessarily reject d in the same step (as evidenced by υ_3). Finally, after rejecting d the argument c can now be accepted.

With slight abuse of notation, we also have the serialisation sequence $(\upsilon_1, \upsilon_2, \upsilon_4, \upsilon_6)$ with $\upsilon_6 = \{\boldsymbol{c} \mapsto \mathbf{f}\}$ corresponding to the admissible model

$$\upsilon' = \{\boldsymbol{a} \mapsto \mathbf{t}, \boldsymbol{b} \mapsto \mathbf{f}, \boldsymbol{c} \mapsto \mathbf{f}, \boldsymbol{d} \mapsto \mathbf{t}, \boldsymbol{e} \mapsto \mathbf{f}\}.$$

Clearly there are many more admissible serialisation sequences of $\mathcal{D}_5$, for instance $(\upsilon_3, \upsilon_1, \upsilon_2)$, (υ_4, υ_6) or (υ_4).

Notably, this example shows that even for ADFs corresponding to AFs, the serialisation sequences of the ADF are more expressive than those of the corresponding AF.

5 Characterising Serialisable Semantics

Besides characterising admissible models through serialisation sequences, we can also characterise other admissibility-based semantics for ADFs. For that, we utilise the distinction between initial models introduced in Definition 10. First, we consider the preferred semantics which can be characterised by the maximal serialisation sequences, i. e., the serialisation sequences $(\upsilon_1, \dots \upsilon_n)$ where the final reduct $\mathcal{D}^{\upsilon_1 \sqcup \dots \sqcup \upsilon_n}$ possesses no further initial models.

Theorem 3. *Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF and $\upsilon : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ is an interpretation. We have that $\upsilon \in \mathit{pr}(\mathcal{D})$ if and only if there is a serialisation sequence $(\upsilon_1, \dots \upsilon_n)$ of $\mathcal{D}$ with $\upsilon = \upsilon_1 \sqcup \dots \sqcup \upsilon_n$ and it holds that $\mathit{is}(\mathcal{D}^{\upsilon_1 \sqcup \dots \sqcup \upsilon_n}) = \emptyset$.*

Example 10. We continue Example 5 with the ADF $\mathcal{D}_4$ in Figure 5. There is, for instance, the preferred serialisation sequence $(\upsilon_3, \upsilon_5, \upsilon_6, \upsilon_7)$ with

$$\upsilon_3 = \{\boldsymbol{a} \mapsto \mathbf{f}, b \mapsto \mathsf{u}, c \mapsto \mathsf{u}, d \mapsto \mathsf{u}, e \mapsto \mathsf{u}, f \mapsto \mathsf{u}\},$$
$$\upsilon_5 = \{\boldsymbol{b} \mapsto \mathbf{t}, c \mapsto \mathsf{u}, d \mapsto \mathsf{u}, e \mapsto \mathsf{u}, f \mapsto \mathsf{u}\},$$
$$\upsilon_6 = \{\boldsymbol{c} \mapsto \mathbf{f}, d \mapsto \mathsf{u}, e \mapsto \mathsf{u}, \boldsymbol{f} \mapsto \mathbf{f}\},$$
$$\upsilon_7 = \{\boldsymbol{d} \mapsto \mathbf{f}, \boldsymbol{e} \mapsto \mathbf{t}\}.$$

With slight abuse of notation, we also have the preferred serialisation sequences $(\upsilon_6, \upsilon_5, \upsilon_3, \upsilon_7)$ and $(\upsilon_6, \upsilon_5, \upsilon_7, \upsilon_3)$, all corresponding to the same preferred model of $\mathcal{D}_4$. We obtain a serialisation sequence for a different preferred model if we, for example, include the initial model $\upsilon_8 \in \mathcal{D}_4^{\upsilon_3 \sqcup \upsilon_5}$, with $\upsilon_8 = \{\boldsymbol{d} \mapsto \mathbf{t}, \boldsymbol{e} \mapsto \mathbf{f}\}$, i. e., the serialisation sequences $(\upsilon_3, \upsilon_5, \upsilon_8, \upsilon_6)$ or $(\upsilon_3, \upsilon_5, \upsilon_6, \upsilon_8)$. Finally, if we consider the initial model $\upsilon_9 = \{\boldsymbol{c} \mapsto \mathbf{t}, \boldsymbol{f} \mapsto \mathbf{t}\}$ (instead of υ_6) of $\mathcal{D}_4$ and the reduct wrt. different models, we can construct similar sequences for the two other preferred models of $\mathcal{D}_4$, e. g., $(\upsilon_9, \upsilon_3, \upsilon_5, \upsilon_7)$ or $(\upsilon_3, \upsilon_9, \upsilon_5, \upsilon_8)$.

We continue with the complete semantics, which can be characterised by the serialisation sequences that maximise only the unattacked initial models. Meaning, a serialisation sequence $(\upsilon_1, \dots \upsilon_n)$ is only complete iff there does not exist an unattacked initial model in the final reduct $\mathcal{D}^{\upsilon_1 \sqcup \dots \sqcup \upsilon_n}$. This nicely generalises the intuition behind the complete semantics that a complete extension should contain every argument defended by it. So, for ADFs, a complete model should contain all of the information (in the sense of truth value assignments) that is implied by it.

Theorem 4. *Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF and $\upsilon : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ is an interpretation. We have that $\upsilon \in \mathit{co}(\mathcal{D})$ if and only if there is a serialisation sequence $(\upsilon_1, \dots \upsilon_n)$ of $\mathcal{D}$ with $\upsilon = \upsilon_1 \sqcup \dots \sqcup \upsilon_n$ and $\mathit{is}^{\not\leftarrow}(\mathcal{D}^{\upsilon_1 \sqcup \dots \sqcup \upsilon_n}) = \emptyset$.*

For the grounded semantics, a serialisation sequence is then comprised only of unattacked initial models and there must be no further unattacked initial model left. As already mentioned, there may be multiple serialisation sequences corresponding to the unique grounded model of an ADF.

Theorem 5. *Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF and $\upsilon : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ is an interpretation. We have that $\upsilon \in \mathit{gr}(\mathcal{D})$ if and only if there is a serialisation sequence $(\upsilon_1, \dots \upsilon_n)$ of $\mathcal{D}$ with $\upsilon = \upsilon_1 \sqcup \dots \sqcup \upsilon_n$ and for all υ_i, $i = 1, \dots, n$, it holds that $\upsilon_i \in \mathit{is}^{\not\leftarrow}(\mathcal{D}^{\upsilon_1 \sqcup \dots \sqcup \upsilon_{i-1}})$ and $\mathit{is}^{\not\leftarrow}(\mathcal{D}^{\upsilon_1 \sqcup \dots \sqcup \upsilon_n}) = \emptyset$.*

Example 11. We continue Example 10 with the ADF $\mathcal{D}_4$ in Figure 5. Besides the preferred serialisation sequences, which are obviously also complete, there are, for instance, the complete serialisation sequences (v_3, v_5, v_7) and (v_3, v_5, v_8). Note that, we have $\mathsf{is}^{\not\leftarrow}(\mathcal{D}_4^{v_3 \sqcup v_5 \sqcup v_7}) = \emptyset$ since v_6 and v_9 are challenged initial models. The same holds for $\mathcal{D}_4^{v_3 \sqcup v_5 \sqcup v_8}$.

Finally, we obtain the following characterisation for the two valued models of ADFs.

Theorem 6. *Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF and $v : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ is an interpretation. We have that $v \in \mathit{val}_2(\mathcal{D})$ if and only if there is a serialisation sequence $(v_1, \dots v_n)$ of $\mathcal{D}$ with $v = v_1 \sqcup \dots \sqcup v_n$ and it holds that $\mathcal{D}^{v_1 \sqcup \dots \sqcup v_n} = (\emptyset, \emptyset)$.*

The following result for AF-like ADFs follows then directly from Theorem 6 and [Brewka *et al.*, 2013].

Corollary 3. *Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF and $\mathcal{D}_\mathcal{F}$ is the corresponding ADF. For any interpretation $v : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ the following are equivalent:*

(1) *There is a serialisation sequence $(v_1, \dots v_n)$ with $v = v_1 \sqcup \dots \sqcup v_n$ and $\mathcal{D}_\mathcal{F}^{v_1 \sqcup \dots \sqcup v_n} = (\emptyset, \emptyset)$,*

(2) $E_v \in \mathit{st}(\mathcal{F})$,

(3) $v \in \mathit{st}(\mathcal{D}_\mathcal{F})$,

(4) $v \in \mathit{val}_2(\mathcal{D}_\mathcal{F})$.

It should be noted that, in the general case, Theorem 6 only characterises the two-valued models and not the stable semantics described in Definition 7. The difference lies in the handling of self-supporting arguments and it is an open question whether the stable semantics of [Brewka *et al.*, 2013] can be characterised in terms of serialisation sequences.

6 Computational Complexity

We assume familiarity with the basic concepts of computational complexity, in particular with the basic complexity classes P, NP and coNP, cf. [Papadimitriou, 1994]. We also consider the classes Σ_2^P and Π_2^P. The class $\Sigma_2^P = \mathsf{NP}^{\mathsf{NP}}$ denotes decision problems that are solvable in polynomial time by a non-deterministic algorithm that has access to an NP-oracle and Π_2^P is the complementary class, i. e., $\Pi_2^P = \mathsf{coNP}^{\mathsf{NP}}$. Furthermore, we consider the class $\mathsf{P}_{\parallel}^{\mathsf{NP}}$ whose problems can be solved by a deterministic polynomial-time algorithm that can make polynomially many *non-adaptive* (or *parallel*) queries to an NP-oracle [Eiter and Gottlob, 1997]. Note that $\mathsf{P}_{\parallel}^{\mathsf{NP}}$ is sometimes denoted as Θ_2^P and is equivalent to $\mathsf{P}^{\mathsf{NP}[log]}$ [Papadimitriou, 1994].

We consider the following computational tasks, cf. [Strass and Wallner, 2015]:

- $\textsc{Ver}_\sigma$ Given $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ and $v : \mathcal{A} \to \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$, decide whether $v \in \sigma(\mathcal{D})$,
- $\textsc{Exists}_\sigma$ Given $\mathcal{D} = (\mathcal{A}, \mathcal{C})$, decide whether $\sigma(\mathcal{D}) \neq \emptyset$,
- $\textsc{Exists}_\sigma^{\neg\emptyset}$ Given $\mathcal{D} = (\mathcal{A}, \mathcal{C})$, decide whether there exists $v \in \sigma(\mathcal{D})$ with $v \neq v_\mathsf{u}$,
- $\textsc{Cred}_\sigma$ Given $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ and $a \in \mathcal{A}$, decide whether there exists $v \in \sigma(\mathcal{D})$ with $v(a) = \mathsf{t}$,
- $\textsc{Skept}_\sigma$ Given $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ and $a \in \mathcal{A}$, decide whether for all $v \in \sigma(\mathcal{D})$ we have $v(a) = \mathsf{t}$.

	ad($\mathcal{F}$)	is($\mathcal{F}$)	ad($\mathcal{D}$)	is($\mathcal{D}$)
$\textsc{Ver}_\sigma$	in L	in P	coNP-c	in $\mathsf{P}_{\parallel}^{\mathsf{NP}}$, coNP-h
$\textsc{Exists}_\sigma$	trivial	NP-c	trivial	Σ_2^P-c
$\textsc{Exists}_\sigma^{\neg\emptyset}$	NP-c	NP-c	Σ_2^P-c	Σ_2^P-c
$\textsc{Cred}_\sigma$	NP-c	NP-c	Σ_2^P-c	Σ_2^P-c
$\textsc{Skept}_\sigma$	trivial	coNP-c	trivial	Π_2^P-c

Table 1: Complexity of tasks related to initial models in ADFs (right two columns), in comparison to their respective counterpart for AFs (left two columns) and to the respective tasks related to ad in AFs and ADFs, cf. [Thimm, 2022; Dvorák and Dunne, 2017; Strass and Wallner, 2015]. The last column is the contribution of this work.

Theorem 7 summarises the results of our analysis, with the exact complexity of $\textsc{Ver}_{\mathsf{is}}$ still an open problem. Table 1 gives an overview over the complexity results in comparison to related problems in AFs and ADFs.

Theorem 7.

1. *$\textsc{Ver}_{\mathsf{is}}$ is in $\mathsf{P}_{\parallel}^{\mathsf{NP}}$ and coNP-hard.*
2. *$\textsc{Exists}_{\mathsf{is}}$ and $\textsc{Exists}_{\mathsf{is}}^{\neg\emptyset}$ are Σ_2^P-complete.*
3. *$\textsc{Cred}_{\mathsf{is}}$ is Σ_2^P-complete.*
4. *$\textsc{Skept}_{\mathsf{is}}$ is Π_2^P-complete.*

7 Conclusion

In this work, we defined initial models for ADFs as a notion of minimal acceptable models that resolve atomic conflicts. We showed that for ADFs that correspond to AFs the initial models coincide with the initial sets of the AF. Subsequently, we generalised the principle of serialisability of argumentation semantics to ADFs. Based on this principle we then characterised all of the characteristic operator-based semantics as well as the two-valued models for ADFs in terms of serialisation sequences, i. e., sequences of initial models of the respective reducts. These serialisation sequences then provide a fine-grained view into why the corresponding model is accepted. We concluded with some results on the computational complexity of task related to initial models.

In future work, we intend to study in more detail the relation of serialisability and the approximation fixpoint theory-based approach to ADF semantics [Strass, 2013; Strass and Wallner, 2015]. In that approach approximation operators for the characteristic operator $\Gamma_\mathcal{D}$ are analysed, whose fixpoints correspond to admissible models. These operators then also allow for a step-by-step evaluation of an ADF. There also exist further semantics for ADFs that can be considered. A promising candidate is the strongly admissible semantics of [Keshavarzi Zafarghandi *et al.*, 2022], which is serialisable for AFs. The stable semantics, as defined in [Brewka *et al.*, 2013], is a relevant candidate, but its characterisation would require a new distinction for the initial models to identify and filter out models that require self-support. Completing the complexity results for the tasks related to initial sets is left for future work. For the development of algorithms for computing initial sets, the work on splitting ADFs [Linsbichler, 2014] is of particular interest, as it provides useful results on partitioning ADFs which helps to limit the search space.

Acknowledgements

We thank Johannes P. Wallner for some discussions related to the computational complexity of the problems discussed in this paper. The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grant 550735820).

References

[Baroni *et al.*, 2018] Pietro Baroni, Martin Caminada, and Massimiliano Giacomin. Abstract argumentation frameworks and their semantics. In Pietro Baroni, Dov Gabbay, Massimiliano Giacomin, and Leendert van der Torre, editors, *Handbook of Formal Argumentation*, pages 159–236. College Publications, 2018.

[Baumann *et al.*, 2020a] Ringo Baumann, Gerhard Brewka, and Markus Ulbricht. Comparing weak admissibility semantics to their dung-style counterparts - reduct, modularization, and strong equivalence in abstract argumentation. In Diego Calvanese, Esra Erdem, and Michael Thielscher, editors, *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning, KR 2020*, pages 79–88, 2020.

[Baumann *et al.*, 2020b] Ringo Baumann, Gerhard Brewka, and Markus Ulbricht. Revisiting the foundations of abstract argumentation - semantics based on weak admissibility and weak defense. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*, pages 2742–2749. AAAI Press, 2020.

[Bengel and Thimm, 2022] Lars Bengel and Matthias Thimm. Serialisable semantics for abstract argumentation. In Francesca Toni, Sylwia Polberg, Richard Booth, Martin Caminada, and Hiroyuki Kido, editors, *Computational Models of Argument - Proceedings of COMMA 2022*, pages 80–91. IOS Press, 2022.

[Bengel *et al.*, 2024] Lars Bengel, Julian Sander, and Matthias Thimm. Characterising serialisation equivalence for abstract argumentation. In Ulle Endriss, Francisco S. Melo, Kerstin Bach, Alberto José Bugarín Diz, Jose Maria Alonso-Moral, Senén Barro, and Fredrik Heintz, editors, *ECAI 2024 - 27th European Conference on Artificial Intelligence, 2024*, volume 392 of *Frontiers in Artificial Intelligence and Applications*, pages 3340–3347. IOS Press, 2024.

[Bengel, 2022] Lars Bengel. On serialisability for argumentative explanations. *Online Handbook of Argumentation for AI: Volume 3*, abs/2212.07996, 2022.

[Brewka and Woltran, 2010] Gerhard Brewka and Stefan Woltran. Abstract dialectical frameworks. In Fangzhen Lin, Ulrike Sattler, and Miroslaw Truszczynski, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, KR 2010*. AAAI Press, 2010.

[Brewka *et al.*, 2013] Gerhard Brewka, Hannes Strass, Stefan Ellmauthaler, Johannes Peter Wallner, and Stefan Woltran. Abstract dialectical frameworks revisited. In Francesca Rossi, editor, *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, 2013*, pages 803–809. IJCAI/AAAI, 2013.

[Brewka *et al.*, 2014] Gerhard Brewka, Sylwia Polberg, and Stefan Woltran. Generalizations of dung frameworks and their role in formal argumentation. *IEEE Intelligent Systems*, 29(1):30–38, 2014.

[Brewka *et al.*, 2017] Gerhard Brewka, Stefan Ellmauthaler, Hannes Strass, Johannes Peter Wallner, and Stefan Woltran. Abstract dialectical frameworks. an overview. *Handbook of Formal Argumentation*, 1, 2017.

[Dung, 1995] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[Dvorák and Dunne, 2017] Wolfgang Dvorák and Paul E. Dunne. Computational problems in formal argumentation and their complexity. *Handbook of Formal Argumentation*, 1, 2017.

[Eiter and Gottlob, 1997] Thomas Eiter and Georg Gottlob. The complexity class theta$_2{}^{\text{p}}$: Recent results and applications in AI and modal logic. In Bogdan S. Chlebus and Ludwik Czaja, editors, *Fundamentals of Computation Theory, 11th International Symposium, FCT '97, 1997, Proceedings*, pages 1–18. Springer, 1997.

[Hage, 2000] Jaap Hage. Dialectical models in artificial intelligence and law. *Artificial Intelligence and Law*, 8(2/3):137–172, 2000.

[Keshavarzi Zafarghandi *et al.*, 2022] Atefeh Keshavarzi Zafarghandi, Rineke Verbrugge, and Bart Verheij. Strong admissibility for abstract dialectical frameworks. *Argument & Computation*, 13(3):249–289, 2022.

[Leofante *et al.*, 2024] Francesco Leofante, Hamed Ayoobi, Gabriel Freedman Adam Dejl, Deniz Gorur, Junqi Jiang, Guilherme Paulino-Passos, Antonio Rago, Anna Rapberger, Fabrizio Russo, Xiang Yin, Dekai Zhang, and Francesca Toni. Contestable AI needs Computational Argumentation. In *Proceedings of the 21st International Conference on Principles of Knowledge Representation and Reasoning (KR'24)*, 2024.

[Linsbichler, 2014] Thomas Linsbichler. Splitting abstract dialectical frameworks. In Simon Parsons, Nir Oren, Chris Reed, and Federico Cerutti, editors, *Computational Models of Argument - Proceedings of COMMA 2014, 2014*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 357–368. IOS Press, 2014.

[Papadimitriou, 1994] Christos H. Papadimitriou. *Computational complexity*. Addison-Wesley, 1994.

[Polberg *et al.*, 2013] Sylwia Polberg, Johannes Peter Wallner, and Stefan Woltran. Admissibility in the abstract dialectical framework. In João Leite, Tran Cao Son, Paolo Torroni, Leon van der Torre, and Stefan Woltran, editors, *Computational Logic in Multi-Agent Systems - 14th International Workshop, CLIMA XIV, 2013. Proceedings*, pages 102–118. Springer, 2013.

[Rescher, 1977] Nicholas Rescher. *Dialectics: A controversy-oriented approach to the theory of knowledge*. Suny Press, 1977.

[Strass and Wallner, 2015] Hannes Strass and Johannes Peter Wallner. Analyzing the computational complexity of abstract dialectical frameworks via approximation fixpoint theory. *Artificial Intelligence*, 226:34–74, 2015.

[Strass, 2013] Hannes Strass. Approximating operators and semantics for abstract dialectical frameworks. *Artificial Intelligence*, 205:39–70, 2013.

[Thimm, 2022] Matthias Thimm. Revisiting initial sets in abstract argumentation. *Argument & Computation*, 13(3):325–360, 2022.

[Verheij, 1996] Bart Verheij. Two approaches to dialectical argumentation: admissible sets and argumentation stages. *Proc. NAIC*, 96:357–368, 1996.

[Xu and Cayrol, 2018] Yuming Xu and Claudette Cayrol. Initial sets in abstract argumentation frameworks. *Journal of Applied Non-Classical Logics*, 28(2-3):260–279, 2018.