Initial Models and Serialisability in Abstract Dialectical Frameworks (Extended Abstract)

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

1 Introduction

In this work we consider the *abstract dialectical framework* (ADF) [Brewka *et al.*, 2013], which has emerged as a very powerful generalisation of the *abstract argumentation framework* (AF) [Dung, 1995]. An AF is a directed graph, where the nodes represent abstract arguments and the edges represent attacks among them. In an ADF, the relations between the arguments are more general and 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 beyond 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 sequentially 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.

An approach to address the above case is *serialisability* for AFs [Thimm, 2022]. It provides a non-deterministic construction scheme for extensions, where initial sets are selected iteratively. An initial set [Xu and Cayrol, 2018] is thereby defined as a non-empty, minimal admissible set, essentially representing a minimal semantical unit of the AF. An extension can be represented by *serialisation sequences*, i. e., sequences of initial sets that represent an order in which the extension can be built. In this work, we characterise initial models for ADFs and generalise the notion of serialisation sequences to ADFs. The full version of the presented work has been published in [Bengel and Thimm, 2025].

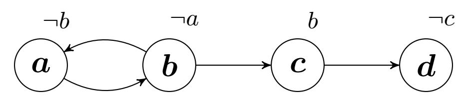

Figure 1: An ADF $\mathcal{D}$, the arguments a and b are in a conflict, while b supports c and the argument d may only be accepted if c is rejected.

2 Background

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. 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 $(a, b) \in \mathcal{R}$. A set S *defends* an argument $b \in \mathcal{A}$ iff for all a with $(a, b) \in \mathcal{R}$ there is $c \in S$ with $(c, a) \in \mathcal{R}$. Furthermore, a set S is called *admissible* (ad) iff it is conflict-free and S defends all $a \in S$. 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 $S' \neq \emptyset$.

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*.[1] Reasoning in ADFs is performed via three-valued propositional interpretations $v : \mathcal{A} \mapsto \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ that satisfy all acceptance conditions called *models*. Under three-valued interpretations a formula is then evaluated inductively as usual.

[1] 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.

We will also write $a \mapsto \mathsf{x}$ instead of $\upsilon(a) = \mathsf{x}$ and omit assignments to u for readability.

We consider the following partial order $\leq_i$: $\mathsf{u} <_i \mathsf{t}$, $\mathsf{u} <_i \mathsf{f}$ and no other pair in $<_i$. For two interpretations υ_1, υ_2 we define $\upsilon_1 \leq_i \upsilon_2$ iff $\upsilon_1(a) \leq_i \upsilon_2(a)$ for all $a \in \mathcal{A}$. For two models υ_1, υ_2, let $\upsilon_1 \sqcap \upsilon_2$ be the *consensus*, i. e., the model that takes the assignments where both υ_1 and υ_2 coincide and assigns u otherwise. For some three-valued interpretation υ, we define the set of *completions* $[\upsilon]_2$ as $[\upsilon]_2 = \{\upsilon' \mid \upsilon \leq_i \upsilon', (\upsilon')^{-1}(\mathsf{u}) = \emptyset\}$.

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 $a \in \mathcal{A}$ as

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

A model υ is then called *admissible* in $\mathcal{D}$ iff $\upsilon \leq_i \Gamma_{\mathcal{D}}(\upsilon)$.

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

Definition 2. 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 $\{\phi_a = \bigwedge_{b \in a^-} \neg b\}_{a \in \mathcal{A}}$.

3 Characterising Initial Models and Serialisation Sequences in ADFs

We now introduce *initial models* for ADFs and formulate the notion of *serialisation sequences* based on them. First, for some 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

$$\upsilon_S(a) = \begin{cases} \mathsf{t} & \text{if } a \in S \\ \mathsf{f} & \text{if } a\mathcal{R}S \\ \mathsf{u} & \text{otherwise} \end{cases} \tag{1}$$

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 3. Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF. An interpretation $\upsilon : \mathcal{A} \mapsto \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ 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 $\upsilon' <_i \upsilon$. $\mathsf{is}(\mathcal{D})$ denotes the initial models of $\mathcal{D}$.

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 1. Consider again the ADF $\mathcal{D}$ in Figure 1. There are two initial models for $\mathcal{D}$: $\upsilon_1 = \{\boldsymbol{a} \mapsto \mathbf{t}, \boldsymbol{b} \mapsto \mathbf{f}\}$ and $\upsilon_2 = \{\boldsymbol{a} \mapsto \mathbf{f}, \boldsymbol{b} \mapsto \mathbf{t}\}$. Note that for both a and b assigning them to t means that the other must necessarily be assigned f in order to have a valid admissible model. Note also that for the latter initial model υ_2, while $\upsilon_2(\phi_c) = \mathsf{t}$, due to the minimality c is still assigned u.

We now characterise admissibility for ADFs in terms of serialisation sequences, i. e. sequences of initial models. First, we generalise the notion of the *reduct* in the sense of [Baumann *et al.*, 2020] 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 4. Let $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ be an ADF and $\upsilon : \mathcal{A} \mapsto \{\mathsf{t}, \mathsf{f}, \mathsf{u}\}$ 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 $\phi'_a = \phi_a^{[b/\mathsf{x} \,:\, \upsilon(b) = \mathsf{x}]}$.

For two non-conflicting models υ_1, υ_2, let $\upsilon_1 \sqcup \upsilon_2$ be the union of υ_1 and υ_2. We now define the concept of the *serialisation sequence* for ADFs as a series of initial models of the respective reducts.

Definition 5. A *serialisation sequence* for $\mathcal{D} = (\mathcal{A}, \mathcal{C})$ is a sequence $\mathcal{Y} = (\upsilon_1, \ldots \upsilon_n)$ with $\upsilon_1 \in \mathsf{is}(\mathcal{D})$ and for each $2 \leq i \leq n$ we have that $\upsilon_i \in \mathsf{is}(\mathcal{D}^{\upsilon_1 \sqcup \cdots \sqcup \upsilon_{i-1}})$.

Based on the above results, we can then show that the union of all initial models υ_i in some serialisation sequence $\mathcal{Y} = (\upsilon_1, \ldots \upsilon_n)$ corresponds directly to an admissible model. In particular, we can characterise the admissible models for ADFs in this way.

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

Example 2. Consider again the ADF $\mathcal{D}$ in Figure 1. Consider for instance the sequence

$$(\{\boldsymbol{a} \mapsto \mathbf{t}, \boldsymbol{b} \mapsto \mathbf{f}\}, \{\boldsymbol{c} \mapsto \mathbf{f}\}, \{\boldsymbol{d} \mapsto \mathbf{t}\}).$$

As shown in Example 1, $\upsilon_1 = \{\boldsymbol{a} \mapsto \mathbf{t}, \boldsymbol{b} \mapsto \mathbf{f}\}$ is an initial model of $\mathcal{D}$. We then consider the reduct $\mathcal{D}^{\upsilon_1}$ with $\phi_c = \bot$ and $\phi_d = \neg c$. It is easy to see that after resolving a and b, the only initial model of the reduct is $\upsilon_2 = \{\boldsymbol{c} \mapsto \mathbf{f}\}$. Finally, in the reduct $\mathcal{D}^{\upsilon_1 \sqcup \upsilon_2}$, we have only $\phi_d = \top$ and thus the initial model $\upsilon_3 = \{\boldsymbol{d} \mapsto \mathbf{t}\}$. This matches exactly the intuition described in the introduction and is also the only serialisation sequence for the admissible model $\upsilon_1 \sqcup \upsilon_2 \sqcup \upsilon_3$. Alternatively there is also the unique serialisation sequence

$$(\{\boldsymbol{a} \mapsto \mathbf{f}, \boldsymbol{b} \mapsto \mathbf{t}\}, \{\boldsymbol{c} \mapsto \mathbf{t}\}, \{\boldsymbol{d} \mapsto \mathbf{f}\})$$

for the other maximal admissible model of $\mathcal{D}$.

Note that in general there can be multiple serialisation sequences for the same admissible model. By putting restrictions on the serialisation sequences, e. g., maximal length, we can then also characterise other admissibility-based semantics for ADFs.

4 Conclusion

In this work, we considered abstract dialectical frameworks and introduced the notion of initial models. Moreover, we showed that they correspond to initial sets in AFs. Finally, we generalised serialisation sequences to ADFs which allows us to characterise admissibility-based semantics in a sequence-based form. In the full paper, we also investigated the computational complexity of tasks related to initial models.

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.*, 2020] 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, 2025] Lars Bengel and Matthias Thimm. Initial models and serialisability in abstract dialectical frameworks. In *Proceedings of the Thirty-fourth International Joint Conference on Artificial Intelligence, IJCAI 2025*. ijcai.org, 2025.

[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.*, 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.

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

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

[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.