Towards Generalising Serialisability to other Argumentation Formalisms

Lars Bengel

Artificial Intelligence Group, University of Hagen

Abstract

We consider the notion of *serialisability*, a non-deterministic construction scheme for admissibility-based semantics in abstract argumentation frameworks. It provides a decomposition of the extension into a sequence of minimal acceptable sets that offers a good foundation for generating argumentative explanations. In this paper, we discuss the goal of generalising serialisability to other types of semantics, to more expressive approaches in formal argumentation, for instance abstract dialectical frameworks, and to other related formalisms.

1 Introduction

Recent advances in *artificial intelligence* (AI) have again made the need for *eXplainable AI* (XAI) more apparent. While there are many different approaches in the domain of XAI [Adadi and Berrada, 2018, Miller, 2019], the field of *formal argumentation* has been shown to provide a natural way for explaining decisions [Antaki and Leudar, 1992]. Recent works on argumentative XAI are concerned with explaining black-box AI models [Vassiliades et al., 2021] as well as providing an inherently explainable model using approaches from computational argumentation [Čyras et al., 2021].

In argumentation, the goal is to model human reasoning and discourse. The main formalism in the field is the *abstract argumentation framework* (AF) by [Dung, 1995]. Central to abstract argumentation is the semantical concept of *admissibility*, with which we define a notion of acceptance for sets of arguments (called extensions) in the sense that they must be conflict-free and defend themselves against all attackers. Many semantics have been introduced in order to refine the set of acceptable sets of arguments (see e. g. [Baroni et al., 2018] for an overview).

In this work, we will consider the notion of serialisability [Thimm, 2022]. Serialisability is a non-deterministic construction scheme for extensions with which we can characterise all classical argumentation semantics. The serialised construction yields a sequence of non-empty minimal admissible sets, each of which we can understand to resolve an atomic conflict of the AF. Generally, we obtain a more comprehensive result than just an extension, which is well suited for creating a detailed explanation for argument acceptance. For example, in the AF in Figure 1 the set $\{a, c\}$ is a complete extension, but the corresponding serialisation sequence $(\{c\}, \{a\})$ additionally tells us that c must be accepted first. Essentially, serialisability can provide an insight into the order in which arguments must be presented in order to be accepted together. Therefore, we propose to generalise serialisability to other, more expressive formalisms to utilise its explanatory capabilities in those frameworks.

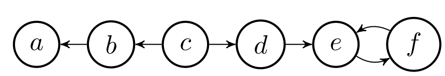

Figure 1: An AF with two initial sets $\{c\}$ and $\{f\}$.

Over the years many extensions of the original abstract AF have been proposed. For example, *bipolar AFs* where we have an additional support relation [Cayrol and Lagasquie-Schiex, 2005], *SetAFs* which are able to model collective attacks [Nielsen and Parsons, 2006] or the *abstract dialetical frameworks* [Brewka and Woltran, 2010] which associate with each argument a logical formula and subsume many other AF formalisms. Another closely related field is *structured argumentation* where an AF is induced from a logical knowledge base, e. g., *ABA* [Dung et al., 2009] or *ASPIC+* [Modgil and Prakken, 2014].

We first take a look at our previous work on the relation of serialisability to other principles [Bengel and Thimm, 2022] as well as the application of serialisability in an algorithm that employs parallel computing for enumerating extensions of AFs [Bengel and Thimm, 2023]. Following that, we examine serialisability in more detail and highlight some novel ideas on how serialisability can be generalised. For that, we consider four directions, including the generalisation to non-admissible semantics and to some of the above mentioned related argumentation formalisms.

The remainder of this work is structured as follows. In Section 2, we introduce the necessary background on abstract argumentation and serialisability. We start in Section 3 by recalling some previous work that we have done on serialisability and then discuss in more detail the idea of generalising serialisability to other semantics and formalisms. With Section 4 we conclude the paper.

2 Method

In this work we consider the *abstract argumentation framework* (AF) introduced by [Dung, 1995]. An AF is defined as a tuple $\mathsf{AF} = (\mathsf{A}, \mathsf{R})$ where A is a finite set of arguments and R is a relation $\mathsf{R} \subseteq \mathsf{A} \times \mathsf{A}$. For two arguments $a, b \in \mathsf{A}$, the relation $a\mathsf{R}b$ means that argument a attacks argument b. For a set $X \subseteq \mathsf{A}$, we denote by $\mathsf{AF}|_X = (X, \mathsf{R} \cap (X \times X))$ the projection of AF on X. For a set $S \subseteq \mathsf{A}$ we define via $S^+_{\mathsf{AF}} = \{a \in \mathsf{A} \mid \exists b \in S : b\mathsf{R}a\}$ and $S^-_{\mathsf{AF}} = \{a \in \mathsf{A} \mid \exists b \in S : a\mathsf{R}b\}$ the set of arguments attacked by S and attacking S respectively. If S is a singleton set, we omit brackets for readability. For two sets S and S' we write $S\mathsf{R}S'$ iff $S' \cap S^+_{\mathsf{AF}} \neq \emptyset$.

We say that a set $S \subseteq \mathsf{A}$ is *conflict-free* iff for all $a, b \in S$ we have that $(a, b) \notin \mathsf{R}$. Furthermore, a set S *defends* an argument $b \in \mathsf{A}$ iff for all a with $a\mathsf{R}b$ there is some $c \in S$ with $c\mathsf{R}a$. A conflict-free set S is called *admissible* if S defends all $a \in S$. Let $\mathsf{ad}(\mathsf{AF})$ denote the set of admissible sets of AF.

Non-empty minimal admissible sets have been coined *initial sets* by [Xu and Cayrol, 2016].

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

$\mathsf{IS}(\mathsf{AF})$ denotes the set of initial sets of AF. We differentiate between exactly three types of initial sets [Thimm, 2022].

Definition 2. *For* $\mathsf{AF} = (\mathsf{A}, \mathsf{R})$ *and* $S \in \mathsf{IS}(\mathsf{AF})$*, we say that*

1. *S is* unattacked *iff* $S^- = \emptyset$*,*
2. *S is* unchallenged *iff* $S^- \neq \emptyset$ *and* $\nexists S' \in \mathsf{IS}(\mathsf{AF})$ *with* $S'\mathsf{R}S$*,*
3. *S is* challenged *iff* $\exists S' \in \mathsf{IS}(\mathsf{AF})$ *with* $S'\mathsf{R}S$*.*

In the following, we will denote with $\mathsf{IS}^{\not\leftarrow}(\mathsf{AF})$, $\mathsf{IS}^{\not\leftrightarrow}(\mathsf{AF})$, and $\mathsf{IS}^{\leftrightarrow}(\mathsf{AF})$ the set of unattacked, unchallenged, and challenged initial sets, respectively. We also consider the *reduct* of an AF wrt. some set $S \subseteq \mathsf{A}$, which is computed by removing all arguments in S and those attacked by S [Baumann et al., 2020].

Definition 3. *For* $\mathsf{AF} = (\mathsf{A}, \mathsf{R})$ *and* $S \subseteq \mathsf{A}$*, the* S*-reduct* AF^S *is defined via* $\mathsf{AF}^S = \mathsf{AF}|_{A \setminus (S \cup S^+)}$*.*

Based on the notion of initial sets and the reduct, we now define a *serialisation sequence* $\mathcal{S}$ as follows [Thimm, 2022, Blümel and Thimm, 2022].

Definition 4. *A* serialisation sequence *for* $\mathsf{AF} = (\mathsf{A}, \mathsf{R})$ *is a sequence* $\mathcal{S} = (S_1, \ldots, S_n)$ *with* $S_1 \in \mathsf{IS}(\mathsf{AF})$ *and for each* $2 \leq i \leq n$ *we have that* $S_i \in \mathsf{IS}(\mathsf{AF}^{S_1 \cup \cdots \cup S_{i-1}})$*.*

As shown in [Blümel and Thimm, 2022], a serialisation sequence $(S_1, \ldots, S_n)$ induces an admissible set $E = S_1 \cup \cdots \cup S_n$ and for every admissible set there is at least one such sequence. Utilising the distinction between initial sets from Definition 2, we can restrict the serialisation sequences to characterise existing admissibility-based semantics, namely strong admissible, complete, grounded, preferred and stable semantics [Thimm, 2022]. For instance, grounded and preferred semantics are serialised as follows.

Theorem 1. *For* $\mathsf{AF} = (\mathsf{A}, \mathsf{R})$*,* $E \subseteq \mathsf{A}$*, we have that:*

- $E \in \mathit{gr}(\mathsf{AF})$ *iff there is a serialisation sequence* $(S_1, \ldots, S_n)$ *with* $E = S_1 \cup \cdots \cup S_n$ *and for all* S_i*,* $i = 1, \ldots, n$*, it holds that* $S_i \in \mathit{IS}^{\not\leftarrow}(\mathsf{AF}^{S_1 \cup \cdots \cup S_{i-1}})$ *and it holds that* $\mathit{IS}^{\not\leftarrow}(\mathsf{AF}^{S_1 \cup \cdots \cup S_n}) = \emptyset$*.*
- $E \in \mathit{pr}(\mathsf{AF})$ *iff there is a serialisation sequence* $(S_1, \ldots, S_n)$ *with* $E = S_1 \cup \cdots \cup S_n$ *and it holds that* $\mathit{IS}(\mathsf{AF}^{S_1 \cup \cdots \cup S_n}) = \emptyset$*.*

Intuitively, we can understand a serialisation sequence to explain its corresponding extension by providing a decomposition into a sequence of atomic parts that each solve an atomic conflict in the AF and together form an acceptable position in the argumentation framework.

Example 1. *Consider the AF in Figure 1. The set* $\{c\}$ *is an unattacked initial set and* $\{f\}$ *is an unchallenged initial set. We have that* $\mathcal{S}_1 = (\{c\}, \{a\})$ *is the only grounded serialisation sequence, since in the* $\{c\}$*-reduct* $\mathsf{AF}^{\{c\}} = (\{a, e, f\}, \{(e, f), (f, e)\})$ *the set* $\{a\}$ *is unattacked initial and in the reduct* $\mathsf{AF}^{\{a,c\}} = (\{e, f\}, \{(e, f), (f, e)\})$ *we have no more unattacked initial sets. On the other hand,* $\mathcal{S}_1$ *is not a preferred sequence, since* $\mathit{IS}(\mathsf{AF}^{\{a,c\}}) \neq \emptyset$*. The sequence* $\mathcal{S}_2 = (\{c\}, \{a\}, \{f\})$ *is preferred in* AF*, just like* $\mathcal{S}_3 = (\{c\}, \{a\}, \{e\})$*, since in both cases the reduct is the empty AF and thus has no initial sets. Note that* $\mathcal{S}_4 = (\{f\}, \{c\}, \{a\})$ *is also a preferred serialisation sequence, corresponding to the preferred extension* $\{a, c, f\}$*, like* $\mathcal{S}_2$*.*

3 Discussion

We first recall some of our previous work on serialisability. In [Bengel and Thimm, 2022] we investigated the relation of serialisability to other principles from the literature (see e. g. [Baroni et al., 2005] for definitions of these principles). Our analysis showed that serialisability is generally independent from SCC-recursiveness. Generally, serialisability implies Conflict-Freeness, Admissibility and Modularization. We also showed that it implies directionality, under the condition that the serialisation for the semantics is closed, i. e., every serialisation sequence is either valid for the semantics or it can be extended into a valid sequence. All serialisable semantics are closed, except stable, which is not directional. Furthermore, we took a closer look at the unchallenged semantics (uc) from [Thimm, 2022], which essentially amounts to exhaustively accepting only unattacked or unchallenged initial sets.

Theorem 2. *For* $\mathsf{AF} = (\mathsf{A}, \mathsf{R})$ *and* $E \subseteq \mathsf{A}$*, we have that* $E \in \mathit{uc}(\mathsf{AF})$ *iff there is a serialisation sequence* $(S_1, \ldots, S_n)$ *with* $E = S_1 \cup \cdots \cup S_n$ *and for all* S_i *it holds that* $S_i \in \mathit{IS}^{\not\leftarrow}(\mathsf{AF}^{S_1 \cup \cdots \cup S_{i-1}}) \cup \mathit{IS}^{\not\leftrightarrow}(\mathsf{AF}^{S_1 \cup \cdots \cup S_{i-1}})$ *and it holds that* $\mathit{IS}^{\not\leftarrow}(\mathsf{AF}^{S_1 \cup \cdots \cup S_n}) \cup \mathit{IS}^{\not\leftrightarrow}(\mathsf{AF}^{S_1 \cup \cdots \cup S_n}) = \emptyset$*.*

This semantics is closed and thus satisfies directionality, but also reinstatement. On the other hand, it does for example not satisfy SCC-recursiveness or I-maximality. Concerning the computational complexity, we showed that all tasks related to the unchallenged semantics are situated on the second level of the polynomial hierarchy and thus more complex than most semantics.

In recent work [Bengel and Thimm, 2023] we utilised the non-determinism of serialisability to develop an algorithm for enumerating extensions, on the example of the unchallenged semantics. The algorithm essentially follows multiple serialisation sequences in parallel and reuses partial results to minimise duplicate computations. Our experimental evaluation showed a significant performance increase compared to a naive approach.

We now turn to the main idea that we plan to explore in the scope of the PhD thesis. Serialisability is a very powerful concept and there are multiple useful applications for it, for example in algorithms or in generating argumentative explanations. So far, serialisability is only defined for admissibility-based semantics in abstract argumentation. We intend to generalise serialisability in order to benefit from it in other contexts as well. We consider four main domains for which we want to define serialisability and the related concepts:

(1) non-admissibility-based semantics in abstract argumentation,

(2) generalisations of abstract argumentation,

(3) structured argumentation systems,

(4) other (non-monotonic) logic formalisms.

Serialisability is based on two building blocks: initial sets and the reduct. Initial sets are inherently admissible, however there exist semantics that are not based on admissibility, such as conflict-free-based (cf) semantics [Baroni et al., 2018] and the weakly-admissible (w-adm) semantics [Baumann et al., 2020]. To characterise these semantics via serialisability, we have to adapt its building blocks. For example, we can capture some cf-based semantics relaxing to non-empty minimal conflict-free sets and modifying the S-reduct to remove further arguments. The w-adm semantics are more difficult. A promising approach would be to redefine initial sets in such a way that we allow them to be undefended against self-defeating arguments.

There exist many generalisations of the abstract argumentation framework. A very expressive formalism is the *abstract dialectical framework* (ADF) where each argument is associated with a propositional formula (called acceptance condition) [Brewka and Woltran, 2010]. With these conditions, we can represent, among other things, simple attacks on an argument, a support relation or even a collective attack on an argument, which means ADFs subsume many formalisms like AFs, BAFs and SetAFs [Brewka et al., 2013b]. Admissibility-based semantics have been defined for ADFs as well, based on a three valued logic and a generalisation of the characteristic function for AFs [Brewka et al., 2013a]. Since serialisability itself is a generalisation of the construction scheme for the grounded extension via the characteristic function [Thimm, 2022], it seems only natural to also define serialisability for ADFs.

Thirdly, we want to define serialisability for structured argumentation approaches. One prominent formalism in the field is *assumption based argumentation* (ABA), where arguments and attacks are computed from a set of logical rules and assumptions [Dung et al., 2009]. In ABA the semantics are for instance defined to return sets of assumptions instead of sets of arguments, thus defining a notion of serialisability for them may be more challenging.

Finally, we also intend to consider other non-monotonic logic formalisms, for instance *logic programming* (LP) [Gelfond and Lifschitz, 1991]. In his seminal paper [Dung, 1995], Dung already discussed how argumentation can be understood as a special form of LP. On the other hand, it has been shown that we can translate a logic program to argumentation [Caminada et al., 2015]. Moreover, there are multiple works showing a correspondence between admissibility-based semantics in argumentation and stable models in LP [Wu et al., 2009, Nieves et al., 2008].

Generally, these four targets can be pursued independently, however some domains may profit from insights gained in others, e. g., insight from definitions for structured argumentation might be relevant when looking at other logic formalisms. Additionally, definitions from (1) could also be adapted to define these semantics in the other domains.

4 Conclusion

In this work we discussed the notion of serialisability for argumentation semantics, a non-deterministic procedure for constructing extensions. We considered some previous results relating serialisability to other principles from the literature. We also discussed our work on utilising serialisability for the parallelisation of extension enumeration. For my PhD thesis, we outlined the plan to generalise serialisability to characterise other non-admissible semantics and to define serialisability for other frameworks in formal argumentation, like ADFs or ABA, in order to utilise its explanatory capabilities. Finally, we discussed the possibility of defining serialisability in the context of logic programming.

Acknowledgements

The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grant 375588274).

References

[Adadi and Berrada, 2018] Adadi, A. and Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, 6:52138–52160.

[Antaki and Leudar, 1992] Antaki, C. and Leudar, I. (1992). Explaining in conversation: Towards an argument model. *European Journal of Social Psychology*, 22(2):181–194.

[Baroni et al., 2018] Baroni, P., Caminada, M., and Giacomin, M. (2018). Abstract argumentation frameworks and their semantics. In Baroni, P., Gabbay, D., Giacomin, M., and van der Torre, L., editors, *Handbook of Formal Argumentation*, pages 159–236. College Publications.

[Baroni et al., 2005] Baroni, P., Giacomin, M., and Guida, G. (2005). SCC-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence*, 168(1–2):162–210.

[Baumann et al., 2020] Baumann, R., Brewka, G., and Ulbricht, M. (2020). Revisiting the foundations of abstract argumentation–semantics based on weak admissibility and weak defense. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2742–2749.

[Bengel and Thimm, 2022] Bengel, L. and Thimm, M. (2022). Serialisable semantics for abstract argumentation. In *Computational Models of Argument*, pages 80–91. IOS Press.

[Bengel and Thimm, 2023] Bengel, L. and Thimm, M. (2023). Towards Parallelising Extension Construction for Serialisable Semantics in Abstract Argumentation. In *Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning*, pages 732–736.

[Blümel and Thimm, 2022] Blümel, L. and Thimm, M. (2022). A ranking semantics for abstract argumentation based on serialisability. *Computational Models of Argument: Proceedings of COMMA 2022*, 353:104.

[Brewka et al., 2013a] Brewka, G., Ellmauthaler, S., Strass, H., Wallner, J. P., and Woltran, S. (2013a). Abstract dialectical frameworks revisited. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 803–809.

[Brewka et al., 2013b] Brewka, G., Polberg, S., and Woltran, S. (2013b). Generalizations of dung frameworks and their role in formal argumentation. *IEEE Intelligent Systems*, 29(1):30–38.

[Brewka and Woltran, 2010] Brewka, G. and Woltran, S. (2010). Abstract dialectical frameworks. In *Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning*, pages 102–111.

[Caminada et al., 2015] Caminada, M., Sá, S., Alcântara, J., and Dvořák, W. (2015). On the equivalence between logic programming semantics and argumentation semantics. *International Journal of Approximate Reasoning*, 58:87–111.

[Cayrol and Lagasquie-Schiex, 2005] Cayrol, C. and Lagasquie-Schiex, M.-C. (2005). On the acceptability of arguments in bipolar argumentation frameworks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 378–389. Springer.

[Čyras et al., 2021] Čyras, K., Rago, A., Albini, E., Baroni, P., and Toni, F. (2021). Argumentative xai: a survey. *arXiv preprint arXiv:2105.11266*.

[Dung, 1995] Dung, P. M. (1995). On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence*, 77(2):321–358.

[Dung et al., 2009] Dung, P. M., Kowalski, R. A., and Toni, F. (2009). Assumption-based argumentation. *Argumentation in artificial intelligence*, pages 199–218.

[Gelfond and Lifschitz, 1991] Gelfond, M. and Lifschitz, V. (1991). Classical negation in logic programs and disjunctive databases. *New generation computing*, 9:365–385.

[Miller, 2019] Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, 267:1–38.

[Modgil and Prakken, 2014] Modgil, S. and Prakken, H. (2014). The aspic+ framework for structured argumentation: a tutorial. *Argument & Computation*, 5(1):31–62.

[Nielsen and Parsons, 2006] Nielsen, S. H. and Parsons, S. (2006). A generalization of dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *International Workshop on Argumentation in Multi-Agent Systems*, pages 54–73. Springer.

[Nieves et al., 2008] Nieves, J. C., Cortés, U., and Osorio, M. (2008). Preferred extensions as stable models. *Theory and Practice of Logic Programming*, 8(4):527–543.

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

[Vassiliades et al., 2021] Vassiliades, A., Bassiliades, N., and Patkos, T. (2021). Argumentation and explainable artificial intelligence: a survey. *The Knowledge Engineering Review*, 36:e5.

[Wu et al., 2009] Wu, Y., Caminada, M., and Gabbay, D. M. (2009). Complete extensions in argumentation coincide with 3-valued stable models in logic programming. *Studia logica*, 93:383–403.

[Xu and Cayrol, 2016] Xu, Y. and Cayrol, C. (2016). Initial sets in abstract argumentation frameworks. In *Proceedings of the 1st Chinese Conference on Logic and Argumentation (CLAR'16)*, volume 1811, pages 72–85.