Graphical representations of KLM-style defeasible justifications for propositional logic

Jane Imrie[1]

[1]University of Cape Town

Abstract

KLM-style defeasible justifications, as they are understood in the literature, are displayed mainly in a textual format. This has happened regardless of the logic under consideration [1]. Though this representation is needed for their computation, this might not be the best way for users to understand why a justification holds. Other forms (particularly graphical ones) might be easier for people to understand. But before that hypothesis can be tested, we need to understand if it is even possible to represent justifications graphically in a sensible manner. Our initial idea is attempting to do so in graph format - here "graph" refers to a construction with nodes and edges. A similar notion has already been developed for abstract argumentation frameworks [2]. We plan to leverage some existing mechanisms that have already been developed in the areas of assumption-based argumentation and abstract argumentation.

Keywords

KLM, Defeasible reasoning, Assumption-based argumentation, Defeasible justifications

1. Related Work

Defeasible reasoning (sometimes referred to as "common-sense" reasoning) is a form of non-monotonic reasoning. It is essentially reasoning where the information may contain exceptions or be incomplete - humans can intuitively reason about information in this form. There are many different frameworks which attempt to formalise this type of reasoning. A well studied version is that of Kraus, Lehmann and Magidor [3]. Their approach adds a defeasible implication $\mid\sim$ to the language of propositional logic - it is essentially the defeasible counterpart to logical implication. A statement of the form "birds $\mid\sim flies$" can be understood to mean "birds typically fly". Lehamnn and Magidor [4] further defined a semantics to describe what they thought is a *rational* defeasible entailment relation, which has a corresponding algorithm named $RationalClosure$ for the computation of defeasible entailments in this paradigm. An in-depth overview of the KLM approach to defeasible reasoning can be found in [5].

There is also the notion of a *justification*, that is, the minimum amount of statements needed for an entailment to hold. Justification computation has been particularly well studied for classical description logics, and there are a number of algorithms which exist for such purposes [6]. However, there is not much research into the computation of justifications in the defeasible case, understandably, as the non-monotonicity property adds a high-degree of complexity. Only more recently has work been published on this very topic [1]. Most notably, the notion that in the defeasible case, at least for KLM, that it is possible to computre more than one type of justification: weak and strong. These two types are not inherently related i.e. not all strong justifications are extensions of weak justifications. An overview of the state of research of KLM-style justifications can be found in [1].

Finally, there is argumentation. There are many approaches to argumentation, such as abstract argumentation and assumption-based argumentation [7] (hereafter referred to as ABA for brevity). ABA is a type of structured argumentation where information is organised into rules (which can be thought of as immutable), assumptions (which can be either strict or defeasible) or contraries (basically the opposite of a given assumption). We then reason about information in terms of which statement

Doctoral Consortium of the 23rd International Conference on Principles of Knowledge Representation and Reasoning (KR 2026 DC), July 20-23, 2026, Lisbon, Portugal

imrjan001@myuct.ac.za (J. Imrie)

0000-0001-6063-8418 (J. Imrie)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-06-30

CEUR-WS.org/Vol-4218/paper2.pdf

(or sets of statements) attack another. Consider the following statements: there are people who drive their cars to get to work (rule); there are people who ride their bicycles to work (rule); cars pollute the air (assumption); bicycles do not pollute the air (assumption); air pollution is bad for the environment (rule). Informally, we could argue that people who drive their cars to work do more harm than those who ride bicyles, because cars are worse for the environment as they pollute the air and air pollution is bad. ABA frameworks with specific properties are referred to as "flat".

2. Motivation

The motivation for this research is primarily to determine if it is possible to integrate KLM defeasible reasoning with propositional logic and some form of argumentation. In KLM we are able to divide our information into ranks, with each subsequent rank containing more specific information than in the preceding rank. This allows us to discard less specific information in favour of more specific information when we perform our reasoning. If we refer back to the cars, bicyles and pollution example from the previous section 1, and adding one more statement - electric cars do not produce air pollution. We could then state that since electric cars do not produce any harmful fumes, that maybe people who drive electric cars and ride their bikes to work do less harm to the environment than than those who drive other types of cars. Since the information about electric cars is *more-specific* than that of just general cars, it would be placed in a higher rank and therefore is more valued. This notion of sorting information by specificity is a very useful one, and is not implemented in ABA. We want to integrate this and other useful mechanisms from KLM into ABA so that we can leverage the wealth of research into argument graphs already existing in the argumentation literature [2]. With the use of these highly-developed and researched graphing semantics, we can then attempt to convert justifications from textual forms that are diffcult to parse and sometimes convoluted, into something that users of explanation systems can easily digest.

3. Research plan and Progress

The plan of research is to first investigate the literature to familiarise ourselves with the different types of argumentation frameworks, then to select one that best fits our requirements. We will then attempt to combine the two formalisms, to produce a new argumentation framework that has the core underlying machinery from KLM, but adapted to fit into the context of argumentation. In other words, we want as much as possible of the benefits of KLM and argumentation, fashioned into one cohesive formalism. If that proves to be possible, we will then work on representing weak and strong justifications in the context of this new argumentation framework. If either or just one are possible, then we shall continue by leveraging existing literature on converting our new ABA-style framework justifications into forms resembling argument graphs.

We decided to select assumption-based argumentation as our foundational framework. There are numerous advantages to ABA, it is well-studied and there have been a number of extensions and modifications to ABA that have desirable properties for our particular context [7]. Namely, that of ABA^+, which allows you to specify preferences over assumptions [7], as well as simple contrapositive assumption-based argumentation, which dealt with semantics for contrapositive logics [8].

What we are working on at present is the *first* part of the problem: sensibly combining KLM-style defeasible reasoning with assumption-based argumentation.

One of the best features of rational closure is that we can discard less-specific information in favour of more specific information. We want to leverage this mechanism in assumption-based argumentation, without losing the unique advantages of ABA. That is to say, this notion of attacking, which is not present in rational closure.

In order to do this, we are using the principles of ABA^+ and simple contrapositive (SC) ABA framework (ABF), to define a new framework.

We define this new ABF in a familiar way, as a tuple, $\langle \mathscr{L}^*, \Gamma, AB, \sim, \leq \rangle$ with:

∞	$p \rightarrow b, sp \rightarrow p$
0	$b \mathrel{\mid\!\sim} w, b \mathrel{\mid\!\sim} f$
1	$p \mathrel{\mid\!\sim} \neg f$
2	$sp \mathrel{\mid\!\sim} f$

Table 1
Simple ranking

- $\mathscr{L}^*$ as our language
- Γ as our set of strict assumptions
- AB as our defeasible assumptions
- $\sim$ as the contrariness operator
- $\leq$ as our preference ordering

We define $\mathscr{L}^* = \mathscr{L} \cup \{\alpha \mathrel{\mid\!\sim} \beta \mid \alpha, \beta \in \mathscr{L}\}$. In other words, the combination of propositional logic and the defeasible connective $\mathrel{\mid\!\sim}$. Our strict assumptions will contain formulas, as well as atoms - that is, we add the antecedent of each of the formulas into our strict assumptions.

There are a number of ways to define the contrariness operator. For our context, let's say we have some defeasible assumption $\alpha \mathrel{\mid\!\sim} \beta \in AB$, we define the contrariness operator as $\sim (\alpha \mathrel{\mid\!\sim} \beta) = \alpha \wedge \neg\beta$. We thus define contrariness as "the condition stated in the assumption does not hold". Note that our choice was not influenced by any particular semantics or mathematical reason. Often, the choice of what constitutes a "contrary" statement boils down to whatever logic is being used and what makes sense for the domain of information under consideration.

In rational closure, we take our formulas and then rank them, producing a ranking as shown in table 1. We then have some query we would like to consider, e.g. "do penguins have wings?". Similar to what we do in rational closure, we want to partition the set of defeasible assumptions into ranked, disjoint, conflict-free sets. We will do this according to each assumption's level of specificity, with assumptions that have more specific information in them about the domain under consideration falling into higher ranks. This will operate in the same way as rational closure i.e. specific relative to some query. For our current example the (defeasible) information about special penguins would be in a higher rank than information about penguins.

This ranking can be produced algorithmically, by ranking the formulas according to their exceptionality in relation to the antecedent of some defeasible statement. We will use the same algorithm as it is defined in the literature for rational closure.

We also specify a preference ordering over the ranks of assumptions, $\leq$, which states that assumptions in higher ranks are preferred more than in lower ranks i.e. $\forall\ A_i$ in $\mathcal{A}$, $A_i \leq A_{i+1}$ where $0 \leq i < n$ where n is the number of defeasible assumptions. This differs to ABA^+, where the ordering is generally over single assumptions.

We then want to see if it is possible to compute some deduction for our query. If we cannot, then we return the empty set. At this stage, it does not matter if there is a set of assumptions (or multiple) which attack our query, we are simply trying to prove that with all the given information, the query can be derived. If we do find that supporting set or sets, we compute all possible attacks for all assumptions in our knowledge base. We refer to the sets of assumptions used in the attack relation as arguments i.e. an argument S attacks an assumption w. We check to see if within these arguments there is one or more assumptions $\mathcal{A}$ that satisfies one of the following criteria:

- $\mathcal{A}$ attacks our query.
- $\mathcal{A}$ attacks all of the arguments that supports our query.

If such an argument exists, then we remove the lowest rank in our set of defeasible assumptions and perform this step again. We continue in this way until there is no argument in our remaining set of assumptions that satisfy any of the above mentioned criteria. We then check again if it is possible to

compute some deduction of our query. This mirrors what we do with rational closure when we check if we can conclude the negation of the antecedent of our query and removing ranks when that is the case.

We will show this by means of a simple example. We start with some query, $sp \mid\sim w$.

For clarity, our strict assumptions are $\Gamma = \{p, sp, p \rightarrow b, sp \rightarrow p\}$. Our defeasible assumptions are $AB = \{b \mid\sim w, b \mid\sim f, p \mid\sim \neg f, sp \mid\sim f\}$. The result of the ranking process (if this was rational closure) is shown in table 1, but for our context the result would look like: $AB = \{A_0, A_1, A_2\}$ where $A_0 = \{b \mid\sim w, b \mid\sim f\}, A_1 = \{p \mid\sim \neg f\}, A_2 = \{sp \mid\sim f\}$.

Next, we see if we can find a set of assumptions that entails our query. For this example, the answer is yes - $\{sp \rightarrow p, p \rightarrow b, b \mid\sim w\}$.

Here, with $\Gamma \cup AB$ we can deduce $sp \mid\sim w$. So, now we test for conflicts. We start with the set $\Gamma \cup A_0 \cup A_1 \cup A_2$ and check if it is conflict free according to the criteria specified above.

Here are all possible attacks, with $S \mapsto x$ read as "some assumption or set of assumptions S attacks an assumption x ". Note that when we say "all possible attacks", we are only considering minimal attacks, noting that we always have to include the strict assumptions in our attacks. The list is as follows:

1. For $\Gamma \cup A_0 \cup A_1 \cup A_2$:
   a) $\{p, p \rightarrow b, p \mid\sim \neg f\} \mapsto b \mid\sim f$
   b) $\{p, sp, sp \rightarrow p, sp \mid\sim f\} \mapsto p \mid\sim \neg f$
   c) $\{p, p \rightarrow b, b \mid\sim f\} \mapsto p \mid\sim \neg f$
   d) $\{p, sp, sp \rightarrow p, p \rightarrow b, p \mid\sim \neg f\} \mapsto sp \mid\sim f$
   e) $\{p, p \rightarrow b, b \mid\sim f, p \mid\sim \neg f\} \mapsto$ every assumption (can conclude $f \wedge \neg f$)
   f) $\{sp, sp \rightarrow p, sp \mid\sim f, p \mid\sim \neg f\} \mapsto$ every assumption (can conclude $f \wedge \neg f$)
2. For $\Gamma \cup A_1 \cup A_2$:
   a) $\{p, sp, sp \rightarrow p, sp \mid\sim f\} \mapsto p \mid\sim \neg f$
   b) $\{p, p \rightarrow b, sp, sp \rightarrow p, p \mid\sim \neg f\} \mapsto sp \mid\sim f$
   c) $\{sp, sp \rightarrow p, sp \mid\sim f, p \mid\sim \neg f\} \mapsto$ every assumption (can conclude $f \wedge \neg f$)
3. For $\Gamma \cup A_2$:
   a) No conflicts.

Working through this, we first note that there are three statements in the deduction that supports our query: $\{sp \rightarrow p, p \rightarrow b, b \mid\sim w\}$. Starting with $\Gamma \cup A_0 \cup A_1 \cup A_2$ (point 1) from above, we evaluate each attack according to our two criteria. If any attack fits either one of these, then that is sufficient for the rank to be removed. We will be checking all of them to demonstrate the process but computationally of course it might make more sense to stop checking once you find an attack which fits the removal criteria. For point 1, we can see that attacks (e) and (f), which show inconsistency, attacks all arguments and therefore attack the assumptions that support our query. We see this again once A_0 is removed in point 2, that we still have inconsistency. It is only once A_0 and A_1 have been removed that the criteria are no longer satisfied. We can then compute the final entailment, which returns "no", since $b \mid\sim w$ was removed.

We can also show these attacks by means of a graph. We represent the sets of assumptions used in the attack as nodes and the edges as representing the attacks between them. See figure 1. This is the most basic representation: the nodes are the sets of strict assumptions together with some subset of the defeasible assumptions. For the edges: the black lines indicating which arguments attack each other, and the grey lines show which individual assumption is used in which argument.

From this point in the research, the plan is as follows:

1. We need to create a formal proof of equivalence between our new framework and rational closure. This is to say, we need to show that the results we get when proving/disproving an entailment in the new framework will be the same as when we perform rational closure.
2. From there, we would then consider other forms of defeasible reasoning, such as lexicographic closure and relevant closure.

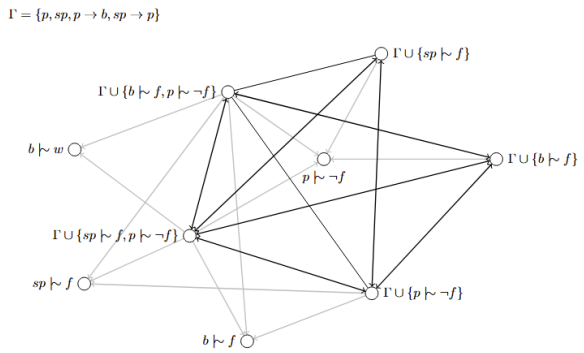

Figure 1: Graph of the attacks

3. We would then also need to determine what a justification looks like in this new framework, and how that compares to a justification in the case of what has already been researched. We want to show that with this new framework we can compute the same justifications as what can already be done in KLM for defeasible reasoning for propositional logic.
4. Finally, we can then investigate how to transform these justifications in the new framework from the written form to a useful graph form.

Declaration on Generative AI

The author has not employed any Generative AI tools.

References

[1] J. Imrie, Justifications for KLM-style defeasible reasoning, Master's thesis, University of Cape Town, Cape Town, South Africa, 2025. URL: http://hdl.handle.net/11427/42351, faculty of Science, Department of Computer Science.

[2] P. Besnard, S. Doutre, A. Herzig, Encoding argument graphs in logic, in: A. Laurent, O. Strauss, B. Bouchon-Meunier, R. R. Yager (Eds.), Information Processing and Management of Uncertainty in Knowledge-Based Systems, Springer International Publishing, Cham, Switzerland, 2014, pp. 345–354. doi:10.1007/978-3-319-08855-6_35.

[3] S. Kraus, D. Lehmann, M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, Artificial Intelligence 44 (1990) 167–207. doi:10.1016/0004-3702(90)90101-5.

[4] D. Lehmann, M. Magidor, What does a conditional knowledge base entail?, Artificial Intelligence 55 (1992) 1–60. doi:10.1016/0004-3702(92)90041-U.

[5] A. Kaliski, An overview of KLM-style defeasible entailment, Master's thesis, University of Cape Town, Cape Town, South Africa, 2020. URL: http://hdl.handle.net/11427/32743.

[6] M. Horridge, Justification based explanation in ontologies, Ph.D. thesis, University of Manchester, Manchester, UK, 2011. URL: https://www.escholar.manchester.ac.uk/uk-ac-man-scw:131699.

[7] X. Fan, K. Cyras, C. Schulz, F. Toni, Assumption-based argumentation: Disputes, explanations, preferences, IFCoLog Journal of Logics and Their Applications 4 (2017) 2407–2456.

[8] O. Arieli, J. Heyninck, Simple contrapositive assumption-based argumentation frameworks with preferences: Partial orders and collective attacks, International Journal of Approximate Reasoning 178 (2025) 109340. URL: https://doi.org/10.1016/j.ijar.2024.109340. doi:10.1016/j.ijar.2024.109340.