# CLEAR: Argumentation Framework for Constructing and Evaluating Deductive Mathematical Proofs

Nadira Boudjani, Abdelkader Gouaich, Souhila Kaci

# CLEAR: Argumentation Frameworks for Constructing and Evaluating Deductive Mathematical Proofs

Nadira BOUDJANI [a] Abdelkader GOUAICH [a] and Souhila KACI [a]
[a] *LIRMM-Montpellier, France*
{boudjani,gouaich,kaci}@lirmm.fr

**Abstract.** This paper presents a tool for constructing and evaluating deductive mathematical proofs using formal argumentation called CLEAR (**C**onstructing and eva**L**uating d**E**ductive m**A**thematical p**R**oofs). This tool has a twofold objective: (i) allows students to construct deductive proofs collaboratively using a structured argumentative debate; and (ii) helps instructors to evaluate these proofs and all intermediary steps in order to provide constructive feedbacks to students. This paper focuses on objective (i) and presents results of an experimental study conducted with undergraduate students. The behavior of students during the construction of deductive proofs is analyzed to show whether formal argumentation frameworks allow students to build deductive proofs and measure students' acceptance of CLEAR.

**Keywords.** deductive mathematical proof, argumentation frameworks, construction of proofs, argumentative debate

## 1. Introduction

Learning deductive proofs is fundamental for mathematics education [13,20]. Yet, many students have difficulties to understand and write deductive mathematical proofs which has severe consequences for problem solving as highlighted by several studies [18,24]. To tackle this problem, several approaches in mathematical didactics have used a social approach in classrooms where students are engaged in a debate and use argumentation in order to build proofs [12,14,15,23]. The term "argumentation" in this context refers to the use of informal discussions in classrooms to allow students to publicly express claims and justify them to build proofs for a given problem [2]. The underlying hypotheses are that argumentation: (i) enhances critical thinking and meta-cognitive skills such as self monitoring and self assessment; (ii) increases student's motivation by social interactions; and (iii) allows learning among students. From instructors' point of view, some difficulties arise with these approaches for assessment. In fact, the evaluation of outcomes – that includes not only the final proof but also all intermediary steps and aborted attempts – introduces an important work overhead. We hypothesis that this evaluation step of all produced data is important to capture students' misconceptions and provide them with constructive feedbacks; however, the introduced overhead can limit the acceptability by instructors of approaches based on informal argumentation.

We present a tool for constructing and evaluating deductive mathematical proofs using formal argumentation called CLEAR (**C**onstructing and eva**L**uating d**E**ductive m**A**thematical p**R**oofs) that has been outlined in [5]. CLEAR has a twofold objective: (i) allow students to build deductive mathematical proofs using structured argumentative debate; (ii) help the instructors to evaluate these proofs and assess all intermediary steps in order to identify misconceptions and provide a constructive feedback to students. Our approach uses AI argumentation frameworks to represent an argumentative debate as a graph with support and defeat relations to respectively express deduction and conflict. This graph will be analyzed using Dung's semantics to identify relevant arguments that will form the final deductive proof. The instructors will have access to this final proof. They can also get access to all steps that led to this proof in order to get more insights.

In this paper, we focus on the first objective, namely construction of deductive mathematical proofs using structured argumentative debate. Our aim is to answer the following questions: (i) are argumentation frameworks suitable to build deductive mathematical proofs? (ii) are actions used in CLEAR sufficient to build deductive mathematical proofs? (iii) how do students evaluate the usability of CLEAR?

## 2. Formal argumentation frameworks for deductive mathematical proofs

The deductive proof is one of the fundamental techniques that students need to master for mathematical problem solving. Other techniques include: proof by contrapositive, contradiction and induction. What characterizes the deductive proof is its direct and sequential process. It starts from a list of hypotheses written as "assuming ..."; followed by a sequence of deduction steps expressed as "if ... then ..."; and finally concludes with a result written as "therefore ..." In its basic form, we can model a deductive proof by the following structure $<(a_i),(d_j),c>$ where $(a_i)$ are hypotheses, $(d_j)$ is a sequence of deduction steps and $c$ represents the conclusion. As previously said, $(d_j)$ is an ordered sequence and this order is important since it makes it possible to use the conclusion of a line $d_i$ as a premise of line $d_j$ given $i<j$. If all deduction steps $(d_j)$ are accepted then $<(a_i),(d_j),c>$ is a proof for the following theorem: "$a_0$ and... $a_n$ implies $c$"

Artificial intelligence witnesses a large amount of contributions in argumentation theory. In particular Dung's argumentation framework is a pioneer work in the topic [11].

**Definition 1 (Dung's Framework)** *An* argumentation framework *(AF) is a tuple* $\langle \mathscr{A}, \mathrm{Def}\rangle$*, where* $\mathscr{A}$ *is a finite set of arguments and* $\mathrm{Def} \subseteq \mathscr{A} \times \mathscr{A}$ *is a binary defeat*[1] *relation. Given* $A,B \in \mathscr{A}$*,* $A$ Def $B$ *stands for "A defeats B".*

The outcome of Dung's AF is a set of sets of arguments, called *extensions*, that are robust against defeats. We distinguish several definitions of extension (e.g. grounded extension, preferred extensions, stable extensions), each corresponding to an acceptability semantics that formally rules the argument evaluation process. For details, see [11].

In addition of the defeat relation, several authors have considered a *support* relation [19,17,9].

[1] called *attack* in [11].

**Definition 2 (Bipolar Argumentation Framework)** *An* abstract bipolar argumentation framework *(BAF) is a tuple* $\langle \mathscr{A}, \mathrm{Def}, \mathrm{Supp} \rangle$*, where* $\langle \mathscr{A}, \mathrm{Def} \rangle$ *is Dung's AF and* $\mathrm{Supp} \subseteq \mathscr{A} \times \mathscr{A}$ *is a binary support relation. For* $A, B \in \mathscr{A}$*,* $A$ Supp $B$ *means "A supports B".*

Three interpretations have been proposed for the support relation [19,17,9]: deductive support in which "$A$ supports $B$" stands for "the acceptance of $A$ implies the acceptance of $B$", necessary support in which "$A$ supports $B$" stands for "the acceptance of $A$ is necessary for the acceptance of $B$", and evidential support. Deductive and necessary support relations show duality like classical implication. In fact, "$A$ is a deductive support for $B$" if and only if "$B$ is a necessary support for $A$". As far as this paper is concerned, we concentrate on the deductive support. This focus will be motivated later in the paper. One way to deal with BAF is to compute a new Dung's AF consisting of the set of arguments of the BAF and whose defeat relation is the defeat relation of the BAF augmented with new defeat relations. The latter are obtained by combining the support relations and defeat relations of the BAF at hand [8,4].

So far Dung's AF and its extensions have mainly considered interactions between arguments only. However in practice one may also need to defeat or support the relations between arguments. Such relations are called recursive. Different proposals have been made to deal with these relations [3,10,7]. We briefly recall the framework presented in [7] called Defeat-Support AF[2] (DSAF). In this framework, both defeat and support (which is necessary support) relate an argument and another argument, a defeat relation or a support relation. Handling DSAF consists in computing an associated BAF with necessary support relation. The obtained BAF is then transformed into a Dung's AF. Although DSAF deals with necessary support relations, it can be directly used in our setting (recall that we focus on deductive support relations) thanks to the duality between necessary and deductive support relations.

Deductive proofs come in the form "if *hypothesis* then *conclusion*". In order to translate these proofs in a formal AF we first need to provide a structure to the arguments. The prominent framework for structured argumentation is the ASPIC+ framework [16]. As far as this paper is concerned we only need a simple fragment of this framework (e.g. we do not need defeasible rules).

**Definition 3 (Argument)** *Let* $\Gamma$ *be a set of formulas constructed from a given language* $\mathscr{L}$*. An argument over* $\Gamma$ *is a pair* $A = \langle \Delta, \alpha \rangle$ *s.t. (i)* $\Delta \subseteq \Gamma$*, (ii)* $\Delta \nvdash_* \bot$*, (iii)* $\Delta \vdash_* \alpha$ *and, (iv) for all* $\Delta' \subset \Delta$*,* $\Delta' \nvdash_* \alpha$*, where* $\vdash_*$ *is the inference symbol.*

**Definition 4 (Defeat)** *Let* $A = \langle \Delta, \alpha \rangle$ *and* $B = \langle \Delta', \alpha' \rangle$ *be two arguments. We say that A undercuts B iff for some* $\phi \in \Delta'$*,* $\alpha$ *and* $\phi$ *are contradictory w.r.t. the language at hand. A rebuts B iff* $\alpha$ *and* $\alpha'$ *are contradictory. Then,* $A$ Def $B$ *iff A rebuts or undercuts B.*

The definition of the support relation depends on the context. In our setting, its semantics derives from the structure of the deductive proofs presented in Sec. 2. Roughly, a deductive proof will be encoded by a set of arguments related with the support relation. More precisely, an argument $A$ is mapped to a deduction step $d_i$. The support relation between two arguments $A$ and $B$ is interpreted as a transition between the corresponding deduction steps. Suppose that the argument A is mapped to a deduction step $d_i$ and B is mapped to a deduction step $d_j$, then $A$ supports $B$ means that $i < j$.

[2]Called Attack-Support AF in [7].

## 3. CLEAR tool

CLEAR is based on the following core elements: collaboration among students, arguments, relations and actions. For the collaboration, students build collaboratively a deductive proof by taking turns to construct arguments and relations. The arguments can be either formal or informal: formal arguments are structured following Definition 3. Checking their validity w.r.t. this definition is left to students as this is part of the problem solving task. Informal arguments are expressed as a free text. Three types of relations are available: support, defeat and append. Support relation stands for deduction and defeat stands for conflict. These relations can connect either two arguments or an argument with a relation. The append relation creates a conjoint support from two or more arguments to another argument. Finally, CLEAR makes available the following actions for students: *Add argument*, *Edit argument*, *Add relation*, *Delete relation*, *Pass turn* and *Debate end*. Since arguments represent students' reasoning, CLEAR does not allow their deletion to keep them in the graph. Only an update is available to correct input errors. To remove an argument, it has to be either defeated by another argument or not connected to the graph by any relation.

### *3.1. CLEAR process*

***Construction step*** The system provides the theorem to be proved as a formal argument and a set of propositions $\mathscr{P}$ from which formal arguments are constructed following Definition 3. $\mathscr{P}$ contains all propositions needed to prove the theorem but also frequent mistakes done by students. Since $\mathscr{P}$ contains frequent mistakes, students have to choose the right propositions to construct a correct proof. The output of the argumentative debate is a graph called *argumentative debate graph*. It contains arguments and relations that led to the deductive proof.

***Analysis step*** This is an intermediate step between the construction and the evaluation steps. The input of this step is the argumentative debate graph and the output is one or multiple proof graphs. The analysis step is done in three stages:

1. *Representation of the argumentative debate graph by an argumentation framework*: Given the argumentation graph submitted by students, a formal AF (BAF or DSAF depending on the type of interaction used by the students) is used to model the interaction between the students.
2. *Computation of Dung's AF and preferred extensions*: Dung's AF associated to the AF (BAF or DSAF) obtained in the previous step is computed. Acceptable extensions associated to Dung's AF should correspond to the proof if students succeeded to write a correct proof. At first sight, we may be tented to use the grounded extension as it corresponds to the set of arguments that are "safe". However this intuition is misleading in our setting. In fact, there is generally no single proof but multiple proofs that may be conflicting. Preferred extensions are appropriate to deal with multiple proofs.
3. *Filtering extensions and constructing the proof graphs*: It is worth noticing that a preferred extension does not necessarily correspond to a proof. For example, in order to express the fact that an argument $A_1$ is not valid, a student has to provide a defeating argument $A_2$. Now, suppose that $A_2$ is not defeated; $A_2$ belongs to the

acceptable extensions but does not contribute to the construction of the proof. Thus it must not appear in the *proof graph*. We say that an argument contributes in the construction of the proof if there is a path of support relations from that argument to the theorem to be proved. Each filtered extension gives rise to a proof graph. The latter is obtained by projecting the arguments of the extension at hand on the argumentative debate graph and the support relation relating these arguments. Each proof graph is submitted to the instructor for evaluation step.

***Evaluation step*** CLEAR is used by the instructors during the evaluating step to correct the proof graph and to provide a constructive feedback to students. In case of erroneous proofs, the instructor is able to access to the initial debate graph in order to identify students' mistakes and thus write a constructive feedback.

## 4. Experimental study

**Population:** 16 undergraduate students in computer science (6 females and 10 males) took part to the experiment.
**Choice of exercises:** Three exercises treating each a different subject have been conceived in collaboration with the teachers of mathematics in order to ensure that students had the prerequisites to understand and solve these problems.
**Procedure:** The experimental procedure comprises the following stages: (i) the 16 volunteers have been randomly grouped into 8 pairs; (ii) concepts such as formal argument, support and defeat relations are briefly introduced; (iii) the CLEAR system, available in [1], was presented; (iv) finally, all pairs began to solve the exercises with a free resolution order for a maximum duration of 2 hours.
**Dependant variables:** The observed variables are divided into two groups: the first group of variables is concerned with students' activity analysis. The second group is concerned with their perceived usability. For activity analysis, the following dependent variables are observed: number of formal and informal arguments; the number of relations; the number of actions; and the correctness of the proof. The perceived usability is evaluated through the standard Usability Scale (SUS) [6] that comprises 10 standardized items with 5 response options from *strongly agree* to *strongly disagree*. To get more insights, we have developed a specific questionnaire in order to evaluate comprehension of the argumentation concepts; expressiveness of the debate graph to build a proof; utility of actions; and finally, the usefulness of collaboration.

### *4.1. Results & Discussion*

Table 1.a presents the average and standard deviation of number of arguments and relations per exercise. Table 1.b presents the average and standard deviation of number of append, delete relation, edit argument and pass turn actions used per exercise. Table 2 presents students' average rating of a self reported questionnaire that explores different aspects such as: understanding of the argumentation concepts, the easiness of constructing formal arguments and adding relations, relevance of CLEAR's actions and the contribution of collaboration.

Now, we provide answers to the three main questions previously presented:
*i) Are argumentation frameworks suitable to build deductive proofs?* The response to this

| **(a)** The average and standard deviation of number of arguments and relations. | | | | |
|---|---|---|---|---|
| | Formal argument | Informal argument | Defeat | Support |
| Exercise 1 | 10(2.2) | 0(0) | 0.25(0.7) | 9.37(5.57) |
| Exercise 2 | 7.37(3.11) | 0.5(0.92) | 0.5(0.75) | 6.75(2.76) |
| Exercise 3 | 4.85(2.54) | 0.14(0.37) | 0(0) | 3.57(2.5) |
| **(b)** The average and standard deviation of actions. | | | | |
| | Append | Delete relation | Edit argument | Pass turn |
| Exercise 1 | 4.25(3.24) | 3(3.81) | 1.87(2.35) | 0.12(0.35) |
| Exercise 2 | 2.37(2.26) | 0.5(1.41) | 2(2.67) | 0.12(0.34) |
| Exercise 3 | 0.85(1.06) | 0.71(1.49) | 1.14(0.89) | 0.28(0.75) |

**Table 1.** The average and standard deviation of number of arguments, relations and actions per exercise.

| Item | average rating /5 |
|---|---|
| C1: Comprehension of formal argument and argumentation theory | 3,46 |
| C2: Representation of proof by graph and its visualization: arguments and relations | 3,34 |
| C3: Building formal argument by selecting premise(s) and conclusion | 3,37 |
| C4: Adding relations: support, defeat and append | 2,94 |
| C5: Importance of having append relation | 3,63 |
| C6: Importance of having edit argument action | 4,43 |
| C7: Importance of having delete relation action | 4,75 |
| C8: Importance of having pass action | 1,78 |
| C9: Importance of having informal arguments | 2,56 |
| C10: Relevance of building proofs in pair | 3,78 |

**Table 2.** The average rating of a self reported questionnaire on the different aspects of CLEAR.

question is quite important as it affects the acceptability, by students, of any argumentation based system to express deductive proofs. Besides, it is necessary to ensure that students having necessary skills to prove a result are able to express it correctly through the system. The activity analysis shows that students, without prior knowledge on argumentation theory, were able to express correct proofs (13 out of 24) using formal arguments and support relations. It is worth noticing that some proofs (6 out of 24) although not wrong, they were considered as incomplete since not all cases were considered or some intermediary results were assumed without justification. The analysis of wrong proofs shows that the argumentation concepts were correctly used and that the errors are mainly due to misunderstanding of mathematical concepts. The marks obtained for components C1 and C2 concerning comprehension and representation of the deductive proofs within the framework of the argumentation theory are 3.5/5 and 3.3/5 respectively. This is consistent with the results of activity analysis and illustrates an acceptance of this framework by students. Therefore we can conclude that the proposed tool allows students, with no prior knowledge on the argumentation, to construct deductive proofs.
*ii) Are actions used in CLEAR sufficient to build proofs?* From Table 1 we observe that the students used all actions available in CLEAR with however a much low extent for passing turn action, informal arguments and defeat relation. The students have mainly used the following actions: add formal argument, add support relation, append an existing

support relation, edit of an argument and delete of a relation. The passing turn action, informal arguments and defeat relation have been rarely used.

Concerning informal arguments, both activity analysis and questionnaire (C9) have shown that they have been rarely used. This is quite a surprise since we were expecting that building a formal argument by selecting premises and conclusion from a predefined list was more difficult than writing down a free text argument.

Most of relations are of type ’support’ and only few ‘defeat’ relations have been used. This is coherent with the fact that deductive proof building is mainly concerned with the support relations among arguments. The defeat relations are only used to express conflicts whenever they arise. All features have been judged as important (append C5, edit C6, delete C7 and build the proofs in pairs C10) except the passing turn action (C8). As a conclusion, we think that CLEAR includes all required features to build deductive proofs using an argumentation debate.
*iii) How do students evaluate the usability of CLEAR?* The system has obtained an average score of 58 on SUS scale. This represents the "low marginal" category that ranges from 50 to 62. Consequently, the usability of the system is evaluated as being "ok" by students [6].

## 5. Conclusion and perspectives

This paper presented an experimental study that shows that: (i) students, with no prior knowledge on formal argumentation, are able to build deductive proofs using formal argumentation frameworks; (ii) and the usability of CLEAR is evaluated as being "ok" on the SUS scale. To the best of our knowledge the closest work to ours is given in [21] which considers the Lakatos’s method [21,22]. Lakatos proposed a novel approach of mathematics in which mathematical reasoning is defeasible. Moreover he promoted the social processes of proof construction. In [21] the authors proposed an implementation of the Lakatos’s method in the form of a dialog game between a proponent, who aims at proving a conjecture, and an opponent, who aims at invalidating the conjecture. To prove the conjecture the proponent uses lemmas. The opponent attacks with counterexamples. The proponent defends the conjecture either by correcting the proof which leads to a modified conjecture, by showing that counterexamples are incorrect, or by modifying lemmas used to prove the conjecture. While the ideas are close to ours, our work differs from [21] in the following aspects: *(i)* the Lakatos game proposed in [21] is a persuasion dialog between a proponent and an opponent. The dialog builds on a conjecture that can be accepted or rejected at the end of the dialog. It can also be modified during the dialog. Our system is both persuasion and inquiry dialog. A group of students collaborate to prove a theorem (not a conjecture). Each student may have the role of proponent or opponent during the dialog. The output of the dialog is a proof of the theorem at hand. During the construction of the proofs, our system allow students to build informal arguments, as in [22], but also formal arguments which better structures the proof. Formal arguments make proof reasoning and relation between arguments explicit, determined and clear; *(ii)* the approach proposed in [21] has not been formalized in terms of structured arguments and formal relations (defeat and/or support); *(iii)* the output of the dialog in [22] is the acceptance of the conjecture or not. There is no evaluation of the dialog: detecting the correct arguments and relations that allowed the proof. On the contrary, the evaluation is one of the main objective of our proposition.

As future work, we will assess experimentally the acceptance and usability of CLEAR by instructors.

## References


[1] CLEAR tool. http://www.lirmm.fr/mathgame/tmp/.

[2] Andrew Aberdein and Ian J Dove. *The argument of mathematics*. Springer, 2013.

[3] Pietro Baroni, Federico Cerutti, Massimiliano Giacomin, and Giovanni Guida. Afra: Argumentation framework with recursive attacks. *International Journal of Approximate Reasoning*, 52(1):19–37, 2011.

[4] Guido Boella, Dov M Gabbay, Leon van der Torre, and Serena Villata. Support in abstract argumentation. In *COMMA'10*, pages 40–51, 2010.

[5] Nadira Boudjani, Abdelkader Gouaich, and Souhila Kaci. Debate-based learning game for constructing mathematical proofs. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 36–45. Springer, 2017.

[6] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[7] Claudette Cayrol, Andrea Cohen, and Marie-Christine Lagasquie-Schiex. Towards a new framework for recursive interactions in abstract bipolar argumentation. In *6th International Conference on Computational Models of Argument (COMMA 2016)*, pages pp–191, 2016.

[8] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Coalitions of arguments: A tool for handling bipolar argumentation frameworks. *International Journal of Intelligent Systems*, 25(1):83–109, 2010.

[9] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Bipolarity in argumentation graphs: Towards a better understanding. *IJAR*, 54(7):876–899, 2013.

[10] Andrea Cohen, Sebastian Gottifredi, Alejandro J García, and Guillermo R Simari. An approach to abstract argumentation with recursive attack and support. *Journal of Applied Logic*, 13(4):509–533, 2015.

[11] 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):321–357, 1995.

[12] Viviane Durand-Guerrier, Paolo Boero, Nadia Douek, Susanna S Epp, and Denis Tanguay. Argumentation and proof in the mathematics classroom. In *Proof and proving in mathematics education*, pages 349–367. Springer, 2011.

[13] Gila Hanna, Michael de Villiers, International Program Committee, et al. Icmi study 19: Proof and proving in mathematics education. *ZDM*, 40(2):329–336, 2008.

[14] Tricia Hedge. *Teaching and learning in the language classroom*, volume 106. Oxford University Press Oxford,, UK, 2001.

[15] Julie Kirkpatrick. The effects of argumentation on student motivation in mathematics. 2016.

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

[17] Farid Nouioua and Vincent Risch. Bipolar argumentation frameworks with specialized supports. In *ICTAI'10*, pages 215–218, 2010.

[18] OECD. PISA 2015 Results (Volume I).

[19] Nir Oren and Timothy J. Norman. Semantics for evidence-based argumentation. In *COMMA'08*, pages 276–284, 2008.

[20] Dominique Pastre. Automated theorem proving in mathematics. *Annals of Mathematics and Artificial Intelligence*, 8(3-4):425–447, 1993.

[21] Alison Pease, Katarzyna Budzynska, John Lawrence, and Chris Reed. Lakatos games for mathematical argument. In *Computational Models of Argument - Proceedings of COMMA 2014, Atholl Palace Hotel, Scottish Highlands, UK, September 9-12, 2014*, pages 59–66, 2014.

[22] Alison Pease, John Lawrence, Katarzyna Budzynska, Joseph Corneli, and Chris Reed. Lakatos-style collaborative mathematics through dialectical, structured and abstract argumentation. *Artificial Intelligence*, 246:181–219, 2017.

[23] Bettina Pedemonte. Argumentation and algebraic proof. *ZDM*, 40(3):385–400, 2008.

[24] Cedric Villani and Charles Torossian. 21 mesures pour enseignement des mathématiques. 2018.