Titolo

June 10, 2024

Contents

Chapter 1

Intro

In 2001 Pym and Harland publish a paper [3] where they propose a new way to tackle the problem of splitting sequents during linear logic proof search using boolean constraints. The paper proposes a new calculus for linear logic that associates to each formula a boolean variable, and enforces linearity by constraints on said variables. This way the complexity shifts from choosing the right set of formulas to prove a certain branch, to solving for boolean assignment – a problem for which there are much more sophisticated algorithmns.

We examine the efficiency of this method and we compare it to other provers for different substets of linear logic.

1.1 Linear logic

Linear logic is a logic proposed by Jean-Yves Girard in his seminal paper of 1987 [2]. The distintive trait of this logic is that its formulae cannot be copied (called weakening) or discarded (called contraction), but instead they are consumed. And a certain sequent it true if and only if all its formulae get consumed exactly once. For this reason this logic is sometimes called a logic of resources, in the same way classical logic is a logic of truths and intuitionistic logic is a logic of proofs.

In linear logic each connective of classical logic is doubled. To better see this let's analyze classic conjuction, this can be defined as

$$\frac{\Delta \vdash \phi_2, \Gamma \qquad \Delta \vdash \phi_1, \Gamma}{\Delta \vdash \phi_1 \wedge \phi_2, \Gamma} \qquad \frac{\Delta'' \vdash \phi_2, \Gamma'' \qquad \Delta' \vdash \phi_1, \Gamma'}{\Delta', \Delta'' \vdash \phi_1 \wedge \phi_2, \Gamma', \Gamma''}$$

On the other hand, these two rules are not equivalent in linear logic, since the former implies some weakening and contraction. This is exactly the reason why in linear logic all connectives have two versions: an additive one – where the two branches keep the same context, and a multiplicative one – where the context gets partioned between the two branches. Obviously the constants $\top$ and $\bot$ also have two versions. We have that

	Add.	Mult.
$\wedge$	$\&$	$\otimes$
$\vee$	$\oplus$	$⅋$
$\top$	$\top$	1
$\bot$	0	$\bot$

It is the multiplicative side which brings the most complexity. The action of partitioning the context – called splitting – implies an exponential number of attempts to find which subset of the multiset is right for a certain branch.

Linear logic defined as of right now, albeit having the added complexity of splitting, is nonetheless decidable: since formulae are finite and they cannot be copied, it is possible to explore all the possibilities. To make linear logic as strong as classical logic two new connectives are added: $!\phi$ and $?\phi$ – called respectively bang and why-not. These are called exponentials and their purpose is to localize uses of contraction and weakening. For example, formulas marked with ! can be used any number of times.

1.2 State of the art

Most forward provers for classic linear logic use some combination of focusing and normalization to structure their proofs, with the notable exception of llprover not using normalization. We confront our prover with two other provers: llprover (1997,) and APLL (circa 2019,).

Usually the splitting is handled in two ways: trying every partition possible, or using something called the method of input/output . The latter tries to do one branch of the proof of a multiplicative, and then feeds the remaining formulae in the sequent of the other branch.

We now give a deeper look at the provers we confront with.

1.2.1 APLL

APLL is the underlying prover of click&collect. It provides 4 different searches – forward and backwards for classic and intuitionistic linear logic. We will focus on the backwards algorithm for classic linear logic.

Before diving into the analysis of this prover, we give the definition of *why-not height*

Definition 1.2.1 (Why-not height)**.** Why-not height is the maximum number of nested "why-not"s in a formlua, or

$$\mathrm{wnh}(\phi) = \begin{cases} 0 & \text{if } \phi \in \{\bot, \top, 1, 0\} \\ \max(\mathrm{wnh}(\phi_1), \mathrm{wnh}(\phi_2)) & \text{if } \phi \in \{\phi_1 \otimes \phi_2, \phi_1 ⅋ \phi_2, \phi_1 \oplus \phi_2, \phi_1 \,\&\, \phi_2\} \\ \mathrm{wnh}(\phi_1) & \text{if } \phi \in \{\phi_1{}^{\bot}, !\phi_1\} \\ 1 + \mathrm{wnh}(\phi_1) & \text{if } \phi \in \{?\phi_1\} \end{cases}$$

This measure is used in their particular way of dealing with uncontrained formulae, but is also used as a way to decide which branch to prove first for any normal operator.

The program is written in OCaML and implements a standard focused proof search on normalized formulae as seen in [4]. In this section we will illustrate two noteworthy charateristics of its implementation:

- Sequent splitting when encountering a tensor is done by generating all the numbers up to $2^{|\Delta|}$ – where Δ is the sequent – and using the bit representation of those to create the two subsets. This can be seen in the function `split_list`, which in turn calls `split_list_aux`

```
let rec split_list_aux (acc1, acc2) l k = match l with
  | [] -> acc1, acc2
  | hd :: tl ->
      if k mod 2 = 0
      then split_list_aux (acc1, hd :: acc2) tl (k / 2)
      else split_list_aux (hd :: acc1, acc2) tl (k / 2)
```

  where the argument `k` is the number that determines the decomposition of the sequent. This function is called recursively when a tensor is encountered during proof search, starting at $k = 2^{|\Delta|}$ and decreasing by one at each iteration

```
(* ... *)
| Tensor (g, h) ->
  let rec split_gamma k =
    if k = -1 then None
    else
      let gamma1, gamma2 = split_list gamma k in
        try
          (* ... *)
        with NoValue ->
          split_gamma (k - 1)
  in
    let k = fast_exp_2 (List.length gamma) - 1 in
      (* ... *)
```

  As we will see in 4.2 this implementation choice will result in a degradation of performance on formulae with a high number of multiplicatives.

- This prover does not use a simple limit to the number of applications of the contraction rule in a branch, instead an initially empty queue of unrestricted formulae (`select_d2`) and a counter (`max_d2`) are kept during the search. Two cases arise:

  – if `select_d2 = []` and `max_d2 > 0` then the sequent of unrestricted formulae is taken, negative terms are filtered out and it is sorted based on why-not height.

```
(* ... *)
if select_d2 = [] then begin
  (if max_d2 = 0 then (bl := true; raise NoValue));
  let select_d2' =
    sort_whynot (List.filter (fun x -> not (is_neg x))
```

```
6       (Set_formula.elements theta)) in
7     if select_d2' = [] then None
8     else
9       apply_d2 select_d2' (max_d2 - 1) end
10      (* ... *)
```

This new list of unrestricted formule becomes the new `select_d2`. Otherwise if `select_d2` is still empty after being refilled (line 7) or if `max_d2` is 0 (line 3) the branch fails.

- if `select_d2` is not empty, then the first formula in the queue is extracted and added to the working set. If the branch fails the formula gets discarded and the next one in the queue is tried.

The main purpose of this whole process is to avoid infinite loops that always contract on the same formula. Instead all the formulae are tried one by one, starting from the simplest (lower why-not height).

The counter `max_d2` is a local bound, since decreasing it in a branch does not affect other branches.

1.2.2 `llprover`

`llprover` is a prover by Naoyuki Tamura. Where APLL had different provers for implicative and classical linear logic, this prover encodes all the rules as the same predicate `rule/6`, using the first argument as a selector for the system. Using classical logic as the system uses all the rules, included the ones for implicative linear logic. For this reason the prover does not implement normalization.

Another particular characteristic of llprover is that it uses a local bound with iterative deepening, so in the benchmarks for formulae which need a lot of contractions, it will perform slightly worse.

1.3 Why Prolog

Prolog as a language and as an environment has been historically tied to automated theorem proving for its ability to express these kind of algorithmns naturally. One particularly conventient characteristic of Prolog is its automatic managing of backtracking, in most other languages we would have had to use exceptions to walk down the stack, or a queue of unfinished computations, which would have made the code much less readable.

Most Prolog implementations also support CLP or constraint logic programming. This allows to have constraints referencing some attributes of variables in the body of clauses, in our case we use CLP($\mathcal{B}$) for boolean constraints and an handy interface to a sat-solver. The library exposes operators to compose boolean formulas made of prolog variables

```
X = (X =:= 1),
Y = (X * X)
```

and to check the satisfaiability of said formulas

```
?- sat(X * Y).
X = Y, Y = 1.
```

One other characteristic of Prolog which revealed to be very handy for our prover is unification. Using this we didn't have to explicitly propagate the solutions of the SAT-solver, which instead were automatically propagated between branches.

Chapter 2

The focused calculus

Before describing the calculus we must give some preliminary definitions

2.1 Normalization

Since in linear logic negation is symmetric and involutive, it is usual to work only with formulae in negated normal form.

Definition 2.1.1 (Negated Normal Form – NNF)**.** A farmula is in NNF if all its linear implications ($\multimap$) are expanded to pars ($\parr$) and negation is found only at atoms. By extension, a sequent is in NNF if all its formulae are in NNF.

A generic formula is then normalized applying recursively the DeMorgan rules for linear logic, untill NNF is reached. The process of normalization takes a two-sided judgement, of the form

$$\Delta \vdash \Gamma$$

and transforms it into a one-sided judgement

$$\vdash \Delta'$$

where the right side is composed of the normalization of Γ and $\Delta^{\perp}$.

This choice has some implementation-wise advantages, but for now we will only care about the fact that it shrinks the size of the complete calulus by roughly half since we only have to deal with the right rules of the connectives.

ϕ	$::=$	1	$\mid$	$\phi \otimes \phi$	$\mid$	$\perp$	$\mid$	$\phi \parr \phi$	(Multiplicatives and their constants)
	$\mid$	0	$\mid$	$\phi \oplus \phi$	$\mid$	$\top$	$\mid$	$\phi \,\&\, \phi$	(Additives and their constants)
	$\mid$	$!\phi$	$\mid$	$?\phi$					(Exponentials)
	$\mid$	$\alpha^{\perp}$	$\mid$	α					(Where α is a term)

Figure 2.1: Normalized linear logic formulae

As seen in Figure 2.1 we will use ϕ for formulae and α for terms.

2.2 Focusing

Focusing is a technique described by Andreoli in his seminal paper [1]. There he recognizes two alternating phases in a proof: a deterministic phase, where the order of rule application to the sequent does not matter; and a non deterministic phase, where several choices may be tried. These two phases are respectively called asynchronous and synchronous phase.

Definition 2.2.1. Given a formula ϕ we define the following predicates "ϕ asy" indicates that the rule for the toplevel connective of the formula ϕ is asynchronous on the right, these connectives are

$$⅋, \&, ?, \top, \bot$$

conversely we define "ϕ sync" to indicate that the rule for the toplevel connective of ϕ is synchronous on the right, these are

$$\otimes, \oplus, !, 1$$

Furthermore in focusing to everything is assigned a polarity of positive or negative. Connectives are negative if are asynchronous on the right, or positive if are asynchronous on the left. Terms also have a polarity, which may be assigned with some complex mechanisms. We will follow [4] and simply assing atoms with a negative polarity and negated atoms with a positive one.

Definition 2.2.2. "α neg lit" is a predicate that is true only when α is a negative literal (i.e. an atom). Conversely "α pos lit" is a predicate that is only when α is a positive literal (i.e. a negated atom).

2.3 Constraints

Our calculus uses constraints to manage the resources.

Definition 2.3.1 (Variables, expressions)**.** A boolean variable is simply a symbol to which one can associate a value of true or false. A boolean expression, in our case, is just a conjunction of possibly negated boolean variables.

Definition 2.3.2 (New)**.** Sometimes we will write

$$x \text{ new}, X \text{ new}$$

these mean respectively that:

- the variable name x has not yet occurred in any expression in the proof;
- each variable name $x_i \in X$ has not yet occurred in the proof and

  $$\forall i, j \mid i \neq j \Rightarrow x_i \neq x_j$$

  each variable in X is distinct.

$$
\begin{array}{rcll}
x & ::= & x_i \quad | \quad \overline{x_i} & \text{(Variable)} \\
e & ::= & x \wedge e \quad | \quad x & \text{(Expression)}
\end{array}
$$

Figure 2.2: Definition of a boolean variable and expression

As seen in Figure 2.2 we will call e such a conjuction and x the single boolean variables.

Definition 2.3.3 (Annotated formula)**.** Given a formula ϕ defined as in Figure 2.1 and a boolean expression e defined as in Definition 2.3.1, an *annotated formula* is simply a term

$$\text{af}(\phi, e)$$

that associates the formula to the expression. We denote

$$\exp(\text{af}(\phi, e)) = e$$

the operation of extracting the boolean expression associated to a given formula, and then extend this notation to sequents such that $\exp(\Delta)$ is the set of all boolean expressions of Δ.

We now define the constraints on these expressions:

Definition 2.3.4 (Constraints)**.** Given an annotated formula $\text{af}(\phi, e)$ as in Definition **??**, a constraint may be of two kinds

- "e used" states that the formula ϕ gets consumed in this branch of the proof, this corresponds to saying the constraint e is true or

$$x_i \wedge \cdots \wedge x_j = \top$$

- "e avail" states that the formula ϕ does not get consumed in this branch of the proof, this corresponds to saying the constraint e is not true or

$$x_i \wedge \cdots \wedge x_j = \bot$$

We then extends these predicates to sequents

$$\Delta \text{ used} = \{e_2 \text{ used} \mid e_2 \in \exp(\Delta)\}$$
$$\Delta \text{ avail} = \{e_2 \text{ avail} \mid e_2 \in \exp(\Delta)\}$$

The purpose of putting formulae and expressions together in the annotated formula is twofold:

- the actions taken on the formula determine the constraints that will be generated, and these depend on the variables associated to said formula;
- after the constraints are solved we can query the assignement of the variables and find out if the associated formula is used or not in a certain branch of a proof.

Definition 2.3.5 (Satisfaiability of constraints)**.** Given a set of constraints Λ and a function V

$$\mathrm{V} : \{x_1, x_2, \ldots, x_n\} \to \{\top, \bot\}$$

mapping variables to their value, such that

$$\Lambda = \{\ldots, e_i \text{ used}, e_j \text{ avail}, \ldots\}$$
$$\mathrm{V} = \{\ldots, (x_i, \top), (x_j, \bot), \ldots\}$$

we write $\Lambda{\downarrow}\mathrm{V}$ if

$$\bigwedge_{e \in \Lambda} e[x_1/\mathrm{V}(x_1), x_2/\mathrm{V}(x_2), \ldots] = \top$$

Here we are using the translation given in Definition 2.3.4 such that

$$e \text{ used} \Leftrightarrow e = \top$$
$$e \text{ avail} \Leftrightarrow e = \bot$$

It is worth noting that simply by rewriting the function V to

$$\mathrm{V} = \{\ldots, x_i = \top, x_j = \bot, \ldots\}$$

we get back a set of constraints.

We now expand the definition of triadic sequent of [1], by adding constraints

Definition 2.3.6 (Members of the sequent)**.** Given any sequent this can be in either two forms:

- focused or in the synchronous phase, written:

$$\vdash \Psi : \Delta \Downarrow \phi \parallel \Lambda : \mathrm{V}$$

- in the asynchronous phase, written:

$$\vdash \Psi : \Delta \Uparrow \Phi \parallel \Lambda : \mathrm{V}$$

These two have more or less the same members, which are

- the set Ψ of unrestricted formulae, or all formulae that can be freely discarded or duplicated;
- the multisets Δ and Φ of linear (annotated) formulae, these are respectively the formulas "put to the side" and the formulae which are being "worked on" during a certain moment of the asynchronous phase;
- the sets Λ and V as defined in Definition 2.3.5. This approach to constraints helps to make the flow of the variables and solutions through the proof tree more explicit and clear and leaves no ambiguity to where the constraints should be checked. The choice of letters is mainly a mnemonic or visual one, constraints Λ "go-up" the proof tree and solutions V "come down" from the leaves.

As explained in Definition 2.3.5, V may be used as a set of constraint itself; this states that a certain solution must be respected in a new one.

Definition 2.3.7 (Splitting)**.** Given a sequent of annotated formulae Δ and a set of variables X such that $|\Delta| = |X|$ we define the operation of splitting it as a function

$$\text{split}(\Delta, X) \mapsto (\Delta^L, \Delta^R)$$

where

$$\Delta^L = \{\text{af}(\phi_i, x_i \wedge e_i) \mid i \in \{1, \dots, n\}\}$$
$$\Delta^R = \{\text{af}(\phi_i, \overline{x_i} \wedge e_i) \mid i \in \{1, \dots, n\}\}$$

with n the cardinality of Δ, and ϕ_i (resp. e_i) the formula (resp. the expression) of the i-eth annotated formula in Δ using an arbitrary order. The same holds for x_i and X.

With a slight abuse of notation we will write $\text{split}(\Delta, X)_L$ and $\text{split}(\Delta, X)_R$ respectively as the left projection and the right projection of the pair (Δ_L, Δ_R).

As a small example for clarity, given the sequent

$$\Delta = \text{af}(a \otimes b, x_1), \text{af}(c^\perp, x_2)$$
$$X = \{x_3, x_4\}$$

this is split into

$$\text{split}(\Delta, X)^L \mapsto \text{af}(a \otimes b, x_3 \wedge x_1), \text{af}(c^\perp, x_4 \wedge x_2)$$
$$\text{split}(\Delta, X)^R \mapsto \text{af}(a \otimes b, \overline{x_3} \wedge x_1), \text{af}(c^\perp, \overline{x_4} \wedge x_2)$$

We are now ready to present the full calculus.

$$[\,\parr\,] \frac{\vdash \Psi : \Delta \Uparrow \text{af}(\phi_1, e), \text{af}(\phi_2, e), \Phi \parallel e \text{ used}, \Lambda : \text{V}}{\vdash \Psi : \Delta \Uparrow \text{af}(\phi_1 \parr \phi_2, e), \Phi \parallel \Lambda : \text{V}}$$

$$[\bot] \frac{\vdash \Psi : \Delta \Uparrow \Phi \parallel e \text{ used}, \Lambda : \text{V}}{\vdash \Psi : \Delta \Uparrow \text{af}(\bot, e), \Phi \parallel \Lambda : \text{V}} \qquad [\top] \frac{}{\vdash \Psi : \Delta \Uparrow \text{af}(\top, -), \Phi \parallel \Lambda : \text{V}}$$

$$[\,\&\,] \frac{\vdash \Psi : \Delta \Uparrow \text{af}(\phi_2, e), \Phi \parallel e \text{ used}, \Lambda : \text{V}' \qquad \vdash \Psi : \Delta \Uparrow \text{af}(\phi_1, e), \Phi \parallel e \text{ used}, \Lambda : \text{V}''}{\vdash \Psi : \Delta \Uparrow \text{af}(\phi_1 \,\&\, \phi_2, e), \Phi \parallel \Lambda : \text{V}', \text{V}''}$$

$$[\,?\,] \frac{\vdash \Psi, \phi : \Delta \Uparrow \Phi \parallel \Lambda : \text{V}}{\vdash \Psi : \Delta \Uparrow \text{af}(?\phi, -), \Phi \parallel \Lambda : \text{V}}$$

$$[R\Uparrow] \frac{\neg\phi \text{ asy} \qquad \vdash \Psi : \Delta, \text{af}(\phi, e) \Uparrow \Phi \parallel \Lambda : \text{V}}{\vdash \Psi : \Delta \Uparrow \text{af}(\phi, e), \Phi \parallel \Lambda : \text{V}}$$

(a) Asynchronous rules

$$[\otimes]\ \frac{X \text{ new} \qquad \vdash \Psi : \text{split}(\Delta, X)^L \Downarrow \text{af}(\phi_1, e) \parallel e \text{ used}, \Lambda : \text{V}' \qquad \vdash \Psi : \text{split}(\Delta, X)^R \Downarrow \text{af}(\phi_2, e) \parallel \text{V}' : \text{V}''}{\vdash \Psi : \Delta \Downarrow \text{af}(\phi_1 \otimes \phi_2, e) \parallel \Lambda : \text{V}''}$$

$$[\oplus_L]\ \frac{\vdash \Psi : \Delta \Downarrow \text{af}(\phi_1, e) \parallel e \text{ used}, \Lambda : \text{V}}{\vdash \Psi : \Delta \Downarrow \text{af}(\phi_1 \oplus \phi_2, e) \parallel \Lambda : \text{V}} \qquad [\oplus_R]\ \frac{\vdash \Psi : \Delta \Downarrow \text{af}(\phi_2, e) \parallel e \text{ used}, \Lambda : \text{V}}{\vdash \Psi : \Delta \Downarrow \text{af}(\phi_1 \oplus \phi_2, e) \parallel \Lambda : \text{V}}$$

$$[1]\ \frac{e_1 \text{ used}, \Delta \text{ avail}, \Lambda \downarrow \text{V}}{\vdash \Psi : \Delta \Downarrow \text{af}(1, e_1) \parallel \Lambda : \text{V}} \qquad [!]\ \frac{\vdash \Psi : \Delta \Downarrow \text{af}(\phi, e_1) \parallel e_1 \text{ used}, \Delta \text{ avail}, \Lambda : \text{V}}{\vdash \Psi : \Delta \Downarrow \text{af}(!\phi, e_1) \parallel \Lambda : \text{V}}$$

$$[R\Downarrow]\ \frac{\phi \text{ asy} \vee \phi \text{ neg lit} \qquad \vdash \Psi : \Delta \Uparrow \text{af}(\phi, e) \parallel \Lambda : \text{V}}{\vdash \Psi : \Delta \Downarrow \text{af}(\phi, e) \parallel \Lambda : \text{V}}$$

(b) Synchronous rules

$$[I_1]\ \frac{e_1 \text{ used}, e_2 \text{ used}, \Delta \text{ avail}, \Lambda \downarrow \text{V}}{\vdash \Psi : \Delta, \text{af}(\phi, e_2) \Downarrow \text{af}(\phi^{\perp}, e_1) \parallel \Lambda : \text{V}} \qquad [D_1]\ \frac{\phi \text{ pos lit} \qquad \vdash \Psi : \Delta \Downarrow \text{af}(\phi, e) \parallel \Lambda : \text{V}}{\vdash \Psi : \Delta, \text{af}(\phi, e) \Uparrow . \parallel \Lambda : \text{V}}$$

$$[I_2]\ \frac{e_1 \text{ used}, \Delta \text{ avail}, \Lambda \downarrow \text{V}}{\vdash \Psi, \phi : \Delta \Downarrow \text{af}(\phi^{\perp}, e_1) \parallel \Lambda : \text{V}} \qquad [D_2]\ \frac{\phi \text{ pos lit} \qquad e \text{ new} \qquad \vdash \Psi : \Delta \Downarrow \text{af}(\phi, e) \parallel e \text{ used}, \Lambda : \text{V}}{\vdash \Psi, \phi : \Delta \Uparrow . \parallel \Lambda : \text{V}}$$

(c) Identity and decide rules

Figure 2.3: Focused constraint calculus for Linear Logic

Chapter 3

Implementation

3.1 Formula transformations

Before beginning the proof a sequent passes through a number of transformations. These transformations both preprocess the sequent to a more convenient form, and also add information about the subformulae.

As a first transformation the sequent gets normalized into a sequent in negated normal form (NNF). NNF is the form where all negations are pushed down to atoms and all linear implications ($\multimap$) are expanded into pars ($\parr$) using the following tautology

$$a \multimap b \Leftrightarrow a^{\perp} \parr b$$

Normalization is a common technique – used in all the provers we compare with. The process is composed of just two steps

1. the left sequent is negated and appended to the right sequent, implemented by the predicate `negate_premises/3`;

2. the predicate `nnf/2`, which encodes the DeMorgan rules, is mapped recursively over the new sequent

This is possible since classic linear logic is symmetric and negation is involutive.

The purpose of this process is to cut away a great deal of possible rules applicable to the sequent, sacrificing some of the structure of the sequent. In fact the number of rules we need to implement after normalization is more than halved, since we now need just the right rules, without the ones for negation and linear implication.

As a second transformation, to each formula we assign its why-not height, as defined in Definition 1.2.1. Why-not height is used during proof search to decide which branch to do first and which exponential to decide first. This is a technique borrowed from APLL and its obvious purpose is to try first the branches with less exponentials: in the case the first probably simpler branch fails, we do not have to try the other. After this transformation formulae are attribute trees, with at each node the why not height of the subformula.

As a third and final transformation, each formula gets annotated. This means we associate a variable to each formula in the sequent as in Definition 2.3.3. Given

a sequent Δ we obtain

$$\hat{\Delta} = \{\text{af}(\phi, x) \mid x \text{ new}, \phi \in \Delta\}$$

To be clear, a variable is only assigned to the "top-level" formula, and subformulae are left unchanged.

In the implementation the concept of variable is split in two: the name of the variable – represented by a Prolog atom, and the value of the variable – represented by a Prolog variable. This is needed since, after checking the constraints, the SAT-solver unifies the variable to its value if it finds a satisfaiable solution, so the purpose of the atom is to associate the variable value to its name if the final proof. The process of annotation is implemented by the predicate `annotate/3`

```
%! annotate(+[Formulae], -[AFs], -[Constraints]) is det.
annotate(Fs, Afs, Cns) :-
  maplist([ F
          , af(F, X, Var)
          , v(Var) =:= 1
          ]>>(gensym(x, X)), Fs, Afs, Cns).
```

which is a simple map over the sequent. The constraints returned state that each formula must have its variable to one, that is to say each formulae must be used. These are the constraints the proof search starts with.

3.2 Constraint propagation

Some care is to be given to explaining how the constraints propagate. In fact, in constrast to Figure 2.3 the implementation does not have explicit propagation of the solution to the constraints. This is because Prolog's unification implicitly propagates a solution from one branch to another.

3.3 Helper predicates

We now define some helper predicates to work with the constraints. What we defined as Δ avail in Definition 2.3.4 corresponds to the predicate `set_to_zero/2`

```
%! set_to_zero(+[AFs], -[Constr]) is det.
set_to_zero(Fs, Cns) :-
    maplist([af(_, _, E), v(E) =:= 0]>>true, Fs, Cns).
```

The other helper predicate implements the split function defined in Definition 2.3.7.

```
%! split_ctx(+[AFs], -[AFs], -[AFs], -[Cns], -[Cns]) is det.
split_ctx(Afs, Pos, Neg, PCns, NCns) :-
  maplist([ af(F, N, E)
          , af(F, VarPos, Y)
          , af(F, VarNeg, Z)
          , v(Y) =:= v(X) * v(E)
          , v(Z) =:= (~ v(X)) * v(E)
          ]>>(
            gensym(x, V),
```

```
            atomic_list_concat([N, V], '.', VarPos),
            atomic_list_concat([N, V], '.~', VarNeg)
  ), Afs, Pos, Neg, PCns, NCns).
```

It is again a map over the list of formulae, that generates the new formulae and the constraints accordingly. Three new Prolog variables are introduced: `X`, `Y` and `Z`. `X` is the new variable, the annotated formulae refer to the variable `Y` and `Z`, and constraints are added so that

$$y = x \wedge e$$
$$y = \overline{x} \wedge e$$

Compare this to the original definition of Definition 2.3.7, one can see that the two are basically identical other than the fact that here the name of the variable (the atom) and its value are treated separately.

3.4 Focusing

When explaining the code we will use some common names for variables, these are

- `A` is a set of unrestricted atoms;
- `U` is a set of unrestriched formulae;
- `F`, `F1`, ..., are formulae, and `Fs` and `D` are a lists of them;
- `S` is the queue of currently usable unrestricted formulae;
- `In` is a list of constraints.

3.4.1 Asynchronous and focusing phase

During the asynchronous phase we have a list of formulae which are being worked on and a list of formulae which are put to the side. With the former being called `Fss` and the latter `D`. At each step we analyze the first element of the list `Fs`, and we keep scomposing the memebers of the list untill we cannot anymore. This process can be seen for example in the predicate for &

```
1 async(A, U, D, [F|Fs], S, M, In, _) :-
2   F = af(((F1–H1) & (F2–H2))–_), N, E), !,
3   ( H2 > H1
4   -> async(A, U, D, [af((F1–H1), N, E)|Fs], S, M, [v(E) =:= 1|In], _),
5      async(A, U, D, [af((F2–H2), N, E)|Fs], S, M, [v(E) =:= 1|In], _)
6   ;  async(A, U, D, [af((F2–H2), N, E)|Fs], S, M, [v(E) =:= 1|In], _),
7      async(A, U, D, [af((F1–H1), N, E)|Fs], S, M, [v(E) =:= 1|In], _)
8   ).
```

Here we can see both the choice being made based on the why-not height of the two subformulae, and how the with is scomposed. Compare this with the & rule in Figure 2.3. The cut at line 2 represents the main concept of the asynchronous

phase: if an asynchronous connective is encountered the only thing we ought to do is to scompose it.

If a formula cannot be further be broken apart – i.e. it is either an atom, a negated atom, or it has a toplevel synchronous connective – then it is put to the side in `D`. This can be seen in the rule `to_delta` which implements the rule $R\Uparrow$

```
async(A, U, D, [F|Fs], S, M, In, _) :-
  F = af((F1-_), _, _),
  not(is_asy(F1)), !,
  async(A, U, [F|D], Fs, S, M, In, _).
```

This process goes as long as `Fs` has still formulae inside.

When `Fs` is empty the phase switches, and the focusing process begins: we choose a formula – called `decide` – from either `D` or `U` and we scompose it untill either an asynchronous connective is left or a negated atom. This is represented by the rules D_1 and D_2 that will be discussed further ahead in Section 3.4.3.

3.4.2 Identity rules

This process of alternating asynchronous and synchronous phases in classic focusing goes on untill we have a only positive literal (in our case a negated atom) in `Fs` and the corresponding negative literal (in our case just an atom) in either `U` or `D`. When this happens the axioms – rules I_1 or I_2 – are applied to close the branch. In our case when we are focusing and we have a positive literal in `Fs`, we check if the corresponding negative literal exists in `D`. If this is true, then the variables of all the other formulae in `D` are set to zero using the predicate `set_to_zero/2` defined in Section 3.3, and the constraints are checked. This is encoded in the clause

```
focus(A, U, D, F, _, _, In, node(id_1, A, U, D, F, [])) :-
  F = af((~ T)-_), _, E1),
  is_term(T),
  select(af((T-_), _, E2), D, D1),
  set_to_zero(D1, Dz),
  append([v(E1) =:= 1, v(E2) =:= 1|Dz], In, Cns),
  check(Cns).
```

A slightly different process happens if instead a correspondence is found in `A` instead of `D`. Here `A` is a special set containing just unrestricted atoms. This is a small modification to APLL's approach based on the fact that once negative literals are put in a sequent they can never leave it, and is due to the fact that since `U` may be sorted many times, we try to keep the number of formulae in it small.

3.4.3 Decide rules

For the decide rules, particularly for D_2, we use a modified version of APLL's algorithm defined in Section 1.2.1. Like APLL we keep a queue of ordered unrestricted formulae which can be refilled only a certain number of times per-branch. This can be seen in the definition of the rule `decide_2`

```
async(A, U, D, [], [H|T], M, In, _) :-
  \+ U = [],
  gensym(x, X),
```

```
4    focus(A, U, D, af(H, X, E), T, M, [v(E) =:= 1|In], _).
5  async(A, U, D, [], [_|T], M, In, _) :-
6    \+ U = [],
7    async(A, U, D, [], T, M, In, _).
8  async(A, U, D, [], [], M, In, _) :-
9    \+ U = [],
10   refill(U, M, S, M1),
11   early_stop(A, U, D, S, M1, In, _).
```

In particular, if the queue `S` is refilled, we do not directly call `async/8`, but instead call another predicate: `early_stop/7` (line 11).

```
early_stop(_, _, _, [], _, _, _) :-
  false.
early_stop(A, U, D, [H|T], M, In, _) :-
  gensym(x, X),
  focus(A, U, D, af(H, X, E), T, M, [v(E) =:= 1|In], _).
early_stop(A, U, D, [_|T], M, In, _) :-
  early_stop(A, U, D, T, M, In, _).
```

This is due the simple fact that if the branch was not provable and we instead called directly `async/8` at line 11, we would try to refill the branch `M` times. What `early_stop/7` does is fail if the queue has just been refilled and it turns out the branch was not provable.

All the rules D_1, I_1 are defined before the unrestricted counterparts, so that they are tried first.

3.5 Building the tree

In the listings above we omitted one parameter of the predicates, which purpose is to build the proof tree. At each call of async and focus one node of the proof tree is built. This contains the context of the call. For example in

```
async(A, U, D, [F|Fs], S, M, In, node(par, A, U, D, [F|Fs], [Tree])) :-
  F = af(((F1 / F2)-_), N, E), !,
  Fs1 = [af(F1, N, E), af(F2, N, E)|Fs],
  async(A, U, D, Fs1, S, M, [v(E) =:= 1|In], Tree).
```

we can see clearly the structure of the node: a label, the context and an – optionally empty – list of subtrees. A leaf is just a node with an empty list of subtrees.

This tree can be used in the end to reconstruct the actual proof tree, by visiting it and – for each formula – querying whether its variable is set to one, and deleting it otherwise. A more sophisticated algorithm may even cancel out unwanted unrestricted formulae, that otherwise remain lingering in the sequent.

Chapter 4

Testing

!!! VERY MUCH STILL WIP !!!

4.1 Infrastructure

4.1.1 Reproducibility

The prover's tests and benchmarks are organized in a jupyter notebook. To ensure reproducibility we use Nix which is a build system based on declarative recipes, called *derivations*. A docker image is also given, which in turn calls nix without the need to install it on the system.

4.1.2 Prefix format

Since for benchmarking we will interface with a lot of different provers, each with its own syntax for expressions, the need for a common format which was easy to parse and translate arose. For this purpose we define a prefix format for linear logic formulae inspired by the format used by [5] for implicational formulae

formula	symbol
$\phi_A \otimes \phi_B$	`*AB`
$\phi_A \parr \phi_B$	`|AB`
$\phi_A \oplus \phi_B$	`+AB`
$\phi_A \,\&\, \phi_B$	`&AB`
$\phi_A \multimap \phi_B$	`@AB`
$\phi_A{}^{\perp}$	`^A`
$?\phi_A$	`?A`
$!\phi_A$	`!A`

Furthermore each single character not representing an operator is considered as a variable name. Longer names can be specified by enclosing them in single apices

as in `'varname'`. As an example we give the translation of DeMorgan for the tensor:

$$\text{trans}((a \otimes b)^{\perp} \multimap a^{\perp} \parr b^{\perp}) = \texttt{@\^{}ab|\^{}a\^{}b}$$

4.1.3 File formats

We use json as a standard format to store the tests because of its vast adoption by most programming languages. A test suite is thus defined as a list of test cases

```
TestCase ::= {
  "id": <Number>,
  "premises": [ <PrefixFormula>* ],
  "conclusions": [ <PrefixFormula>* ]
}
```

where

id
: is a number with the sole purpose of tracing back the test case from the output;

premises
: is a list of premises as prefix formulae;

conclusions
: a list of conclusions as prefix formulae.

Other arbitrary fields may be present, for example we will use the following optional fields:

thm
: whether this test case is a tautology or not, may be null;

***, &, +, ...**
: the number of times a specific connective appears in the test case;

notes
: human readable text about the test case, for example its infix representation;

size
: an indicative number of the size of the formula;

atoms
: the upper bound on the number of atoms.

4.1.4 Formula generation

One of the sources of formulae we'll use in Section 4.2 is APLL's random formula generator. The version we'll use is a slight modification of it where:

- the output is in the json format described in Section 4.1.3
- one can choose to generate normalized formulae or not;

- one can choose which connectives appear in the generated formula.

A noteworthy detail is how the parameters `size` and `atoms` mentioned in Section 4.1.3 are defined, since these are directly related to how the formulae are generated:

- when one specifies a number of atoms `atoms`, the generator initializes an array containing `atoms` atoms, their negations, and the constants $\bot, \top, \ldots$. During the generation of the formula this array is randomly accessed, choosing an element when needed. This means that `atoms` represents an upper bound to the number of different atoms that may appear in the formula, not their exact number.
- when a formula is generated, at each step it is chosen whether to generate a unary or binary connective based on a threshold:
  - if a unary connective is chosen, the process continues with a size of `size` $-$ 1;
  - if a binary connective is chosen, the program chooses a random value between 0 and `size`, and it generates the two branches of the formula, with size respectively k and `size` $- k$.

4.2 Benchmarking

We'll mainly use three sources for formulae:

- llprover's tests
- LLTP, especially the translations of Kleene's intuitionistic formulae
- randomly generated formulae made by the genrator described in 4.1.4

llprovers tests are composed mainly by simple linear logic tautologies, e.g. the DeMorgan rules, for this reason these tests are used more as a simple and fast suite to catch the most obvious bugs between iterations of the prover.

We now show the results of running the provers on two datasets: KLE-cbn and KLE-cbv, respectively the call-by-name and call-by-value translations of Kleene's theorems. These translations introduce a high number of exponentials, and this causes – other than timeouts because of the added complexity – some failures. These failures are not due to bugs, but instead happen because the prover has reached its bound for that formula. The benchmarks are done using a timeout of 60 s and a bound of 3.

Kleene CBN

prover	timeouts	failures	successes	success rate	avg. time (succ.)	avg. time (tot.)
APLL	0	16	72	≈ 0.80	0.037 s	0.055 s
llprover	20	6	62	≈ 0.70	1.709 s	3.253 s
sat-ll	4	18	66	≈ 0.75	0.130 s	0.185 s

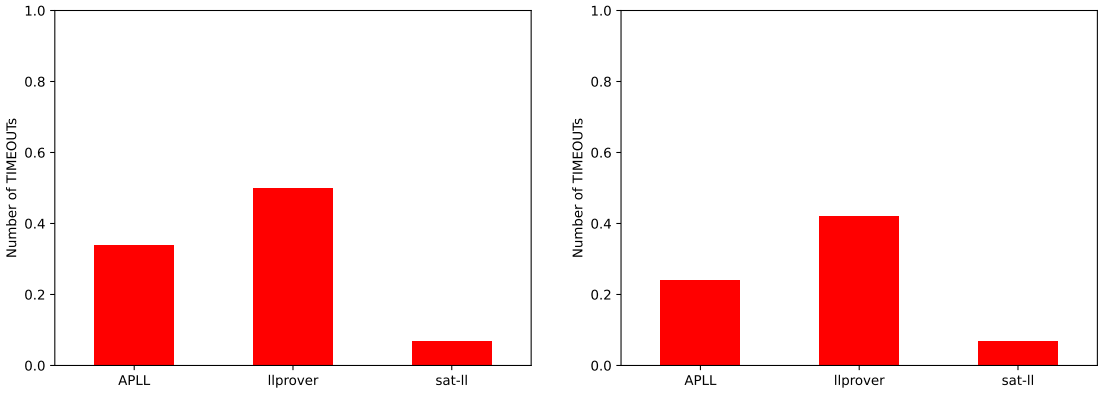
Kleene CBV

prover	timeouts	failures	successes	success rate	avg. time (succ.)	avg. time (tot.)
APLL	0	17	71	≈ 0.80	0.035 s	0.326 s
llprover	20	6	62	≈ 0.70	0.981 s	2.179 s
sat-ll	5	15	68	≈ 0.77	0.443 s	0.496 s

There are mainly two downsides with using random formulae:

- most cases do not hold any structure,
- ???

Using random formulae we can clearly see that our prover outperforms APLL (and llprover) when dealing with formulae rich in multiplicatives.



(a) Multiplicative case (b) Multiplicative and additive case

Figure 4.1: Percentage of number of timeouts out of a hundred formulae

We can see that in the multiplicative and additive case the difference begin to level. The additive case is not that significant as the formulae remain manageable and no major differences can be seen.

As soon as exponentials come into play the differences level out. It can be seen in 4.2 that in full linear logic our prover performs slightly worse than APLL.

All these test were done generating suits of 100 tests. Figure 4.1a's tests were composed of normalized formulae of `size` 100, and `atoms` 50, with just the connectives $\otimes$ and $\parr$. Figure 4.1b's tests were composed of normalized formulae formulae

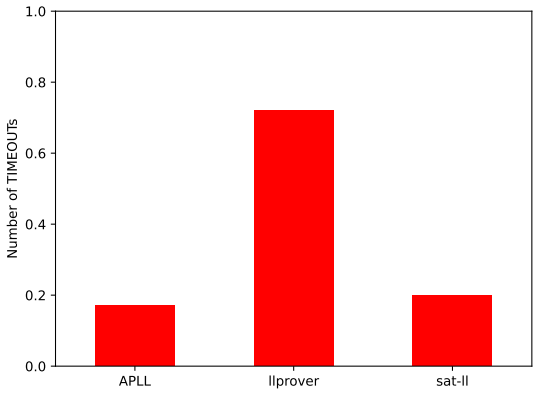


Figure 4.2: Full linear logic

of `size` 500, and `atoms` 250, with the connectives $\otimes$, $\parr$, $\&$ and $\oplus$. Finally 4.2's tests were composed of normalized formulae of `size` 500, and `atoms` 250, with all the connectives.

When looking at the results of full linear logic, it is important to note that unlike the tests with multiplicatives and additives, some of the results may be early failures because of the bound. Since llprover uses incremental search, its times are often the slowest. Similarly our prover is consistently slightly slower that APLL, this difference is negligible and due to the fact that APLL is compiled, whereas our prover is interpreted.

Bibliography

[1] Jean-Marc Andreoli. Logic programming with focusing proofs in linear logic. *J. Log. Comput.*, 2(3):297–347, 1992.

[2] Jean-Yves Girard. Linear logic. *Theor. Comput. Sci.*, 50:1–102, 1987.

[3] James Harland and David J. Pym. Resource-distribution via boolean constraints. *ACM Trans. Comput. Log.*, 4(1):56–90, 2003.

[4] Chuck C. Liang and Dale Miller. Focusing and polarization in linear, intuitionistic, and classical logics. *Theor. Comput. Sci.*, 410(46):4747–4768, 2009.

[5] Paul Tarau and Valeria de Paiva. Deriving theorems in implicational linear logic, declaratively. In Francesco Ricca, Alessandra Russo, Sergio Greco, Nicola Leone, Alexander Artikis, Gerhard Friedrich, Paul Fodor, Angelika Kimmig, Francesca A. Lisi, Marco Maratea, Alessandra Mileo, and Fabrizio Riguzzi, editors, *Proceedings 36th International Conference on Logic Programming (Technical Communications), ICLP Technical Communications 2020, (Technical Communications) UNICAL, Rende (CS), Italy, 18-24th September 2020*, volume 325 of *EPTCS*, pages 110–123, 2020.