Lecture Notes on Inversion

15-317: Constructive Logic
Frank Pfenning*

Lecture 12
October 12, 2017

1 Introduction

The contraction-free sequent calculus can be seen as describing a decision procedure for intuitionistic propositional logic. Great! But as soon as we have a sequent with many antecedents, there are many choices of rules to apply. Unless we somehow "optimize" we would have to try them all for each sequent we are trying to prove. This turns out not to be feasible except for very small examples.

Fortunately, some rules have the property that we can *always* apply them without having to consider alternatives. Loosely speaking, this is because whenever the conclusion is provable, so are all the premises. We call such rules *invertible* and the proof search strategy that first applies all such rules *inversion*. In this lecture we develop a calculus in which inversion is "built-in" in the sense that the only legal deductions are those that do apply inversions eagerly.

2 Invertible Rules

The restrictive sequent calculus in the previous section is a big improvement, but if we use it directly to implement a search procedure it is hopelessly inefficient. The problem is that for any goal sequent, any left or right rule might be applicable. But the application of a rule changes the sequent

*With edits by André Platzer

just a little—most formulas are preserved and we are faced with the same choices at the next step. Eliminating this kind of inefficiency is crucial for a practical theorem proving procedure.

The first observation, to be refined later, is that certain rules are *invertible*, that is, the premises hold iff the conclusion holds. This is powerful, because we can apply the rule and never look back and consider any other choice.

As an example of an invertible rule, consider $\wedge R$ again:

$$\frac{\Gamma \longrightarrow A \quad \Gamma \longrightarrow B}{\Gamma \longrightarrow A \wedge B}\ \wedge R$$

The premises already imply the conclusion since the rule is sound. So for $\wedge R$ to be invertible means that if the conclusion holds then both premises hold as well. That is, we have to show: *If* $\Gamma \longrightarrow A \wedge B$ *then* $\Gamma \longrightarrow A$ *and* $\Gamma \longrightarrow B$, which is the opposite of what the rule itself expresses. Fortunately, this follows easily by cut, since $\Gamma, A \wedge B \longrightarrow A$ and $\Gamma, A \wedge B \longrightarrow B$.

$$\dfrac{\Gamma \longrightarrow A \wedge B \qquad \dfrac{\dfrac{}{\Gamma, A, B \longrightarrow A}\ \mathsf{id}}{\Gamma, A \wedge B \longrightarrow A}\ \wedge L}{\Gamma \longrightarrow A}\ \mathsf{cut}$$

In order to formalize the strategy of applying inversions eagerly, without backtracking over the choices of which invertible rules to try, we refine the restricted sequent calculus further into two, mutually dependent forms of sequents.

$$\begin{array}{ll} \Gamma^- \ ; \Omega \xrightarrow{R} C & \text{Decompose } C \text{ on the right} \\ \Gamma^- \ ; \Omega \xrightarrow{L} C^+ & \text{Decompose } \Omega \text{ on the left} \end{array}$$

Here, Ω is an *ordered context* (say, a stack) that we only access at the right end. To make this stand out in the notation, we write $\Omega \cdot A$ instead of Ω, A when building up or decomposing ordered contexts.

Γ^- is a context restricted to those formulas whose left rules are *not* invertible, and C^+ is a formula whose right rule is *not* invertible. Both types of sequents can also contain atoms.

Only left decompositions $\Gamma^- \ ; \Omega \xrightarrow{L} C^+$ are restricted to have a formula with a connective of a non-invertible right-rule. Right decompositions $\Gamma^- \ ; \Omega \xrightarrow{R} C$ are unrestricted. The idea is that decompositions in the

ordered context Ω should be preferred when the succedent is of the non-invertible form C^+ so does not have a canonical search-free decomposition. Overall, actions in the ordered context Ω will turn out to be deterministic while those for Γ^- involve decisions and search. That gives eager invertible decompositions and lazy search for non-invertibles.

After we have developed the rules we will summarize the forms of Γ^- and C^+. We refer to this as the *inversion calculus*. Rather than organizing the presentation by connective, we will follow the judgments, starting on the right. That presentation order will enable us to emphasize the intended search order and exhaustiveness of the resulting procedure.

Right inversion. We decompose conjunction, truth, and implication eagerly on the right and on the left, because both rules are invertible and can easily be checked.

$$\frac{\Gamma^- \,;\, \Omega \xrightarrow{R} A \quad \Gamma^- \,;\, \Omega \xrightarrow{R} B}{\Gamma^- \,;\, \Omega \xrightarrow{R} A \wedge B}\, \wedge R \qquad \frac{}{\Gamma^- \,;\, \Omega \xrightarrow{R} \top}\, \top R \qquad \frac{\Gamma^- \,;\, \Omega \cdot A \xrightarrow{R} B}{\Gamma^- \,;\, \Omega \xrightarrow{R} A \supset B}\, \supset R$$

If we encounter an atomic formula, we switch to inverting on the left, since the identity rule is not invertible. Once could optimize further by checking if P is in the context and only start left inversion if it is not.

$$\frac{\Gamma^- \,;\, \Omega \xrightarrow{L} P}{\Gamma^- \,;\, \Omega \xrightarrow{R} P}\, \mathsf{LR}_P$$

If we encounter disjunction or falsehood, we punt and switch to left inversion.

$$\frac{\Gamma^- \,;\, \Omega \xrightarrow{L} A \vee B}{\Gamma^- \,;\, \Omega \xrightarrow{R} A \vee B}\, \mathsf{LR}_\vee \qquad \frac{\Gamma^- \,;\, \Omega \xrightarrow{L} \bot}{\Gamma^- \,;\, \Omega \xrightarrow{R} \bot}\, \mathsf{LR}_\bot$$

Disjunctions would need a commitment whether their left or their right disjunct is proved. Switching to right decomposition postpones that choice until we maximize what we know. Note how the right inversion rules really only switch to left decomposition for non-invertible succedents C^+. Also suddenly there is a rule for $\bot$ on the right, but it merely switches mode to left inversion, so no need to panic.

Left inversion. The next phase performs left inversion at the right end of the ordered context Ω. Note that for each logical connective or constant, there is exactly one rule to apply.

$$\frac{\Gamma^- \,;\, \Omega \cdot A \cdot B \xrightarrow{L} C^+}{\Gamma^- \,;\, \Omega \cdot (A \wedge B) \xrightarrow{L} C^+}\, \wedge L \qquad \frac{\Gamma^- \,;\, \Omega \xrightarrow{L} C^+}{\Gamma^- \,;\, \Omega \cdot \top \xrightarrow{L} C^+}\, \top L$$

$$\frac{\Gamma^- \,;\, \Omega \cdot A \xrightarrow{L} C^+ \quad \Gamma^- \,;\, \Omega \cdot B \xrightarrow{L} C^+}{\Gamma^- \,;\, \Omega \cdot (A \vee B) \xrightarrow{L} C^+}\, \vee L \qquad \frac{}{\Gamma^- \,;\, \Omega \cdot \bot \xrightarrow{L} C^+}\, \bot L$$

Observe how helpful it is that the succedent of $\vee L$ is already decomposed to C^+ so has no invertible right rule, otherwise we would have to repeat the same effort decomposing the succedent by right inversion on both premises of $\vee L$. For atomic formulas, we just move them into the noninvertible context, since the identity is not invertible. We could optimize further by looking of P was equal to C^+ succeed if so.

$$\frac{\Gamma^-, P \,;\, \Omega \xrightarrow{L} C^+}{\Gamma^- \,;\, \Omega \cdot P \xrightarrow{L} C^+}\, \mathsf{shift}_P$$

Finally, in the inversion phase, if the formula on the left is an implication, which can not be inverted, we move it into Γ^-.

$$\frac{\Gamma^-, A \supset B \,;\, \Omega \xrightarrow{L} C^+}{\Gamma^- \,;\, \Omega \cdot (A \supset B) \xrightarrow{L} C^+}\, \mathsf{shift}_\supset$$

Search. The proof process described so far is deterministic and either succeeds finitely with a deduction, or we finally have to make a decision we might regret. Such decisions become necessary when the ordered context has become empty (marked $\cdot$). At this point either identity or one of the $\vee R$

or $\supset L$ rules must be tried.

$$\frac{P \in \Gamma}{\Gamma \,;\, \cdot \xrightarrow{L} P}\ \mathsf{id}$$

$$\frac{\Gamma^- \,;\, \cdot \xrightarrow{R} A}{\Gamma^- \,;\, \cdot \xrightarrow{L} A \vee B}\ \vee R_1 \qquad \frac{\Gamma^- \,;\, \cdot \xrightarrow{R} B}{\Gamma^- \,;\, \cdot \xrightarrow{L} A \vee B}\ \vee R_2$$

$$\frac{\Gamma^-, A \supset B \,;\, \cdot \xrightarrow{R} A \quad \Gamma^- \,;\, B \xrightarrow{L} C^+}{\Gamma^-, A \supset B \,;\, \cdot \xrightarrow{L} C^+}\ \supset L$$

After making a choice, we go back to a phase of inversion, either on the right (in the first premise or only premise) or on the left (in the second premise of $\supset L$). Right inversion is the appropriate phase for $\vee R_1$, $\vee R_2$ and the first premise of $\supset L$, since the resulting formula A or B, respectively, might very well have an invertible connective so should be handled with the deterministic search first. For the second premise of $\supset L$, right inversion would be pointless, because its succedent C^+ is already known to have a non-invertible connective. Finally observe how all inversion rules make some progress to simplify the sequents, which, in the propositional setting, can happen only finitely often.

Again, it is easy to see that the inversion calculus is sound, since it is a further restriction on the rules from the sequent calculus. It is more difficult to see that it is complete. We will not carry out this proof, but just mention that it revolves around the invertibility of the rules excepting only $\vee R_1$, $\vee R_2$, and $\supset L$.

The inversion calculus is a big step forward, but it does not solve the problem with the left rule for implication, where the principal formula is copied to the first premise. We will address this in the next lecture with a so-called *contraction-free* calculus.

3 Soundness and Completeness of Inversion

We define the translation between ordered an unordered contexts via

$$\begin{array}{lcl} \overline{\cdot} & = & \cdot \\ \overline{\Omega \cdot A} & = & \overline{\Omega}, A \end{array}$$

Then the soundness theorem states

1. If $\Gamma^- \,;\, \Omega \xrightarrow{R} A$ then $\Gamma^-, \overline{\Omega} \longrightarrow A$, and
2. if $\Gamma^- \,;\, \Omega \xrightarrow{L} C^+$ then $\Gamma^-, \overline{\Omega} \longrightarrow C^+$

The proof is straightforward by induction over the structure of the given sequent deduction. This is because the new rules just distinguish and limit certain inferences, but the otherwise the rules remain intact.

The completness is a much more complex theorem. What we want is

1. If $\Gamma^-, \overline{\Omega} \Longrightarrow A$ then $\Gamma^- \,;\, \Omega \xrightarrow{R} A$, and
2. if $\Gamma^-, \overline{\Omega} \Longrightarrow C^+$ then $\Gamma^- \,;\, \Omega \xrightarrow{L} C^+$.

The key to this property, as for many completeness theorems, is the admissibility of cut and identity in the more restricted system. Both of these are significantly more complicated than for ordinary sequent calculus. Simple properties, such as weakening, no longer hold in the strong form we had earlier. For example, we might have

$$\frac{}{\Gamma^- \,;\, \Omega \cdot \bot \xrightarrow{L} C^+}\ \bot L$$

but if we weaken, for example, as

$$\Gamma^- \,;\, \Omega \cdot \bot \cdot ((A \vee B) \wedge (C \vee D)) \xrightarrow{L} C^+$$

we are now *forced* to break down $(A \vee B) \wedge (C \vee D)$ completely before we can apply $\bot L$ in each branch.

We do not replicate the proof here, but the interested reader is referred to Rob Simmons' elegant solution for an even more restricted system [Sim14].

4 The Contraction-Free Sequent Calculus, Revisited

At this point we need to reexamine the question from last lecture: where do we really need to make choices in this sequent calculus? We ask the question slight differently this time, although the primary tool will still be the invertibility of rules. The question we want to ask this time: if we consider a formula on the right or on the left, can we always apply the corresponding rule without considering other choices? The difference between the two questions becomes clear, for example, in the $P{\supset}L$ rule.

$$\frac{P \in \Gamma \quad \Gamma, B \longrightarrow C}{\Gamma, P \supset B \longrightarrow C}\ P{\supset}L$$

This rule is clearly invertible, because $P \wedge (P \supset B) \equiv P \wedge B$. Nevertheless, when we consider $P \supset B$ we cannot necessarily apply this rule because P may not be in the remaining context Γ. It might become available in the context later, though, after decomposing Γ. So we may have to wait with applying $P{\supset}L$ until $P \in \Gamma$.

Formulas whose left or right rules can always be applied are called left or right *asynchronous*, respectively, otherwise *synchronous*, because we may have to wait until they can be applied. We can see by examining the rules and considering the equivalences above and the methods from the last lecture, that the following formulas are asynchrounous:

Right asynchronous	$A \wedge B$, $\top$, $A \supset B$
Left asynchronous	$A \wedge B$, $\top$, $A \vee B$, $\bot$,
	$(A_1 \wedge A_2) \supset B$, $\top \supset B$, $(A_1 \vee A_2) \supset B$, $\bot \supset B$

This leaves

Right synchronous	P, $A \vee B$, $\bot$
Left synchronous	P, $P \supset B$, $(A_1 \supset A_2) \supset B$

Atomic propositions are synchronous, because we may have to wait until it shows up in the antecedent and succedent. Disjunction is right synchronous because of the honest choice that $\vee R_1$ versus $\vee R_2$ imposes. Falsum is right synchronous because it needs to wait for $\bot$ to appear in the antecedent (no $\bot R$ rule). Atomic implication $P \supset B$ is left synchronous, because its rule $P{\supset}L$ waits for a $P \in \Gamma$. Nested implication $(A_1 \supset A_2) \supset B$ could be considered left synchronous in the sense of waiting, because it is useful to handle the remaining context Γ before applying $\supset{\supset}L$, because any rules on Γ would otherwise have to be repeated in the first and second premise. But that is not actual reason! Nested implication $(A_1 \supset A_2) \supset B$ is left synchronous, because $\supset{\supset}L$ is not invertible. In its first premise, rule $\supset{\supset}L$ sets out to prove E from some assumptions, which may be unsuccessful, e.g., if C is a disjunction $P \vee Q$ for which the other synchronous cases $\vee R_1$ or $\vee R_2$ succeed without expanding $(A_1 \supset A_2) \supset B$ in the antecedent.

Similar to the idea behind the inversion calculus from the previous lecture, proof search proceeds in phases. Proof search begins by breaking down all asynchronous formulas, leaving us with a situation where we have a synchronous formula on the right and only synchronous formulas on the left. We now check if id or $P{\supset}L$ can be applied and use them if possible. Since these rules are invertible, this, fortunately, does not require a choice. But, of course, if they do not apply, we have to check again later as

more facts became available in Γ. When no more of these rules are applicable, we have to choose between $\vee R_1$, $\vee R_2$ or $\supset\supset L$, if the opportunity exists; if not we fail and backtrack to the most recent choice point. This makes intuitive sense. If we have a disjunction on the right and an implication with an implicational assumption on the left, there is a tradeoff of whether proof search should try proving the assumption via $\supset\supset L$ or try proving one of the two disjuncts by $\vee R_1$ or $\vee R_2$.

This strategy is complete and efficient for many typical examples, although in the end we cannot overcome the polynomial-space completeness of the intuitionistic propositional logic [Sta79]. Indeed, the search will only ever keep strictly smaller subformulas of the input (in the well-founded order) in the sequents. But we need to search through different choices to find the right combination of $\vee R_1$, $\vee R_2$ or $\supset\supset L$ that yield a proof.

The metatheory of the contraction-free sequent calculus has been investigated separately from its use as a decision procedure by Dyckhoff and Negri [DN00]. The properties there could pave the way for further efficiency improvements by logical considerations, specifically in the treatment of atoms.

An entirely different approach to theorem proving in intuitionistic propositional logic is to use the *inverse method* [MP08] which is, generally speaking, more efficient on difficult problems, but not as direct on easier problems. We will discuss this technique in a later lecture.

Finally observe a computational interpretation of the identity theorem that $A \longrightarrow A$. In particular, in combination with the weakening theorem, $\Gamma, A \longrightarrow A$, which is in direct competition with the id rule which is of the same form but only applicable if A is an atomic formula. The pragmatics for proof search is that a check for applicability of the identity theorem would lead to frequent formula comparisons of complexity linear in the size of the formulas. In comparison, id is simpler because it is a direct comparison of atoms, so can essentially be made in constant time for a finite number of atoms. The identity theorem shows that it is sufficient to wait for only atomic formulas to be compared in the application of the identity rule.

References

[DN00] Roy Dyckhoff and Sara Negri. Admissibility of structural rules for contraction-free systems of intuitionistic logic. *Journal of Symoblic Logic*, 65:1499–1518, 2000.

[MP08] Sean McLaughlin and Frank Pfenning. Imogen: Focusing the polarized inverse method for intuitionistic propositional logic. In I.Cervesato, H.Veith, and A.Voronkov, editors, *Proceedings of the 15th International Conference on Logic for Programming, Artificial Intelligence, and Reasoning (LPAR'08)*, pages 174–181, Doha, Qatar, November 2008. Springer LNCS 5330. System Description.

[Sim14] Robert J. Simmons. Structural focalization. *Transactions on Computational Logic*, 15(3):21:1–21:33, July 2014.

[Sta79] Richard Statman. Intuitionistic propositional logic is polynomial-space complete. *Theoretical Computer Science*, 9:67–72, 1979.