Lecture Notes on Inversion

15-317: Constructive Logic
Frank Pfenning*

Lecture 11
October 6, 2016

1 Introduction

The sequent calculus we have introduced so far maintains a close correspondence to natural deductions or, more specifically, to verifications. One consequence is *persistence of antecedents*: once an assumption has been introduced in the course of a deduction, it will remain available in any sequent above this point. While this is appropriate in a foundational calculus, it is not ideal for proof search since rules can be applied over and over again without necessarily making progress. We therefore develop a second sequent calculus and then a third in order to make the process of bottom-up search for a proof more efficient by reducing unnecessary choices in proof search. By way of the previous link of the sequent calculus with verification-style natural deductions, this lecture will, thus, give rise to a more efficient way of coming up with natural deduction proofs.

This lecture marks the begin of a departure from the course of the lectures so far, which, broadly construed, focused on understanding what a constructive proof is and what can be read off or done once one has such a proof. Now we begin to move toward the question of how to find such a proof in the first place.

*With edits by André Platzer

2 A More Restrictive Sequent Calculus

Ideally, once we have applied an inference rule during proof search (that is, bottom-up), we should not have to apply the same rule again to the same proposition. Since all rules decompose formulas, if we had such a sequent calculus, we would have a simple and clean decision procedure. As it turns out, there is a fly in the ointment, but let us try to derive such a system.

We write $\Gamma \longrightarrow C$ for a sequent whose deductions try to eliminate principal formulas as much as possible. We keep the names of the rules, since they are largely parallel to the rules of the original sequent calculus, $\Gamma \Longrightarrow C$.

Conjunction. The right rule works as before; the left rule extracts *both* conjuncts so that the conjunction itself is no longer needed.

$$\frac{\Gamma \longrightarrow A \quad \Gamma \longrightarrow B}{\Gamma \longrightarrow A \wedge B}\ \wedge R \qquad \frac{\Gamma, A, B \longrightarrow C}{\Gamma, A \wedge B \longrightarrow C}\ \wedge L$$

Observe that for both rules, all premises have smaller sequents than the conclusion if one counts the number of connectives in a sequent. So applying either rule obviously made progress toward simplifying the sequent.

It is easy to see that these rules are sound with respect to the ordinary sequent calculus rules. Soundness here is the property that if $\Gamma \longrightarrow C$ then $\Gamma \Longrightarrow C$. This is straightforward since $\wedge R$ is the same rule and $\wedge L$ is the same as $\wedge L_1$ followed by $\wedge L_2$ followed by weakening the original $A \wedge B$ away. Completeness if generally more difficult. What we want to show is that if $\Gamma \Longrightarrow C$ then also $\Gamma \longrightarrow C$, where the rules for the latter sequents are more restrictive, by design. The proof of this will eventually proceed by induction on the structure of the given deduction $\mathcal{D}$ and appeal to lemmas on the restrictive sequent calculus. For example:

Case: (of completeness proof)

$$\mathcal{D} = \frac{\begin{array}{c}\mathcal{D}_1 \\ \Gamma, A \wedge B, A \Longrightarrow C\end{array}}{\Gamma, A \wedge B \Longrightarrow C}\ \wedge L_1$$

$\Gamma, A \wedge B, A \longrightarrow C$	By i.h. on $\mathcal{D}_1$
$\Gamma, A, B \longrightarrow A$	By identity for $\longrightarrow$
$\Gamma, A \wedge B \longrightarrow A$	By $\wedge L$
$\Gamma, A \wedge B \longrightarrow C$	By cut for $\longrightarrow$

The induction hypothesis is applicable to $\mathcal{D}_1$ because, even if it is a longer sequent, $\mathcal{D}_1$ is a shorter proof than $\mathcal{D}$. We see that identity and cut for the restricted sequent calculus is needed to show completeness in the sense described above. Fortunately, they hold (see further notes at the end of the lecture). We will not formally justify many of the rules, but give informal justifications or counterexamples.

Truth. There is a small surprise here, in that, unlike in natural deduction which had no elimination rule for $\top$, we can have a left rule for $\top$, which eliminates it from the antecedents to make progress (cleanup). It is analogous to the zero-ary case of conjunction.

$$\frac{}{\Gamma \longrightarrow \top}\ \top R \qquad \frac{\Gamma \longrightarrow C}{\Gamma, \top \longrightarrow C}\ \top L$$

Atomic propositions. They are straightforward, since the initial sequents do not change.

$$\frac{}{\Gamma, P \longrightarrow P}\ \mathsf{init}$$

Disjunction. The right rules to do not change; in the left rule we can eliminate the principal formula.

$$\frac{\Gamma \longrightarrow A}{\Gamma \longrightarrow A \vee B}\ \vee R_1 \qquad \frac{\Gamma \longrightarrow B}{\Gamma \longrightarrow A \vee B}\ \vee R_2 \qquad \frac{\Gamma, A \longrightarrow C \quad \Gamma, B \longrightarrow C}{\Gamma, A \vee B \longrightarrow C}\ \vee L$$

Intuitively, the assumption $A \vee B$ can be eliminated from both premises of the $\vee L$ rule, because the new assumptions A and B are stronger. More formally:

Case: (of completeness proof)

$$\mathcal{D} = \frac{\begin{array}{c}\mathcal{D}_1 \\ \Gamma, A \vee B, A \Longrightarrow C\end{array} \quad \begin{array}{c}\mathcal{D}_2 \\ \Gamma, A \vee B, B \Longrightarrow C\end{array}}{\Gamma, A \vee B \Longrightarrow C}\ \vee L$$

$\Gamma, A \vee B, A \longrightarrow C$	By i.h. on $\mathcal{D}_1$
$\Gamma, A \longrightarrow A$	By identity for $\longrightarrow$
$\Gamma, A \longrightarrow A \vee B$	By $\vee R_1$
$\Gamma, A \longrightarrow C$	By cut for $\longrightarrow$

$$
\begin{array}{ll}
\Gamma, A \vee B, B \longrightarrow C & \text{By i.h. on } \mathcal{D}_2 \\
\Gamma, B \longrightarrow B & \text{By identity for } \longrightarrow \\
\Gamma, B \longrightarrow A \vee B & \text{By } \vee R_2 \\
\Gamma, B \longrightarrow C & \text{By cut for } \longrightarrow \\
\Gamma, A \vee B \longrightarrow C & \text{By rule } \vee L
\end{array}
$$

Falsehood. There is no right rule, and the left rule has no premise, which means it transfers directly.

$$
\text{no } \bot R \text{ rule} \qquad \dfrac{}{\Gamma, \bot \longrightarrow C}\;\bot L
$$

Implication. In all the rules so far, all premises have fewer connectives than the conclusion. For implication, we will not be able to maintain this property.

$$
\dfrac{\Gamma, A \longrightarrow B}{\Gamma \longrightarrow A \supset B}\;{\supset}R \qquad \dfrac{\Gamma, A \supset B \longrightarrow A \quad \Gamma, B \longrightarrow C}{\Gamma, A \supset B \longrightarrow C}\;{\supset}L
$$

Here, the assumption $A \supset B$ persists in the first premise but not in the second. While the assumption B is more informative than $A \supset B$, so only B is kept in the second premise, this is not the case in the first premise. Unfortunately, $A \supset B$ may be needed again in that branch of the proof. An example which requires the implication more than once is $\longrightarrow \neg\neg(A \vee \neg A)$, where $\neg A = A \supset \bot$ as usual. Without that additional assumption (marked in red below), the proof would not work:

$$
\dfrac{
  \dfrac{
    \dfrac{
      \dfrac{
        \dfrac{
          \dfrac{\genfrac{}{}{0pt}{}{\cdots\cdots\cdots\cdots}{\neg(A \vee \neg A), A \longrightarrow A}\;id}{\neg(A \vee \neg A), A \longrightarrow A \vee \neg A}\;\vee R_1
          \quad
          \dfrac{}{A, \bot \longrightarrow \bot}\;init
        }{\neg(A \vee \neg A), A \longrightarrow \bot}\;{\supset}L
      }{\neg(A \vee \neg A) \longrightarrow \neg A}\;{\supset}R
    }{\neg(A \vee \neg A) \longrightarrow A \vee \neg A}\;\vee R_2
    \quad
    \dfrac{}{\bot \longrightarrow \bot}\;init
  }{\neg(A \vee \neg A) \longrightarrow \bot}\;{\supset}L
}{\longrightarrow \neg\neg(A \vee \neg A)}\;{\supset}R
$$

Now all rules have smaller premises (if one counts the number of logical constants and connectives in them) except for the $\supset L$ rule. We will address the issue with $\supset L$ in the next lecture.

Nevertheless, we can interpret the rules as a decision procedure if we make the observation that in bottom-up proof search we are licensed to fail a branch if along it we have a repeating sequent. If there were a deduction, we would be able to find it applying a different choice at an earlier sequent, lower down in the incomplete deduction. If there is a proof with repeating sequents, there also is a proof without repeating sequents, by applying the proof that was used for the later occurrence of the repeating sequent already to the first occurrence of said sequent. If we also apply contraction (which is admissible in the restricted sequent calculus) to argue that we can remove duplicate formulas from the antecedent, then there are only finitely many sequents because antecedents and succedents are composed only of subformulas of our original proof goal.

One can be much more efficient in loop checking than this [**?**, Chapter 4], but just to see that intuitionistic propositional calculus is decidable, this is sufficient. In fact, we could have made this observation on the original sequent calculus, although it would be even further from a realistic implementation.

3 Metatheory of the Restricted Sequent Calculus

We only enumerate the basic properties.

Theorem 1 (Weakening) *If* $\Gamma \longrightarrow C$ *then* $\Gamma, A \longrightarrow C$ *with a structurally identical deduction.*

Theorem 2 (Atomic contraction) *If* $\Gamma, P, P \longrightarrow C$ *then* $\Gamma, P \longrightarrow C$ *with a structurally identical deduction*

Theorem 3 (Identity) $A \longrightarrow A$ *for any proposition* A*.*

Proof: By induction on the structure of A. □

Theorem 4 (Cut) *If* $\Gamma \longrightarrow A$ *and* $\Gamma, A \longrightarrow C$ *then* $\Gamma \longrightarrow C$

Proof: Analogous to the proof for the ordinary sequent calculus in Lecture 8. In the case where the first deduction is initial, we use atomic contraction. □

Theorem 5 (Contraction) *If* $\Gamma, A, A \longrightarrow C$ *then* $\Gamma, A \longrightarrow C$*.*

Proof: $\Gamma, A \longrightarrow A$ by identity and weakening. Therefore $\Gamma, A \longrightarrow C$ by cut. □

Theorem 6 (Soundness wrt. $\Longrightarrow$) *If* $\Gamma \longrightarrow A$ *then* $\Gamma \Longrightarrow A$.

Proof: By induction on the structure of the given deduction. □

Theorem 7 (Completeness wrt. $\Longrightarrow$) *If* $\Gamma \Longrightarrow A$ *then* $\Gamma \longrightarrow A$.

Proof: By induction on the structure of the given deduction, appealing to identity and cut in many cases. See the cases for $\wedge L_1$ and $\vee L$ in the previous section. □

4 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 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 $\Gamma, A \wedge B \longrightarrow B$.

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

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. 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.

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

If we encounter an atomic formula, we succeed if it is among the antecedents; otherwise we switch to left inversion.

$$\frac{P \in \Gamma^-}{\Gamma^-; \Omega \xrightarrow{R} P} \mathsf{init} \qquad \frac{P \notin \Gamma^- \quad \Gamma^-; \Omega \xrightarrow{L} P}{\Gamma^-; \Omega \xrightarrow{R} P} \mathsf{LR}_P$$

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.

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

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

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 look to see if it matches the right-hand side and, if so, succeed. Otherwise, we move it into Γ^- since we cannot operate on the atomic formula P any longer (it is non-invertible since there are no rules for it, besides we hope to match it up via init ultimately).

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

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

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

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 one of the $\vee R$ or $\supset L$ rules must be tried.

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

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

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.