5.7 Forward Sequent Calculus for First-Order Logic

Generalizing the basic ideas of the inverse method as introduced in the preceding sections requires unification (see Section 4.3), although it is employed in a different way than in backward search. The underlying method can be traced directly to Robinson's original work on resolution [Rob65], and precise connections to classical resolution have been established in the literature [Tam97].

The extension of the forward sequent calculus to the first-order case is straightforward.

$$\frac{\Gamma, [t/x]A \longrightarrow \gamma}{\Gamma, \forall x.\ A \longrightarrow \gamma}\ \forall\mathrm{L} \qquad \frac{\Gamma \longrightarrow [a/x]A}{\Gamma \longrightarrow \forall x.\ A}\ \forall\mathrm{R}^a$$

$$\frac{\Gamma, [a/x]A \longrightarrow \gamma}{\Gamma, \exists x.\ A \longrightarrow \gamma}\ \exists\mathrm{L}^a \qquad \frac{\Gamma \longrightarrow [t/x]A}{\Gamma \longrightarrow \exists x.\ A}\ \exists\mathrm{R}$$

Recall the restriction on the $\forall$R and $\exists$L rules: the derivation of premise must be parametric in a. That is, a may not occur in Γ or A. Soundness and completeness of this calculus with respect to the backward sequent calculus extends in a straightforward way.

These rules suggest an extension of the subformula property. We write $A < B$ for A *is an immediate subformula of* B, $^{\pm}$ for an arbitrary sign ($^{+}$ or $^{-}$) and $^{\mp}$ for its complement.

$$\begin{array}{ll}
A^{\pm} < (A \wedge B)^{\pm} & B^{\pm} < (A \wedge B)^{\pm} \\
A^{\pm} < (A \vee B)^{\pm} & B^{\pm} < (A \vee B)^{\pm} \\
A^{\mp} < (A \supset B)^{\pm} & B^{\pm} < (A \supset B)^{\pm} \\
[a/x]A^{+} < (\forall x.\ A)^{+} & \text{for all parameters } a \\
[t/x]A^{-} < (\forall x.\ A)^{-} & \text{for all terms } t \\
[t/x]A^{+} < (\exists x.\ A)^{+} & \text{for all terms } t \\
[a/x]A^{-} < (\exists x.\ A)^{-} & \text{for all parameters } a
\end{array}$$

We write $A <^* B$ for the reflexive and transitive closure of the immediate subformula relation. Also, we write $A <^* \Gamma$ if there is a formula B in Γ such that $A <^* B$, and $\Delta <^* \Gamma$ if for every A in Δ, $A <^* \Gamma$.

The signed subformula property (Theorem 5.5) directly extends to the first-order case, using the definitions above:

> *For all sequents* $\Delta^- \longrightarrow A^+$ *or* $\Delta^- \longrightarrow \cdot$ *in a derivation of* $\Gamma^- \longrightarrow C^+$ *or* $\Gamma^- \longrightarrow \cdot$ *we have* $\Delta^-, A^+ <^* \Gamma^-, C^+$.

Before formalizing the first-order inverse method, we now go through several examples which show how to take advantage of this extended subformula property in order to construct a search algorithm.

The first example is

$$(\forall x.\ P(x) \supset P(g(x))) \longrightarrow P(c) \supset P(g(g(c)))$$

for a unary predicate P, function f and constant c. We begin by enumerating and naming subformulas. First, the atomic subformulas, from left to right.

$$
\begin{array}{lll}
(i) & P(\underline{t})^+ & \text{for all terms } \underline{t} \\
(ii) & P(g(\underline{s}))^- & \text{for all terms } \underline{s} \\
(iii) & P(c)^- & \\
(iv) & P(g(g(c)))^+ &
\end{array}
$$

Now, we have to consider all initial sequents $Q \longrightarrow Q$ where Q is a subformula of the goal sequent above. To this end we *unify* positive and negative atomic propositions, treating $\underline{t}$ and $\underline{s}$ as variables, since they stand for arbitrary terms. We obtain:

$$
\begin{array}{lll}
1. & P(g(\underline{s}))^- \longrightarrow P(g(\underline{s}))^+ & \text{for all term } \underline{s}\text{, from } (ii) \text{ and } (i) \\
2. & P(g(g(c)))^- \longrightarrow P(g(g(c)))^+ & \text{from } (ii) \text{ and } (iv) \\
3. & P(c)^- \longrightarrow P(c)^+ & \text{from } (iii) \text{ and } (i)
\end{array}
$$

Note that the sequent (1) above represents a schematic judgment in the same way that inferences rules are schematic, where $\underline{s}$ is a schematic variable ranging over arbitrary terms. This will be true not only of the initial sequents, but of the sequents we derive. This is one of the major generalizations from the propositional case of the inverse method.

We can see that the initial sequents described in line (1) includes those in line (2), since we can use $g(c)$ for $\underline{s}$. This is an extended form of subsumption: not only do we check is one sequent can be weakened to another, but we also have to allow for instantiation of variables ($\underline{s}$, in this case).

Next, we introduce names for compound subformulas.

$$
\begin{array}{lcll}
L_1(\underline{t})^- & = & P(\underline{t})^+ \supset P(g(\underline{t}))^- & \text{for terms } \underline{t} \\
L_2^- & = & \forall x.\, L_1(x)^- & \\
L_3^+ & = & P(c)^- \supset P(g(g(c)))^+ &
\end{array}
$$

From the general forward sequent rules, we can now construct versions of the inference rules specialized to subformulas of the goal sequent.

$$
\frac{\Gamma_1 \longrightarrow P(\underline{t})^+ \qquad \Gamma_2, P(g(\underline{t}))^- \longrightarrow \gamma}{\Gamma_1 \cup \Gamma_2, L_1(\underline{t})^- \longrightarrow \gamma} \supset\mathrm{L}
$$

$$
\frac{\Gamma, L_1(t)^- \longrightarrow \gamma}{\Gamma, L_2^- \longrightarrow \gamma} \forall\mathrm{L}
$$

$$
\frac{\Gamma, P(c)^- \longrightarrow P(g(g(c)))^+}{\Gamma \longrightarrow L_3^+} \supset\mathrm{R}_1
$$

$$
\frac{\Gamma \longrightarrow P(g(g(c)))^+}{\Gamma \longrightarrow L_3^+} \supset\mathrm{R}_2 \qquad \frac{\Gamma, P(c)^- \longrightarrow \cdot}{\Gamma \longrightarrow L_3^+} \supset\mathrm{R}_3
$$

The notation distinguishes the cases where an arbitrary term t is involved in the rule because of the principal connective (in the $\forall$L rule) and where an arbitrary term $\underline{t}$ is involved because of subformula considerations (in the $\supset$L rule).

We can now use these rules, starting from the remaining two initial sequents to derive the goal sequent $L_2^- \longrightarrow L_3^+$. We omit some, but not all sequents that could be generated, but do not contribute to the final derivation.

1.	$P(g(\underline{s}))^- \longrightarrow P(g(\underline{s}))^+$	init, for all terms $\underline{s}$
3.	$P(c)^- \longrightarrow P(c)^+$	init
4.	$P(c)^-, L_1(c)^- \longrightarrow P(g(c))^+$	$\supset$L 3 1$[c/\underline{s}]$
5.	$P(g(\underline{t}))^-, L_1(g(\underline{t}))^- \longrightarrow P(g(g(\underline{t})))^-$	$\supset$L 1$[\underline{t}/\underline{s}]$ 1$[g(\underline{t})/\underline{s}]$, for all $\underline{t}$
6.	$P(g(g(c)))^- \longrightarrow L_3^+$	$\supset$R$_2$ 1$[g(c)/\underline{s}]$
7.	$P(g(\underline{t}))^-, L_2^- \longrightarrow P(g(g(\underline{t})))^+$	$\forall$L 5, for all $\underline{t}$
8.	$P(c)^-, L_2^-, L_1(c)^- \longrightarrow P(g(g(c)))^+$	$\supset$L 3 7$[c/\underline{t}]$
9.	$P(c)^-, L_2^- \longrightarrow P(g(g(c)))^+$	$\forall$L 8, with contraction
10.	$L_2^- \longrightarrow L_3^+$	$\supset$R$_1$ 9

Inference previously involved matching a sequents against the premises of an inference rule. As this example shows, we now have to *unify* derived sequents with the premises of the inference rules. The schematic variables in the sequent as well as in the inference rule may be instantiated in this process, thereby determining the most general conclusion. It is important in this process to note that the scope of each schematic variable includes only a particular sequent or inference rule. Schematic variables called $\underline{t}$ in different sequents are different—usually this is accounted for by systematically renaming variables before starting unification.

The example above does not involve any parameters, only schematic variables. We now consider another example involving parameters,

$$\exists y.\, \forall x.\, P(x, y) \longrightarrow \forall x.\, \exists y.\, P(x, y)$$

for a binary predicate P. Clearly, this judgment should be derivable. Again, we first generate positive and negative atomic subformulas.

(i)	$P(\underline{t}, \underline{a})^-$	for all terms $\underline{t}$ and parameters $\underline{a}$
(ii)	$P(\underline{b}, \underline{s})^+$	for all parameters $\underline{b}$ and terms $\underline{s}$

Because of the negative existential and positive universal quantification the allowed instances of the atomic subformulas are restricted to parameters in certain places. However, it should be understood that $\underline{a}$ in line (i) is only a schematic variable ranging over parameters and may be instantiated to different parameters for different uses of a negative formula $P(\_, \_)^-$.

Next we generate all possible atomic initial sequents. This means we have to look for common instances of the positive and negative atomic formulas schemas listed above. The only possible instances have the form

1. $P(\underline{b}, \underline{a})^- \longrightarrow P(\underline{b}, \underline{a})^+$ for all parameters $\underline{b}$ and terms $\underline{s}$

Now we list the possible compound subformulas.

$$\begin{array}{lcll} L_1(\underline{a})^- & = & \forall x.\, P(x, \underline{a})^- & \text{for parameters } \underline{a} \\ L_2^- & = & \exists y.\, L_1(y)^- & \\ L_3(\underline{b})^+ & = & \exists y.\, P(\underline{b}, y)^+ & \text{for parameters } \underline{b} \\ L_4^+ & = & \forall x.\, L_3(x)^+ & \end{array}$$

The specialized inference rules read:

$$\frac{\Gamma, P(t, \underline{a})^- \longrightarrow \gamma}{\Gamma, L_1(\underline{a})^- \longrightarrow \gamma}\forall\mathrm{L} \qquad \frac{\Gamma, L_1(a)^- \longrightarrow \gamma}{\Gamma, L_2^- \longrightarrow \gamma}\exists\mathrm{L}^a$$

$$\frac{\Gamma \longrightarrow P(\underline{b}, s)^+}{\Gamma \longrightarrow L_3(\underline{b})^+}\exists\mathrm{R} \qquad \frac{\Gamma \longrightarrow L_3(b)^+}{\Gamma \longrightarrow L_4^+}\forall\mathrm{R}^b$$

Note that the $\exists$L and $\forall$R rules have parametric premises, which means we have to enforce the side condition that parameter a or b do not occur elsewhere in the premises of these two rules, respectively. The derivation takes the following simple form. We omit signs for brevity, and it should be understood that $\underline{b}$ and $\underline{a}$ are quantified *locally* in each sequent.

1.	$P(\underline{b}, \underline{a}) \longrightarrow P(\underline{b}, \underline{a})$	init
2.	$L_1(\underline{a}) \longrightarrow P(\underline{b}, \underline{a})$	$\forall$L 1
3.	$P(\underline{b}, \underline{a}) \longrightarrow L_3(\underline{b})$	$\exists$R 1
4.	$L_1(\underline{a}) \longrightarrow L_3(\underline{b})$	$\exists$R 2
5.	$L_1(\underline{a}) \longrightarrow L_3(\underline{b})$	$\forall$L 3 (subsumed by 4)
6.	$L_2 \longrightarrow L_3(\underline{b})$	$\exists\mathrm{L}^a$ 4
7.	$L_1(\underline{a}) \longrightarrow L_4$	$\forall\mathrm{R}^b$ 4
8.	$L_2 \longrightarrow L_4$	$\forall\mathrm{R}^b$ 6 or $\exists\mathrm{L}^a$ 7

Note that the $\exists$L and $\forall$R rule are not applicable to sequents (2) or (3), because the side conditions on the parameters would be violated.

Next we consider the converse, which should *not* be derivable.

$$\forall x.\, \exists y.\, P(x, y) \longrightarrow \exists y.\, \forall x.\, P(x, y)$$

Again, we first generate the atomic subformulas.

$$\begin{array}{lll} (i) & P(\underline{t}, \underline{a})^- & \text{for all terms } \underline{t} \text{ and parameters } \underline{a} \\ (ii) & P(\underline{b}, \underline{s})^+ & \text{for all parameters } \underline{b} \text{ and terms } \underline{s} \end{array}$$

Then the possible initial sequents.

1. $P(\underline{b}, \underline{a})^- \longrightarrow P(\underline{b}, \underline{a})^+$ for all parameters $\underline{b}$ and terms $\underline{a}$

Then, the compound subformulas.

$$\begin{array}{lcll} L_1(\underline{t})^- & = & \exists y.\, P(\underline{t}, y)^- & \text{for terms } \underline{t} \\ L_2^- & = & \forall x.\, L_1(x)^- & \\ L_3(\underline{s})^+ & = & \forall x.\, P(x, \underline{s})^+ & \text{for terms } \underline{s} \\ L_4^+ & = & \exists y.\, L_3(y)^+ & \end{array}$$

From this we derive the specialized rules of inference.

$$\frac{\Gamma, P(\underline{t}, a)^- \longrightarrow \gamma}{\Gamma, L_1(\underline{t})^- \longrightarrow \gamma}\exists\mathrm{L}^a \qquad \frac{\Gamma, L_1(t)^- \longrightarrow \gamma}{\Gamma, L_2^- \longrightarrow \gamma}\forall\mathrm{L}$$

$$\frac{\Gamma \longrightarrow P(b, \underline{s})^+}{\Gamma \longrightarrow L_3(\underline{s})^+}\forall\mathrm{R} \qquad \frac{\Gamma \longrightarrow L_3(s)^+}{\Gamma \longrightarrow L_4^+}\exists\mathrm{R}$$

Given an initial sequent

1. $P(\underline{b}, \underline{a})^- \longrightarrow P(\underline{b}, \underline{a})^+$ for all parameters $\underline{b}$ and terms $\underline{a}$

we see that no inference rules are applicable, because the side condition on parameter occurrences would be violated. Therefore the goal sequent cannot be derivable.

5.8 Factoring

The examples in the previous section suggest the following algorithm:

1. Determine all signed schematic atomic subformulas of the given goal sequent.
2. Unify positive and negative atomic subformulas after renaming variables so they have none in common. This yields a set of initial sequents from which subsumed copies should be eliminated.
3. Name all signed compound subformulas as new predicates on their free variables.
4. Specialize the inference rules to these subformulas.
5. Starting from the initial sequents, apply the specialized inference rules in a fair way by unifying (freshly renamed) copies of sequents derived so far with premises of the inference rules, generating most general conclusions as a new schematic sequents.
6. Stop with success when the goal sequent has been derived.

Perhaps somewhat surprisingly, this method is incomplete using only the rules given so far. As a counterexample, consider

$$\cdot \longrightarrow \exists x.\, P(x) \supset P(x) \wedge P(c)$$

for a unary predicate P and constant c. Initial sequents:

1. $P(\underline{t}) \longrightarrow P(\underline{t})$ for all terms t
2. $P(c) \longrightarrow P(c)$ (subsumed by (1))

Signed subformulas:

$$\begin{array}{lcl} L_1^+(\underline{s}) & = & P(\underline{s})^+ \wedge P(c)^+ \\ L_2^+(\underline{s}) & = & P(\underline{s})^- \supset L_1(\underline{s})^+ \\ L_3^+ & = & \exists x.\, L_2^+(x) \end{array}$$

Specialized rules (omitting polarities and the irrelevant $\supset$R$_3$):

$$\frac{\Gamma_1 \longrightarrow P(\underline{s}) \qquad \Gamma_2 \longrightarrow P(c)}{\Gamma_1 \cup \Gamma_2 \longrightarrow L_1} \wedge\mathrm{I}$$

$$\frac{\Gamma, P(\underline{s}) \longrightarrow L_1(\underline{s})}{\Gamma \longrightarrow L_2(\underline{s})} \supset\mathrm{R}_1 \qquad \frac{\Gamma \longrightarrow L_1(\underline{s})}{\Gamma \longrightarrow L_2(\underline{s})} \supset\mathrm{R}_2$$

$$\frac{\Gamma \longrightarrow L_2(t)}{\Gamma \longrightarrow L_3} \exists\mathrm{R}$$

Initially, we can only apply $\wedge$I, after renaming a copy of (1).

$$\begin{array}{lll} 1. & P(\underline{t}) \longrightarrow P(\underline{t}) & \text{init, for all terms } \underline{t} \\ 3. & P(\underline{t}), P(c) \longrightarrow L_1(\underline{t}) & \wedge\mathrm{R}\; 1[\underline{t}/\underline{t}]\; 1[c/\underline{t}], \text{for all terms } \underline{t} \end{array}$$

Now there are two ways to apply the $\supset$R$_1$ rule, but either $P(\underline{t})$ or $P(c)$ is left behind as an assumption, and the goal sequent cannot be derived.

The problem is that even though the sequent

$$P(c) \longrightarrow L_1(c)$$

should be derivable, it is only the contraction of an instance of sequent (3). We therefore extend the system with an explicit rule which permits contraction after instantiation, called *factoring*. That is, after we derive a new sequent, we consider possible most general unifiers among antecedents of the sequent and add the results (while continuing to check for subsumption).

In the example above, we proceed as follows:

$$\begin{array}{lll} 1. & P(\underline{t}) \longrightarrow P(\underline{t}) & \text{init, for all terms } \underline{t} \\ 3. & P(\underline{t}), P(c) \longrightarrow L_1(\underline{t}) & \wedge\mathrm{R}\; 1[\underline{t}/\underline{t}]\; 1[c/\underline{t}], \text{for all terms } \underline{t} \\ 4. & P(c) \longrightarrow L_1(c) & \mathit{factor}\; 3[c/\underline{t}] \\ 5. & \cdot \longrightarrow L_2(c) & \supset\mathrm{R}_1\; 4 \\ 6. & \cdot \longrightarrow L_3 & \exists\mathrm{R} \end{array}$$

Usually, this is done eagerly for each rule which unions assumptions and therefore might allow new factors to be derived. It is also possible to delay this until the rules which require factoring (such as $\supset$R), but this might require factoring to be done repeatedly and may prohibit some subsumption.

In our inference rule notation, where unification of sequents with premises of rules is implicit, this factoring rule would simply look like a contraction.

$$\frac{\Gamma, A, A \longrightarrow C}{\Gamma, A \longrightarrow C}\,\text{contract}$$

Previously, this was implicit, since we maintained assumptions as sets.