1 Abstract notions on natural deduction systems

We assume a standard formal language $\mathcal{L}$ for propositional logic with the usual logical operators $\land$, $\lor$, $\neg$, $\rightarrow$ and the usual denumerable set of atomic formulae augmented with the constant $\bot$ denoting falsehood (or absurdity). We use the lower case letters $p, q, r$, etc., with or without subscripts, for the atomic formulae of $\mathcal{L}$ and the upper case letters $A, B, C$, etc., with or without subscripts, as metalinguistic variables ranging over arbitrary formulae of $\mathcal{L}$. Moreover, we use $\Gamma, \Delta$, etc., possibly with subscripts, as metalinguistic variables for finite set of $\mathcal{L}$-formulae. The complexity of a formula $A$, denoted by $|A|$, is defined as the number of occurrences of logical operators and atomic formulae in $A$.

For every formula $A$, a subformula of $A$ is defined inductively as follows: (i) $A$ is a subformula of $A$, (ii) for every binary operator $\circ$, if $B \circ C$ is a subformula of $A$, then so are $B$ and $C$, (iii) if $\neg B$ is a subformula of $A$, then so is $B$; (iv) nothing else is a subformula of $A$. A proper subformula of $A$ is any subformula of $A$ that is different from $A$. An immediate subformula of $A$ is any proper subformula of $A$ that is not a proper subformula of any proper subformula of $A$. Finally, $B$ is a weak subformula of $A$ if either $B$ is a subformula of $A$ or $B = \neg C$ for some subformula $C$ of $A$.

**Remark 1.** Observe that while the relation “$A$ is a subformula of $B$” is transitive, the relation “$A$ is a weak subformula of $B$” is not, as witnessed by the triple $\neg\neg A, \neg A, A$, where $\neg\neg A$ is a weak subformula of $\neg A$ and $\neg A$ is a weak subformula of $A$, but $\neg\neg A$ is not a weak subformula of $A$. However, if $A$ is a weak subformula of $B$ and $B$ is a subformula of $C$, then $A$ is a weak subformula of $C$.

By a formula scheme for $\mathcal{L}$ we mean any expression that results from a formula of $\mathcal{L}$ by replacing the atomic formulas with metalinguistic variables for arbitrary formulae ($A, B, C$, etc.). Let $\rho$ be a uniform substitution of these metalinguistic variables with formulae of $\mathcal{L}$. An instance of a formula scheme $\varphi$ is any formula obtained from $\varphi$ by replacing the metalinguistic variables with $\mathcal{L}$-formulae.

In natural deduction systems proofs are often represented as trees of occurrences of formulae with the conclusion occurring as root root and the assumptions occurring
as leaves. (In computational applications they may be more efficiently represented in other formats, e.g. as directed acyclic graphs.) A formula-tree is a tree $T$ of occurrences of $L$-formulae in which some or all of the formulae occurring in the leaves may be enclosed in square brackets. (The meaning of enclosing a formula in square brackets will be explained below.) A subtree of $T$ is a tree consisting of a node of $T$ and all its descendants. An immediate subtree of $T$ is a subtree whose root is a child of $T$’s root. We shall use the notation

$$T_1, \ldots, T_n \psi$$

to represent the tree $T$ with (an occurrence of) $\psi$ as root and $T_1, \ldots, T_n$ as immediate subtrees; and the notation

$$T_1 \cdots T_n \psi$$

when we want to emphasize that the root of $T_i$ ($i = 1, \ldots, n$) is an (occurrence of) $A_i$. The height of a formula tree $T$, denoted by $h(T)$, is the maximum length of its branches.

Proofs are formula-trees constructed in accordance with certain derivation rules. For this purpose, we distinguish between inference rules and proof rules. An inference rule is a relation between tuples of formulae and formulae represented by an expression of the following form:

$$\varphi_1, \ldots, \varphi_n \psi$$

where $\varphi_1, \ldots, \varphi_n, \psi$ are formula schemes. We say that $A$ follows from a set $\Gamma$ of formulae by an application of the inference rule if there is a uniform substitution of the metalinguistic variables with $L$-formulae that turns $\{\varphi_1, \ldots, \varphi_n\}$ into $\Gamma$ and $\psi$ into $A$. The formulae in $\Gamma$ and the formula $A$ are called, respectively, the premises and the conclusion of the rule application.

**Example 1.** The expression $A \rightarrow B$ is a formula scheme, since it can be obtained from $p \rightarrow q$ by replacing $p$ with $A$ and $q$ with $B$. The $L$-formula $(p \land q) \rightarrow (q \lor r)$ is an instance of $A \rightarrow B$. The expression

$$\begin{array}{cc}
A \rightarrow B & A \\
B & MP
\end{array}$$

is an inference rule. The $L$-formula $q \lor r$ follows from $(p \land q) \rightarrow (q \lor r)$ and $(p \land q)$ by an application of MP.

A proof rule is a relation between formula trees and formula trees represented by an expression of the following form:

$$\begin{array}{cccc}
[\theta_1], \ldots, [\theta_{k_1}] & [\theta_1], \ldots, [\theta_{k_n}] \\
\vdots & \vdots & \vdots \\
\varphi_1 & \cdots & \varphi_n \\
\psi & R
\end{array}$$
where \( k_i \geq 0 \), for all \( i = 1, \ldots, n \). Such a rule says that if, for some uniform substitution \( \rho \) of the metalinguistic variables with \( \mathcal{L} \)-formulae, there are formula trees \( T_1, \ldots, T_n \), such that \( T_i \) represents a proof of \( \rho(\varphi_i) \), then the formula tree obtained by adding an occurrence of \( \rho(\psi) \) immediately below the roots of \( T_1, \ldots, T_n \) is a proof of \( \rho(\psi) \). Moreover, the ending formula \( \rho(\psi) \) no longer depends on the assumptions \( \rho(\theta_{ij}) \), with \( 1 \leq i \leq k \) and \( 1 \leq j \leq n \), enclosed in square brackets, which are said to be discharged by the application of the rule, but only on the remaining undischarged assumptions (if any) of \( T_1, \ldots, T_n \). In general, some of the discharged assumptions may not actually occur in the leaves of the subproofs. In this case we say that the assumptions are vacuously discharged.

Although, strictly speaking, the premises and the conclusions of an application of a proof rule are proofs and not formulae, we shall often loosely speak of the conclusion of an application of the rule to refer to the ending formula of the proof that results from this application.

**Example 2.** The expression:

\[
\begin{array}{c}
[A] \\
B \\
\hline
A \rightarrow B
\end{array}
\]

represents a proof rule. It says that any proof of \( B \) depending on assumptions \( \Gamma \cup \{ A \} \) counts as a proof of \( B \) depending on the assumptions \( \Gamma \) only. For every application of the rule, the assumption instantiating \( A \) is discharged by the rule application. As a special case, a proof of \( B \) depending on \( \Gamma \) counts as a proof of \( A \rightarrow B \) depending on \( \Gamma \) even if the discharged assumption \( A \) is not used in it. In this case \( n = 1 \) and \( k_1 = 1 \).

The expression

\[
\begin{array}{ccc}
[A] & [B] \\
\hline
A \lor B & C & C \\
\hline
C
\end{array}
\]

is also a proof rule which says that a proof of \( A \lor B \), combined with a proof of \( C \) depending a set of assumptions possibly including \( A \), and with a proof of \( C \) depending on a set of assumptions possibly including \( B \), counts as a proof of \( C \) that does not depend on the discharged assumptions \( A \) and \( B \), but only on the undischarged ones. In this case \( n = 3 \), \( k_1 = 0 \), that is, none of the assumptions on which the proof of \( A \lor B \) depends are discharged by the rule application, \( k_2 = k_3 = 1 \).

As a last example, the following expression:

\[
\begin{array}{ccc}
[A] & [\neg A] \\
\hline
\hline
B & B \\
\hline
B
\end{array}
\]

is a proof rule which says that a proof of \( B \) depending on a set of assumptions possibly including \( A \) combined with a proof of \( B \) depending on a set of assumptions possibly in-
cluding \(\neg A\), counts as a proof of \(B\) that does not depend on the discharged assumptions \(A\) and \(\neg A\), but only on the undischarged ones. In this case \(n = 2\) and \(k_1 = k_2 = 1\).

A natural deduction system \(S\) is a pair \(\langle \mathcal{I}, \mathcal{P} \rangle\) where \(\mathcal{I}\) is a finite set of inference rules and \(\mathcal{P}\) is a finite set of proof rules. The notion of \(S\)-proof of a formula \(\psi\) depending on a set of assumptions \(\Gamma\) is defined inductively as follows:

**Definition 1.**

1. The one-node formula tree consisting of a single occurrence of a formula \(\psi\) is an \(S\)-proof of \(\psi\) depending on \(\{\psi\}\).

2. If the inference rule
   \[
   \begin{array}{c}
   \varphi_1, \ldots, \varphi_n \\
   \hline
   \psi
   \end{array}
   \]
   belongs to \(\mathcal{I}\) and \(T_1, \ldots, T_n\) are formula trees such that, for some uniform substitution \(\rho\) of the metalinguistic variables in the formula schemes, \(T_i\) is an \(S\)-proof of \(\rho(\varphi_i)\) depending on \(\Gamma_i\), then the tree
   \[
   T = \frac{T_1, \ldots, T_n}{\rho(\psi)}
   \]
   is an \(S\)-proof of \(\rho(\psi)\) depending on \(\Gamma_1 \cup \cdots \cup \Gamma_n\).

3. If the proof rule
   \[
   \begin{array}{c}
   \begin{array}{c}
   [\theta_1_1], \ldots, [\theta_{k_1}] \\
   \vdots
   \end{array} \quad [\theta_{1_n}], \ldots, [\theta_{k_n}] \\
   \vdots
   \end{array}
   \begin{array}{c}
   \varphi_1 \\
   \vdots
   \end{array}
   \begin{array}{c}
   \varphi_n \\
   \end{array}
   \begin{array}{c}
   R
   \end{array}
   \end{array}
   \]
   \[
   \begin{array}{c}
   \hline
   \psi
   \end{array}
   \]
   belongs to \(\mathcal{P}\) and the following conditions hold for some uniform substitution \(\rho\) of the metalinguistic variables in the formula schemes:
   - \(T_1, \ldots, T_n\) are formula trees such that \(T_i\) is an \(S\)-proof of \(\rho(\varphi_i)\) depending on \(\Gamma_i\),
   - \(T'_1, \ldots, T'_n\) are the formula trees obtained from \(T_1, \ldots, T_n\) by enclosing every occurrence of \(\rho(\theta_{j_i})\) in square brackets,

then, the tree
   \[
   T = \frac{T'_1, \ldots, T'_n}{\rho(\psi)}
   \]
   is an \(S\)-proof of \(\rho(\psi)\) depending on
   \[
   (\Gamma_1 \setminus \{\rho(\theta_{1_1}), \ldots, \rho(\theta_{k_1})\}) \cup \cdots \cup (\Gamma_n \setminus \{\rho(\theta_{1_n}), \ldots, \rho(\theta_{k_n})\}).
   \]
In the sequel we shall use the following notation:

\[ \tau \]
\[ A \]

to mean that \( \tau \) is a formula-tree whose root is an occurrence of \( A \);

\[ [B_1], \ldots, [B_m] \]
\[ \tau \]
\[ A \]

to mean that \( \tau \) is a formula-tree whose last step is an application of a proof rule that yields \( A \) as conclusion and discharges the assumptions \( B_1, \ldots, B_m \);

\[ \tau_1 \]
\[ B \]
\[ \tau_2 \]
\[ A \]

to denote the result of replacing all the occurrences of \( B \) in the leaves of \( \tau_2 \) with the tree \( \tau_1 \) whose root is an occurrence of \( B \).

**Definition 2.** Given an S-proof \( \tau \), an immediate subproof of \( \tau \) is defined as follows:

1. if \( \tau \) has the form

\[
\begin{array}{c}
\tau_1 \ldots \tau_n \\
A_1 \ldots A_n \\
B
\end{array}
\]

where \( A_1, \ldots, A_n \) is an instance of an inference rule, then \( \tau_1, \ldots, \tau_n \) are immediate subproofs of \( \tau \);

2. if \( \tau \) has the form

\[
\begin{array}{c}
[A_1], \ldots, [A_{k_1}] \\
\tau_1 \\
B_1 \\
\vdots \\
[A_{k_n}], \ldots, [A_{k_n}] \\
\tau_n \\
B_n \\
C
\end{array}
\]

where \( \tau_1, \ldots, \tau_n/C \) is an instance of a proof rule, then \( \tau_1, \ldots, \tau_n \) are immediate subproofs of \( \tau \).

3. nothing else is an immediate subproof of \( \tau \).

We also say that \( \tau' \) is a subproof of \( \tau \) if there is a sequence \( \tau_0, \ldots, \tau_k \), with \( k \geq 0 \), such that (i) \( \tau_0 = \tau \), (ii) \( \tau_k = \tau' \), (iii) \( \tau_{i+1} \) is an immediate subproof of \( \tau_i \).

Note that, according to the above definition, every S-proof is a subproof of itself (taking \( k = 0 \)). We say that \( \tau' \) is a strict subproof of \( \tau \) when \( \tau' \neq \tau \).

The paradigmatic example of a natural deduction system is Prawitz’s system investigated in [Prawitz 1965].
Prawitz’s rules for natural deduction are shown in Table 1. In the intuitionistic system, the set \( \mathcal{I} \) of the inference rules consists of \( \land \), \( \land \), \( \lor \), \( \rightarrow \), \( \rightarrow \) and \( \land \), while the set \( \mathcal{P} \) of the proof rules consists of \( \forall \), \( \forall \), \( \rightarrow \), \( \land \). This system coincides with Gentzen’s system NJ of [Gentzen 1935]. A system for classical logic is obtained by replacing the inference rule \( \land \) with the proof rule \( \land \). Note that the intuitionistic falsum rule is a special case of the classical one that arises when the assumption \( \neg A \) is vacuously discharged. As mentioned above, Gentzen’s system NK used the law of excluded middle \( A \lor \neg A \) as an axiom, i.e. an extra assumption that can be freely introduced in proof, instead of \( \land \). He also mentioned that this axiom could have been equivalently replaced by the inference rule \( \neg \neg A \).

**Example 3.** Here is a proof in Gentzen-Prawitz style of the following inference

\[
A \rightarrow (B \rightarrow C) \vdash (A \land B) \rightarrow C
\]
A proof the theorem:

\[ \vdash A \rightarrow (B \rightarrow C) \rightarrow (A \land B) \rightarrow C \]

(intended as an inference from the empty set of assumptions) can be immediately obtained by discharging the only assumption yet undischarged in the above proof by means of an introduction of \( \rightarrow \):

\[
\begin{align*}
\frac{\frac{\frac{[A \land B]^1}{A \rightarrow (B \rightarrow C) \quad A \quad A \land B}{B \rightarrow C \quad B}}{C \quad (A \land B) \rightarrow C^1}{A \rightarrow (B \rightarrow C) \rightarrow (A \land B) \rightarrow C^2}
\end{align*}
\]

2 Gentzen-style natural deduction for classical logic

It is well-known that Gentzen-style natural deduction provides a natural formalization of intuitionistic logic, but a quite unnatural formalization of classical logic. Gentzen himself observed that his calculus \( NK \) was obtained by adding to the intuitionistic calculus \( NJ \) the law of excluded middle in “a purely external manner” \( [Szabo\,1969,\,p.\,81] \). This approach is illuminating to clarify the relationship between the two logical systems. However, it does so from the point of view of intuitionistic logic. Then, it is only to be expected that the \( NK \) proof of an inference that is classically, but not intuitionistically, valid turns out to be rather unnatural. The main reason is that the introduction and elimination rules for the logical operators, being tailored to their intuitionistic meaning, inevitably fail to represent faithfully the inner symmetries that emerge from their classical meaning. The same holds true for Prawitz’s variant that consists in replacing the intuitionistic ex-falso rule and the law of excluded middle with classical reductio. To illustrate this point, let us consider a proof in Prawitz’s system of the intuitionistically valid inference:

\[ \neg A \lor \neg B \vdash \neg (A \land B) \quad (1) \]

We use numerals to keep track of the assumptions that are discharged by the application of a proof rule. The numerals corresponding to the discharged assumptions are shown
beside the inference line.

\[
\begin{align*}
&\quad \text{[} A \land B \text{]}^3 \\
&\quad \text{[} \neg A \text{]}^1 \quad A \quad \text{[} \neg B \text{]}^2 \quad B \\
\hdashline
\neg A \lor \neg B \\
\quad \text{[} \neg (A \land B) \text{]}^3 \\
\quad \text{[} A \land B \text{]}^3 \\
\quad \text{[} A \land B \text{]}^3_{1,2}
\end{align*}
\] (2)

Let us now consider a standard proof of the reverse inference:

\[
\neg (A \land B) \vdash \neg A \lor \neg B
\] (3)

which is not intuitionistically valid.

\[
\begin{align*}
&\quad \text{[} \neg A \text{]}^2 \\
&\quad \text{[} \neg (\neg A \lor \neg B) \text{]}^1 \quad \neg (\neg A \lor \neg B) \quad \text{[} \neg B \text{]}^3 \\
\hdashline
\quad \text{[} A \lor \neg B \text{]}^1_{1,2} \\
\quad A \\
\quad B \\
\end{align*}
\] (4)

This proof is quite unnatural from the classical point of view in that it does not exploit the inner symmetries of classical logic. Indeed, it is very different from the previous one although, in a classical setting, the two proofs should be essentially the same, modulo the duality of \( \lor \) and \( \land \).

As suggested by these two examples, standard natural deduction is really natural from the point of view of intuitionistic logic, but it is not natural from the point of view of classical logic. The reason is that its introduction and elimination rules are faithful to the intuitionistic meaning of the logical operators and not to their classical meaning. If we are interested in a deduction system that is really natural for classical logic, we need introduction and elimination rules that closely reflect the classical meaning of the logical operators and the way in which they are used in classical proofs. In such a system a formula and its negation should be treated symmetrically. Moreover, the conjunction and disjunction operators should be governed by dual rules. This is the case, for instance, of the tableau method, where we have tableau rules for a complex formula of a given logical form and for its negation and where the rules for \( \land \) and \( \lor \) are dual of each other. However, the interplay between introduction and elimination rules as well as the possibility of generating direct proofs, that are typical of natural deduction, are inevitably lost in the tableau method that uses only elimination rules and can prove a conclusion from a set of assumptions only by refuting its negation.

\[2\] Indeed, Smullyan once claimed that tableaux could indeed be presented as a sort of natural deduction for classical logic (Smullyan [1965]).
References


