Note that the second premise of the $\supset$L rule is an unfocused sequent. From a practical point of view it is important to continue with the focusing steps in the first premise before attempting to prove the second premise, because the decomposition of B may ultimately fail when an atomic proposition is reached. Such a failure would render the possibly difficult proof of A useless.

It is possible to extend the definition of $\overline{L}^+$ to include conjunction and $\top$ and remove the left focus rules for conjunction. In some situations this would clearly lead to shorter proofs, but the present version appears to have less disjunctive non-determinism.[3]

Initial Sequents. There is a slight, but important asymmetry in the initial sequents: we require that we have focused on the left proposition.

$$\frac{}{\Delta; P \gg \cdot; P}\,\text{init}$$

Since this is the only rule which can be applied when the left focus formula is atomic, a proof attempt fails in a situation where $\Delta; P \gg \cdot; Q$ for $P \neq Q$. This is a very important property of the search, limiting non-determinism in focusing.

If one shows only applications of the decision rules in a derivation, the format is very close to *assertion-level proofs* as proposed by Huang [Hua94]. His motivation was the development of a formalism appropriate for the presentation of mathematical proofs in a human-readable form. This provides independent evidence for the value of focusing proofs. Focusing derivations themselves were developed by Andreoli [And92] in the context of classical linear logic. An adaptation to intuitionistic linear logic was given by Howe [How98] which is related the calculus LJT devised by Herbelin [Her95]. Herbelin's goal was to devise a sequent calculus whose derivations are in bijective correspondence to normal natural deductions. Due to the $\vee$, $\perp$ and $\exists$ elimination rules, this is not the case here.

The search procedure which works with focusing sequents is similar to the one for inversion: it mixes conjunctive non-determinism for active rules with disjunctive non-determinism for choice and focused rules. After the detailed development of inversion proofs, we will not repeat or extend the development here, but refer the interested reader to the literature. The techniques are very similar to the ones shown in Section 4.1.

4.3 Unification

When proving a proposition of the form $\exists x.\ A$ by its right rule in the sequent or focusing calculus, we must supply a term t and then prove $[t/x]A$. The domain of quantification may include infinitely many terms (such as the natural

[3] *[evaluate]*

numbers), so this choice cannot be resolved simply by trying all possible terms t. Similarly, when we use a hypothesis of the form $\forall x.\, A$ we must supply a term t to substitute for x. We refer to this a *existential non-determinism.*

Fortunately, there is a technique called *unification* which is sound and complete for syntactic equality between terms. The basic idea is quite simple: we postpone the choice of t and instead substitute a new *existential variable* (often called *meta-variable* or *logic variable*) X for x and continue with the bottom-up construction of a derivation. When we reach initial sequents we check if there is a substitution for the existential variables such that the hypothesis matches the conclusion. If so, we apply this instantiation globally to the partial derivation and continue to search for proofs of other subgoals. Finding an instantiation for existential variables under which two propositions or terms match is called *unification.* It is decidable if a unifying substitution or *unifier* exists, and if so, we can effectively compute it in linear time. Moreover, we can do so with a minimal commitment and we do not need to choose between various possible unifiers.

Because of its central importance in both backward- and forward-directed search, unification has been thoroughly investigated. Herbrand [Her30] is given credit for the first description of a unification algorithm in a footnote of his thesis, but it was not until 1965 that it was introduced into automated deduction through the seminal work by Alan Robinson [Rob65, Rob71]. The first algorithms were exponential, and later almost linear [Hue76, MM82] and linear algorithms [MM76, PW78] were discovered. In the practice of theorem proving, generally variants of Robinson's algorithm are still used, due to its low constant overhead on the kind of problems encountered in practice. For further discussion and a survey of unification, see [Kni89]. We describe a variant of Robinson's algorithm.

Before we describe the unification algorithm itself, we relate it to the problem of proof search. We use here the sequent calculus with atomic initial sequents, but it should be clear that precisely the same technique of *residuation* applies to focused derivations. We enrich the judgment $\Gamma \stackrel{-}{\Longrightarrow} A$ by a *residual proposition* F such that

1. if $\Gamma \stackrel{-}{\Longrightarrow} A$ then $\Gamma \stackrel{-}{\Longrightarrow} A \setminus F$ and F is true, and

2. if $\Gamma \stackrel{-}{\Longrightarrow} A \setminus F$ and F is true then $\Gamma \stackrel{-}{\Longrightarrow} A$.

Generally, we cannot prove such properties directly by induction, but we need to generalize them, exhibiting the close relationship between the derivations of the sequents and residual formulas F.

Residual formulas F are amenable to specialized procedures such as unification, since they are drawn from a simpler logic or deductive system than the general propositions A. In practice they are often solved *incrementally* rather than collected throughout a derivation and only solved at the end. This is important for the early detection of failures during proof search. Incremental solution of residual formulas is the topic of Exercise **??**.

What do we need in the residual propositions so that existential choices and equalities between atomic propositions can be expressed? The basic proposition is one of equality between atomic propositions, $P_1 \doteq P_2$. We also have conjunction $F_1 \wedge F_2$, since equalities may be collected from several subgoals, and $\top$ if there are no residual propositions to be proven. Finally, we need the existential quantifier $\exists x.\, F$ to express the scope of existential variables, and $\forall x.\, F$ to express the scope of parameters introduced in a derivation. We add equality between terms, since it is required to describe the unification algorithm itself. We refer to the logic with these connectives as *unification logic*, defined via a deductive system.

$$\textit{Formulas} \quad F \quad ::= \quad P_1 \doteq P_2 \mid t_1 \doteq t_2 \mid F_1 \wedge F_2 \mid \top \mid \exists x.\, F \mid \forall x.\, F$$

The main judgment "F *is valid*", written $\models F$, is defined by the following rules, which are consistent with, but more specialized than the rules for these connectives in intuitionistic natural deduction (see Exercise **??**).

$$\frac{}{\models P \doteq P}\doteq\mathrm{I} \qquad\qquad \frac{}{\models t \doteq t}\doteq\mathrm{I}'$$

$$\frac{\models F_1 \qquad \models F_2}{\models F_1 \wedge F_2}\wedge\mathrm{I} \qquad\qquad \frac{}{\models \top}\top\mathrm{I}$$

$$\frac{\models [t/x]F}{\models \exists x.\, F}\exists\mathrm{I} \qquad\qquad \frac{\models [a/x]F}{\models \forall x.\, F}\forall\mathrm{I}^a$$

The $\forall\mathrm{I}^a$ rule is subject to the usual proviso that a is a new parameter not occurring in $\forall x.\, F$. There are no elimination rules, since we do not need to consider hypotheses about the validity of a formula F which is the primary reason for the simplicity of theorem proving in the unification logic.

We enrich the sequent calculus with residual formulas from the unification logic, postponing all existential choices. Recall that in practice we merge residuation and solution in order to discover unprovable residual formulas as soon as possible. This merging of the phases is not represented in our system.

Initial Sequents. Initial sequents residuate an equality between its principal propositions. Any solution to the equation will unify P' and P, which means that this will translate to a correct application of the initial sequent rule in the original system.

$$\frac{}{\Gamma, P' \stackrel{-}{\Longrightarrow} P \setminus P' \doteq P}\mathrm{init}$$

Propositional Connectives. We just give a few sample rules for the connectives which do not involve quantifiers, since all of them simply propagate or combine unification formulas, regardless whether they are additive, multiplicative, or exponential.

$$\frac{\Gamma, A \overset{=}{\Longrightarrow} B \setminus F}{\Gamma \overset{=}{\Longrightarrow} A \supset B \setminus F}\supset\!\mathrm{R} \qquad \frac{}{\Gamma \overset{=}{\Longrightarrow} \top \setminus \top}\top\mathrm{R}$$

$$\frac{\Gamma, A \supset B \overset{=}{\Longrightarrow} A \setminus F_1 \qquad \Gamma, A \supset B, B \overset{=}{\Longrightarrow} C \setminus F_2}{\Gamma, A \supset B \overset{=}{\Longrightarrow} C \setminus F_1 \wedge F_2}\supset\!\mathrm{L}$$

Quantifiers. These are the critical rules. Since we residuate the existential choices entirely, the $\exists$R and $\forall$L rules instantiate a quantifier by a new *parameter*, which is existentially quantified in the residual formula in both cases. Similarly, the $\forall$R and $\exists$L rule introduce a parameter which is universally quantified in the residual formula.

$$\frac{\Gamma \overset{=}{\Longrightarrow} [a/x]A \setminus [a/x]F}{\Gamma \overset{=}{\Longrightarrow} \forall x.\ A \setminus \forall x.\ F}\forall\mathrm{R}^a \qquad \frac{\Gamma, \forall x.\ A, [a/x]A \overset{=}{\Longrightarrow} C \setminus [a/x]F}{\Gamma, \forall x.\ A \overset{=}{\Longrightarrow} C \setminus \exists x.\ F}\forall\mathrm{L}^a$$

$$\frac{\Gamma \overset{=}{\Longrightarrow} [a/x]A \setminus [a/x]F}{\Gamma \overset{=}{\Longrightarrow} \exists x.\ A \setminus \exists x.\ F}\exists\mathrm{R}^a \qquad \frac{\Gamma, \exists x.\ A, [a/x]A \overset{=}{\Longrightarrow} C \setminus [a/x]F}{\Gamma, \exists x.\ A \overset{=}{\Longrightarrow} C \setminus \forall x.\ A}\exists\mathrm{L}^a$$

The soundness of residuating equalities and existential choices in this manner is straightforward.

Theorem 4.9 (Soundness of Equality Residuation)
If $\Gamma \overset{=}{\Longrightarrow} A \setminus F$ *and* $\models F$ *then* $\Gamma \overset{=}{\Longrightarrow} A$.

Proof: By induction on the structure of the given derivation $\mathcal{R}$. We show the critical cases. Note how in the case of the $\exists$R rule the derivation of $\models \exists x.\ F$ provides the essential witness term t.

Case:

$$\mathcal{R} = \frac{}{\Gamma, P' \overset{=}{\Longrightarrow} P \setminus P' \doteq P}\mathrm{init}$$

$\models P' \doteq P$	By assumption
$P' = P$	By inversion
$\Gamma, P' \overset{=}{\Longrightarrow} P$	By rule init

Case:

$$\mathcal{R} = \frac{\begin{array}{c}\mathcal{R}_1\\ \Gamma \overset{=}{\Longrightarrow} [a/x]A_1 \setminus [a/x]F_1\end{array}}{\Gamma \overset{=}{\Longrightarrow} \exists x.\ A_1 \setminus \exists x.\ F_1}\exists\mathrm{R}^a$$

$\models \exists x.\ F_1$	By assumption
$\models [t/x]F_1$ for some t	By inversion
$\Gamma \overset{-}{\Longrightarrow} [t/x]A_1 \setminus [t/x]F_1$	By substitution for parameter a
$\Gamma \overset{-}{\Longrightarrow} [t/x]A_1$	By i.h.
$\Gamma \overset{-}{\Longrightarrow} \exists x.\ A_1$	By rule $\exists$R

Case:

$$\mathcal{R} = \cfrac{\begin{array}{c}\mathcal{R}_1\\ \Gamma \overset{-}{\Longrightarrow} [a/x]A_1 \setminus [a/x]F_1\end{array}}{\Gamma \overset{-}{\Longrightarrow} \forall x.\ A_1 \setminus \forall x.\ F_1}\ \forall\mathrm{R}^a$$

$\models \forall x.\ F_1$	By assumption
$\models [b/x]F_1$ for a new parameter b	By inversion
$\models [a/x]F_1$	By substititution of a for b
$\Gamma \overset{-}{\Longrightarrow} [a/x]A_1$	By i.h.
$\Gamma \overset{-}{\Longrightarrow} \forall x.\ A_1$	By rule $\forall$R

$\square$

The opposite direction is more difficult. The desired theorem:

If $\Gamma \overset{-}{\Longrightarrow} A$ *then* $\Gamma \overset{-}{\Longrightarrow} A \setminus F$ *for some* F *with* $\models F$

cannot be proved directly by induction, since the premisses of the two derivations are different in the $\exists$R and $\forall$L rules. However, one can be obtained from the other by substituting terms for parameters. Since this must be done simultaneously, we introduce a new notation.

$$\textit{Parameter Substitution} \quad \rho \quad ::= \quad \cdot \mid \rho, t/a$$

We assume all the parameters a substituted for by ρ are distinct to avoid ambiguity. We write $A[\rho]$, $F[\rho]$, and $\Gamma[\rho]$, for the result of applying the substitution ρ to a proposition, formula, or context, respectively.

Lemma 4.10 *If* $\Gamma \overset{-}{\Longrightarrow} A$ *where* $A = A'[\rho]$, $\Gamma = \Gamma'[\rho]$ *then* $\Gamma' \overset{-}{\Longrightarrow} A' \setminus F$ *for some* F *such that* $\models F[\rho]$.

Proof: The proof proceeds by induction on the structure of the given derivation $\mathcal{D}$. We show only two cases, the second of which required the generalization of the induction hypothesis.

Case:

$$\mathcal{D} = \cfrac{}{\Gamma_1, P \overset{-}{\Longrightarrow} P}\ \mathrm{init}$$

$\Gamma_1 = \Gamma_1'[\rho]$, $P = P'[\rho]$, and $P = P''[\rho]$ — Assumption
$\Gamma_1', P' \stackrel{=}{\Longrightarrow} P'' \setminus P' \doteq P''$ — By rule init
$\models P'[\rho] \doteq P''[\rho]$ — By rule $\doteq$I

Case:

$$\mathcal{D} = \cfrac{\begin{array}{c}\mathcal{D}_1\\ \Gamma \stackrel{=}{\Longrightarrow} [t/x]A_1\end{array}}{\Gamma \stackrel{=}{\Longrightarrow} \exists x.\ A_1}\ \exists\mathrm{R}$$

$\exists x.\ A_1 = A'[\rho]$ — Assumption
$A' = \exists x.\ A_1'$ for a new parameter a with
$[a/x]A_1 = ([a/x]A_1')[\rho, a/a]$ — By definition of substitution
$[t/x]A_1 = ([a/x]A_1')[\rho, t/a]$ — By substitution for parameter a
$\Gamma = \Gamma'[\rho]$ — Assumption
$\Gamma'[\rho] = \Gamma'[\rho, t/a]$ — Since a is new
$\Gamma' \stackrel{=}{\Longrightarrow} [a/x]A_1' \setminus [a/x]F_1$, and
$\models ([a/x]F_1)[\rho, t/a]$ — By i.h.
$\Gamma' \stackrel{=}{\Longrightarrow} \exists x.\ A_1' \setminus \exists x.\ F_1$ — By rule $\exists$R
$\models (\exists x.\ F_1)[\rho]$ — By rule $\exists$R and definition of substitution

□

Theorem 4.11 (Completeness of Equality Residuation)
If $\Gamma \stackrel{=}{\Longrightarrow} A$ *then* $\Gamma \stackrel{=}{\Longrightarrow} A \setminus F$ *for some* F *and* $\models F$.

Proof: From Lemma 4.10 with $A' = A$, $\Gamma' = \Gamma$, and ρ the identity substitution on the parameters in Γ and A. □

4.4 Exercises

Exercise 4.1 Give an alternative proof of the inversion properties (Theorem 4.1) which does not use induction, but instead relies on admissibility of cut in the sequent calculus (Theorem 3.11).

Exercise 4.2 Formulate one or several cut rules directly on inversion sequents as presented in Section 4.1 and prove that they are admissible. Does this simplify the development of the completeness result for inversion proofs? Show how admissibility might be used, or illustrate why it is not much help.

Bibliography

[And92] Jean-Marc Andreoli. Logic programming with focusing proofs in linear logic. *Journal of Logic and Computation*, 2(3):197–347, 1992.

[Byr99] John Byrnes. *Proof Search and Normal Forms in Natural Deduction*. PhD thesis, Department of Philosophy, Carnegie Mellon University, May 1999.

[Cur30] H.B. Curry. Grundlagen der kombinatorischen Logik. *American Journal of Mathematics*, 52:509–536, 789–834, 1930.

[Gen35] Gerhard Gentzen. Untersuchungen über das logische Schließen. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1935. Translated under the title *Investigations into Logical Deductions* in [Sza69].

[Her30] Jacques Herbrand. Recherches sur la théorie de la démonstration. *Travaux de la Société des Sciences et de Lettres de Varsovic*, 33, 1930.

[Her95] Hugo Herbelin. *Séquents qu'on calcule*. PhD thesis, Universite Paris 7, January 1995.

[Hil22] David Hilbert. Neubegründung der Mathematik (erste Mitteilung). In *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, pages 157–177, 1922. Reprinted in [Hil35].

[Hil35] David Hilbert. *Gesammelte Abhandlungen*, volume 3. Springer-Verlag, Berlin, 1935.

[How69] W. A. Howard. The formulae-as-types notion of construction. Unpublished manuscript, 1969. Reprinted in To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism, 1980.

[How98] Jacob M. Howe. *Proof Search Issues in Some Non-Classical Logics*. PhD thesis, University of St. Andrews, Scotland, 1998.

[Hua94] Xiarong Huang. *Human Oriented Proof Presentation: A Reconstructive Approach*. PhD thesis, Universität des Saarlandes, 1994.

[Hue76] Gérard Huet. *Résolution d'équations dans des langages d'ordre* $1, 2, \ldots, \omega$. PhD thesis, Université Paris VII, September 1976.

[Kle52] Stephen Cole Kleene. *Introduction to Metamathematics*. North-Holland, 1952.

[Kni89] Kevin Knight. Unification: A multi-disciplinary survey. *ACM Computing Surveys*, 2(1):93–124, March 1989.

[LS86] Joachim Lambek and Philip J. Scott. *Introduction to Higher Order Categorical Logic*. Cambridge University Press, Cambridge, England, 1986.

[ML85a] Per Martin-Löf. On the meanings of the logical constants and the justifications of the logical laws. Technical Report 2, Scuola di Specializzazione in Logica Matematica, Dipartimento di Matematica, Università di Siena, 1985.

[ML85b] Per Martin-Löf. Truth of a proposition, evidence of a judgement, validity of a proof. Notes to a talk given at the workshop *Theory of Meaning*, Centro Fiorentino di Storia e Filosofia della Scienza, June 1985.

[ML94] Per Martin-Löf. Analytic and synthetic judgements in type theory. In Paolo Parrini, editor, *Kant and Contemporary Epistemology*, pages 87–99. Kluwer Academic Publishers, 1994.

[MM76] Alberto Martelli and Ugo Montanari. Unification in linear time and space: A structured presentation. Internal Report B76-16, Ist. di Elaborazione delle Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy, July 1976.

[MM82] Alberto Martelli and Ugo Montanari. An efficient unification algorithm. *ACM Transactions on Programming Languages and Systems*, 4(2):258–282, April 1982.

[Par92] Michel Parigot. $\lambda\mu$-calculus: An algorithmic interpretation of classical natural deduction. In A. Voronkov, editor, *Proceedings of the International Conference on Logic Programming and Automated Reasoning*, pages 190–201, St. Petersburg, Russia, July 1992. Springer-Verlag LNCS 624.

[Pfe95] Frank Pfenning. Structural cut elimination. In D. Kozen, editor, *Proceedings of the Tenth Annual Symposium on Logic in Computer Science*, pages 156–166, San Diego, California, June 1995. IEEE Computer Society Press.

[Pra65] Dag Prawitz. *Natural Deduction*. Almquist & Wiksell, Stockholm, 1965.

[PW78] M. S. Paterson and M. N. Wegman. Linear unification. *Journal of Computer and System Sciences*, 16(2):158–167, April 1978.

[Rob65] J. A. Robinson. A machine-oriented logic based on the resolution principle. *Journal of the ACM*, 12(1):23–41, January 1965.

[Rob71] J. A. Robinson. Computational logic: The unification computation. *Machine Intelligence*, 6:63–72, 1971.

[Sza69] M. E. Szabo, editor. *The Collected Papers of Gerhard Gentzen*. North-Holland Publishing Co., Amsterdam, 1969.