Herbrand expansion proofs and proof identity

Richard McKinley*

Theoretische Informatik und Logik, Institut für Informatik und angewandte Mathematik,
Neubrückstrasse 10, CH-3012 Bern, Switzerland, mckinley@iam.unibe.ch

Abstract. We present Herbrand expansion proofs: an abstract representation of proofs in first-order classical logic, derived from a form of Herbrand's theorem. These objects are essentially Miller's *expansion tree proofs* with a notion of cut. We show a function mapping sequent calculus proofs to Herbrand expansion proofs, and a system of reductions which we conjecture gives cut-elimination for a subclass of HEPs including those which arise as the translations of sequent proofs.

1 Introduction

The question of when two proofs in (some formalism for) classical logic are morally equal has seen significant treatment in recent years [3, 4, 11, 10, 14, 9, 8, 12]; each treatment must find its own way of avoiding *Joyal's Paradox*; that the naïve way of extending the identity on proofs in intuitionistic natural deduction leads to collapse (every proof of a given sequent is identified). These studies are almost exclusively focused on propositional classical logic, for two reasons. First is that this problem is already hard; and the well-known barriers to a good semantics for classical logic exist in a quantifier (and indeed negation-) free setting. The second is a prejudice that the proof-theory of first-order quantifiers is uninteresting.

This workshop paper introduces *Herbrand expansion proofs* (HEPs) – an abstract representation of proofs in first-order classical logic. (The cut-free Herbrand expansion proofs can be seen as an alternative presentation of Miller's expansion tree proofs [13].) We see Herbrand expansion proofs as a modular semantics for proofs in first-order classical logic, in the sense that, given any notion of proof-identity on propositional proofs, we may produce a class of *augmented* Herbrand expansion proofs, incorporating that notion. What is surprising is that the notion of proof-identity we obtain from HEPs themselves is non-trivial, even when the notion of identity on propositional proofs is trivial. The problem of constructive cut-elimination for HEPs is nontrivial; the intuitive cut-reduction rules that we will present in this paper exhibit both non-confluence, and non-termination. We conjecture that, as in the sequent calculus, there is a strategy for cut-elimination, and that further, unlike the sequent calculus, this strategy can be made confluent in a non-trivial way, giving rise to a category of Herbrand expansion proofs.

* Supported by the Swiss National Science Foundation grant "Algebraic and Logical Aspects of Knowledge Processing."

1.1 Related work

Cut elimination for a system of expansion-tree-like objects for second-order propositional multiplicative linear logic may also be found in Strassburger [15], where the procedure is simpler owing to the lack of contractions. An abstract form of Herbrand's theorem is suggested in [8], which has a different character to that discussed here. The author has very recently become aware of the work of Willem Heijltjes, who is studying similar objects to those considered here.

1.2 Conventions

We work over a signature Σ of functions and relation symbols, in first-order logic without equality and without non-logical axioms. All formulae are assumed to be written in negation normal form (with negation pushed down to the atoms). Within a formula, we assume that bound variables are distinct from free variables, and that each quantifier binds a distinct bound variable , which we call its *eigenvariable*. We will say that a sequence S of formulae *satisfies the eigenvariable condition* if each quantifier appearing in a formula of S binds a distinct bound variable. We take the following definition of a sequent:

Definition 1.1. *A sequent Γ is a sequence $A_1, \dots, A_n$ of formulae satisfying the eigenvariable condition, such that no variable appears both free and bound in Γ.*

A lower case q will be used to denote a quantifier (either $\forall$ or $\exists$), and in context the symbol $\bar{q}$ will denote the dual quantifier. The metavariable X will stand for L or R, and $\bar{X}$ will stand, in context, for L if X is R, and vice versa.

2 Herbrand expansion proofs

Herbrand's theorem, as originally devised [7], introduces a system $\mathbf{H}$ (what we would now call a Hilbert system) for first-order classical logic. The axioms are all tautologies, and the inference rules are antiprenexing, universal and existential generalization, $\exists$-contraction (replacing a subformula $\exists x_1 A(x_1) \vee \exists x_2 A(x_2)$ by $\exists x A(x)$). In fact, any valid first-order formula has a proof in which first come the generalizations, then the antiprenexing operations, and finally the $\exists$-contractions. We may see such a proof as a triple of expansion $\hat{A}$, a prenexification of $\hat{A}$, and a substitution. An *expansion-tree proof* can be seen as a generalization of this proof system, where the explicit prenexification is replaced by a correctness criterion. A pair of expansion and substitution is correct if a certain *dependence relation* is acyclic. An *Herbrand expansion proof* will be a generalization of this structure, where a proof can in addition contain cuts. In this section, we describe this generalization.

Definition 2.1. *(a) An* n-hole-context *is a sequent with precisely n occurrences each of the special atom* $\{\}$*(the* hole*). We write* $\Gamma\{\}\dots\{\}$ *to denote an* n*-hole context.*

(b) If $\Gamma\{\}$ *is an* n*-hole context, we write* $\Gamma\{A_1\}\dots\{A_n\}$ *for the sequent given by replacing the leftmost hole with* A_1*, the next leftmost with* A_2*, and so on.*

We define the dual $\neg\Gamma\{\}$ of a context $\Gamma\{\}$ by setting $\neg\{\} := \{\}$ and using the usual definition of negation. The negation of $\Gamma\{A\}$ can be easily seen to be $\neg\Gamma\{\neg A\}$.

Definition 2.2. *Let A be a formula in negation normal form. An* expansion *of A is defined as follows: A is an expansion of A, and if $C\{\exists x.B(x)\}$ is an expansion of A, then $C\{\exists x_1.B(x_1) \vee \exists x_2.B(x_2)\}$ is an expansion of A.*

An expansion of a sequent $\Gamma = A_1 \ldots A_n$ is a sequent $\hat{\Gamma} = \hat{A}_1 \ldots \hat{A}_n$ of expansions of the members of Γ. Note that an expansion $\hat{A}$ of A induces a surjective mapping from the bound variables of $\hat{A}$ to the bound variables of A. Given a variable x in $\hat{A}$, denote the unique corresponding variable in A by $\check{x}$.

Definition 2.3. *Let A be a formula satisfying our variable restrictions. Define the* matrix *A^* of A by induction over the structure of A. If A is quantifier-free, then $A^* = A$, and otherwise*

$$(A * B)^* = A^* * B^*; \qquad (qx.B)^* = B^*.$$

Definition 2.4. *A* validating substitution *σ for a formula A is a substitution which assigns to each existentially bound variable of A a term, such that σA^* is valid.*

Given a formula B and a validating substitution σ for B, we say that they are *compatible* if there is a proof in system **H** starting with σB^* (an axiom of **H**), in which first we derive a formula $Q.B$ in prenex normal form using generalization, and then derive A from $Q.B$ using anti-prenexification. Clearly, if there is a compatible substitution for B, the B is valid. We can replace the existential condition of compatibility by using the *dependence relation*

Definition 2.5. *Let B be a formula, let V be the set of bound variables occurring in B, and let σ be a validating substitution for B. Define the following relations on V:*

*(a) **(scoping)** $y \ll x$ if $q.x$ occurs within the scope of $q.y$ in A;*
*(b) **(instantiation)** $y \sqsubset x$ if $y \in$ fv $\sigma(x)$;*
*(c) **(dependence relation)** $\lhd$ is the transitive closure of $(\ll \cup \sqsubset)$.*

Theorem 2.1. *Let B be a formula, and let σ be a validating substitution for B. Then σ and B are compatible if and only if $\lhd$ is irreflexive.*

Informally, a cut-free Herbrand expansion proof of a formula A is a pair of an expansion $\hat{A}$ of A, and a compatible substitution σ for $\hat{A}$. We extend this definition to sequents in an evident fashion (to be made formal when we introduce cuts).

Example 2.1. A cut-free Herbrand expansion proof of $(\exists x.A \vee \exists x.B) \vdash \exists x.(A \vee B)$ (written in one-sided form) is given by the pair of the sequent

$$(\forall y.(\neg A(y)) \wedge \forall z.(\neg B(z))), (\exists x_L(A(x_L) \vee B(x_L)) \vee \exists x_R(A(x_R) \vee B(x_R))).$$

and the substitution $[x_L := y, x_R := z]$.

The general definition of an Herbrand expansion proof involves cuts. We will represent a cuts by conjunctions $A \wedge \neg A$ of dual formulae. The intuition is that a proof of a sequent Γ with a cut is a proof of $\Gamma, A \wedge \neg A$. Note, however, that in a sequent proof, we can form a cut on a formula A in which a variable x is free, and then later universally quantify over x. We will need a similar freedom to have a variable free in a cut, but bound in the conclusion.

Definition 2.6. *Let Θ be a sequence $A_1|\ldots|A_n$ of formulae satisfying the eigenvariable condition (but not necessarily the separation of free and bound variables). We will call a pair of a sequent and such a sequence Θ an* extension *of Γ, writing $\Gamma \mid \Theta$ for the pair, if together they satisfy the eigenvariable condition. A* cut-extension *of Γ will be an extension $\Gamma \mid \Theta$ in which Θ is a sequence of conjunctions of dual formulae (a sequence of* cut-pairs*).*

The definition of the expansion of a cut B is a little delicate, since a variable might occur free in a cut but not in the conclusion.

Definition 2.7. *Let $\Gamma \mid \Theta$ be a cut-extension of Γ, and let $B \in \Theta$ be a cut-pair. Let $Y = \{y_1, \ldots, y_n\}$ be the set of variables free in B but bound elsewhere in $\Gamma \mid \Theta$.*

(a) An expansion of a cut B in context $\Gamma \mid \Theta$ *is a formula $\hat{B}'$, where $\hat{B}'$ is the expansion of a formula B', and $B' = B[y_1 := z_1, \ldots y_n := z_n]$, for any variables $z_1, \ldots, z_n$. An expansion of $\Theta = B_1, \ldots, B_n$ is a sequence $\hat{B}'_1, \ldots, \hat{B}'_n$.*
(b) An expansion *of $\Gamma \mid \Theta$ is a pair of an expansion of Γ and an expansion of Θ, written $\hat{\Gamma} \mid \hat{\Theta}$. The matrix $(\Gamma \mid \Theta)^*$ of $\Gamma \mid \Theta$ is the matrix of the sequence Γ, Θ.*

Where possible, we will use the following displayed notation for a pair of an extension and an expansion of Γ:

$$\begin{array}{c}\hat{\Gamma} \mid \hat{\Theta} \\ \curlyvee \\ \Gamma \mid \Theta\end{array}$$

Definition 2.8. *A* validating substitution *σ for $\Gamma \mid \Theta$ is a substitution which assigns to each existentially bound variable of $\hat{\Gamma} \mid \hat{\Theta}$ a term, such that $\sigma(\Gamma \mid \Theta)^*$ is valid.*

We use the notion of pre-proof (also often referred to as a proof-structure) for something which has the form of a proof, but which does not necessarily correspond to a genuine proof.

Definition 2.9 (Herbrand pre-proof). *An Herbrand pre-proof Φ of a sequent Γ is a triple of a cut-extension $\Gamma \mid \Theta$ of Γ, an expansion $\hat{\Gamma} \mid \hat{\Theta}$ of $\Gamma \mid \Theta$, and a validating substitution σ_Φ for $\hat{\Gamma} \mid \hat{\Theta}$.*

To see that a pre-proof is correct (that it should be regarded as a proof) we generalize the notion of a *dependence relation* for expansion-tree proofs: $\hat{\Gamma} \mid \hat{\Theta}$, taking into account the fact that variables may be free in a cut but bound elsewhere.

Definition 2.10. *Let $\Phi = (\Gamma \mid \Theta,\ \hat{\Gamma} \mid \hat{\Theta}, \sigma_\Phi)$ be an Herbrand pre-proof. Define the following relations on the variables bound in $\hat{\Gamma} \mid \hat{\Theta}$:*

(a) ***(scoping)*** *$y \ll x$ if $q.x$ occurs within the scope of $q.y$; and*
(b) ***(instantiation)*** *$y \sqsubset x$ if $y \in fv\ \sigma(x)$, or if x is bound in the expansion of a cut-pair $A_i \wedge \neg A_i$ in which y is free.*
(c) ***(dependence relation)*** *$\lhd_\Phi$ of Φ is the transitive closure of the union of $\ll$ and $\sqsubset$.*

Definition 2.11 (Correctness). *An Herbrand expansion proof (HEP) is an Herbrand pre-proof Φ for which $\lhd_\Phi$ is irreflexive. We will call an HEP of a sequent Γ* cut-free *if the extension of Γ in the proof is the trivial extension.*

We establish completeness for this system by showing that any sequent possessing a proof in the sequent calculus also has an HEP. This has the bonus of giving a function mapping sequent proofs to HEPs. Most of the cases of this proof are trivial, but the case of contraction is more problematic. We show instead the admissibility of a "deep" contraction rule, and will need the following easy fact about expansions:

Fact 1. An expansion of a sequent $\Gamma\{A\}$ has the form $\hat{\Gamma}\{A_1\}\ldots\{A_n\}$, where $A_1 \ldots A_n$ are expansions of A, and $\hat{\Gamma}\{\}\ldots\{\}$ is an expansion of $\Gamma\{\}$.

Theorem 2.2. *Let **GS** be the usual context splitting, one-sided, sequent calculus for first-order classical logic, with explicit structural rules. For each rule ρ of **GS**, and for the cut-rule, given Herbrand expansion proofs of the premises we may construct an Herbrand expansion proof of the conclusion.*

Proof. Ax: An Herbrand proof of $a, \neg a$ (where a is atomic) is given by

$$(a, \neg a,\ a, \neg a,\ \emptyset).$$

Since there are no cuts, it is trivially correct.

$\vee$R: Given a correct Herbrand proof

$$(\Gamma, A \vee B \mid \Theta,\ \hat{\Gamma}, \hat{A} \vee \hat{B} \mid \hat{\Theta},\ \sigma)$$

of Γ, A, B, the pair

$$(\Gamma, A \vee B \mid \Theta,\ \hat{\Gamma}, \hat{A} \vee \hat{B} \mid \hat{\Theta},\ \sigma)$$

is a correct Herbrand proof of $\Gamma, A \vee B$.

$\wedge$R: Given a correct Herbrand expansion proof $\Phi_1 = (\Gamma, A \mid \Theta,\ \hat{\Gamma}, \hat{A} \mid \hat{\Theta},\ \sigma_1)$ of Γ, A, and a correct Herbrand expansion proof $\Phi_2 = (\Delta, B \mid \Theta',\ \hat{\Delta}, \hat{B} \mid \hat{\Theta}',\ \sigma_2)$ of Δ, B, the triple

$$\Phi_3 = (\Gamma, \Delta, A \wedge B \mid \Theta \mid \Theta',\ \hat{\Gamma}, \hat{\Delta}, \hat{A} \wedge \hat{B} \mid \hat{\Theta} \mid \hat{\Theta}',\ \sigma_1 \cup \sigma_2)$$

is an Herbrand pre-proof of $\Gamma, \Delta, A \wedge B$. To see correctness, observe that the reduction ordering on Φ_3 is the disjoint union of the reduction orderings on Φ_1 and Φ_2.

$\exists$R: Given an HEP

$$\Phi := (\Gamma, A[x := t] \mid \Theta,\ \hat{\Gamma}, \hat{A}[x := t] \mid \hat{\Theta},\ \sigma)$$

of $\Gamma, A(t)$,

$$\Psi := (\Gamma, \exists x.A \mid \Theta,\ \hat{\Gamma}, \exists x.\hat{A} \mid \hat{\Theta},\ \sigma \cup \{\langle y, t\rangle\})$$

is an Herbrand pre-proof. The relation $\lhd_\Psi$ is irreflexive as there is no bound variable y such that $y \sqsubset x$ (since otherwise y would appear both free and bound in the conclusion of Φ.)

$\forall R$: Given an Herbrand proof

$$\Phi := (\Gamma, A \mid \Theta,\ \hat{\Gamma}, \hat{A} \mid \hat{\Theta},\ \sigma)$$

of Γ, A,

$$\Phi := (\Gamma, \forall x.A \mid \Theta,\ \hat{\Gamma}, \forall x.\hat{A} \mid \hat{\Theta},\ \sigma)$$

is a pre-proof. To see correctness, observe that there is no y such that $y \ll x$.

W: Given an Herbrand proof

$$\Phi := (\Gamma \mid \Theta,\ \hat{\Gamma} \mid \hat{\Theta},\ \sigma)$$

The triple

$$\Psi := (\Gamma, A \mid \Theta,\ \hat{\Gamma}, A \mid \hat{\Theta},\ \sigma')$$

is a correct HEP of Γ, A, where σ' is the union of σ and a substitution assigning a fresh unbound variable to each existentially bound variable of A.

C: We show admissibility for correct HEPs of a deep contraction rule

$$\frac{\Gamma\{A \vee A\}}{\Gamma\{A\}}\,\mathrm{DeepC}$$

by induction over the structure of A. Admissibility of usual (shallow) contraction follows as a corollary.

Consider first the case where A is atomic. Suppose we have a correct HEP

$$(\Gamma\{a \vee a\} \mid \Theta,\ \hat{\Gamma}\{a \vee a\} \ldots \{a \vee a\} \mid \hat{\Theta},\ \sigma)$$

of $\Gamma\{a \vee a\}$. Then clearly

$$(\Gamma\{a\} \mid \Theta,\ \hat{\Gamma}\{a\} \ldots \{a\} \mid \hat{\Theta},\ \sigma)$$

is a correct HEP of $\Gamma\{a\}$. Now assume that DeepC is admissible for formula B and C. We show that it holds for $B \wedge C$, $B \vee C$, $\exists z.B$ and $\forall z.B$.

(a) An correct HEP of $\Gamma\{(B \vee C) \vee (B \vee C)\}$ has the form

$$(\Gamma\{(B{\vee}C){\vee}(B{\vee}C)\} \mid \Theta,\ \hat{\Gamma}\{(B_1{\vee}C_1){\vee}(B_2{\vee}C_2)\} \ldots \{(B_{n-1}{\vee}C_{n-1}){\vee}(B_n{\vee}C_n)\} \mid \hat{\Theta},\ \sigma).$$

Clearly

$$(\Gamma\{(B{\vee}B){\vee}(C{\vee}C)\} \mid \Theta,\ \hat{\Gamma}\{(B_1{\vee}B_2){\vee}(C_1{\vee}C_2)\} \ldots \{(B_{n-1}{\vee}B_n){\vee}(B_{n-1}{\vee}C_n)\} \mid \hat{\Theta},\ \sigma)$$

is an HEP of $\Gamma\{(B \vee B) \vee (C \vee C)\}$. This HEP is correct, since no new cuts or quantifiers have been formed. Apply the induction hypothesis to obtain an HEP of $\Gamma\{B \vee C\}$.

(b) An HEP of $\Gamma\{\exists x.B \vee \exists y.B\}$ has the form

$$(\Gamma\{\exists x.B \vee \exists y.B\} \mid \Theta,\ \hat{\Gamma}\{A_1 \vee A_2\} \dots \{A_{n-1} \vee A_n\} \mid \hat{\Theta},\, \sigma).$$

But this is also an HEP of $\Gamma\{\exists x.B\}$, since if C and D are expansions of an existential formula $\exists x.B$, then so is $C \vee D$. This HEP is correct, since no new cuts or quantifiers have been formed.

(c) An HEP of $\Gamma\{\forall x.B \vee \forall y.B\}$ has the form

$$\Phi := (\Gamma\{\forall x.B \vee \forall y.B\} \mid \Theta,\ \hat{\Gamma}\{\forall x_1 B_1 \vee \forall y_1.B_2\} \dots \{\forall x_n B_{2n-1} \vee \forall y_n.B_{2n}\} \mid \hat{\Theta},\, \sigma_1).$$

The following is then clearly a pre-proof:

$$\begin{aligned}\Psi := (&\Gamma\{\forall z.(B \vee B)\} \mid \Theta[x := z, y := z],\\ &\hat{\Gamma}\{\forall z_1(B_1 \vee B_2)\} \dots \{\forall z_m(B_{2n-1} \vee B_{2n})\} \mid \hat{\Theta}[x_i := z_i, y_i := z_i],\, \sigma_2).\end{aligned}$$

where σ_2 is $[x_i := z_i, y_i := z_i] \circ \sigma_1$. Suppose that Ψ is not correct. Then there is a path in $\ll_\Psi \cup \sqsubset_\Psi$, which wlog begins and ends with z_i, since if the path does not touch any z_i then it is also a path in $\ll_\Phi \cup \sqsubset_\Phi$. That path has the form

$$z_i \sqsubset a_1 \sqsubset \dots a_n \ll z_i$$

This path lifts to an evident path in $\ll_\Phi \cup \sqsubset_\Phi$. Observe that, a_n must be a variable appearing in Φ, and that $a_n \ll_\Phi x_i, y_i$. An easy induction on n yields that there is a cycle on $\ll_\Psi \cup \sqsubset_\Psi$ if and only id there is a cycle on $\ll_\Phi \cup \sqsubset_\Phi$, and therefore $\triangle_\Psi$ is irreflexive.

By the induction hypothesis, we derive an Herbrand expansion proof of $\Gamma\{\forall z.B\}$.

(d) An HEP of $\Gamma\{(B \wedge C) \vee (B \wedge C)\}$ has the form

$$\begin{aligned}\Phi :=(&\Gamma\{(B \wedge C) \vee (B \wedge C)\} \mid \Theta,\\ &\hat{\Gamma}\{(B_1 \wedge C_1) \vee (B_2 \wedge \mathrm{C}_2)\} \dots \{(B_{n-1} \wedge \mathrm{C}_{n-1}) \vee (B_n \wedge C_n)\} \mid \hat{\Theta},\, \sigma).\end{aligned}$$

Given such a proof,

$$\begin{aligned}\Psi :=(&\Gamma\{(B \vee B) \wedge (C \vee C)\} \mid \Theta,\\ &\hat{\Gamma}\{(B_1 \vee B_2) \wedge (C_1 \vee C_2)\} \dots \{(B_{n-1} \vee B_n) \wedge (C_{n-1} \vee C_n)\} \mid \hat{\Theta},\, \sigma)\end{aligned}$$

is a preproof of $\Gamma\{(B \vee B) \wedge (C \vee C)\}$ (noting that since

$$(A \wedge B) \vee (C \wedge D) \Rightarrow (A \vee C) \wedge (B \vee D)$$

is a tautology, if the substituted matrix of Φ is valid, then so is the substituted matrix of Ψ). It is correct, since $\lhd_\Phi = \lhd_\Psi$. By the induction hypothesis we obtain an Herbrand expansion proof of $\Gamma\{B \wedge C\}$.

Finally, we show how to simulate the application of a sequent calculus cut:

CUT Given an Herbrand expansion proof $\Phi_1 := (\Gamma_1, A \mid \Theta,\ \hat{\Gamma}, A_1 \mid \hat{\Theta_1}, \sigma_1)$ of Γ, A, and an Herbrand expansion proof $\Phi_2 := (\Gamma_2, \neg A \mid \Theta',\ \hat{\Gamma}, \neg A_2 \mid \hat{\Theta_2}, \sigma_2)$ of $\Delta, \neg A$, the triple

$$\Psi = (\Gamma, \Gamma_2 \mid \Theta \mid \Theta_2 \,|A \wedge \neg A,\ \hat{\Gamma}_1, \hat{\Gamma}_2, \hat{A} \wedge \hat{B} \mid \hat{\Theta}_1 \mid \hat{\Theta_2} | A_1 \wedge \neg A_2, \sigma_3)$$

(where $\sigma_3 = \sigma_1 \cup \sigma_2$) is a preproof of Γ_1, Γ_2. This proof is correct since any cycle in $\ll_\Psi \cup \sqsubset_\Psi$ must be a cycle in either $\ll_{\Phi_1} \cup \sqsubset_{\Phi_1}$ or in $\ll_{\Phi_2} \cup \sqsubset_{\Phi_2}$. But Φ_1 and Φ_2 are both correct, so there can be no such cycle.

Since the cut-rule is admissible, we have a notion of *composition* for Herbrand proofs.

Example 2.2. Recall the proof in Example 2.1. A proof of the opposite entailment is given by the pair

$$(\exists w.(A(w)) \vee \exists v.(B(v)), \forall u.(\neg A(u) \wedge \neg B(u)),\ [w, v := u])$$

By admissibility of cut, we obtain a proof Ψ of, essentially, $\exists x.A \vee \exists x.B) \vdash \exists x.A \vee \exists x.B$. Written in one-sided notation with the eigenvariable condition, the conclusion of Ψ is

$$\Gamma = \exists w.(A(w)) \vee \exists v.(B(v)), \forall y.(\neg A(y)) \wedge \forall z.(\neg B(z))$$

and Ψ has extension and expansion

$$\begin{array}{c} \Gamma \mid (\exists x_L(A(x_L) \vee B(x_L)) \vee \exists x_2(A(x_L) \vee B(x_L))) \wedge \forall u.(\neg A(u) \wedge \neg B(u)) \\ \curlyvee \\ \Gamma \mid (\exists x(A(x) \vee B(x)) \wedge \forall u.(\neg A(u) \wedge \neg B(u)) \end{array}$$

and substitution $[w, v := u, x_L := y, x_L := z]$.

Theorem 2.3. *Herbrand expansion proofs are sound and complete for first-order classical logic.*

Proof. Consider first the class of cut-free HEPs. These are sound, since instantiation, prenexification, and deep $\exists$-contraction are implications in classical logic. They are complete, since any sequent provable in the classical sequent calculus has an HEP.

Completeness for general HEPs follows immediately. For soundness, suppose that an HEP Φ has an unsound conclusion Γ. Then Φ must contain at least one cut. For each cut C_i in Φ, let $C'_i = \exists x_1 \ldots \exists x_n C[\boldsymbol{y} := \boldsymbol{x}]$, where $\boldsymbol{y}$ are the variables free in C but bound elsewhere in the extension of Φ. Similarly, let $\hat{C}'_i = \exists z_1 \ldots \exists z_m \hat{C}[\boldsymbol{w} := \boldsymbol{z}]$, where $\hat{C}_i$ is the expansion of C in Φ, and where $\boldsymbol{w}$ are the variables free in $\hat{C}$ but bound elsewhere in the expansion of Φ. Then

$$(\Gamma, C'_1, \ldots C'_m,\ \hat{\Gamma}, \hat{C}'_1, \ldots \hat{C}'_m,\ [\boldsymbol{z} := \boldsymbol{w}] \circ \sigma)$$

is a cut-free Herbrand proof with an unsound conclusion – contradiction.

Remark 2.1. That this is a structured version of Herbrand's theorem is clear – a first-order sentence A is provable if and only if it has an cut-free Herbrand expansion proof, if and only if there is a valid substitution instance of an expansion of A. The approach extends to cases with equality and non-logical universal axioms. Most proof-theoretic demonstrations of Herbrand's theorem begin with Gentzen's midsequent theorem for formulae in prenex normal form, and so give the theorem for prenex sentences only; here we derive the theorem for all sentences.

3 Cut-reduction for Herbrand expansion proofs

3.1 Cut-correct proofs

Since the cut-free Herbrand proofs are complete, it is clear that cut-elimination holds. The evident much harder problem is to give a constructive proof of this fact.

Problem 3.1. Provide an algorithm which, when given an Herbrand expansion proof with cuts of Γ, constructs a cut-free Herbrand expansion proof of Γ.

This is a more general problem than cut-elimination, since there are HEPs which are not the image of a sequent calculus proof. To see this, consider the correct HEP

$$\Phi = \begin{array}{c} \exists x.A, \forall y.\neg A \mid (\exists z_1.A \vee \exists z_2.A) \wedge \forall w.\neg A \\ \curlyvee \\ \exists x.A, \forall y.\neg A \mid \exists z.A \wedge \forall w.\neg A \end{array}, \ [x, z_1 := w, z_2 := y] \tag{1}$$

It is not possible for the "cut" in Φ to originate from a sequent calculus cut, because of the dependency between the variables on the left and right sides of the cut.

The cut-reduction step QSTRUCT in the next section requires that if x appears in one side of a cut, and y in the other, then $x \not\lhd y$. We will call such proofs *cut-correct*:

Definition 3.1. *Let $B = A \wedge \neg A$ be a cut pair in an Herbrand expansion proof Φ, and $\hat{B}$ its expansion in Φ. The* left branch *of B is the left conjunct of $\hat{B}$. The right branch is the right conjunct of $\hat{B}$. Define S_0^B to be the set of all pairs (x, y) such that x is bound in the left branch of B and y in the right branch of B and S_1^B to be the set of all pairs (x, y) such that x is bound in the right branch of B, and y is bound in the left branch of B.*

Definition 3.2. *(a) An* orientation *for a pre-proof $\Phi = (\Gamma \mid \Theta,\ \hat{\Gamma} \mid \hat{\Theta}, \sigma_\Phi)$ is a function f from the elements of Θ to $\{0, 1\}$.*
(b) A pre-proof Φ is cut-correct if, for each $i \in \Theta$, and each member (x, y) of (S_X^i), we have that $\lhd_\Phi \cup (x, y)$ is irreflexive.
(c) A pre-proof Φ is sequent-correct *if, for each orientation f of Φ, $\lhd_\Phi \cup \bigcup_{i \in \Theta} (S_{f(i)}^i)$ is irreflexive.*

Clearly, if a proof is sequent-correct, it is cut-correct. Intuitively, checking sequent-correctness amounts to checking the existence of $2^{|\Theta|}$ prenexifications of $\hat{\Gamma} \mid \hat{\Theta}$ compatible with σ, one for each orientation f; if $f(B) = 0$, then all quantifiers from the left branch of B occur before all quantifiers from the right branch, and vice versa if $f(B) = 1$. The following is an easy induction on the structure of sequent-calculus proofs:

Lemma 3.1. *The translation of any sequent calculus proof is sequent-correct.*

We are left with the question of the status of HEPs that are not sequent-correct. Consider the following sequent calculus proof:

$$\dfrac{\dfrac{\vdash \neg A(y), A(y)}{\vdash \forall y.\neg A, \exists z_2.A}\ \forall\mathrm{R}, \exists\mathrm{R} \qquad \dfrac{\dfrac{\dfrac{\vdash A(w_2), \neg A(w_2)}{\vdash \exists z_1.A, \forall w_2.A}\ \forall\mathrm{R}, \exists\mathrm{R} \qquad \dfrac{\vdash A(w_1), \neg A(w_1)}{\vdash \exists x.A, \forall w_1.A}\ \forall\mathrm{R}, \exists\mathrm{R}}{\vdash \exists x.A, \forall w_2.\neg A, \exists z_1.A \wedge \forall w_1 \neg A}\ \wedge\mathrm{R}}{}}{\dfrac{\vdash \exists xA, \forall y \neg A, \exists z_1.A \wedge \forall w_1.\neg A, \exists z_2.A \wedge \forall w_2.\neg A}{\vdash \exists xA, \forall y \neg A, \exists z.A \wedge \forall w.\neg A}\ \mathrm{CR}}\ \wedge\mathrm{R}$$

Notice that we cannot replace the conjunctions in this proof with cuts, owing to the contraction. The translation into Herbrand expansion proofs of this proof is

$$\Phi = \begin{array}{c} \exists x.A, \forall y.\neg A, (\exists z_1.A \vee \exists z_2.A) \wedge \forall w.\neg A \\ \curlyvee \\ \exists x.A, \forall y.\neg A \exists z.A, \wedge \forall w.\neg A \end{array}, \ [x, z_1 := w, z_2 := y]$$

that is, the same as (1) but with $\exists z.A \wedge \forall w.\neg A$ as part of the conclusion rather than as a cut. What we see is that non sequent-correct HEPs can arise as a result of "contracting cuts". This kind of behaviour is not allowed in the sequent calculus, but is allowed in, for example, the Calculus of Structures [1]. In particular, [2] gives syntactic cut-elimination for a system of first-order classical logic with such features.

3.2 Cut-reductions steps for cut-correct HEPs

We give, in this section, three reductions, TRIV, QLOG and QSTRUCT, on HEPs. These rules operate where the cut-formula is in prenex normal form; a similar treatment works in the general case, but we omit the details for space reasons.

For each reduction, we identify a cut (the *principal cut*, which for simplicity we assume is the first appearing) in a proof Φ_1, and give a proof Φ_2 in which this cut is either removed, or replaced by a cut(s) (the *resulting cuts*) such that the proof has lower complexity. For all but one case, this reduction of complexity will be the rank:

Definition 3.3 (Rank of an HEP). *The rank of an HEP Φ is the sum of the number of logical symbols appearing in all cuts in Φ.*

<u>TRIV</u> The simplest reduction (TRIV) is the removal of a purely propositional cut (i.e. a cut containing no quantifiers). Let a be quantifier free in :

$$\Phi_1 = (\Gamma \mid a \wedge \neg a \mid \Theta,\ \hat{\Gamma} \mid\ a \wedge \neg a \mid \hat{\Theta},\ \sigma)$$

Such a cut may be deleted: the matrix of the resulting expansion remains valid, and Φ_2 has lower rank than Φ_1.:

$$\Phi_2 = (\Gamma \mid \Theta,\ \hat{\Gamma} \mid \hat{\Theta},\ \sigma)$$

QLOG We will call a cut between $\exists x.A$ and $\forall y.\neg A$ *logical* if the expansion of $\exists x.A$ in Φ has the form $\exists x.\hat{A}$. Given a proof Φ_1 containing such a cut

$$\Phi_1 = (\Gamma \mid \exists x.A \wedge \forall y.\neg A \mid \Theta,\ \hat{\Gamma} \mid \exists x.A_1 \wedge \forall y.\neg A_2 \mid \hat{\Theta},\, \sigma_{\Phi_1})$$

we construct a proof Φ_2

$$\Phi_2 = (\Gamma \mid A \wedge \neg A \mid \Theta[y := \sigma_{\Phi_1}(x)],\ \hat{\Gamma} \mid A_1 \wedge \neg A_2 \mid \hat{\Theta}[y := \sigma_{\Phi_1}(x)],\, \sigma_{\Phi_2})$$

where $\sigma_{\Phi_2} = [y := \sigma_{\Phi_1}(x)] \circ \sigma_{\Phi_1}$. The rank of Φ_2 is smaller than that of Φ_1. Call this reduction QLOG.

QSTRUCT This leaves the case where the existential branch of a cut is of the form $B \vee C$ (where it has a nontrivial expansion at the top level); we will call such cuts *structural cuts*. We now detail the reduction QSTRUCT for such cuts. The redex of this reduction will have the following shape

$$\Phi_1 = (\Gamma \mid \exists y.A \wedge \forall x.\neg A \mid \Theta,\ \hat{\Gamma} \mid (B \vee C) \wedge (\forall x.\neg\hat{A}) \mid \hat{\Theta},\, \sigma_{\Phi_1}) \qquad (2)$$

and we will refer to the variable x as the principal variable of the cut.

The reduction of this cut is similar to the naïve pushing of a cut past a contraction in **LK** [5]; to form the reduct Φ_2 we replace the cut with two cuts $\exists y_L.A_L \wedge \forall x_L.\neg A_L$ and $\exists y_R.A_R \wedge \forall x_R.\neg A_R$, with the expansion of $\exists y_L.A_L$ given by B, and the expansion of $\exists y_R.A_R$ being C. In the sequent calculus this requires the duplication of a whole subproof, and then the contraction of all the conclusions of that subproof. Here we can be more subtle – we only duplicate subformulae in the context which depend hereditarily on the principal variable x. Informally, we replace each maximal subformula in Γ of the form $\exists z.D$, such that $x \lhd z$, by $\exists z_L.D_L \vee \exists z_R.D_R$. The formal definition of Φ_2 is a little more involved. We describe the extension, expansion and substitution in the appendix.

Proposition 3.1. *If Φ is a correct proof, the result Ψ of reducing a cut in Φ is also a correct proof.*

Example 3.1. Recall the HEP Ψ in Example 2.2; this is a proof with a prenex cut, and so falls under our cut-reduction system. Ψ is a redex of QSTRUCT. We will see now how our system eliminates the cut. Define

$$\begin{aligned}\hat{\Gamma} := (\exists w_L.(A(w_R)) \vee \exists w_R.(A(w_R))) \vee \exists v_L.(B(v_L)) \vee \exists v_R.(B(v_R)),\\ \forall y.(\neg A(y)) \wedge \forall z.(\neg A(z))\end{aligned}$$

and

$$B_X := \exists x_X(A(x_L) \vee B(x_X)) \wedge \forall u.(\neg A(u_X) \wedge \neg B(u_X))$$

Reducing the cut in Ψ results in an HEP

$$\Phi = (\Gamma \mid B_X \mid B_R,\ \hat{\Gamma} \mid B_X \mid B_R,\, [w_X, v_X := u_X, x_L := y, x_R := z])$$

Two applications of QLOG and TRIV yields the cut-free HEP

$$\Psi = (\Gamma,\ \hat{\Gamma},\ [v_L, w_L := y, v_R, w_R := z])$$

Note that this proof is clearly not the "identity" proof on $\exists x.A \vee \exists y.A$; this suggests that, in algebraic models built over this system, $\exists x.A \vee \exists x.B$ will not be isomorphic to $\exists x.(A \vee B)$. By duality, $\forall x.(A \wedge B)$ is not isomorphic to $\forall x.A \wedge \forall x.B$, contrary to intuitionistic logic where $\forall x.(A \wedge B) \cong \forall x.A \wedge \forall x.B$.

3.3 The challenge of normalization

Just as in the sequent calculus, there are infinite chains of reductions in the above system; unlike the sequent calculus, there are also non-normalizing HEPs. All the known examples of non-normalizing terms fall outside of the class of sequent-correct proofs, and we conjecture that all sequent-correct proofs are weakly normalizing. What is unfortunate is that sequent-correctness is not preserved by QSTRUCT:

Example 3.2. Consider an sequent-correct HEP of the form

$$\Phi = \begin{array}{c} \hat{\Gamma}_1 \mid \forall z.B \wedge \exists \bar{z} \neg B \mid \forall x \exists y A \wedge (\exists v_1 \forall w_1 \neg A \vee \exists v_2 \forall w_2 \neg A) \\ \curlyvee \\ \Gamma \mid \forall z.B \wedge \exists \bar{z} \neg B \mid \forall x \exists y A \wedge \exists v \forall w \neg A \end{array} \quad [y := z, \bar{z} := x, v_2 := w_1]$$

Reducing the right-hand cut using QSTRUCT results in a proof with expansion

$$\begin{array}{c} \hat{\Gamma}_2 \mid \forall z.B \wedge (\exists \bar{z}_L \neg B \vee \exists \bar{z}_R \neg B) \mid \forall x_L \exists y_L A \wedge \exists v_1 \forall w_1 \neg A \mid \forall x_R \exists y_R A \wedge \exists v_2 \forall w_2 \neg A) \\ \curlyvee \\ \Gamma \mid \forall z.B \wedge \exists \bar{z} \neg B \mid \forall x \exists y A \wedge \exists v \forall w \neg A \mid \forall x \exists y A \wedge \exists v \forall w \neg A \end{array}$$

with substitution $[y_X := z, \bar{z}_X := x_X, v_2 := w_1]$. This HEP is not sequent-correct; there is a path $z, \bar{z}_R, x_R, v_2, w_1, y_L, z$, with evident orientation.

We consider the best course of attack for giving normalization for sequent-correct HEPs to be identifying those applications of QSTRUCT which preserve sequent-correctness (since every known non-normalizing proof is not sequent correct) This is the subject of ongoing work.

4 Further work

In addition to the problem of normalization, we wish to study the properties of these proofs as proof-net-like objects. HEPs exhibit many of the properties associated with proof-nets [6], but they clearly identify too many proofs (in particular, all proofs of a given tautology). Nevertheless, we may consider the question of sequentialization. A very general proof for the cut-free case is given in [13]; another may be easily found by treating HEPs as strategies for proof search in G3 [16]. Both of these give sequentialization for the cut-free, cut-correct HEPs. A good sequentialization for HEPs with cuts (where we use a context splitting cut) seems impossible, since there is no information about how to split the context. For this reason, it seems natural to consider sequentialization and other proof-net like properties in the setting of *augmented* HEPs, where we replace validity of the substituted matrix with a proof or proof-like object. Two obvious cases are the following:

(a) A *Proof-net for first-order classical logic* is a pair Φ, ϕ of an HEP Φ and a proof-net [11] ϕ with conclusion the substituted matrix of Φ.
(b) A *Combinatorial proof for first-order classical logic* is a pair Φ, ϕ of an HEP Φ and a combinatorial proof [8] ϕ with conclusion the substituted matrix of Φ.

In both cases, we replace our reduction TRIV with the relevant propositional cut-elimination. The study of these systems is ongoing work.

Thanks The idea for this work arose from discussions with David Pym, while the author was at the University of Bath. Special thanks go to Willem Heijltjes for spotting an serious error the development of this work. The author is also grateful to Kai Brünnler, Lutz Strassburger, Dominic Hughes and Francois Lamarche for useful discussions.

References

1. Kai Brünnler. *Deep Inference and Symmetry in Classical Proofs*. Logos Verlag, Berlin, 2004.
2. Kai Brünnler. Cut elimination inside a deep inference system for classical predicate logic. *Studia Logica*, 82:51–71, 2006.
3. C. Führmann and D. Pym. Order-enriched categorical models of the classical sequent calculus. *J. Pure Applied Algebra*, 204(1):21–78, January 2006.
4. Carsten Führmann and David Pym. On categorical models of classical logic and the geometry of interaction. *Mathematical. Structures in Comp. Sci.*, 17(5):957–1027, 2007.
5. G. Gentzen. Untersuchungen über das logische Schließen. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1934.
6. J.-Y. Girard. *Proof-nets: the parallel syntax for proof theory*. In Urisini and Agliano, editors, *Logic and Algebra*. Marcel Dekker, New York, 1996.
7. Jacques Herbrand. *Recherches sur la theorie de la demonstration*. PhD thesis, Universite de Paris, 1930.
8. Dominic Hughes. Towards Hilbert's 24th problem: Combinatorial proof invariants. In *Proceedings of WoLLIC*, 2006.
9. Dominic J.D. Hughes. Proofs Without Syntax. *Annals of Mathematics*, 143(3):1065–1076, November 2006.
10. F. Lamarche. Exploring the gap between linear and classical logic. *Theory and Application of Categories*, 18:473–535, 2007.
11. François Lamarche and Lutz Straßburger. Naming proofs in classical propositional logic. In Paweł Urzyczyn, editor, *Typed Lambda Calculi and Applications, TLCA 2005*, volume 3461 of *Lecture Notes in Computer Science*, pages 246–261. Springer-Verlag, 2005.
12. R.I. McKinley. *Categorical models of First-Order classical proofs*. PhD thesis, University of Bath, 2006.
13. Dale Miller. A compact representation of proofs. *Studia Logica*, 46(4):347–370, 1987.
14. Lutz Strassburger. On the axiomatization of Boolean categories with and without medial. *Theory and Application of Categories*, 18:536–601, 2007. To appear, Theory and Application of Catgories.
15. Lutz Strassburger. Some observations on the proof theory of second order propositional multiplicative linear logic. Submitted, 2008.
16. A.S. Troelstra and H. Schwichtenberg. *Basic Proof Theory*. Cambridge University Press, Cambridge, 1996.

A The reduction QSTRUCT: formal definition

Given a cut-correct HEP

$$\Phi_1 = (\Gamma \mid \exists y.A \wedge \forall x.\neg A \mid \Theta,\ \hat{\Gamma} \mid (B \vee C) \wedge (\forall x.\neg \hat{A}) \mid \hat{\Theta},\ \sigma_{\Phi_1}) \tag{3}$$

we define its reduct $\Phi_2 = (K, \hat{K}, \sigma_2)$ under QSTRUCT.

Definition A.1. *If A is a formula, then A_L is the result of alpha renaming each bound variable x as x_L, and similarly for A_R.*

Suppose that $x \lhd z$, and that there is a cut B appearing in Φ_1 such that the variable z is bound in Φ_1 but free in the expansion of B. This cut must be duplicated in Φ_2. Let Θ_2 be a sequence of cut-pairs derived from Θ by replacing each cut pair B such that $z \in \text{fv}(\hat{B})$ and $x \lhd z$ by $B_L \mid B_R$. Then the extension of Φ_2 is

$$K = \Gamma \mid \exists y_L.A_L \wedge \forall x_L.\neg A_L \mid \exists y_R.A_R \wedge \forall x_R.\neg A_R \mid \Theta_2.$$

In the expansion, whether or not a subformula is duplicated depends on whether it depends hereditarily on x. We define two functions $\mathbf{T}_L, \mathbf{T}_R$ from variables bound in Φ_1 to variables bound in Φ_2, which are the identity on variables not copied by QSTRUCT, and which otherwise give us one of the two copies.

$$\mathbf{T}_X(x) = x \tag{4}$$

$$\mathbf{T}_X(z) = \begin{cases} z & x \not\lhd z, z \neq x \\ z_X & x \lhd z \end{cases} \tag{5}$$

Extend that function to all terms as follows

$$\mathbf{T}_X(a) = a \quad a \text{ a constant} \tag{6}$$

$$\mathbf{T}_X(f(t_1, \ldots, t_n)) = f(\mathbf{T}_X(t_1), \ldots, \mathbf{T}_X(t_n)) \tag{7}$$

We now extend these functions to formulae; on quantified formulae we rename those bound variables depending on x:

$$\mathbf{T}_X(qz.B) = \begin{cases} qz_X.\mathbf{T}_X(B) & x \lhd z \\ qz.\mathbf{T}_X(B) & x \not\lhd z \end{cases}$$

with the value of $\mathbf{T}_X$ on other formulae given by

$$\begin{aligned} \mathbf{T}_X(R(t_1 \ldots t_n)) &= R(\mathbf{T}_X(t_1), \ldots \mathbf{T}_X(t_n)) \quad R \text{ a relation symbol} \\ \mathbf{T}_X(\neg B) &= \neg \mathbf{T}_X(B) \\ \mathbf{T}_X(B * C) &= \mathbf{T}_X(B) * \mathbf{T}_X(C) \qquad * = \wedge, \vee \end{aligned}$$

We define now a function from $\mathbf{HD}$ (hereditary duplication) from formulae to sequences of formulae. This function will help define the expansion of Φ_2. We ensure that each duplicated cut has an expansion by setting

$$\mathbf{HD}(A) = \mathbf{T}_L(A) \mid \mathbf{T}_R(A) \quad z \in \text{fv}(A), x \lhd z.$$

For all other formulae A, $\mathbf{HD}(A)$ is a formula. On an existentially bound formulae A, $\mathbf{HD}(A)$ depends on the eigenvariable of the quantifier:

$$\mathbf{HD}(\exists z.B) = \begin{cases} \exists z.\mathbf{HD}(B) & x \not\lhd z \\ \exists z_L.\mathbf{T}_L(B) \vee \exists z_R.\mathbf{T}_R(B) & x \lhd z \end{cases}$$

Otherwise

$$\begin{aligned} \mathbf{HD}(R(t_1 \dots t_n)) &= R(t_1 \dots t_n) \\ \mathbf{HD}(\neg B) &= \neg \mathbf{HD}(B) \\ \mathbf{HD}(B * C) &= \mathbf{HD}(B) * \mathbf{HD}(C) \quad * = \wedge, \vee \\ \mathbf{HD}(\forall z.B) &= \forall z.\mathbf{HD}(B) \end{aligned}$$

Extend $\mathbf{HD}$ to a function from sequences to sequences in the obvious way. The expansion of Φ_2 is

$$\hat{K} = \mathbf{HD}(\hat{\Gamma}) \mid B \wedge (\forall x_L.\neg \mathbf{T}_L(\hat{A})) \mid C \wedge (\forall x_R.\neg \mathbf{T}_R(\hat{A})) \mid \mathbf{HD}(\hat{\Theta}).$$

Finally, let σ_2 be defined as follows:

$$\sigma_2(z) = \sigma(z) \text{ if } x \not\lhd_\Phi z \tag{8}$$

$$\sigma_2(z_L) = \mathbf{T}_L(\sigma(z)) \tag{9}$$

$$\sigma_2(z_R) = \mathbf{T}_R(\sigma(z)). \tag{10}$$

We have, in conclusion

$$\Phi_2 = (K,\ \hat{K},\ \sigma_2)$$