On extracting variable Herbrand disjunctions

Andrei Sipoș

Research Center for Logic, Optimization and Security, University of Bucharest
Institute of Mathematics of the Romanian Academy

August 29, 2022
International Conference on Applied Proof Theory 2022 (APT22),
honouring Ulrich Kohlenbach's 60th birthday
D'Annunzio University of Chieti–Pescara
Pescara, provincia di Pescara, Abruzzo, Italy

Herbrand's theorem

The classical Herbrand theorem states, more or less, that for every first-order formula φ over a signature containing at least one constant symbol, such that φ has at most one free variable denoted by x and such that in a purely universal theory of classical first-order logic it can be proven that $\exists x \varphi$, there is a finite number of closed terms $t_1, \ldots, t_n$ over that signature – extractable in some way from the proof – such that it is also derivable from that theory that

$$\bigvee_{i=1}^{n} \left(\varphi[x := t_i]\right),$$

this kind of expression being usually called a *Herbrand disjunction*.

An example

Example (Ulrich Berger)

Consider the signature σ having a constant 0 and two unary function symbols S, f. Then in the theory consisting of just the axiom $\forall x \neg(Sx = 0)$, one can prove that $\exists x \neg(fSfx = x)$.

Proof (outside the system)

Let $\mathcal{M}$ be a σ-structure with universe M, such that $\mathcal{M} \models \forall x \neg(Sx = 0)$ and $\mathcal{M} \models \forall x(fSfx = x)$. Then $f^{\mathcal{M}}$ is injective, but also (since, for all $x \in M$, $S^{\mathcal{M}}(x) \neq 0^{\mathcal{M}}$) it is surjective when restricted to $\{x \in M \mid x \neq 0^{\mathcal{M}}\}$ and thus non-injective. Contradiction!

From (a careful examination of) the above proof, one can 'extract' the terms $t_1 := 0$, $t_2 := f0$, $t_3 := Sff0$ (exercise!).

Extensions

Such 'term extraction' results have later been obtained for systems which include non-logical axioms: systems which may serve as a foundation of mathematics, like first-order arithmetic.

Among the first were Georg Kreisel's no-counterexample interpretation and the various flavours of Gödel's functional ('*Dialectica*') interpretation, the latter of which playing nowadays a central role in the research program of *proof mining* (which everyone here is familiar with at least by name).

One striking feature of this sort of theorems is that they generally do not extract terms expressible in the original system under discussion, but usually go beyond it – e.g. Gödel's functional interpretation, which worked originally for systems of the strength of first-order arithmetic, interprets such systems into some which involve higher-type functionals.

The recursor

One classical feature of this kind of system is the *recursor* term – for each 'type' ρ, one has a term R_ρ of type

$$\rho \to ((\mathbb{N} \to (\rho \to \rho)) \to (\mathbb{N} \to \mathbb{N})$$

such that (informally speaking) for each a of type ρ, each b of type $\mathbb{N} \to (\rho \to \rho)$ and each $n \in \mathbb{N}$,

$$R_\rho ab0 = a, \quad R_\rho ab(n+1) = bn(R_\rho abn),$$

i.e. R_ρ expresses some sort of higher-type primitive recursion.

This rule may be formally coded in two ways (keep this in mind for any further kinds of terms we may introduce):

- by giving a set-theoretic denotation in terms of actual higher-type functionals;
- by specifying a term rewriting system, yielding a 'term-model' semantics.

Proofs of Herbrand's theorem

Back to first-order logic, along the years there were many and quite varied proofs of Herbrand's theorem:

"Herbrand's original proof was syntactic and provides an algorithm for the extraction of the Herbrand terms from a given [first-order logic] proof [...]. However, two lemmas in his proof need a correction as was discovered first by K. Gödel in the 40s [...] and in the 60s by B. Dreben et al. After Herbrand's work, alternative syntactic proofs were given by D. Hilbert and P. Bernays (using Hilbert's ε-substitution method) and by G. Gentzen (using his cut elimination procedure for a sequent calculus formulation of [first-order logic]). Most textbook treatments of Herbrand's theorem nowadays are model-theoretic and do not yield any term extraction algorithm (with Shoenfield as a notable exception)."
(Kohlenbach, 2008)

Kreisel's suggestion

Shoenfield's 1967 book mentioned above gives a complicated syntactic proof of Herbrand's theorem, and Kreisel in his review of the book suggested that one might use the functional interpretation instead:

"To give some coherence to the subject [a lecturer] should explain how (i) the cut elimination methods in predicate calculus generalize to number theory or how (ii) Gödel's functional interpretation for number theory specializes to predicate logic (the author uses (i) for predicate logic and (ii) for number theory)."

This suggestion was followed up by Gerhardy and Kohlenbach in this paper:

Philipp Gerhardy, Ulrich Kohlenbach, Extracting Herbrand disjunctions by functional interpretation. *Arch. Math. Logic* 44, no. 5, 633–644, 2005.

Gerhardy and Kohlenbach's system

Gerhardy and Kohlenbach stripped down Gödel's higher-typed system – formulated as a λ-calculus and not using combinators as in Kohlenbach's book *Applied Proof Theory* – by eliminating the recursors (since those aren't needed for pure logic).

One is then left with a simply-typed λ-calculus, having, alongside variables, λ-expressions and applications, also the following constants:

- for each function symbol of arity n of the given first-order signature, one has a constant of type $0^n \to 0$ (0 denotes the base type, corresponding to elements of a structure);
- one also has some special 'case distinction' constants.

Given that this is an exact instantiation of the simply-typed λ-calculus, we can use classical results like normalization.

Case distinction constants

For each quantifier-free formula φ with m free variables $x_1, \ldots, x_m$, in order of their appearance in φ, we have a constant term c_φ of type $0^{m+2} \to 0$, such that (again informally) for every $a_1, \ldots, a_m$, b_1, b_2, we have that

$$c_\varphi(a_1)\ldots(a_m)(b_1)(b_2) = \begin{cases} b_1, & \text{if } \varphi[x_1 := a_1, \ldots, x_m := a_m], \\ b_2, & \text{otherwise.} \end{cases}$$

A 'subtle' point is that we only have these constants for formulas φ in the original language, and not for those involving higher-order terms (since that would make a mess for more than one reason), but we shall see that this is enough.

Sketch of the Gerhardy/Kohlenbach proof

In their proof of Herbrand's theorem, they first apply Gödel's interpretation to $\exists x \varphi$ to obtain a higher-typed term t such that $\varphi[x := t]$ in their custom higher-typed first-order logic.

To interpret the 'contraction' rule $A \vee A \vdash A$ one might suspect one needs a case distinction term corresponding to a higher-typed formula ψ but the formula that shows up can always be written as a substitution instance of an ordinary first-order formula χ (its 'skeleton'), and the terms which go into the substitution can be written somewhere in the arguments of c_χ.

One then **normalizes** t to obtain a term s which can be shown to have a sufficiently nice form that one can read the Herbrand terms directly off it.

Other treatments

In recent years, there have been other higher-order treatments of Herbrand's theorem, including

Fernando Ferreira, Gilda Ferreira, A herbrandized functional interpretation of classical first-order logic. *Arch. Math. Logic* 56, no. 5-6, 523–539, 2017,

which uses a 'star-combinatory' calculus of their own devising, and

Bahareh Afshari, Stefan Hetzl, Graham E. Leigh, Herbrand's theorem as higher order recursion. *Ann. Pure Appl. Logic* 171, no. 6, 102792 [45 pp.], 2020,

which uses higher-order recursion schemes.

My visit: going higher

On the last day of my visit to Ulrich Kohlenbach in July 2021, I expressed my disappointment that all these results concerned pure logic and not arithmetical theories.

Kohlenbach then suggested that I should look into Tait's infinitary calculus.

Tait's calculus

The Tait calculus, in its simplest form, features, instead of recursors, an infinitary term constructor $(t_n)_{n\in\mathbb{N}}$ which seeks to be the computational counterpart of the Schütte-style ω-rule – thus, we call the resulting terms ω-terms. Such a term is of type $\mathbb{N} \to \rho$ (we will write the type $\mathbb{N}$ simply as 0 from now on) and its intended interpretation is the mapping $n \mapsto t_n$.

One can then replace a recursor R_ρ by a term like $\lambda a.\lambda b.(t_n)_{n\in\mathbb{N}}$, where $t_0 = a$, and, for each $n \in \mathbb{N}$, $t_{n+1} = bnt_n$.

We notice that this kind of term does not contain 'atomic' subterms of the type level of the recursor. Thus, there is a **trade-off** between the type level and the passage to the infinite.

I realized after reading this that a careful use of this calculus would explain an old phenomenon of proof mining.

Metastability

Terence Tao, during his work on multiple ergodic averages, rediscovered in 2007 the Herbrand normal form of the Cauchy property for bounded monotone sequences, a property which was then dubbed 'metastability' at the suggestion of Jennifer Chayes.

The property is expressed as follows (denoting, for all $f : \mathbb{N} \to \mathbb{N}$ and for all $n \in \mathbb{N}$, by $\widetilde{f}(n)$ the quantity $n + f(n)$ and by $f^{(n)}$ the n-fold composition of f with itself): if we take $k \in \mathbb{N}$, $g : \mathbb{N} \to \mathbb{N}$, and (a_n) a, say, nonincreasing sequence in the interval $[0, 1]$, then there is an N such that for all $i, j \in [N, \widetilde{g}(N)]$,

$$|a_i - a_j| \leq \frac{1}{k+1},$$

and, moreover, N can be taken to be an element of the finite sequence $0, \widetilde{g}(0), \ldots, \widetilde{g}^{(k)}(0)$.

A variable Herbrand disjunction

The conclusion of the statement before can be written as

$$\bigvee_{i=0}^{k} \left(\left(\forall i, j \in [x, \widetilde{g}(x)] \, |a_i - a_j| \leq \frac{1}{k+1} \right) \left[x := \widetilde{g}^{(i)}(0) \right] \right).$$

It was immediately noticed that the above can be said to be a Herbrand disjunction "of variable length" (Kohlenbach, 2008) – more precisely, it is variable in k, but does not depend on the g, which only shows up in the terms themselves (a fact to which we shall later return).

This was all known to Kreisel in the 1950s!

THE JOURNAL OF SYMBOLIC LOGIC
Volume 17, Number 1, March 1952

ON THE INTERPRETATION OF NON-FINITIST PROOFS

PART II. INTERPRETATION OF NUMBER THEORY. APPLICATIONS.

G. KREISEL

Example. Consider the theorem that a bounded, monotone increasing sequence of reals a_n converges. To simplify notation let them lie between 0 and 1, and let us use binary scale, in which rationals $n/2^m$ terminate.

$$\frac{a(n, m)}{2^m} \leq a_n < \frac{a(n, m) + 1}{2^m}.$$

Then convergence means:

$$(m)(En_0)(n)[n > n_0 \rightarrow a(n, m) = a(n_0, m)] \cdot$$
$$\vee\ (Er)(Es)[a(r + 1, s) < a(r, s) \vee a(r, 0) > 1 \vee a(r, 0) < 0].$$

A counter-example would be a number $\mathfrak{m}$, and a function $N(n_0)$ so that for all n_0

$$N(n_0) > n_0 \quad \text{and} \quad a[N(n_0), \mathfrak{m}] \neq a(n_0, \mathfrak{m}), \text{ also}$$
$$a(r, \mathfrak{m}) \leq a(r + 1, \mathfrak{m}) \quad \text{and} \quad 0 \leq a(r, 0) \leq 1. \qquad 39.3$$

This is impossible: take $n_0 = 0$, $n_1\ (= N(n_0)), \cdots, n_{i+1}(= N(n_i))$; then if for all r, $a(r, \mathfrak{m}) \leq a(r + 1, \mathfrak{m})$,

$$a(0, \mathfrak{m}) + 1 \leq a(n_1, \mathfrak{m})$$
$$a(n_1, \mathfrak{m}) + 1 \leq a(n_2, \mathfrak{m})$$
$$\vdots$$
$$a(n_{2^m}, \mathfrak{m}) + 1 \leq a(n_{2^m+1}\ \mathfrak{m})$$

so that $a(n_{2^m+1}, \mathfrak{m}) > 2^m + a(0, \mathfrak{m})$, hence $a(n_{2^m+1}, 0) > 1$. In our notation, for some n_0, $0 \leq n_0 \leq \omega_x[N(x), 0, 2^m + 1]$, 39.3 breaks down.

The extension

Let us now see what the logical framework will be in our effort to extend Herbrand's theorem to systems of the strength of first-order arithmetic PA.

We shall denote by σ_{ar} the signature of arithmetic, containing the symbols 0, S, $<$ having their usual natures and arities, together with symbols for all primitive recursive functions (in particular, the $+$ and $\cdot$ symbols; we shall need, though, to keep $<$ as a **relation** symbol), and by $\mathcal{N}$ the standard σ_{ar}-structure with universe $\mathbb{N}$ and with the symbols having their natural interpretation. We shall fix a signature σ which contains σ_{ar}.

Why extend arithmetic

The point of considering a signature greater than σ_{ar} is to model stuff which may not be purely arithmetical, a sort of miniaturized version of Kohlenbach's metatheorem framework where one works both with foundational and structural (e.g. Banach space) axioms within the same system. We shall see at the end of the talk a concrete example of how this gets to be applied.

Our logical theory will essentially be PA, but with the induction axiom extended to all σ-formulas, and hence we denote the resulting theory by PA^{σ}.

We will also add a set Γ of purely universal σ-sentences.

Let's start

Let A be a quantifier-free σ-formula such that it has at most one free variable denoted by x. Assume that

$$\mathsf{PA}^{\sigma} + \Gamma \vdash \exists x A.$$

Then, by Gödel's functional interpretation but with Tait's calculus as its destination, there is a closed ω-term of type 0, extractable from the proof, such that for every σ-structure $\mathcal{M}$ such that its σ_{ar}-reduct is $\mathcal{N}$ and $\mathcal{M} \models \Gamma$, we have that

$$\mathcal{M} \models A[x := t].$$

More on case distinctions

Since for σ_{ar}-formulas one can use the tried and true method of interpreting contraction through actual computations instead of case distinctions (again, see Kohlenbach's book *Applied Proof Theory*), we reserve the latter option for formulas which contain something not in σ_{ar}.

From our paper:
"We shall prefer using [purely arithmetical case distinctions] instead of the case distinction constants which we have introduced before for formulas involving symbols not in σ_{ar} essentially by replacing those terms in the formula at hand which involve those kind of symbols with variables, with the only non-arithmetic symbols which we will not be able to get rid of being the relation symbols."

Term rewriting

If we again follow Tait and consider the term rewriting system on the ω-terms generated by the reduction relations

$$(\lambda x.t)(s) \rightsquigarrow t[x := s], \quad ((t_n)_{n\in\mathbb{N}} r)\, s \rightsquigarrow ((t_n s)_{n\in\mathbb{N}})\, r,$$

$$(t_n)_{n\in\mathbb{N}} \underline{m} \rightsquigarrow t_m,$$

as extended by compatibility with the term constructors, we have that each term t has a normal form s having the same free variables, such that for every σ-structure $\mathcal{M}$ whose σ_{ar}-reduct is $\mathcal{N}$, we have that $\mathcal{M} \models t = s$.

It therefore makes sense to study how normal forms in this system look like. This will be the meat of our proof.

The form of an ω-term

Remember that an ω-term is either:

- a variable;
- a constant of type $0^m \to 0$;
- a λ-expression or an application;
- an infinitary term $(t_n)_{n\in\mathbb{N}}$ of type $0 \to \rho$ – if $\rho = 0$, we shall call it *zero*, otherwise we shall call it *non-zero*.

We get the following lemma, which can be thought of as our main technical result.

Lemma

Let s be a closed ω-term of type 0 in normal form and let u be a subterm of s. Then there is a $k \in \mathbb{N}$ and terms $s_1, \ldots, s_k$ of type 0 such that there is a term f which is either a constant or a zero $(t_n)_{n\in\mathbb{N}}$ with $u = fs_1 \ldots s_k$ (and thus $s_1, \ldots, s_k$ are also in normal form).

Proof: one page of combinatorial arguments!

The set of terms

We now fix $\mathcal{M}$ be a σ-structure whose σ_{ar}-reduct is $\mathcal{N}$ and $\mathcal{M} \models \Gamma$.

Based on the characterization in the previous lemma, we now define, for each closed ω-term s of type 0 in normal form, a finite set $T_s^{\mathcal{M}}$ of closed σ-terms, recursively in the length of s:

- if s is of the form $fs_1 \dots s_n$, we put

$$T_s^{\mathcal{M}} := \{f(t_1, \dots, t_n) \mid \text{for all } i,\ t_i \in T_{s_i}^{\mathcal{M}}\};$$

- if s is of the form $c_\varphi s_1 \dots s_m s_{m+1} s_{m+2}$, we put

$$T_s^{\mathcal{M}} := T_{s_{m+1}}^{\mathcal{M}} \cup T_{s_{m+2}}^{\mathcal{M}};$$

- if for each $n \in \mathbb{N}$, t_n is a term of type 0 and s is of the form $((t_n)_{n\in\mathbb{N}})\, s_1$, we put

$$T_s^{\mathcal{M}} := \bigcup_{r \in T_{s_1}^{\mathcal{M}}} T_{t_{r^{\mathcal{M}}}}^{\mathcal{M}}.$$

The end of the proof

Let s be the closed ω-term which is the normal form of t (the term originally extracted from the proof of $\exists x A$). Since $\mathcal{M} \models t = s$, we have that

$$\mathcal{M} \models A[x := s].$$

Taking now $S := T_s^{\mathcal{M}}$, a finite set of closed σ-terms, we have that

$$\mathcal{M} \models \bigvee_{r \in S} s = r,$$

so

$$\mathcal{M} \models \bigvee_{r \in S} A[x := r],$$

which we now tell you it is the conclusion of our theorem.

We note that, like in the classical Herbrand theorem, "the Herbrand terms do not depend on the predicate symbols" (Kohlenbach, 2008).

Issues

Of course, this final statement is completely trivial as it stands, since to obtain a finite set of closed σ-terms S such that

$$\mathcal{M} \models \bigvee_{r \in S} A[x := r],$$

we can just take an appropriate $n \in \mathbb{N}$ such that $\mathcal{M} \models A[x := \underline{n}]$ and take $S := \{\underline{n}\}$.

The real content of our result lies in the way S is constructed, since one does not necessarily "call" the whole of $\mathcal{M}$ in the course of the recursive definition, so one may hope to recover some amount of uniformity in $\mathcal{M}$, i.e. one might not care about the interpretations of some of the symbols **not** in σ_{ar}.

We will now illustrate this with the metastability example from before.

An example

Let, therefore, $k \in \mathbb{N}$, $g : \mathbb{N} \to \mathbb{N}$ and $(a_n)_{n\in\mathbb{N}} \subseteq [0, 1]$ be nonincreasing. We want to show that there is an $N \in \mathbb{N}$ such that for all $i, j \in [N, \widetilde{g}(N)]$,

$$|a_i - a_j| \leq \frac{1}{k+1}.$$

It is enough to show that there is an N such that

$$a_N - a_{\widetilde{g}(N)} \leq \frac{1}{k+1},$$

since, then, taking the same N and $i, j \in [N, \widetilde{g}(N)]$ and assuming w.l.o.g. that $i \leq j$, we have that

$$|a_i - a_j| = a_i - a_j \leq a_N - a_{\widetilde{g}(N)} \leq \frac{1}{k+1}.$$

An example

Assume towards a contradiction that, for all $N \in \mathbb{N}$,

$$a_N - a_{\widetilde{g}(N)} > \frac{1}{k+1}.$$

We now show that for all $x \in \mathbb{N}$ there is an $y \in \mathbb{N}$ such that

$$a_0 - a_y > \frac{x+1}{k+1}.$$

We prove this by induction on x. For $x = 0$, we have that

$$a_0 - a_{\widetilde{g}(0)} > \frac{1}{k+1},$$

so we can take $y := \widetilde{g}(0)$. Let now $x \in \mathbb{N}$ and assume that there is a y such that

$$a_0 - a_y > \frac{x+1}{k+1}.$$

We want to show that there is a w such that

$$a_0 - a_w > \frac{x+2}{k+1}.$$

An example

Since

$$a_y - a_{\widetilde{g}(y)} > \frac{1}{k+1},$$

we have that

$$a_0 - a_{\widetilde{g}(y)} = a_0 - a_y + a_y - a_{\widetilde{g}(y)} > \frac{x+1}{k+1} + \frac{1}{k+1} = \frac{x+2}{k+1},$$

so we can take $w := \widetilde{g}(y)$ and the induction is finished.

Now, if we take $x := k$, we have that there is an $y \in \mathbb{N}$ such that

$$a_0 - a_y > \frac{k+1}{k+1} = 1,$$

but $a_0 - a_y \leq a_0 \leq 1$, which yields a contradiction.

Formalization

To formalize the previous proof in our framework, we construct the first-order signature σ by adjoining to σ_{ar} the function symbols k and g of arity (obviously) 0 and 1, respectively, together with a relation symbol P of arity 3 such that $P(v, w, l)$ signifies

$$a_v - a_w \leq \frac{l}{k+1}.$$

Then, after examining the proof, one can see that if we take

$$\Gamma := \{\forall v \forall p \forall r \forall b((\neg P(v, p, b) \wedge \neg P(p, r, S0)) \rightarrow \neg P(v, r, Sb)),$$

$$\forall t P(0, t, Sk)\},$$

which is a set of universal σ-sentences, then the whole of the argument is formalizable in $\text{PA}^\sigma + \Gamma$.

The final result

After applying our extraction algorithm, we obtain the set

$$\left\{ g^{(i)}0 \mid 0 \leq i \leq \max\left(k^{\mathcal{M}}, 1\right) \right\},$$

which yields a Herbrand disjunction of the form exhibited for the metastability statement, and we can clearly see that the computed set depends on $k^{\mathcal{M}}$ but not on $g^{\mathcal{M}}$ or $P^{\mathcal{M}}$, and thus we obtain uniformity in each class of σ-structures whose σ_{ar}-reduct is $\mathcal{N}$, which satisfy Γ and which interpret k the same way.

These results may all be found in:

A. Sipoș, On extracting variable Herbrand disjunctions.
Studia Logica, Volume 110, Issue 4, 1115–1134, 2022.

Future work

We now suggest several avenues for future work in this vein:

- one could find more concrete proof mining examples which fit this schema (we suggest looking at statements involving non-constructive convergence towards 0);
- one could recast this work into some other framework (to obtain more insights): either use the star-combinatory calculus of Ferreira/Ferreira or (as suggested by Paulo Oliva) use a monadic translation;
- one could extend it to computational calculi inspired by stronger logical systems.

Thank you for your attention.