Logique & Analyse **242** (2018), 183-199

NOVIKOV'S CUT ELIMINATION

Luca Bellotti

Abstract

This is an exposition of Novikov's cut-elimination procedure for a Hilbert-style formulation of the first-order predicate calculus, which depends on a property of formulas introduced by him, called 'regularity'. A comparison with other methods is outlined.

Keywords: Cut elimination, regular formula, Novikov

1. Introduction

Our purpose in this paper is to explain the interesting, original reduction procedure invented by Pëtr Sergeevich Novikov[1] by means of the notion of regularity (*regulyarnost'*) of formulas, to obtain what amounts to (a sort of) cut elimination for a Hilbert-style formulation of the first-order predicate calculus.

The notion of regularity (a form of cut-free derivability) was first introduced by Novikov in (1939), and then developed in (1943), with reference to infinitary derivations in a system for the propositional calculus admitting countable conjunctions and disjunctions.[2] Novikov's work was reviewed by Church in (1946), a bit tepidly and apparently without appreciating the novelty of the approach, and, much later, well described (and compared with similar results in the West, on which it does not depend) by Jon Barwise (1981, esp. Appendix), Grigori Mints (in his survey of Soviet proof theory, 1991, 387 ff.) and Sergei Tupailo (1992).[3] Thierry Coquand (1995) deeply

[1] We only recall here that P. S. Novikov (1901–1975) was one of the foremost Soviet mathematicians of the last century, especially renowned for his contributions to classical descriptive set theory and his work on the borderline of logic and group theory (work of epic difficulty, done in part in collaboration with his pupil Adian).

[2] One is strongly reminded of Tait's work (1968), carried out independently (and in greater generality) almost thirty years later. A comparison could be useful, as an anonymous referee rightly remarked, but will not be given here, since I chose to focus on Novikov's specific work on finitary systems, neglected in the literature.

[3] Tupailo considered an infinitary system also for predicate logic (embedding finitary predicate logic into it) and showed that cut-free derivations of the same form can be obtained

doi: 10.2143/LEA.242.0.3284750.

explained further, and reformulated in terms of games, the reduction technique developed by Novikov (also illuminating the nontrivial matter of the interaction between the classical and the intuitionist point of view in this setting).

In (1949) Novikov first defined regularity for the usual *finitary* first-order predicate logic, and stated a theorem to the effect that every formula which is provable[4] there is regular. But only in the book *Elements of mathematical logic* (1959), the first original logic textbook in Russian (a beautiful, very clear introduction, sometimes a bit unfamiliar for the western reader), Novikov finally published a *proof* of the result that regularity is preserved by the rules of a (usual) formal system for predicate logic, in particular by Modus Ponens. In fact, the result is thoroughly proved there (1959, Ch. 6) for a stronger system (viz. a form of 'restricted' arithmetic, based on full first-order logic with identity, with successor and order axioms, definitions for all primitive recursive functions, without any form of induction), but the point of the proof-theoretic reduction concerns the purely logical part of the system.[5]

To our knowledge, *this* proof has received scarce attention in the literature, and a full exposition of it is lacking. In view of the fact that Novikov's method is very different from the usual ones, this could be useful. Moreover, a comparison of this original technique with other forms of cut elimination for *finitary* systems (mainly Gentzen's one, but also the methods employed in Herbrand's Thesis) could be interesting.

As we shall see, the proof is nontrivial and requires a long series of lemmas, whose proofs are sometimes a bit involved; its (perhaps) most important peculiarity is that no induction *stricto sensu* on the complexity of (what corresponds to) the cut formula is needed: the formula is modified but (in a sense to be seen below) preserved.

After the description and explanation of Novikov's procedure, a brief comparison with other classical cut-elimination procedures for predicate logic (inspired by Gentzen 1934) is given. Of course, a comparison with

by Gentzen-type reductions and Novikov's transformations. Barwise, on the other hand, employed mainly semantic arguments.

[4] But Novikov always used the Russian word for 'true' (*istinnaya*) for this notion (except in the second edition, 1973, of Novikov 1959), just like Herbrand, who always used the French *vraie* in his works.

[5] Thus we shall restrict to this part, and we shall ignore questions of consistency, with some ensuing simplifications, e.g. avoiding the notion of weak regularity (see ibid.). As results from Mints's survey, Novikov's finitary proof systems based on regularity did not gain much favor, perhaps in view of the greater efficiency of other systems, such as those based on Maslov's 'inverse method' (an independently invented form of resolution, see Mints 1991). Apparently, the only further development (after 1959) of Novikov's finitary methods for applications in the proof theory of arithmetic can be found in the work of a pupil of his, Tsinman (1968).

the relevant methods and results contained in Herbrand's Thesis (see Herbrand 1968, 35-154), or deriving from it, would be even more interesting (since the formulation of predicate logic and the techniques adopted there are much closer[6] to Novikov's ones), but in view of the notorious difficulties of Herbrand's original work (even after all the corrections), and of my limited competence on the topic, this will be left as a task for future research (except for a few remarks below: see Section 3). On the other hand, another possible direction for further work would be a comparison with certain more recent developments in proof theory.[7]

2. The procedure

2.1. *Basic definitions*

Our system is a Hilbert-style axiomatic formulation of the first-order predicate calculus without identity (though it is not difficult to extend the results below to the case with identity and even to the restricted arithmetic mentioned above), without constants or function symbols, with countably many predicate letters for each arity, and both quantifiers treated as primitive. The axiom schemas[8] are those of any usual complete system for predicate logic, e.g. we take a complete system for propositional logic and, as axioms for quantifiers, $(\forall x)A(x) \to A(y)$ and the dual axiom for the existential quantifier. The quantifier rules are: from $A \to B(x)$ infer $A \to (\forall x)B(x)$ (x not free in A), and the dual rule for the existential quantifier. The only propositional rule is Modus Ponens. As usual, '$A \wedge B \vee C$' abbreviates '$(A \wedge B) \vee C$' (conjunction binds stronger than disjunction).

Following Novikov, we shall use the term *product* for conjunction (*factor* for conjunct) and *sum* for disjunction (*summand* for disjunct). If a factor is not a product, we say that it is a *prime* factor; similarly, if a summand is

[6] E.g., there is an analogy, the exact extent of which is unclear, between Novikov's notion of regularity and Herbrand's notion of Property A of provable propositions, defined in Ch. 5, Sec. 2 of his thesis (see Herbrand 1968, 118 ff. for the definition, which is too involved to be given here). One cannot exclude among Novikov's sources also Bernays' reformulation of Herbrand's approach in *Grundlagen* II §3.3 (Hilbert-Bernays 1939, 149 ff.), although this is not mentioned in 1959, despite the overtly Hilbertian overall setup. See however Section 3 below.

[7] E.g., expansion trees (Miller 1987), or deep inference, in particular the calculus of structures (see, e.g., this page, maintained by A. Guglielmi: http://alessio.guglielmi.name/res/cos/, especially references to the contributions of Brünnler and McKinley, among others). These developments do not descend from Novikov's approach, but could be related to it (again, a specific competence would be needed for the comparison). I owe these suggestions to an anonymous referee.

[8] Novikov has axioms and substitution rules, but this is immaterial.

not a sum, it is a prime summand. By associativity, any product is equivalent to a product whose factors are all prime; similarly for sums. A formula is *reduced* if it contains as connectives only conjunction, disjunction and negation, the latter applied only to atomic formulas (i.e., it is in negation normal form). Any formula has an equivalent reduced formula, its *reduct*. Without loss of generality, we shall deal with reduced formulas below. A^- denotes the reduct of $\neg A$.

A formula is *primitive* if it contains no quantifiers. A primitive formula is *primitively true* if any formula which results from an arbitrary uniform substitution of its atomic subformulas with propositional letters is a tautology.[9]

A formula is *elementary regular* if either it is primitively true or it is a sum in which at least one summand is primitively true.

Every formula can be put in the following form (by means of the usual transformations):

$$(\forall x_1) \cdots (\forall x_n)\, ((A_{11} \vee \cdots \vee A_{1p_1}) \wedge \cdots \wedge (A_{k1} \vee \cdots \vee A_{kp_k})).$$

Here $n \geq 0$, $k \geq 1$, $p_i \geq 1$. The members of the conjunction are (by definition) prime factors and are called *exterior factors*, while the members of the disjunctions are (by definition) prime summands and are called *exterior summands*; the quantifiers are called *exterior quantifiers*.

Main definition (Regular formula). A formula is *regular* if, when put in the above form, every exterior factor of the formula is either elementary regular, or it can be made elementary regular by finitely iterated application of the following operations (1)-(3).

Operation (1): shift of the universal quantifier. Any universal quantifier of a prime summand of an exterior factor is moved backwards out of the parentheses and put at the end of the sequence of the exterior quantifiers (renaming the variable if necessary).[10]

Operation (2): separation from the existential quantifier. If an exterior summand has the form $(\exists x)B(x)$, a further summand of the form $B(t)$ is added to the sum in which it occurs, where t is any variable not occurring bound in the formula $(A_{11} \vee \cdots \vee A_{1p_1}) \wedge \cdots \wedge (\mathrm{A}_{k1} \vee \cdots \vee A_{kp_k})$.

Operation (3): distributive operation. If an exterior factor has a summand (in which quantifiers occur) which is a product, i.e. if the factor has the form $A_1 \wedge \cdots \wedge A_{n_1} \vee B$ (where the A's are prime factors), then that factor is replaced by the product $(A_1 \vee B) \wedge \cdots \wedge (A_n \vee B)$.

This concludes the definition of regularity.

[9] A primitive formula is *primitively false* if its negation is primitively true.

[10] I.e., if it coincides with another variable in the formula.

A *regularity series* of a formula A is a sequence $A_0, \cdots, A_n = A$, where all the exterior factors of the formula A_0 are elementary regular, and for all i, $0 < i \leq n$, the formula A_{i-1} is obtained from the formula A_i by a single application of one of the operations (1)-(3).[11]

All the formulas of a regularity series are regular. Properties of regular formulas can be proved to hold by induction on the length of the regularity series of the given formula.

Let us see a simple example of a regular formula. Take the following formula: $(\exists x)(A(x) \rightarrow (\forall y)A(y))$. Put in the above form, it becomes $(\exists x)\ (\neg A(x) \vee (\forall y)\ A\ (y))$, with a single exterior factor containing a single exterior summand, without exterior quantifiers. We have the (possible) regularity series:

[A_n=A] $(\exists x)\ (\neg A(x) \vee (\forall y)\ A(y))$;

[A_{n-1}] $\neg A(t) \vee (\forall y)\ A(y) \vee (\exists x)\ (\neg A(x) \vee (\forall y)\ A(y))$, by Operation (2);

[A_{n-2}] $(\forall z)\ (\neg A(t) \vee A\ (z) \vee (\exists x)\ (\neg A(x) \vee (\forall y)\ A(y)))$ by Operation (1), renaming the variable y to avoid collision; finally,

[A_{n-3}=A_0] $(\forall z)\ (\neg A(t) \vee A\ (z) \vee \neg A(z) \vee (\forall y)\ A(y) \vee (\exists x)\ (\neg A(x) \vee (\forall y)\ A(y)))$,

again by Operation (2), this time choosing the variable z, which is useful for the (sort of) *unification* we need here, and on the other hand can be freely used, since it respects the restriction in the formulation of Operation (2) above. The only exterior factor of the latter formula is elementary regular, since it contains the primitively true summand $A(z) \vee \neg A(z)$. Thus, the formula [A] is regular (here $n = 3$).

Note that after each application of Operation (2), the affected existentially quantified formula is *retained*, so that it is still available and can be used again when needed.[12] It should be clear that by means of the notion of regularity Novikov in fact independently introduced a new, original notion of cut-free provability (with invertible deduction rules given in reverse order in his definition, in terms of the three operations), since a regularity series for a certain formula is a (form of) cut-free proof of the formula, more Herbrand-style (or Bernays-Quine-Dreben-Craig-Lyndon style, just to cite a few possible analogues, all descending from that source;

[11] In view of the obvious non-uniqueness, in general, of the regularity series of a formula, when we say below 'the regularity series of the formula A' we always mean a fixed chosen series in the given context; non-uniqueness is immaterial for all the proofs given or sketched below.

[12] Something similar happens in Gentzen's first consistency proof for arithmetic, when he considers his reduction rules for sequents. It is also a common feature in classical analytic systems.

see, e.g., Craig 1957) than Gentzen-style (either in the sense of sequent calculi[13] or variants of natural deduction) or in the style of resolution methods. The possible steps according to Operations (1)-(3) (in particular, the first two) transform a given formula by means of a process which can be interpreted as producing something very close to the various stages of a kind of Herbrand expansion (with variables) of the formula, here aiming (in each factor) at a final disjunction containing a disjunct which is an instance of a tautology.[14]

Every formula can also be put in the following form (dual to the form given above):

$$(\exists x_1) \cdots (\exists x_n) \, (A_{11} \wedge \cdots \wedge A_{1p_1} \vee \cdots \vee A_{k1} \wedge \cdots \wedge A_{kp_k}).$$

As above, dually, the members of the disjunction are prime summands and are called *exterior summands*, while the members of the conjunctions are prime factors and are called *exterior factors*; the quantifiers are called *exterior quantifiers*.

We can then define three operations, (1*)-(3*), dual to the operations (1)-(3) defined above (they will have a decisive role in the main theorem below).

Operation (1*): shift of the existential quantifier. If an exterior factor has the form $(\exists x)B(x)$, the existential quantifier is moved backwards out of the parentheses, and put at the end of the sequence of the exterior quantifiers (renaming the variable if necessary).

Operation (2*): separation from the universal quantifier. If an exterior factor has the form $(\forall x)B(x)$, a further factor of the form $B(t)$ is added to the product in which it occurs, where t is any variable not occurring bound in the formula $A_{11} \wedge \cdots \wedge A_{1p_1} \vee \cdots \vee A_{k1} \wedge \cdots \wedge A_{kp_k}$.

Operation (3*): dual distributive operation. If an exterior summand has a factor (in which quantifiers occur) which is a sum, i.e. if the summand has the form $(A_1 \vee \cdots \vee A_n) \wedge B$ (where the A's are prime summands), then that summand is replaced by the sum $A_1 \wedge B \vee \cdots \vee A_n \wedge B$.

2.2. *Lemmas on regular formulas*

Our aim is to prove that any formula provable in the predicate calculus is regular. Several lemmas are needed. First we show some basic properties of regular formulas (we only give in brackets hints for the proofs).

[13] But there is a closer resemblance (visible in the above example) with the sequent form of analytic tableaux (see, e.g., Smullyan 1968), in which one uses a system of reversed rules for sequents embodying a reformulation (à la Schütte and Kanger) of the tableaux methods (of Beth and Hintikka).

[14] A similar remark was made by Mints with respect to some aspects of Shanin's 'majorant semantics', which is independent of Novikov's work (see Mints 1991, 413).

Property 0. The formula $(\forall x)A(x)$ is regular iff the formula $A(x)$ is regular. (Proof: by considering the regularity series of the formula.)

Remark 0. The regularity of a formula is determined only by the exterior factors of the formula. By replacing Operation (1) with Operation (1′): elimination (instead of shift) of the universal quantifier, by renaming variables (if necessary), and by applying the other operations in the same order, we can transform any regularity series into another series without altering the regularity of the occurring formulas; thus, in the definition of regularity, (1′) can replace (1). (Proof: by the previous property.)

Property 1. Any exterior factor of a regular formula is a regular formula. (Proof: by considering the regularity series of the formula.)

Property 2. If every exterior factor of a formula is regular, then the formula is regular. (Proof: by operating on the product of the exterior factors of the formula.)

Property 3. A product is regular iff all its factors are regular. (Proof: by Properties 1 and 2).

Property 4. Every regular formula is provable in the predicate calculus. (Proof: by observing first that the operations (1)-(3) transform any formula into a logically equivalent one; this is proved by simple logical equivalences provable in the predicate calculus, e.g., for (2), $(\exists x)A(x) \equiv A\ (y) \vee (\exists x)A(x)$; secondly, that elementary regular formulas are provable by definition; finally, that taking products and adding exterior quantifiers preserves provability.)

Property 5. If the formula which results from the application of one of the operations (1)-(3) to a given formula is regular, then the latter formula was regular. (Proof: by definition of regularity.)

Property 6. If any universal quantifier is deleted from a regular formula, the result is a regular formula. (Proof: by considering the regularity series of the formula, by the eliminability of exterior universal quantifiers, see above.)

Now we give the necessary lemmas. Their proofs are usually lengthy and tedious, and sometimes nontrivial, though not conceptually difficult. We only sketch them below (except for the fundamental Lemmas 5, 1* and 4*). The usual restrictions on free occurrences of variables are assumed without mention.

Lemma 1. If some summands of the exterior factors of a regular formula are products, and we delete from every such product any number of factors, but not every one, we obtain a regular formula. (Proof: by induction on the length of the regularity series of the formula, and by cases in the induction step.)

Lemma 2. If in a regular formula we add any summands to the summands of the exterior factors of the formula, we obtain a regular formula. (Proof: as above, Lemma 1.)

Lemma 3. If in a regular formula we replace a free variable with a term, we obtain a regular formula. (Proof: as above, Lemma 1.)

Remark 1. If we apply one of the operations (1)-(3) to a regular formula, we obtain a regular formula. (Proof: by showing that the modified exterior factor remains regular, by Property 6 above, Lemma 2, Lemma 1, resp.)

Lemma 4. If both formulas $A \vee K$ and $B \vee K$ are regular, then the formula $A \wedge B \vee K$ is regular. (Proof: by cases, nontrivial.)

Corollary 1. If A is a regular formula whose exterior factors are of the form $A' \vee B' \vee B''$, and the formula $C \vee B'$ is regular, then by replacing in A any number of exterior factors with $A' \wedge C \vee B' \vee B''$ (renaming variables if necessary) we obtain a regular formula. (Proof: by Lemma 4.)

Remark 2. The formula $A_1 \wedge \cdots \wedge A_n \vee B$ is regular iff the formula $(A_1 \vee B) \wedge \cdots \wedge (A_n \vee B)$ is regular. (Proof: by the previous lemmas.)

The following lemma will have a crucial role below, and we give a full proof.

Lemma 5. If both formulas $(\exists x)A(x) \vee L$ and $B \vee L$ are regular, then the formula $(\exists x)(A(x) \wedge B) \vee L$ is regular.

Proof. The proof is by induction on the length of the regularity series of the formula, and by cases in the induction step. We have to show the admissibility (for regularity) of a form of movement[15] of the existential quantifier over conjunction (a basic prenex operation); in other words, that reductions for conjunction and the existential quantifier commute.

Consider the regularity series $K_0, \ldots, K_m$ of $(\exists x)A(x) \vee L$, in which the first summand is always present (since no operation can delete it). For K_0, if the formula is elementary regular, then L must contain a primitively true summand, and this makes also $(\exists x)(A(x) \wedge B) \vee L$ elementary regular. Assume the lemma holds for K_{i-1}. In K_i we have a factor of the form $(\exists x)A(x) \vee L^i$, and in K_{i-1} we have either the same factor (if another factor is affected), or a factor of the form $(\exists x)A(x) \vee L^{i-1}$ (if L^i is affected) or finally a factor of the form $A(t) \vee (\exists x)A(x) \vee L^i$. The first case is trivial; for the second case we apply directly the induction hypothesis, since $(\exists x)A(x)$ is not affected. In the third case, since by hypothesis both $(\exists x)A(x) \vee L^i$ and

[15] In fact, a form of distributivity (since x is not free in B), if one reads existentially quantified formulas as countable sums, as in Novikov (1943). The result of this lemma is a notorious sore point in Herbrand's Thesis (Ch. 5, Sec. 3.3).

$B \vee L^i$ are regular, also $A(t) \vee (\exists x)A(x) \vee L^i$ and $A(t) \vee B \vee L^i$ are regular (by adding the same summand $A(t)$ to both regular formulas, Lemma 2 above), and then by induction hypothesis also $A(t) \vee (\exists x)\,(A(x) \wedge B) \vee L^i$ is regular. Hence, by Corollary 1 above (recall that $B \vee L^i$ is regular), $A(t) \wedge B \vee (\exists x)(A(x) \wedge B) \vee L^i$ is regular, but then (by definition of Operation (2)) also $(\exists x)(A(x) \wedge B) \vee L^i$ is regular. This concludes the proof. □

The following remark will have a similar role as the previous lemma.

Remark 3. If the formula $(\exists x)A(x) \vee L' \vee L''$ is regular, then the formula $(\exists x)(A(x) \vee L') \vee L''$ is regular. (Proof: by applying Operation (2) to the second formula.)

Recall that A^- denotes the reduced form of $\neg A$.

Remark 4. For every formula A, the formula $A \vee A^-$ is regular. (Proof: by induction on the structure of the reduced formula.)

The following lemmas (on the dual operations (1*)-(3*)) are also crucial for the reduction procedure on our system. We fully prove the first (the others are proved similarly).

Lemma 1*. If $A \vee H$ is a regular formula, and A' is obtained from A by Operation (1*), then $A' \vee H$ is regular.

Proof. The proof is by double induction:[16] on the number of exterior existential quantifiers in the formula and on the length of its regularity series.

We have to prove that if the formula

$$(*)\ \ (\exists x_1) \cdots (\exists x_n)((\exists y)A(y) \wedge B \vee C) \vee H$$

is regular, then the formula

$$(**)\ \ (\exists x_1) \cdots (\exists x_n)(\exists y)(A(y) \wedge B \vee C) \vee H,$$

obtained by applying Operation (1*), is also regular.

If there are no exterior quantifiers, then since $(\exists y)A(y) \wedge B \vee C \vee H$ is regular, both $(\exists y)A(y) \vee C \vee H$ and $B \vee C \vee H$ are regular, and thus by the previous lemmas (crucially, Lemma 5 and Remark 3 above) $(\exists y)\,(A(y) \wedge B \vee C) \vee H$ is regular.

[16] That we need a double induction on ω in the proof of a lemma for (a form of) cut elimination for the predicate calculus is hardly surprising.

If the formula is elementary regular, then Operation (1*) cannot be applied to the relevant primitively true summand, and regularity is preserved.

The first induction is on the number of exterior quantifiers. For the induction step, let us consider $K_0, \ldots, K_m$, the regularity series of the formula (*). We need a second induction, on the length of this regularity series. For the induction step of the second induction, we have to prove that applying Operation (1*) to any summand of the form $(\exists x_1) \cdots (\exists x_n)((\exists y) A(y) \wedge B \vee C)$ preserves regularity.

If the operation leading from K_i to K_{i-1} does not apply to any such summand, then regularity is preserved. The crucial case is the one in which one such summand is affected. But the only operation which could have been applied to it, in that case, is Operation (2). Then we have a formula (in K_{i-1}) of the form

$$(\exists x_2) \cdots (\exists x_n)((\exists y) A(y, t) \wedge B(t) \vee C(t)) \vee (\exists x_1) \cdots \\ (\exists x_n)((\exists y) A(y) \wedge B \vee C \vee H.$$

By the second induction hypothesis, the formula

$$(\exists x_2) \cdots (\exists x_n)((\exists y) A(y, t) \wedge B(t) \vee C(t)) \vee (\exists x_1) \cdots \\ (\exists x_n)(\exists y)(A(y) \wedge B \vee C) \vee H,$$

which results by applying Operation(1*) to the second main summand, is still regular, and by the first induction hypothesis also the formula

$$(\exists x_2) \cdots (\exists x_n)(\exists y)(A(y, t) \wedge B(t) \vee C(t)) \vee (\exists x_1) \cdots \\ (\exists x_n)(\exists y)(A(y) \wedge B \vee C) \vee H,$$

obtained by further applying the same operation, this time to the first main summand, is regular. The latter formula can be obtained by Operation (2) from a formula of the form

$$(\exists x_1) \cdots (\exists x_n)(\exists y)(A(y) \wedge B \vee C) \vee H,$$

which is thus regular, as we needed. This concludes the proof. □

Lemma 2*. The same as Lemma 1*, with Operation (2*) in place of Operation (1*). (Proof: as above, Lemma 1*.)

Lemma 3*. The same as Lemma 1*, with Operation (3*). (Proof: as above, Lemma 1*.)

Finally, we need to show that deleting from regular disjunctions certain subformulas containing primitively false components preserves regularity.

Lemma 4*. If the formula $(\exists x_1) \cdots (\exists x_n)(A_1 \wedge B_1 \vee \cdots \vee A_p \wedge B_p) \vee H$ is regular, and every formula A_i is primitively false, then the formula H is regular.

Proof. The proof is again by double induction (number of exterior quantifiers and length of the regularity series). If there are no exterior quantifiers, if the formula is regular then (by Lemma 1 above) also $A_1 \vee \cdots \vee A_p \vee H$ is regular; if we can delete all the primitively false summands A_i preserving regularity, H must be regular. That this can be done is proved by showing (by induction on the length of the regularity series of the formula) that in any regular sum any primitively false summand can be deleted preserving regularity. For the base case, if the primitively false summand occurs in a primitively true component (the only nontrivial case), then the latter must have the form of a sum of that summand with another summand that must be primitively true, and then the elimination can be done. For the induction step, it is sufficient to recall that none of the operations (1)-(3) can be applied to a primitive (here, primitively false) formula.

If there is at least an exterior quantifier,[17] we have to prove (induction step of the first induction) that H is regular if the hypotheses of the lemma are satisfied, assuming that this holds when one initial quantifier (the first one, without loss of generality) is absent; and this is proved by a second induction, on the length of the regularity series of the formula of the lemma (where the base case is trivial, since in that case H is elementary regular). The only nontrivial case in the induction step of the latter induction is the one in which our formula is affected, and in that case (being the formula existentially quantified, as in the proof of Lemma 1*) we must have an application of Operation (2). Thus we obtain a regular formula of the form:

$$(\exists x_2) \cdots (\exists x_n)(A_1(t) \wedge B_1(t) \vee \cdots \vee A_p(t) \wedge B_p(t)) \vee \\ \vee (\exists x_1) \cdots (\exists x_n)(A_1 \wedge B_1 \vee \cdots \vee A_p \wedge B_p) \vee H,$$

where all formulas $A_i(t)$ remain primitively false. Hence, by the second induction hypothesis

$$(\exists x_2) \cdots (\exists x_n)(A_1(t) \wedge B_1(t) \vee \cdots \vee A_p(t) \wedge B_p(t)) \vee H$$

is regular; then, by the first induction hypothesis, H is regular. This concludes the proof. □

2.3. *The main theorem*

Theorem. *Every formula provable in the predicate calculus is regular.*

Proof. The (reduced forms of the) propositional axioms are tautologies, hence they are primitively true and (by definition) regular. For the axiom

[17] This case is not explicitly treated by Novikov.

(schema) on the universal quantifier, $(\forall x)A(x) \rightarrow A(y)$, we have the reduced form[18] $(\exists x)\ \neg A(x) \vee A(y)$, from which by Operation (2) the elementary regular formula $\neg A(y) \vee (\exists x)\neg A(x) \vee A(y)$ is obtained. The proof for the dual axiom is similar.

For the rule on the universal quantifier, let us assume that the premise $A \rightarrow B(x)$ (x not free in A) is regular. Then its reduct (assuming without loss of generality that B is reduced) $A^{-} \vee B(x)$ is also regular, and $(\forall x)$ $(A^{-} \vee B(x))$ remains regular (by adding a universal quantifier, see above). But this formula can be obtained by Operation (1) from $A^{-} \vee (\forall x)\ B(x)$, which is the reduct of the conclusion of the rule, which is thus regular. An analogous proof can be given for the dual rule on the existential quantifier.

The only interesting case is the one concerning Modus Ponens. Assume that $A \rightarrow B$ and A are regular. The reduced form of A can be written in the form

$$(\forall x_1) \cdots (\forall x_n)\ ((A_{11} \vee \cdots \vee A_{1p_1}) \wedge \cdots \wedge (A_{k1} \vee \cdots \vee A_{kp_k})).$$

This formula is regular, thus it can be reduced by means of Operations (1)-(3) to the form

$$(\forall x_1) \cdots (\forall x_{n+r})\ ((A_0^1 \vee C_0^1) \wedge \cdots \wedge (A_0^m \vee C_0^m)).$$

where all the formulas A_0^i are primitively true. On the other hand, the reduced form of $A \rightarrow B$ is $A^{-} \vee B'$ where B' is the reduct of B. The form of the formula A^{-} is

$$(\exists x_1) \cdots (\exists x_n)\ (A_{11}^{-} \wedge \cdots \wedge A_{1p_1}^{-} \vee \cdots \vee A_{k1}^{-} \wedge \cdots \wedge A_{kp_k}^{-})),$$

dual with respect to the reduced form of A given above. By applying Operations (1*)-(3*) to this form in the dual way with respect to the above applications of Operations (1)-(3), we obtain the form

$$(\exists x_1) \cdots (\exists x_{n+r})\ (A_0^{1-} \wedge C_0^{1-} \vee \cdots \vee A_0^{m-} \wedge C_0^{m-}).$$

The application of Operations (1*)-(3*) to A^{-} in $A^{-} \vee B'$ preserves regularity of the whole formula (by the lemmas on these operations, see above); hence the formula

$$(\exists x_1) \cdots (\exists x_{n+r})\ (A_0^{1-} \wedge C_0^{1-} \vee \cdots \vee A_0^{m-} \wedge C_0^{m-}) \vee B'$$

is regular. But all the formulas A_0^{i-} are primitively false, since all the formulas A_0^i are primitively true; then, by the last lemma of the previous subsection, B', and thus B, are regular. This concludes the proof. $\square$

[18] This is the simplest case, easily generalized by Lemma 2 and Remark 4.

3. A comparison

Consider the 'cut formula' A in the above proof that Modus Ponens preserves regularity. We have seen that its reduct can be written in the form

$$(A)\ \ (\forall x_1) \cdots (\forall x_n)\,((A_{11} \vee \cdots \vee A_{1p_1}) \wedge \cdots \wedge (A_{k1} \vee \cdots \vee A_{kp_k})).$$

while the reduct of its negation can be written as

$$(A^{-})\ \ (\exists x_1) \cdots (\exists x_n)\,(A^{-}_{11} \wedge \cdots \wedge A^{-}_{1p_1} \vee \cdots \vee A^{-}_{k1} \wedge \cdots \wedge A^{-}_{kp_k})).$$

Let us see what would happen in the case of a Gentzen-style cut elimination in which A is indeed the cut formula. To avoid unnecessary complications and to make the comparison easier, we shall use (after the well-known approach of Schütte and Tait; see Schütte 1951) the formulation with derivable disjunctions (instead of ordinary sequents) and negation for compound formulas *defined* by DeMorgan Laws. We shall also ignore some details without loss of generality (e.g., freely using associativity, commutativity and idempotence of disjunction) and use cut in the (equivalent) simplified form which allows to obtain B from A and $\neg A \vee B$.

First observe that the (exact) statements of the crucial lemmas above are provable without difficulties: in any reasonable formulation of a sequent calculus the cut-free derivability of $(\exists x)A(x) \vee L$ and $B \vee L$ almost immediately implies the cut-free derivability of $(\exists x)\,(A(x) \wedge B) \vee L$ (Lemma 5),[19] and this is the main ingredient in the induction which proves Lemma 1*; while the point of the crucial Lemma 4* (the fact that in any cut-free derivable sum any primitively false summand can be deleted without destroying cut-free derivability) is obtained by simply observing that the primitively false summand could have been introduced only by weakening. Here the comparison is rather trivial.

A little more interesting is the case of Modus Ponens in the main theorem. We have to show (directly taking reduced formulas, without loss of generality) that B can be proved without cuts, assuming that the same holds for A and $A^{-} \vee B$. In sequent calculi (in the formulation we have chosen) this would be carried out, of course, by induction on the complexity of A (together with a sub-induction, as usual, on cut rank, which however is minimal by hypothesis in the case we have chosen), taking the relevant subformulas and applying the cut rule to them. In view of the form of A (given above) this can be done as follows.

If all initial (external) quantifiers are absent, the procedure is straightforward. On the left, we have $(A_{11} \vee \cdots \vee A_{1p_1}) \wedge \cdots \wedge (A_{k1} \vee \cdots \vee A_{kp_k})$,

[19] We noted above that, on the contrary, this is problematic in Herbrand's Thesis (Ch. 5, Sec. 3.3).

which can be (meta-theoretically) decomposed (by inversion), as a cut-free derivable conjunction, into its cut-free derivable component disjunctions. On the right, we have its negation with added the further summand B, i.e. the formula $(A^-_{11} \land \cdots \land A^-_{1p_1}) \lor \cdots \lor (A^-_{k1} \land \cdots \land A^-_{kp_k}) \lor B$. This disjunction can be directly used as the right premiss of a new cut (this is a typical advantage of Schütte's formulation, which we apply to our further simplified form of cut), where the left premiss of this new cut is one of the component disjunctions (say, the first) just obtained from the decomposed conjunction which was on the left. This is precisely what we need for the induction step: we simply iterate the new cuts, eliminating one component at a time from the disjunction (the new right premiss), until only B is left. Schematically, we have:

$$\dfrac{\dfrac{\cdots \qquad \dfrac{\dfrac{\cdots}{(A_{21} \lor \cdots \lor A_{2p_2})} \qquad \dfrac{\dfrac{\cdots}{(A_{11} \lor \cdots \lor A_{1p_1})} \qquad \dfrac{\cdots}{(A^-_{11} \land \cdots \land A^-_{1p_1}) \lor \cdots \lor B}}{(A^-_{21} \land \cdots \land A^-_{2p_2}) \lor \cdots \lor B}}{(A^-_{31} \land \cdots \land A^-_{3p_3}) \lor \cdots \lor (A^-_{k1} \land \cdots \land A^-_{kp_k}) \lor B}}{\cdots \quad \cdots}}{B}$$

There is still the case in which some initial (external) quantifiers are present. The induction step is not especially troublesome in this case, since the scope of the remaining quantifiers is preserved as a whole (except for the newly indicated eigenvariable of the universal quantifier which has been removed) and the whole remaining formula (after substitution of the eigenvariable with the respective variable occurring in the right subderivation, where the corresponding existential quantifier has been removed) becomes the cut formula.

On the other hand, in Novikov's proof given above we do not directly make any induction on the complexity of the cut formula, but we *transform* it, reducing the formula to a form in which its primitively true summands are explicitly isolated. Then we exploit the primitively false components in A^- (which has undergone the dual transformation with respect to A), recognized as such by means of the corresponding primitively true components in A, to get B, still regular, by Lemma 4* (which is indeed proved by double induction).[20] In order to carry out this procedure, we must be assured that the transformation of A^- does not destroy regularity (i.e. cut-free derivability) of $A^- \lor B$, but this is warranted precisely by the crucial Lemmas 1*-3* (also proved by double induction).[21]

[20] Involving two measures (one direct: length of regularity series, and one indirect: number of exterior quantifiers) of length of proofs, rather than of complexity of formulas.

[21] See the preceding footnote.

Which way is simpler and more elegant is arguably a matter of taste: neither is short (they both require a long series of lemmas or distinct cases, though without conceptual trouble); Novikov's procedure shows us (so to say) the axioms incorporated in the cut formula; the usual one has the advantage of a uniform and direct inductive treatment. In any case, since we do not have any decisive argument in one direction or the other, we suspend judgement.

We have seen that Novikov transforms the cut formula in such a way that a certain tautological instantiation results, from which the formula could be ordinarily derived by introducing quantifiers in a suitable way. It is remarkable that something fairly similar happens in Herbrand's Thesis (Ch. 5, Sec. 5.1), where elimination of Modus Ponens for the predicate calculus is achieved *indirectly*, as a consequence of the fact (which is part of Herbrand's 'Fundamental Theorem' modulo the usual necessary corrections) that every provable formula has Property A, i.e. (roughly) that the result of a certain instantiation (*without* constant or functional terms) of the variables in a disjunction obtained from a prenex equivalent of the formula is a quantifier-free tautology (see, for the rather involved definition, ibid. Sec. 2), and that formulas with Property A can be derived from their corresponding tautologies without Modus Ponens.[22]

Another striking analogy is with Bernays' treatment of (a suitable adaptation of) Herbrand's theorem and related results in *Grundlagen* II §3.3 (Hilbert-Bernays 1939, 149 ff.).[23] Although his proofs are based on the second Epsilon-theorem (see ibid. §3.1), hence ultimately (if one tracks the dependence of theorems) on the use of epsilon-substitution techniques (which have no analogue in Novikov), we find (ibid., 160 (b), 158 in the first ed.) the following result (with some generalizations), which certainly should sound familiar at this point: to each prenex formula provable in the predicate calculus we can associate a disjunction with the following properties: (1) each disjunct is obtained from the formula by deleting all quantifiers, replacing universal variables with free variables, and existential variables with terms built on free variables and constants and function symbols

[22] One definitely has the impression that Herbrand (whose name is mentioned only once in passing in Novikov 1949, never in 1939, 1943, not even in the textbook of 1959) is the true 'stone guest' in this whole story. Indeed, what Novikov gives is basically a reformulation in Herbrandian terms of Gentzen's result. We note in passing that a variant of Herbrand's Property A resurfaces in Ackermann (1954, 90 ff.), but in connection with decision problems, thus *employing* the fact that Schütte's system mentioned above (i.e., the first system in Schütte 1951) is semantically complete and enjoys cut elimination.

[23] Although it is often quite difficult to trace Novikov's (and some other Soviet authors') sources, it is clear from his works (e.g. the introduction of Novikov 1959) that he knew (beyond Hilbert-Ackermann) the *Grundlagen* (but he apparently never cites Vol. II), although his first results on regularity (which concern infinitary propositional derivations), published already in 1939, are independent and utterly original.

occurring in the formula; (2) the disjunction is an instance of a tautology; (3) the formula can be derived from the disjunction by applying the ordinary rules of introduction of quantifiers in members of disjunctions[24] and a rule of contraction.[25]

4. Concluding remarks

We hope to have given sufficient evidence, at least, of the originality of Novikov's method of elimination. If, on the one hand, the interest of his procedure is mainly theoretical, since it seems to allow no gain in efficiency, on the other hand, the fact that in the ordinary[26] setting of a Hilbert proof system he needs for reduction no form of induction directly on (what amounts to) the cut formula is a peculiar feature, whose consequences are not immediately clear, and surely deserve further investigation.

Acknowledgements

I wish to thank Paolo Mancosu, Enrico Moriconi, Michele Pagano, Jan von Plato and the referees.

References

[1] Ackermann, W. (1954), S*olvable cases of the decision problem*, Amsterdam.
[2] Barwise, J. (1981), Infinitary logics, in *Modern logic – a survey*, Dordrecht, 93-112.
[3] Church, A. (1946), Review of Novikov (1943), *Journal of Symbolic Logic*, 11, 129-131.
[4] Coquand, T. (1995), A semantics of evidence for classical arithmetic, *Journal of Symbolic Logic*, 60, 325-337.
[5] Craig, W. (1957), Linear reasoning. A new form of the Herbrand-Gentzen theorem, *Journal of Symbolic Logic*, 22, 250-268.
[6] Gentzen, G. (1934), Untersuchungen über das logische Schließen, *Mathematische Zeitschrift*, 39, 176-210, 405-431.
[7] Herbrand, J. (1968): *Écrits logiques*, Paris; English translation, Dordrecht 1971.
[8] Hilbert, D. and P. Bernays (1939), *Grundlagen der Mathematik*, Vol. II, Berlin; 2nd ed., 1970.

[24] I.e., the derivable rules $A \vee B\ (t)$ / $A \vee (\exists x)\ B(x)$, resp. $A \vee B\ (x)$ / $A \vee (\forall x)\ B(x)$ under the usual suitable restrictions.

[25] In a formulation which allows to eliminate any repetitions of members in disjunctions.

[26] Viz., neither with the Herbrand proof method nor with the Hilbert-Bernays epsilon calculus.

[9] Miller, D. (1987), A compact representation of proofs, *Studia Logica*, 46, 347-370.

[10] Mints, G. (1991), Proof theory in the USSR 1925-1969, *Journal of Symbolic Logic*, 56, 385-424.

[11] Novikov, P. S. (1939), Sur quelques théorèmes d'existence, *Comptes Rendus (Doklady) de l'Academie des Sciences de l'URSS*, 23, 438-440 (French).

[12] Novikov, P. S. (1943), On the consistency of certain logical calculus, *Matematicheskii Sbornik*, 12, 231-261 (English).

[13] Novikov, P. S. (1949), O klassakh regulyarnosti (On classes of regularity), *Doklady Akademii Nauk SSSR*, 64, 293-295 (Russian).

[14] Novikov, P. S. (1959), *Elementy matematicheskoi logiki* (Elements of mathematical logic), Moscow (Russian); 2nd ed., 1973; English translation, Edinburgh-London and Reading (Mass.) 1964.

[15] Novikov, P. S. (1979), *Izbrannye trudy. Teoriya mnozhestv i funktsii. Matematicheskaya logika i algebra* (Selected works. Theory of sets and functions. Mathematical logic and algebra), Moscow (Russian).

[16] Schütte, K. (1951), Schlußweisen-Kalküle der Prädikatenlogik, *Mathematische Annalen*, 122, 47-65.

[17] Smullyan, R. (1968), *First-order logic*, Berlin; 2nd ed., 1995.

[18] Tait, W. W. (1968), Normal derivability in classical logic, in *The syntax and semantics of infinitary languages*, LNM 72, Berlin, 204-236.

[19] Tsinman, L. L. (1968), On the role of the induction principle in a formal arithmetic system, *Matematicheskii Sbornik*, 77, 71-104 (Russian); English translation, *Mathematics of the USSR – Sbornik*, 6, 65-95.

[20] Tupailo, S. (1992), Gentzen-style and Novikov-style cut-elimination, in *Logical foundations of computer science – Tver '92*, LNCS 620, Berlin, 493-502.

Luca Bellotti
Department CFS (Philosophy), University of Pisa
Via Paoli 15, 56126 Pisa, Italy
Email: luca.bellotti@unipi.it