Cut elimination in sequent calculi with implicit contraction, with a conjecture on the origin of Gentzen's altitude line construction

Sara Negri and Jan von Plato
University of Helsinki

Abstract: Sequent calculi are given in which contexts represent finite sets of formulas. Standard cut elimination will not work if the principal formula of a logical rule is already found in a premiss, i.e., if there is an implicit contraction on it. A procedure is given in which cut with the original cut formula is first permuted up, followed by cuts on its immediate subformulas. It is next adapted to sequent calculi with multisets and explicit contraction, by which Gentzen's mix rule trick is avoided, a procedure strikingly similar to the peculiar "altitude line" construction that Gentzen used in his second proof of the consistency of arithmetic in 1938. The conjecture is close at hand that this is indeed the way Gentzen originally proved cut elimination in 1933.

1. Introduction

Axiomatic logic began with Frege and was perfected in the Hilbert school in the 1920s, with the aim to apply it in the formalization of mathematical proofs. The crucial *deduction theorem* shows how this goes: Given some axiomatic system with axioms H expressed in the language of logic, it should be possible to treat it, not as a collection of *truths* added to the logical axioms, but as *hypotheses.* Thus, derivations in axiomatic logic were extended to include derivations under hypotheses, as in the book that first defined a complete formal system of predicate logic, [Hilbert and Ackermann, 1928], and very soon the aim was reached: A proof of a proposition P by the mathematical axioms can be converted into a purely logical proof of the implication $H \supset P$. The matter is not as trivial as it might seem: Say, if $A(x)$ is provable without assumptions, $\forall x A(x)$ can be concluded, whereas if $A(x)$ is proved under assumptions, nothing of the kind need follow.

Turing in a manuscript of 1944, *The reform of mathematical notation and phraseology*, was keenly aware of how mathematical proof hangs on the deduction theorem. A version of Turing's manuscript is found printed in [Cooper and van Leeuwen, 2013] to which we refer. The first of his two central points was (p. 246): 1. "Free and bound variables should be understood by all and properly respected." He then gives an example of constants and variables and adds: "The difference between the constants and the free variables is somewhat subtle. The constants appear in the formula as if they were free variables, but we cannot substitute for them. In these cases there has always been some assumption made about the variable (or constant) previously."

Turing's second point was (*ibid.*): 2. "The deduction theorem should be taken account of." The deduction theorem is the main way of handling free variables: "This process whereby we pass from P proved under an assumption H to 'If H then P' may be called 'absorption of hypotheses'. The process converts constants or 'restricted variables' into free variables."

His example, slightly rephrased, is: Let the radius a and volume v of a sphere be given. Then $v = \frac{4}{3}\pi a^3$.

> The 'deduction theorem' states that in such a case, where we have obtained a result by means of some assumptions, we can state the result in a form in which the assumptions are included in the result, e.g., 'If a is the radius and v is the volume of the sphere then $v = \frac{4}{3}\pi a^3$. In this statement a and v are no longer constants.

A sufficiently detailed proof of the deduction theorem, as in [Hakli and Negri, 2012] or [von Plato, 2013, section 3.6] gives an algorithm for converting hypothetical proofs that use mathematical axioms as assumptions into purely logical proofs without assumptions, but the algorithm is hopelessly clumsy for any practical use. Axiomatic logic was put to rest as a proof system of logic in an act that rendered the famous deduction theorem a proof-theoretical triviality: it became the rule of implication introduction of natural deduction. Indeed, if a translation from the latter to axiomatic logic is defined in sufficient detail, the deduction theorem will come out simply as the image of implication introduction under that translation (as in [von Plato, 2013, p. 55]). When written in terms of sequent calculus with the sequent arrow $\Gamma \rightarrow C$ that indicates the *derivability* of C under the collection of assumptions Γ, the deduction theorem becomes the rule: If $H \rightarrow P$, then $\rightarrow H \supset P$, with no assumptions left.

In proving the deduction theorem of axiomatic logic, there are two limiting cases: an axiom H was not at all used in a proof, and secondly, it was used repeatedly. The former is easy and the latter routinely skimmed over in logic books. Now we have come to the central problem: How to count the multiplicity of assumptions in proofs? Many books in proof theory are as light-hearted about this problem as are books on axiomatic logic. The latter may include Frege's *contractive axiom* $(A \supset (A \supset B)) \supset (A \supset B)$ that can be used for reducing a multiple deduction theorem to a singular one.

The easiest way with assumptions would be if they could be just named and counted once. In a second approach, we could list them in the order in which they are taken, in a third, we could repeat in this list assumptions that were used more than once. Finally, we could ignore the order that seems perhaps not to be intrinsic but at least in part imposed by the linearity of writing proofs as texts, which makes the first and second approach collapse into one that uses *finite sets*. The third uses *lists with multiplicity*, or *lists* for short, and the fourth *finite multisets*.

Logical rules are next to the formulas the central part of the syntax of a logical proof system; thus, the rules act on *expressions*. Lists in contrast to sets seem to be, by their very nature, already formal objects of a syntax. We shall therefore begin with a little theory of expressions for finite sets, then put up a logical sequent calculus and prove the crucial cut elimination theorem directly by the use of sets. In a second part, starting with Sec. 4, we draw conclusions about our proof that point at so far unnoticed passages and constructions in Gentzen's original work connected to his cut elimination procedure: first, his actual use of the "sequents with sets" approach in the proof of decidability of intuitionistic propositional logic in 1933 that can be easily turned into a new kind of cut elimination procedure, detailed out in Sec. 5; whether Gentzen actually saw this possibility is not known at present. Secondly there is his, quite opposite, strict adherence to lists in the "altitude line" construction in the consistency proof of arithmetic of 1938, presented in detail in Sec. 6, that gives an alternative cut elimination procedure known to Gentzen.

2. Sequents with sets

Sequent calculus is the proper setting for a precise discussion of the treatment of assumptions in proofs. The rules of sequent calculus display the collection of *open assumptions* Γ at the left of an arrow and the *consequence* C of those assumptions at right, in a sequent $\Gamma \to C$. Each logical rule modifies derivability relations, as in rule $R\supset$ that gives the result $\Gamma \to A \supset B$ from the premiss $A, \Gamma \to B$, as in

$$\frac{A, \Gamma \to B}{\Gamma \to A \supset B}\, R\supset$$

Here A and B are the *active* formulas of the rule, $A \supset B$ its *principal* formula, and Γ the *context*.

It is customary to speak of derivations even in sequent calculus; no harm, if the double sense of derivability is kept in mind. The word "conclusion" can have a similar double use, either as the consequence C in a sequent $\Gamma \to C$, or the conclusion of a rule of sequent calculus. Assumptions A are presented in sequent calculus as *initial sequents* $A \to A$, as a limiting case in which the conclusion A depends on the assumption A.

In Gentzen's doctoral work of 1933, published as [Gentzen, 1934–35], the assumptions Γ were given as a *list* of formulas, thus, the order and multiplicity of occurrences of formulas was counted. Later, Katudi Ono [Ono, 1938] introduced lists in which order is not counted and proved that the resulting calculi are equivalent to Gentzen's, a discovery known to the latter (cf. his review of Ono's work, [Gentzen, 1939]), and reinvented by Kleene, Curry, and others in the 1950s. In sequent calculi as well as natural deduction, formal derivations have the form of a tree, and there is no intrinsic linear order of assumptions. Perhaps in reflection of this latter fact, it has become customary to formulate sequent calculi so that the open assumptions form finite *multisets*, and similarly for the open cases if a symmetric sequent calculus with more than one succedent formula is used. Below we pay only marginal attention to Gentzen's original exchange rules by which the order of formula occurrences in lists of assumptions and cases can be changed.

The idea has been often entertained of dispensing, next to order, also with multiplicity, known as the "sequents with sets" idea. There will be some difficulties in carrying the idea through on the level of formalization: Logical languages are inductively defined classes of formulas, and a system of logical rules gives an inductive definition of the class of derivations. Sets are objects and two sets are equal if they have the same members. However, two different expressions for the same set cannot be just like that substituted one for the other in logical rules, because syntactically incorrect rule instances can be produced: For example, a single left rule of conjunction has the premiss $A, B, \Gamma \to C$ and the conclusion $A \,\&\, B, \Gamma \to C$. If A and B are identical, the premiss collapses into $A, \Gamma \to C$ with no match with the premiss of the rule. Therefore a little theory of expressions for sets needs to be put up. This theory will contain rules by which, say, an expression $\{A, \ldots, B, B, \ldots, C\}$ for a set of formulas can be replaced by the expression $\{A, \ldots, B, \ldots, C\}$, thus, a *rule of contraction.*

Details such as the above have a bearing on the modularity of sequent calculus proof systems, as expressed by the rule of *cut* that shows how two derivations can be combined, and how it is removed in a process of cut elimination. Formally, we have a derivation of some result A, and another derivation in which A occurs as an assumption, combined as in the scheme:

$$\frac{\Gamma \to A \quad A, \Delta \to C}{\Gamma, \Delta \to C}\,{\scriptstyle Cut}$$

In the cut elimination procedure to be presented we shall, instead of a theory of expressions for sets, use *canonical expressions* for sets, as lists without repetition and with a prescribed lexicographical order, and the principle:

> *At each rule instance, any potential multiplication of formula occurrences in the conclusion is erased when the conclusion is written down.*

Initial sequents have already the property of a single occurrence, and our principle makes conclusions of logical rules inherit that property. The effect is that the rule of contraction, needed when different expressions for the same set are allowed, will become *implicit.*

Cut elimination for sequent calculi with implicit contraction is different from the Gentzen-style cut elimination procedure in which, whenever the cut formula is concluded in both premisses of cut by the rules that correspond to its logical form, cuts on shorter formulas replace the given cut. The difference is seen when, for example, the right premiss of cut has been concluded by the left conjunction rule, with an implicit contraction on the cut formula $A \,\&\, B$. In this case, the cut on $A \,\&\, B$ does not get removed, but is first permuted up, followed by cuts on the two immediate subformulas of $A \,\&\, B$. We show first that the cut elimination procedure to be defined works for a single-succedent calculus, then extend it to a calculus with symmetric sequents of the form $\Gamma \to \Delta$ that can have several cases in the succedent part. These proofs of cut elimination use the standard rule of cut, whereas in Gentzen's original work, cut elimination is done through the "mix rule," or rule of multicut as one often says, in which any number of cut formulas in the two premisses can be deleted:

$$\frac{\Gamma \rightarrow A \quad A^*, \Delta \rightarrow C}{\Gamma, \Delta \rightarrow C}\,\scriptstyle{Mix}$$

Here A^* denotes any number $n \geqslant 1$ of copies of A. [Gentzen, 1934–35] has the comment that the rule is used "to make the proof easier." The problematic case of cut elimination that led Gentzen to the mix rule is that the right premiss has been concluded by a rule of *contraction:*

$$\frac{\begin{matrix}\vdots \\ \Gamma \rightarrow A\end{matrix} \quad \dfrac{A, A, \Delta \rightarrow C}{A, \Delta \rightarrow C}\,\scriptstyle{Ctr}}{\Gamma, \Delta \rightarrow C}\,\scriptstyle{Cut}$$

The obvious idea of cutting twice with $\Gamma \rightarrow A$ gives:

$$\frac{\begin{matrix}\vdots \\ \Gamma \rightarrow A\end{matrix} \quad \dfrac{\begin{matrix}\vdots \\ \Gamma \rightarrow A\end{matrix} \quad A, A, \Delta \rightarrow C}{A, \Gamma, \Delta \rightarrow C}\,\scriptstyle{Cut}}{\Gamma, \Gamma, \Delta, \Theta \rightarrow C}\,\scriptstyle{Cut}$$

The derivation of the right premiss of the lower cut has clearly grown in some sense, with an instance of contraction replaced by an instance of cut, and no simple solution has been found to correct the situation. Intrigued by Gentzen's remark about making the proof easier, the second author produced in [von Plato, 2001] a proof of cut elimination with the standard cut rule, based on an analysis of how the premiss of contraction was derived. Contraction was then reduced by *inversion lemmas* to contractions on shorter formulas. These lemmas require subtle changes in the sequent calculi used, unknown to Gentzen.

3. Cut elimination for sequents with sets

The cut elimination procedure to be defined is rather straightforward for a single-succedent sequent calculus. Logical rules with two premisses have contexts that are added up in the conclusion, with possible multiplications of formulas erased. When a rule is written, no premiss has duplications of formulas. A "bra–ket" notation $\langle\Gamma\rangle$, adopted from a manuscript of Gentzen's of 1944, indicates a context Γ in the conclusion of a rule such that no duplications in the antecedent remain. Such duplications can arise either because the principal formula of a rule was already found in Γ or because the antecedents of two premisses of a rule overlapped with some formulas. Rules $R\&$ and $L\vee$ are written in the manner of rule $L\supset$, with contexts added up in the conclusion.

The sequent calculus LIS:

Initial sequents have the forms, with C an arbitrary formula:

$$C \to C \qquad \bot \to C$$

The logical rules are:

$$\frac{A, \Gamma \to C}{\langle A \,\&\, B, \Gamma \rangle \to C} L\& \qquad \frac{B, \Gamma \to C}{\langle A \,\&\, B, \Gamma \rangle \to C} L\& \qquad \frac{\Gamma \to A \quad \Delta \to B}{\langle \Gamma, \Delta \rangle \to A \,\&\, B} R\&$$

$$\frac{A, \Gamma \to C \quad B, \Delta \to C}{\langle A \vee B, \Gamma, \Delta \rangle \to C} L\vee \qquad \frac{\Gamma \to A}{\Gamma \to A \vee B} R\vee \qquad \frac{\Gamma \to B}{\Gamma \to A \vee B} R\vee$$

$$\frac{\Gamma \to A \quad B, \Delta \to C}{\langle A \supset B, \Gamma, \Delta \rangle \to C} L\supset \qquad \frac{A, \Gamma \to B}{\Gamma \to A \supset B} R\supset$$

$$\frac{A(t), \Gamma \to C}{\langle \forall x A(x), \Gamma \rangle \to C} L\forall \qquad \frac{\Gamma \to A(y)}{\Gamma \to \forall x A(x)} R\forall$$

$$\frac{A(y), \Gamma \to C}{\langle \exists x A(x), \Gamma \rangle \to C} L\exists \qquad \frac{\Gamma \to A(t)}{\Gamma \to \exists x A(x)} R\exists$$

The structural rules are weakening and cut:

$$\frac{\Gamma \to C}{\langle A, \Gamma \rangle \to C} Wk \qquad \frac{\Gamma \to A \quad A, \Delta \to C}{\langle \Gamma, \Delta \rangle \to C} Cut$$

Note that in rule $L\&$, for example, the principal formula $A \,\&\, B$ can very well occur in Γ, whereas if there is no such occurrence, $\langle A \,\&\, B, \Gamma \rangle$ could be written as $A \,\&\, B, \Gamma$.

We have chosen to have two left rules for conjunction, instead of a single rule with the premiss $A, B, \Gamma \to C$, to avoid the complication that occurs if A and B are identical. As mentioned, this case would make the premiss collapse into $A, \Gamma \to C$ with no match with the conclusion.

Theorem 1. Cut elimination for LIS. *The rule of cut can be eliminated from derivations in* **LIS**.

Proof. We show that uppermost occurrences of rule *Cut* can be permuted upwards until they hit initial sequents and get removed. The *height of a cut* is the sum of the heights of the derivations of its premisses, i.e., of the longest branches in each. In each case, a cut is replaced by cuts that can have the same cut formula but a lesser height or else they have a shorter cut formula. The cases are:

1. One premiss of cut is an initial sequent.

1.1. If it is of the form $C \to C$, the conclusion is identical to the other premiss and the cut deleted.

1.2. The left premiss of cut is an initial sequent of the form $\bot \to C$ and we have:

$$\frac{\bot \to D \quad D, \Gamma \to C}{\langle \bot, \Gamma \rangle \to C} Cut$$

The conclusion is obtained from the initial sequent $\perp \rightarrow C$ by repeated weakenings.

1.3. The right premiss of cut is an initial sequent of the form $\perp \rightarrow C$ and we have:

$$\frac{\Gamma \rightarrow \perp \quad \perp \rightarrow C}{\Gamma \rightarrow C}\,{\scriptstyle Cut}$$

The cut formula cannot be principal in the left premiss and cut is permuted up until a sequent of the form $\perp, \Delta \rightarrow \perp$ is reached:

$$\frac{\perp, \Delta \rightarrow \perp \quad \perp \rightarrow C}{\langle \perp, \Delta \rangle \rightarrow C}\,{\scriptstyle Cut}$$

The conclusion of that cut is obtained from the right premiss by weakenings.

2. The left premiss of cut is a conclusion of a left rule. Cut is permuted up until case 1 is met, or the left premiss has been concluded by a right rule:

3. There are five cases of right rules in the left premiss of cut.

3.1. The first is rule $R\&$ with the cut:

$$\frac{\dfrac{\Gamma \rightarrow A \quad \Delta \rightarrow B}{\langle \Gamma, \Delta \rangle \rightarrow A \,\&\, B}\,{\scriptstyle R\&} \qquad A \,\&\, B, \Theta \rightarrow C}{\langle \Gamma, \Delta, \Theta \rangle \rightarrow C}\,{\scriptstyle Cut}$$

Cut is permuted up at right and its height reduced until case 1 is met or $A \,\&\, B$ is the principal formula in rule $L\&$. In the latter case we have, say:

$$\frac{\dfrac{\Gamma \rightarrow A \quad \Delta \rightarrow B}{\langle \Gamma, \Delta \rangle \rightarrow A \,\&\, B}\,{\scriptstyle R\&} \qquad \dfrac{A, \Theta \rightarrow C}{\langle A \,\&\, B, \Theta \rangle \rightarrow C}\,{\scriptstyle L\&}}{\langle \Gamma, \Delta, \Theta \rangle \rightarrow C}\,{\scriptstyle Cut}$$

If $A \,\&\, B$ does not occur in Θ, cut can be permuted as in [Gentzen, 1934–35], into:

$$\frac{\Gamma \rightarrow A \quad A, \Theta \rightarrow C}{\langle \Gamma, \Theta \rangle \rightarrow C}\,{\scriptstyle Cut}$$

The conclusion of the original cut is now obtained through weakenings.

If $A \,\&\, B$ instead does occur in Θ, rule $L\&$ produces an implicit contraction on the cut formula, and the Gentzen-style transformation leaves one copy of $A \,\&\, B$ in the conclusion: The cut after the transformation can be written with the notation $A \,\&\, B, \Theta' \equiv \Theta$ as:

$$\frac{\Gamma \rightarrow A \quad A, A \,\&\, B, \Theta' \rightarrow C}{\langle \Gamma, A \,\&\, B, \Theta' \rangle \rightarrow C}\,{\scriptstyle Cut}$$

This problematic case is resolved as follows:

$$\frac{\Gamma \rightarrow A \qquad \dfrac{\dfrac{\Gamma \rightarrow A \quad \Delta \rightarrow B}{\langle \Gamma, \Delta \rangle \rightarrow A \,\&\, B}\,{\scriptstyle R\&} \qquad A, A \,\&\, B, \Theta' \rightarrow C}{\langle A, \Gamma, \Delta, \Theta' \rangle \rightarrow C}\,{\scriptstyle Cut}}{\langle \Gamma, \Delta, \Theta' \rangle \rightarrow C}\,{\scriptstyle Cut}$$

The height of the upper cut is diminished by one. The new cut is on a shorter formula.

3.2. The case of rule $R\vee$ is similar.

3.3. With rule $R\supset$ we have as in 3.1 the worst case:

$$\dfrac{\dfrac{A,\Gamma_1 \to B}{\Gamma_1 \to A \supset B}\,{\scriptstyle R\supset} \qquad \dfrac{A \supset B,\Gamma_2 \to A \quad B, A \supset B, \Gamma_3 \to C}{\langle A \supset B,\Gamma_2,\Gamma_3\rangle \to C}\,{\scriptstyle L\supset}}{\langle \Gamma_1,\Gamma_2,\Gamma_3\rangle \to C}\,{\scriptstyle Cut}$$

The transformed derivation is, with the $\langle\ \rangle$ notation left out to make it fit:

$$\dfrac{\dfrac{\dfrac{\dfrac{A,\Gamma_1 \to B}{\Gamma_1 \to A \supset B}\,{\scriptstyle R\supset} \quad A \supset B,\Gamma_2 \to A}{\Gamma_1,\Gamma_2 \to A}\,{\scriptstyle Cut} \quad A,\Gamma_1 \to B}{\Gamma_1,\Gamma_1,\Gamma_2 \to B}\,{\scriptstyle Cut} \quad \dfrac{\dfrac{A,\Gamma_1 \to B}{\Gamma_1 \to A \supset B}\,{\scriptstyle R\supset} \quad B, A \supset B,\Gamma_3 \to C}{B,\Gamma_1,\Gamma_3 \to C}\,{\scriptstyle Cut}}{\Gamma_1,\Gamma_1,\Gamma_1,\Gamma_2,\Gamma_3 \to C}\,{\scriptstyle Cut}$$

Contractions are now made to reproduce the original conclusion. There are two cuts with the cut formula $A \supset B$ and a lesser cut height, followed by cuts on the shorter formulas A and B.

3.4. The cut formula is $\neg A$ and the transformation nearly a special case of the above.

3.5. The cut formula is $\forall x A(x)$ with the derivation:

$$\dfrac{\dfrac{\Gamma_1 \to A(y)}{\Gamma_1 \to \forall x A(x)}\,{\scriptstyle R\forall} \qquad \dfrac{A(t), \forall x A(x), \Gamma_2 \to C}{\forall x A(x), \Gamma_2 \to C}\,{\scriptstyle L\forall}}{\langle \Gamma_1,\Gamma_2\rangle \to C}\,{\scriptstyle Cut}$$

The transformed derivation is, with the substitution $[t/y]$ in the derivation of the premiss of rule $R\forall$ of the original derivation that gives as a result the derivable sequent $\Gamma_1 \to A(t)$:

$$\dfrac{\Gamma_1 \to A(t) \qquad \dfrac{\dfrac{\Gamma_1 \to A(y)}{\Gamma_1 \to \forall x A(x)}\,{\scriptstyle R\forall} \quad \forall x A(x), A(t), \Gamma_2 \to C}{\langle A(t),\Gamma_1,\Gamma_2\rangle \to C}\,{\scriptstyle Cut}}{\langle \Gamma_1,\Gamma_2\rangle \to C}\,{\scriptstyle Cut}$$

The upper cut has a lesser cut height, the lower a shorter cut formula.

3.6. The cut formula is $\exists x A(x)$. This case is dual to the previous.

4. One premiss of cut is derived by rule *Wk*. If it is the left premiss, cut can be permuted up, and the same with the right premiss except when the cut formula is principal in weakening. In this case the conclusion is obtained from the premiss of weakening without any cut. **QED**

The above proof of cut elimination goes through also for a calculus with more than one succedent formula, even if the details turn out somewhat intricate. Initial sequents of this classical sequent calculus with sets of formulas **LKS** have the form $C \to C$, and negation is treated as a primitive connective:

Logical rules of the calculus LKS:

$$\frac{A, \Gamma \rightarrow \Delta}{\langle A \,\&\, B, \Gamma \rangle \rightarrow \Delta} L\& \qquad \frac{B, \Gamma \rightarrow \Delta}{\langle A \,\&\, B, \Gamma \rangle \rightarrow \Delta} L\& \qquad \frac{\Gamma_1 \rightarrow \Delta_1, A \quad \Gamma_2 \rightarrow \Delta_2, B}{\langle \Gamma_1, \Gamma_2 \rangle \rightarrow \langle \Delta_1, \Delta_2, A \,\&\, B \rangle} R\&$$

$$\frac{A, \Gamma_1 \rightarrow \Delta_1 \quad B, \Gamma_2 \rightarrow \Delta_2}{\langle A \vee B, \Gamma_1, \Gamma_2 \rangle \rightarrow \langle \Delta_1, \Delta_2 \rangle} L\vee \qquad \frac{\Gamma \rightarrow \Delta, A}{\Gamma \rightarrow \langle \Delta, A \vee B \rangle} R\vee \qquad \frac{\Gamma \rightarrow \Delta, B}{\Gamma \rightarrow \langle \Delta, A \vee B \rangle} R\vee$$

$$\frac{\Gamma_1 \rightarrow \Delta_1, A \quad B, \Gamma_2 \rightarrow \Delta_2}{\langle A \supset B, \Gamma_1, \Gamma_2 \rangle \rightarrow \langle \Delta_1, \Delta_2 \rangle} L\supset \qquad \frac{A, \Gamma \rightarrow \Delta, B}{\Gamma \rightarrow \langle \Delta, A \supset B \rangle} R\supset$$

$$\frac{\Gamma \rightarrow \Delta, A}{\langle \neg A, \Gamma \rangle \rightarrow \Delta} L\neg \qquad \frac{A, \Gamma \rightarrow \Delta}{\Gamma \rightarrow \langle \Delta, \neg A \rangle} R\neg$$

$$\frac{A(t), \Gamma \rightarrow \Delta}{\langle \forall x A(x), \Gamma \rangle \rightarrow \Delta} L\forall \qquad \frac{\Gamma \rightarrow \Delta, A(y)}{\Gamma \rightarrow \langle \Delta, \forall x A(x) \rangle} R\forall$$

$$\frac{A(y), \Gamma \rightarrow \Delta}{\langle \exists x A(x), \Gamma \rangle \rightarrow \Delta} L\exists \qquad \frac{\Gamma \rightarrow \Delta, A(t)}{\Gamma \rightarrow \langle \Delta, \exists x A(x) \rangle} R\exists$$

The **structural rules** are left and right weakening and cut:

$$\frac{\Gamma \rightarrow \Delta}{\langle A, \Gamma \rangle \rightarrow \Delta} LW \qquad \frac{\Gamma \rightarrow \Delta}{\Gamma \rightarrow \langle \Delta, A \rangle} RW \qquad \frac{\Gamma_1 \rightarrow \Delta_1, A \quad A, \Gamma_2 \rightarrow \Delta_2}{\langle \Gamma_1, \Gamma_2 \rangle \rightarrow \langle \Delta_1, \Delta_2 \rangle} Cut$$

Note that if the weakening formula is found in the context, the conclusion is identical to the premiss.

Theorem 2. Cut elimination for LKS. *The rule of cut can be eliminated from derivations in* **LKS**.

Proof. The proof is similar to that of theorem 1 in which uppermost instances of cut are eliminated. The cases are:

1. One premiss of cut is an initial sequent. The conclusion is identical to the other premiss and the cut is deleted.

2. Cut is permuted up in the left and right premisses until case 1 is met.

3. Cut is permuted up in the left and right premisses until the cut formula is principal in both premisses of cut. There are six cases:

3.1. The cut formula is $A \,\&\, B$, and we may assume that the principal formula in rules $L\&, R\&$ occurs in each premiss. Should it lack from some, either a simpler cut elimination procedure is made, but with more case distinctions, or, for a uniform procedure, possible missing principal formulas in premisses of $L\&, R\&$ are added through weakenings. Further, to fit the derivations on a page, we leave out the $\langle\ \rangle$ -notation:

$$\frac{\dfrac{\Gamma_1 \rightarrow \Delta_1, A \,\&\, B, A \quad \Gamma_2 \rightarrow \Delta_2, A \,\&\, B, B}{\Gamma_1, \Gamma_2 \rightarrow \Delta_1, \Delta_2, A \,\&\, B} R\& \qquad \dfrac{A, A \,\&\, B, \Gamma_3 \rightarrow \Delta_3}{A \,\&\, B, \Gamma_3 \rightarrow \Delta_3} L\&}{\Gamma_1, \Gamma_2, \Gamma_3 \rightarrow \Delta_1, \Delta_2, \Delta_3} Cut$$

The transformed derivation is:

$$\frac{\dfrac{\Gamma_1 \rightarrow \Delta_1, A \,\&\, B, A \qquad \dfrac{A, A \,\&\, B, \Gamma_3 \rightarrow \Delta_3}{A \,\&\, B, \Gamma_3 \rightarrow \Delta_3} L\&}{\Gamma_1, \Gamma_3 \rightarrow \Delta_1, \Delta_3, A} Cut \qquad \dfrac{\dfrac{\Gamma_1 \rightarrow \Delta_1, A \,\&\, B, A \quad \Gamma_2 \rightarrow \Delta_2, A \,\&\, B, B}{\Gamma_1, \Gamma_2 \rightarrow \Delta_1, \Delta_2, A \,\&\, B} R\& \qquad A, A \,\&\, B, \Gamma_3 \rightarrow \Delta_3}{A, \Gamma_1, \Gamma_2, \Gamma_3 \rightarrow \Delta_1, \Delta_2, \Delta_3} Cut}{\Gamma_1, \Gamma_2, \Gamma_3 \rightarrow \Delta_1, \Delta_2, \Delta_3} Cut$$

The two upper cuts have a lesser height of cut, the lower cut a shorter cut formula. The transformation is similar with the second form of rule $L\&$.

3.2. The cut formula is $A \vee B$. This case is dual to cut formula $A \,\&\, B$.

3.3. The cut formula is $A \supset B$ with the derivation:

$$\dfrac{\dfrac{A, \Gamma_1 \to \Delta_1, A \supset B, B}{\Gamma_1 \to \Delta_1, A \supset B}\,{\scriptstyle R\supset} \qquad \dfrac{A \supset B, \Gamma_2 \to \Delta_2, A \quad B, A \supset B, \Gamma_3 \to \Delta_3}{A \supset B, \Gamma_2, \Gamma_3 \to \Delta_2, \Delta_3}\,{\scriptstyle L\supset}}{\Gamma_1, \Gamma_2, \Gamma_3 \to \Delta_1, \Delta_2, \Delta_3}\,{\scriptstyle Cut}$$

The transformed derivation is too broad to be displayed as such. We shall show the active and principal formulas and indicate the contexts by numbers on top of the sequent arrows so that, for example, $A \xrightarrow{1} A \supset B, B$ stands for the sequent $A, \Gamma_1 \to \Delta_1, A \supset B, B$:

$$\dfrac{\dfrac{\dfrac{\dfrac{A \xrightarrow{1} A \supset B, B}{\xrightarrow{1} A \supset B}\,{\scriptstyle R\supset} \quad A \supset B \xrightarrow{2} A}{\xrightarrow{1,2} A}\,{\scriptstyle Cut} \quad \dfrac{A \xrightarrow{1} B, A \supset B \quad \dfrac{A \supset B \xrightarrow{2} A \quad B, A \supset B \xrightarrow{3}}{A \supset B \xrightarrow{2,3}}\,{\scriptstyle L\supset}}{A \xrightarrow{1,2,3} B}\,{\scriptstyle Cut}}{\xrightarrow{1,2,3} B}\,{\scriptstyle Cut} \quad \dfrac{\dfrac{A \xrightarrow{1} A \supset B, B}{\xrightarrow{1} A \supset B}\,{\scriptstyle R\supset} \quad B, A \supset B \xrightarrow{3}}{B \xrightarrow{1,3}}\,{\scriptstyle Cut}}{\xrightarrow{1,2,3}}\,{\scriptstyle Cut}$$

There are three cuts with the cut formula $A \supset B$ and a lesser cut height, followed by cuts on the shorter formulas A and B.

3.4. The cut formula is $\neg A$ and the transformation nearly a special case of the above.

3.5. The cut formula is $\forall x A(x)$ with the derivation:

$$\dfrac{\dfrac{\Gamma_1 \to \Delta_1, \forall x A(x), A(y)}{\Gamma_1 \to \Delta_1, \forall x A(x)}\,{\scriptstyle R\forall} \quad \dfrac{A(t), \forall x A(x), \Gamma_2 \to \Delta_2}{\forall x A(x), \Gamma_2 \to \Delta_2}\,{\scriptstyle L\forall}}{\Gamma_1, \Gamma_1, \Gamma_2, \Gamma_2 \to \Delta_1, \Delta_1, \Delta_2, \Delta_2}\,{\scriptstyle Cut}$$

The transformed derivation is, with the substitution $[t/y]$ in the derivation of the premiss of rule $R\forall$ of the original derivation that gives as a result the derivable sequent $\Gamma_1 \to \Delta_1, \forall x A(x), A(t)$:

$$\dfrac{\dfrac{\Gamma_1 \to \Delta_1, \forall x A(x), A(t) \quad \dfrac{A(t), \forall x A(x), \Gamma_2 \to \Delta_2}{\forall x A(x), \Gamma_2 \to \Delta_2}\,{\scriptstyle L\forall}}{\Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2, A(t)}\,{\scriptstyle Cut} \quad \dfrac{\dfrac{\Gamma_1 \to \Delta_1, \forall x A(x), A(y)}{\Gamma_1 \to \Delta_1, \forall x A(x)}\,{\scriptstyle R\forall} \quad \forall x A(x), A(t), \Gamma_2 \to \Delta_2}{A(t), \Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2}\,{\scriptstyle Cut}}{\Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2}\,{\scriptstyle Cut}$$

The two upper cuts have a lesser cut height, the last cut a shorter cut formula.

3.6. The cut formula is $\exists x A(x)$. This case is dual to the previous.

4. One premiss of cut has been derived by weakening. If the weakening formula is the cut formula, the weakening and cut are removed. Else cut is permuted above the weakening. **QED**

4. Proof of decidability of intuitionistic propositional logic

The first published application of sequent calculus was a proof of the decidability of intuitionistic propositional logic. Gentzen's argument here is purely verbal, without a single formula [Gentzen, 1934–35, IV Sec. 1]. A sequent is defined to be *reduced* if it has at most three copies of the same formula on either side. Given a derivation of a reduced sequent, all contexts are rewritten so that they have no multiple occurrences of formulas. As Gentzen notes, sequents that appear as conclusions of a rule and as premisses of a successive rule need after this modification not conform to the form of the rules as defined earlier in the setting up of the calculus. He adds, however, that a syntactically correct derivation can always be produced, by the structural rules of weakening, contraction, and exchange.

> *Gentzen's derivations reduced to single occurrences of each formula are precisely derivations in a sequent calculus with sets.*

His main observation is that the occurrences of formulas on either side of a sequent in a derivation in his calculi **LI** and **LK**, the rule of cut included, can be limited to at most three, provided this holds for the sequent to be derived. It is easy to see where this number comes from: In Gentzen's rule $L\supset$, we can have one occurrence of $A \supset B$ in both premisses, with the condition of a single occurrence in each context respected. With $\Gamma \equiv A \supset B, \Gamma'$ and $\Delta \equiv A \supset B, \Delta'$, respectively, we have:

$$\frac{A \supset B, \Gamma' \to A \quad B, A \supset B, \Delta' \to C}{A \supset B, A \supset B, A \supset B, \Gamma', \Delta' \to C}{\scriptstyle L\supset}$$

Next the conclusion is rewritten as $A \supset B, \Gamma', \Delta' \to C$, a sequent that can be equally well obtained from the conclusion of $L\supset$ by two contractions on $A \supset B$. Thus, the result is that all derivations of a reduced sequent can be transformed into a reduced form, with at most three copies of a formula in each antecedent or succedent of a sequent and, as we may add, with at most two successive contractions on the same formula. Gentzen's other two-premiss rules $R\&$ and $L\vee$ have shared contexts and behave like one-premiss rules in this respect, with at most two copies of the same formula in the conclusion and at most one contraction. Cut has no principal formula and can produce at most duplications.

5. Cut elimination without the mix rule

The observations at the end of the previous Section can be turned into a cut elimination procedure for Gentzen's calculi **LI** and **LK**, with no need for the mix rule trick. Sequents will have multisets as antecedents and succedents, with logical rules as in the above tables but without the bra-ket notation, and the insubstantial difference that we have independent contexts in rules $R\&$ and $L\vee$. Gentzen has shared contexts in these rules, which takes away one step in cut elimination. His rule $L\supset$ is as above, with independent contexts for the reason that that is the only way to arrive at a single-succedent instance of the rule; the proof of cut elimination was designed so that a proof for the intuitionistic calculus **LI** came out as a single-succedent special case.

To the structural rules are added left and right contraction:

$$\frac{A, A, \Gamma \to \Delta}{A, \Gamma \to \Delta}{\scriptstyle LC} \qquad \frac{\Gamma \to \Delta, A, A}{\Gamma \to \Delta, A}{\scriptstyle RC}$$

Theorem 3. Cut elimination for LI. *The rule of cut can be eliminated from derivations in* **LI**.

Proof. By the above, a given derivation with cuts can be transformed into one in which there are at most two successive contractions on the principal formula with rule $L\supset$ and one with the rest.

All cases except when the right premiss of cut has been derived by contraction have been covered in the proof of theorem 1. We show the most involved case of rule $L\supset$. If between the rule and the cut there are other rules than the two contractions, they can be permuted, so we have the derivation:

$$\frac{\Gamma_1 \to A \supset B \qquad \dfrac{\dfrac{\dfrac{A \supset B, \Gamma_2 \to A \quad B, A \supset B, \Gamma_3 \to C}{A \supset B, A \supset B, A \supset B, \Gamma_2, \Gamma_3 \to C}{\scriptstyle L\supset}}{A \supset B, A \supset B, \Gamma_2, \Gamma_3 \to C}{\scriptstyle LC}}{A \supset B, \Gamma_2, \Gamma_3 \to C}{\scriptstyle LC}}{\Gamma_1, \Gamma_2, \Gamma_3 \to C}{\scriptstyle Cut}$$

Cut is permuted up at left until the cut formula is principal, concluded by rule $R\supset$, and the transformation of case 3.3 in theorem 1 applies, and similarly for all the other cases. **QED**

As Gentzen notes, the notion of a reduced sequent applies as well to the symmetric classical calculus **LK**. Therefore theorem 2 and its proof turns equally well into a proof for **LK**:

Theorem 4. Cut elimination for LK. *The rule of cut can be eliminated from derivations in* **LK**.

Proof. Similar to the previous. For reasons of comparison in Sec. 7, we show the case of cut formula $\forall x A(x)$ and assume the worst case, with the principal formula found already in the premisses of rules $R\forall, L\forall$. After suitable permutations, we have single contractions right after these logical rules, followed by a cut:

$$\frac{\dfrac{\dfrac{\Gamma_1 \to \Theta_1, \forall x A(x), A(y)}{\Gamma_1 \to \Theta_1, \forall x A(x), \forall x A(x)}{\scriptstyle R\forall}}{\Gamma_1 \to \Theta_1, \forall x A(x)}{\scriptstyle RC} \qquad \dfrac{\dfrac{A(t), \forall x A(x), \Gamma_2 \to \Theta_2}{\forall x A(x), \forall x A(x), \Gamma_2 \to \Theta_2}{\scriptstyle L\forall}}{\forall x A(x), \Gamma_2 \to \Theta_2}{\scriptstyle LC}}{\Gamma_1, \Gamma_2 \to \Theta_1, \Theta_2}{\scriptstyle Cut}$$

The premiss of $R\forall$ has the eigenvariable y that can be changed into the term t throughout the derivation of the premiss, after which the transformation is into:

$$\dfrac{\Gamma_1 \to \Theta_1, \forall x A(x), A(t) \qquad \dfrac{\dfrac{A(t), \forall x A(x), \Gamma_2 \to \Theta_2}{\forall x A(x), \forall x A(x), \Gamma_2 \to \Theta_2}{\scriptstyle L\forall}}{\forall x A(x), \Gamma_2 \to \Theta_2}{\scriptstyle LC}}{\Gamma_1, \Gamma_2 \to \Theta_1, \Theta_2, A(t)}{\scriptstyle Cut} \qquad \dfrac{\dfrac{\dfrac{\Gamma_1 \to \Theta_1, \forall x A(x), A(t)}{\Gamma_1 \to \Theta_1, \forall x A(x), \forall x A(x)}{\scriptstyle R\forall}}{\Gamma_1 \to \Theta_1, \forall x A(x)}{\scriptstyle RC} \qquad A(t), \forall x A(x), \Gamma_2 \to \Theta_2}{A(t), \Gamma_1, \Gamma_2 \to \Theta_1, \Theta_2}{\scriptstyle Cut}$$

$$\overline{\Gamma_1, \Gamma_1, \Gamma_2, \Gamma_2 \to \Theta_1, \Theta_1, \Theta_2, \Theta_2}\ {\scriptstyle Cut}$$

A sufficient number of contractions gives the original conclusion of cut. **QED**

Another case in the proof is covered in the discussion in Sec. 7. It might seem at a first sight that nothing is gained by the above transformation, for the contractions are still there. However, in the left branch, the left premiss of the upper cut is not derived by right contraction but by some other rule, and similarly for the right branch. The elimination of these upper cuts proceeds now by an analysis of how the respective left and right premisses were derived.

6. The notion of an altitude line

The method of cut elimination for **LKS** above, in theorem 2, is quite similar to one connected with the intricate "altitude line" construction (*Höhenlinie*) found in Gentzen's 1938 proof of the consistency of Peano arithmetic [Gentzen, 1938], the origin of which has been wondered by many. It is quite plausible, in the light of how cut elimination is adapted to a calculus with an explicit rule of contraction, theorem 4, that Gentzen had done his original cut elimination theorem of 1933 along the lines given here, then changed for the simpler mix rule proof for expository reasons. As we shall point out, the mix rule cannot be used in the proof theory of arithmetic, so five years later, the original methods of cut elimination resurfaced in his consistency proof.

After the new consistency proof of 1938 that used the classical calculus **LK**, Gentzen worked hard with the proof theory of intuitionistic arithmetic and analysis. In the Summer of 1944, he prepared a summary of calculi and reductions in consistency proofs, the extant shorthand series **WKRd** that is explained in some detail in [von Plato, 2012, pp. 356–358] and in detail in the introduction to the first volume of Gentzen's shorthand notes, [Gentzen, 2016]. The central problem was always the multiplication of formulas in steps of reduction, and one suggested remedy in **WKRd** was a "contractive cut" (*Zusammenziehungsschnitt*). None of Gentzen's attempts at keeping contraction a business separate from the logical rules worked, though. Above we used already Gentzen's notation, here given directly from the edited shorthand manuscript **WKRd** [Gentzen, 1944, p. 5]:

$$\frac{\Gamma \to \mathfrak{D} \quad \mathfrak{D}\,\Delta \to \mathfrak{C}}{\langle \Gamma\,\Delta \rangle \to \mathfrak{C}}$$

Gentzen explains the $\langle\ \rangle$ notation by: "Meaning: the formulas $\Gamma\,\Delta$ contracted throughout." On p. 4, he uses a similar notation $[\Gamma\,\Delta]$ for lists without order, with the explanation: "[ ]: arbitrary order of the formulas in question."

The altitude line construction of 1938: Gentzen's 1938 'New formulation of the consistency proof for pure number theory' is in terms of the classical sequent calculus **LK**. In a letter to Bernays dated 12 May 1938, included in [Gentzen, 2016], he writes that the proof of consistency is organized similarly to the way in which he found the original proof of cut elimination in 1933. Inconsistency is expressed as the derivability of the empty sequent $\to$. Logical rules import a formula on one or the other side of a sequent. Therefore, if the empty sequent is derivable, there is in the derivation an *endpiece* in which only structural rules and steps of inductive inference are found. The uppermost sequents of the endpiece are conclusions of logical rules. Instances of the rule of induction in the endpiece become reduced into repeated cuts, after which contractions and cuts are used to arrive at the empty endsequent. A crucial lemma 3.4.3 in the proof states that there is at least one cut formula in the endpiece such that the cut formula is principal in a left, resp. a right logical rule that delimits the endpiece. Such a cut is permuted so that it applies right after the logical rules, by which the cut formula becomes the principal formula in both premisses of cut. Thus, there is a precise analogy to cut elimination for pure logic, but with a price:

In the proof that the reduction of derivations of the empty sequent through cut elimination terminates, Gentzen uses a strange proof transformation and the related notion of "altitude line." Here is his example of an altitude line, with the given part of a derivation [Gentzen, 1938, p. 34]:

$$
\begin{array}{c}
\dfrac{\Gamma_1 \to \Theta_1, F(a)}{\Gamma_1 \to \Theta_1, \forall x F(x)}\,R\forall \qquad \dfrac{F(n), \Gamma_2 \to \Theta_2}{\forall x F(x), \Gamma_2 \to \Theta_2}\,L\forall \\
\vdots \qquad\qquad\qquad\qquad \vdots \\
\dfrac{\Gamma \to \Theta, \forall x F(x) \qquad\qquad \forall x F(x), \Delta \to \Lambda}{\Gamma, \Delta \to \Theta, \Lambda}\,\mathit{Cut} \\
\vdots \\
\overline{\Gamma_3 \to \Theta_3}\,\mathit{Altitude\ line} \\
\vdots \\
\overline{\quad \to \quad}
\end{array}
$$

An *altitude line* in a derivation is any inference line of a cut with a formula such that all cut formulas from the line to the endsequent are shorter. Gentzen uses the word *Höhe* (altitude) even for the length of a cut formula that determines an altitude line, a notion not to be confused with other uses of English equivalents of *Höhe*, such as in the height of a cut in Section 3 or the height of a derivation.

The cut formula $\forall x F(x)$ of the example has a length ϱ that is also assumed to be an altitude; thus, the line of cut is an altitude line. At the indicated *Altitude line* lower down, this altitude diminishes from ϱ to some σ through another cut, so we have $\sigma < \varrho$.

The transformed derivation is, disregarding rules of exchange in the original [Gentzen, 1938, p. 35]:

$$
\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\Gamma_1 \rightarrow \Theta_1, F(n)}}{\Gamma_1 \rightarrow \Theta_1, F(n), \forall x F(x)}\,RW \quad \dfrac{\dfrac{\vdots}{F(n), \Gamma_2 \rightarrow \Theta_2}}{\forall x F(x), \Gamma_2 \rightarrow \Theta_2}\,L\forall}{\dfrac{\vdots}{\dfrac{\dfrac{\Gamma \rightarrow \Theta, F(n), \forall x F(x) \qquad \forall x F(x), \Delta \rightarrow \Lambda}{\Gamma, \Delta \rightarrow \Theta, \Lambda, F(n)}}{\vdots}}}
$$

$$
\dfrac{\dfrac{\dfrac{\vdots}{\Gamma_1 \rightarrow \Theta_1, F(a)}}{\Gamma_1 \rightarrow \Theta_1, \forall x F(x)}\,R\forall \quad \dfrac{\dfrac{\vdots}{F(n), \Gamma_2 \rightarrow \Theta_2}}{\forall x F(x), F(n), \Gamma_2 \rightarrow \Theta_2}\,LW}{\dfrac{\vdots}{\dfrac{\dfrac{\Gamma \rightarrow \Theta, \forall x F(x) \qquad \forall x F(x), F(n), \Delta \rightarrow \Lambda}{F(n), \Gamma, \Delta \rightarrow \Theta, \Lambda}}{\vdots}}}
$$

$$
\dfrac{\dfrac{\dfrac{\Gamma_3 \rightarrow \Theta_3, F(n) \qquad F(n), \Gamma_3 \rightarrow \Theta_3}{\Gamma_3, \Gamma_3 \rightarrow \Theta_3, \Theta_3}\,\textit{New Cut}}{\Gamma_3 \rightarrow \Theta_3}\,\textit{Contractions}}{\dfrac{\vdots}{\rightarrow}}
$$

Gentzen indicates only in the text but not in the figure that altitude lines with a lowered altitude $< \varrho$ appear higher up in the transformed derivation, above the sequents $\Gamma_3 \rightarrow \Theta_3, F(n)$ and $F(n), \Gamma_3 \rightarrow \Theta_3$. On the whole, the idea is that altitude lines are being pushed up in the endpiece. The empty endsequent has altitude 0, and the transformations force this to be the case for the whole endpiece, by which there is no derivation of the empty sequent.

When explaining the transformation, Gentzen writes (p. 34):

> The inference figures represent an introduction and an elimination of $\forall$ in $\forall x F(x)$. Following the original basic idea, both should be put aside and $\forall x F(x)$ should be substituted by $F(n)$ – its grade [length] is one less; in place of the cut with the cut formula $\forall x F(x)$ there would occur a cut with the cut formula $F(n)$. There appears, however, the difficulty already mentioned, namely that the formula $\forall x F(x)$ can have been used and even introduced in several places.

To resolve the difficulty, the mix rule cannot be used, as will be made clear in our closing paragraph. Instead, cuts on the formula $\forall x F(x)$ are maintained, but Gentzen notes that the essential point of the transformation is that above the two cuts on the old cut formula, the inferences have been simplified because one step of logical inference has been removed in each. This, precisely, is what happened in the transformations in our proofs of cut elimination for **LIS** and **LKS**, theorems 1 and 2 of Section 3. As to why the steps of weakening are added in which the formula $\forall x F(x)$ reappears, Gentzen notes that "this is a matter of convenience, for one would have to count anyway with its appearance further down, so in this way the new form of the derivation is taken over from the old one in the most convenient way."

7. Contraction and altitude lines

As Gentzen notes, copies of the formula $\forall x F(x)$ can appear through branches above the step of the original cut other than the branch shown. (In fact, Gentzen's figure has three-pronged upward-pointing "forks" that we failed to reproduce.) These formula occurrences would have to be contracted before the cut. Assume now that instead of a weakening, the extra copy of the cut formula was already there in the premisses of the logical rules of the original derivation to be transformed, and vanished right after the step in which it became principal, in a step of contraction. As noted above in Sec. 4, with a one-premiss rule, there need be at most one such contraction present:

$$
\dfrac{
\begin{array}{c}
\dfrac{\dfrac{\Gamma_1 \to \Theta_1, \forall xF(x), F(a)}{\Gamma_1 \to \Theta_1, \forall xF(x), \forall xF(x)}{\scriptstyle R\forall}}{\Gamma_1 \to \Theta_1, \forall xF(x)}{\scriptstyle RC} \\ \vdots \\ \Gamma \to \Theta, \forall xF(x)
\end{array}
\qquad
\begin{array}{c}
\dfrac{\dfrac{F(n), \forall xF(x), \Gamma_2 \to \Theta_2}{\forall xF(x), \forall xF(x), \Gamma_2 \to \Theta_2}{\scriptstyle L\forall}}{\forall xF(x), \Gamma_2 \to \Theta_2}{\scriptstyle LC} \\ \vdots \\ \forall xF(x), \Delta \to \Lambda
\end{array}
}{
\begin{array}{c}
\Gamma, \Delta \to \Theta, \Lambda \\ \vdots \\ \overline{\Gamma_3 \to \Theta_3}\,{\scriptstyle Altitude\ line} \\ \vdots \\ \overline{\quad \to \quad}
\end{array}
}{\scriptstyle Cut}
$$

Permuting *Cut* up, the essential case to consider is:

$$
\dfrac{
\dfrac{\dfrac{\Gamma_1 \to \Theta_1, \forall xF(x), F(a)}{\Gamma_1 \to \Theta_1, \forall xF(x), \forall xF(x)}{\scriptstyle R\forall}}{\Gamma_1 \to \Theta_1, \forall xF(x)}{\scriptstyle RC}
\qquad
\dfrac{\dfrac{F(n), \forall xF(x), \Gamma_2 \to \Theta_2}{\forall xF(x), \forall xF(x), \Gamma_2 \to \Theta_2}{\scriptstyle L\forall}}{\forall xF(x), \Gamma_2 \to \Theta_2}{\scriptstyle LC}
}{
\begin{array}{c}
\Gamma_1, \Gamma_2 \to \Theta_1, \Theta_2 \\ \vdots \\ \Gamma, \Gamma_2 \to \Theta, \Theta_2 \\ \vdots \\ \Gamma, \Delta \to \Theta, \Lambda \\ \vdots \\ \overline{\Gamma_3 \to \Theta_3} \\ \vdots \\ \overline{\quad \to \quad}
\end{array}
}{\scriptstyle Cut}
$$

After the cut, there follow the steps that led originally from $\Gamma_1 \to \Theta_1, \forall xF(x)$ to $\Gamma \to \Theta, \forall xF(x)$, with $\forall xF(x)$ kept intact, then the steps that led originally from $\forall xF(x), \Gamma_2 \to \Theta_2$ to $\forall xF(x), \Delta \to \Lambda$. This derivation is transformed following strictly Gentzen's general procedure for the specific case of contractions. We have the substitution $[n/a]$ in the derivation of the sequent $\Gamma_1 \to \Theta_1, \forall xF(x), F(a)$ and just one step of logical inference instead of two above each of the cuts on $\forall xF(x)$:

$$
\dfrac{\dfrac{\Gamma_1 \to \Theta_1, \forall x F(x), F(n) \qquad \dfrac{\dfrac{F(n), \forall x F(x), \Gamma_2 \to \Theta_2}{\forall x F(x), \forall x F(x), \Gamma_2 \to \Theta_2}{\scriptstyle L\forall}}{\forall x F(x), \Gamma_2 \to \Theta_2}{\scriptstyle LC}}{\Gamma_1, \Gamma_2 \to \Theta_1, \Theta_2, F(n)}{\scriptstyle Cut} \qquad \dfrac{\dfrac{\dfrac{\Gamma_1 \to \Theta_1, \forall x F(x), F(a)}{\Gamma_1 \to \Theta_1, \forall x F(x), \forall x F(x)}{\scriptstyle R\forall}}{\Gamma_1 \to \Theta_1, \forall x F(x)}{\scriptstyle RC} \qquad F(n), \forall x F(x), \Gamma_2 \to \Theta_2}{F(n), \Gamma_1, \Gamma_2 \to \Theta_1, \Theta_2}{\scriptstyle Cut}}{\Gamma_1, \Gamma_2 \to \Theta_1, \Theta_2}{\scriptstyle Cut}
$$

$$
\begin{array}{c}
\vdots \\
\Gamma, \Gamma_2 \to \Theta, \Theta_2 \\
\vdots \\
\Gamma, \Delta \to \Theta, \Lambda \\
\vdots \\
\hline
\Gamma_3 \to \Theta_3 \\
\vdots \\
\hline
\to
\end{array}
$$

As in the proof of theorem 4, the two contractions in the derivation of the premisses of the upper cuts remain, but the other premiss has in both cases a reduced derivation.

In the proof of theorem 2, cut elimination for **LKS**, we assumed the principal formula to occur in each premiss. Should it be lacking from some, either a simpler cut elimination procedure is made, but with more case distinctions as in the proof of theorem 1, or, for a uniform procedure as in our proof of theorem 2, possible missing principal formulas in premisses are added through weakenings, just as in Gentzen's 1938 proof.

If contraction is an explicit rule, we have now precisely the problematic case of cut elimination in the calculus **LK**, namely the one in which the cut formula has been contracted and that led Gentzen to introduce the mix rule idea [Gentzen, 1934–35]:. For a second example, consider a cut on $A \supset B$ with the "worst case" of $A \supset B$ repeated in each of the three premisses, as in the proof of theorem 2 (case 3.3), and with contractions before the instance of *Cut*. Even this case would be one in a detailed proof of theorem 4:

$$
\dfrac{\dfrac{\dfrac{A, \Gamma_1 \to \Delta_1, A \supset B, B}{\Gamma_1 \to \Delta_1, A \supset B, A \supset B}{\scriptstyle R\supset}}{\Gamma_1 \to \Delta_1, A \supset B}{\scriptstyle RC} \qquad \dfrac{\dfrac{\dfrac{A \supset B, \Gamma_2 \to \Delta_2, A \quad B, A \supset B, \Gamma_3 \to \Delta_3}{A \supset B, A \supset B, A \supset B, \Gamma_2, \Gamma_3 \to \Delta_2, \Delta_3}{\scriptstyle L\supset}}{A \supset B, A \supset B, \Gamma_2, \Gamma_3 \to \Delta_2, \Delta_3}{\scriptstyle LC}}{A \supset B, \Gamma_2, \Gamma_3 \to \Delta_2, \Delta_3}{\scriptstyle LC}}{\Gamma_1, \Gamma_2, \Gamma_3 \to \Delta_1, \Delta_2, \Delta_3}{\scriptstyle Cut}
$$

The contracted premisses of *Cut* cannot be resolved by simply cutting several times, say twice in the premiss of *RC*, as in (with duplications of the contexts omitted):

$$
\dfrac{\dfrac{\dfrac{A, \Gamma_1 \to \Delta_1, A \supset B, B}{\Gamma_1 \to \Delta_1, A \supset B, A \supset B}{\scriptstyle R\supset} \qquad \begin{array}{c}\vdots \\ A \supset B, \Gamma_2, \Gamma_3 \to \Delta_2, \Delta_3\end{array}}{\Gamma_1, \Gamma_2, \Gamma_3 \to \Delta_1, \Delta_2, \Delta_3, A \supset B}{\scriptstyle Cut} \qquad \begin{array}{c}\vdots \\ A \supset B, \Gamma_2, \Gamma_3 \to \Delta_2, \Delta_3\end{array}}{\Gamma_1, \Gamma_2, \Gamma_3 \to \Delta_1, \Delta_2, \Delta_3}{\scriptstyle Cut}
$$

The "vertical dots" stand for a derivation of the right premiss of the original cut, and it is seen that the lower cut is not reduced, because the derivation of the left premiss has grown. Therefore the derivation with contracted premisses of cut has to be transformed analogously to case 3.3 of our theorem 2. There will be altogether five cuts in the new derivation, three of them on $A \supset B$ with the height of cut reduced in each, followed by two on the shorter formulas A and B. The transformation is displayed for typographical reasons in parts, with the first part:

$$\dfrac{\dfrac{\dfrac{A, \Gamma_1 \to \Delta_1, A \supset B, B}{\Gamma_1 \to \Delta_1, A \supset B, A \supset B}{\scriptstyle R\supset}}{\Gamma_1 \to \Delta_1, A \supset B}{\scriptstyle RC} \qquad A \supset B, \Gamma_2 \to \Delta_2, A}{\Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2, A}{\scriptstyle Cut}$$

The second part is:

$$\dfrac{A, \Gamma_1 \to \Delta_1, A \supset B, B \qquad \dfrac{\dfrac{\dfrac{A \supset B, \Gamma_2 \to \Delta_2, A \quad B, A \supset B, \Gamma_3 \to \Delta_3}{A \supset B, A \supset B, A \supset B, \Gamma_2, \Gamma_3 \to \Delta_2, \Delta_3}{\scriptstyle L\supset}}{A \supset B, A \supset B, \Gamma_2, \Gamma_3 \to \Delta_2, \Delta_3}{\scriptstyle LC}}{A \supset B, \Gamma_2, \Gamma_3 \to \Delta_2, \Delta_3}{\scriptstyle LC}}{A, \Gamma_1, \Gamma_2, \Gamma_3 \to \Delta_1, \Delta_2, \Delta_3, B}{\scriptstyle Cut}$$

The third part is:

$$\dfrac{\dfrac{\dfrac{A, \Gamma_1 \to \Delta_1, A \supset B, B}{\Gamma_1 \to \Delta_1, A \supset B, A \supset B}{\scriptstyle R\supset}}{\Gamma_1 \to \Delta_1, A \supset B}{\scriptstyle RC} \qquad B, A \supset B, \Gamma_3 \to \Delta_3}{B, \Gamma_1, \Gamma_3 \to \Delta_1, \Delta_3}{\scriptstyle Cut}$$

These parts are combined by two cuts as follows, with the duplications of the contexts omitted:

$$\dfrac{\dfrac{\overset{\vdots}{\Gamma_1, \Gamma_2 \to \Delta_1, \Delta_2, A} \qquad \overset{\vdots}{A, \Gamma_1, \Gamma_2, \Gamma_3 \to \Delta_1, \Delta_2, \Delta_3, B}}{\Gamma_1, \Gamma_2, \Gamma_3 \to \Delta_1, \Delta_2, \Delta_3, B}{\scriptstyle Cut} \qquad \overset{\vdots}{B, \Gamma_1, \Gamma_3 \to \Delta_1, \Delta_3}}{\Gamma_1, \Gamma_2, \Gamma_3 \to \Delta_1, \Delta_2, \Delta_3}{\scriptstyle Cut}$$

As can be seen from this transformation, from the one for the cut formula $\forall x F(x)$, as well as from the transformations displayed in the proof of theorem 2, the procedure is somewhat involved, even if in principle clear.

8. Why all the trouble?

The proof transformations we have found for sequent calculi with implicit contraction are not altogether unknown: In the contraction-free intuitionistic sequent calculus **G3i**, the principal formula $A \supset B$ of rule $L\supset$ is repeated in the antecedent of the left premiss, with the effect of an implicit contraction at that point. In the proof transformation, there is first a cut on $A \supset B$

with a reduced height of cut, followed by cuts on A and B (as in [Negri and von Plato, 2001, p. 40]).

We cited Gentzen's reason for the use of the mix rule in the published proof, namely to make the proof easier. What he writes clearly indicates that he had some previous proof at hand that was changed. The most likely candidate for such a proof is the one presented here. Gentzen had, as is now well known, written down a detailed proof of normalization for intuitionistic natural deduction in 1933 [Gentzen, 2008]. Somehow the level of complexity escaped out of hands with cut elimination for **LK**, in comparison to natural deduction, so for reasons of exposition, Gentzen took into use the mix rule.

Finally, one may wonder why Gentzen went to the trouble of introducing the cut rule and the altitude line construction in 1938, instead of the well-behaving mix rule of 1933. It might seem a routine matter to treat the proof theory of arithmetic with the sequents-as-sets idea. Instead, what happens is that the cut elimination procedures of Section 3 are useless, because two cuts cannot necessarily be permuted with each other when sets are used, contrary to a calculus with multisets or lists and explicit contractions. The same is true of Gentzen's mix rule, so here is the true reason for the intricacies of the 1938 paper: The essential step in the 1938 proof of consistency, namely the permutation of a "suitable cut" as given by Gentzen's lemma 3.4.3 to the upper limit of the endpiece, would get blocked. Were it possible to do the consistency proof with sets and thus no contractions in the endpiece, an ordinal assignment could be given that beats Gödel's theorem! This particular detail about "sequents with sets" was found out in 2005 and remained a puzzle until Michael Rathjen pointed out that two instances of the mix rule do not necessarily permute.

References

B. Cooper and J. van Leeuwen. *Alan Turing: His Work and Impact.* Elsevier, 2013.

G. Gentzen. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift*, 39:176–210, 1934–35.

G. Gentzen. Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften*, 4:19–44, 1938.

G. Gentzen. Review of Ono (1938). *Zentralblatt für Mathematik und ihre Grenzgebiete*, 19:242, 1939.

G. Gentzen. **WKRd** (Widerspruchsfreiheitsbeweis: Kalküle und Reduktionen). Shorthand notes dated VIII.44, 8 pages. Available at the Philosophisches Archiv der Universität Konstanz. 1944.

G. Gentzen. The normalization of derivations. *The Bulletin of Symbolic Logic*, 14:245–257, 2008.

G. Gentzen. *Saved from the Cellar: Gerhard Gentzen's Shorthand Notes on Logic and Foundations of Mathematics*. With an Introduction and English Translation by Jan von Plato, 2016.

R. Hakli and S. Negri. Does the deduction theorem fail for modal logic? *Synthese*, 187:849–867, 2012.

D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik*. Springer, 1928.

S. Negri and J. von Plato. *Structural Proof Theory*. Cambridge University Press, 2001.

K. Ono. Logische Untersuchungen über die Grundlagen der Mathematik. *Journal of the Faculty of Science, Imperial University of Tokyo*, 3:329–389, 1938.

J. von Plato. A proof of Gentzen's *hauptsatz* without multicut. *Archive for Mathematical Logic*, 40:9–18, 2001.

J. von Plato. Gentzen's proof systems: byproducts in a work of genius. *The Bulletin of Symbolic Logic*, 18:313–367, 2012.

J. von Plato. *Elements of Logical Reasoning*. Cambridge University Press, 2013.

A. Turing. *The reform of mathematical notation and phraseology*. Manuscript in the Turing Archives, 1944.