The concept of proof

Matthias Baaz
Vienna University of Technology

1903

David Hilbert, Grundlagen der Geometrie

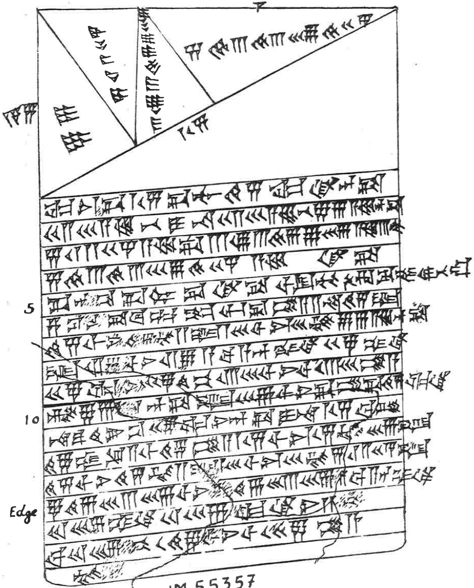
$\Downarrow$

the axiomatic method

the Hilbertian revolution

A proof is a sequence of formulas $A_1 \ldots A_n$ such that for all i

- A_i is an instance of an axiom or
- A_i follows from $A_{j_1} \ldots A_{j_k}$ by application of a rule $(j_1, \ldots, j_k < i)$.
- A_n is the result of the proof.



IM.55357

Euler became famous by deriving

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \tag{1}$$

Let us consider Eulers reasoning. Consider the polynomial of even degree

$$b_0 - b_1x^2 + b_2x^4 - \ldots + (-1)^n b_n x^{2n} \tag{2}$$

If it has the $2n$ roots $\pm\beta_1, \ldots, \beta_n \neq 0$ then (2) can be written as

$$b_0 \left(1 - \frac{x^2}{\beta_1^2}\right)\left(1 - \frac{x^2}{\beta_2^2}\right)\ldots\left(1 - \frac{x^2}{\beta_n^2}\right) \tag{3}$$

By comparing coefficients in(2) and (3) one obtains that

$$b_1 = b_0\left(\frac{1}{\beta_1^2} + \frac{1}{\beta_2^2} + \ldots + \frac{1}{\beta_n^2}\right). \tag{4}$$

Next Euler considers the Taylor series

$$\frac{\sin x}{x} = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n+1)!} \tag{5}$$

which has as roots $\pm\pi, \pm 2\pi, \pm 3\pi, \ldots$ Now by way of analogy Euler assumes that the infinite degree polynomial (5) behaves in the same way as the finite polynomial (2). Hence in analogy to (3) he obtains

$$\frac{\sin x}{x} = \left(1 - \frac{x^2}{\pi^2}\right)\left(1 - \frac{x^2}{4\pi^2}\right)\left(1 - \frac{x^2}{9\pi^2}\right)\ldots \tag{6}$$

and in analogy to (4) he obtains

$$\frac{1}{3!} = \left(\frac{1}{\pi^2} + \frac{1}{4\pi^2} + \frac{1}{9\pi^2} + \ldots\right) \tag{7}$$

which immediately gives

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} \tag{8}$$

The structure of Euler's argument is the following

(a) $(2) = (3)$	(mathematically derivable)
(b) $(2) = (3) \supset (4)$	(mathematically derivable)
(c) $(2) = (3) \supset (5) = (6)$	(analogical hypothesis)
(d) $(5) = (6) \supset (4)$	(modus ponens)
(e) $((2) = (3) \supset (4)) \supset ((5) = (6) \supset (7))$	(analogical hypothesis)
(f) $(5) = (6) \supset (7)$	(modus ponens)
(g) (7)	(modus ponens)
(h) $(7) \supset (1)$	(mathematically derivable)
(a) (1)	(modus ponens)

Petitio principii - circular reasoning

$$A$$

$$\vdots$$

$$B$$

$$A \leftrightarrow B$$

Ravans are black, therefore white birds are no ravens.

Fermat's méthode descente

Π : 2 is square of a rational

$$\sqrt{2} = \frac{p}{q} \Leftrightarrow \Pi$$

$$\Downarrow$$

$$2q^2 = p^2$$

$$\Downarrow$$

$$2q^2 = (2p_0)^2$$

$$\Downarrow$$

$$q^2 = 2p_0^2$$

$$\Downarrow$$

$$(2q_0)^2 = 2p_0^2$$

$$\Downarrow$$

$$2q_0^2 = p_0^2$$

$$\Downarrow$$

$$\Pi$$

$q_0 + p_0 < q + p$! (cf CLKID$^\omega$ by Brotherson and Simpson)

(1) That Kurt Gödel is Austrian entails that Kurt Gödel is Austrian.
(2) Hence, that Kurt Gödel is Austrian entails that everyone is Austrian.
(3) That is, if Kurt Gödel is Austrian, then all people are Austrian.
(4) Therefore, there exists a person such that, if that person is Austrian, then all people are Austrian.

$$\dfrac{\dfrac{\dfrac{A(a) \vdash A(a)}{A(a) \vdash \forall y A(y)}}{\vdash A(a) \to \forall y A(y)}}{\vdash \exists x (A(x) \to \forall y A(y))}$$

The traditional way to ensure soundness

- Inferences are sound, i.e. only true conclusions result from true premises.
- Derivations are hereditary, i.e. initial segments of proofs are proofs themselves.

Weak regularity

$$\dfrac{\dfrac{\dfrac{\dfrac{A(a) \vdash A(a)}{A(a) \vdash \forall x A(x)} \qquad \dfrac{A(f(a)) \vdash A(f(a))}{\forall x A(x) \vdash A(f(a))}}{A(a) \vdash A(f(a))}}{\vdash A(a) \rightarrow A(f(a))}}{\vdash \exists x (A(x) \rightarrow A(f(x)))}$$

Side variables

b is a side variable of a in π (written $a <_\pi b$) if π contains a strong-quantifier inference of the form

$$\frac{\Gamma \vdash \Delta, A(a, b, \vec{c})}{\Gamma \vdash \Delta, \forall x A(x, b, \vec{c})}$$

or of the form

$$\frac{A(a, b, \vec{c}), \Gamma \vdash \Delta}{\exists x A(x, b, \vec{c}), \Gamma \vdash \Delta}$$

Skolemization sk(A)

The Skolemization of a first-order formula is defined by replacing every strongly quantified variable y with a new function symbol $f_y(x_1, \ldots, x_n)$, where $x_1, \ldots, x_n$ are the weakly quantified variables such that Q_y appears in the scope of their quantifiers, and removing the quantifier Q_y.

Suitable quantifier inferences

A quantifier inference is suitable for a proof π if either it is a weak-quantifier inference, or the following three conditions are satisfied:

- (substitutability) the characteristic variable does not appear in the conclusion of π.
- (side-variable condition) the relation $<_\pi$ is acyclic.
- (weak regularity) the characteristic variable is not the characteristic variable of another strong-quantifier inference in π.

(LK^+)

The calculus LK^+ is defined like LK, except that we instead allow all weak and strong quantifier inferences with the proviso that they be suitable for the proof.

Weakly suitable quantifier inference

A quantifier inference is weakly suitable for a proof π if either it is a weak-quantifier inference or it satisfies substitutability, the side-variable condition, and

- (very weak regularity) whenever the characteristic variable is also the characteristic variable of another strong-quantifier inference in π, then it has the same critical formula.

LK^{++}

The calculus LK^{++} is the extension of LK^{+} that results from allowing all weakly suitable quantifier inferences.

Soundness

Theorem.
If a sequent is LK^{++}-derivable, then it is already LK-derivable.

Proof. Let π be an LK^{++}-proof. Replace every unsound universal quantifier inference by an $\rightarrow L$ inference:

$$\frac{\Gamma \vdash \Delta, A(a) \qquad \forall x A(x) \vdash \forall x A(x)}{\Gamma, A(a) \rightarrow \forall x A(x) \vdash \Delta, \forall x A(x)}$$

Similarly replace every unsound existential quantifier by an $\rightarrow L$ inference

$$\frac{\exists x A(x) \vdash \exists x A(x) \qquad A(a), \Gamma \vdash \Delta}{\Gamma, \exists x A(x), \exists x A(x) \rightarrow A(a) \vdash \Delta}$$

By doing this, we obtain a proof of the desired sequent, together with many formulae of the form $A(a) \rightarrow \forall x A(x)$ or $\exists x A(x) \rightarrow A(a)$ on the left-hand side. Introduce existential quantifiers left. This is sound in LK by properties of $<_\pi$.

Corollary.

If a sequent is derivable in LK^+ or LK^{++}, then it is already derivable in LK.

$$\dfrac{\dfrac{\dfrac{\dfrac{A(a,b) \vdash A(a,b)}{A(a,b) \vdash \forall y A(a,y)}}{A(a,b) \vdash \exists x \forall y A(x,y)}}{\exists x A(x,b) \vdash \exists x \forall y A(x,y)}}{\forall y \exists x A(x,y) \vdash \exists x \forall y A(x,y)}$$

$$a <_\pi b \quad b <_\pi a \ !$$

LK

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A(a) \vdash A(a)}{A(a) \vdash A(a), B}}{\vdash A(a), A(a) \to B}}{\vdash A(a), \exists x\,(A(x) \to B)}}{\vdash \exists x\,(A(x) \to B), A(a)}}{\vdash \exists x\,(A(x) \to B), \forall x\, A(x)} \qquad B \vdash B}{\forall x\, A(x) \to B \vdash \exists x\,(A(x) \to B), B}}{\forall x\, A(x) \to B, A(b) \vdash \exists x\,(A(x) \to B), B}}{\forall x\, A(x) \to B \vdash \exists x\,(A(x) \to B), A(b) \to B}}{\forall x\, A(x) \to B \vdash \exists x\,(A(x) \to B), \exists x\,(A(x) \to B)}}{\forall x\, A(x) \to B \vdash \exists x\,(A(x) \to B)}
$$

LK$^+$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{A(a) \vdash A(a)}{A(a) \vdash \forall x\, A(x)} \qquad B \vdash B}{A(a), \forall x\, A(x) \to B \vdash B}}{\forall x\, A(x) \to B, A(a) \vdash B}}{\forall x\, A(x) \to B \vdash A(a) \to B}}{\forall x\, A(x) \to B \vdash \exists x\, (A(x) \to B)}$$

Theorem.
There is no elementary function bounding the length of the shortest cut-free LK-proof of a formula in terms of its shortest cut-free LK^{+}-proof.

An immediate consequence is the following:

Corollary.
There is no elementary function bounding the length of the shortest cut-free LK-proof of a formula in terms of its shortest cut-free LK^{++}-proof.

The calculus $\mathrm{LK}_{\mathrm{shift}}$ is obtained by extending LK with the following rules:

$$\frac{\Gamma, \kappa[Qx\, A \lhd B] \vdash \Delta}{\Gamma, \kappa[Q'x\,(A \lhd B)] \vdash \Delta} \qquad \frac{\Gamma, \kappa[A \lhd Qx\, B] \vdash \Delta}{\Gamma, \kappa[Q'x\,(A \lhd B)] \vdash \Delta}$$

$$\frac{\Gamma \vdash \Delta, \kappa[Qx\, A \lhd B]}{\Gamma \vdash \Delta, \kappa[Q'x\,(A \lhd B)]} \qquad \frac{\Gamma \vdash \Delta, \kappa[A \lhd Qx\, B]}{\Gamma \vdash \Delta, \kappa[Q'x\,(A \lhd B)]}$$

where $\kappa[\cdot]$ is a context, $\lhd \in \{\wedge, \vee, \rightarrow\}$ and $Q' = Q$, except if $\lhd$ is $\rightarrow$ and Q is taken from the antecedent, in which case Q' is opposite. We refer to these rules as *deep quantifier shifts*.

Proposition.

Cut-free LK^+ simulates cut-free LK_{shift} double-exponentially, i.e., every LK_{shift}-provable sequent is LK^+-provable and there is a double exponential function that bounds the length of the least cut-free LK^+-proof of a LK^+-provable sequent in terms of its least cut-free LK_{shift}-proof.

Theorem.
There is no elementary function bounding the length of the shortest cut-free LK-proof of a formula in terms of its shortest cut-free LK_{shift}-proof.

e.g. Statman's sequence $\{s_j\}_{j<\omega}$

1. the size of S_i is polynomial in i;
2. there is no bound on the size of their smallest cut-free LK-proofs that is elementary in i;
3. the size of these proofs (with cuts), however, is polynomially bounded in i;
4. all cut formulae are closed; we can also assume they are prenex by, e.g., Theorem 3.3 in [BaazLeitsch94][1].

[1] M. Baaz and A. Leitsch, On Skolemization and Proof Complexity, Fund. Inform., 20 (1994), 353-379.

Transform this proof in $\mathrm{LK}_{\mathrm{shift}}$

$$\frac{\Gamma \vdash \Delta, A \qquad A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta}$$

$$\Downarrow$$

$$\frac{\Gamma \vdash \Delta, A \qquad \Gamma, A \vdash \Delta}{\Gamma, A \to A \vdash \Delta} \to L$$

obtaining

$$A_0 \to A_0, \ldots, A_m \to A_m, \Gamma_i \vdash \Delta_i$$

cut-free.

And by $\mathrm{LK}_{\mathrm{shift}}$-rules cut-free

$$\forall x_0^0 \exists x_1^0 \ldots (\hat{A}_0 \to \hat{A}_0), \ldots, \forall x_0^m \exists x_1^m (\hat{A}_m \to \hat{A}_m), \Gamma_i \vdash \Delta_i.$$

Claim.

There is no elementary function bounding the size of the smallest cut-free LK-proofs of

$$\forall x_0^0 \exists x_1^0 \ldots (\hat{A}_0 \to \hat{A}_0), \ldots, \forall x_0^m \exists x_1^m (\hat{A}_m \to \hat{A}_m), \Gamma_i \vdash \Delta_i.$$

Skolemize, extract a Herbrand sequent and replace all Skolem terms stepwise by $f(t_1, \ldots, t_n) \to t_i$ sucht that the instances of

$$\forall x_0^0 \exists x_1^0 \ldots (\hat{A}_0 \to \hat{A}_0), \ldots, \forall x_0^m \exists x_1^m$$

Skolemized become of the form $c \to c$.

$\mathbf{LJ}^{+}$ and $\mathbf{LJ}^{++}$

Proposition

LJ^{+} and LJ^{++} do not admit cut elimination.

$$
\dfrac{\dfrac{A(a) \vdash A(a)}{A(a) \vdash \forall x A(x)} \qquad \dfrac{A(f(a)) \vdash A(f(a))}{\forall x A(x) \vdash A(f(a))}}{\dfrac{A(a) \vdash A(f(a))}{\dfrac{\vdash A(a) \to A(f(a))}{\vdash \exists x (A(x) \to A(f(x)))}}}
$$

Quantifier shifts not valid intuitionistically

1. $\forall x\,(A \vee B(x)) \vdash A \vee \forall x\,B(x)$;
2. $(\forall x\,A(x) \to B) \vdash \exists x\,(A(x) \to B)$;
3. $(A \to \exists x\,B(x)) \vdash \exists x\,(A \to B(x))$.

Proposition.

A sequent is provable in LJ^{++} if and only if it is provable in LJ with all quantifier shifts added as axioms.

No elementary Skolemization for cut-free LK^+ and LK^{++} proofs. (But quadratic Skolemization using additional cuts.)

No elementary extraction of Skolemized Herbrand disjunctions from cut-free LK^+ and LK^{++} proofs.

No Gentzen-style cut-elimination (as Gentzen-style cut-elimination would transform LJ^+ (LJ^{++}) proofs into cut-free LJ^+ (LJ^{++}) proofs).

Relation to the ε-calculus

$$\exists x A(x) \sim A(\varepsilon_x A(x))$$

$$\forall x A(x) \sim A(\varepsilon_x \neg A(x)) \sim A(\tau_x A(x))$$

LK_ε

$$\frac{\Gamma, A(t) \vdash \Delta}{\Gamma, A(\tau_x A(x)) \vdash \Delta}\ \tau$$

$$\frac{\Gamma \vdash \Delta, A(t)}{\Gamma \vdash \Delta, A(\varepsilon_x A(x))}\ \varepsilon$$

Relation to the ε-calculus

Another soundness proof for LK^{+} and LK^{++}

But e.g.

$$\dfrac{\dfrac{\dfrac{\begin{matrix}(\varphi)\\ \Gamma \vdash \Delta, A(s(t))\end{matrix}}{\Gamma \vdash \Delta, A(s(\varepsilon_x A(s(x))))}}{\Gamma' \vdash \Delta', A(s(\varepsilon_x A(s(x))))}}{\Gamma' \vdash \Delta', A(\varepsilon_x A(x))}$$

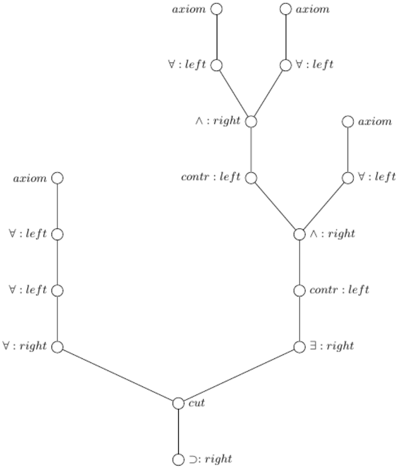
Not represented in LK^{+} and LK^{++}.

Corollary. For a filter in a compact space to converge it is necessary and sufficient that is has a single cluster point.

Necessity by §8 , no. 1, Proposition 1; sufficiency by Theorem 1 above.

Bourbaki, General Topology 1, page 85.

An LK-proof representation by name (aka skeleton, proof analysis) consists of a labelled tree where the inner nodes are labelled with rule names and the bottom node is labelled with the result. For exchange left and exchange right the occurrence of the exchanged formulas is denoted.



$\rightarrow \forall x \forall y P(x, y) \supset \exists z(P(0, z) \wedge P(z, a) \wedge P(Sz, Sa))$.

There is a proof underlying this representation by name iff a is replaced by $S^{2n}(0)$.

It is undecidable whether a proof representation by name corresponds to a proof

Let L be a set of function symbols, $a_1, \ldots, a_m$ variables. Let $\underline{T} = (T, Sub_1, \ldots, Sub_m)$ be the algebra of terms where T is the set of terms in L, $a_1, \ldots, a_m$ and for $i = 1, \ldots, m$

$$Sub_i(\delta, \sigma) := \delta(a_i/\sigma)$$

are substitutions as binary operations on T. A second order unification problem is a finite set of equations in the language $T \cup \{Sub_1, \ldots, Sub_m\}$ plus free variables for elements of T. The free variables will be called the term variables. By introducing new term variables we can transform any such system into an equivalent one where all equations have form

$$\delta(a_i/\sigma) = \rho,$$

where δ, σ, ρ are terms of term variables.

Theorem

Let L contain a unary function symbol S, a constant 0 and a binary function symbol. Let τ_0 be a term variable. Then for every recursively enumerable set $X \subseteq \omega$ there exists a second order unification problem Ω such that $\Omega \cup \{\tau_0 = S^n(0)\}$ has a solution iff $n \in X$.

Theorem

Let L be a language containing a unary function symbol S, a constant 0 and a binary function symbol. Then for every recursively enumerable set $X \subseteq \omega$ there exists a sequent $A \rightarrow A, P(a)$ and a proof description by name S such that $n \in X$ iff $A \rightarrow A, P(S^n(0))$ has an LK-proof with skeleton S.

Construct a derivation such that $P(a) \vee P(d), P(s) \vee P(t)$ occur on the right side enforced by the end-sequent. Quantify both formulas by ∃-right (one after the other). Afterwards infer ∃-left with eigenvariable a such that the position of a has to be bound on the right side. The two formulas can be constructed iff

$$d(a/s) = t.$$

Cut the description of the contracted formula F with the description of $F \rightarrow A \supset A$ directly obtained from an axiom by $\supset$: *right* and *weakening* : *left*.

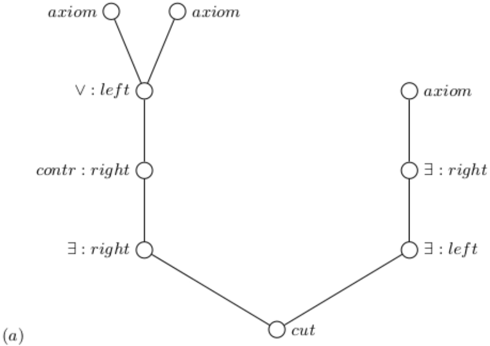
Corollary.

There is no recursive bound of the symbol complexity of an LK-proof in terms of the symbol complexity of its proof description by name.

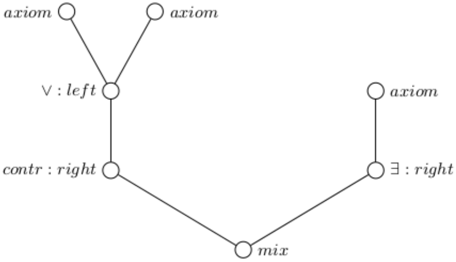
Theorem

There is a procedure, which transforms any proof representation by name P into a proof representation P' without reference of the cut-rule and with the same bottom node. If there is a proof realizing P there is a proof realizing P'.

Consider $P(c) \vee P(d) \to \exists x P(x)$ (all trees have this formula as bottom node):

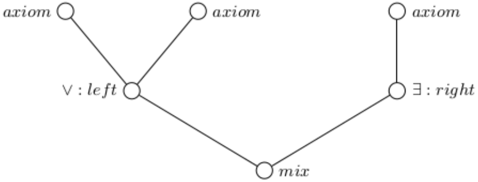


$\Downarrow$



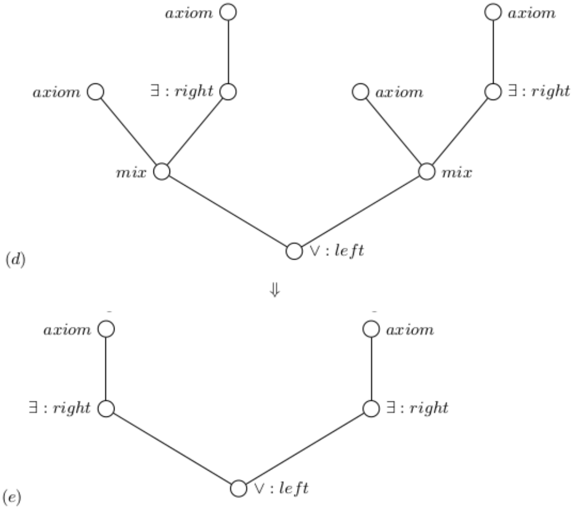
(b)

$\Downarrow$



(c)

$\Downarrow$



$(a) - (c)$ cannot be realized ($P(c)$ and $P(d)$ are forced to be contracted). (d) is realized by

$$\dfrac{P(c) \to P(c) \qquad \dfrac{P(c) \to P(c)}{P(c) \to \exists x P(x)}}{P(c) \to \exists x P(x)} \qquad \dfrac{P(d) \to P(d) \qquad \dfrac{P(d) \to P(d)}{P(d) \to \exists x P(x)}}{P(d) \to \exists x P(x)}$$

$$\dfrac{P(c) \to \exists x P(x) \qquad P(d) \to \exists x P(x)}{P(c) \vee P(d) \to \exists x P(x)}$$

(e) is realized by

$$\dfrac{\dfrac{P(c) \to P(c)}{P(c) \to \exists x P(x)} \qquad \dfrac{P(d) \to P(d)}{P(d) \to \exists x P(x)}}{P(c) \vee P(d) \to \exists x P(x)}$$

Theorem

Assume atomic axioms:
In a proof representation by name without reference to the cut-rule it can be determined whether there is a proof realizing the description.

It is a future challenge for proof theory to develop alternative proof formats to support the solution of specific mathematical problems.