Course on Proof Theory - Lecture 4

A proof of the cut-elimination theorem

Gianluca Curzi, Marianna Girlando

University of Birmingham

Midlands Graduate School
Nottingham, 10-14 April 2022

The rules of LK propositional

$$\text{init}\ \frac{}{p, \Gamma \vdash \Delta, p}$$

$$\neg_{\mathbf{L}}\ \frac{\Gamma \vdash \Delta, A}{\neg A, \Gamma \vdash \Delta} \qquad \neg_{\mathbf{R}}\ \frac{A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg A}$$

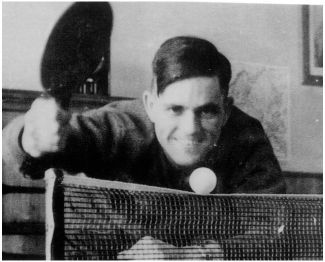
$$\wedge_{\mathbf{L}}\ \frac{A, B, \Gamma \vdash \Delta}{A \wedge B, \Gamma \vdash \Delta} \qquad \wedge_{\mathbf{R}}\ \frac{\Gamma \vdash \Delta, A \quad \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \wedge B}$$

$$\vee_{\mathbf{L}}\ \frac{A, \Gamma \vdash \Delta \quad B, \Gamma \vdash \Delta}{A \vee B, \Gamma \vdash \Delta} \qquad \vee_{\mathbf{R}}\ \frac{\Gamma \vdash \Delta, A, B}{\Gamma \vdash \Delta, A \vee B}$$

$$\rightarrow_{\mathbf{L}}\ \frac{\Gamma \vdash \Delta, A \quad B, \Gamma \vdash \Delta}{A \rightarrow B, \Gamma \vdash \Delta} \qquad \rightarrow_{\mathbf{R}}\ \frac{A, \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \rightarrow B}$$

$$\text{cut}\ \frac{\Gamma \vdash \Delta, A \quad A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta}$$

Today goal



Theorem (*Hauptsatz*, Gentzen 1934)

Every theorem of LK *has a proof that does not use the cut rule.*

Corollary (Analyticity)

Every theorem of LK *has a proof that contains only subformulas of it (up to substitution of free variables).*

Informal example 1

$$\mathrm{cut}\ \dfrac{\wedge_{\mathbf{R}}\ \dfrac{\overset{\mathcal{D}_1}{\Gamma \vdash \Delta, A} \qquad \overset{\mathcal{D}_2}{\Gamma \vdash \Delta, B}}{\Gamma \vdash \Delta, A \wedge B} \qquad \wedge_{\mathbf{L}}\ \dfrac{\overset{\mathcal{D}_3}{A, B, \Gamma \vdash \Delta}}{A \wedge B, \Gamma \vdash \Delta}}{\Gamma \vdash \Delta}\ *$$

Let's eliminate the occurrence of cut marked by $*$

$$\mathrm{cut}\ \dfrac{\overset{\mathcal{D}_1}{\Gamma \vdash \Delta, A} \qquad \mathrm{cut}\ \dfrac{\mathrm{wk}\ \dfrac{\overset{\mathcal{D}_2}{\Gamma \vdash \Delta, B}}{A, \Gamma \vdash \Delta, B} \qquad \overset{\mathcal{D}_3}{A, B, \Gamma \vdash \Delta}}{A, \Gamma \vdash \Delta}}{\Gamma \vdash \Delta}$$

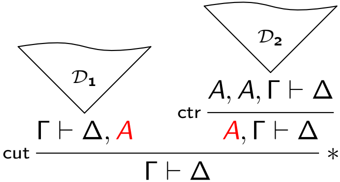
Informal example 2

$$
\text{cut}\;\cfrac{\forall_{\mathbf{R}}\;\cfrac{\begin{matrix}\mathcal{D}_1\\ \Gamma \vdash \Delta, B[x/y]\end{matrix}}{\Gamma \vdash \Delta, {\color{red}\forall x B}} \qquad \forall_{\mathbf{L}}\;\cfrac{\begin{matrix}\mathcal{D}_2\\ B[x/t], \forall x.B, \Gamma \vdash \Delta\end{matrix}}{{\color{red}\forall x.B}, \Gamma \vdash \Delta}}{\Gamma \vdash \Delta}\; *
$$

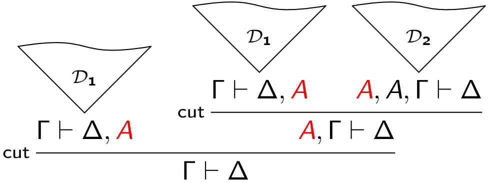
Let's eliminate the occurrence of cut marked by $*$

$$
\text{cut}\;\cfrac{\text{subst}\;\cfrac{\begin{matrix}\mathcal{D}_1\\ \Gamma \vdash \Delta, B[x/y]\end{matrix}}{\Gamma \vdash \Delta, {\color{purple}B[x/t]}} \qquad \text{cut}\;\cfrac{\text{wk}\;\cfrac{\forall_{\mathbf{R}}\;\cfrac{\begin{matrix}\mathcal{D}_1\\ \Gamma \vdash \Delta, B[x/y]\end{matrix}}{\Gamma \vdash \Delta, \forall x B}}{B[x/t], \Gamma \vdash \Delta, {\color{red}\forall x.B}} \qquad \begin{matrix}\mathcal{D}_2\\ B[x/t], {\color{red}\forall x.B}, \Gamma \vdash \Delta\end{matrix}}{{\color{purple}B[x/t]}, \Gamma \vdash \Delta}}{\Gamma \vdash \Delta}
$$

Informal example 3

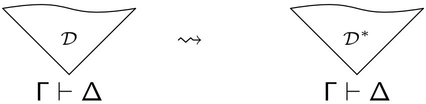


Let's eliminate the occurrence of cut marked by $*$



General strategy of the proof

LK derivation $\rightsquigarrow$ cut-free LK derivation



- ▷ Apply the cut on smaller formulas, until they disappear!
- ▷ Push the cuts upwards in the proof, until they disappear!
- ▷ We need a "measure" on formulas and on derivations, to ensure that the cut-elimination procedure terminates.

... The cut-elimination proof is quite complex.

We are going to sketch the proof for propositional LK (no quantifiers rules).

References

Several proofs of cut-elimination exist in the literature, using slightly different procedures and for slightly different systems:

- ▷ [Buss, 1998]. *Handbook of Proof Theory*.
- ▷ [Troelstra and Schwichtenberg, 1996]. *Basic Proof Theory*.
- ▷ [Negri and von Plato, 2001]. *Structural Proof Theory*.
- ▷ ...

A measure of formulas

The degree of a formula A, $\mathsf{deg}(\mathsf{A})$, is the number of logical connectives occurring in it.

Inductive definition on the structure of the formula:

$$\begin{aligned}\mathsf{deg}(\mathsf{p}) &:= 0 \\ \mathsf{deg}(\mathsf{A} \star \mathsf{B}) &:= \mathsf{deg}(\mathsf{A}) + \mathsf{deg}(\mathsf{B}) + 1\end{aligned}$$

A measure of derivations

The height of $\mathcal{D}$, $\mathrm{ht}(\mathcal{D})$, is the length of its longest branch, minus one.

The rank of $\mathcal{D}$, $\mathrm{rk}(\mathcal{D})$, is the maximal degree of the cut formulas occurring in $\mathcal{D}$, plus 1.

We write

$$\Gamma \vdash_p^m \Delta$$

meaning

> *There is a derivation of $\Gamma \vdash \Delta$ of height* ***at most*** *m and rank* ***at most*** *p.*

A measure of derivations (more formally)

Height and rank can be inductively defined on the structure of $\mathcal{D}$:

$$\mathcal{D} = \text{init}\,\frac{}{\Gamma \vdash \Delta} \qquad \mathrm{ht}(\mathcal{D}) = \mathrm{rk}(\mathcal{D}) = 0$$

$$\mathcal{D} = \text{R}\,\frac{\begin{matrix}\mathcal{D}_1 \\ \Gamma_1 \vdash \Delta_1\end{matrix}}{\Gamma \vdash \Delta} \qquad \mathrm{ht}(\mathcal{D}) = \mathrm{ht}(\mathcal{D}_1) + 1 \qquad \mathrm{rk}(\mathcal{D}) = \mathrm{rk}(\mathcal{D}_1)$$

$$\mathcal{D} = \text{R}\,\frac{\begin{matrix}\mathcal{D}_1 \\ \Gamma_1 \vdash \Delta_1\end{matrix} \qquad \begin{matrix}\mathcal{D}_2 \\ \Gamma_2 \vdash \Delta_2\end{matrix}}{\Gamma \vdash \Delta} \qquad \begin{aligned}\mathrm{ht}(\mathcal{D}) &= \max(\mathrm{ht}(\mathcal{D}_1) + 1, \mathrm{ht}(\mathcal{D}_2) + 1) \\ \mathrm{rk}(\mathcal{D}) &= \max(\mathrm{rk}(\mathcal{D}_1), \mathrm{rk}(\mathcal{D}_2))\end{aligned}$$

$$\mathcal{D} = \text{cut}\,\frac{\begin{matrix}\mathcal{D}_1 \\ \Gamma \vdash \Delta, A\end{matrix} \qquad \begin{matrix}\mathcal{D}_2 \\ A, \Gamma \vdash \Delta\end{matrix}}{\Gamma \vdash \Delta} \qquad \begin{aligned}\mathrm{ht}(\mathcal{D}) &= \max(\mathrm{ht}(\mathcal{D}_1) + 1, \mathrm{ht}(\mathcal{D}_2) + 1) \\ \mathrm{rk}(\mathcal{D}) &= \max(\mathrm{rk}(\mathcal{D}_1), \mathrm{rk}(\mathcal{D}_2), \mathrm{deg}(\mathrm{A}) + 1)\end{aligned}$$

Some preliminary lemmas

1. Lemma: Closure under weakening

   ▷ If $\Gamma \vdash_p^m \Delta$, then $\Gamma', \Gamma \vdash_p^m \Delta, \Delta'$, for any Γ', Δ'.

   Proof. Easy induction on the height m of the derivation.

2. Lemma: Invertibility All the rules are invertible:

   ($\wedge_L$) If $A \wedge B, \Gamma \vdash_p^m \Delta$, then $A, B, \Gamma \vdash_p^m \Delta$.

   ($\wedge_R$) If $\Gamma \vdash_p^m \Delta, A \wedge B$, then $\Gamma \vdash_p^m \Delta, A$ and $\Gamma \vdash_p^m \Delta, A$.

   (... and so on for all the rules)

   Proof. Induction on m, using closure under weakening.

3. Lemma: Closure under contraction

   ▷ If $A, A, \Gamma \vdash_p^m \Delta$, then $A, \Gamma \vdash_p^m \Delta, A$.

   ▷ If $\Gamma \vdash_p^m \Delta, A, A$, then $\Gamma \vdash_p^m \Delta, A$.

   Proof. Induction on m, using invertibility.

NB: all the above preserve height and rank of the derivation.

The plan

- Principal Lemma (most of the work)

  Intuitively: we can compose two derivations $\Gamma \vdash \Delta, A$ and $A, \Gamma \vdash \Delta$ *without* using the *cut* rule, possibly introducing cut on formulas of rank smaller than A in the process.

- Reduction Lemma (follows from PrL)

  Intuitively: we can decrease the rank of derivations, by applying PrL to the cut formulas of highest degree.

- Cut-elimination Theorem (follows from RedL)

  Intuitively: we can iterate the procedure in RedL until all occurrences of cut have been eliminated.

Principal Lemma

Lemma Let $\Gamma \vdash_p^m \Delta, A$ and $A, \Gamma \vdash_p^n \Delta$ with $p = \deg(A)$:

$$\text{cut}\ \dfrac{\overset{\mathcal{D}_1}{\Gamma \vdash_p^m \Delta, A} \qquad \overset{\mathcal{D}_2}{A, \Gamma \vdash_p^n \Delta}}{\Gamma \vdash_{p+1}^{\max(m,n)+1} \Delta}$$

Then, we can construct a derivation $\Gamma \vdash_p^{m+n} \Delta$:

$$\overset{\mathcal{D}}{\Gamma \vdash_p^{m+n} \Delta}$$

Proof of the Principal Lemma

$$\begin{array}{c} \mathcal{D}_1 \\ \mathsf{R_1}\ \dfrac{\Gamma' \vdash_p^{m-1} \Delta'}{\Gamma \vdash_p^{m} \Delta, A} \end{array} \qquad\qquad \begin{array}{c} \mathcal{D}_2 \\ \mathsf{R_2}\ \dfrac{\Gamma'' \vdash_p^{n-1} \Delta''}{A, \Gamma \vdash_p^{n} \Delta} \end{array}$$

Induction on $m + n$. We distinguish cases:

1. R_1 is init (R_2 is init)
2. A is principal in both R_1 and R_2
3. A is not principal in R_1 (A is not principal in R_2)

R_1 is init

$$\mathcal{D}_1 = \text{init}\,\dfrac{}{{\color{red}A}, \Gamma' \vdash_{\color{green}p}^{\color{orange}m} \Delta, {\color{red}A}} \qquad\qquad \begin{array}{c} \mathcal{D}_2 \\ \vdots \end{array} \atop \mathrm{R}_2\,\dfrac{\Gamma'' \vdash_{\color{green}p}^{\color{orange}n-1} \Delta''}{{\color{red}A}, \Gamma \vdash_{\color{green}p}^{\color{orange}n} \Delta}$$

with $\Gamma = {\color{red}A}, \Gamma'$ We construct the following derivation $\mathcal{D}$ of $\Gamma \vdash_{\color{green}p}^{\color{orange}m+n} \Delta$:

$$\begin{array}{c} \mathcal{D}_2 \\ \vdots \end{array}$$

$$\mathrm{ctr}\,\dfrac{\mathrm{R}_2\,\dfrac{\Gamma'' \vdash_{\color{green}p''}^{\color{orange}n-1} \Delta''}{{\color{red}A}, {\color{red}A}, \Gamma' \vdash_{\color{green}p}^{\color{orange}n} \Delta}}{{\color{red}A}, \Gamma' \vdash_{\color{green}p}^{\color{orange}n} \Delta}$$

A is principal in both R_1 and R_2

R_1 is $\to_R$ and R_2 is $\to_L$

$$\to_R \dfrac{\overset{\mathcal{D}_1}{B, \Gamma \vdash_p^{m-1} \Delta, C}}{\Gamma \vdash_p^{m} \Delta, B \to C} \qquad \to_L \dfrac{\overset{\mathcal{D}_2'}{\Gamma \vdash_p^{n1} \Delta, B} \qquad \overset{\mathcal{D}_2''}{C, \Gamma \vdash_p^{n2} \Delta}}{B \to C, \Gamma \vdash_p^{n} \Delta}$$

with $n1, n2 < n$. We construct the following derivation $\mathcal{D}$ of $\Gamma \vdash_p^{m+n} \Delta$:

$$\text{cut}\ \dfrac{\text{cut}\ \dfrac{\text{wk}\ \dfrac{\overset{\mathcal{D}_2'}{\Gamma \vdash_p^{n1} \Delta, B}}{\Gamma \vdash_p^{n1} \Delta, B, C} \qquad \overset{\mathcal{D}_1}{B, \Gamma \vdash_p^{m-1} \Delta, C}}{\Gamma \vdash_p^{k} \Delta, C} \qquad \overset{\mathcal{D}_2''}{C, \Gamma \vdash_p^{n2} \Delta}}{\Gamma \vdash_p^{m+n} \Delta}$$

$k = \max(n1, m-1) + 1 \leq \max(m, n)$ $\qquad \max(k, n2) + 1 \leq \max(m, n) + 1 \leq \max(m, n)$

$\mathrm{rk}(B) < \mathrm{rk}(B \to C)$ $\qquad \mathrm{rk}(C) < \mathrm{rk}(B \to C)$

Cases for the other rules ...

A is not principal in R_1

R_1 is a one-premiss rule

$$
\mathrm{R}_1\;\dfrac{\begin{array}{c}\mathcal{D}_1\\ \Gamma' \vdash^{m-1}_p \Delta', A\end{array}}{\Gamma \vdash^{m}_p \Delta, A}
\qquad\qquad
\mathrm{R}_2\;\dfrac{\begin{array}{c}\mathcal{D}_2\\ \Gamma'' \vdash^{n-1}_p \Delta''\end{array}}{A, \Gamma \vdash^{n}_p \Delta}
$$

We construct the following derivation $\mathcal{D}$ of $\Gamma \vdash^{m+n}_p \Delta$:

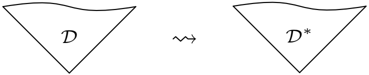
$$
\mathrm{ctr}\;\dfrac{\mathrm{R}_1\;\dfrac{\mathrm{IH}\;\dfrac{\mathrm{wk}\;\dfrac{\begin{array}{c}\mathcal{D}_1\\ \Gamma' \vdash^{m-1}_p \Delta', A\end{array}}{\Gamma', \Gamma \vdash^{m-1}_p \Delta, \Delta', A} \qquad \mathrm{wk}\;\dfrac{\mathrm{R}_2\;\dfrac{\begin{array}{c}\mathcal{D}_2\\ \Gamma'' \vdash^{n-1}_p \Delta''\end{array}}{A, \Gamma \vdash^{n}_p \Delta}}{A, \Gamma', \Gamma \vdash^{n}_p \Delta, \Delta'}}{\Gamma', \Gamma \vdash^{(m-1)+n}_p \Delta, \Delta'}}{\Gamma, \Gamma \vdash^{m+n}_p \Delta, \Delta}}{\Gamma \vdash^{m+n}_p \Delta}
$$

R_1 is a two-premisses rule . . .

End of the proof

Reduction Lemma

Reduction Lemma If $\Gamma \vdash^{m}_{p+1} \Delta$, we can construct $\Gamma \vdash^{2^m}_{p} \Delta$.



$$\Gamma \vdash^{m}_{p+1} \Delta \qquad\qquad \Gamma \vdash^{2^m}_{p} \Delta$$

Proof. Induction on m

Base case Just set $\mathcal{D} = \mathcal{D}^*$

Induction step Case distinction according to the last rule R applied in $\mathcal{D}$.

We show just one case: R is cut.

Proof of the Reduction Lemma: key case

The last rule R applied in $\mathcal{D}$ is cut, with $\deg(A) = p$

$$\text{cut}\ \dfrac{\begin{array}{c}\mathcal{D}_1\\ \Gamma \vdash^{m1}_{p+1} \Delta, A\end{array} \qquad \begin{array}{c}\mathcal{D}_2\\ A, \Gamma \vdash^{m2}_{p+1} \Delta\end{array}}{\Gamma \vdash^{m}_{p+1} \Delta} \qquad \rightsquigarrow \qquad \begin{array}{c}\mathcal{D}^*\\ \Gamma \vdash^{2^m}_{p} \Delta\end{array}$$

We construct $\mathcal{D}^*$ as follows:

Since $m1, m2 < m$, by IH, we have:

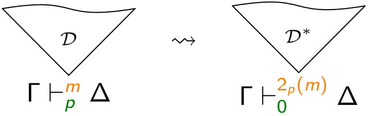
$$\text{PL}\ \dfrac{\begin{array}{c}\mathcal{D}_1^*\\ \Gamma \vdash^{2^{m1}}_{p} \Delta, A\end{array} \qquad \begin{array}{c}\mathcal{D}_2^*\\ A, \Gamma \vdash^{2^{m2}}_{p} \Delta\end{array}}{\Gamma \vdash^{2^{m1}+2^{m2}}_{p} \Delta}$$

$2^{m1} + 2^{m2} \leq 2^{m-1} + 2^{m-1} = 2(2^{m-1}) = 2^m$

We have constructed $\mathcal{D}^*$ of $\Gamma \vdash^{2^m}_{p} \Delta$. End of the proof.

Cut-elimination Theorem

Cut-elimination Theorem If $\Gamma \vdash^m_p \Delta$, we can construct $\Gamma \vdash_0^{2_p(m)} \Delta$, that is, a derivation **where** cut **does not occur.**

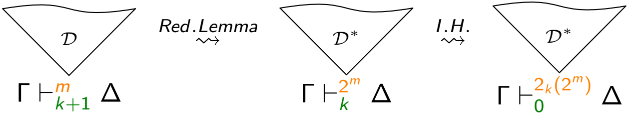


$$2_p(n) \quad = \quad \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2^n}}}}}}_{p}$$

Proof. Induction on p

Base case $p = 0$ Just set $\mathcal{D} = \mathcal{D}^*$

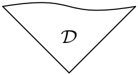
Induction step $p = k+1$



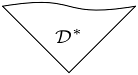
and we're done: $2_k(2^m) = 2_{k+1}(m) = 2_p(m)$ $\square$

Putting it all together

LK derivation $\rightsquigarrow$ cut-free LK derivation



$\rightsquigarrow$



$$\Gamma \vdash_p^m \Delta \qquad \rightsquigarrow \qquad \Gamma \vdash_0^{2_p(m)} \Delta$$

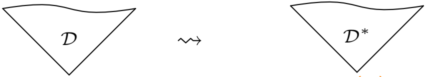
Proof sketch of the Cut-elimination Theorem

By induction on the rank od a proof:

- Identify the cuts of highest rank, say $p+1$, in the proof, and apply the Reduction Lemma to them.
- The Principal Lemma ensures that we might only introduce cuts of rank at most p in the process.
- Thus, the rank of the derivation decreases to p.
- We may conclude by the inductive hypothesis.

The cost of cut-elimination

LK derivation $\rightsquigarrow$ cut-free LK derivation



$$\Gamma \vdash_p^m \Delta \qquad \rightsquigarrow \qquad \Gamma \vdash_0^{2_p(m)} \Delta$$

Eliminating cuts from a propositional LK derivation leads to a hyperexponential blow-up of the size of the proof.

Can we do better?

- For propositional LK: yes, we can get an exponential bound in proof size.
- For full LK (with quantifiers rules): no.

  Theorem (Statman '79, Orevkov '82). Cut-elimination for predicate logic necessarily has a non-elementary cost in proof size.

Cut-elimination for predicate logic

The method presented before can be extended to full LK, modulo assuming a renaming of variables and the following:

Substitution Lemma If $\Gamma \vdash_p^m \Delta$, then for each x variable and t term, $\Gamma[x/t] \vdash_p^m \Delta[x/t]$.

Proof. Easy induction on p.

Summing up: is it worth to eliminate cuts?

Drawbacks:

- Exponential or hyper-exponential blow-up of proof size w.r.t. input size
- Headache proof

Benefits:

- Analyticity: automated proof search
-

References

- ...

Exercises for Lecture 4

1. ...

Appendix

Well-ordered sets

A well-ordered set is a pair $\langle W, \leq_W \rangle$ such that:

- W is a set
- $\leq_W$ is a linear order on W:
  - reflexivity: $x \leq_W x$
  - antysimmetry: $x \leq_W y$ and $y \leq_W y$ implies $x = y$
  - transitivity: $x \leq_W y$ and $y \leq_W z$ implies $x \leq_W z$
  - strong connectedness: $x \leq_W y$ or $y \leq_W x$
- There is no infinite descending chain:

$$\ldots <_W x_{n+1} <_W x_n <_W \ldots <_W x_0 \qquad (x <_W y := x \leq_W y \wedge x \neq$$

Examples of orders on $W := \mathbb{N}$:

- $x \leq_{\mathbb{N}} y$ if $x = 0$ or $x = y$: $0 \leq_{\mathbb{N}} 0 \quad 0 \leq_{\mathbb{N}} 1 \quad 0 \leq_{\mathbb{N}} 2 \quad \ldots$
- $x \leq_{\mathbb{N}} y$ if $y \leq x$: $\ldots \leq_{\mathbb{N}} 4 \leq_{\mathbb{N}} 3 \leq_{\mathbb{N}} 2 \leq_{\mathbb{N}} 1 \leq_{\mathbb{N}} 0$
- $x \leq_{\mathbb{N}} y$ if x, y have same parity or (x even and y odd):

$$0 \leq_{\mathbb{N}} 2 \leq_{\mathbb{N}} 4 \leq_{\mathbb{N}} 6 \leq_{\mathbb{N}} \ldots \leq_{\mathbb{N}} 1 \leq_{\mathbb{N}} 3 \leq_{\mathbb{N}} 5 \leq_{\mathbb{N}} 7 \leq_{\mathbb{N}} \ldots$$