Mining Human Proofs from Machine Proofs

Big Proof / Isaac Newton Institute
Tuesday, 18 July 2017

Paulo Oliva
Queen Mary University of London
(joint work with Rob Arthan)

formula

logic

Is A provable in L?

reduce

recover natural deduction proof

Is E true in C?

equation

class of algebras

machine finds equational proof

Case Studies

Uniqueness of halving in (minimal) continuous logic

Double negation of (double negation elimination)

Double negation translations (sub-structurally)

Logics and Algebras

Minimal Affine Logic

$$\Gamma, A \vdash A$$

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

$$\frac{\Gamma \vdash A \quad \Delta \vdash B}{\Gamma, \Delta \vdash A \otimes B} \qquad \frac{\Delta, A, B \vdash C \quad \Gamma \vdash A \otimes B}{\Gamma, \Delta \vdash C}$$

Further Axioms

*EFQ: $\bot \vdash A$

DNE: $\neg\neg A \vdash A$

*DIV: $A, A \to B \vdash B \otimes (B \to A)$

CON: $A \vdash A \otimes A$

* Assuming weakening

Nine Logics

$$\begin{tikzcd}
\text{Classical Affine Logic} \arrow[r, "\mathrm{DIV}"] & \text{Classical Lukasiewicz Logic} \arrow[r, "\mathrm{CON}"] & \text{Classical Logic} \\
\text{Intuitionistic Affine Logic} \arrow[u, "\mathrm{DNE}"'] \arrow[r, "\mathrm{DIV}"] & \text{Intuitionistic Lukasiewicz Logic} \arrow[u, "\mathrm{DNE}"'] \arrow[r, "\mathrm{CON}"] & \text{Intuitionistic Logic} \arrow[u, "\mathrm{DNE}"'] \\
\text{Minimal Affine Logic} \arrow[u, "\mathrm{EFQ}"'] \arrow[r, "\mathrm{DIV}"] & \text{Minimal Lukasiewicz Logic} \arrow[u, "\mathrm{EFQ}"'] \arrow[r, "\mathrm{CON}"] & \text{Minimal Logic} \arrow[u, "\mathrm{EFQ}"']
\end{tikzcd}$$

pocrim

$$\langle X, \otimes, \rightarrow, \geq, 0 \rangle$$

partially ordered
commutative
residuated integral
monoid

hoops

pocrims satisfying

If $A \geq B$ then $A = B \otimes (B \rightarrow A)$

Buchi/Owens'74

Nine Algebras

$$
\begin{tikzcd}
\text{involutive pocrims} \arrow[r] & \text{involutive hoops} \arrow[r] & \text{boolean algebras} \\
\text{bounded pocrims} \arrow[r] \arrow[u] & \text{bounded hoops} \arrow[r] \arrow[u] & \text{Heyting algebras} \arrow[u] \\
\text{pocrims} \arrow[r] \arrow[u] & \text{hoops} \arrow[r] \arrow[u] & \text{Brouwer algebras} \arrow[u]
\end{tikzcd}
$$

Case Study I

Continuous Logic

Continuous Logic

- Lukasiewicz logic with a halving operator axiomatised as:

$$\frac{A}{2} \Leftrightarrow \frac{A}{2} \to A$$

- Classically it's easy to show this uniquely defines the operation
- But how about in minimal logic?

...prover9

prover9

- Automated theorem prover for first-order and equational logic
- Successor of Otter
- Developed by Bill McCune
- Uses resolution and paramodulation

http://www.cs.unm.edu/~mccune/prover9/

Continuous Logic

- Wanted to show

$$\frac{X \leftrightarrow (X \to A) \quad Y \leftrightarrow (Y \to A)}{X \leftrightarrow Y}$$

- Proof found in about 3 mins (by prover9)
- Subsequently massaged into human-readable form by us

Lemma 2 *Let* $\mathbf{M} = (M, 0, +, \to; \leq)$ *be a hoop and let* $a, b, c, x, y \in M$. *If* $a \to b = a$ *and* $c \to b = c$, *then the following hold:*

(1) $b \geq a$ *and* $b \geq c$.

(2) $a + a = b$.

(3) $a \to (a \to c) = 0$.

(4) $(x \to y) + z \geq x \to (y + (y \to x) + z)$.

(5) $c \to (a + a + x) \geq c$.

(6) $c \to a \geq a \to c$.

(7) $c \to a = a \to c$.

(8) $c + (c \to a) + ((a \to c) \to a) = b$.

(9) $a + c = b$.

Theorem 3 *In any hoop the following holds: if* $a \to b = a$ *and* $c \to b = c$ *then* $a = c$.

Proof: By symmetry it is enough to show $c \geq a$. By Lemma 2 (9) we have $c \geq a \to b$ and hence $c \geq a$.

Case Study II

$$\neg\neg(\neg\neg A \to A)$$

Deriving $\neg\neg(\neg\neg A \to A)$ in IL:

$$\dfrac{\dfrac{[\neg(\neg\neg A \to A)]_\alpha \qquad \dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[\neg(\neg\neg A \to A)]_\alpha \qquad \dfrac{[A]_\beta}{\neg\neg A \to A}\,(\text{WKN})}{\bot}}{\neg A}\,\beta \qquad [\neg\neg A]_\delta}{\bot}}{A}\,(\text{EFQ})}{\neg\neg A \to A}\,\delta}{}}{\bot}}{\neg\neg(\neg\neg A \to A)}\,\alpha\ (\text{CON})$$

Is

$$\neg\neg(\neg\neg A \to A)$$

provable in intuitionistic Lukasiewz logic?

Does

$$\neg\neg(\neg\neg x \to x) = 0$$

hold in all bounded hoops?

Demo!

```
40 x + (x ==> 1) = 1.  [copy(39),flip(a)].
41 1 ==> x = y ==> ((y ==> 1) ==> x).  [para(40(a,1),5(a,1,1))].
42 x ==> ((x ==> 1) ==> y) = 1 ==> y.  [copy(41),flip(a)].
43 x + 1 = y + (x + (y ==> 1)).  [para(40(a,1),18(a,1,2))].
44 1 = y + (x + (y ==> 1)).  [para(9(a,1),43(a,1))].
45 x + (y + (x ==> 1)) = 1.  [copy(44),flip(a)].
46 x + (y ==> (x ==> z)) = (y ==> z) + ((y ==> z) ==> x).  [para(22(a,1),6(a,1,2))].
47 (x ==> y) + ((x ==> y) ==> z) = z + (x ==> (z ==> y)).  [copy(46),flip(a)].
48 x ==> 0 = y ==> (x ==> y).  [para(7(a,1),22(a,1,2))].
49 0 = y ==> (x ==> y).  [para(8(a,1),48(a,1))].
50 x ==> (y ==> x) = 0.  [copy(49),flip(a)].
51 x ==> 0 = y ==> (x ==> ((y ==> z) ==> z)).  [para(29(a,1),22(a,1,2))].
52 0 = y ==> (x ==> ((y ==> z) ==> z)).  [para(8(a,1),51(a,1))].
53 x ==> (y ==> ((x ==> z) ==> z)) = 0.  [copy(52),flip(a)].
54 1 ==> x = 0.  [para(37(a,1),7(a,1))].
55 x ==> ((x ==> 1) ==> y) = 0.  [para(54(a,1),42(a,2))].
56 1 = x + ((x ==> y) + (y ==> 1)).  [para(45(a,1),24(a,1))].
57 x + ((x ==> y) + (y ==> 1)) = 1.  [copy(56),flip(a)].
58 x ==> (0 ==> y) = (z ==> x) ==> (((z ==> x) ==> x) ==> y).  [para(50(a,1),26(a,1,2,1))].
59 x ==> y = (z ==> x) ==> (((z ==> x) ==> x) ==> y).  [para(33(a,1),58(a,1,2))].
60 (x ==> y) ==> (((x ==> y) ==> y) ==> z) = y ==> z.  [copy(59),flip(a)].
61 x ==> ((x ==> y) ==> ((y ==> z) ==> z)) = 0.  [para(26(a,1),53(a,1))].
62 x + (0 + (((x ==> y) ==> y) ==> 1)) = 1.  [para(29(a,1),57(a,1,2,1))].
63 x + (((x ==> y) ==> y) ==> 1) = 1.  [para(20(a,1),62(a,1,2))].
64 x + ((y ==> z) + ((y ==> z) ==> u)) = u + (x + (y ==> (u ==> z))).  [para(22(a,1),38(a,2,2,2))].
65 1 ==> x = y ==> ((((y ==> z) ==> z) ==> 1) ==> x).  [para(63(a,1),5(a,1,1))].
66 0 = y ==> ((((y ==> z) ==> z) ==> 1) ==> x).  [para(54(a,1),65(a,1))].
67 x ==> ((((x ==> y) ==> y) ==> 1) ==> z) = 0.  [copy(66),flip(a)].
68 (x ==> y) + ((x ==> y) ==> (x ==> 1)) = (x ==> 1) + 0.  [para(55(a,1),47(a,2,2))].
69 (x ==> y) + (x ==> ((x ==> y) ==> 1)) = (x ==> 1) + 0.  [para(22(a,1),68(a,1,2))].
70 (x ==> y) + (x ==> ((x ==> y) ==> 1)) = 0 + (x ==> 1).  [para(3(a,1),69(a,2))].
71 (x ==> y) + (x ==> ((x ==> y) ==> 1)) = x ==> 1.  [para(20(a,1),70(a,2))].
```

Certain derived connectives kept appearing:

weak conjunction

$$A \wedge B \equiv A \otimes (A \to B)$$

strong disjunction

$$A \vee B \equiv (B \to A) \to A$$

strong implication

$$A \Rightarrow B \equiv A \to A \otimes B$$

NOR, Peirce's ampheck

$$A \downarrow B \equiv \neg A \otimes (B \to A)$$


prover9
prover9 proof
spot
"defined" connective
spot
def. conn. properties
prover9
proof of lemmas
expand theory
extended theory

Lemma 4.2 ($\text{ŁL}_\mathbf{i}$) $A \otimes B \;\leftrightarrow\; A \otimes (B \vee (A \Rightarrow B))$

Theorem 4.7 ($\text{ŁL}_\mathbf{i}$) $B \downarrow A \;\leftrightarrow\; A \downarrow B$

Corollary 4.8 ($\text{ŁL}_\mathbf{i}$) $(A^{\perp\perp} \multimap A)^{\perp\perp}$

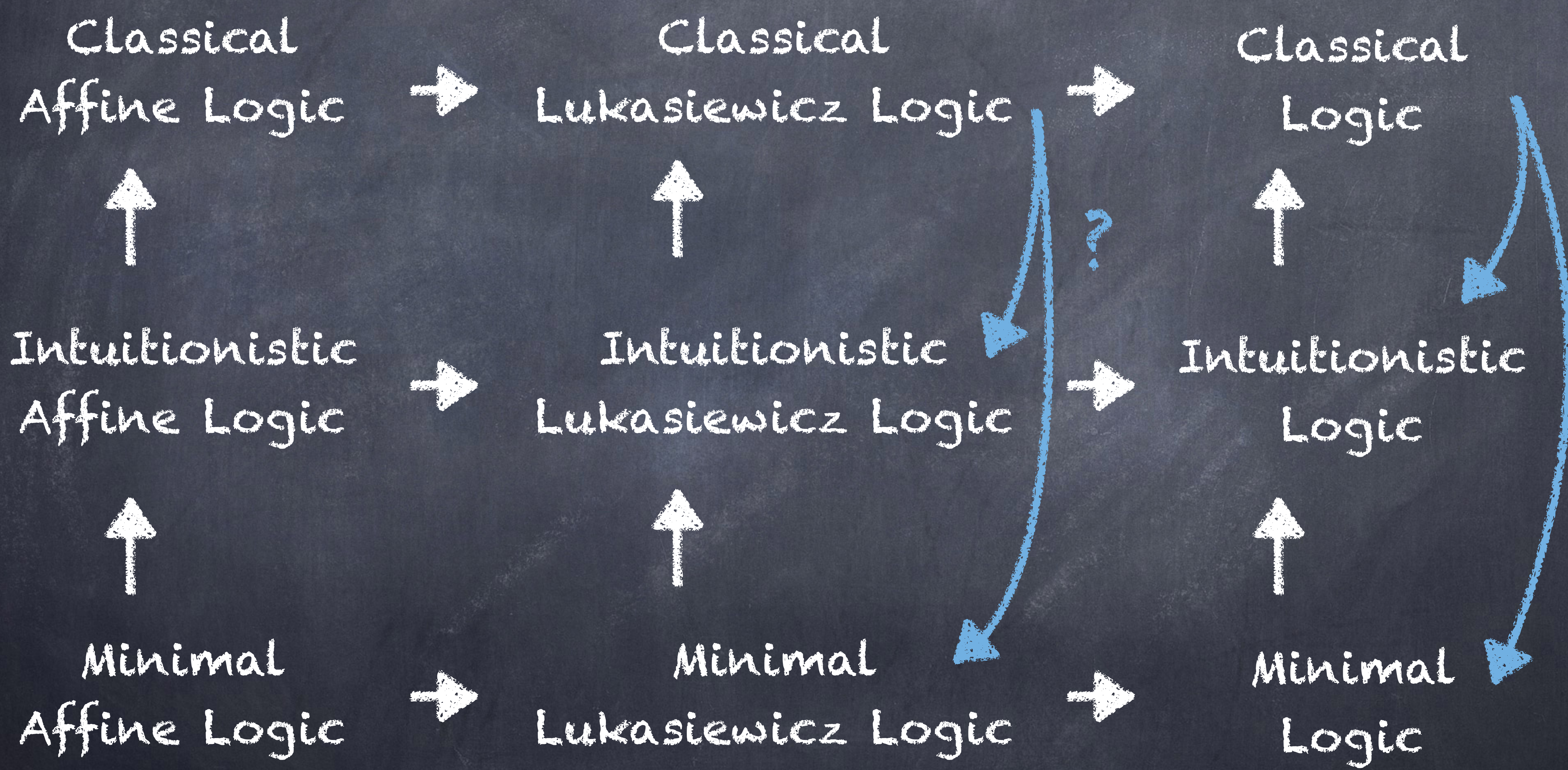
Proof: Note that, since $\perp \;\leftrightarrow\; A \otimes A^{\perp}$ we have $(*)$ $A^{\perp\perp} \;\leftrightarrow\; A^{\perp} \Rightarrow A$. Moreover, it is easy to check that $(**)$ $X \downarrow (Y \multimap X) \;\leftrightarrow\; X^{\perp} \otimes (X \vee Y)$, for all X and Y. Hence

$$\begin{aligned}
(A^{\perp\perp} \multimap A)^{\perp} &\leftrightarrow ((A^{\perp} \Rightarrow A) \multimap A)^{\perp} && (*)\\
&\leftrightarrow ((A^{\perp} \Rightarrow A) \multimap A)^{\perp} \otimes \underline{(A \multimap ((A^{\perp} \Rightarrow A) \multimap A))} && ([\mathsf{WK}])\\
&\leftrightarrow ((A^{\perp} \Rightarrow A) \multimap A) \downarrow A && (\text{def} \downarrow)\\
&\leftrightarrow A \downarrow ((A^{\perp} \Rightarrow A) \multimap A) && (\text{Theorem } 4.7)\\
&\leftrightarrow A^{\perp} \otimes (A \vee (A^{\perp} \Rightarrow A)) && (**)\\
&\leftrightarrow A^{\perp} \otimes A && (\text{Lemma } 4.2)\\
&\leftrightarrow \perp . && \blacksquare
\end{aligned}$$

Case Study III

Double Negation Translations

Negative Translations



Translations of $P \otimes (P \to Q)$

- Kolmogorov

$$\neg\neg(\neg\neg P \otimes \neg\neg(\neg\neg P \to \neg\neg Q))$$

- Gentzen

$$\neg\neg P \otimes (\neg\neg P \to \neg\neg Q)$$

- Glivenko

$$\neg\neg(P \otimes (P \to Q))$$

Translations of $P \otimes (P \to Q)$

$$\neg\neg(\neg\neg P \otimes \neg\neg(\neg\neg P \to \neg\neg Q))$$

simplifications

$$\neg\neg P \otimes (\neg\neg P \to \neg\neg Q) \qquad\qquad \neg\neg(P \otimes (P \to Q))$$

Using CON we easily have:

$$\neg\neg P \otimes \neg\neg Q \;\Leftrightarrow\; \neg\neg(P \otimes Q)$$

$$\neg\neg P \to \neg\neg Q \;\Leftrightarrow\; \neg\neg(P \to Q)$$

which allows us to simplify Kolmogorov and obtain Gentzen and Glivenko

Ferreira/O.'12

Can the same be done with DIV?

Examples of lemmas:

"De Morgan" like properties:

$$\neg(A \otimes B) \equiv A \rightarrow \neg B$$

$$\neg(A \rightarrow B) \equiv \neg\neg A \otimes \neg B$$

$$\neg(A \wedge B) \equiv A \Rightarrow \neg B$$

$$\neg(A \Rightarrow B) \equiv \neg\neg A \wedge \neg B$$

$$\neg(A \wedge B) \equiv \neg A \vee \neg B$$

$$\neg(A \vee B) \equiv \neg A \wedge \neg B$$

Ampheck is definable in terms of conjunction and negation:

$$A \downarrow B \equiv \neg A \wedge \neg B$$

Desired homomorphism properties:

$$\neg\neg P \otimes \neg\neg Q \Leftrightarrow \neg\neg(P \otimes Q)$$

$$\neg\neg P \rightarrow \neg\neg Q \Leftrightarrow \neg\neg(P \rightarrow Q)$$

Weak conjunction residuates strong implication:

$$(A \wedge B) \Rightarrow C \;\equiv\; A \Rightarrow (B \Rightarrow C)$$

- Found by Bob Veroff
- Yet to tease out human readable proof

Conclusions

- Successfully mined human-readable proofs from machine proofs
- Human input is identifying the "right" abstractions:
  - Find useful derived concepts
  - Recover an intuitive proof plan
- Automated support for proof refactoring?
- AI to automate human aspect?
- The late Bill McCune is the real star!