(CHOOSE a b A)

IF	RETURN
a = b	T
Neither a nor b occurs in A	T
a and b are numbers	T
a is a number	a/b
b is a number	b/a
a occurs in b	a/b
b occurs in a	b/a
(ATOM a)	a/b
(ATOM b)	b/a
ELSE	a/b

(SUB=LE EQ)

EQ is a list of equality units, (a b).
DB, H, and C are from IMPLY

ALGORITHM:

For each element (a b) of EQ

1. Put $\theta := $ (CHOOSE a b (DB $\wedge$ H $\wedge$ C))
2. Put DB: = DBθ, H: = Hθ, C: = Cθ
   Repeat steps 1-2 for all elements of EQ

If two elements ('int' x a b) and ('int' x a' b') have the same second member x, after the above substitutions, then they are combined into one ('int' x (max a a') (min b b')).

(NOTL A)

$A \equiv (B \wedge C)$	((NOTL B) $\vee$ (NOTL C))
$A \equiv (B \vee C)$	((NOTL B) $\wedge$ (NOTL C))
$A \equiv \sim B$	B
$A \equiv (B \rightarrow C)$	(B $\wedge$ (NOTL C))
$A \equiv (\leq a\ b)$	$(< b\ a)$
$A \equiv (< a\ b)$	$(\leq b\ a)$
ELSE	$(\sim A)$

NOTL pushes the $\sim$ to the inside as far as it will go, so that only atomic formulas contain '$\sim$'. In the case of pure inequalities, if A has no '$\sim$', then (NOTL A) has none. For example if A is

$$(B \wedge D \rightarrow (E \rightarrow (a < b)))$$

then (NOTL A) is

$$(B \wedge D \wedge E \wedge b \leq a) .$$

(GROUND A)

Variables, to be instantiated by the prover, are represented as atoms. Thus in any expression, an atom, which is not a number and which does not occupy the initial position of a list, is considered to be a variable. Thus x and y are variables in (+ x y) and (- x), but not in (x a).

An expression A is said to be "Ground" if it has no such variables.

3. Examples

In the following examples we will depict a call to

(IMPLY DB H C TL LT PV)

as follows: where DB ≡ (A-UNIT RL TY),

(TL) $\quad\quad$ ([TY] $\wedge$ H $\Rightarrow$ C) $\quad$ {RL} .

Thus: the theorem label TL is at the left; the TYPELIST TY, being part of the hypothesis is shown with H in implying C; the restriction list RL is shown at the right. A-UNIT will be omitted here. PV is explained below.

In the descriptions of proofs of the examples we will often leave out some of the steps when clarity is not impaired, and often, the components TL, TY, H, C, and RL, will not all be written at each step.

The theorem label TL starts as an empty list ( ) and grows, as the depth of the proofs increases. It consists of a sequence of symbols representing the actions that have been taken. Some of the symbols used in TL are

1	first branch of an And-Split, Rule 4 of IMPLY
2	second branch of an And-Split, Rule 4 of IMPLY
P→	Promote; Rule 6 of IMPLY
P2→	Reverse Promote; Rule 6.1 of IMPLY
L	L is the lemma name of a hypothesis used to obtain the current subgoal from the last
S=	Equal substitution Rule 5
O	Or-split, Rule 8
CHECK	Check that a proposed binding is consistent with RL

Even though the prover obtains a substitution

$$\theta = (t_1/x_1 \ t_2/x_2 \ \ldots \ t_n/x_n)$$

from a call to IMPLY, in actuality it returns only the part of θ which will affect a later part of the proof. For example, in proving a theorem of the form

$$(H \rightarrow A \wedge B),$$

if $\theta = (a/x \ b/y)$ is obtained for the proof of the subgoal $(H \rightarrow A)$, and if x does not occur in B and y does, then the prover will return only

$$\theta = (b/y)$$

for the subgoal $(H \rightarrow A)$. This prevents the proliferation of substitution units which will have no further use in the proof. (See the Examples below, especially Example 7.) This feature is implemented by the use of the parameter, PV, which is a list of variables ("protected variables"). If a substitution unit (t/x) from IMPLY has nothing in common with PV (and no connection to it) then (t/x) is dropped from the answer. PV starts as NIL and has variables added to it by AND-C (see p. 9, footnote *). (In AND-C, when proving the first half of a conjunction $(A \wedge B)$, the variables common to both A and B are added to PV.)

This PV feature was not present in our earlier provers [4,3,1].

In the following examples all symbols a,b,c,f,g,ℓ, etc., that are not quantified by $\forall$ or $\exists$ will be treated as constants ("skolem" constants).

These examples were proved automatically.

Ex. 1. $(a \leq b \rightarrow \exists x\ (x \leq b))$

TL	H	C
( )	(NIL $\Rightarrow$	$(a \leq b \rightarrow x \leq b)$)

This shows the theorem as it is first called by IMPLY. It has been skolemized and sits wholly in C. TY and RL (being NIL) are omitted. The theorem label TL is at this point the empty list ( ).

Rule 6 ("Promote") of IMPLY is now applied to obtain

TL	TY	C
P→)	([$a \leq b$] $\Rightarrow$	$x \leq b$)

Notice that the ground formula $a \leq b$, when promoted, was inserted into the TYPELIST TY. The hypothesis H, being NIL is not shown.

Rule 7 of IMPLY is now triggered which calls (PROVE-LE x b $\leq$). PROVE-LE first calls LESS=, but that fails, and then (RESTRICTION-LE x b $\leq$) which "solves" the theorem by placing

$$('int'\ x\ (< -\infty)\ (\leq b))$$

in the data base; and

$$(T\ (NIL\ (('int'\ x\ (< -\infty)\ (\leq b)))\ NIL))$$

is gotten as the answer to complete the proof. Q.E.D.

We will now rewrite the above description of Ex. 1 in a more abbreviated form.

Ex. 1. $(a \leq b \rightarrow \exists x\ (x \leq b))$

(P→) $([a \leq b] \Rightarrow x \leq b)$

Use RLE (RESTRICTION-LE) $\{x:\ -\infty < x \leq b\}$

Q.E.D.

Notice that we have written $\{x:\ -\infty < x \leq b\}$ in place

$$(\text{'int'}\ x\ (<-\infty)\ (\leq b))\ .$$

In the sequel we will use the following abbreviations:

PLE	for	PROVE-LE
RLE	for	RESTRICTION-LE
GLE	for	PROVE-LE-GROUND-CASE
MTY	for	MATCH-IN-TYPELIST
MLE	for	MATCH-LE

In Ex. 1, we did not use the hypothesis $a \leq b$ in the proof because RLE solved $x \leq b$ without any help. Also we note that we never actually bound x to any particular value (such as a), but only restricted it to an interval of values; any value in that interval would satisfy the theorem.

If in Ex. 1, we had not used RLE then PROVE-LE would have next tried the ground prover GLE but that does not apply since $x \leq b$ is not ground, and then tried MTY which succeeds with the substitution a/x. In this last case we would have written

P→) $([a \leq b] \rightarrow x \leq b)$

Use MTY: $a \leq b$: a/x. Q.E.D.

<u>Ex. 2.</u> $(f(\ell) < 0 \wedge 0 \leq f(b) \wedge b \leq \ell \rightarrow b < \ell)$

<u>Proof.</u>

( ) $$(f(\ell) < 0 \wedge 0 \leq f(b) \wedge b \leq \ell \rightarrow b < \ell)$$

(P→) $$([f(\ell) < 0 \wedge 0 \leq f(b) \wedge b \leq \ell] \Rightarrow b < \ell)$$

PROVE-LE is called, which calls LESS= and RLE, which fail.

PROVE-LE then calls the Ground prover GLE which succeeds as follows:

The $\sim(b < \ell) \equiv (\ell \leq b)$ is inserted into TY which becomes

$$[f(\ell) < 0 \wedge 0 \leq f(b) \wedge b \leq \ell \wedge \ell \leq b]$$

CONTRADICTION is called but it finds no contradiction in TY; however, it does find and return the equality unit $(b = \ell)$, which is then applied to TY to obtain

$$[f(\ell) < 0 \wedge 0 \leq f(\ell)]$$

which has a contradiction. Q.E.D.

In our description of Ex.'s 3,4,... below, we will omit much of this explanation, giving mainly the resulting formulas and an indication of the agent causing the changes. In particular, we will usually not mention a routine that has been called by failed (such as LESS= and RLE in the above).

Ex. 3. $(f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < c \wedge b \leq \ell$

$\longrightarrow \exists y [\forall z (z \leq b \wedge f(z) \leq 0 \rightarrow z \leq y) \wedge y < \ell])$

(P→) $([f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < c \wedge b \leq \ell]$

$\Rightarrow (Zy \leq b \wedge f(Zy) \leq 0 \rightarrow Zy \leq y) \wedge y < \ell)$

Note: y is a variable, Zy is a skolem function of y, and the other letters represent constants.

(P→ 1) $([f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < c \wedge b \leq \ell]$

$\Rightarrow (Zy \leq b \wedge f(Zy) \leq 0 \rightarrow Zy \leq y))$

(P→ 1 P→) $([f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < c \wedge b \leq \ell] \wedge Zy \leq b \wedge f(Zy) \leq 0$

$\Rightarrow Zy \leq y)$

LESS=, RLE, GLE, and MTY fail.

MLE: $Zy \leq b$: b/y

Note: the substitution b/y has succeeded on subgoal (P→ 1), and then prover will now proceed to subgoal (P→ 2) with y replaced by b.

(P→ 2) $([f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < c \wedge b \leq \ell] \Rightarrow b < \ell)$

GLE inserts $\sim(b < \ell)$ into TY which becomes

$[f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < c \wedge b \leq \ell \wedge \ell \leq b]$

CONTRADICTION discovers the equality $(b = \ell)$ in TY, getting

$[f(\ell) < 0 \wedge 0 \leq f(\ell) \wedge \ell < c]$

which holds a contradiction. Q.E.D.

Ex. 4. $(a \leq 2 \leq b) \rightarrow \exists x\ (0 \leq x \leq 5 \wedge a \leq x))$

-( ) $(a \leq 2 \wedge 2 \leq b \rightarrow (0 \leq x \wedge x \leq 5) \wedge a \leq x)$

(P→) $([a \leq 2 \wedge 2 \leq b] \Rightarrow (0 \leq x \wedge x \leq 5) \wedge a \leq x)$

(P→ 1) $([\qquad\qquad] \Rightarrow (0 \leq x \wedge x \leq 5))$

(P→ 1 1) $([\qquad\qquad] \Rightarrow 0 \leq x)$

PROVE-LE is called which calls LESS= which fails.

PROVE-LE then calls RLE which succeeds

with $\{x\colon 0 \leq x < \infty\}$

(P→ 1 2) $([\qquad\qquad] \Rightarrow x \leq 5)$

RLE $\{x\colon 0 \leq x \leq 5\}$

(P→ 2) $([a \leq 2 \wedge 2 \leq b] \Rightarrow a \leq x)$

RLE $\{x\colon (\max\ 0\ a) \leq x \leq 5\}$

Here RLE has intersected the interval $\{x\colon a \leq x < \infty\}$ with $\{x\colon 0 \leq x < 5\}$ to obtain the desired result. However, before it can return this answer it must verify that it is not empty: i.e., that $a \leq 5$.

(P→ 2 CHECK) $([a \leq 2 \wedge 2 \leq b] \Rightarrow a \leq 5)$

PROVE-LE calls GLE, which inserts $\sim(a \leq 5) \equiv (5 < a)$ into TY to get

$$[5 < a \leq 2 \wedge 2 \leq b]$$

which has a contradiction (by routine CONTRADICTION). Q.E.D.

The following lemmas will be used in Examples 5,6,7,...

LUB: $$([\exists u \forall t(t \leq b \wedge f(t) \leq 0 \rightarrow t \leq u) \wedge \exists r(r \leq b \wedge f(r) \leq 0)]$$
$$\longrightarrow \exists \ell[\forall x(x \leq b \wedge f(x) \leq 0 \rightarrow x \leq \ell)$$
$$\wedge \forall y(\forall z(z \leq b \wedge f(z) \leq 0 \rightarrow z \leq y) \rightarrow \ell \leq y)]).^{*}$$

In its skolemized form this becomes

LUB: $(B \wedge NE \rightarrow LUB1 \wedge LUB2)$

where,

B: $(t_u \leq b \wedge f(t_u) \leq 0 \rightarrow t_u \leq u)$ **BOUNDED**

NE: $(r \leq b \wedge f(r) \leq 0)$ **NOT EMPTY**

LUB1: $(xL \leq b \wedge f(xL) \leq 0 \rightarrow xL \leq \ell)$

LUB2: $((z_{yL} \leq b \wedge f(z_{yL}) \leq 0 \rightarrow z_{yL} \leq yL) \rightarrow \ell \leq yL)$

Note that z_{yL} is a skolem function of the variable yL

L1: $\forall x(a \leq x \leq b \wedge 0 < f(x) \rightarrow \exists t(t < x \wedge \forall s(t < s \leq x \rightarrow 0 < f(s))))$

Skolemized: $(a \leq x1 \leq b \wedge 0 < f(x1) \rightarrow (t_{x1} < x1 \wedge (t_{x1} < s \wedge s \leq x1 \rightarrow 0 < f(s))))$

L2: $\forall x(a \leq x \leq b \wedge f(x) < 0 \rightarrow \exists t(x < t \wedge \forall s(x \leq s < t \rightarrow f(s) < 0)))$

Skolemized: $(a \leq x2 \leq b \wedge f(x2) < 0 \rightarrow (x2 < t_{x2} \wedge (x2 \leq s2 \wedge s2 < t_{x2} \rightarrow$
$f(s2) < 0))).$

---

*LUB is the result of substituting the set $\{z: z \leq b \wedge f(z) \leq 0\}$ for the set variable A, in the least upper bound axiom: Any set A which is bounded above and non-empty has a least upper bound.

Ex. 5. $(LUB1 \wedge LUB2 \wedge f(\ell) < 0 \wedge 0 \leq f(b)$

$\longrightarrow (\forall s(0 \leq f(s) \wedge \ell \leq s \rightarrow t \leq s) \rightarrow t \leq \ell))$

where ℓ, b and t are constants.

(P→) $([f(\ell) < 0 \wedge 0 \leq f(b)] \wedge LUB1 \wedge LUB2$

$\Rightarrow ((0 \leq f(s) \wedge \ell \leq s \rightarrow t \leq s) \rightarrow t \leq \ell)$.

It now "promotes" (Rule 3 of IMPLY)

(P→ P→) $([f(\ell) < 0 \wedge 0 \leq f(b)] \wedge LUB1 \wedge LUB2$

$\wedge \quad (0 \leq f(s) \wedge \ell \leq s \rightarrow t \leq s) \Rightarrow t \leq \ell)$

This eventually fails, so it "reverse promotes" (Rule 6.1 of IMPLY)

(P→ P2→) $([f(\ell) < 0 \wedge 0 \leq f(b)] \wedge LUB1 \wedge LUB2$

$\wedge \quad \ell < t \Rightarrow (0 \leq f(s) \wedge \ell \leq s) \wedge s < t)$

$([f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < t] \wedge LUB1 \wedge LUB2$

$\Rightarrow \quad (0 \leq f(s) \wedge \ell \leq s) \wedge s < t)$

(P→ P2→ 1) $([\qquad\qquad] \wedge LUB1 \wedge LUB2 \Rightarrow 0 \leq f(s) \wedge \ell \leq s)$

(P→ P2→ 1 1) $([f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < t] \wedge LUB1 \wedge LUB2 \Rightarrow 0 \leq f(s))$

MTY: $0 \leq f(b)$: b/s

This eventually fails. (See note at the end of the proof.)

MLE uses LUB1: $0 < f(xL)$: s/xL. (See MATCH-LE 4.1.)

$\underbrace{\text{(P→ P2→ 1 1 LUB1)}}_{\lambda}$ $([\qquad\qquad] \wedge LUB1 \wedge LUB2 \Rightarrow s \leq b \wedge \ell < s)$

($\wedge$ 1) $\quad$ ([$\quad$] $\wedge$ " $\Rightarrow s \leq b$)

RLE $\quad$ $\{s: -\infty < s \leq b\}$

($\wedge$ 2) $\quad$ ([$\quad$] $\wedge$ " $\Rightarrow \ell < s$)

RLE $\quad$ $\{s: \ell < s \leq b\}$

The goal $(\ell < s)$ has the solution $\{s: \ell < s < \infty\}$, and this is intersected with the existing restriction on s, namely $\{s: -\infty < s \leq b\}$ to obtain the desired $\{s: \ell < s \leq b\}$. However, this is valid only if this latter interval is not empty, so it must check $(\ell < b)$.

($\wedge$ 2 CHECK) $\quad$ ([$\quad$] $\wedge$ " $\Rightarrow \ell < b$)

MLE uses LUB2: $\ell \leq yL$: b/yL. (See MATCH-LE, Rule 8.)

($\wedge$ 2 CHECK LUB2) $\quad$ ([$\quad$] $\wedge$ " $\Rightarrow (z_b \leq b \wedge f(z_b) \leq 0 \rightarrow z_b \leq b) \wedge \ell \neq b$)

($\wedge$ 2 CHECK LUB2 1) $\quad$ ([$\quad$] $\wedge$ " $\Rightarrow (z_b \leq b \wedge f(z_b) \leq 0 \rightarrow z_b \leq b)$)

($\wedge$ 2 CHECK LUB2 1 P$\rightarrow$) ($[f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < t \wedge z_b \leq b \wedge f(z_b) \leq 0] \Rightarrow z_b \leq b$)

Proved by GLE.

($\wedge$ 2 CHECK LUB2 2) ($[f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < t] \wedge$ LUB1 $\wedge$ LUB2 $\Rightarrow \ell \neq b$)

Use IMPLY Rule 8

($\wedge$ 2 CHECK LUB2 2 S=) ($[f(\ell) < 0 \wedge 0 \leq f(\ell) \wedge \ell < t] \wedge$ LUB1 $\wedge$ LUB2 $\Rightarrow \ell < b$)

GLE returns TRUE (contradiction in TY)

So goal ($\wedge$ 2 CHECK) is proved, with the substitution b/yL, and hence the interval $\{s: \ell < s \leq b\}$ is valid for goal ($\wedge$ 2), and for ($\wedge$) = (P$\rightarrow$ P2$\rightarrow$ 1 1 LUB1).

Note that IMPLY has obtained the bindings $((s/xL, b/yL)(\text{NIL}\{s: \ell \leq s \leq b,\ s < t\}\text{NIL}))$ in proving the subgoal $(\wedge) = (P\!\rightarrow P2\!\rightarrow 1\ 1)$. However, it <u>returns</u> only $(T\ (\text{NIL}\{s: \ell < s \leq b,\ s < t\}\text{NIL}))$, dropping the bindings s/xL and b/yL because the variables xL and yL do not occur in the remaining subgoals to be proved. See explanation at the beginning of this section, page 27.

(P→ P2→ 1 2) $([\qquad\qquad] \wedge\ \text{"}\ \Rightarrow\ \ell \leq s)$

RLE returns TRUE, because s already has the restriction $\{s: \ell < s \leq b\}$ which implies the desired goal $\ell \leq s$. Or, said another way, the intersection of $\{s: \ell < s \leq b\}$ and $\{s: \ell \leq s < \infty\}$ is again $\{s: \ell < s \leq b\}$.

(P→ P2→ 2) $([\qquad\qquad] \wedge\ \text{"}\ \Rightarrow\ s < t)$

RLE returns $\{s: \ell < s \leq b \wedge s < t\}$. But it must check that $\ell < t$.

(P→ P2→ 2 CHECK) $([f(\ell) < 0 \wedge 0 \leq f(\ell) \wedge \ell < t] \wedge \text{LUB1} \wedge \text{LUB2} \Rightarrow \ell < t)$

GLE inserts $\sim(\ell < t)$ into TY and detects a contradiction. Q.E.D.

NOTE: In the above proof of Ex. 5, at Step (P→ P2→)

(P→ P2→) $([f(\ell) < 0 \wedge 0 \leq f(b) \wedge \ell < t] \wedge \text{LUB1} \wedge \text{LUB2} \Rightarrow (0 \leq f(s) \wedge \ell \leq s) \wedge s < t)$

it was necessary to use "backtracking". (See algorithm AND-C, and the explanation thereafter.)

This was accomplished by AND-C as follows:

AND-C called the first subgoal

(P→ P2→ 1) $([\qquad] \wedge LUB1 \wedge LUB2 \Rightarrow (0 \leq f(s) \wedge \ell \leq s))$

which returned the substitution $\theta = (b/s)$, and this was applied to the second subgoal

(P→ P2→ 2) $([\qquad] \wedge LUB1 \wedge LUB\ 2 \Rightarrow (s < t)\ \theta)$

$([\qquad] \wedge LUB1 \wedge LUB\ 2 \Rightarrow (b < t))$

which failed.

At this point AND-C again called the second subgoal but without applying the substitution θ.

P→ P2→ 2) $([\qquad] \wedge LUB1 \wedge LUB2 \Rightarrow s < t)$

This succeeded using RLE with $\lambda = \{s: -\infty < s < t\}$.

θ and λ are not compatible and CONFLICT is called, which returns the unit b/s. Then the proof of (P→ P2→) is restarted, with the set, EXCLUDE, now containing the unit b/s, which prevented s from being bound to b in the sequel, and the proof continued successfully as shown above.

Ex. 6. $(a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b) \wedge LUB \wedge L1 \wedge L2 \rightarrow 0 \leq f(\ell))$

where a, b, ℓ, and f are constants.

(P→) $([a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b)] \wedge \underbrace{LUB \wedge L1 \wedge L2}_{H} \Rightarrow 0 \leq f(\ell))$

GLE tries and fails but the TYPELIST TY is changed,

$([a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b) \wedge f(\ell) < 0] \wedge H \Rightarrow 0 \leq f(\ell))$

MLE tries LUB1; fails.

MLE uses L_2: $0 \leq f(x2)$: $\ell/x2$

($P\rightarrow L_2$) $([\quad] \wedge H \Rightarrow$

$(a \leq \ell \leq b \wedge (t_\ell \leq \ell \vee (\ell \leq s2 \wedge s2 < t_\ell \wedge 0 \leq f(s2))))$

($P\rightarrow L_2$ 1) $([\quad] \wedge H \Rightarrow a \leq \ell \wedge \ell \leq b)$

($P\rightarrow L_2$ 1 1) $([\quad] \wedge H \Rightarrow a \leq \ell)$

Tries GLE; fails.

$([a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b) \wedge f(\ell) < 0 \wedge \ell < a] \wedge H \Rightarrow a \leq \ell)$

MLE uses LUB1: $xL \leq \ell$: a/xL

$\underbrace{(P\rightarrow L_2 1\ 1\ LUB1)}_{\lambda}$ $([\quad] \wedge H \Rightarrow (a \leq b \wedge f(a) \leq 0) \wedge (B \wedge NE))$

(λ 1) $([\quad] \wedge H \Rightarrow a \leq b \wedge f(a) \leq 0)$

(λ 1 1) $([a \leq b \wedge \ldots] \wedge H \Rightarrow a \leq b)$

True by GLE.

(In the sequel when the conclusion C appears explicitly in TY we will just state "True by GLE", without saying how it is done.)

(λ 1 2) $([a \leq b \wedge f(a) \leq 0 \wedge \ldots] \wedge H \Rightarrow f(a) \leq 0)$

True by GLE.

(λ 2) $([\qquad] \wedge H \Rightarrow B \wedge NE)$

(λ 2 1) $([\qquad] \wedge H \Rightarrow \underbrace{(t'_u \leq b \wedge f(t'_u) \leq 0 \rightarrow t'_u \leq u)}_{B})$

(λ 2 1 P$\rightarrow$) $([\qquad] \wedge H \wedge t'_u \leq b \wedge f(t'_u) \leq 0 \Rightarrow t'_u \leq u)$

MLE uses $t'_u \leq b$: b/u

(λ 2 2) $([\qquad] \wedge H \Rightarrow NE)$

$([\qquad] \wedge H \Rightarrow r \leq b \wedge f(r) \leq 0)$

(λ 2 2 1) $([\qquad] \wedge H \Rightarrow r \leq b)$

RLE $\{r: -\infty < r \leq b\}$

(λ 2 2 2) $([a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b) \wedge \ldots] \wedge H \Rightarrow f(r) \leq 0)$

MTY: $f(a) \leq 0$: a/r

But it must check that $a \leq b$ since r is restricted by $\{r: -\infty < r \leq b\}$.

(λ 2 2 2 CHECK) $([a \leq b \wedge \ldots] \wedge H \Rightarrow a \leq b)$

Proved previously (in (λ 1 1)). Rule 1 of IMPLY. Thus subgoals (λ) = (P$\rightarrow$ L_2 1 1 LUB1) and (P$\rightarrow$ L_2 1 1) are proved.

$(P\to L_2\ 1\ 2)$ $([\qquad\qquad\qquad] \wedge H \Rightarrow \ell \leq b)$

MLE uses LUB2: $\ell \leq yL$: b/yL

$(P\to L_2\ 1\ 2\,LUB2)$ $([\qquad\qquad\qquad] \wedge H \Rightarrow (z_b \leq b \wedge f(z_b) \leq 0 \to z_b \leq b) \wedge (B \wedge NE))$

$(P\to L_2\ 1\ 2\,LUB2\ 1)$ $([\qquad\qquad\qquad] \wedge H \Rightarrow (z_b \leq b \wedge f(z_b) \leq 0 \to z_b \leq b))$

$(P\to L_2\ 1\ 2\,LUB2\ 1\,P\to)$ $([\ldots \wedge z_b \leq b \wedge f(z_b) \leq 0] \wedge H \Rightarrow z_b \leq b)$

TRUE by GLE.

$(P\to L_2\ 1\ 2\,LUB2\ 2)$ $([\qquad\qquad\qquad] \wedge H \Rightarrow B \wedge NE)$

Proved previously $(\wedge\ 2)$.

Thus $(P\to L_2\ 1\ 2)$ and $(P\to L_2\ 1)$ are proved.

$(P\to L_2\ 2)$ $([a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b) \wedge f(\ell) < 0] \wedge H$

$\Rightarrow t_\ell \leq \ell \vee (\ell \leq s2 \wedge s2 < t_\ell \wedge 0 \leq f(s2))$

Rule 5 of IMPLY.

$(P\to L_2\ 2\ 0)$ $([\qquad\qquad\qquad] \wedge H \Rightarrow (\ell < t_\ell \to (\ell \leq s2 \wedge \ldots)))$

$\underbrace{(P\to L_2\ 2\ 0\,P\to)}_{\lambda_1}$ $[a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b) \wedge f(\ell) < 0 \wedge \ell < t_\ell] \wedge H$

$\Rightarrow (\ell \leq s2 \wedge s2 < t_\ell) \wedge 0 \leq f(s2))$

$(\lambda_1\ 1)$ $([\qquad\qquad\qquad] \wedge H \Rightarrow (\ell \leq s2 \wedge s2 < t_\ell))$

$(\lambda_1\ 1\ 1)$ $([\qquad\qquad\qquad] \wedge H \Rightarrow \ell \leq s2)$

RLE $\{s2\!: \ell \leq s2 < \infty\}$

$(\lambda_1\ 1\ 2)$ $([\quad] \wedge H \Rightarrow s2 < t_\ell)$

RLE $\{s2: \ell \leq s2 < t_\ell\}$

$(\lambda_1\ 1\ 2\ \text{CHECK})$ $([a \leq b \wedge \ldots \wedge \ell < t_\ell] \wedge H \Rightarrow \ell < t_\ell)$

True by GLE.

$(\lambda_1\ 2)$ $([\quad] \wedge H \Rightarrow 0 \leq f(s2)$

MLE uses LUB1: $0 < f(xL)$: $s2/xL$

$(\lambda_1\ 2\ \text{LUB1})$ $([\quad] \wedge H \Rightarrow (s2 \leq b \wedge \ell < s2) \wedge (B \wedge NE))$

$(\lambda_1\ 2\ \text{LUB1}\ 1)$ $([\quad] \wedge H \Rightarrow (s2 \leq b \wedge \ell < s2))$

$(\lambda_1\ 2\ \text{LUB1}\ 1\ 1)$ $([\quad] \wedge H \Rightarrow s2 \leq b)$

RLE $\{s2: \ell \leq s2 \begin{array}{l} \leq b \\ < t_\ell \end{array}\}$

But it must check $\ell \leq b$.

$(\lambda_1\ 2\ \text{LUB1}\ 1\ 1\ \text{CHECK})$ $([\quad] \wedge H \Rightarrow \ell \leq b)$

MLE uses LUB2: $\ell \leq yL$: b/yL.

$(\lambda_1\ 2\ \text{LUB1}\ 1\ \text{CHECK LUB2})$

$([\quad] \wedge H \Rightarrow (z_b \leq b \wedge f(z_b) \leq 0 \rightarrow z_b \leq b) \wedge (B \wedge NE))$

$\lambda_1\ 2\ \text{LUB}\ 1\ 1\ \text{CHECK LUB2}\ 1)$

$([\quad] \wedge H \Rightarrow (z_b \leq b \wedge f(z_b) \leq 0 \rightarrow z_b \leq b))$

True. Promote and use GLE.

$(\lambda_1$ 2 LUB1 1 1 CHECK LUB2 2)

$([\qquad\qquad] \wedge H \Rightarrow B \wedge NE)$

Proved previously $(\lambda\ 2)$.

Thus $(\lambda_1$ 2 LUB1 1 1 CHECK LUB2) and $(\lambda_1$ 2 LUB1 1 1) are proved and

$(T\ (NIL\{s2:\ \ell\quad s2 \begin{smallmatrix}\leq b\\ < t_\ell\end{smallmatrix}\}\ NIL))$ is returned.

$(\lambda_1$ 2 LUB1 1 2) $([\qquad\qquad] \wedge H \Rightarrow \ell < s2)$

RLE $\{s2:\ \ell < s2 \begin{smallmatrix}\leq b\\ < t_\ell\end{smallmatrix}\}$

Must check $\ell < b$.

$(\lambda_1$ 2 LUB1 1 2 CHECK)

$([\qquad\qquad] \wedge H \Rightarrow \ell < b)$

MLE uses LUB2: $\ell \leq yL$: b/yL

$\underbrace{(\lambda_1 \text{ 2 LUB1 1 2 CHECK LUB2})}_{\lambda_1'}$ $(\qquad\qquad \wedge H \quad ((z_b \leq b \wedge f(z_b) \leq 0 \rightarrow z_b \leq b) \wedge \ell \neq b) \wedge (B \wedge NE))$

$(\lambda_1'\ 1)$ $([\qquad\qquad] \wedge H \Rightarrow (z_b \leq b \wedge f(z_b) \leq 0 \rightarrow z_b \leq b)) \wedge \ell \neq b)$

$(\lambda_1'\ 1\ 1)$ $([\qquad\qquad] \wedge H \Rightarrow (z_b \leq b \wedge f(z_b) \leq 0 \rightarrow z_b \leq b))$

Proved previously $(\lambda_1$ 2 LUB1 1 1 CHECK LUB2 1).

$(\lambda_1'\ 1\ 2)$ $([a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b) \wedge f(\ell) < 0 \wedge \ell < t_\ell] \wedge H \Rightarrow \ell \neq b)$

Use Rule 8 of IMPLY, replacing ℓ by b.

$(\lambda_1'\ 1\ 2\ S=)$ $([a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b) \wedge f(b) < 0 \wedge b < t_b] \wedge H \Rightarrow \ell < b)$

GLE detects a contradiction in TY.

$(\wedge_1' \ 2)$ $([\qquad\qquad] \wedge H \Rightarrow B \wedge NE)$

Proved previously $(\wedge \ 2)$.

Thus $(\wedge_1')$ is proved, and $(\wedge_1 \ 2\,LUB1\ 1\ 2)$ and $(\wedge_1 \ 2\,LUB1\ 1)$ are proved and $(T\ (NIL\ \{s2\colon \ell < s2 \underset{<}{\leq} \underset{t_\ell}{b}\}\ NIL))$ is returned.

$(\wedge_1 \ 2\,LUB1\ 2)$ $([\qquad\qquad] \wedge H \Rightarrow B \wedge NE)$

Proved previously $(\wedge \ 2)$.

Thus $(\wedge_1)$ and $(P \rightarrow L_2 \ 2)$ are proved, and hence the theorem itself is proved. Q.E.D.

Ex. 7. $(a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b) \wedge LUB \wedge L1 \wedge L2$
$\longrightarrow \exists x(f(x) \leq 0 \wedge 0 \leq f(x)))$.

This is the intermediate value theorem, where LUB is the result of substituting the set

$$\{z: z \leq b \wedge f(z) \leq 0\}$$

for the set variable in the least upper bound axiom. L_1 and L_2 are two lemmas that follow from the continuity of f. (See page 33, just before Example 5.)

(P→) $([a \leq b \wedge f(a) \leq 0 \wedge 0 \leq f(b)] \wedge LUB \wedge L_1 \wedge L_2 \Rightarrow f(x) \leq 0 \wedge 0 \leq f(x))$.

SPLIT.

(P→ 1) $([\qquad\qquad] \wedge H \Rightarrow f(x) \leq 0)$

MTY uses $f(a) \leq 0$: a/x. Fails (eventually).

MLE uses L_1: $f(x1) \leq 0$: x/x1

(P→ 1 L_1) $([\qquad\qquad] \wedge H \Rightarrow a \leq x \leq b \wedge (x \leq t_x \vee (t_x < s \wedge s \leq x \wedge f(s) \leq 0)))$

(P→ 1 L_1 1) $([\qquad\qquad] \wedge H \Rightarrow a \leq x \leq b)$

(P→ 1 L_1 1 1) $([\qquad\qquad] \wedge H \Rightarrow a \leq x)$

RLE $\{x: a \leq x < \infty\}$

(P→ 1 L_1 1 2) $([\qquad\qquad] \wedge H \Rightarrow x \leq b)$

RLE $\{x: a \leq x \leq b\}$

($P \rightarrow 1 L_1$ 1 2 CHECK)

$([a \leq b \wedge \ldots] \wedge H \Rightarrow a \leq b)$

TRUE by GLE.

($P \rightarrow 1 L_1$ 2) $([\qquad] \wedge H \Rightarrow (x \leq t_x \vee (t_x < s \wedge s \leq x \wedge f(s) \leq 0)))$

($P \rightarrow 1 L_1$ 2 o) $([\qquad] \wedge H \Rightarrow (t_x < x \rightarrow (\qquad)))$

($P \rightarrow 1 L_1$ 2 o $P\rightarrow$) $([\qquad] \wedge H \wedge t_x < x \Rightarrow (t_x < s \wedge s \leq x) \wedge f(s) \leq 0)$

Fail (eventually).

Reverse PROMOTE (Rule 6.1 of IMPLY).

($P \rightarrow 1 L_1$ 2 o $P2\rightarrow$)

$([\qquad] \wedge H \wedge \sim(t_x < s \wedge s \leq x \wedge f(s) \leq 0) \Rightarrow x \leq t_x))$

$\underbrace{(P \rightarrow 1 L_1 \text{ 2 o } P2\rightarrow)}_{\lambda}$

$([\qquad] \wedge H \wedge \underbrace{(s \leq t_x \vee x < s \vee 0 < f(s))}_{\alpha} \Rightarrow x \leq t_x)$

MLE uses LUB2: $\ell \leq yL$: ℓ/x, t_ℓ/yL

(λ LUB2) $([\qquad] \wedge H \wedge \alpha \Rightarrow (z_{t_\ell} \leq b \wedge f(z_{t_\ell}) \leq 0 \rightarrow z_{t_\ell} \leq t_\ell))$

Before it can proceed, it must check that this substitution, ℓ/x, is consistent with the current restriction on x

$$\{x:\ a \leq x \leq b\}.$$

I.e., it must check that $a \leq \ell \leq b$.

(λ CHECK) $([\qquad] \wedge H \wedge \alpha \Rightarrow a \leq \ell \wedge \ell \leq b)$

(λ CHECK 1) $\quad$ ([$\quad\quad$] $\wedge$ H $\wedge$ $\alpha \Rightarrow a \leq \ell$)

Trys GLE; fails.

$([\ldots \wedge \ell < a] \wedge H \wedge \alpha \Rightarrow a \leq \ell)$

MLE uses LUB1: $xL \leq \ell$: a/xL

$\underbrace{\text{(λ CHECK 1 LUB1)}}_{\lambda'}$ $\quad$ ([$\quad\quad$] $\wedge H \wedge \alpha \Rightarrow (a \leq b \wedge f(a) \leq 0) \wedge (B \wedge NE))$

(λ' 1) $\quad$ ([$\quad\quad$] $\wedge H \wedge \alpha \Rightarrow a \leq b \wedge f(a) \leq 0)$

(λ' 1 1) $\quad$ $([a \leq b \wedge f(a) \leq 0 \wedge \ldots] \wedge H \wedge \alpha \Rightarrow a \leq b)$

TRUE by GLE.

(λ' 1 2) $\quad$ $([a \leq b \wedge f(a) \leq 0 \wedge \ldots] \wedge H \wedge \alpha \Rightarrow f(a) \leq 0)$

TRUE by GLE.

(λ' 2) $\quad$ ([$\quad\quad$] $\wedge H \wedge \alpha \Rightarrow B \wedge NE)$

(λ' 2 1) $\quad$ ([$\quad\quad$] $\wedge H \wedge \alpha \Rightarrow (t'_u \leq b \wedge f(t'_u) \leq 0 \rightarrow t'_u \leq u)$)

(λ' 2 1 P→) $\quad$ ([$\quad\quad$] $\wedge H \wedge \alpha \wedge t'_u \leq b \wedge f(t'_u) \leq 0 \Rightarrow t'_u \leq u)$

MLE uses $t'_u \leq b$: b/u

(λ' 2 2) $\quad$ ([$\quad\quad$] $\wedge H \wedge \alpha \Rightarrow NE)$

([$\quad\quad$] $\wedge H \wedge \alpha \Rightarrow r \leq b \wedge f(r) \leq 0)$

(λ' 2 2 1) $\quad$ ([$\quad\quad$] $\wedge H \wedge \alpha \Rightarrow r \leq b)$

RLE $\quad\quad$ $\{r: -\infty < r \leq b\}$

(λ' 2 2 2) $([a \leq b \wedge f(a) \leq 0 \wedge \ldots] \wedge H \wedge \alpha \Rightarrow f(r) \leq 0)$

MTY uses $f(a) \leq 0$: a/r

But it must check that $a \leq b$.

(λ' 2 2 2 CHECK) $([a \leq b \wedge \ldots] \wedge H \wedge \alpha \Rightarrow a \leq b)$

Proved previously (λ' 1 1).

Thus (λ') = (λ CHECK 1 LUB1) and (λ CHECK 1) are proved.

(λ CHECK 2) $([\qquad\qquad] \wedge H \wedge \alpha \Rightarrow \ell \leq b)$

Trys GRE; fails.

$([\ldots \wedge b < \ell] \wedge H \wedge \alpha \Rightarrow \ell \leq b)$

MLE uses LUB2: $\ell \leq yL$: b/yL

$\underbrace{(\lambda \text{ CHECK 2 LUB2})}_{\lambda''}$ $([\qquad\qquad] \wedge H \wedge \alpha \Rightarrow (z_b \leq b \wedge f(b) \leq 0 \rightarrow z_b \leq b) \wedge (B \wedge NE))$

(λ" 1) $([\qquad\qquad] \wedge H \wedge \alpha \Rightarrow (z_b \leq b \wedge f(b) \leq 0 \rightarrow z_b \leq b))$

TRUE (promote and use GLE).

(λ" 2) $([\qquad\qquad] \wedge H \wedge \alpha \Rightarrow B \wedge NE)$

Proved previously (λ' 2).

Thus (λ CHECK) is proved, and it can now proceed with

(λ LUB2) $([\qquad\qquad] \wedge H \wedge \alpha \Rightarrow (z_{t_\ell} \leq b \wedge f(z_{t_\ell}) \leq 0 \rightarrow z_{t_\ell} \leq t_\ell))$

PROMOTE

(We will now recall the definition of α)

$(\wedge$ LUB2 P$\rightarrow)$ $([\ldots\ z_{t_\ell} \leq b \wedge f(z_{t_\ell}) \leq 0] \wedge H \wedge \underbrace{(s \leq t_x \vee x < s \vee 0 < f(s))}_{\alpha} \Rightarrow z_{t_\ell} \leq t_\ell)$

MLE uses α: $s \leq t_x$: ℓ/x, z_{t_ℓ}/s

But it must first check that $(a \leq \ell \leq b)$.

$(\wedge$ LUB2 P$\rightarrow$ CHECK) $([\qquad] \wedge H \wedge \alpha \Rightarrow a \leq \ell \wedge \ell \leq b)$

Proved previously $(\wedge$ CHECK).

Now it can proceed with the use of α on $(\wedge$ LUB2 P$\rightarrow)$

$(\wedge$ LUB2 P$\rightarrow \alpha)$ $([\qquad] \wedge H \wedge \alpha \Rightarrow z_{t_\ell} \leq \ell \wedge f(z_{t_\ell}) \leq 0)$

(Note: We have backchained on α, getting the subgoal $\sim(x < s \vee 0 < f(s))$ with the substitution $(\ell/x,\ z_{t_\ell}/s)$ applied to it.)

$(\wedge$ LUB2 P$\rightarrow \alpha$ 1) $([\qquad] \wedge H \wedge \alpha \Rightarrow z_{t_\ell} \leq \ell)$

MLE uses LUB1: $xL \leq \ell$: z_{t_ℓ}/xL

$(\wedge$ LUB2 P$\rightarrow \alpha$ 1 LUB1)

$([\ldots\ z_{t_\ell} \leq b \wedge f(z_{t_\ell}) \leq 0 \wedge \ldots] \wedge H \wedge \alpha \Rightarrow z_{t_\ell} \leq b \wedge f(z_{t_\ell}) \leq 0)$

TRUE (Split, and use GLE).

$(\wedge$ LUB2 P$\rightarrow \alpha$ 2) $([\ldots\ f(z_{t_\ell}) \leq 0 \wedge \ldots] \wedge H \wedge \alpha \Rightarrow f(z_{t_\ell}) \leq 0)$

TRUE by GLE.

Thus $(\wedge$ LUB2) and $(\wedge)$ are proved and $\theta = (\ell/x)$ is returned for $(\wedge)$.

Since $(\wedge) = (P\rightarrow 1 L_1\ 2\ 0 P2\rightarrow)$, $\theta = (\ell/x)$ is also returned for $(P\rightarrow 1)$

$(P\rightarrow 2)$ $([a \leq b \wedge f(s) \leq 0 \wedge 0 \leq f(b)] \wedge LUB \wedge L_1 \wedge L_2 \Rightarrow 0 \leq f(\ell))$

This is exactly Ex. 6. Q.E.D.

Remark

In the proof of Example 7, the following substitution units were obtained for various subgoals

$$(\ell/x, a/xL, Z_{t_\ell}/xL, b/yL, t_\ell/yL, b/u, a/r, Z_{t_\ell}/s),$$

and this list contains some apparent inconsistencies such as a/xL and Z_{t_ℓ}/xL. But there was never such an inconsistency in the proof because, as was explained at the beginning of this section, page 27, IMPLY returns only those substitution units that will be used later in the proof. Thus for example when it proved

$$(P\rightarrow 1) \qquad (f(x) \leq 0)$$

it returned only (ℓ/x) because x is the only variable which occurs in the final subgoal

$$(P\rightarrow 2) \qquad (0 \leq f(x)).$$

In fact, for this reason, throughout the proof of Example 7, the substitutions actually returned for the various subgoals were usually the "empty" substitution "T". Exceptions to this were in subgoals like $(P\rightarrow 1\ L_1\ 1\ 1)$, $(P\rightarrow 1\ L_1\ 1\ 2)$, and (Λ) where only a binding for x is returned.

In Example 6, a binding for s2 was returned for subgoal $(\Lambda_1\ 2\ \text{LUB1}\ 1\ 1)$ (because subgoal $(\Lambda_1\ 2\ \text{LUB1}\ 1\ 2)$ contains the variable s2), but this too is dropped from the return from subgoal $(\Lambda_1\ 2\ \text{LUB1}\ 1)$ because s2 does not occur in subgoal $(\Lambda_1\ 2\ \text{LUB1}\ 2)$.

Backtracking

In Ex. 7 the prover returns the binding a/x for the first subgoal

(P→ 1) $(f(x) \leq 0)$,

and then (in accordance with AND-C) tries to use that value of x in proving the second subgoal $0 \leq f(x)$. That is, it tries to prove

(P→ 2) $(0 \leq f(a))$.

This eventually fails and the prover must "backtrack" to select another value of x. It first verifies that there is a value of x which will satisfy $0 \leq f(x)$ and finds that b/x will do so. These two substitutions, a/x and b/x, are in conflict, so a/x is placed in the EXCLUDE list (which prevents it from being used again as a binding), and IMPLY is again called on the first subgoal

(P→ 1) $(f(x) \leq 0)$.

This time it returns ℓ/x and the proof continues successfully.

4. Remarks

The authors would be interested in hearing how well other provers are able to handle these examples.

One of the reasons for our success here was our use of devices such as an algebraic simplifier, to avoid the explicit use of the axioms of the real number system. Our special handling of inequalities also attempted such a saving by avoiding such axioms as the transitivity of $\leq$,

$$(x \leq y \wedge y \leq z \rightarrow x \leq z),$$

but much more needs to be done if we are to smoothly and economically handle the "low level" calculations associated with proofs in analysis. For example, special techniques are needed for absolute values, especially when they occur in connection with inequalities.

We don't think of this prover as ultimate in any sense. At best it will become a part of a more powerful prover we are now trying to build, which will include many of the features mentioned in [11]. For instance, recent work indicates that the use of automatically generated couterexamples can greatly speed up the proofs of these examples and others like them.

References

1. W.W. Bledsoe. A Maximal Method for Set Variables in Automatic Theorem Proving. The Univ. of Texas Math. Dept. Memo ATP-33A, July 1977. Proc. IJCAI-77, MIT, Aug. 1977, pp. 501-510. To appear in MI-9.

2. Robert Shostak. A Practical Decision Procedure for Arithmetic with Function Symbols. JACM, April 1979.

3. W.W. Bledsoe and Mabry Tyson. Typing and Proof by Cases in Program Verification. Machine Intelligence, 8(1977), Donald Michie and E.W. Elcock (Eds.), Ellis Horwood Limited, Chichester, pp. 30-51.

4. W.W. Bledsoe and Mabry Tyson. The UT Interactive Theorem Prover. The Univ. of Texas Math. Dept. Memo ATP-17A, June 1978.

5. W.W. Bledsoe. Splitting and Reduction Heuristics in Automatic Theorem Proving. A.I. Jour. 2(1971), pp. 55-77.

6. Greg Nelson and Derek Oppen. A Simplifier based on Efficient Decision Algorithms. Proc. 5th ACM Symp. on Principles of Programming Languages, 1978. (Also: Simplification by Cooperating Decision Procedures, Stanford A.I. Memo.)

7. Norihisa Suzuki. Verifying Programs by Algebraic and Logical Reduction. Proc. of Int'l Conf. on Reliable Software, IEEE, 1975, pp. 473-481.

8. J.C. King. A Program Verifier. Ph.D. Thesis, Carnegie-Mellon University, 1969.

9. Louis Hodes. Solving Programs by Formula Manipulation in Logic and Linear Inequality. Proc. IJCAI-71, London (1971), pp. 553-559.

10. D.C. Cooper. Programs for Mechanical Program Verification. Machine Intelligence, 6(1971), American Elsevier, New York, pp. 43-59.

11. W.W. Bledsoe. Non-Resolution Theorem Proving. AI Jour. 9(1977), 1-35.