TH∃OREM∀

A System for Supporting Mathematical Proving Implemented in Mathematica

What is TH∃OREM∀ ?

TH∃OREM∀ provides

- a front−end for composing formal mathematical text consisting of a hierarchy of axioms, definitions, propositions, algorithms etc. in a common logic frame (a version of higher order predicate logic) with user−extensiblesyntax and
- a library of well−establishedand new provers (solvers, and simplifiers) for proving (solving, and simplifying) mathematical formulae.

TH∃OREM∀ is implemented in Mathematica and, hence, can be used on all platforms on which Mathematica is available. TH∃OREM∀ heavily uses both advanced features of the Mathematica front−end and also the pattern match paradigm of the Mathematica language.

Some of the design features of TH∃OREM∀:

- TH∃OREM∀ is mainly meant for supporting the working mathematicians, computer scientists, and teachers in *routine proving*. We do not see the main value of computer−supportedproving in attacking proof problems that have resisted attempts by humans. (The notion of "routine" is, however, permanently evolving towards higher and higher levels of sophistication.)
- TH∃OREM∀ emphasizes a *theory exploration* paradigm as opposed to *an isolated theorem proving* paradigm. Hence, the convenient management of hierarchical mathematical knowledge bases is considered to be a crucial aspect of the project.
- TH∃OREM∀ is a multi−methodsystem. We think it is not realistic to tackle proving in all areas of mathematics by just one (general) proof method as, for example, resolution. Rather, a large arsenal of general and special provers must be made available from which the user can choose the appropriate method for the proof problem at hand.
- TH∃OREM∀ emphasizes the possibility of proving, solving, and simplifying and their interaction within one common logic and system frame.
- TH∃OREM∀ emphasizes the importance of attractive proof presentation styles, for example, the automatic generation of intermediate natural language explanatory text and various tools for proof browsing.

Who is TH∃OREM∀ ?

TH∃OREM∀ was designed by Bruno Buchberger and is currently developed, under his direction, by the TH∃OREM∀ Working Group at the Research Institute for Symbolic Computation (RISC), Johannes Kepler University, A4040 Linz, Austria.

Information and contact:

- www.*Theorema*.org
- Bruno.Buchberger@RISC.Uni-Linz.ac.at

Current members of the TH∃OREM∀ Working Group: B. Buchberger, T. Jebelean, F. Kriftner, T. Kutsia, K. Nakagawa, F. Piroi, J. Robu, W. Windsteiger.

A Simple Example

Let us suppose the user wants to explore the notion of limit. For this, he may start to define this notion by entering, after having loaded the TH∃OREM∀ package on top of Mathematica, the following text into an input cell

Definition["limit:", any[f, a],

$$\text{limit}[f, a] \Longleftrightarrow \underset{\epsilon>0}{\forall} \underset{N}{\exists} \underset{n\geq N}{\forall} |f[n] - a| < \epsilon$$

]

(Various tools are available for facilitating typing this text.) Note that keywords like 'Definition' is auxiliary text and strings in quotation marks are considered to be labels that are not part of the actual logical formulae. The 'any' construct declares the free variables in the formula.

Then he may enter propositions he conjectures, for example the proposition

Proposition["limit of sum", any[f, a, g, b],

$$(\text{limit}[f, a] \wedge \text{limit}[g, b]) \Longrightarrow \text{limit}[f + g, a + b]]$$

Now he may compose an appropriate knowledge base from which he thinks the proposition(s) can be proved, for example the following formulae.

Definition["+:", any[f, g, x],
$(f + g)[x] = f[x] + g[x]]$

Lemma["|+|", any[x, y, a, b, δ, ϵ],
$(|(x + y) - (a + b)| < (\delta + \epsilon)) \quad \Longleftarrow \quad (|x - a| < \delta \wedge |y - b| < \epsilon)]$

Lemma["max", any[m, M1, M2],
$m \geq \max[M1, M2] \quad \Rightarrow \quad (m \geq M1 \wedge m \geq M2)]$

Theory["limit",
Definition["limit"]
Definition["+:"]
Lemma["|+|"]
Lemma["max"]
]

Note that the Theory construct enables the user to compose knowledge bases from axioms, definitions, lemmas etc. but also from other theories so that the hierarchical composition of well−structured large knowledge bases is possible.

For trying to prove the proposition from the knowledge base, the only thing the user has to do is to enter the command

Prove[Proposition["limit of sum"], using → Theory["limit"], by → PCS]

'PCS' is the name of one of the automated provers currently implemented in TH∃OREM∀. It is based on a proof method that has recently be proposed by B. Buchberger for proving typical formulae in analysis. One of the ideas in this method is the reduction of the 'prove' problem to a 'solve' problem over the reals, which can be solved by advanced computer algebra methods. Note that the proof generated produces also explicit information on the construction of, for example, the bound 'N' in the goal formula, which is interesting from the mathematical, algorithmic, as well as didactic aspect.

After entering the above Prove call, in a few seconds, TH∃OREM∀ will generate the proof shown below. Note that every individual symbol in this proof is generated completely automatically wihtout any user interaction, i.e. both the logical formulae appearing in the proof and all explanatory intermediate text including the labels are generated by the system. At the beginning of the proof text, the proof goal and the knowledge base used is echoed.

Prove:

(Proposition (limit of sum)) $\forall_{f,a,g,b} (\text{limit}[f, a] \wedge \text{limit}[g, b] \Rightarrow \text{limit}[f + g, a + b])$,

under the assumptions:

(Definition (limit)) $\forall_{f,a} \left(\text{limit}[f, a] \Leftrightarrow \forall_{\epsilon > 0} \exists_{N} \forall_{n \geq N} (|f[n] - a| < \epsilon) \right)$,

(Definition (+:)) $\forall_{f,g,x} ((f + g)[x] = f[x] + g[x])$,

(Lemma (|+|)) $\forall_{x,y,a,b,\delta,\epsilon} (|(x + y) - (a + b)| < \delta + \epsilon \Leftarrow (|x - a| < \delta \wedge |y - b| < \epsilon))$,

(Lemma (max)) $\forall_{m,M1,M2} (m \geq \max[M1, M2] \Rightarrow m \geq M1 \wedge m \geq M2)$.

We assume

(1) $\text{limit}[f_0, a_0] \wedge \text{limit}[g_0, b_0]$,

and show

(2) $\text{limit}[f_0 + g_0, a_0 + b_0]$.

Formula (1.1), by (Definition (limit)), implies:

(3) $\forall_{\epsilon > 0} \exists_{N} \forall_{n \geq N} (|f_0[n] - a_0| < \epsilon)$.

By (3), we can take an appropriate Skolem function such that

(4) $\forall_{\epsilon>0} \forall_{n\geq N_0[\epsilon]} (|f_0[n]-a_0|<\epsilon),$

Formula (1.2), by (Definition (limit)), implies:

(5) $\forall_{\epsilon>0} \exists_{N} \forall_{n\geq N} (|g_0[n]-b_0|<\epsilon).$

By (5), we can take an appropriate Skolem function such that

(6) $\forall_{\epsilon>0} \forall_{n\geq N_1[\epsilon]} (|g_0[n]-b_0|<\epsilon),$

Formula (2), using (Definition (limit)), is implied by:

(7) $\forall_{\epsilon>0} \exists_{N} \forall_{n\geq N} (|(f_0+g_0)[n]-(a_0+b_0)|<\epsilon).$

We assume

(8) $\epsilon_0>0,$

and show

(9) $\exists_{N} \forall_{n\geq N} (|(f_0+g_0)[n]-(a_0+b_0)|<\epsilon_0).$

We have to find N_2^* such that

(10) $\forall_{n} (n\geq \mathrm{N}_2^* \Rightarrow |(f_0+g_0)[n]-(a_0+b_0)|<\epsilon_0).$

Formula (10), using (Definition (+:)), is implied by:

(11) $\forall_{n} (n\geq \mathrm{N}_2^* \Rightarrow |(f_0[n]+g_0[n])-(a_0+b_0)|<\epsilon_0).$

Formula (11), using (Lemma (|+|)), is implied by:

(12) $\exists_{\delta+\epsilon=\epsilon_0} \forall_{n} (n\geq \mathrm{N}_2^* \Rightarrow |f_0[n]-a_0|<\delta \wedge |g_0[n]-b_0|<\epsilon).$

We have to find δ_0^*, ϵ_1^* and N_2^* such that

(13) $(\delta_0^*+\epsilon_1^*=\epsilon_0) \bigwedge \forall_{n} (n\geq \mathrm{N}_2^* \Rightarrow |f_0[n]-a_0|<\delta_0^* \wedge |g_0[n]-b_0|<\epsilon_1^*).$

Formula (13), using (6), is implied by:

$(\delta_0^*+\epsilon_1^*=\epsilon_0) \bigwedge \forall_{n} (n\geq \mathrm{N}_2^* \Rightarrow |f_0[n]-a_0|<\delta_0^* \wedge (\epsilon_1^*>0 \wedge n\geq N_1[\epsilon_1^*])),$

which, using (4), is implied by:

$(\delta_0^*+\epsilon_1^*=\epsilon_0) \bigwedge \forall_{n} (n\geq \mathrm{N}_2^* \Rightarrow (\delta_0^*>0 \wedge n\geq N_0[\delta_0^*]) \wedge (\epsilon_1^*>0 \wedge n\geq N_1[\epsilon_1^*])),$

which, using (Lemma (max)), is implied by:

(14) $(\delta_0^* + \epsilon_1^* = \epsilon_0) \bigwedge \forall_n (n \geq \mathrm{N}_2^* \Rightarrow \delta_0^* > 0 \bigwedge \epsilon_1^* > 0 \bigwedge n \geq \max[N_0[\delta_0^*], N_1[\epsilon_1^*]])$.

Formula (14) is implied by

(15) $(\delta_0^* + \epsilon_1^* = \epsilon_0) \bigwedge \delta_0^* > 0 \bigwedge \epsilon_1^* > 0 \bigwedge \forall_n (n \geq \mathrm{N}_2^* \Rightarrow n \geq \max[N_0[\delta_0^*], N_1[\epsilon_1^*]])$.

Partially solving it, formula (15) is implied by

(16) $(\delta_0^* + \epsilon_1^* = \epsilon_0) \bigwedge \delta_0^* > 0 \bigwedge \epsilon_1^* > 0 \bigwedge (\mathrm{N}_2^* = \max[N_0[\delta_0^*], N_1[\epsilon_1^*]])$.

Now,

$(\delta_0^* + \epsilon_1^* = \epsilon_0) \bigwedge \delta_0^* > 0 \bigwedge \epsilon_1^* > 0$

can be solved for δ_0^* and ϵ_1^* by a call to Collins cad-method yielding a solution

$\delta_0^* \leftarrow \frac{1}{2} * \epsilon_0$,

$\epsilon_1^* \leftarrow \frac{1}{2} * \epsilon_0$.

Let us take

$\mathrm{N}_2^* \leftarrow \max\left[N_0\left[\frac{1}{2} * \epsilon_0\right], N_1\left[\frac{1}{2} * \epsilon_0\right]\right]$.

Formula (16) is solved. Hence, we are done.