MECHANIZED MATHEMATICS AND ITS APPLICATIONS, VOL. 6, NO. 1, NOVEMBER 2007, 47:54

Towards a Coherent Repository of Knowledge ⋆

Adam Grabowski

Institute of Mathematics
University of Białystok
ul. Akademicka 2, 15-267, Poland
adam@math.uwb.edu.pl

Abstract – In this paper, we describe the development of the Mizar Mathematical Library used to formally prove the Jordan Curve Theorem. More general issues of knowledge reusability are also raised with special attention paid to the logical equivalence between some formal apparatus of topology and real analysis.

Keywords – homotopy theory, Jordan Curve Theorem, formalization of mathematics.

1. Introduction

During the past five years, the current state of the development and usefulness of mathematical proof assistants seems to be measured in the most popular way by formulating great challenges for the proof encoding. Obviously, it does not mean that significant formalization projects (or proofs given with the help of computer) are that new.

As examples of such classic works we can point out the mechanization of Landau's *Grundlagen der Analysis* done in de Bruijn's AUTOMATH system by Jutting in the 1970s or the proof of the Four Colour Theorem formalized in Coq by Werner and Gonthier in 2004. The famous proof of the Robbins conjecture about the new short axiomatic base for Boolean algebras done with the EQP/Otter theorem prover by McCune could convince many people that the computer can solve arbitrarily hard mathematical problems but such a judgment is obviously false. However the role of machine checking for correctness of mathematical proofs is growing (see for example the "Formal Proof of Kepler" project by Hales) and it is quite natural to expect such software to face some real-life problems.

The Jordan Curve Theorem (JCT) [19] was selected as one such challenge for math-assistants; its formal proof is rather complicated and its encoding is a testbed both for the expressive power of the formalization language and for the quality of the database of as yet formalized knowledge. On the other hand, the abilities of the reasoner were also examined, and one can view the development of the proof of JCT as the source of many benefits.

2. Development of the Library

As the starting point of the Mizar Mathematical Library the date of acceptance of the article by Święczkowska and Trybulec *Boolean properties of sets* is claimed (January 1, 1989). This

⋆ This work has been partially supported by the FP6 IST project no. 510996 *Types for Proofs and Programs* (TYPES).

Manuscript received August 23, 2007; revised October 31, 2007.

year and the following year were most fruitful and influential for the development of the library as a whole and the formalization of general topology started shortly thereafter. This choice was by no means accidental, since the MML is based on set theory (Andrzej Trybulec, the designer of Mizar, PhD in topology), and our interest put just there is not surprising then. We were responsible for two parts of the proof of JCT: one of them, contained mainly in [10], is a direct translation of a fragment of the proof sketch in [19]. This work was done during our stay in Nagano in 1997; the other, auxiliary parts for it are from the `JORDAN5` and `BORSUK` series: [11], [9], [17], and [3]. Our other activity leading to the completion of the proof of JCT was the development of algebraic topology within the MML ([3], [15]) described in more detail in the next section.

The most recent work in the topic was the authorship of some `TOPGEN` articles aiming at formalizing Kuratowski's monograph *General Topology*, but this did not have any influence on JCT.

3. Formalizing Topology via Real Analysis

As one of our first developments (although `TOPRNS_1` which is also of some interest here, was the first Mizar article of the author) towards formalizing the proof of JCT, we translated a part of the proof sketch by Yatsuka Nakamura based on [19] and started to study the parts of the MML which formed a base for this formalization. However, we paid also special attention to these few articles which had very little in common with this Mizar sketch at first sight, at least if we count directives given in the environment declaration. The aim was to prove that all continuous functions possesing some additional properties are monotone.

3.1 The Continuity of Real Functions

We focus here on articles from the `FCONT` and `FDIFF` series, devoted mainly to continuous functions and the differentiation operator, respectively; all functions act on $\mathbb{R}$ (`REAL` in the Mizar notation). A sort of the natural topology of the real line was also defined, but the notion of structure was not used at all (as of now, all these articles are placed in the concrete part of the MML, a counterpart of the abstract one, i.e., dealing with the notion of structure). Because JCT-articles are placed mainly within the abstract part, where the approach to the topology is totally different, the results obtained in `FCONT` seemed to be useless.

```
theorem :: FCONT_2:18
  f is one-to-one & p <= g & f is_continuous_on [.p,g.] implies
    f is_increasing_on [.p,g.] or f is_decreasing_on [.p,g.];
```

Although the primary definition of the continuity is given in Heine style, its Cauchy (epsilon-delta) equivalent form was also already present in the MML; we found however another lemma even more useful (neighbourhoods, similarly as open sets, are defined as intervals with non-empty radius):

```
theorem :: FCONT_1:5
```

```
  for f,x0 holds f is_continuous_in x0 iff x0 in dom f &
  for N1 being Neighbourhood of f.x0
   ex N being Neighbourhood of x0 st f.:N c= N1;
```

For the details of the formalization we refer to [18] and [16].

3.2 General Topology in Mizar

As the backbone notion of the general topology formalized in the MML is the structure of the topological space, all “topological" articles are in the abstract part of the library. We do not describe here very basic notions, they can be found in [7].

For the JCT, the most important notions are as follows:

- the predicate `is_homeomorphism`, i.e., the bijective function between two topological spaces, which is continuous as well as its inverse is;
- the predicate `is_an_arc` which is defined for a subset of the Euclidean topological plane P as shown below.

```
  pred P is_an_arc_of p1,p2 means
:: TOPREAL1:def 2
    ex f being Function of I[01], T|P st
              f is_homeomorphism & f.0 = p1 & f.1 = p2;
    ex f being Function of I[01], T|P st
       f is_homeomorphism & f.0 = p1 & f.1 = p2;
```

3.3 The Correspondence

In the `JORDAN5` series we formalized a part of the proof sketch of JCT based on [19]. At the very beginning the notion of `First_Point` had to be defined, i.e., the point of an arc P in which P intersects a subset of a Euclidean plane Q for the first time (with respect to the natural parametrization from 0 to 1). Of course, the definition of the dual `Last_Point` afterwards was easy.

```
theorem :: JORDAN5A:23
   for P, Q being Subset of TOP-REAL 2,
      p1, p2 being Point of TOP-REAL 2 st
  P meets Q & P /\ Q is closed & P is_an_arc_of p1,p2 holds
   ex EX be Point of TOP-REAL 2 st
    ( EX in P /\ Q &
     ex g being Function of I[01], (TOP-REAL 2)|P, s2 being Real st
      g is_homeomorphism & g.0 = p1 & g.1 = p2 &
       g.s2 = EX & 0 <= s2 & s2 <= 1 &
        (for t being Real st 1 >= t & t > s2 holds not g.t in Q));
```

For our development in [11], the crucial point was the proof of the fact that the notions of an open set with respect to the topology and being an open set in the set of the real line coincide. Also, since continuity is defined in terms of open sets, we can relatively easily extend this equivalence into some other notions.

To summarize, we have shown the following correspondence:

- `X is open Subset of REAL` vs. `X in Family_open_set RealSpace`
- `f is continuous Function of R^1, R^1` vs. `f is_continuous_on REAL`
- the equivalence of closed subsets of `REAL` and `closed Subset of R^1`

In some sense, one could try to duplicate the work from [16], but as we soon noticed it was written in rather old style, we applied a way which looked more elegant.

```
definition let PM be MetrStruct;
  func Family_open_set PM -> Subset-Family of PM means
:: PCOMPS_1:def 5
  for V holds V in it iff for x st x in V holds
      ex r st r>0 & Ball(x,r) c= V;
end;
```

We should recall here that such neighbourhoods `Ball` are given in a natural way with the radius bounding the distance between the center and the remaining point.

4. Algebraic Topology

Algebraic topology [13], [12], developed in `TOPALG` series [15], is strongly connected with group theory and with general topology. Both disciplines were already widely formalized within the MML, so the choice of this topic was rather natural.

4.1 Paths

Let us start with the very basic, although important notion, of a continuous function from the unit interval into a given topological space, which satisfies some additional properties:

```
definition let T be TopStruct; let a, b be Point of T;
  assume a, b are_connected;
  mode Path of a, b -> Function of I[01], T means
:: BORSUK_2:def 2
  it is continuous & it.0 = a & it.1 = b;
end;
```

The assumption allowing us to show that an object stated in the definiens exists, is of course needed to prove the correctness of the definition. The definition itself is permissive – it states that if the mapping connecting two points, say a and b, exists, we call it a `Path`, otherwise the definiens is not accessible, even if the Mizar analyser accepts the type `Path of a, b` with no errors reported.

```
definition let T be arcwise_connected TopStruct;
  let a, b be Point of T;
  redefine mode Path of a, b means
:: BORSUK_2:def 4
  it is continuous & it.0 = a & it.1 = b;
end;
```

Although the earlier definition is of a slightly more general type (i.e., it assumes the existence of the mapping connecting only two fixed points), every time that the path returns proper values on its limits should be justified via the predicate `are_connected` which seems to be too high a price for that minor generalization. Hence if the considered space possesses the property of being `arcwise_connected` (which truly speaking should be rather *pathwise connected*), paths behave as expected (e.g., the real Euclidean line or the unit interval has this adjective added automatically to its type) and the reasoning simplifies as the user can forget about the assumption from the original definition of a path.

4.2 Homotopies

A very similar trick to the aforementioned, i.e., first guaranteeing that the desired object exists, then proving the correctness of the appropriate functor definition, was applied to the definition of a homotopy.

```
definition let T be non empty TopStruct;
  let a, b be Point of T;
  let P, Q be Path of a, b;
  pred P, Q are_homotopic means
:: BORSUK_2:def 7
  ex f being Function of [:I[01],I[01]:], T st
    f is continuous &
    for s being Point of I[01] holds f.(s,0) = P.s & f.(s,1) = Q.s &
    for t being Point of I[01] holds f.(0,t) = a & f.(1,t) = b;
  symmetry;
end;
```

Under this assumption the existence of a continuous deformation can be proven; of course not its uniqueness, hence the definition of the `mode` and not the functor.

```
definition let T be non empty TopSpace;
           let a, b be Point of T;
           let P, Q be Path of a, b;
  assume P, Q are_homotopic;
  mode Homotopy of P, Q -> Function of [:I[01],I[01]:], T means
:: BORSUK_6:def 13
  it is continuous &
  for s being Point of I[01] holds it.(s,0) = P.s & it.(s,1) = Q.s &
```

```
  for t being Point of I[01] holds it.(0,t) = a & it.(1,t) = b;
end;
```

Homotopies [8] were used to define the fundamental group later in [15] and were also effectively used in the proof of the Brouwer Fixed Point Theorem [14] directly leading to the final proof of JCT.

5. Knowledge Reusability

Of course, the proof of JCT strongly depends on the state of the MML and strictly reuses the facts available there. On the contrary, to formalize the Fundamental Theorem of Algebra within Coq, for example, a special repository (C-CoRN) was created from scratch. The quantitative data about the MML, which is considered the largest repository of computerized mathematical knowledge, besides obvious advantages, brings also certain inconveniences.

In this repository, among nearly a thousand files, there is no unique, coherent style of writing. The authors (and there are about 150 of them as of now) usually follow the pattern of some previous submissions by themselves – or they learn from others. However, many of the articles written in the previous century (just to recall, the MML has over 15 years of growth) were created in a quite different language and due to revisions have sometimes very little in common with their original.

Furthermore, some of the authors do not care about the generality of theorems. Because of that, e.g., unnecessary assumptions should be added to all descendant theorems which use such restrictive facts. The sooner underlying revisions are performed, the fewer unnecessary assumptions remain for authors.

Also different, parallel approaches can cause certain troubles. The perfect example here could be the formalization of category theory, in two distinct approaches, done by Byliński and, more recently, Trybulec. Potential users are free to choose which approach to use, but then we can obtain a duplication of results.

However, luckily, as it seems, the formalization of JCT brought more advantages. As the most important one we can point out the encoding of results of a more general interest, such as the following of the present author: the theorems about subcontinua of a real line and of simple closed curves [5], [6], the compactness of the binary product of topological spaces (restricted version of the Tychonoff theorem) [3], equivalence of analysis and topology [11], and the closure-complement problem of fourteen Kuratowski sets [1]. Many other important results were also developed by other members of the JCT team, e.g., Fashoda Meet Theorem, Tietze Extension Theorem, and Brouwer Fixed Point Theorem.

Unfortunately, it seems that the formal proof of JCT in Mizar took a long time because theories were developed instead of narrow sets of lemmas which were actually used. From the point of view of library developers it is quite natural, but the latter approach is more challenge-oriented.

6. Merging Structures

The method of showing the correspondence of theories described in Section 3, although much better (but usually not easier) than proving everything from scratch without any direct link to the existing implementation, does not fully benefit from the expressive power of the Mizar language.

For example, the functor introduced in [2]:

```
definition let PM be MetrStruct;
  func TopSpaceMetr PM -> TopStruct equals
:: PCOMPS_1:def 6
     TopStruct (# the carrier of PM, Family_open_set PM #);
end;
```

transforms metric space structure into topological space forgetting in some sense about the original metric (distance function). This resulted in at least two parallel paths of proving lemmas and caused some technical difficulties with the overloading of symbols: although the Jordan theorem is formulated in terms of the Euclidean plane viewed as a topological space (`TOP-REAL n` which is equal to `TopSpaceMetr Euclid n`), still many lemmas only for the Euclidean metric space were available thus creating a need for translating between these two approaches. From this point of view, the equivalence of topological and analytical notions could also be expressed better.

It seems more feasible to define a new structure, e.g.,

```
definition
  struct (TopStruct, MetrStruct) TopMetrStruct
        (# carrier -> set,
          topology -> Subset-Family of the carrier
          distance -> Function of [:the carrier,the carrier:],REAL #);
end;
```

which has the fields of its parents `TopStruct` and `MetrStruct` inherited. Then we can define an adjective stating that the distance of such a merged structure is measured according to the usual Euclidean metric and the topology is just its `Family_open_set`. Afterwards, `TOP-REAL n` could be a both topological and metric space.

We hope the Library Committee of the Association of Mizar Users will consider such improvements and revise the whole proof of JCT in the nearest future.

7. Conclusions

Verification of mathematical proofs with the help of computer assistants seems to be more and more important nowadays. As the Jordan Curve Theorem is one of the best known problems relatively easily formulated, but with complicated proofs, the completion of the encoding of the proof of JCT into the Mizar formalism was a great achievement for its developers. Also for the library the project was a big step towards better coherence of the MML. Due to the high level of generality wherever possible and the continuous process of revisions leading to the enhancement of the database, the net of notions and theorems created when proving JCT can be reused later not only in the topology itself, but also in the other parts of mathematics in future developments.

Acknowledgments

I am greatly indebted to Professors Yatsuka Nakamura and Yasunari Shidama of Shinshu University for their help during my stay in Japan. My gratitude goes also to Professor Andrzej Trybulec for acquainting me with the Mizar system. I could not go that far with some critical points of [4] also without constant motivation from Artur Korniłowicz's side.

References

[1] L. Bagińska and A. Grabowski, *On the Kuratowski closure-complement problem*, Formalized Mathematics, 11(**3**), 2003, pp.323–329.
[2] L. Borys, *Paracompact and metrizable spaces*, Formalized Mathematics, 2(**4**), 1991, pp.481–485.
[3] A. Grabowski, *Introduction to the homotopy theory*, Formalized Mathematics, 6(**4**), 1997, pp.449–454.
[4] A. Grabowski, *Properties of the product of compact topological spaces*, Formalized Mathematics, 8(**1**), 1999, pp.55–59.
[5] A. Grabowski, *On the decompositions of intervals and simple closed curves*, Formalized Mathematics, 10(**3**), 2002, pp.145–151.
[6] A. Grabowski, *On the subcontinua of a real line*, Formalized Mathematics, 11(**3**), 2003, pp.313–322.
[7] A. Grabowski, *Solving two problems in general topology via types*, In J.-Ch. Filliâtre, Ch. Paulin-Mohring, and B. Werner, editors, Proceedings of International Workshop, TYPES 2004, LNCS 3839, Springer, 2006, pp.138–153.
[8] A. Grabowski and A. Korniłowicz, *Computer-assisted reasoning about algebraic topology*, In R. Matuszewski and A. Zalewska, editors, From Insight to Proof: Festschrift in Honour of Andrzej Trybulec, 10(**23**) of Studies in Logic, Grammar and Rhetoric, University of Białystok, 2007, pp.179–189.
[9] A. Grabowski and Y. Nakamura, *The ordering of points on a curve. Part I*, Formalized Mathematics, 6(**4**), 1997, pp.461–465.
[10] A. Grabowski and Y. Nakamura, *The ordering of points on a curve. Part II*, Formalized Mathematics, 6(**4**), 1997, pp.467–473.
[11] A. Grabowski and Y. Nakamura, *Some properties of real maps*, Formalized Mathematics, 6(**4**), 1997, pp.455–459.
[12] M. Greenberg, *Lectures on Algebraic Topology*. W. A. Benjamin, Inc., 1973.
[13] A. Hatcher, *Algebraic Topology*. Cambridge University Press, 2002.
[14] A. Korniłowicz and Y. Shidama, *Brouwer fixed point theorem for disks on the plane*, Formalized Mathematics, 13(**2**), 2005, pp.333–336.
[15] A. Korniłowicz, Y. Shidama, and A. Grabowski, *The fundamental group*, Formalized Mathematics, 12(**3**), 2004, pp.261–268.
[16] J. Kotowicz and K. Raczkowski. *Real function uniform continuity*, Formalized Mathematics, 1(**4**), 1990, pp.793–795.
[17] Y. Nakamura and A. Grabowski, *Bounding boxes for special sequences in* $\mathcal{E}^2$, Formalized Mathematics, 7(**1**), 1998, pp.115–121.
[18] K. Raczkowski and P. Sadowski, *Real function continuity*, Formalized Mathematics, 1(**4**), 1990, pp.787–791.
[19] Y. Takeuchi and Y. Nakamura, *On the Jordan curve theorem*, Technical Report 19804, Dept. of Information Eng., Shinshu University, 500 Wakasato, Nagano city, Japan, April 1980.