Formalisation of Algebraic Topology: a report

Julio Rubio

Universidad de La Rioja
Departamento de Matemáticas y Computación

MAP 2012

Konstanz (Germany), September 17th-21th, 2012

Partially supported by Ministerio de Educación y Ciencia, project MTM2009-13842-C02-01, and
by European Commission FP7, STREP project ForMath, n. 243847.

Formalizing mathematics: the European Project ForMath

- European Commission FP7, STREP project ForMath: 2010-2013
- Objective: formalized libraries for mathematical algorithms.
- Four nodes:
  - Gothenburg University: Thierry Coquand, leader.
  - Radboud University.
  - INRIA.
  - Universidad de La Rioja.

Status of ForMath

- Four Work Packages:
  - Infrastructure to formalize mathematics in constructive type theory.
    - `SSReflect` extension of Coq.
      Gonthier's library created for the Four Color Theorem.
      Now extended and applied to simple finite group classification.
    - Mixing deduction and computation, Big-Op library, ...
  - Linear Algebra library.
    - Verified and efficient matrix manipulation.
    - Coherent and strongly discrete rings in type theory.
  - Real numbers and differential equations.
    - Verified and efficient reals in Coq.
    - Numerical integration, Simpson's rule, Newton method, ...
  - Algebraic topology and... (medical) image processing.
- Why formalizing mathematics?

Summary

- Computer-based mathematical error detection.
- Essential building blocks.
  - Eilenberg-Zilber (EZ) theorem.
  - Basic Perturbation Lemma (BPL).
- Formalisation of the EZ theorem.
- Formalisation of the BPL.
- Discrete vector fields.
- Biomedical image processing.
- Formalisation of homological computing.
- Interoperability.
- Persistent homology.
- Another mathematical error.
- Conclusions and further work.

A published "theorem"

Theorem 5.4: *Let* A_4 *be the* 4-*th alternating group.*
Then $\pi_4(\Sigma K(A_4, 1)) = \mathbb{Z}_4$

"On homotopy groups of the suspended classifying spaces".
Algebraic and Geometric Topology 10 (2010) 565-625.

- A_4 = 4-th alternating group.
- $K(A_4, 1)$ = Eilenberg-MacLane space.
- Σ = Suspension.
- $\pi_4()$ = 4-th homotopy group.
- $\mathbb{Z}_4$ = cyclic group with 4 elements.

A computer calculation

After some previous definitions, we define in Kenzo the alternate group A_4:

```
> (setf A4 (group1 (tcc rsltn))) ; rsltn = resolution
[K1 Group]
```

It is a group with *effective homology* (Ana Romero's programs):

```
> (setf (slot-value A4 'resolution) rsltn)
[K10 Reduction K2 => K5]
```

We apply the classifying construction, obtaining $K(A_4, 1)$:

```
> (setf k-A4-1 (k-g-1 A4))
[K11 Simplicial-Group]
```

We apply the suspension construction, obtaining $\Sigma K(A_4, 1)$:

```
> (setf s-k-A4-1 (suspension k-A4-1))
[K23 Simplicial-Set]
```

And finally we compute the controversial homotopy group:

```
> (homotopy s-k-A4-1 4)
Homotopy in dimension 4 :
   Component Z/4Z
   Component Z/3Z
```

Anatomy of a calculation

- In this particular case, Kenzo was right and the mathematical text wrong.
- In general?
- Increasing trust: formal verification of (part of) (the algorithms supporting) the programs.
- $\pi_4(\Sigma K(A_4, 1)) = H_4(K_4)$.
- A homotopy group is computed as a homology group of an space K_4.
- K_4 is the total space of a fibration: $K(\mathbb{Z}_6, 2) \to K_4 \to K_3$.
- ($\mathbb{Z}_6 = H_3(K_3) = \pi_3(\Sigma K(A_4, 1))$.)
- $K_4 = K(\mathbb{Z}_6, 2) \times_\tau K_3$ (twisted Cartesian product).
- The (effective) homology of $K(\mathbb{Z}_6, 2)$ and K_3 are known.
- An effective version of the Serre spectral sequence is needed.

Reductions

- Given two chain complexes $C := \{(C_n, d_n)\}_{n\in\mathbb{Z}}$ and $C' := \{(C'_n, d'_n)\}_{n\in\mathbb{Z}}$ a *reduction* between them is (f, g, h) where
  - $f : C \to C'$ and $g : C' \to C$ are chain morphisms
  - and h is a family of homomorphisms (called *homotopy operator*) $h_n : C_n \to C_{n+1}$.

  satisfying

  1. $f \circ g = 1$
  2. $d \circ h + h \circ d + g \circ f = 1$
  3. $f \circ h = 0$
  4. $h \circ g = 0$
  5. $h \circ h = 0$
- If $(f, g, h) : C \Longrightarrow C'$ is a reduction, then $H(C) \cong H(C')$.
- *Theorem*: From $A \Longrightarrow A'$ and $B \Longrightarrow B'$, an algorithm constructs $A \otimes B \Longrightarrow A' \otimes B'$.
- *Corollary*: If A and B are with effective homology, then $A \otimes B$ is with effective homology.

Essential building blocks

- *Eilenberg-Zilber Theorem*: $C(F \times B) \Longrightarrow C(F) \otimes C(B)$.
- It is the case of a trivial fibration: $F \rightarrow F \times B \rightarrow B$.
- What about the general (twisted) case? $F \rightarrow F \times_\tau B \rightarrow B$.
- Then?
- Given a chain complex (C, d), a *perturbation* for it is a family ρ of group homomorphisms $\rho_n : C_n \rightarrow C_{n-1}$ such that $(C, d + \rho)$ is again a chain complex (that is to say: $(d + \rho) \circ (d + \rho) = 0$).
- *Basic Perturbation Lemma*: Let $(f, g, h) : (C, d) \Longrightarrow (C', d')$ be a reduction and be ρ a perturbation for (C, d) which are locally nilpotent. Then there exists a reduction $(f_\infty, g_\infty, h_\infty) : (C, d + \rho) \Longrightarrow (C', d'_\infty)$.

Putting all together

- Given a fibration $F \to F \times_\tau B \to B$ where
  - F and B are with effective homology (known reductions $C(F) \Longrightarrow HF$ and $C(B) \Longrightarrow HB$) and
  - B is simply connected.
- EZ application: $C(F \times B) \Longrightarrow C(F) \otimes C(B)$.
- BPL application: $C(F \times_\tau B) \Longrightarrow C(F) \otimes_t C(B)$.
- Tensor product application: $C(F) \otimes C(B) \Longrightarrow HF \otimes HB$.
- BPL application (B simply connected):
  $C(F) \otimes_t C(B) \Longrightarrow HF \otimes_{t'} HB$
- Composing it all: $C(F \times_\tau B) \Longrightarrow HF \otimes_{t'} HB$.
- *Conclusion*: The total space $F \times_\tau B$ is with effective homology.

Statement of the EZ theorem

- $(f, g, h) : C(F \times B) \Longrightarrow C(F) \otimes C(B)$
  - $f = AW$ (Alexander-Whitney)
    $AW(x_n, y_n) = \sum_{i=0}^{n} \partial_{i+1} \dots \partial_n x_n \otimes \partial_0 \dots \partial_{i-1} y_n$
  - $g = EML$ (Eilenberg-MacLane)
    $EML(x_p \otimes y_q) =$
    $\sum_{(\alpha,\beta)\in\{(p,q)\text{-shuffles}\}} (-1)^{sg(\alpha,\beta)} (\eta_{\beta_q} \dots \eta_{\beta_1} x_p, \eta_{\alpha_p} \dots \eta_{\alpha_1} y_q)$
  - $h = SHI$ (Shih)
    $SHI(x_n, y_n) =$
    $\sum (-1)^{n-p-q+sg(\alpha,\beta)} (\eta_{\beta_q+n-p-q} \dots \eta_{\beta_1+n-p-q} \eta_{n-p-q-1} \partial_{n-q+1} \dots \partial_n x_n,$

    $\eta_{\alpha_{p+1}+n-p-q} \dots \eta_{\alpha_1+n-p-q} \partial_{n-p-q} \dots \partial_{n-q-1} y_n).$
- where a (p, q)-shuffle $(\alpha, \beta) = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$ is a permutation of the set $\{0, 1, \dots, p+q-1\}$ such that $\alpha_i < \alpha_{i+1}$ and $\beta_j < \beta_{j+1}$.
- EZ is responsible of much of the exponential behaviour of Kenzo.
- It is essentially unique (so unavoidable).
- The formulas are very well-structured and of combinatorial nature.

Formalisation of the EZ theorem

- A proof purely based on induction + rewriting.
- The ACL2 theorem prover is the right tool for the task.
- Main conceptual tool: *simplicial polynomials*.
- It allows one to enhance ACL2 with *algebraic rewriting*.
- Already used in the proof of the *Normalisation Theorem*.
  - $C^D(K) \Longrightarrow C(K)$.
  - L. Lambán, F. J. Martín-Mateos, J. R., J. L. Ruiz-Reina. "Formalization of a normalization theorem in simplicial topology". Annals of Mathematics and Artificial Intelligence 64 (2012) 1-37.
- EZ formalisation by the same team, with proving effort
  - EZ: 13000 lines.
  - Normalisation: 4500 lines.
  - Common infrastructure: 6000 lines.

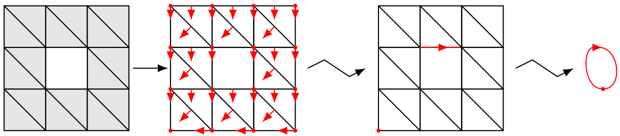
Statement of the BPL

- Let $(f, g, h)\colon (D, d_D) \Longrightarrow (C, d_C)$ be a reduction and $\rho_D\colon D \to D$ a perturbation of the differential d_D satisfying the local nilpotency condition with respect to the reduction (f, g, h). Then, a new reduction $(f', g', h')\colon (D', d_{D'}) \Longrightarrow (C', d_{C'})$ can be obtained, where the underlying graded groups D and D' (resp. C and C') are the same, but the differentials are perturbed: $d_{D'} = d_D + \rho_D$, $d_{C'} = d_C + \rho_C$, where $\rho_C = f\rho_D\psi g$; $f' = f\phi$; $g' = \psi g$; $h' = h\phi$, where $\phi = \sum_{i=0}^{\infty}(-1)^i(\rho_D h)^i$, and $\psi = \sum_{i=0}^{\infty}(-1)^i(h\rho_D)^i$.
- Note the role of the series.
- The graded groups are general (with infinitely many generators, for instance).
- No combinatorial approach possible.

Formalisation of the BPL

- Isabelle/HOL formalisation:
  - J. Aransay, C. Ballarin, J. R.
    "A mechanized proof of the Basic Perturbation Lemma".
    Journal of Automated Reasoning 40 (2008) 271-293.
  - General statement. Ungraded case. General groups (not effective).
- Coq formalisation:
  - C. Domínguez, J. R.
    "Effective homology of bicomplexes, formalized in Coq".
    Theoretical Computer Science 412 (2011) 962-970.
  - Bicomplexes only. Graded case. Locally effective and effective groups.
- SSReflect formalisation:
  - C. Domínguez, J. Heras, M. Poza, J. R.
  - General statement. Graded case. Only finitely generated groups.
  - Based on a shorter and brand new proof by:
    A. Romero, F. Sergeraert. "Discrete Vector Fields and Fundamental Algebraic Topology". ArXiv 2010.

Discrete Vector Fields



- Given a chain complex C_* and a *dvf*, V over C_*
  - $C_* \Longrightarrow C_*^c$
  - generators of C_*^c are *critical cells* of C_*

$$\begin{array}{c} 0 \leftarrow \mathbb{Z}^{16} \stackrel{d_1}{\longleftarrow} \mathbb{Z}^{32} \stackrel{d_2}{\longleftarrow} \mathbb{Z}^{16} \leftarrow 0 \\ \downarrow \\ 0 \leftarrow \mathbb{Z} \stackrel{\widehat{d_1}}{\longleftarrow} \mathbb{Z} \stackrel{\widehat{d_2}}{\longleftarrow} 0 \leftarrow 0 \end{array}$$

DVF Reduction Theorem

- Let $C_* = (C_p, d_p)_{p\in\mathbb{Z}}$ a free chain complex with distinguished $\mathbb{Z}$-basis $\beta_p \subset C_p$. A *discrete vector field* V on C_* is a collection of pairs $V = \{(\sigma_i; \tau_i)\}_{i\in I}$ satisfying the conditions:
  - Every σ_i is some element of β_p, in which case $\tau_i \in \beta_{p+1}$.
  - Every component σ_i is a *regular face* of the corresponding τ_i.
  - Each generator (*cell*) of C_* appears at most once in V.
- *DVF Reduction Theorem*: Let $C_* = (C_p, d_p)_{p\in\mathbb{Z}}$ be a free chain complex and $V = \{(\sigma_i; \tau_i)\}_{i\in I}$ be an admissible discrete vector field on C_*. Then the vector field V defines a canonical reduction $(f, g, h) : (C_p, d_p) \Longrightarrow (C^c_p, d'_p)$ where $C^c_p = \mathbb{Z}[\beta^c_p]$ is the free $\mathbb{Z}$-module generated by the critical p-cells.
- One proof by Romero and Sergeraert uses the BPL.
- Formalised in: J. Heras, M. Poza, J. R. "Verifying an Algorithm Computing Discrete Vector Fields for Digital Imaging". Calculemus 2012, LNCS 7362 (2012) 216-230.

Biomedical image processing

- Constraints in the previous formalisation:
  - Computing over $\mathbb{Z}_2$.
  - Only finitely generated groups (finite dimensional vector spaces, matrices, SSReflect).
- Application: counting synapses.
  - *Synapses* are the points of connection between neurons.
  - *Relevance*: Computational capabilities of the brain.
  - Procedures to modify the synaptic density may be an important asset in the treatment of neurological diseases.
  - An automated and reliable method is necessary.

Counting Synapses

Computing Homology Groups

- Counting synapses:
  - Counting connected components.
  - Computing a homology group: H_0.
- It is a matter of *matrix diagonalisation*.
- Formalisation of Smith Normal Form:
  C. Cohen, M. Dénès, A. Mörtberg, V. Siles.
  "Smith Normal Form and executable rank for matrices".
  `http://wiki.portal.chalmers.se/cse/pmwiki.php/ForMath/`
- Formalisation of homological computing:
  J. Heras, M. Dénès, G. Mata, A. Mörtberg, M. Poza, V. Siles.
  "Towards a certified computation of homology groups for digital images". CTIC 2012, LNCS 7309 (2012) 49-57.
- Results with biomedical images:
  - Without DVF reduction procedure:
    - ⋆ Coq is not able to compute homology of this kind of images.
  - After reduction procedure:
    - ⋆ Coq computes in just 25 seconds.

Interoperability

- Could different proof assistants cooperate in a same proof?
- Matrix computing: essentially a first-order problem.
- Formalisation in Isabelle/HOL: Hermite form (J. Aransay, J. Divasón).
- Could the specification be translated automatically to ACL2?
- Interlingua: OCL, the constraint language for UML.
- Largely based in XML manipulation and already-made tools (Eclipse tools, as Ecore).
- Joint work: J. Aransay, J. Divasón, J. Heras, AL Rubio, J. R.

Persistent Homology

- Another biological problem: neuron recognition (where counting synapses).
- Topological tool: persistent homology.
- Formalisation in SSReflect:
  J. Heras, T. Coquand, A. Mörtberg, V. Siles.
  "Computing Persistent Homology within Coq/SSReflect".
- To define persistent homology a *filtration* of a simplicial complex is required.
- From the same data, a spectral sequence can be defined.
- Ana Romero made Kenzo compute spectral sequences...
- ...and then persistent homology.

Another published "theorem"

Spectral Sequence Theorem:

$$\sum_{p=1}^{n} \text{rank} E^r_{p,q} = \text{card}\{a \in Dgm_{p+q}(f) | pers(a) \geq r\}$$

"Computational Topology".
Americal Mathematical Society, 2010.

- Ana Romero (Kenzo) found a discrepancy.
- The formula was corrected.
- Another more accurate formula was given.
- Computer Algebra is going beyond...
- ...more formal verification is needed.

Conclusions and further work

- Conclusion. . . of the ForMath european project.
  - Infrastructure to formalize mathematics in constructive type theory.
  - Linear Algebra library.
  - Real numbers and differential equations.
  - Algebraic topology.
    - Representation of simplicial complexes.
    - Certified computation of homology groups.
    - Representation of the Basic Perturbation Lemma.
    - Integration with other proofs systems.
    - Applications to medical imagery.
- Future:
  - From certified computing to *efficient* certified computing.
  - More applications.
    - More Topology in biomedical applications.
    - More verification in Topology.