Verifying an algorithm computing Discrete Vector Fields for digital imaging*

J. Heras, M. Poza, and J. Rubio

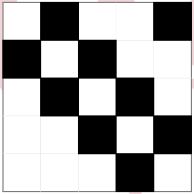
Department of Mathematics and Computer Science, University of La Rioja

Calculemus 2012

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

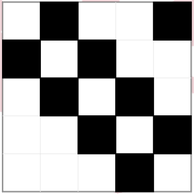
Algebraic Topology and Digital Images

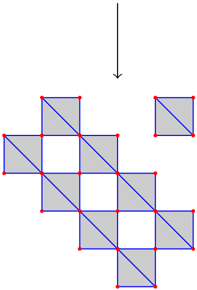
Digital Image



Algebraic Topology and Digital Images

Digital Image

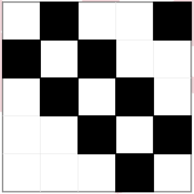


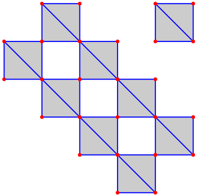


Simplicial complex

Algebraic Topology and Digital Images

Digital Image





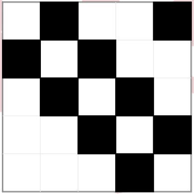
Simplicial complex

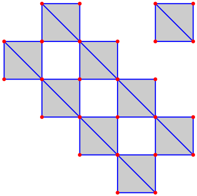
$$C_0 = \mathbb{Z}_2[vertices]$$
$$C_1 = \mathbb{Z}_2[edges]$$
$$C_2 = \mathbb{Z}_2[triangles]$$

Chain complex

Algebraic Topology and Digital Images

Digital Image



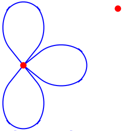


Simplicial complex

$$C_0 = \mathbb{Z}_2[vertices]$$
$$C_1 = \mathbb{Z}_2[edges]$$
$$C_2 = \mathbb{Z}_2[triangles]$$

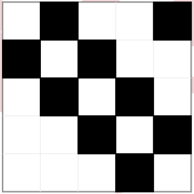
Chain complex

Reduced
Chain Complex

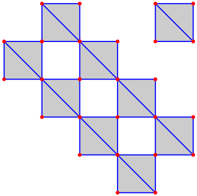


Algebraic Topology and Digital Images

Digital Image

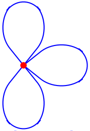


Simplicial complex



Homology groups

$$H_0 = \mathbb{Z}_2 \oplus \mathbb{Z}_2$$
$$H_1 = \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$$



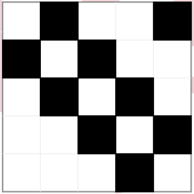
Reduced Chain Complex

$$C_0 = \mathbb{Z}_2[vertices]$$
$$C_1 = \mathbb{Z}_2[edges]$$
$$C_2 = \mathbb{Z}_2[triangles]$$

Chain complex

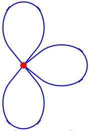
Algebraic Topology and Digital Images

Digital Image

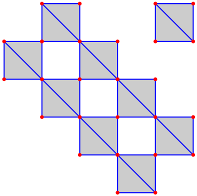


Homology groups

$$H_0 = \mathbb{Z}_2 \oplus \mathbb{Z}_2$$
$$H_1 = \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$$



Reduced Chain Complex



Simplicial complex

$$C_0 = \mathbb{Z}_2[vertices]$$
$$C_1 = \mathbb{Z}_2[edges]$$
$$C_2 = \mathbb{Z}_2[triangles]$$

Chain complex

Goal

- Application:
  - Analysis of biomedical images

Goal

- Application:
  - Analysis of biomedical images
- Requirements:
  - Efficiency
  - Reliability

Goal

- Application:
  - Analysis of biomedical images
- Requirements:
  - Efficiency
  - Reliability

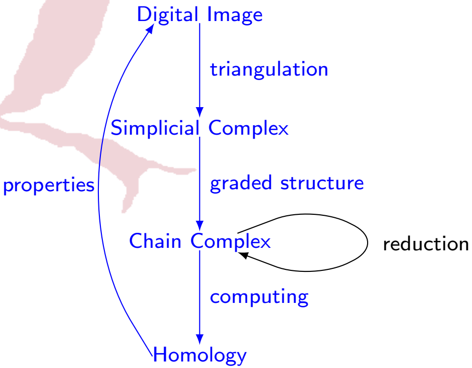
Goal

A formally verified efficient method to compute homology from digital images

Goal

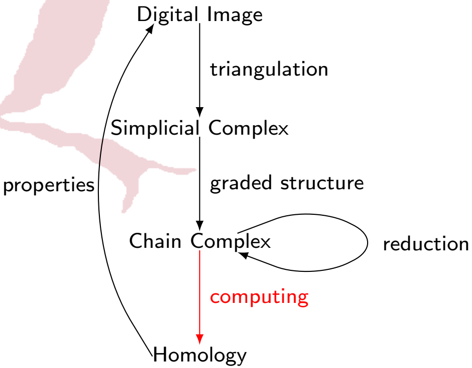
$$
\begin{tikzcd}
\text{Digital Image} \arrow[d, "\text{triangulation}"] \\
\text{Simplicial Complex} \arrow[d, "\text{graded structure}"] \\
\text{Chain Complex} \arrow[d, "\text{computing}"] \arrow[loop right, "\text{reduction}"] \\
\text{Homology} \arrow[uuu, bend left, "\text{properties}"]
\end{tikzcd}
$$

Goal



J. Heras, M. Dénès, G. Mata, A. Mörtberg, M. Poza, and V. Siles. Towards a certified computation of homology groups. In proceedings 4th International Workshop on Computational Topology in Image Context. Lecture Notes in Computer Science, 7309, pages 49–57, 2012.

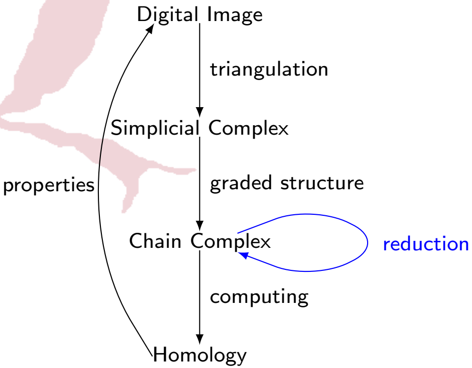
Goal



Bottleneck

Compute Homology from Chain Complexes

Goal



Goal of this work

Formalization in Coq/SSReflect of a procedure to reduce the size of Chain Complexes but preserving homology

Table of Contents

Table of Contents

Chain Complexes

Definition

A chain complex C_* *is a pair of sequences* $C_* = (C_q, d_q)_{q \in \mathbb{Z}}$ *where:*

- *For every* $q \in \mathbb{Z}$, *the component* C_q *is a* $\mathbb{Z}_2$*-module, the chain group of degree* q
- *For every* $q \in \mathbb{Z}$, *the component* d_q *is a module morphism* $d_q : C_q \to C_{q-1}$, *the differential map*
- *For every* $q \in \mathbb{Z}$, *the composition* $d_q d_{q+1}$ *is null:* $d_q d_{q+1} = 0$

Chain Complexes

Definition

A chain complex C_* *is a pair of sequences* $C_* = (C_q, d_q)_{q\in\mathbb{Z}}$ *where:*

- *For every* $q \in \mathbb{Z}$*, the component* C_q *is a* $\mathbb{Z}_2$*-module, the chain group of degree* q
- *For every* $q \in \mathbb{Z}$*, the component* d_q *is a module morphism* $d_q : C_q \to C_{q-1}$*, the differential map*
- *For every* $q \in \mathbb{Z}$*, the composition* $d_q d_{q+1}$ *is null:* $d_q d_{q+1} = 0$

Definition

If $C_* = (C_q, d_q)_{q\in\mathbb{Z}}$ *is a chain complex:*

- *The image* $B_q = im\ d_{q+1} \subseteq C_q$ *is the (sub)module of q-boundaries*
- *The kernel* $Z_q = ker\ d_q \subseteq C_q$ *is the (sub)module of q-cycles*

Chain Complexes

Definition

A chain complex C_* *is a pair of sequences* $C_* = (C_q, d_q)_{q\in\mathbb{Z}}$ *where:*

- *For every* $q \in \mathbb{Z}$*, the component* C_q *is a* $\mathbb{Z}_2$*-module, the chain group of degree* q
- *For every* $q \in \mathbb{Z}$*, the component* d_q *is a module morphism* $d_q : C_q \to C_{q-1}$*, the differential map*
- *For every* $q \in \mathbb{Z}$*, the composition* $d_q d_{q+1}$ *is null:* $d_q d_{q+1} = 0$

Definition

If $C_* = (C_q, d_q)_{q\in\mathbb{Z}}$ *is a chain complex:*

- *The image* $B_q = im\ d_{q+1} \subseteq C_q$ *is the (sub)module of q-boundaries*
- *The kernel* $Z_q = ker\ d_q \subseteq C_q$ *is the (sub)module of q-cycles*

Definition

Let $C_* = (C_q, d_q)_{q\in\mathbb{Z}}$ *be a chain complex. For each degree* $n \in \mathbb{Z}$*, the n-homology module of* C_* *is defined as the quotient module*

$$H_n(C_*) = \frac{Z_n}{B_n}$$

Reduction

Definition

A reduction ρ between two chain complexes C_* y D_* (denoted by $\rho : C_* \Rightarrow D_*$) is a tern $\rho = (f, g, h)$

$$\begin{tikzcd} C_* \arrow[loop left, "h"] \arrow[r, bend left, "f"] & D_* \arrow[l, bend left, "g"] \end{tikzcd}$$

satisfying the following relations:

1) $fg = id_{D_*}$;
2) $d_C h + hd_C = id_{C_*} - gf$;
3) $fh = 0; \quad hg = 0; \quad hh = 0.$

Theorem

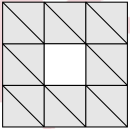
If $C_* \Rightarrow D_*$, then $C_* \cong D_* \oplus A_*$, with A_* acyclic, what implies that $H_n(C_*) \cong H_n(D_*)$ for all n.

Discrete Morse Theory

A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.

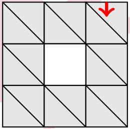
Discrete Morse Theory

A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.



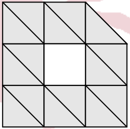
Discrete Morse Theory

A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.



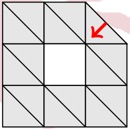
Discrete Morse Theory

A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.



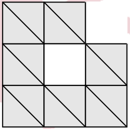
Discrete Morse Theory

A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.



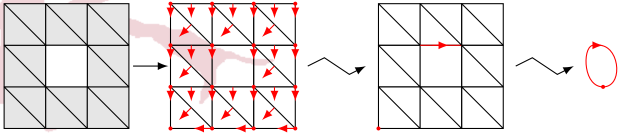
Discrete Morse Theory

A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.



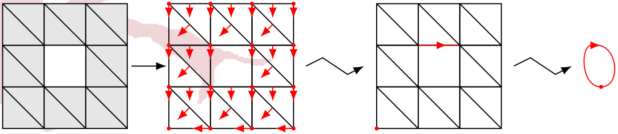
Discrete Morse Theory

A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.



Discrete Morse Theory

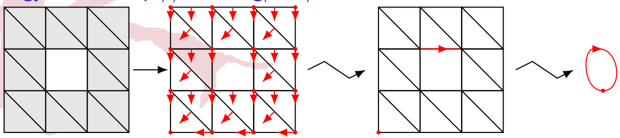
A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.



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

Discrete Morse Theory

A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.



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

$$\begin{array}{ccccccccc} 0 & \leftarrow & \mathbb{Z}_2^{16} & \xleftarrow{d_1} & \mathbb{Z}_2^{32} & \xleftarrow{d_2} & \mathbb{Z}_2^{16} & \leftarrow & 0 \\ & & & & \downarrow & & & & \\ 0 & \leftarrow & \mathbb{Z}_2 & \xleftarrow{\widehat{d_1}} & \mathbb{Z}_2 & \xleftarrow{\widehat{d_2}} & 0 & \leftarrow & 0 \end{array}$$

Discrete Vector Fields

Definition

Let $C_* = (C_p, d_p)_{p\in\mathbb{Z}}$ a free chain complex with distinguished $\mathbb{Z}_2$-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 $\tau_i \in \beta_{p+1}$. The degree p depends on i and in general is not constant.*
- *Every component σ_i is a regular face of the corresponding τ_i.*
- *Each generator (cell) of C_* appears at most once in V.*

Discrete Vector Fields

Definition

Let $C_* = (C_p, d_p)_{p\in\mathbb{Z}}$ a free chain complex with distinguished $\mathbb{Z}_2$-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 $\tau_i \in \beta_{p+1}$. The degree p depends on i and in general is not constant.*
- *Every component σ_i is a regular face of the corresponding τ_i.*
- *Each generator (cell) of C_* appears at most once in V.*

Definition

A V-path of degree p and length m is a sequence $\pi = ((\sigma_{i_k}, \tau_{i_k}))_{0\leq k<m}$ satisfying:

- *Every pair $(\sigma_{i_k}, \tau_{i_k})$ is a component of V and τ_{i_k} is a p-cell.*
- *For every $0 < k < m$, the component σ_{i_k} is a face of $\tau_{i_{k-1}}$, non necessarily regular, but different from $\sigma_{i_{k-1}}$.*

Discrete Vector Fields

Definition

A discrete vector field V is admissible if for every $p \in \mathbb{Z}$, a function $\lambda_p : \beta_p \to \mathbb{N}$ is provided satisfying the following property: every V-path starting from $\sigma \in \beta_p$ has a length bounded by $\lambda_p(\sigma)$.

Discrete Vector Fields

Definition

A discrete vector field V is admissible if for every $p \in \mathbb{Z}$, a function $\lambda_p : \beta_p \to \mathbb{N}$ is provided satisfying the following property: every V-path starting from $\sigma \in \beta_p$ has a length bounded by $\lambda_p(\sigma)$.

Definition

A cell σ which does not appear in a discrete vector field V is called a critical cell.

Discrete Vector Fields

Definition

A discrete vector field V is admissible if for every $p \in \mathbb{Z}$, a function $\lambda_p : \beta_p \to \mathbb{N}$ is provided satisfying the following property: every V-path starting from $\sigma \in \beta_p$ has a length bounded by $\lambda_p(\sigma)$.

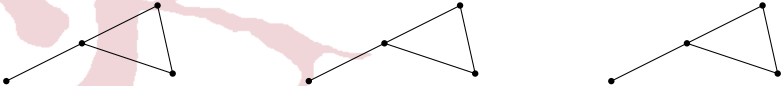
Definition

A cell σ which does not appear in a discrete vector field V is called a critical cell.

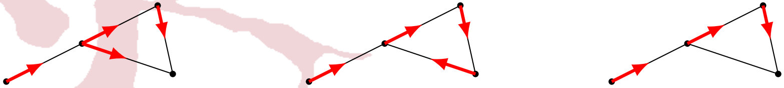
Theorem

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 $\rho = (f, g, h) : (C_p, d_p) \Rightarrow (C_p^c, d_p')$ where $C_p^c = \mathbb{Z}_2[\beta_p^c]$ is the free $\mathbb{Z}_2$-module generated by the critical p-cells.

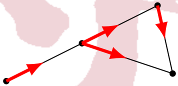
Example: an admissible discrete vector field



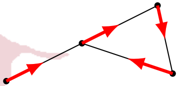
Example: an admissible discrete vector field

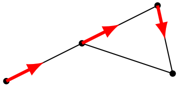


Example: an admissible discrete vector field

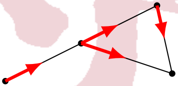


Dvf x

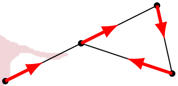




Example: an admissible discrete vector field

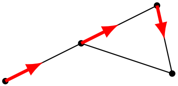


Dvf x

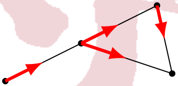


Dvf ✓

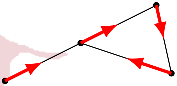
Admissible x



Example: an admissible discrete vector field

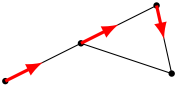


Dvf x



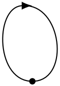
Dvf ✓

Admissible x



Dvf ✓

Admissible ✓



Vector fields and integer matrices

Differential maps of a Chain Complex can be represented as matrices

$$\ldots \leftarrow \mathbb{Z}_2^m \xleftarrow{M} \mathbb{Z}_2^n \leftarrow \ldots$$

Vector fields and integer matrices

Differential maps of a Chain Complex can be represented as matrices

$$\ldots \leftarrow \mathbb{Z}_2^m \xleftarrow{M} \mathbb{Z}_2^n \leftarrow \ldots$$

Definition

An admissible vector field V for M is nothing but a set of integer pairs $\{(a_i, b_i)\}$ satisfying these conditions:

1. $1 \leq a_i \leq m$ *and* $1 \leq b_i \leq n$
2. *The entry* $M[a_i, b_i]$ *of the matrix is* 1
3. *The indices* a_i *(resp.* b_i*) are pairwise different*
4. *Non existence of loops*

Vector fields and integer matrices

Differential maps of a Chain Complex can be represented as matrices

$$\ldots \leftarrow \mathbb{Z}_2^m \xleftarrow{M} \mathbb{Z}_2^n \leftarrow \ldots$$

Definition

An admissible vector field V for M is nothing but a set of integer pairs $\{(a_i, b_i)\}$ satisfying these conditions:

1. $1 \leq a_i \leq m$ *and* $1 \leq b_i \leq n$
2. *The entry* $M[a_i, b_i]$ *of the matrix is* 1
3. *The indices* a_i *(resp.* b_i*) are pairwise different*
4. *Non existence of loops*

Vector fields and integer matrices

Differential maps of a Chain Complex can be represented as matrices

$$\ldots \leftarrow \mathbb{Z}_2^m \xleftarrow{M} \mathbb{Z}_2^n \leftarrow \ldots$$

Definition

An admissible vector field V for M is nothing but a set of integer pairs $\{(a_i, b_i)\}$ satisfying these conditions:

1. $1 \leq a_i \leq m$ *and* $1 \leq b_i \leq n$
2. *The entry $M[a_i, b_i]$ of the matrix is* 1
3. *The indices a_i (resp. b_i) are pairwise different*
4. *Non existence of loops*

Table of Contents

Coq/SSReflect

- Coq:
  - An Interactive Proof Assistant
  - Based on Calculus of Inductive Constructions
  - Interesting feature: program extraction from a constructive proof

Coq/SSReflect

- Coq:
  - An Interactive Proof Assistant
  - Based on Calculus of Inductive Constructions
  - Interesting feature: program extraction from a constructive proof
- SSReflect:
  - Extension of Coq
  - Developed while formalizing the Four Color Theorem by G. Gonthier
  - Currently, it is used in the formalization of Feit-Thompson Theorem

Admissible discrete vector fields in SSReflect

Definition

An admissible discrete vector field V for M is nothing but a set of integer pairs $\{(a_i, b_i)\}$ satisfying these conditions:

1. $1 \leq a_i \leq m$ *and* $1 \leq b_i \leq n$
2. *The entry* $M[a_i, b_i]$ *of the matrix is* 1
3. *The indices* a_i *(resp.* b_i*) are pairwise different*
4. *Non existence of loops*

```
Definition admissible_dvf (M: 'M[Z2]_(m,n))
                  (V: seq ('I_m * 'I_n)) (ords : simpl_rel 'I_m) :=
  all [pred p | M p.1 p.2 == 1] V &&
  uniq (map (@fst _ _) V) && uniq (map (@snd _ _) V) &&
  all [pred i | ~~ (connect ords i i)] (map (@fst _ _) V).
```

The abstract algorithm

```
Fixpoint genDvfOrders M V (ords : simpl_rel _) k :=
  if k is l.+1 then
    let P := [pred ij | admissible (ij::V) M
                         (relU ords (gen_orders M ij.1 ij.2))] in
    if pick P is Some (i,j)
       then genDvfOrders M ((i,j)::V)
                          (relU ords (gen_orders M i j)) l
    else (V, ords)
  else (V, ords).

Definition gen_adm_dvf M :=
  genDvfOrders M [::] [rel x y | false] (minn m n).
```

The abstract algorithm

```
Fixpoint genDvfOrders M V (ords : simpl_rel _) k :=
  if k is l.+1 then
    let P := [pred ij | admissible (ij::V) M
                        (relU ords (gen_orders M ij.1 ij.2))] in
    if pick P is Some (i,j)
       then genDvfOrders M ((i,j)::V)
                         (relU ords (gen_orders M i j)) l
    else (V, ords)
  else (V, ords).

Definition gen_adm_dvf M :=
  genDvfOrders M [::] [rel x y | false] (minn m n).

Lemma admissible_gen_adm_dvf m n (M : 'M[Z2]_(m,n)) :
 let (vf,ords) := gen_adm_dvf M in admissible vf M ords.
```

Problem

It is not an executable algorithm

Table of Contents

Romero-Sergeraert's Algorithm

A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.

Algorithm

Input: A matrix M
Output: An admissible discrete vector field for M

Romero-Sergeraert's Algorithm

A. Romero and F. Sergeraert. Discrete Vector Fields and Fundamental Algebraic Topology, 2010. http://arxiv.org/abs/1005.5685v1.

Algorithm

Input: A matrix M
Output: An admissible discrete vector field for M

Algorithm

Input: A chain complex C_* *and an admissible discrete vector field of* C_*
Output: A reduced chain complex $\hat{C}_*$

From computation to verification through testing

- *Haskell* as programming language
- *QuickCheck* to test the programs
- *Coq/SSReflect* to verify the correctness of the programs

Haskell

Algorithm (*gen_adm_dvf*)

Input: A matrix M
Output: An admissible discrete vector field for M

Algorithm (*reduced_cc*)

Input: A chain complex C_*
Output: A reduced chain complex $\hat{C}_*$

```
> gen_adm_dvf [[1,0,1,1],[0,0,1,0],[1,1,0,1]]
[(0,0),(1,2),(2,1)]
```

QuickCheck

- A specification of the properties which our program must verify
- Testing them
  - Towards verification
  - Detect bugs

```
> quickCheck M -> admissible (gen_adm_dvf M)
+ + + OK, passed 100 tests
```

Coq/SSReflect

SSReflect Theorem:

```
Theorem gen_adm_dvf_is_admissible (M : seq (seq Z2)) :
 admissible (gen_adm_dvf M).
```

SSReflect Theorem:

```
Theorem is_reduction (C : chaincomplex) : reduction C (reduced_cc C).
```

SSReflect Theorem:

```
Theorem reduction_preserves_betti (C D : chaincomplex)
  (rho : reduction C D) : Betti C = Betti D.
```

Experimental results

500 randomly generated matrices

- Initial size of the matrices: 100×300
- Time: 12 seconds

Experimental results

500 randomly generated matrices

- Initial size of the matrices: 100×300
- Time: 12 seconds

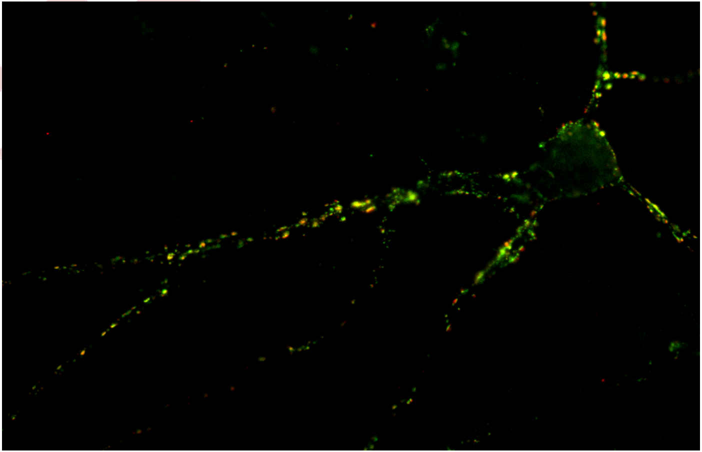
- After reduction: 5×50
- Time: milliseconds

Table of Contents

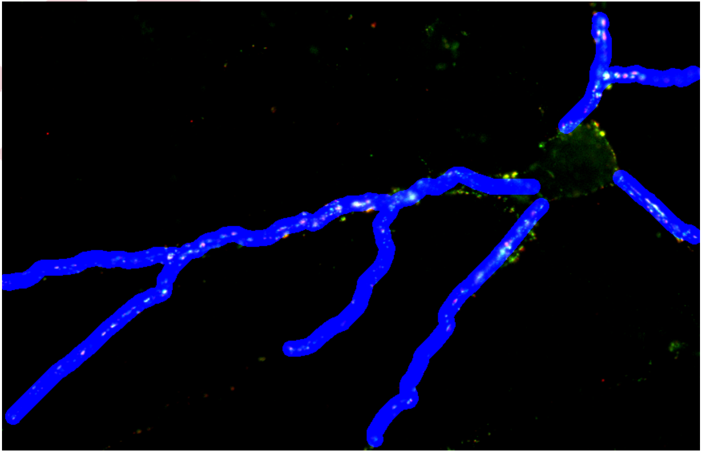
Counting Synapses

- *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



Counting Synapses



Counting Synapses

Results with biomedical images

- Without reduction procedure:
  - Coq is not able to compute homology of this kind of images

Results with biomedical images

- Without reduction procedure:
  - Coq is not able to compute homology of this kind of images
- After reduction procedure:
  - Coq computes in just 25 seconds

Table of Contents

Conclusions and Further work

- Conclusions:
  - Method to reduce big images preserving homology
  - Formalization of admissible discrete vector fields on Coq
  - Remarkable reductions in different benchmarks

Conclusions and Further work

- Conclusions:
  - Method to reduce big images preserving homology
  - Formalization of admissible discrete vector fields on Coq
  - Remarkable reductions in different benchmarks
- Further work:
  - Matrices with coefficients over $\mathbb{Z}$
  - Integration between Coq and ACL2
  - Application of homological methods to biomedical problems

Verifying an algorithm computing Discrete Vector Fields for digital imaging

J. Heras, M. Poza, and J. Rubio

Department of Mathematics and Computer Science, University of La Rioja

Calculemus 2012