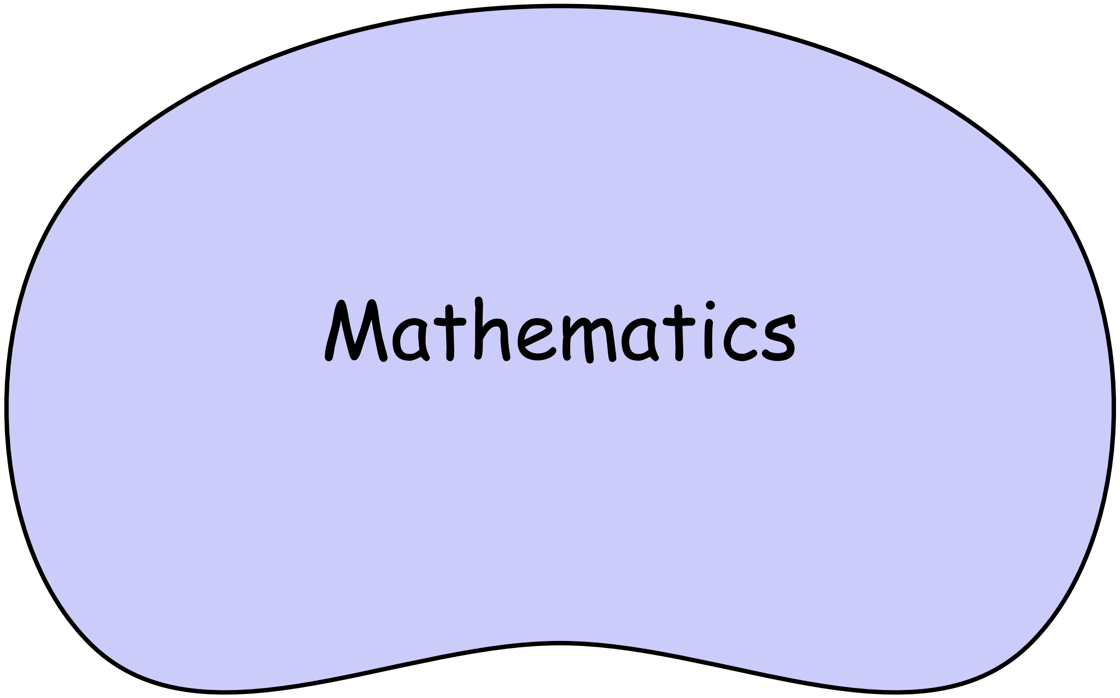
Can we mechanically check any mathematical proof?

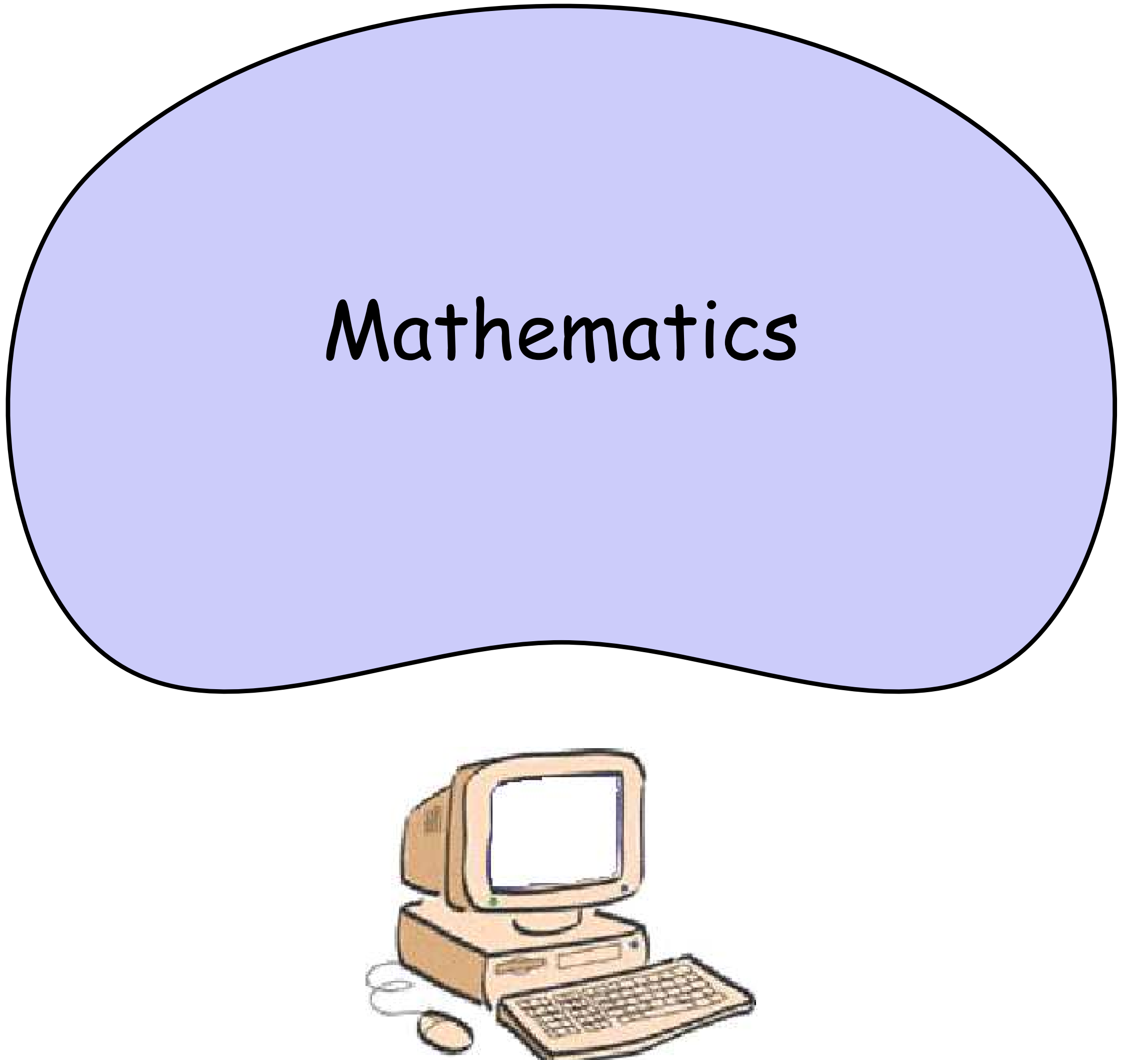
Laurent Théry
Marelle INRIA Sophia-Antipolis France

Mathematics and Computer

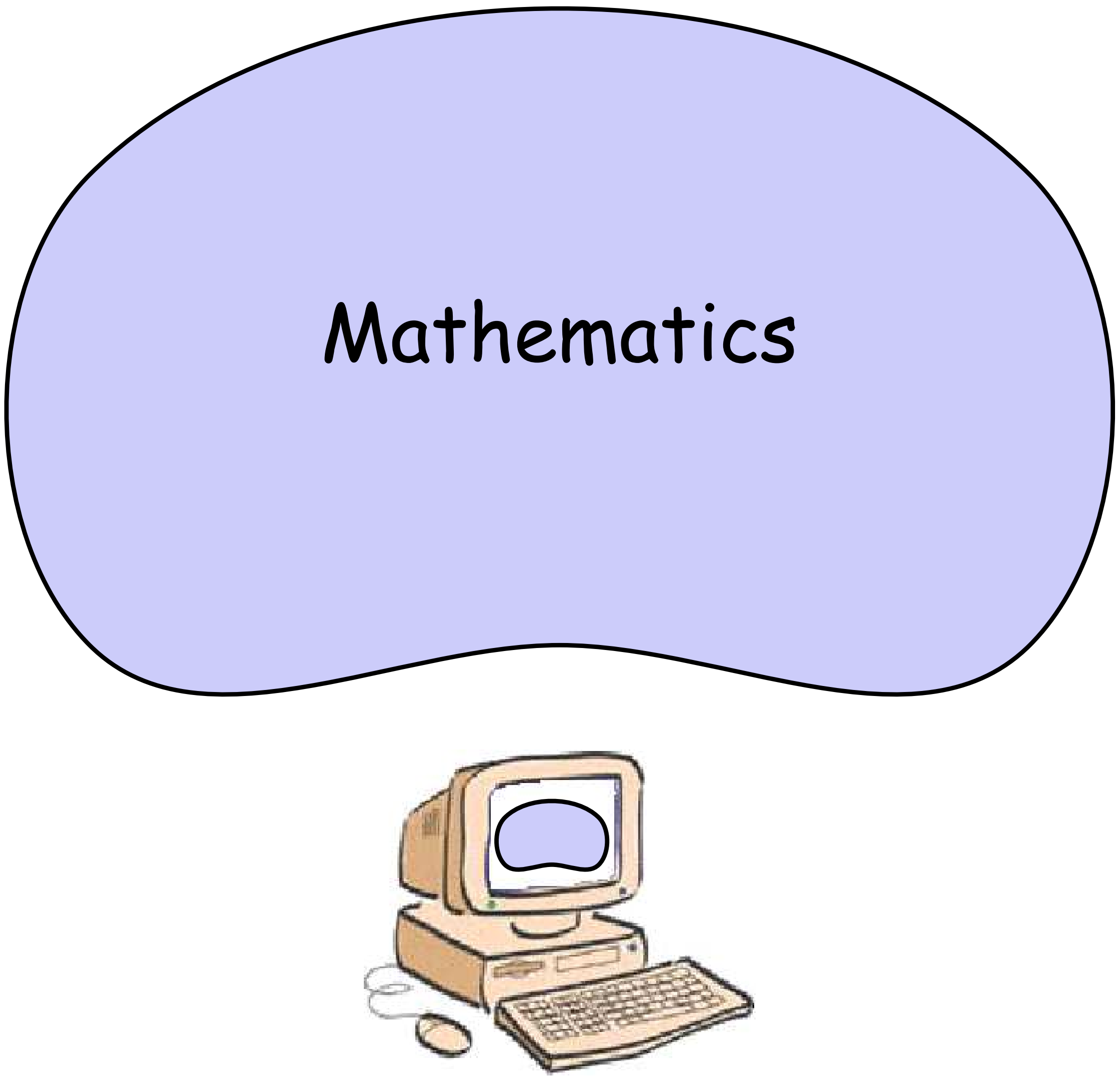
Mathematics and Computer



Mathematics and Computer



Mathematics and Computer



Formal Mathematics

Formal Mathematics

Formal Language

Formal Mathematics

Formal Language

function: $x + y$

predicate: $prime(x)$

relation: $x = y$

logical connective: $\wedge, \vee, \Rightarrow, \leftrightarrow, \neg$

quantifier: $\forall, \exists$

Formal Mathematics

Formal Language

function: $x + y$

predicate: $prime(x)$

relation: $x = y$

logical connective: $\wedge, \vee, \Rightarrow, \leftrightarrow, \neg$

quantifier: $\forall, \exists$

Example

$$\forall nxyz, x^n + y^n = z^n \Rightarrow xyz = 0 \vee n \leq 2$$

Formal Proof

Formal Proof

Proof:
sequence of elementary steps

$$\frac{A \qquad B}{A \wedge B}$$

Formal Proof

Proof:
sequence of elementary steps

$$\frac{A \qquad B}{A \wedge B}$$

$\longrightarrow$ Too tedious for a human-being

Formal Proof

Proof:
sequence of elementary steps

$$\frac{A \qquad B}{A \wedge B}$$

$\longrightarrow$ Too tedious for a human-being

$\longrightarrow$ Ok for a computer

Formal Proof

Proof:
sequence of elementary steps

$$\frac{A \qquad B}{A \wedge B}$$

$\longrightarrow$ Too tedious for a human-being

$\longrightarrow$ Ok for a computer

Origin:
B. Russell, *Principia Mathematica*

Pros and Cons

Pros and Cons

Pros:

- Simple datastructure

Pros and Cons

Pros:

Simple datastructure

Easy to check

Pros and Cons

Pros:

- Simple datastructure
- Easy to check
- Complete check

Pros and Cons

Pros:

- Simple datastructure
- Easy to check
- Complete check

Cons:

- Catch 22

Pros and Cons

Pros:

- Simple datastructure
- Easy to check
- Complete check

Cons:

- Catch 22
- De Millo and al.:
  Social processes and proofs

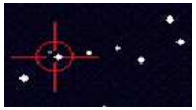
Proof Systems

Proof Systems

Mizar

Proof Systems

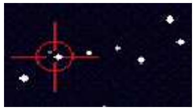
Hol

Mizar

Proof Systems

Coq

Hol



Mizar

Proof Systems

Coq

Hol

Isabelle

Mizar

Proof Systems

Coq

Hol

Isabelle

Mizar

Pvs

Proof Systems

Coq

Hol

Isabelle

Mizar

Pvs

Acl2

Some formalisations

Some formalisations

Prime number theorem

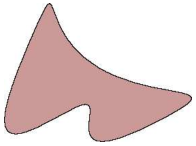
$\pi(x) \sim \frac{x}{\ln x}$ (Isabelle)

Some formalisations

Prime number theorem

$$\pi(x) \sim \frac{x}{\ln x}$$ (Isabelle)

Jordan curve theorem



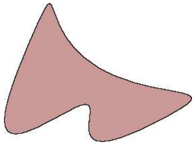
(Hol, Mizar)

Some formalisations

Prime number theorem

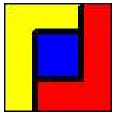
$\pi(x) \sim \frac{x}{\ln x}$ (Isabelle)

Jordan curve theorem



(Hol, Mizar)

Four colour theorem



(Coq)

What's next?

What's next?

Fermat theorem?

What's next?

Fermat theorem?

Poincaré theorem?

What's next?

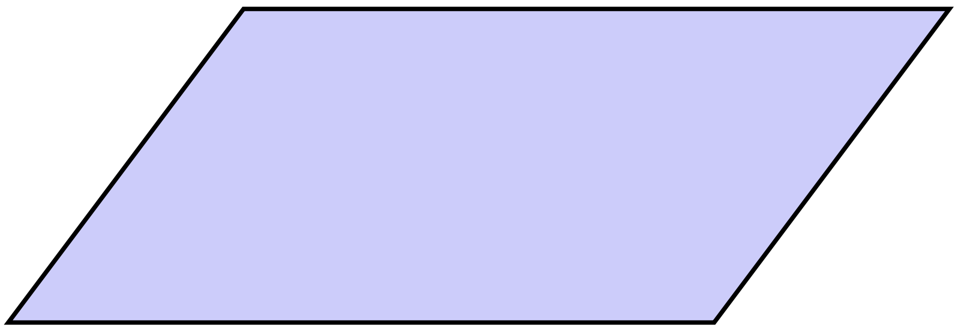
Fermat theorem?

Poincaré theorem?

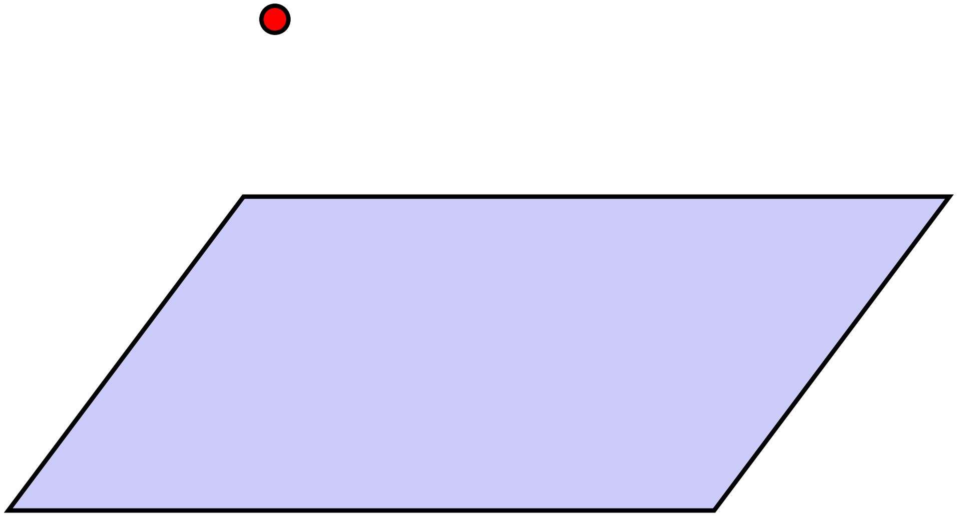
...

Mexican hats

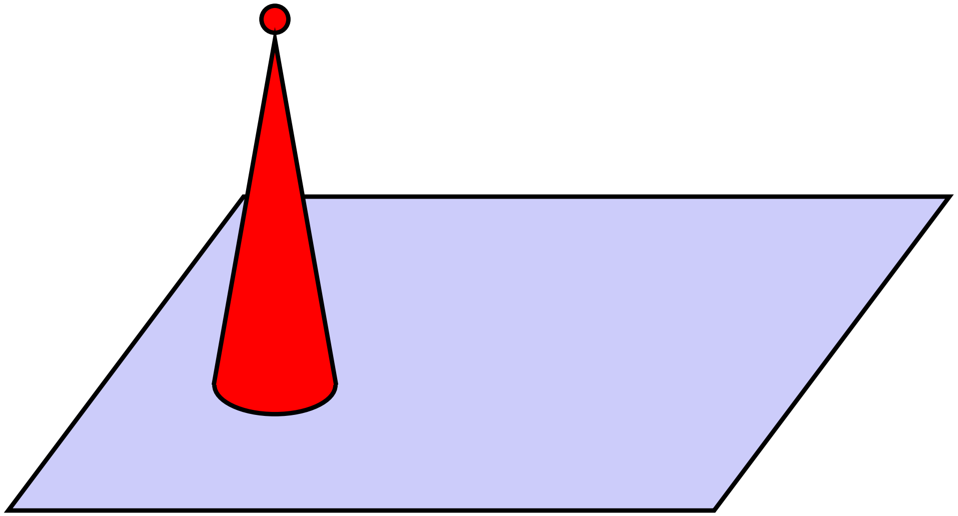
Mexican hats



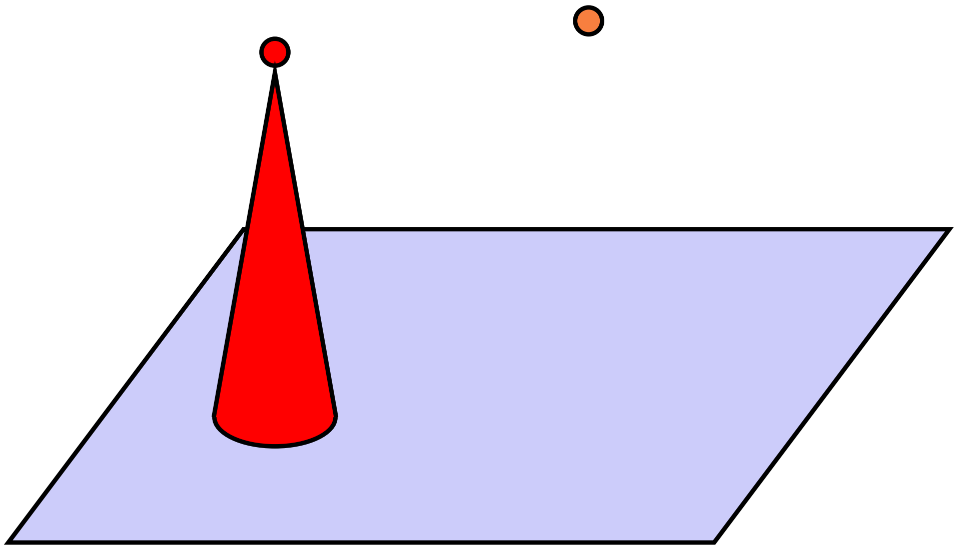
Mexican hats



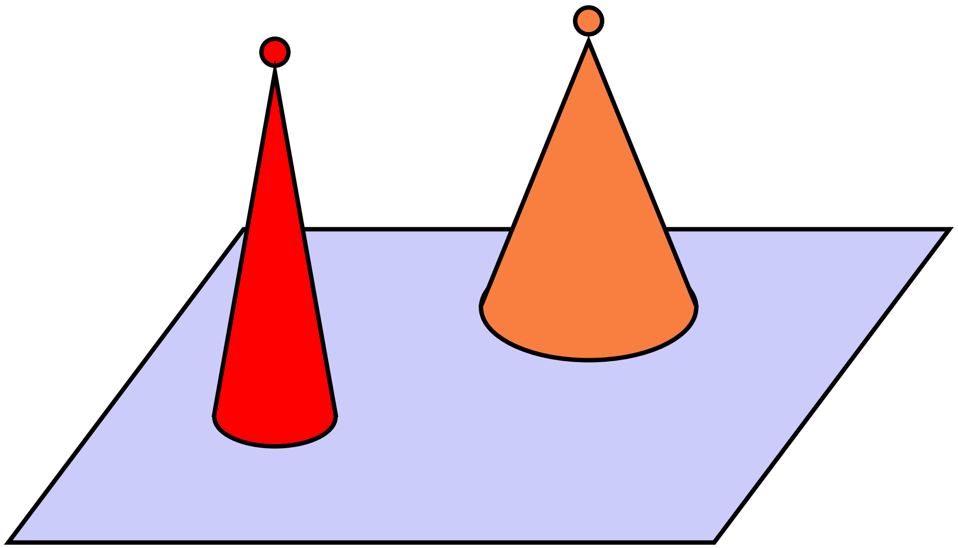
Mexican hats



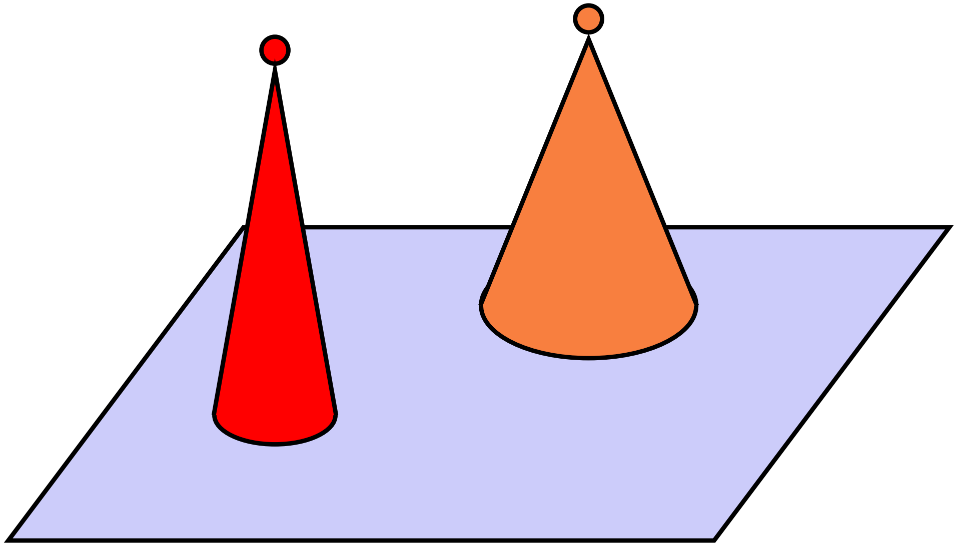
Mexican hats



Mexican hats

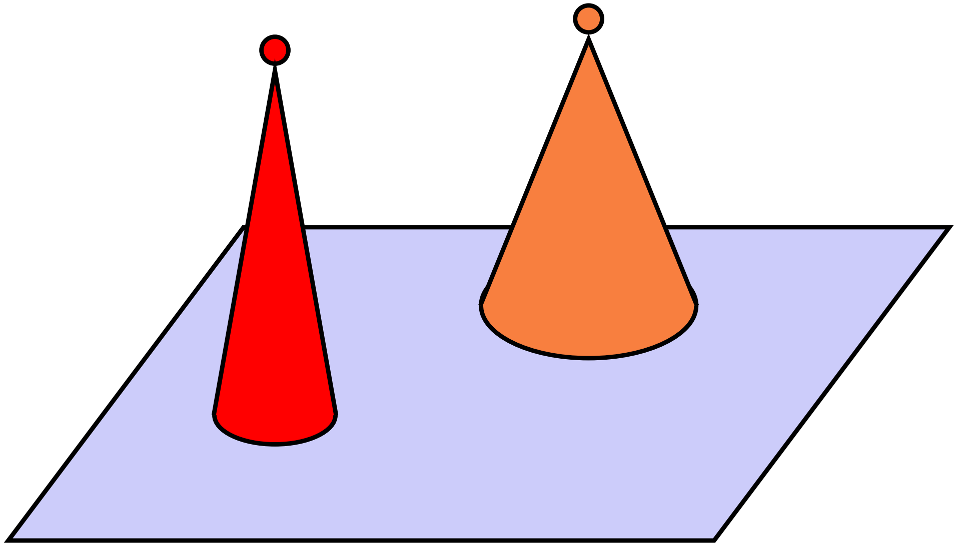


Mexican hats



Cooperative work

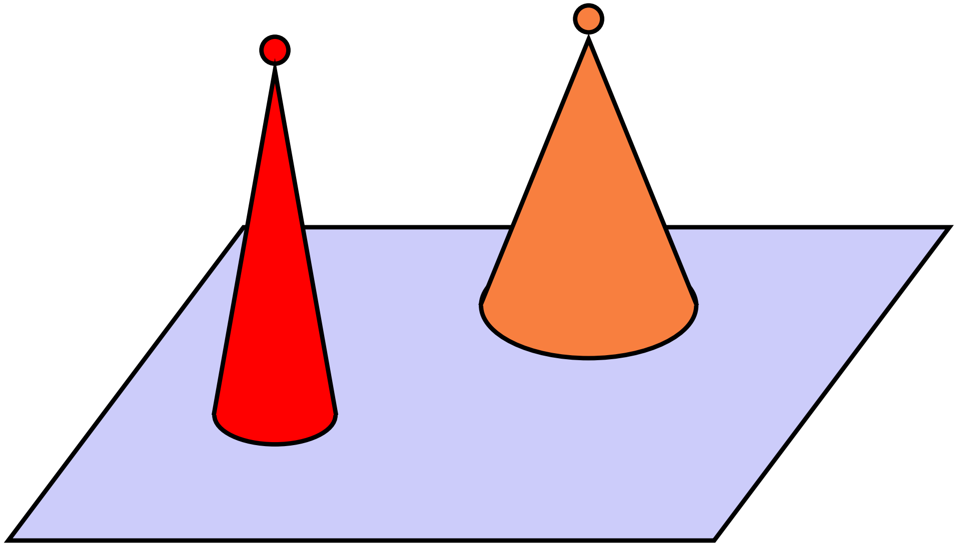
Mexican hats



Cooperative work

Wikipedia effect

Mexican hats



Cooperative work

Wikipedia effect

Education

Technical Proofs

Not every proof is nice and tiny

Technical Proofs

Not every proof is nice and tiny

Example

J. Demmel and Y. Hida,

Accurate floating point summation

19 page proof

Technical Proofs

Property B: The leftmost leading bit of $\widehat{SUM}_{I+1}$ through $\widehat{SUM}_n$ is to the left of the leading bit of SUM_I: $\max_{k>I} E_k > E_I$.

Now we may consider six cases, labeled 1A, 1B, 2A, 2B, 3A and 3B, according to which pair of properties holds. We may also have subcases of these cases depending on the size of n. There may be further subcases depending on when the exponents e_k and E_j further decrease below their initial levels.

We would like to believe a simpler proof exists, but have not managed to find one.

8.1 Case 1A - $n \le \bar{n} + 1$

Property 1 means $e_{I+1} \le E_I - F + f - 1$, so let K be the smallest integer in the range $I \le K \le n$ such that $e_k \le E_I - F + f - 2$ for all $k > K$. In other words, e_{I+1} through e_K are all $E_I - F + f - 1$, and e_{K+1} through e_n are all at most $E_I - F + f - 2$. Note that either list, but not both, can be vacuous. Thus we have the bounds

$$|s_k| \le \begin{cases} 2^{E_I-F+f}(1-2^{-f}) & \text{for } I+1 \le k \le K \\ 2^{E_I-F+f-1}(1-2^{-f}) & \text{for } K+1 \le k \le n \end{cases} \tag{9}$$

Property A implies $E_k \le E_I$ for all $k \ge I$, so let J be the largest integer in the range $I \le J \le n$ such that $E_J = E_I$ but $E_j < E_I$ for all $j > J$. In other words $\widehat{SUM}_J$ is the last computed partial sum with the exponent E_I. This enables us to bound 1 ulp on the partial sums:

$$\text{ulp}(\widehat{SUM}_j) \le \begin{cases} 2^{E_I-F+1} & \text{for } I \le j \le J \\ 2^{E_I-F} & \text{for } J+1 \le j \le n \end{cases} \tag{10}$$

We consider the cases $J \le K$ and $K < J$ separately.

8.1.1 Case $J \le K$

In this case, we have $1 \le I \le J \le K \le n$. The additions of s_{I+1} through s_J, resulting in $\widehat{SUM}_{I+1}$ through $\widehat{SUM}_J$, can yield a maximum roundoff error of half an ulp in each of $\widehat{SUM}_{I+1}$ through $\widehat{SUM}_J$, which is at most 2^{E_I-F} each. If $K \ge J+1$, then addition of s_{J+1} causes *no roundoff*, since $\widehat{SUM}_{J+1}$ is computed by exact cancellation. Additions of s_{J+2} through s_K to the partial sums $\widehat{SUM}_{J+1}$ through $\widehat{SUM}_{K-1}$, resulting in the partial sums $\widehat{SUM}_{J+2}$ through $\widehat{SUM}_K$, also causes *no roundoff*, since all the numbers involved occupies the same F-bit range. Finally, the additions of s_{K+1} through s_n can cause roundoff errors at most 2^{E_I-F-1} each. Thus we have the roundoff error bounds

$$|\epsilon_i| \le \begin{cases} 2^{E_I-F} & \text{for } I+1 \le i \le J \\ 0 & \text{for } J+1 \le i \le K \\ 2^{E_I-F-1} & \text{for } K+1 \le i \le n \end{cases} \tag{11}$$

Thus we can bound the total roundoff error

$$\begin{aligned} |\widehat{SUM}_n - S| &\le \sum_{i=I+1}^{n} |\epsilon_i| \\ &\le (J-I)2^{E_I-F} + (n-K)2^{E_I-F-1} \\ &= (2J - 2I + n - K)2^{E_I-F-1} \\ &= 2^{E_I} N_{1A,\, J\le K}(I, J, K, n), \end{aligned} \tag{12}$$

Technical Proofs

where

$$N_{1A,\ J\le K}(I,J,K,n) = (2J-2I+n-K)2^{-F-1}.$$

We now bound $|\widehat{SUM}_n|$ from below by noting that $|\widehat{SUM}_J| \ge 2^{E_I}$ and using the triangle inequality:

$$\begin{aligned}
|\widehat{SUM}_n| &= |\widehat{SUM}_J + (s_{J+1}+\cdots+s_n) + (\epsilon_{J+1}+\cdots+\epsilon_n)| \\
&\ge |\widehat{SUM}_J| - \sum_{i=J+1}^{n}|s_i| - \sum_{i=J+1}^{n}|\epsilon_i| \\
&\ge 2^{E_I} - (K-J)2^{E_I+f-F}(1-2^{-f}) - (n-K)2^{E_I+f-F-1}(1-2^{-f}) - (n-K)2^{E_I-F-1} \\
&= 2^{E_I}\left[1-(K-2J+n)2^{f-F-1}(1-2^{-f}) - (n-K)2^{-F-1}\right] \\
&= 2^{E_I} D_{1A,\ J\le K}(J,K,n), \qquad (13)
\end{aligned}$$

where

$$D_{1A,\ J\le K}(J,K,n) = 1-(K-2J+n)2^{f-F-1}(1-2^{-f}) - (n-K)2^{-F-1}.$$

The relative error is then bounded by

$$\frac{|\widehat{SUM}_n - S|}{|\widehat{SUM}_n|} \le \frac{N_{1A,\ J\le K}(I,J,K,n)}{D_{1A,\ J\le K}(J,K,n)} \equiv RE_{1A,\ J\le K}(I,J,K,n). \qquad (14)$$

Note that $I = J < K$ cannot occur since means that $E_{I+1} < E_I - 1$ and $\widehat{SUM}_{I+1}$ is computed without roundoff by exact cancellation, contradicting our choice of I. Hence we must have either $I = J = K$ or $I < J \le K$, and the worst case relative error is bounded by the maximum of $RE_{1A,\ J\le K}(I,J,K,n)$ over the domain $U = \{(I,J,K) \mid 1 \le I = J = K \le n \text{ or } 1 \le I < J \le K \le n\}$:

$$\frac{|\widehat{SUM}_n - S|}{|\widehat{SUM}_n|} \le \max_{(I,J,K)\in U} RE_{1A,\ J\le K}(I,J,K,n).$$

We consider the two cases $I = J = K$ and $I < J \le K$ separately.

8.1.1.1 Case $I = J = K$. We first note that the denominator $D_{1A,\ J\le K}(I,I,n)$ becomes

$$D_{1A,\ J\le K}(I,I,n) = 1-(n-I)2^{f-F-1}.$$

Since $(n-I) \le \bar{n}$, we can use bound (7) to get

$$D_{1A,\ J\le K}(I,I,n) \ge 1-\bar{n}2^{f-F-1} > 1 - \frac{2^{-1}+2^{f-F-1}}{1-2^{-f}} \ge 1 - \frac{2^{-1}+2^{-2}}{1-2^{-2}} = 0.$$

Thus $n \le \bar{n}+1$ implies that the denominator is positive.
If $(n-I) \le \bar{n}-1$ (implied by $n \le \bar{n}$), then

$$\begin{aligned}
RE_{1A,\ J\le K}(I,I,I,n) &\le \frac{(\bar{n}-1)2^{-F-1}}{1-(\bar{n}-1)2^{f-F-1}} \\
&= \frac{2^{-1-f}-2^{-F-1-r}}{(1-2^{-f})-(2^{-1}-2^{f-F-r-1})}
\end{aligned}$$

Technical Proofs

$$= \frac{2^{-f}(1-2^{f-F-r})}{1-2^{1-f}+2^{f-F-r}}$$
$$< \frac{2^{-f}}{1-2^{1-f}}. \qquad (15)$$

If $(n-I)=\bar{n}$ (implying $n=\bar{n}+1$), then

$$\begin{aligned} RE_{1A,\,J\le K}(I,I,I,n) &\le \frac{\bar{n}2^{-F-1}}{1-\bar{n}2^{f-F-1}} \\ &= \frac{2^{-1-f}+(1-2^{-f}-2^{-r})2^{-F-1}}{(1-2^{-f})-2^{-1}-(1-2^{-f}-2^{-r})2^{f-F-1}} \\ &= \frac{2^{-f}\left[2^{-1}+(1-2^{-f}-2^{-r})2^{f-F-1}\right]}{(1-2^{-f})-2^{-1}-(1-2^{-f}-2^{-r})2^{f-F-1}} \\ &= \frac{2^{-f}\left[1+(1-2^{-f}-2^{-r})2^{f-F}\right]}{1-2^{1-f}-(1-2^{-f}-2^{-r})2^{f-F}}. \qquad (16) \end{aligned}$$

To bound the last line in the above inequality, we consider the cases $F-f=1$ and $F-f\ge 2$ separately. If $F-f=1$, then $r=f-1$, and so

$$\begin{aligned} RE_{1A,\,J\le K}(I,J,K,n) &\le 2^{-f}\left[\frac{1+(1-2^{-f}-2^{-r})2^{f-F}}{1-2^{1-f}-(1-2^{-f}-2^{-r})2^{f-F}}\right] \\ &= 2^{-f}\left[\frac{1+(1-2^{-f}-2^{1-f})2^{-1}}{1-2^{1-f}-(1-2^{-f}-2^{1-f})2^{-1}}\right] \\ &= 2^{-f}\left[\frac{3(1-2^{-f})}{1-2^{-f}}\right] \\ &= 3\cdot 2^{-f}. \qquad (17) \end{aligned}$$

If $F-f\ge 2$, then

$$\begin{aligned} RE_{1A,\,J\le K}(I,J,K,n) &\le 2^{-f}\left[\frac{1+(1-2^{-f}-2^{-r})2^{f-F}}{1-2^{1-f}-(1-2^{-f}-2^{-r})2^{f-F}}\right] \\ &\le 2^{-f}\left[\frac{1+(1-2^{1-f})2^{-2}}{1-2^{1-f}-(1-2^{1-f})2^{-2}}\right] \\ &= 2^{-f}\frac{1}{3}\left[1+\frac{4}{1-2^{1-f}}\right] \\ &\le 3\cdot 2^{-f}. \qquad (18) \end{aligned}$$

Hence in either case, $RE_{1A,\,J\le K}(I,J,K,n)\le 3\cdot 2^{-f}$.

8.1.1.2 Case $I<J\le K$. We maximize $RE_{1A,\,J\le K}$ as follows. First, we need to confirm that the denominator $D_{1A,\,J\le K}(I,J,K,n)$ remains positive over the range of parameters, so that $RE_{1A,\,J\le K}(I,J,K,n)$ is bounded. Then we compute the derivatives of $RE_{1A,\,J\le K}(I,J,K,n)$ with respect to J and K in order to find the maximum.

Proof and Computation

Computational Mathematics

Proof and Computation

Computational Mathematics:

Write the program

Proof and Computation

Computational Mathematics:

Write the program

Prove it correct

Proof and Computation

Computational Mathematics:

Write the program

Prove it correct

Run it

Proof and Computation

Computational Mathematics:

Write the program

Prove it correct

Run it

Some verified implementations:
Gröbner, CAD, . . .

A Toy Example

How to define primality?

A Toy Example

How to define primality?

$$a \mid b \ \equiv_{def} \ \exists c, b = a * c$$

A Toy Example

How to define primality?

$$a \mid b \;\equiv_{def}\; \exists c,\, b = a * c$$

$$\begin{aligned} prime(p) \;\equiv_{def}\; & \forall c,\ c \mid p \Rightarrow (c = 1 \vee c = p) \\ \wedge \;\; & p \neq 1 \end{aligned}$$

Millenium Prime

16956227128806874788740039322257331454187103117215
25840282275463944443915017665187677648590458000952
76019439238167628964472781614506010476756059209553
52991146912809889393788383275988559054911295888721
82960597685441916678707363819484213131086251607398
28916259849380903171097200622744525334228076766180
55026575947807075176847196621981731534435122000052
36954810760431933960890325207991357855479116978257
24813770921462191448004112124119861318903084696307
90885182138135695302659166091775570804154531803228
62568934813720132935654045610123135360707455696482
05201111917803917089426925046726225659955364931026
16692889439256467875582491686567852505452615238453
74848173118991769725459297170194389891282683509781
98688634535632050868586493434240748304066486419914
61359441312447494379322488285783808178719040181648
61392979973303392716865005707782205062785523380586
42568033176696034210912974807429316748028057862495
35420221209942941757696865516674318082496965158247
33880524302307632732846854292869718923971577651169
23799050487122757054071124036147984242060317055743
21534640226758722180834588946485900779828265504316
34699098451988007731484430206887367460349041126438
05305579536776436596545370995982233304571888022964
36148103186029822738580486520602921469716777408128
90535728376984313845837331631683153208667889787320
54662757387113620757117588811133939675875891109260
92993622826554868172778946611292164358541457015429
81042899718955296718177187607200478963165719294795
40535020439735959026693283761942669724863257628123
19235317644619299197836613393426798742904083056775
41222944571332463609098549099900518313145113955639
57356654879738029278858802439304646158950266071940
44836695846663693272095862707204730604279754937134
76058004161080839413731262386511622381062542796281
60459016991526332735165194734663597772286383643625
30252777456951341275593886397426831655585085243207
96020479134995279005506892191527037067638499420678
41781864906059067455902834520379834660782200872459
36042046285093059704977817748127777709652208140691

is prime

Pocklington's Criterion

Test to assess the primality of a number n

Pocklington's Criterion

Test to assess the primality of a number n

- If there exist an a and $p_1, \ldots p_i$ primes such that

Pocklington's Criterion

Test to assess the primality of a number n

If there exist an a and $p_1, \ldots p_i$ primes such that

$$a^{n-1} \equiv 1[n]$$

Pocklington's Criterion

Test to assess the primality of a number n

If there exist an a and $p_1, \ldots p_i$ primes such that

$$a^{n-1} \equiv 1[n]$$

$$n - 1 = p_1^{k_1} \ldots p_i^{k_i} r \text{ and } \sqrt{n} \leq p_1^{k_1} \ldots p_i^{k_i}$$

Pocklington's Criterion

Test to assess the primality of a number n

If there exist an a and $p_1, \ldots p_i$ primes such that

$$a^{n-1} \equiv 1[n]$$

$n - 1 = p_1^{k_1} \ldots p_i^{k_i} r$ and $\sqrt{n} \leq p_1^{k_1} \ldots p_i^{k_i}$

$gcd(a^{n-1/p_j} - 1, n) = 1$ for all $j = 1 \ldots i$

Pocklington's Criterion

Test to assess the primality of a number n

If there exist an a and $p_1, \ldots p_i$ primes such that

$a^{n-1} \equiv 1[n]$

$n - 1 = p_1^{k_1} \ldots p_i^{k_i} r$ and $\sqrt{n} \leq p_1^{k_1} \ldots p_i^{k_i}$

$gcd(a^{n-1/p_j} - 1, n) = 1$ for all $j = 1 \ldots i$

Then n is prime.

Pocklington's Criterion

Test to assess the primality of a number n

If there exist an a and $p_1, \ldots p_i$ primes such that

$$a^{n-1} \equiv 1[n]$$

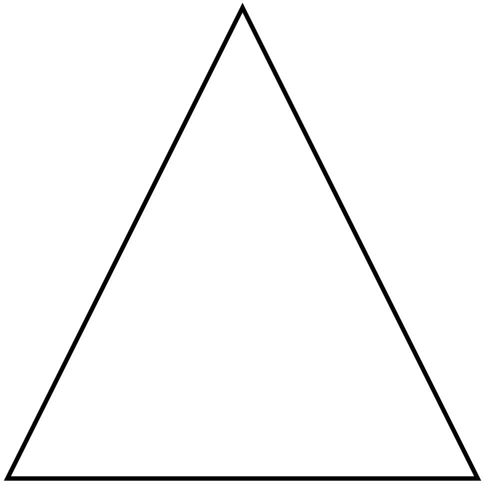
$$n - 1 = p_1^{k_1} \ldots p_i^{k_i} r \text{ and } \sqrt{n} \leq p_1^{k_1} \ldots p_i^{k_i}$$

$$gcd(a^{n-1/p_j} - 1, n) = 1 \text{ for all } j = 1 \ldots i$$

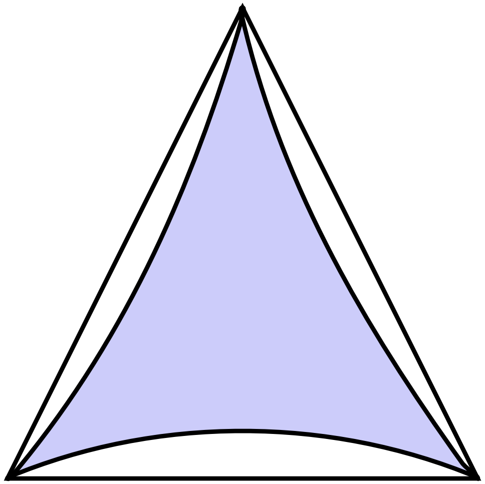
Then n is prime.

For the millenium prime, $a = 2$, $p_1 = 2161$, $p_2 = 2$.

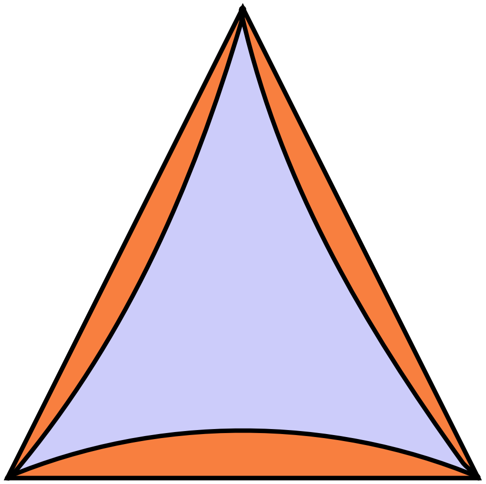
Mechanized Proof



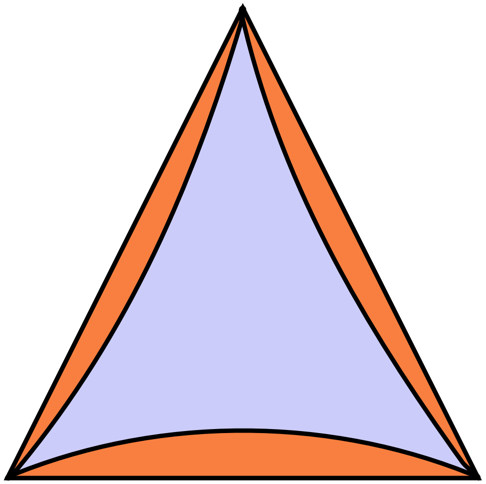
Mechanized Proof



Mechanized Proof

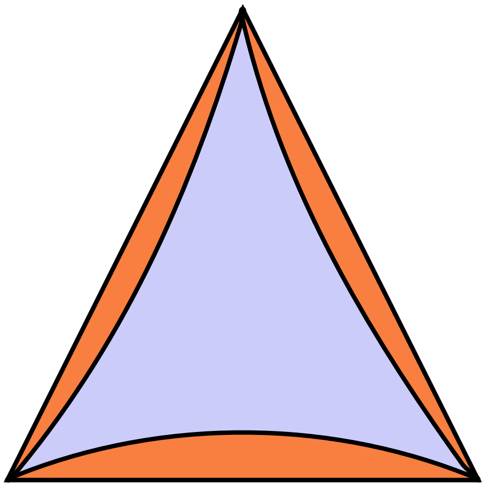


Mechanized Proof



Four Colour Theorem

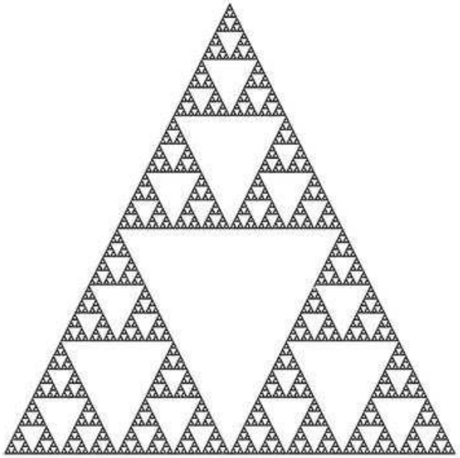
Mechanized Proof



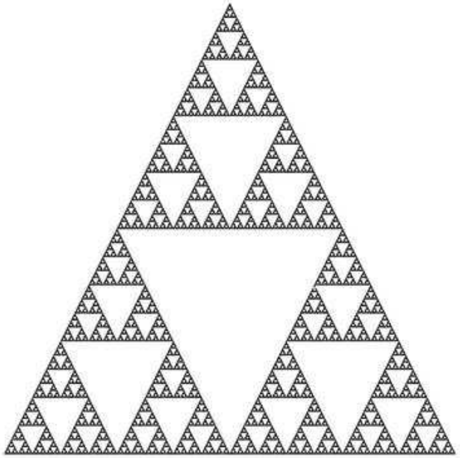
Four Colour Theorem

Flyspeck Project

Mechanized Proof



Mechanized Proof



Enormous Theorem

Challenges

Challenges

Pen and Paper + Formal Proof

Challenges

Pen and Paper + Formal Proof

Proof + Computation

Challenges

Pen and Paper + Formal Proof

Proof + Computation

Collective effort

Challenges

Pen and Paper + Formal Proof

Proof + Computation

Collective effort