Formal Global Optimisation with Taylor Models

TYPES
20 April 2006

Roland Zumkeller
LogiCal project, École Polytechnique, Paris

Formal Global Optimisation with Taylor Models

TYPES
20 April 2006

Roland Zumkeller
LogiCal project, Ecole Polytechnique, Paris

Outline

- Global Optimisation
  - Problem: finding the minimum and maximum value of a given objective function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ on a certain domain $[a_1; b_1] \times \ldots \times [a_n; b_n]$
  - Traditional methods based on interval arithmetic.
- Taylor Models
  - An optimisation method computing with sets of functions of the form "polynomial + error interval"
- Formal
  - Both the optimisation algorithm (-Cal) and its correctness proof (Logi-) are Coq terms. To obtain a formal proof for a particular problem it suffices to execute the algorithm (proof by reflection).

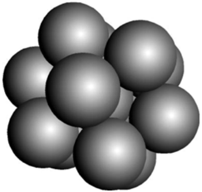
What Is Global Optimisation Good for?

- Engineering: aeronautics, robotics, ...
- Experimental physics: particle motion in accelerators
- Geometry: volumes of cells in space decomposition occuring in T. Hales' proof of the Kepler conjecture

$$x_1 \in [4; 2.168^2] \wedge x_2 \in [6.3001; 2.696^2] \wedge x_3 \in [4; 2.168^2] \wedge$$
$$x_4 \in [4; 6.3001] \wedge x_5 \in [4; 6.3001] \wedge x_6 \in [4; 6.3001] \rightarrow$$
$$\frac{\pi}{2} + \arctan\left(- \frac{-x_1^2 - (x_3 - x_5)(x_2 - x_6) + (x_1(x_2 + x_3 - 2x_4 + x_5 + x_6))}{\sqrt{\begin{array}{l} 4x_1(-x_1^2 * x_4 - x_2^2 * x_5 - (x_3 - x_4)(x_3 - x_5)x_6 - \\ \quad x_3 x_6^2 + x_2((x_3 - x_5)(x_5 - x_4) + (x_3 + x_5)x_6) + \\ \quad x_1((x_3 - x_4)(x_4 - x_5) + (x_3 + x_4)x_6 + x_2(x_4 + x_5 - x_6))) \end{array}}} \right) > 0.74$$

The Kepler Conjecture (1611)

- The maximal density of any sphere packing in 3-space is $\frac{\pi}{\sqrt{18}}$.



- In 1998 Thomas C. Hales has found a proof, which is large in every sense: article of 300 pages, 40.000 lines of code, several weeks of computation
- The "**Flyspeck**" project aims at formalising this proof, in order to eliminate any doubt about its correctness.

Interval Arithmetic

- The set of intervals: $\mathbb{I} = (\mathbb{R} \cup \{-\infty\}) \times (\mathbb{R} \cup \{\infty\})$
- Let $f : \mathbb{R}^n \to \mathbb{R}$. The function $\hat{f} : \mathbb{I}^n \to \mathbb{I}$ is called ...
  - an *extension* of f iff $\quad \forall X \in \mathbb{I}^n.\ \hat{f}(X) \supseteq \{f\ x \mid x \in X\}$
  - a *sharp extension* of f iff $\quad \forall X \in \mathbb{I}^n.\ \hat{f}(X) = \{f\ x \mid x \in X\}$
- Some sharp extensions:

$$
\begin{aligned}
[a;b] \mathbin{\hat{+}} [c;d] &:= (a+c, b+d) \\
[a;b] \mathbin{\hat{-}} [c;d] &:= (a-d, b-c) \\
[a;b] \mathbin{\hat{*}} [c;d] &:= (\min\{ac, ad, bc, bd\}, \max\{ac, ad, bc, bd\}) \\
1 \mathbin{\hat{/}} [a;b] &:= \begin{cases} (\frac{1}{b}, \frac{1}{a}) & \text{if } a \geq 0 \vee b \leq 0 \\ (-\infty, \infty) & \text{if } a \leq 0 \leq b \end{cases}
\end{aligned}
$$

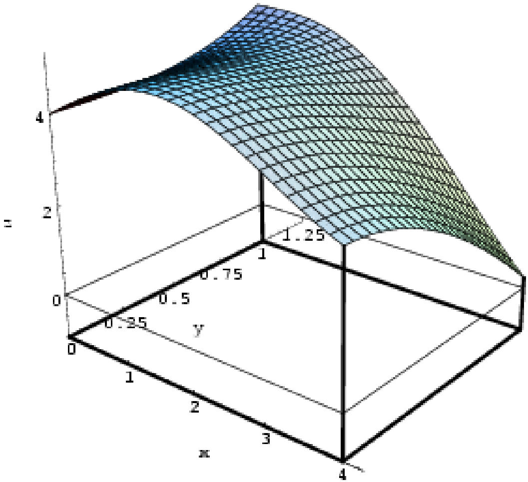
- Structural recursion with these extensions over a term yields its *natural extension*.

The Dependence Problem

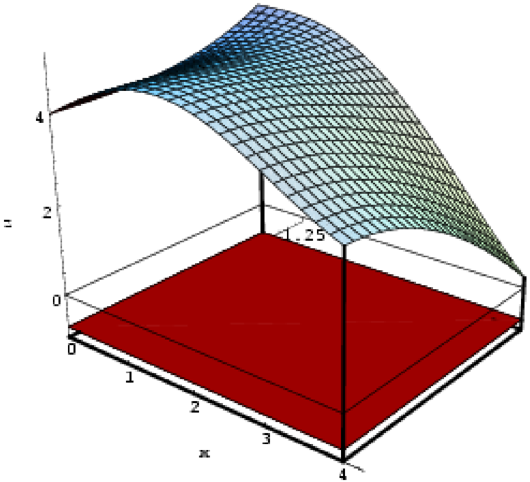
- Goal: $x \in [1;2] \wedge y, z \in [2;3] \rightarrow xy - z \geq -5$
  Proof: $[1;2] \mathbin{\hat{*}} [2;3] \mathbin{\hat{-}} [2;3] = [-1;3] \geq -5$
  structural induction; apply extension property. q.e.d.
- Goal: $x \in [3;5] \rightarrow x - x \geq -1$
  Proof: $[3;5] \mathbin{\hat{-}} [3;5] = [-2;2] \not\geq -1$
  :-(
- Why? For interval arithmetic the second goal looks like:
  $x, y \in [3;5] \rightarrow x - y \geq -1$

A Remedy: Branch & Bound

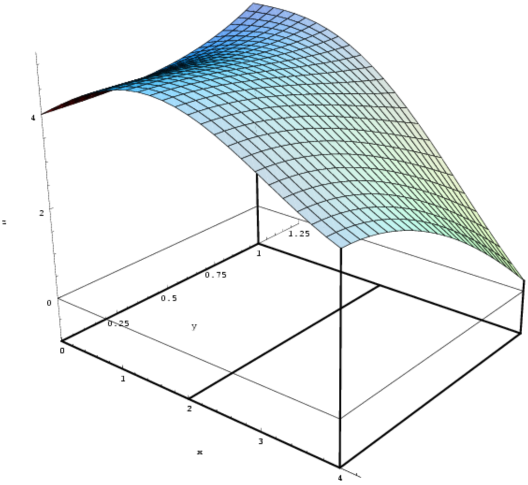
- If a simple evaluation of the natural extension fails, we split the domain X into X_1 and X_2. From the extension property follows: $x \in X_1 \vee x \in X_2 \to x \in \hat{f}(X_1) \cup \hat{f}(X_2)$
- Goal: $x \in [3;5] \to x - x \geq -1$
  Proof: $\left([3;4] \hat{-} [3;4]\right) \cup \left([4;5] \hat{-} [4;5]\right) = [-1,1]$
  ...structural induction; apply extension property. q.e.d.
- Algorithm for proving $x \in X \to f\ x \geq 0$:
  - $\hat{f}(X) \geq 0$: success
  - $\hat{f}(X) \not\geq 0$: split X into X_1 and X_2 and restart for each one



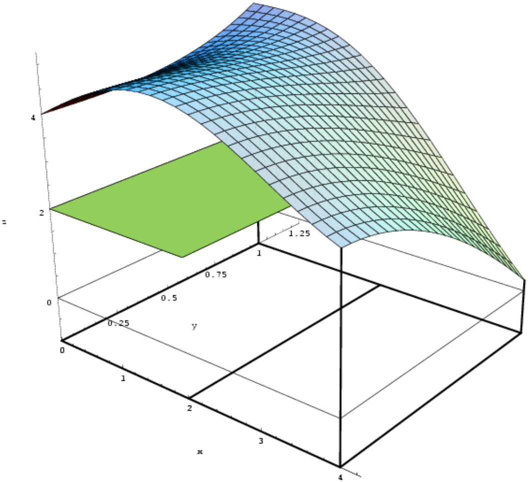
$$\sin x + y^2(y - x) + 4 > 0$$



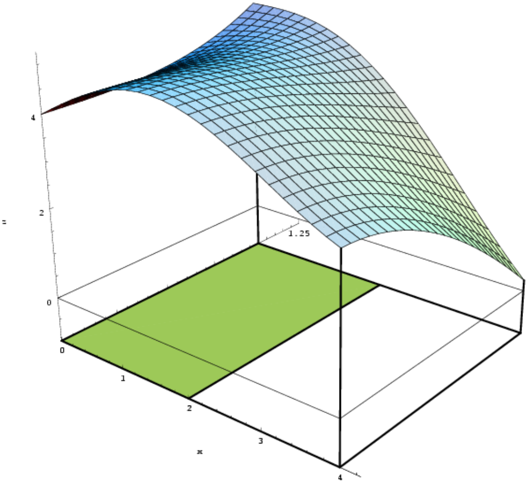
$$\sin x + y^2(y - x) + 4 > 0$$



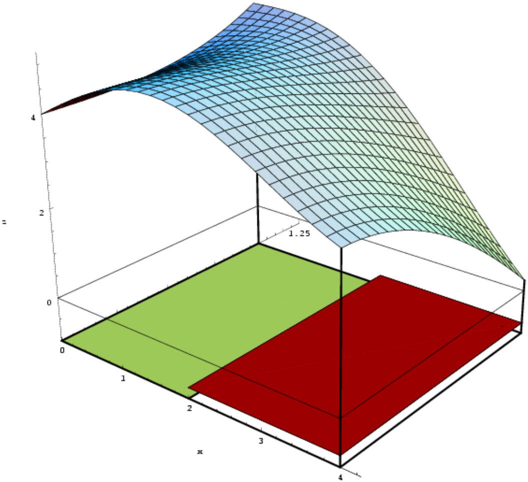
$$\sin x + y^2(y - x) + 4 > 0$$



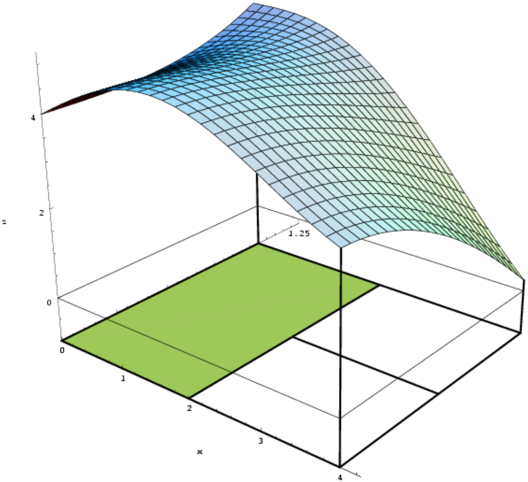
$\sin x + y^2(y - x) + 4 > 0$



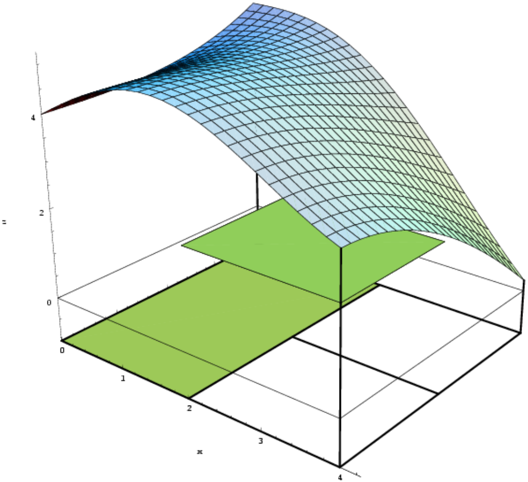
$$\sin x + y^2(y - x) + 4 > 0$$



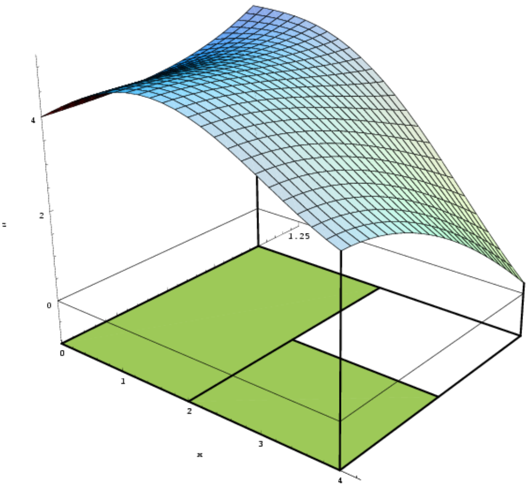
$$\sin x + y^2(y - x) + 4 > 0$$



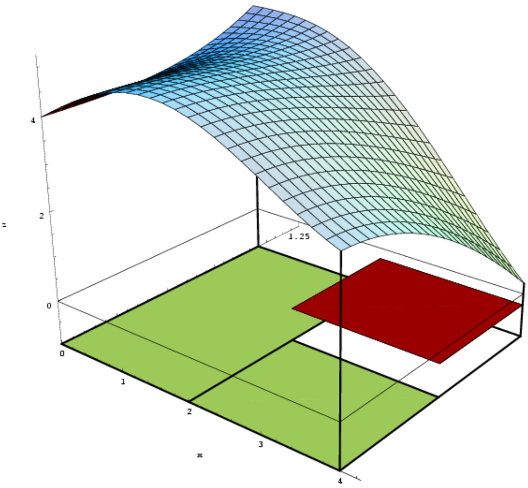
$$\sin x + y^2(y - x) + 4 > 0$$



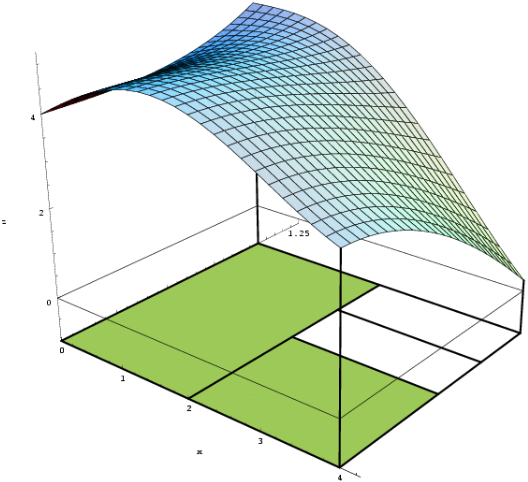
$$\sin x + y^2(y - x) + 4 > 0$$



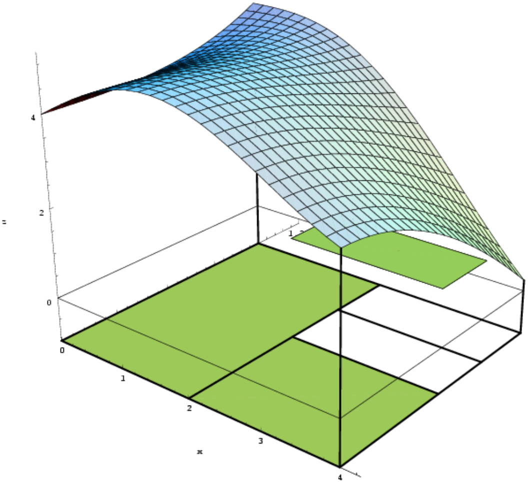
$$\sin x + y^2(y - x) + 4 > 0$$



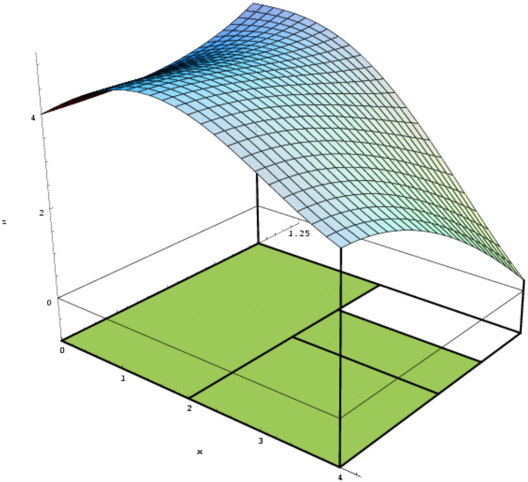
$$\sin x + y^2(y - x) + 4 > 0$$



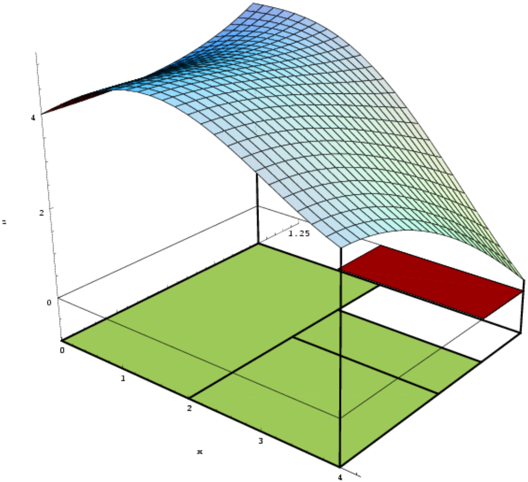
$$\sin x + y^2(y - x) + 4 > 0$$



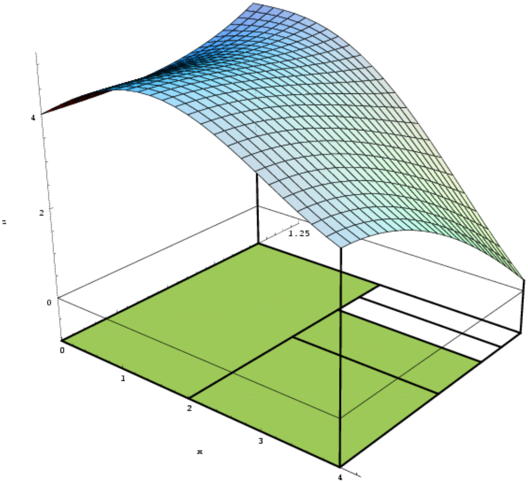
$$\sin x + y^2(y - x) + 4 > 0$$



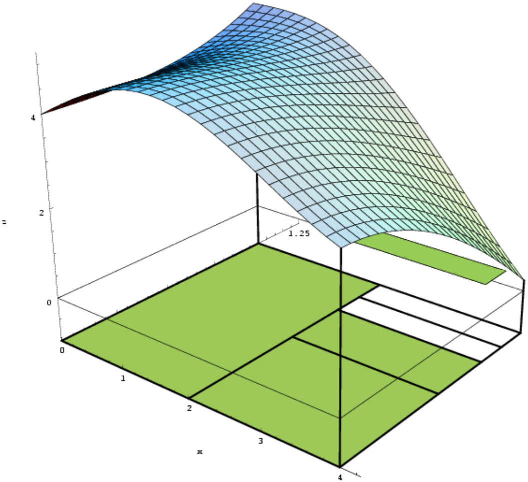
$$\sin x + y^2(y - x) + 4 > 0$$



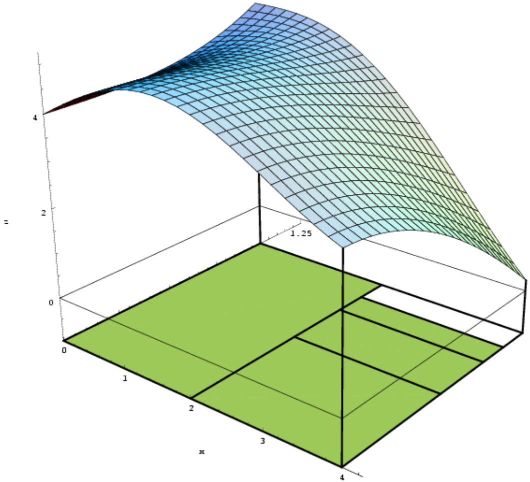
$$\sin x + y^2(y - x) + 4 > 0$$



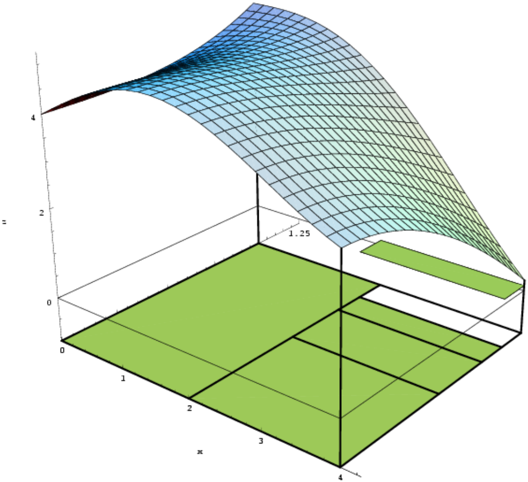
$$\sin x + y^2(y - x) + 4 > 0$$



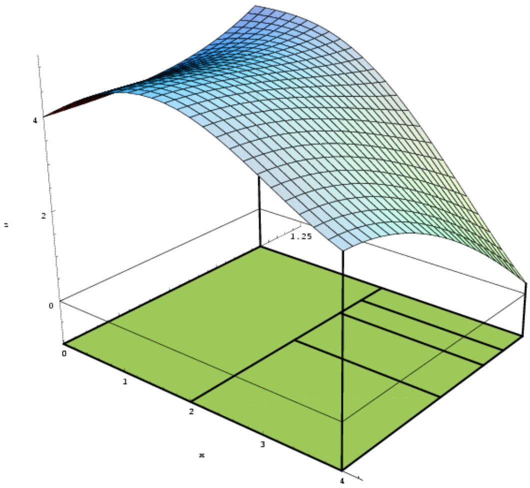
$$\sin x + y^2(y - x) + 4 > 0$$



$\sin x + y^2(y - x) + 4 > 0$



$$\sin x + y^2(y - x) + 4 > 0$$



$$\sin x + y^2(y - x) + 4 > 0$$

Another Remedy: Use of The Gradient

- Fermat/Euler: x is a local extremum $\rightarrow \nabla f(x) = 0$
- $\widehat{\nabla f}(X) \not\ni 0 \rightarrow X$ does not contain a local extremum
- A *global* extremum is either a local one, or it lies on the border of the domain.
- If (after several splits) a sub-domain X does not touch the original domain's borders and $\widehat{\nabla f}(X) \not\ni 0$, then it can be safely forgotten.
- If $\widehat{\nabla f}(X) \not\ni 0$ and it still touches some borders, then there is some i with $\widehat{\partial_i f}(X) > 0$.
  So $x \in X \rightarrow x \in \hat{f}(X[i := \underline{X_i}]) \cup \hat{f}(X[i := \overline{X_i}])$

The Choice of Interval Bounds

- Most implementations use floating-point numbers as interval bounds.
- Therefore irrational functions have to be approximated with a *pre-defined* precision.
- For example with $\sqrt{(0,2)} = (\lfloor\sqrt{0}\rfloor, \lceil\sqrt{2}\rceil) = (0, 1.42)$ the inequality $x \in (0,2) \to \sqrt{x} \leq 1.416$ can't be proved.
- This is the floating-point numbers' fault!

Use of Constructive Real Numbers

- From Russell's talk: $\mathbb{R} \subset \mathbb{Q} \to \mathbb{Q}$
- With $\mathbb{I} = (\mathbb{R} \cup \{-\infty\}) \times (\mathbb{R} \cup \{\infty\})$ the united extension

$$\hat{f}_n\,[a;b] := \bigcup_{i=1}^{n} \hat{f}\,\left[a + (i-1)\frac{b-a}{n}; a + i\frac{b-a}{n}\right]$$

converges towards f's actual bounds, i.e.

$$X \mapsto \lim_{n\to\infty} \hat{f}_n\,X$$

is sharp.

- Constructive reals can be faster than rationals with fixed-precision operations (precisely when the precision actually necessary is less). But they can also be slower ...

Solving the Dependence Problem: Taylor Models

- For interval arithmetic $x - x$ on X looks like $x - y$ with $X := Y$
- A Taylor model represents a set of functions:

$$\mathbb{T} \ni (X, P, \Delta) \cong \{f : X \to \mathbb{R} \mid \forall x \in X.\, f\,x - P\,x \in \Delta\}$$

  - *domain* $X : \mathbb{I}^k$
  - *polynomial* $P : \mathbb{R}[k]$
  - *error bound* $\Delta : \mathbb{I}$
- Assuming we have some polynomial bounder B available, we can bound all functions in a Taylor model by $B\,X\,P + \Delta$
- How to obtain Taylor Models?
  - by Taylor's theorem with Lagrange remainder
  - by composition ...

Arithmetic on Taylor Models

- Constants, variables: trivial ($\Delta := [0;0]$)
- Addition and multiplication:

$$
\begin{aligned}
(X, P_1, \Delta_1) \tilde{+} (X, P_2, \Delta_2) &= (X, P_1 +_{\mathbb{R}[k]} P_2, \Delta_1 \hat{+} \Delta_2) \\
(X, P_1, \Delta_1) \tilde{\cdot} (X, P_2, \Delta_2) &= (X, (P_1 \cdot_{\mathbb{R}[k]} P_2)_{\leq n}, B\, X\, (P_1 \cdot_{\mathbb{R}[k]} P_2)_{>n} \hat{+} \\
&\qquad B\, P_1\, X \hat{\cdot} \Delta_2 \hat{+} \Delta_1 \hat{\cdot} B\, P_2\, X \hat{+} \Delta_1 \hat{\cdot} \Delta_2)
\end{aligned}
$$

- $\tilde{f} : \mathbb{T}^m \to \mathbb{T}$ is a Taylor-extension of $f : \mathbb{R}^m \to \mathbb{R}$ iff:

$$\forall T.\ [|\tilde{f}\ T_1 \ldots T_m|] \supseteq \{x \mapsto f\ (t_1\, x) \ldots (t_2\, x) \mid \forall i \leq m.\ t_i \in [|T_i|]\}$$

Combining Smooth Functions with Taylor Models

Makino/Berz: First apply an addition theorem, depending on the function under consideration. Then apply Taylor's theorem with Lagrange's remainder.

$$
\begin{aligned}
\log \circ F &= \log \circ (c + \bar{F}) \stackrel{\textit{Heureka!}}{=} \log c + \log \circ \left(1 + \frac{\bar{F}}{c}\right) \\
&\in \log c + \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \left(\frac{\bar{F}}{c}\right)^k + \frac{(-1)^n \left(\frac{B\,\bar{F}\,X}{c}\right)^{n+1}}{(n+1)\left(1 + \left[0, \frac{B\,\bar{F}\,X}{c}\right]\right)^{n+1}}
\end{aligned}
$$

where X the domain under consideration.
In [Makino/Berz] this procedure is applied to exp, log, inv, sqrt, sin, cos, sinh, cosh, arcsin, arccos, arctan.

Combining Smooth Functions with Taylor Models without Heureka

$$\log \circ F \subseteq \log y_0 + \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k y_0^k} (F - y_0)^k + \frac{(-1)^n (B\,F\,X - y_0)^{n+1}}{(n+1)[y_0, B\,F\,X]^{n+1}}$$

- for $y_0 = c$ this is equivalent to Makino/Berz's version
- Advantages:
  - Implementation and proofs can be factorised.
  - Better choices for y_0 are possible.

Computational Proof by Reflection

- This approach has been successfully applied to the four colour theorem [Gonthier/Werner] and to Pocklington certificates for prime numbers [Grégoire/Thery/Werner]
- The tactic is a program written in Coq's term language:
  `test : list intvl -> term -> nat -> bool`
- ```
  test_correct :
    forall (X : list intvl) (t : term) (n : nat),
    test X t n = true ->
    forall x:R, contains x X ->
      interpR x t >= 0
  ```
- The trace does not need to be stored:
  ```
  test_correct X t n (refl_equal true) :
    forall x:R, contains x X -> interpR x t >= 0
  ```
  ```

Future Work

- Better polynomial bounding algorithms: vast choice
- `let x = pi^2 in x + x` performs two approximations of `pi^2`
- Provide a user-friendly Coq tactic.