Recent Advances in Taylor Model based Rigorous Global Optimization

Kyoko Makino and Martin Berz

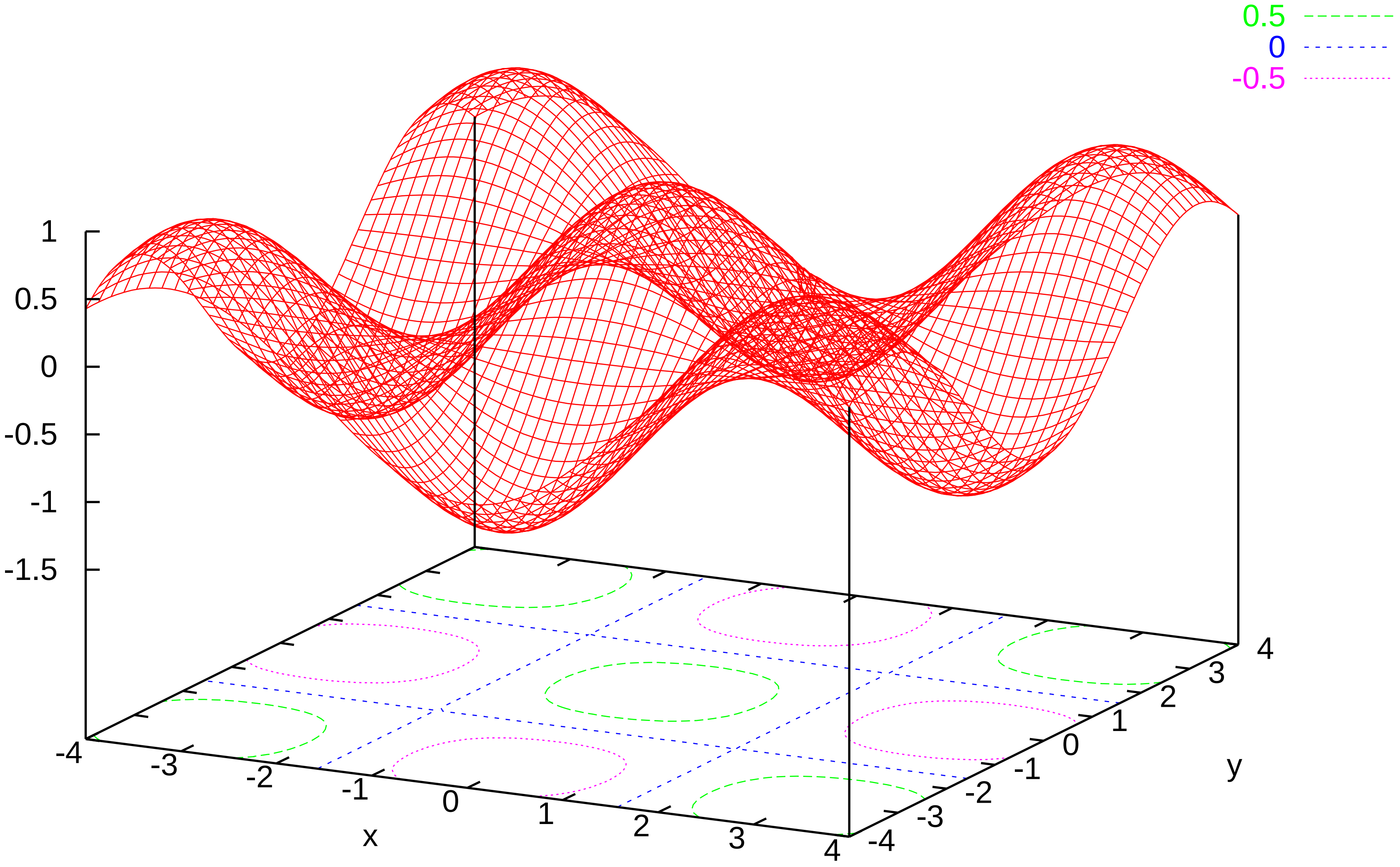
Department of Physics and Astronomy
Michigan State University

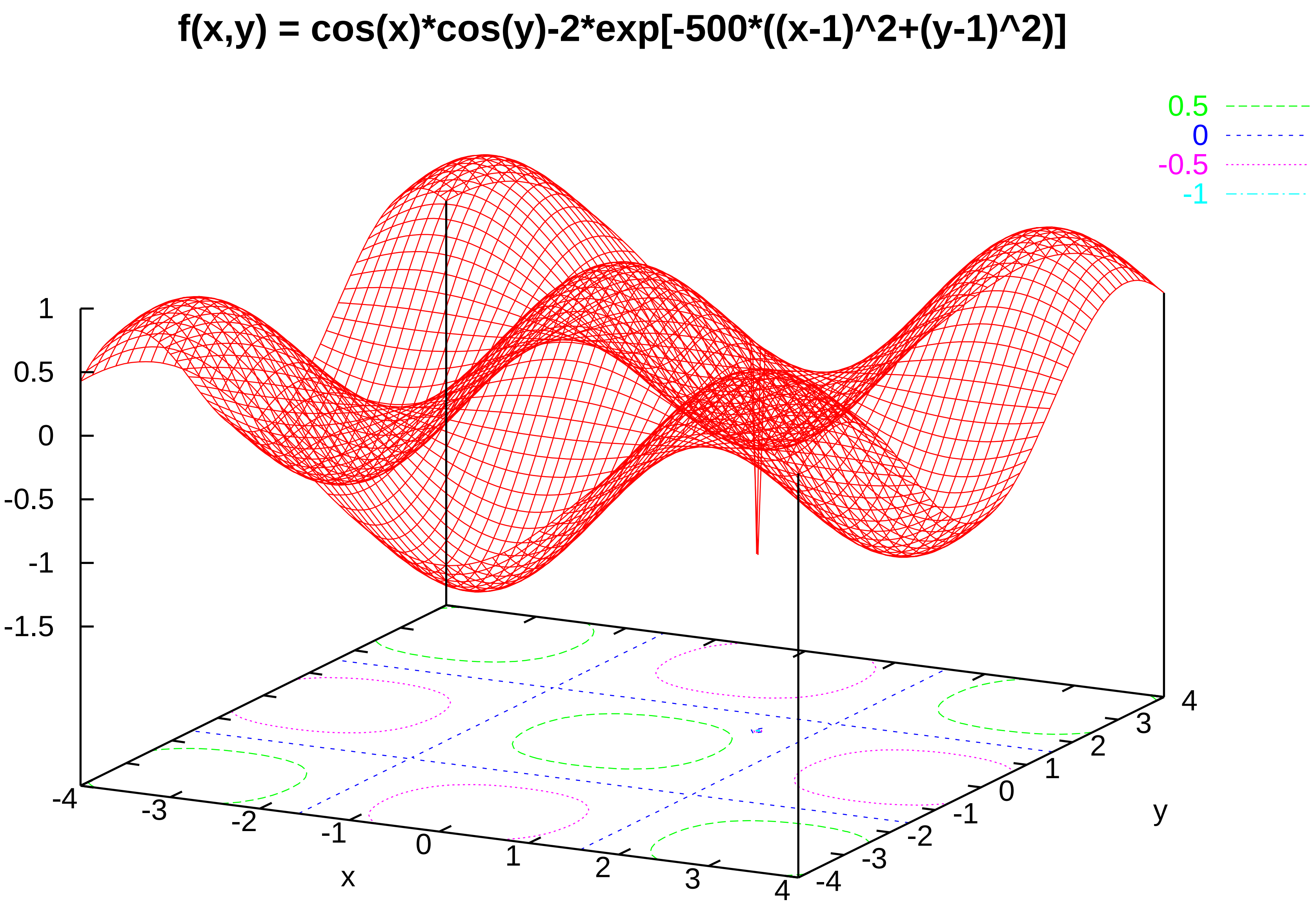
Example

An extremely simplified model of some QCD optimization problems, scaled down to 2D for the illustration. Find the minimum of the function $f(x, y)$:

$$f(x, y) = \cos x \cos y - 2 \exp\left[-500 \cdot \left((x-1)^2 + (y-1)^2\right)\right]$$

f(x,y) = cos(x)*cos(y)-2*exp[-500*((x-1)^2+(y-1)^2)]




f(x,y) = cos(x)*cos(y)-2*exp[-500*((x-1)^2+(y-1)^2)]
0.5
0
-0.5
-1
1
0.5
0
-0.5
-1
-1.5
-4
-3
-2
-1
0
1
2
3
4
x
-4
-3
-2
-1
0
1
2
3
4
y

Example

An extremely simplified model of some QCD optimization problems, scaled down to 2D for the illustration. Find the minimum of the function $f(x, y)$:

$$f(x, y) = \cos x \cos y - 2 \exp\left[-500 \cdot \left((x-1)^2 + (y-1)^2\right)\right]$$

- A spike on top of the periodic cos behavior.

$$\min = -1.70817675216007^{26}_{39} \qquad \text{at } x = y = 1.0002272^{92}_{85}.$$

Example

An extremely simplified model of some QCD optimization problems, scaled down to 2D for the illustration. Find the minimum of the function $f(x, y)$:

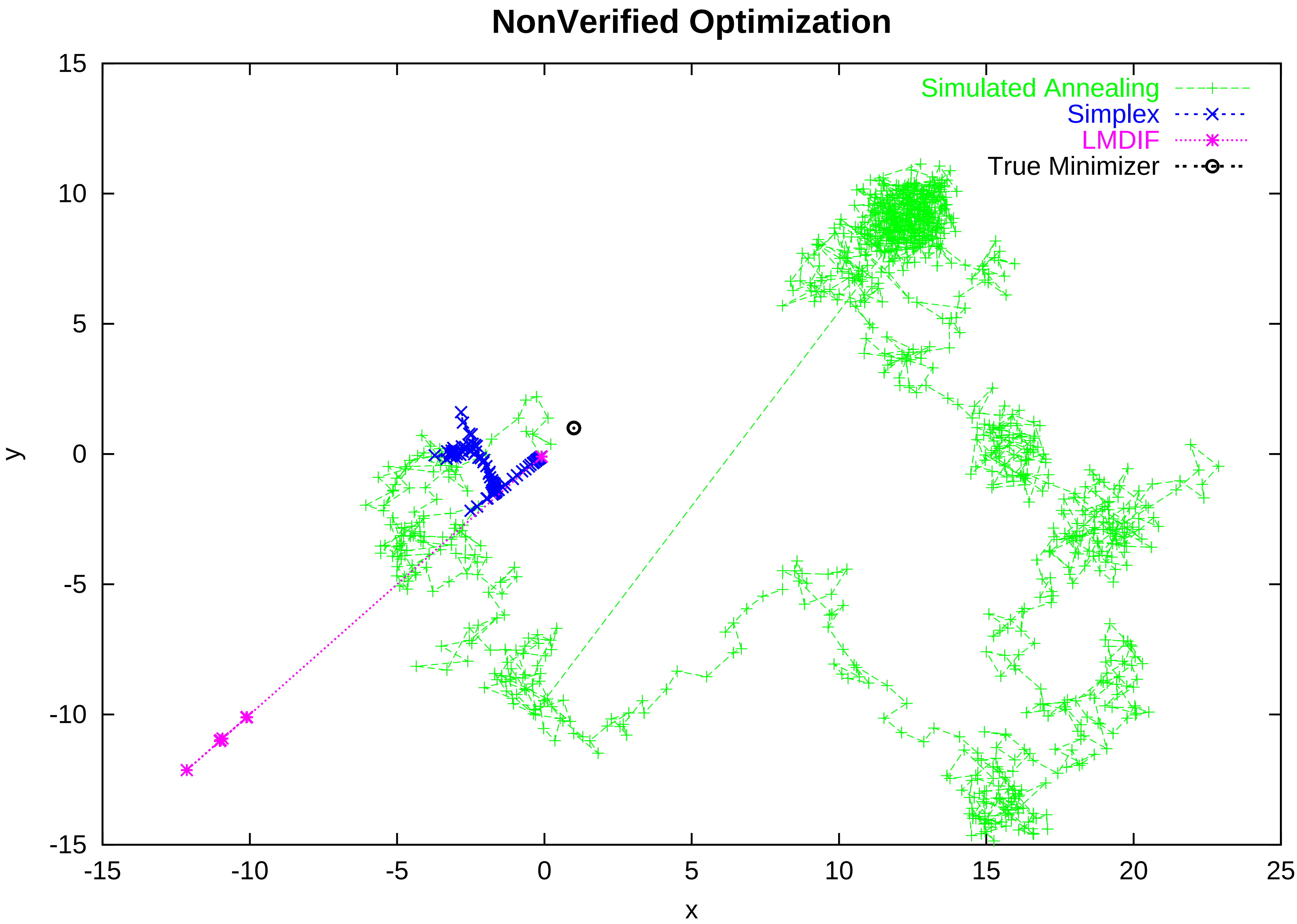
$$f(x, y) = \cos x \cos y - 2 \exp\left[-500 \cdot \left((x-1)^2 + (y-1)^2\right)\right]$$

- A spike on top of the periodic $\cos$ behavior.

$$\min = -1.70817675216607^{26}_{39} \quad \text{at } x = y = 1.0002272^{92}_{85}.$$

- Use various local optimization algorithms
  Simulated Annealing, Simplex, LMDIF (available as "**FIT**" in COSY)

Algorithm	Without Constraint		Constraint $\lvert x\rvert, \lvert y\rvert \leq 4$	
	Steps	min	Steps	min
Simulated Annealing	1000	~ -1		
Simplex	130	~ -1		
LMDIF	27	~ 0		


NonVerified Optimization
Simulated Annealing
Simplex
LMDIF
True Minimizer
15
10
5
0
-5
-10
-15
y
-15
-10
-5
0
5
10
15
20
25
x

Example

An extremely simplified model of some QCD optimization problems, scaled down to 2D for the illustration. Find the minimum of the function $f(x,y)$:

$$f(x,y) = \cos x \cos y - 2\exp\left[-500 \cdot \left((x-1)^2 + (y-1)^2\right)\right]$$

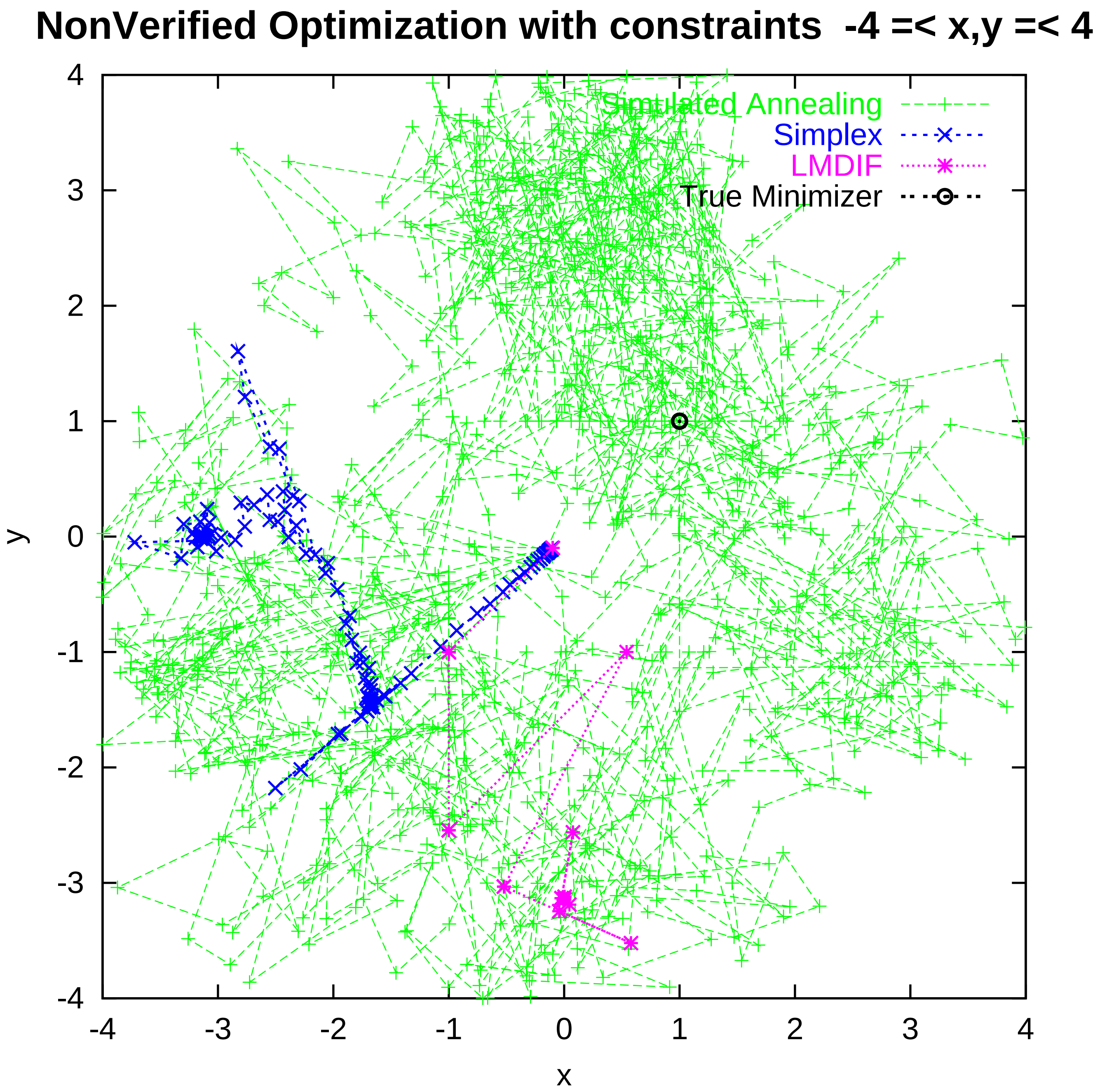
- A spike on top of the periodic $\cos$ behavior.

$$\min = -1.7081767521607^{26}_{39} \qquad \text{at } x = y = 1.0002272^{92}_{85}.$$

- Use various local optimization algorithms

  Simulated Annealing, Simplex, LMDIF (available as “**FIT**” in COSY)

Algorithm	Without Constraint		Constraint $\lvert x\rvert, \lvert y\rvert \le 4$	
	Steps	min	Steps	min
Simulated Annealing	1000	~ -1	1000	10^{-2} accurate
Simplex	130	~ -1	130	~ -1
LMDIF	27	~ 0	57	~ -1


NonVerified Optimization with constraints -4 =< x,y =< 4
Simulated Annealing
Simplex
LMDIF
True Minimizer
x
y
-4
-3
-2
-1
0
1
2
3
4

Example

An extremely simplified model of some QCD optimization problems, scaled down to 2D for the illustration. Find the minimum of the function $f(x, y)$:

$$f(x, y) = \cos x \cos y - 2 \exp\left[-500 \cdot \left((x-1)^2 + (y-1)^2\right)\right]$$

- A spike on top of the periodic cos behavior.

$$\min = -1.7081767521607^{26}_{39} \quad \text{at } x = y = 1.0002272^{92}_{85}.$$

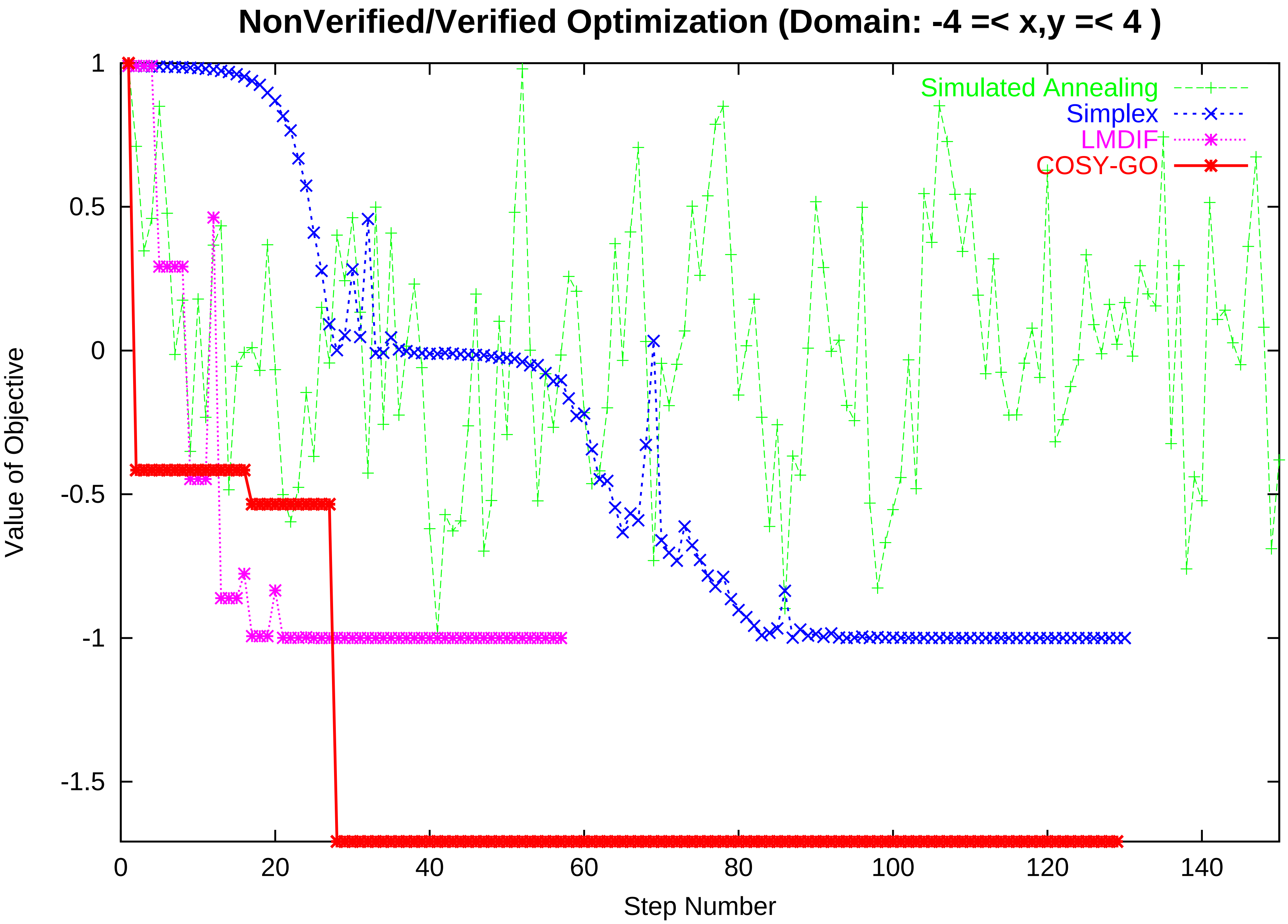
- Use various local optimization algorithms

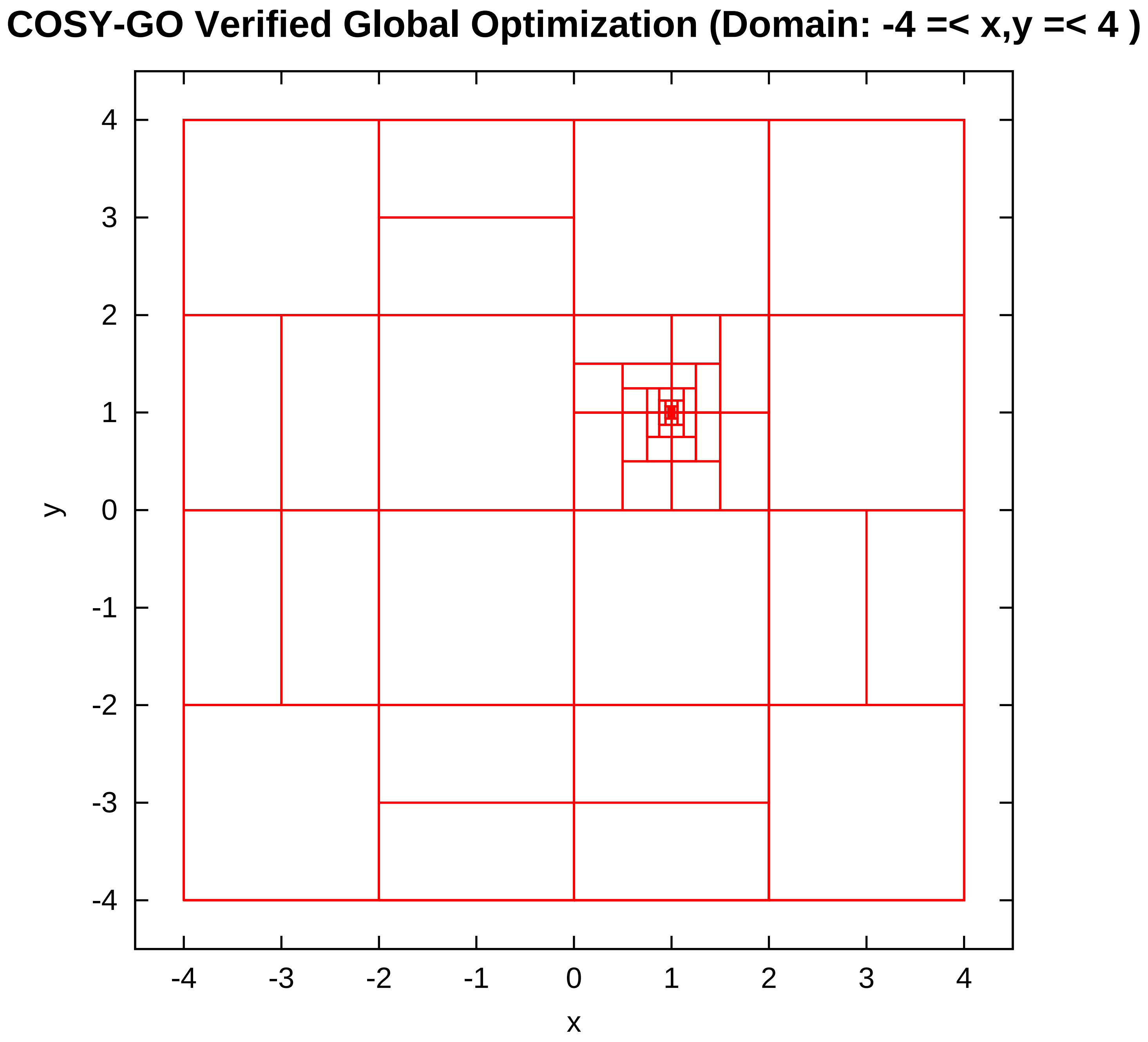
  Simulated Annealing, Simplex, LMDIF (available as “**FIT**” in COSY)

Algorithm	Without Constraint		Constraint $\lvert x\rvert, \lvert y\rvert \le 4$	
	Steps	min	Steps	min
Simulated Annealing	1000	~ -1	1000	10^{-2} accurate
Simplex	130	~ -1	130	~ -1
LMDIF	27	~ 0	57	~ -1

- Use COSY-GO (**verified** global optimizer)

  In the search domain $[-4, 4] \times [-4, 4]$


NonVerified/Verified Optimization (Domain: -4 =< x,y =< 4)
Value of Objective
Step Number
1
0.5
0
-0.5
-1
-1.5
0
20
40
60
80
100
120
140
Simulated Annealing
Simplex
LMDIF
COSY-GO


COSY-GO Verified Global Optimization (Domain: -4 =< x,y =< 4)
4
3
2
1
0
-1
-2
-3
-4
y
-4
-3
-2
-1
0
1
2
3
4
x

Example

An extremely simplified model of some QCD optimization problems, scaled down to 2D for the illustration. Find the minimum of the function $f(x, y)$:

$$f(x, y) = \cos x \cos y - 2 \exp \left[-500 \cdot \left((x-1)^2 + (y-1)^2 \right) \right]$$

- A spike on top of the periodic cos behavior.

$$\min = -1.708176752160 7_{39}^{26} \quad \text{at } x = y = 1.0002272_{85}^{92}.$$

- Use various local optimization algorithms

  Simulated Annealing, Simplex, LMDIF (available as "**FIT**" in COSY)

Algorithm	Without Constraint		Constraint $\lvert x\rvert, \lvert y\rvert \le 4$	
	Steps	min	Steps	min
Simulated Annealing	1000	~ -1	1000	10^{-2} accurate
Simplex	130	~ -1	130	~ -1
LMDIF	27	~ 0	57	~ -1

- Use COSY-GO (**verified** global optimizer)

  In the search domain $[-4, 4] \times [-4, 4]$, the minimum is found with 10^{-14} accuracy in 129 steps. The minimizer is localized in the volume $5 \cdot 10^{-17}$.

The TM based Global Optimizer, COSY-GO

has utilized various algorothms based on Taylor models.

- LDB (Linear Dominated Bounding) bounding and domain reduction
- QFB (Quadratic Fast Bounding) bounding and domain reduction for positive definate cases (Quadratic pruning)
- Various cutoff value update schemes

Quadratic Pruning - The Idea

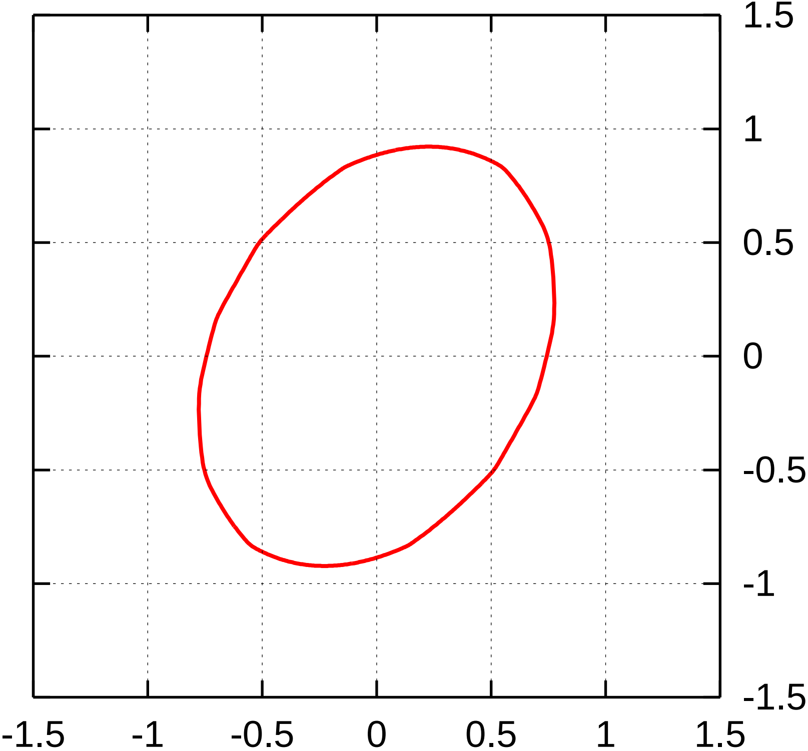
Extract a convex quadratic part P_2 of Taylor model, write

$$f(x) \in P_2(x) + R(x) + I \text{ where}$$
$$P_2(x) = \frac{1}{2} x^t \cdot H \cdot x$$

Want to confine the region $P_2(x) \leq a$ with $a > 0$, by an interval box $[-x_m, x_m]$ with $x_m > 0$.

Because of positive definiteness and convexity, this region is inside a **closed ellipsoidal contour surface** $P_2(x) = a$. The optimal confining interval box touches such a region at each box side surface tangentially, so the condition to find x_m is ∇f is normal to the corresponding box surface; namely for determining the k-th component $x_{mk,}$

$$(\nabla P)_i = 0 \text{ for } \forall i \neq k.$$


1.5
1
0.5
0
-0.5
-1
-1.5
-1.5
-1
-0.5
0
0.5
1
1.5

The TM based Global Optimizer, COSY-GO

has utilized various algorothms based on Taylor models.

- LDB (Linear Dominated Bounding) bounding and domain reduction
- QFB (Quadratic Fast Bounding) bounding and domain reduction for positive definate cases (Quadratic pruning)
- Various cutoff value update schemes

And, we have completed

- Adjustment to pallarel environments with low inter-processoer communication rate
- Restart capability
- Continuation of computations while the underlying arithmetic fails
- COSY INFINITY Version 9.0 has been released

Some General Thoughts about Rigorous Parallel Optimization

1. Performance gains in modern computing are gained through multi-processor architectures, not increased clock speed and more efficient microcode.
2. While the global optimization task does not parallelize trivially, with due care it is manageable

Caveats:

1. Communication mode, in particular for large numbers of processors - point to point, master - slave, common meeting?
2. Load balancing, in particular with many processors and slow connections

COSY-GO in Parallel Environment

Design aspects of COSY-GO-P

1. Utilize MPI and be standard. This is done with a COSY language construct called **PLOOP**, a parallel loop with various types inter-processor updates upon conclusion. Can be **nested**.
2. Should scale from for **different numbers of processors**
   (a) multiple cores in a chip
   (b) large clusters with thousands of processors
3. Should scale for **different connection speeds**
   (a) extremely fast interconnect (multiple cores in one chip)
   (b) very fast (a few cores in a "node" with a nearly bus-like interconnect)
   (c) fast (specialized network for parallel use, at least Gigabit)
   (d) slow (grid-based systems - geographicly dispersed, relying on standard Internet)

Basic Ideas of the COSY-GO Parallel Environment

1. **List Management:** Each processor has two lists:
   (a) **Short List of large boxes**, shared with other processors
   (b) A section of Short List is pre-allocated to each processor.
   (c) **Long List of regular boxes** owned by each processor.
   (d) Long List is kept in moderately strict order of difficulty. Achieved by selection strategy favoring newer boxes
2. **Communication Concept**
   (a) Processors communicate in **sheduled meeting mode** after predetermined fixed time interval T_m.
   (b) Time interval T_m is determined by trial and error for each environment under consideration. Single node: fraction of second, Berkeley NERSC cluster (~6000 processors): 1-2 minutes, Grid systems: fractions of hours.

What Happens in a Meeting

1. **Assess status**. Gather short data from each processor, scatter this information to all others. Cutoff updates, number of remaining large boxes and small boxes

2. **Processing of results.** Global cutoff is updated; it is determined if we can stop code

3. **Processing of status**. Each processor simultaneously identifies

   (a) how many boxes N_r are needed to replenish **Short List**

   (b) Let $N_p = N_r / N_{proc}$

4. **Load balancing.**

   (a) Each processor uploads its N_p largest boxes, if available, to the Short List

   (b) The Short List is randomized, so that the sections allocated to each processor are roughly of similar complexity

What Happens Between Meetings

1. Each processor splits its time between
   - (a) working on its Long List of boxes. For each box, perform a sequence of tests: interval evaluation rejection test; Taylor model evaluation: LDB, QFB bounders, Gradient-based box rejection with Gradient Taylor models
   - (b) performing non-rigorous global search (currently via genetic algorithm) in its assigned search space of global boxes, as well as neighboring global boxes
2. If Long List of boxes is exhausted, retrieve a box from the processor's section on the Short List
3. If processor's section on Short List is exhausted, continue to perform non-rigorous global search as in 1b.
4. After appropriate time, join next meeting.

The TM based Global Optimizer, COSY-GO

has utilized various algorothms based on Taylor models.

- LDB (Linear Dominated Bounding) bounding and domain reduction
- QFB (Quadratic Fast Bounding) bounding and domain reduction for positive definate cases (Quadratic pruning)
- Various cutoff value update schemes

And, we have completed

- Adjustment to pallarel environments with low inter-processoer communication rate
- Restart capability
- Continuation of computations while the underlying arithmetic fails
- COSY INFINITY Version 9.0 has been released

And, what we are doing further...

- High-order derivative based box rejection and the domain reduction
- Supporting high multiple precision computations for TMs

Outer and Inner Enclosures using Taylor Models

Example: A crescent mapping (A. Goldsztejn and L. Jaulin, 2005)
The function f on $R = \{(x, y)|x^2 + y^2 \in [1, 2]\}$ given by

$$f(x, y) = \begin{cases} xy \\ x + y \end{cases}$$

Using Taylor models:

- Representing the domain R using Taylor models as

$$(x, y) = (r\cos(\phi), r\sin(\phi)) \text{ where } (r, \phi) \in [1, \sqrt{2}] \times [0, 2\pi] = D.$$

Outer and Inner Enclosures using Taylor Models

Example: A crescent mapping (A. Goldsztejn and L. Jaulin, 2005)
The function f on $R = \{(x,y) | x^2 + y^2 \in [1,2]\}$ given by

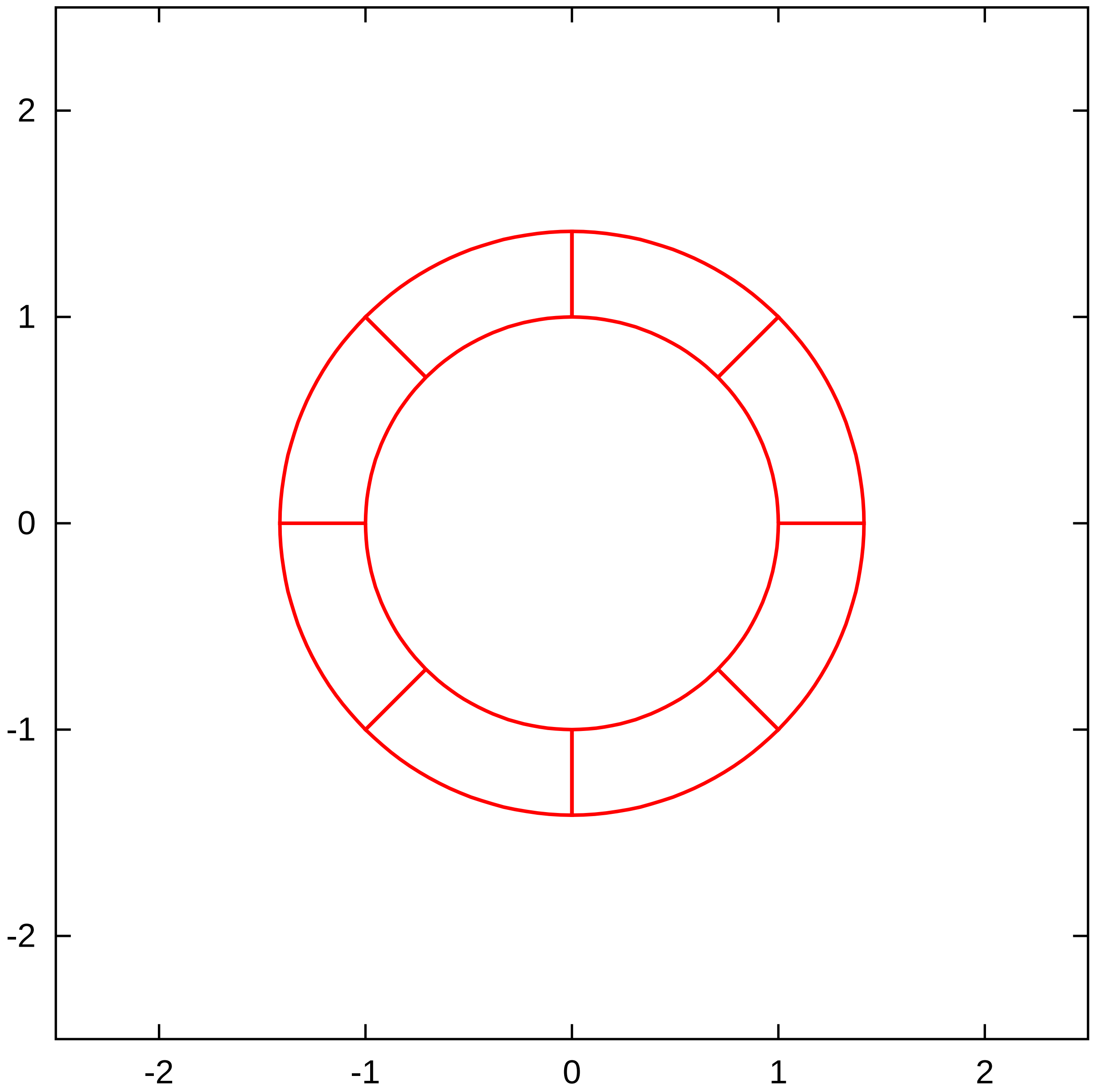
$$f(x,y) = \begin{cases} xy \\ x + y \end{cases}$$

Using Taylor models:

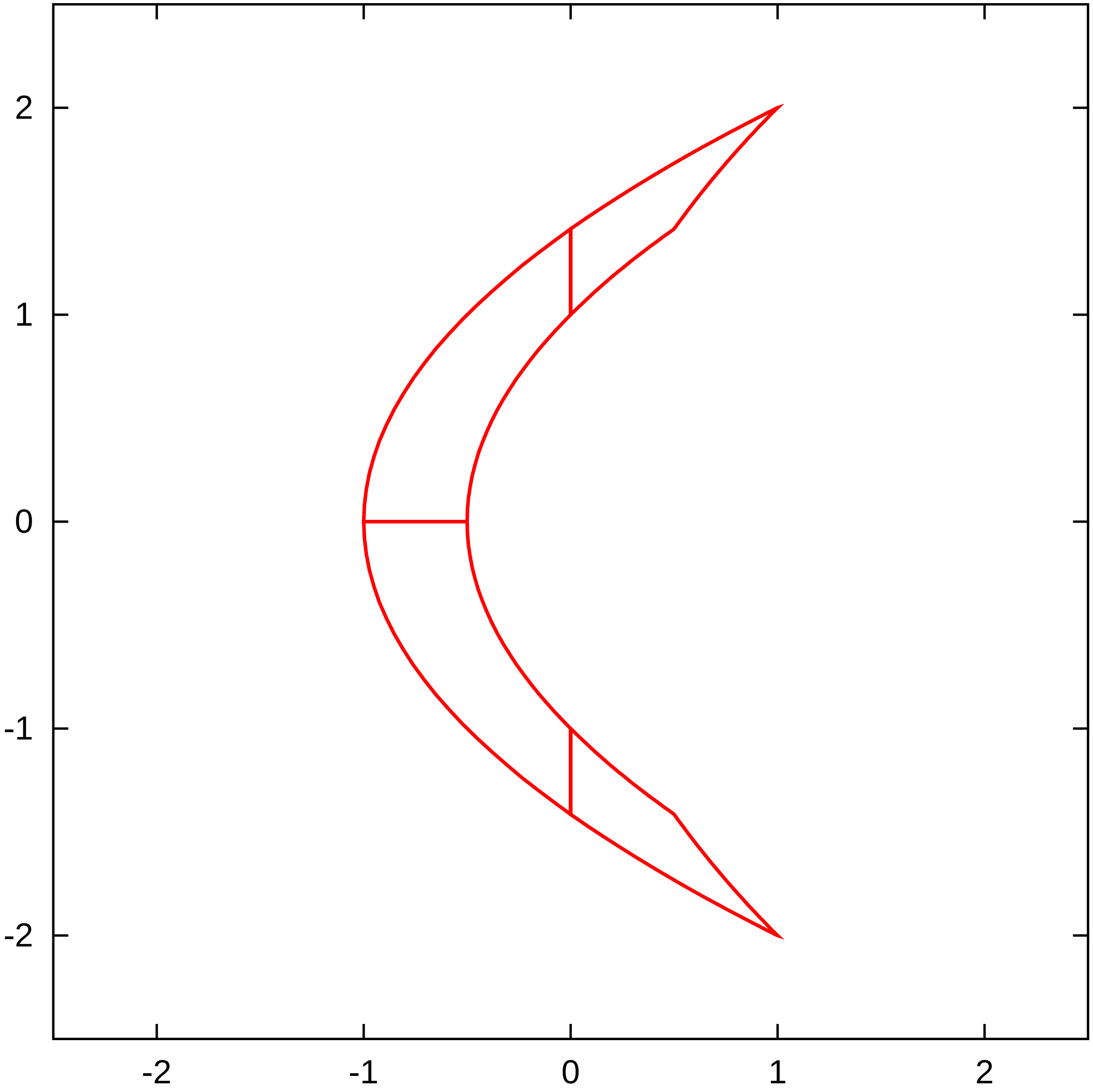
- Representing the domain R using Taylor models as

$$(x,y) = (r\cos(\phi), r\sin(\phi)) \text{ where } (r,\phi) \in [1,\sqrt{2}] \times [0,2\pi] = D.$$

- Split D in 8 subdomains to represent f by 8 Taylor models of order 5.
  → width$(I) < 0.03$; the outer and inner enclosures are indistinguishable.



Region 1=< r^2 =< 2 expressed by TMs (5th order) fx=x*y, fy=x+y, mapped by TMs in 1=< r^2 =< 2 (5th order)



Outer and Inner Enclosures using Taylor Models

Example: A crescent mapping (A. Goldsztejn and L. Jaulin, 2005)
The function f on $R = \{(x,y) | x^2 + y^2 \in [1,2]\}$ given by

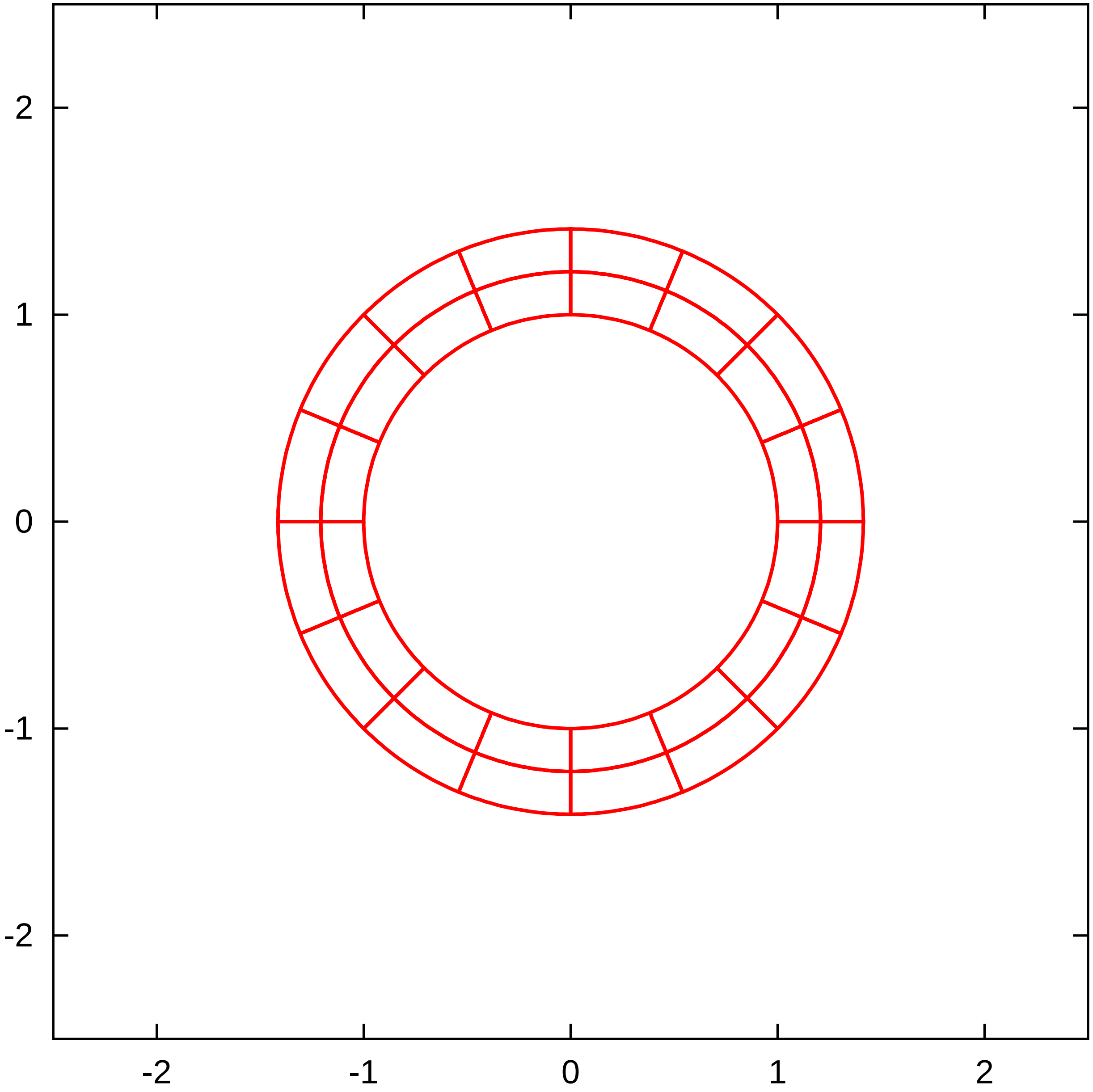
$$f(x,y) = \begin{cases} xy \\ x+y \end{cases}$$

Using Taylor models:

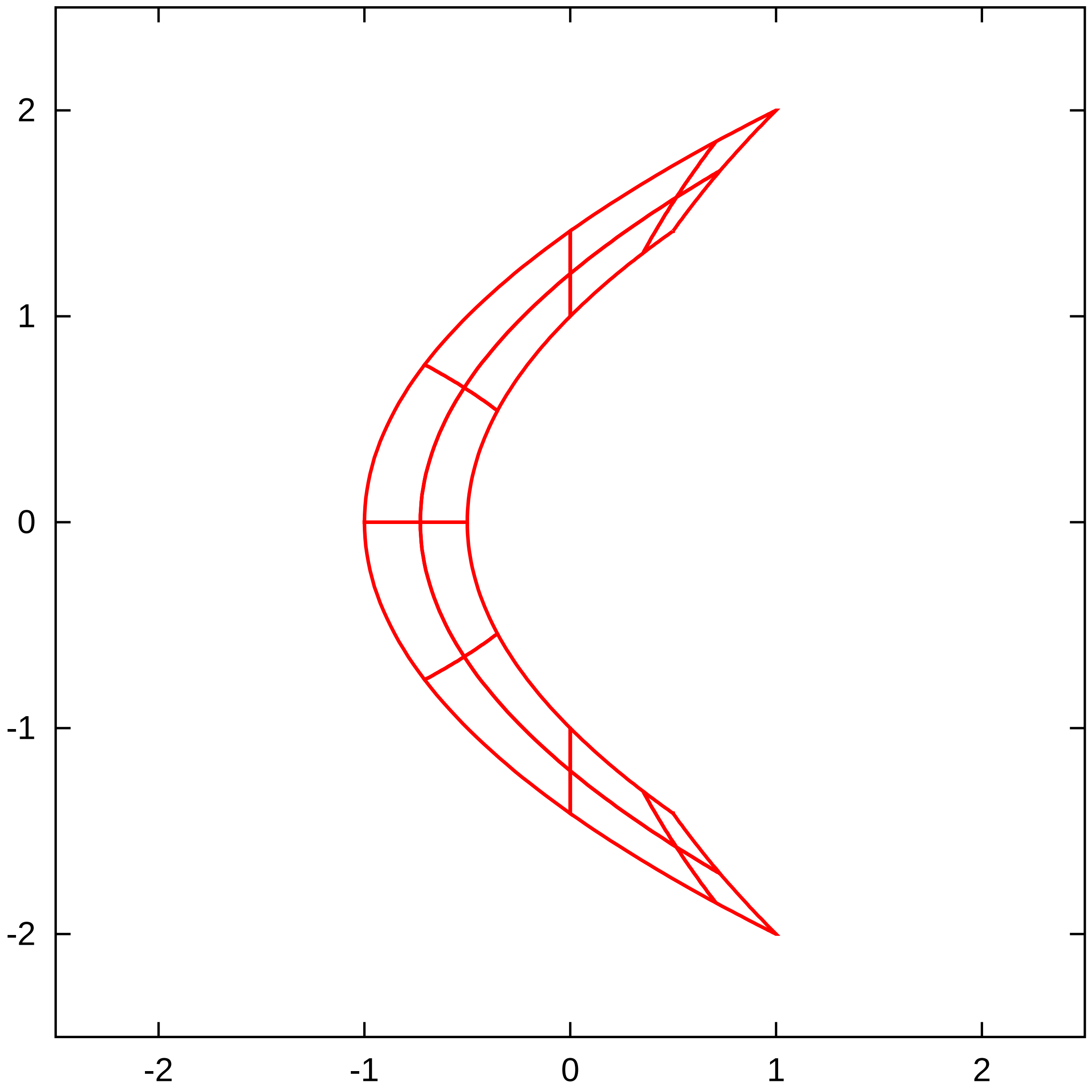
- Representing the domain R using Taylor models as

$$(x,y) = (r\cos(\phi), r\sin(\phi)) \text{ where } (r,\phi) \in [1,\sqrt{2}] \times [0,2\pi] = D.$$

- Split D in 8 subdomains to represent f by 8 Taylor models of order 5.
  → width(I) < 0.03; the outer and inner enclosures are indistinguishable.
- ... 16 × 2 subdomains, ... by 32 TMs of order 5.



Region 1=< r^2 =< 2 expressed by TMs (5th order) fx=x*y, fy=x+y, mapped by TMs in 1=< r^2 =< 2 (5th order)



Outer and Inner Enclosures using Taylor Models

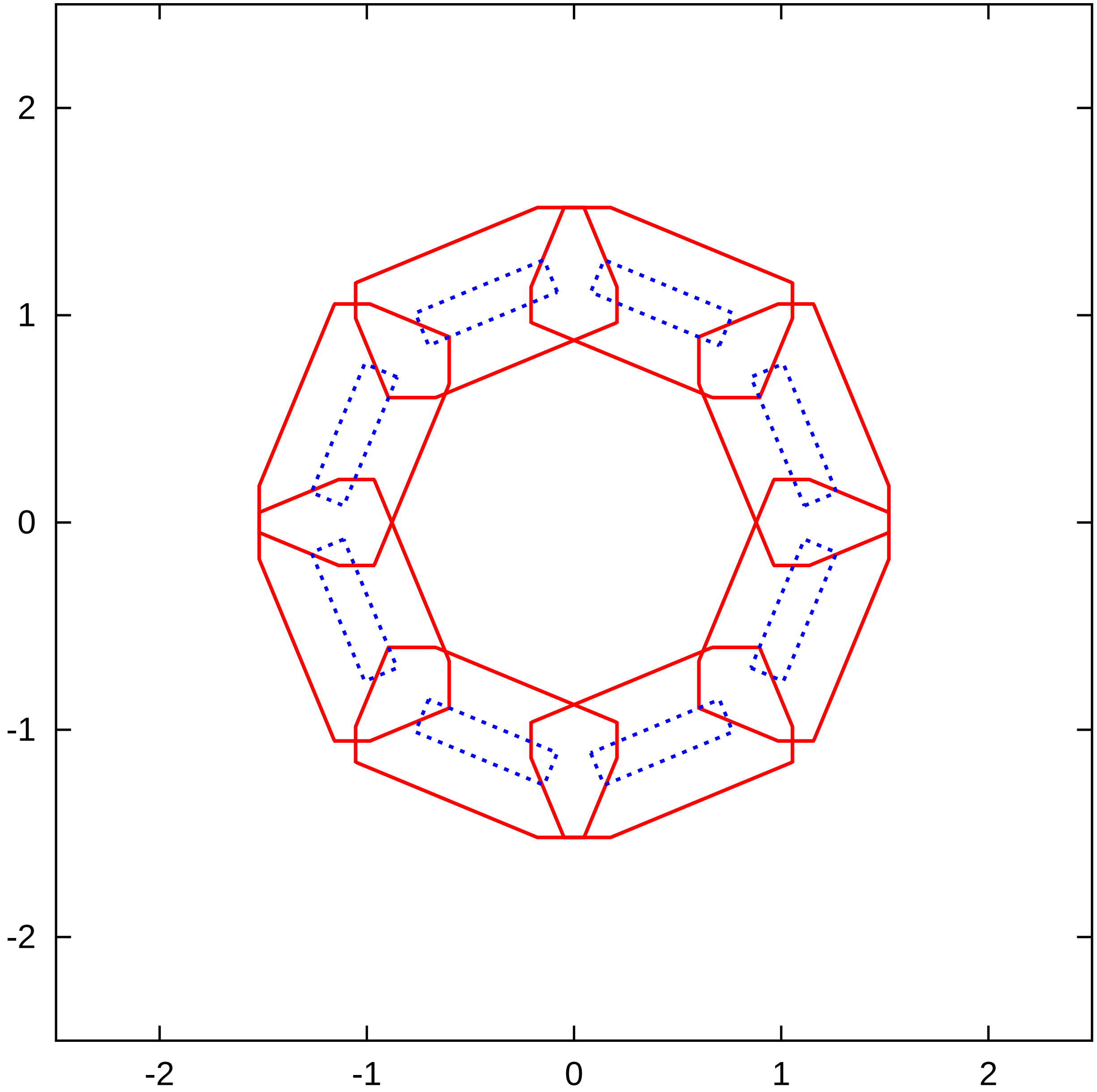
Example: A crescent mapping (A. Goldsztejn and L. Jaulin, 2005)
The function f on $R = \{(x,y) | x^2 + y^2 \in [1,2]\}$ given by

$$f(x,y) = \begin{cases} xy \\ x+y \end{cases}$$

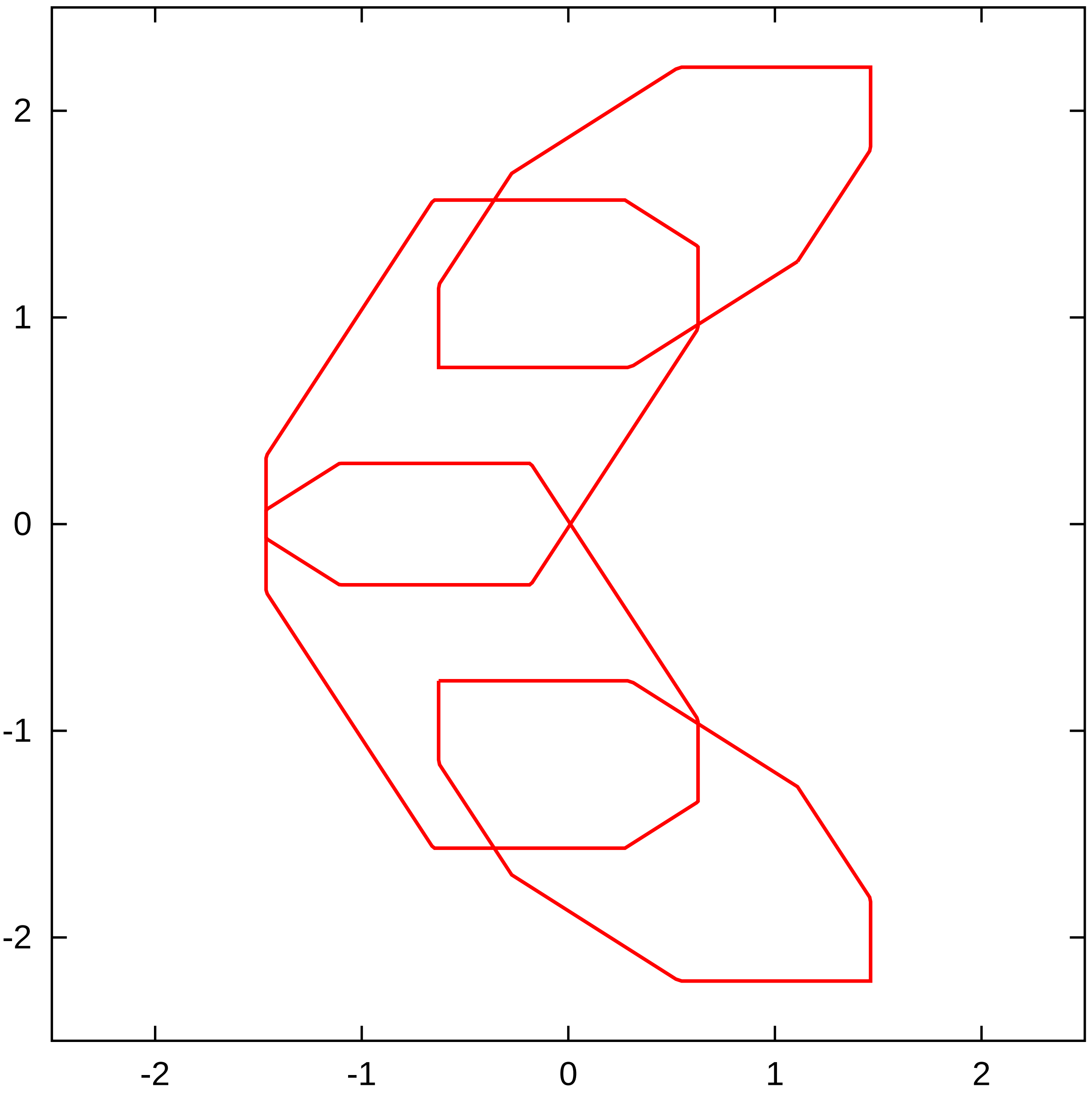
Using Taylor models:

- Representing the domain R using Taylor models as
$$(x,y) = (r\cos(\phi), r\sin(\phi)) \text{ where } (r,\phi) \in [1,\sqrt{2}] \times [0,2\pi] = D.$$
- Split D in 8 subdomains to represent f by 8 Taylor models of order 5.
→ width$(I) < 0.03$; the outer and inner enclosures are indistinguishable.
- ... 16×2 subdomains, ... by 32 TMs of order 5.
- ... 8 subdomains, ... by 8 TMs of order 1.
→ width$(I) \simeq 0.9$; the inner representations of f are empty sets.

Region 1=< r^2 =< 2 expressed by TMs (1st order)



fx=x*y, fy=x+y, mapped by TMs in 1=< r^2 =< 2 (1st order)



Outer and Inner Enclosures using Taylor Models

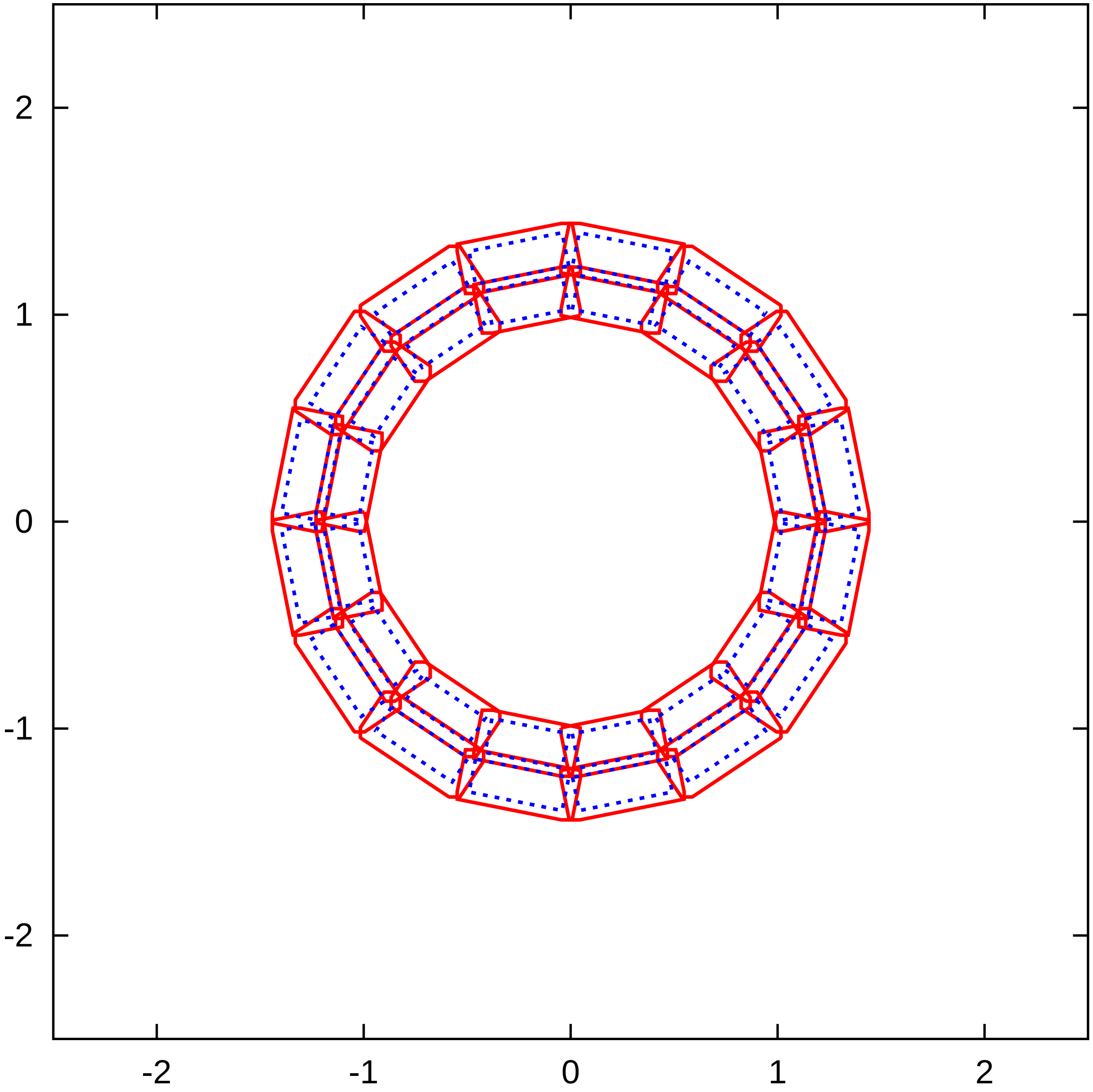
Example: A crescent mapping (A. Goldsztejn and L. Jaulin, 2005)
The function f on $R = \{(x, y) | x^2 + y^2 \in [1, 2]\}$ given by

$$f(x, y) = \begin{cases} xy \\ x + y \end{cases}$$

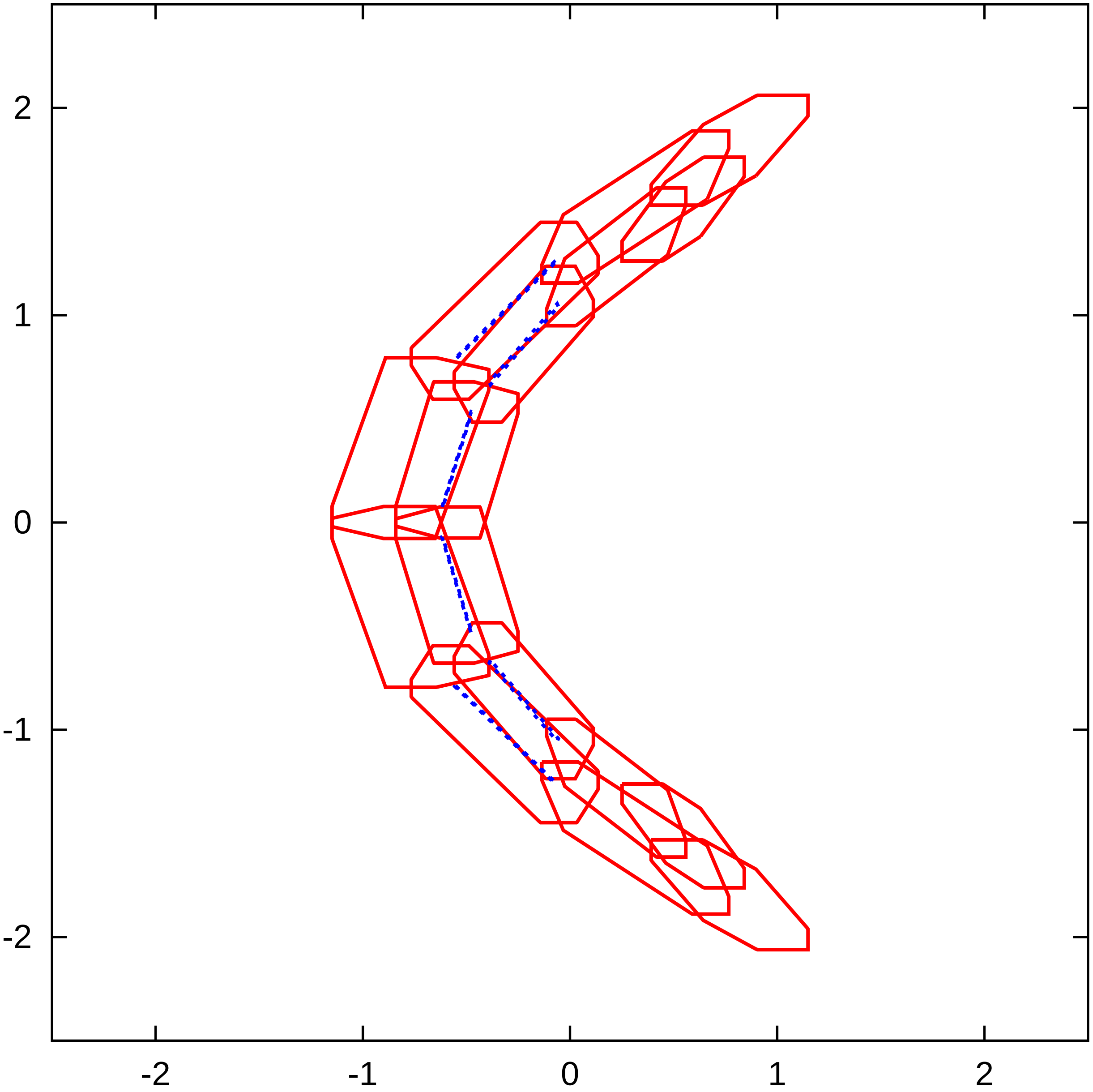
Using Taylor models:

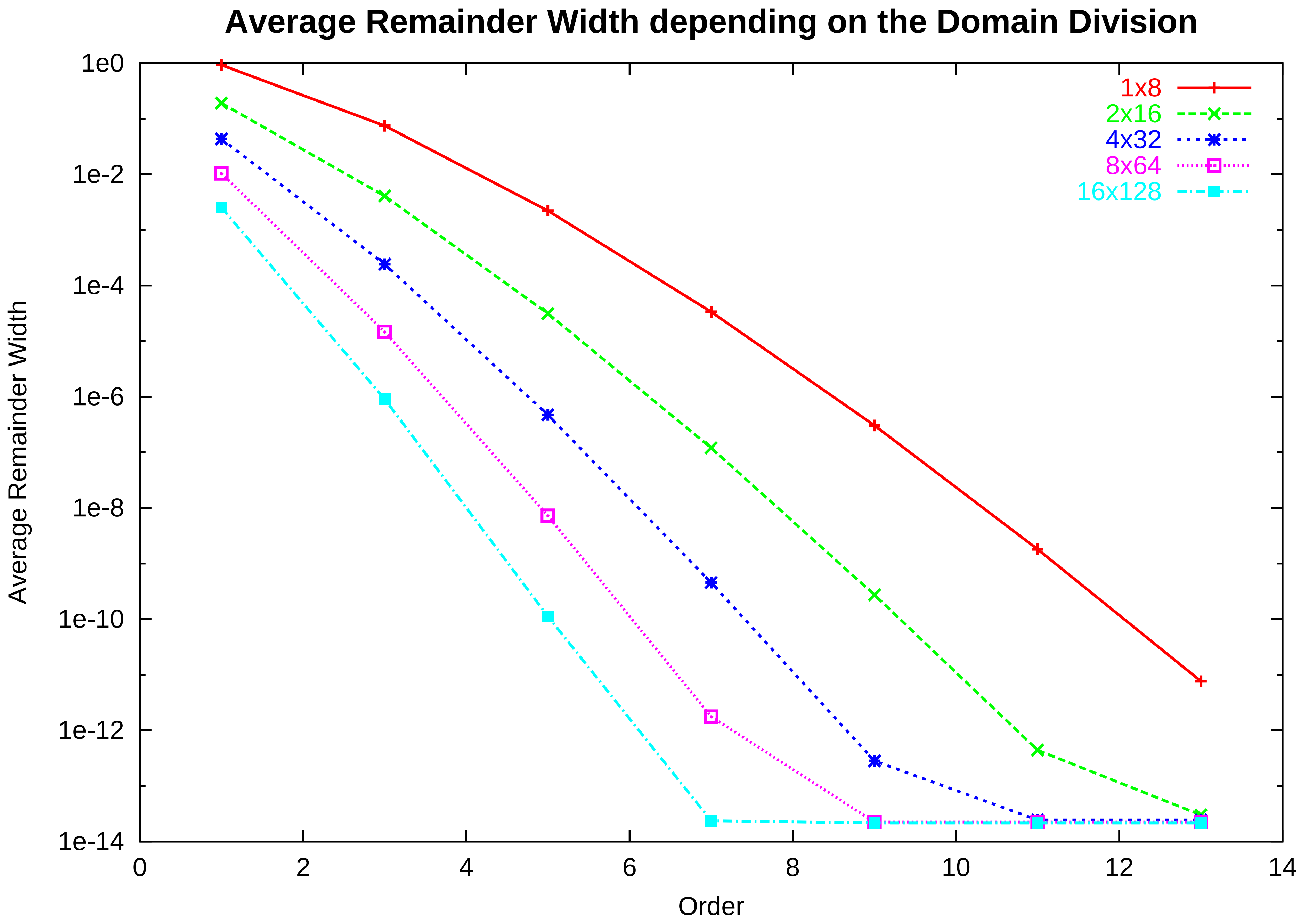
- Representing the domain R using Taylor models as
$$(x, y) = (r\cos(\phi), r\sin(\phi)) \text{ where } (r, \phi) \in [1, \sqrt{2}] \times [0, 2\pi] = D.$$
- Split D in 8 subdomains to represent f by 8 Taylor models of order 5.
$\rightarrow$ width(I) < 0.03; the outer and inner enclosures are indistinguishable.
- ... 16 × 2 subdomains, ... by 32 TMs of order 5.
- ... 8 subdomains, ... by 8 TMs of order 1.
$\rightarrow$ width(I) $\simeq 0.9$; the inner representations of f are empty sets.
- ... 16 × 2 subdomains, ... by 32 TMs of order 1.

Region 1=< r^2 =< 2 expressed by TMs (1st order)



fx=x*y, fy=x+y, mapped by TMs in 1=< r^2 =< 2 (1st order)




Average Remainder Width depending on the Domain Division
Average Remainder Width
1e0
1e-2
1e-4
1e-6
1e-8
1e-10
1e-12
1e-14
Order
0
2
4
6
8
10
12
14
1x8
2x16
4x32
8x64
16x128

High-Order Verification of Exteriority of Points

Goal: Decide whether a point c is outside the range of $f(D)$;

$$c \notin f(D).$$

Scheme:

Have a Taylor model representation $f(x) \in P(x) + I$ for all $x \in D$.
Let $C = P(0)$ and $P^{\geq 1}(x) = P(x) - C$.
Then we want to assert that there is no $x \in D$ with $c - C \in P^{\geq 1}(x) + I$.
Determine the order n inverse $P^{\geq 1*}$ to $P^{\geq 1}$, and apply it to both sides:

$$P^{\geq 1*}(c - C) \in P^{\geq 1*}(P^{\geq 1}(x) + I) \subset \mathcal{I} + I^* \subset D + I^*.$$

So, if it can be shown that

$$P^{\geq 1*}(c - C) \notin D + I^*,$$

then $c \notin f(D)$.

High-Order Verification of Interiority of Points

Goal: Decide whether a point c is inside the range of $f(D)$;

$$c \in f(D).$$

Scheme:

Very similar to the previous scheme for an exteriority test.
If it can be shown that

$$P^{\geq 1*}(c - C) \in D + I^*,$$

then $c \in f(D)$. In this case, however, $P^{\geq 1*}$ has to be shown to be injective (for example by factoring out linear part, and showing that resulting identity plus higher orders has strictly positive slopes).

High-Order Verification of Interiority of Points

Goal: Decide whether a point c is inside the range of $f(D)$;

$$c \in f(D).$$

Scheme 2:

Have a Taylor model representation $f(x) \in P(x)+I$ for all $x \in D$, and let $r(x) = f(x)-P(x)$, where $r(x) \in I$. Let $C = P(0)$ and $P^{\geq 1}(x) = P(x)-C$.

We attempt to assure the existence of a point $s \in D : f(s) = c$, which is equivalent to

$$c - C = P^{\geq 1}(s) + r(s).$$

Determine the order n inverse $P^{\geq 1 *}$ to $P^{\geq 1}$. Let

$$s_0 = P^{\geq 1 *}(c - C),$$

an approximation for a solution s, and it is likely $s_0 \in D$. If not, move s_0 towards the center of D.

Now we introduce a change of variables $\bar{s} = s - s_0$.

Let $\bar{L}$ denote the linear part of $P^{\geq 1}(s_0 + \bar{s})$ in $\bar{s}$, and let $\bar{L}^*$ be an approximate non-singular inverse of $\bar{L}$. Then the problem is equivalent to

$$\begin{aligned} c - C &= P^{\geq 1}(s_0 + \bar{s}) + r(s_0 + \bar{s}), \text{ or} \\ \mathcal{I}(\bar{s}) &= \bar{L}^* \left[c - C - \left(P^{\geq 1}(s_0 + \bar{s}) + r(s_0 + \bar{s}) \right) \right] + \mathcal{I}(\bar{s}). \end{aligned}$$

$$\mathcal{I}(\bar{s}) = \bar{L}^* \left[c - C - (P^{\geq 1}(s_0 + \bar{s}) + r(s_0 + \bar{s}))\right] + \mathcal{I}(\bar{s}).$$

Compute the RHS by TM arithmetic ($r(s_0 + \bar{s})$ by $0 + I$). We obtain a fixed point problem in $\bar{s}$:

$$\mathcal{I}(\bar{s}) = R(\bar{s}) + I_f.$$

Observe that the zeroth order and the linear parts of $R(\bar{s})$ have very small coefficients. By bounding them by an interval I_R,

$$\mathcal{I}(\bar{s}) = R^{\geq 2}(\bar{s}) + I_R + I_f.$$

Now we attempt to find an interval $\bar{S}$:

$$\bar{S} \supset A(\bar{S}) = R^{\geq 2}(\bar{S}) + I_R + I_f.$$

If such $\bar{S}$ is found, $\bar{S}$ contains a fixed point (via Brower FP theorem). Then we have found a solution of the original problem as long as $s_0 + \bar{S} \subseteq D$.

$$c \in f(s_0 + \bar{S}), \text{ thus } c \in f(D).$$

Note: For a small interval $\bar{S}$, $R^{\geq 2}(\bar{S})$ is a much smaller interval. Choosing $\bar{S}$ as a small multiple of the interval $(I_f + I_R)$ will likely lead to an immediate self enclosure, which will scale with order $(n + 1)$ of the domain size.

Properties:

1. Given a TM $P+I$ describing the function f over D, the problem of deciding interiority is transformed to a FP problem. The occurring intervals are comparable in size to the small remainder I. Since only quadratic and higher order terms appear in the FP problem, the method is very likely to succeed as long as I is sufficiently small.
2. The resulting enclosure of the FP scales with the width of I, and hence with order $(n+1)$ of the domain D.
3. The setup of this FP problem requires only limited inexpensive additional TM arithmetic beyond the computation of the TM representing the original function f. In particular, using an intrinsic tool for Horner shifts, no TM multiplications are necessary.
4. Unlike conventional interval-Newton methods, the approach does not require the inversion of any interval matrices, but only floating point approximations to inverses of floating point matrices.

Example: Exteriority and Interiority of Points

Example: A crescent mapping (A. Goldsztejn and L. Jaulin, 2005)
The function f on $R = \{(x,y)|x^2+y^2 \in [1,2]\}$ given by

$$f(x,y) = \begin{cases} xy \\ x+y \end{cases}$$

Using the discussed schemes:

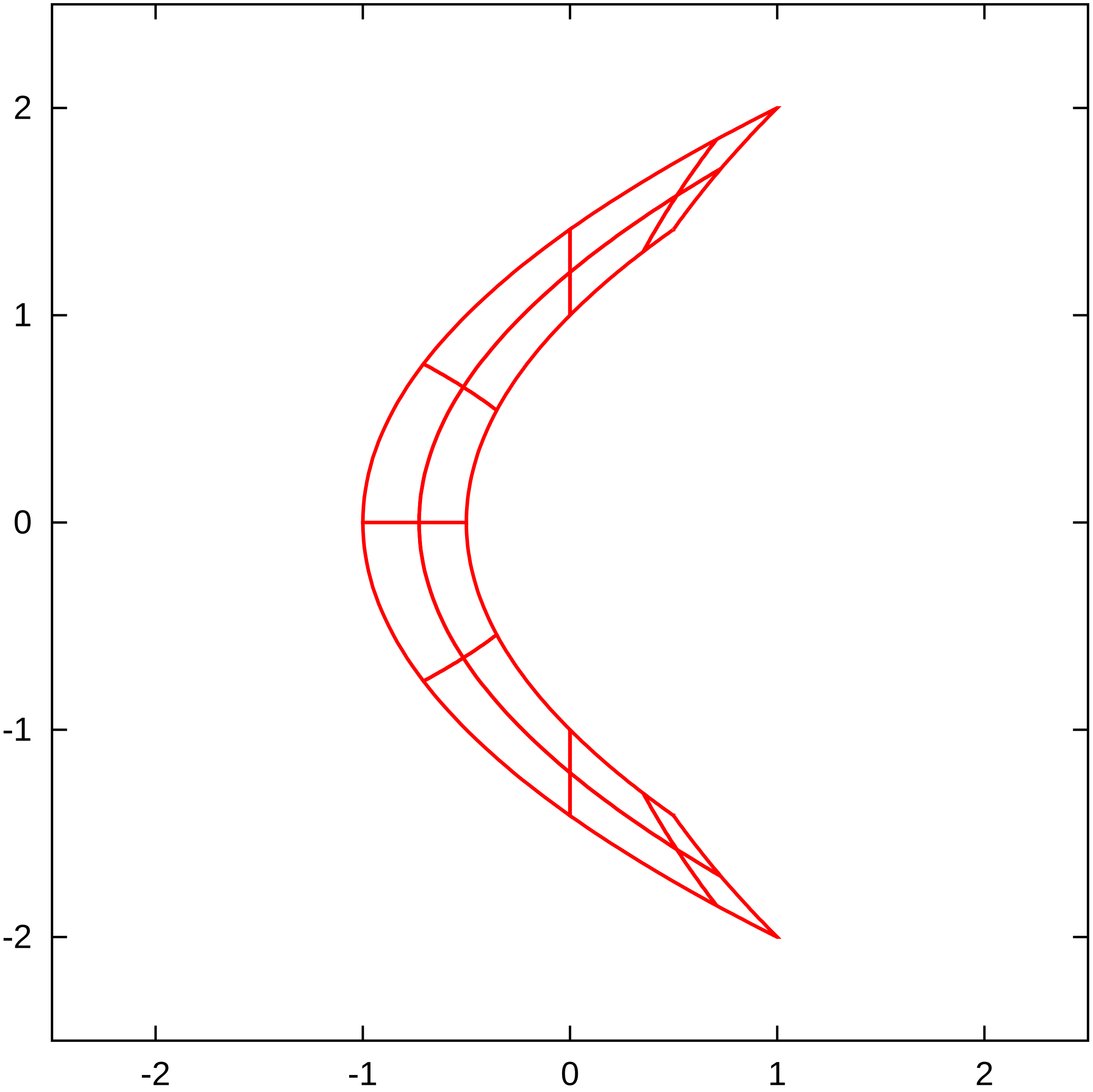
- Study the distance of points verified barely outside and inside the range of the function $f(D)$.

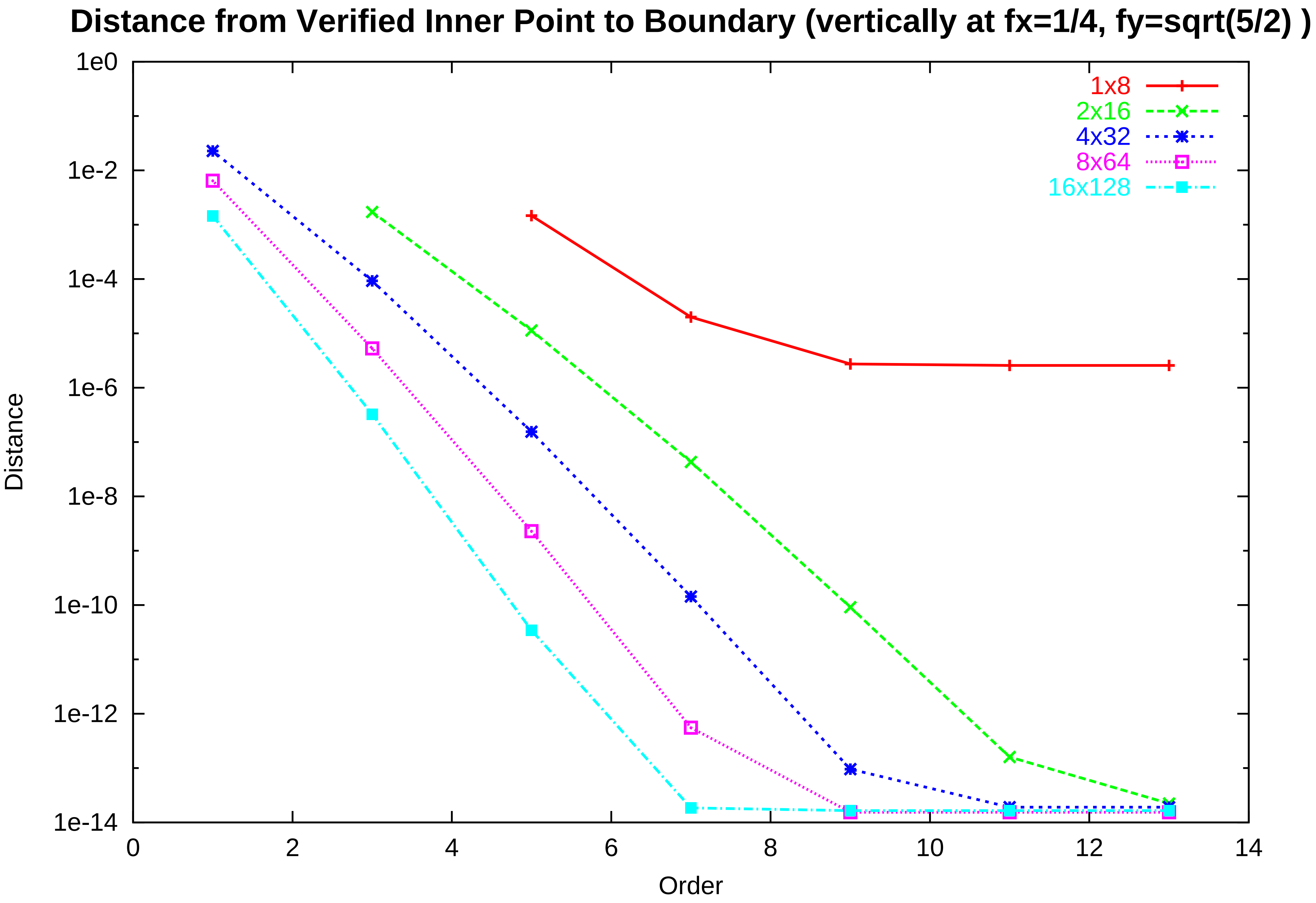
  Measure the distance along the vertical line through the point

  $$(x,y) = \left(\frac{1}{4}, \sqrt{\frac{5}{2}}\right),$$

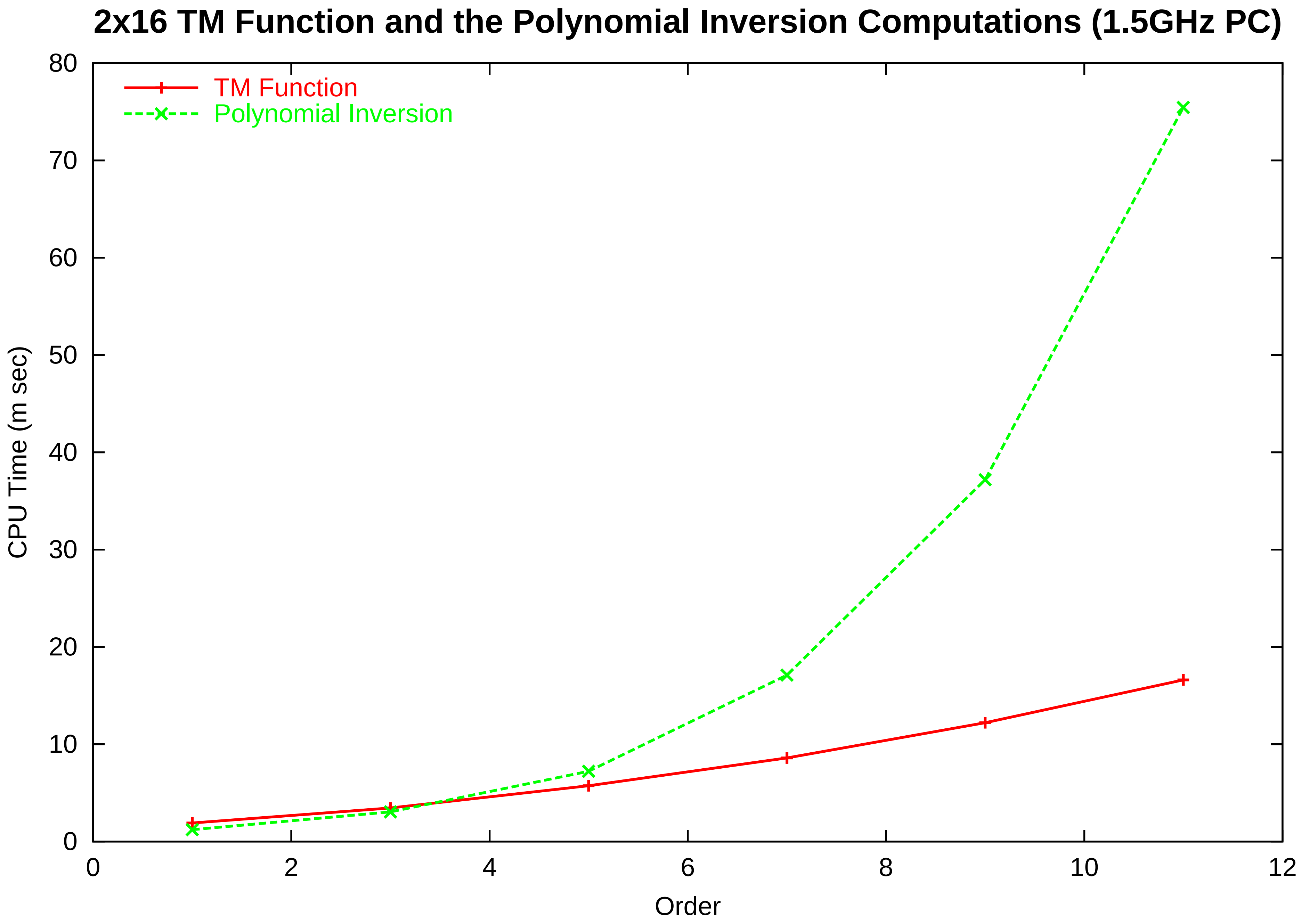
  which is known to lie on the boundary.

fx=x*y, fy=x+y, mapped by TMs in 1=< r^2 =< 2 (5th order)





Distance from Verified Inner Point to Boundary (vertically at fx=1/4, fy=sqrt(5/2))


2x16 TM Function and the Polynomial Inversion Computations (1.5GHz PC)
TM Function
Polynomial Inversion
CPU Time (m sec)
0
10
20
30
40
50
60
70
80
Order
0
2
4
6
8
10
12

Gradient Taylor Models - Motivation

Conventional Taylor Model for f on D: Pair of Taylor polynomial P of f and remainder I such that

$$f(x) \in P(x) + I \text{ for all } x \in D$$

Many verified tools benefit from enclosures for derivatives of functions; for example **Global Optimization**.

Frequently used: **Automatic Differentiation** with interval arguments yields interval enclosure of derivatives

Similarly, can use Automatic Differentiation **with Taylor Model** arguments to get Taylor model enclosures of derivatives.

But: the Taylor polynomial of $\partial f/\partial x_i$ is just $\partial P/\partial x_i$ (to one order less)

Thus: Knowing P from TM of f, we already know the bulk of the information for the TMs of the gradient (to one order less)!

Idea: Save time by calculating only P, and **not** the Taylor polynomials of the derivative.

The Gradient Taylor Model

Definition: Gradient Taylor model

For a function $f(x)$ on D, we call the vector

$$(P, I, I_1, \ldots, I_v)$$

a gradient Taylor model of f, if $\forall x \in D$,

$$\begin{aligned} f(x) &\in P(x) + I \text{ and} \\ \frac{\partial f(x)}{\partial x_i} &\in \frac{\partial P(x)}{\partial x_i} + I_i \text{ for all } i = 1, ..., v \end{aligned}$$

Thus a gradient Taylor model for the function f consists of its Taylor polynomial, a remainder bound, as well as remainder bounds for its partial derivatives based on their Taylor polynomials of one order less.

Addition of Gradient Taylor Models

Assume we know gradient Taylor models for f and g, and thus on D, we have

$$(f, \nabla f) \in \left(P_f + I_f, \frac{\partial P_f}{\partial x_{f1}} + I_{f1}, \ldots, \frac{\partial P_f}{\partial x_{fv}} + I_{fv} \right)$$
$$(g, \nabla g) \in \left(P_g + I_g, \frac{\partial P_g}{\partial x_{g1}} + I_{g1}, \ldots, \frac{\partial P_g}{\partial x_{gv}} + I_{gv} \right).$$

Then we apparently also have

$$(f+g, \nabla f + \nabla g) \in \left(P_f + P_g + I_f + I_g, \frac{\partial P_f}{\partial x_{f1}} + \frac{\partial P_g}{\partial x_{g1}} + I_{f1} + I_{g1}, \right.$$
$$\left. ..., \frac{\partial P_f}{\partial x_{fv}} + \frac{\partial P_g}{\partial x_{gv}} + I_{fv} + I_{gv} \right).$$

and thus $(P_f + P_g, I_f + I_g, I_{f1} + I_{g1}, ..., I_{fv} + I_{gv})$ is a gradient Taylor model for $f + g$.

Multiplication of Gradient Taylor Models

We compute the regular Taylor model part $P_{(f \cdot g)}$ and $I_{(f \cdot g)}$ as we do it for regular Taylor models. For the remainder bounds, there are several ways based on the product rule

$$\frac{\partial}{\partial x_i}(f \cdot g) = \left(\frac{\partial}{\partial x_i} f\right) \cdot g + f \cdot \left(\frac{\partial}{\partial x_i} g\right). \tag{1}$$

Method 1: Use

$$\frac{\partial f}{\partial x_i} \in \frac{\partial P_f}{\partial x_i} + I_{fi}$$

and perform the computation of (1) in regular Taylor model arithmetic, then get $I_{(f \cdot g)i}$ as the remainder of the arithmetic for (1).

This is simple, but wasteful as it always re-computes the polynomial parts of the derivatives, which is unnecessary as they are merely derivatives of $P_{f \cdot g}$.

Multiplication of Gradient TM - Remainders

Studying the details of Taylor Model multiplication, we see: To obtain the remainder bound of the product, all we need are **bounds for each order** $I_{f,i}^{(j)}, I_{g,i}^{(j)}$ of the factors. Indeed, derivative remainder bounds $I_{(f\cdot g),i}$ are then given by

$$I_{(f\cdot g),i} = \sum_{j=0}^{n} I_{f,i}^{(j)} \cdot I_g^{(\geq n-j+1)} + I_f^{(j)} \cdot I_{g,i}^{(\geq n-j+1)}$$

Method 2a: Use the polynomial derivation operation to calculate each of the gradient polynomials

$$\frac{\partial P_f}{\partial x_i}, \frac{\partial P_g}{\partial x_i}$$

and use the already existing tool for order bounding.

Relatively fast (polynomial derivation is essentially coefficient re-shuffling) and **optimally precise**, given available information.

Multiplication of Gradient TM - Remainders

Method 2b: Observe that

$$I_f^{(j+1)} = \sum_{i_1+\cdots+i_v=j+1} |a_{i_1,\ldots,i_i,\ldots,i_v}| \text{ and we also have}$$

$$I_{f,i}^{(j)} = \sum_{i_1+\cdots+i_v=j+1} |i_i \cdot a_{i_1,\ldots,i_i,\ldots,i_v}| = \sum_{i_1+\cdots+i_v=j+1} i_i \cdot |a_{i_1,\ldots,i_i,\ldots,i_v}|$$

So, the order bounds of the derivative polynomials satisfy

$$I_{f,i}^{(j)} \subset I_f^{(j+1)} \cdot (j+1) \tag{2}$$

Extremely fast (only re-use existing order bounds), but **less precise**.

Intrinsic Functions of Gradient Taylor Models

Problem: given gradient TM of f, want a gradient TM for $int(f)$, where $int \in \{\sin, \cos, \exp, \log\text{,sqrt,...}\}$.

Example: sin. Note

$$\frac{d}{dx_i}(\sin(f)) = \cos(f) \cdot \frac{d}{dx_i} f$$

1. First factor: perform one additional intrinsic evaluation, compute order bounds.
2. First factor: bulk of effort lies in computing powers of f, which are already available from $\sin(f)$ evaluation!
3. Second factor: Obtain order bounds directly, as above.
4. Determine derivative remainder bound as above from $\sum_{j=0}^{n} I_{f,i}^{(j)} \cdot I_{\cos(f)}^{(\geq n-j+1)}$

Generalizations of Gradient Taylor Models

1. **Extremely Cheap:** do not store $I_1, \ldots, I_v$, but only **a single** I' such that
$$I_1 \subset I', \ldots, I_v \subset I'$$
Particularly useful if combined with method 2b based on $I_{f,i}^{(j)} \subset I_f^{(j+1)} \cdot (j+1)$, since all $I_{f,i}^{(j)}$ will be the same already.

2. **Higher Orders:** Apparently the approach readily generalizes to Hessian Taylor models, or higher derivatives yet.

3. **Cheap Higher Orders:** Can be done economically, i.e. without large numbers of remainder bounds, by using the "extremely cheap" storage way.

Gradient Taylor Models - Summary

- We have shown how to **simultaneously** obtain Taylor models for the **function** and its entire **gradient.**
- There are **various ways** of obtaining remainders, differing in sharpness, speed, and storage
- Even the least sharp way yields good results, since TM remainder bounds are usually so small.
- In all cases, the extra computational **effort is minor**, usually less than 10% beyond the normal cost of a Taylor model.
- Can be generalized to **higher orders.**

The TM based Global Optimizer, COSY-GO

has utilized various algorothms based on Taylor models.

- LDB (Linear Dominated Bounding) bounding and domain reduction
- QFB (Quadratic Fast Bounding) bounding and domain reduction for positive definate cases (Quadratic pruning)
- Various cutoff value update schemes

And, we have completed

- Adjustment to pallarel environments with low inter-processoer communication rate
- Restart capability
- Continuation of computations while the underlying arithmetic fails
- COSY INFINITY Version 9.0 has been released

And, what we are doing further...

- High-order derivative based box rejection and the domain reduction
- Supporting high multiple precision computations for TMs