Rigorous Global Optimization for Parameter Estimates and Long-Term Stability Bounds

Kyoko Makino and Martin Berz

Department of Physics and Astronomy
Michigan State University
USA

Example

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

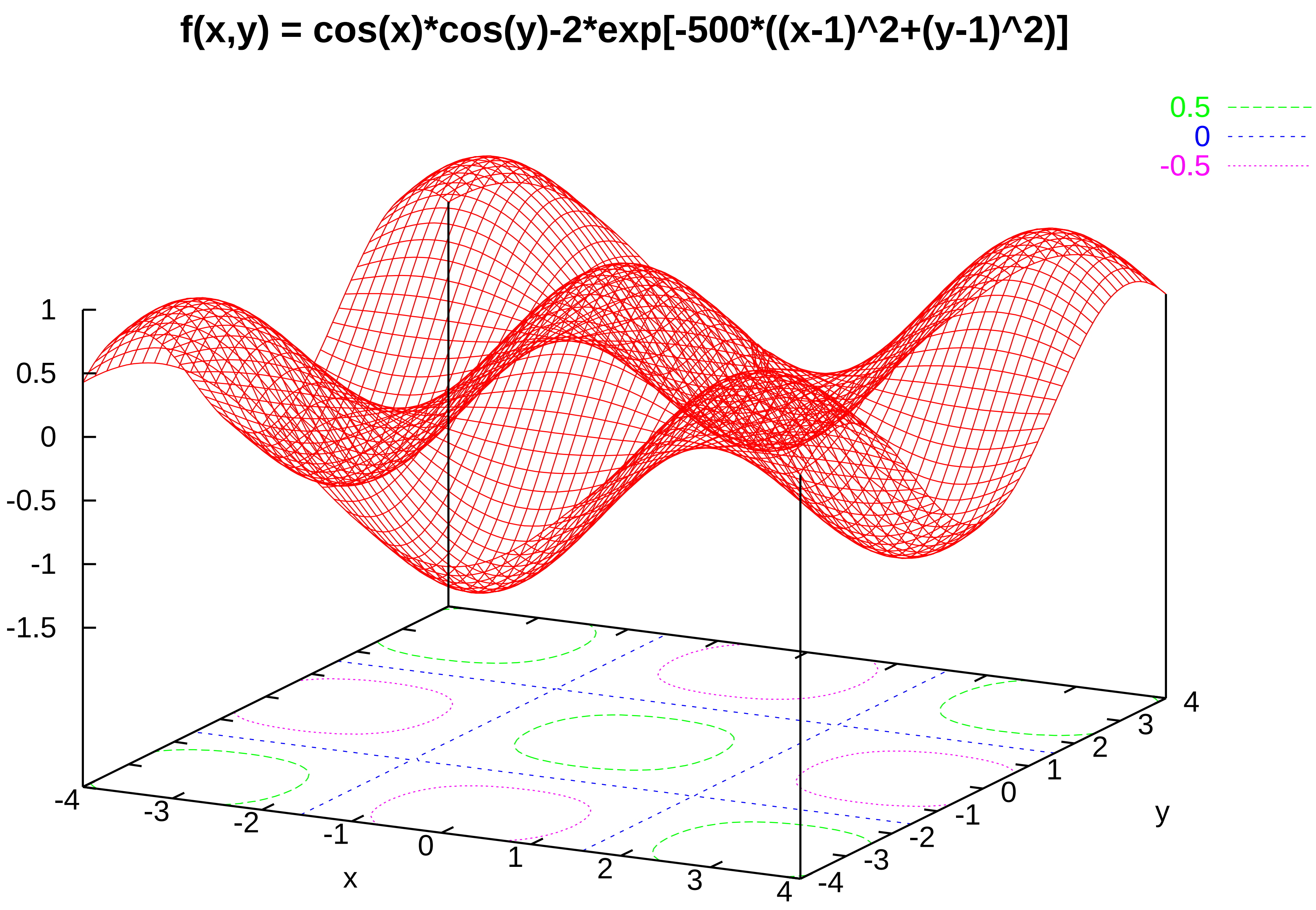
$$f(x, y) = \cos x \cos y - 2 \exp\left[-500 \cdot \left((x-1)^2 + (y-1)^2\right)\right]$$

Example

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

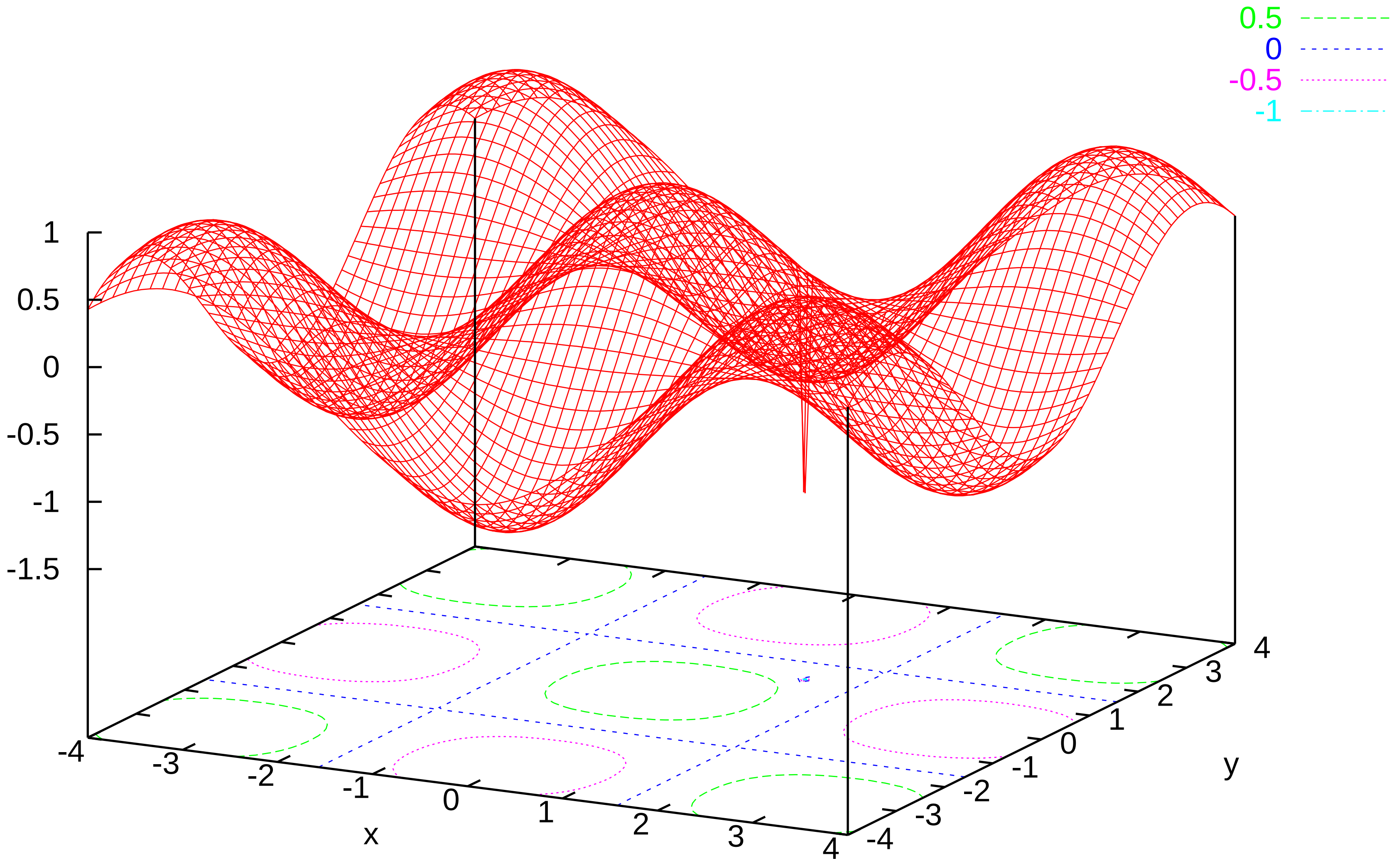
$$f(x, y) = \cos x \cos y - 2 \exp \left[-500 \cdot \left((x-1)^2 + (y-1)^2 \right) \right]$$

- A spike on top of the periodic cos behavior.



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

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



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.

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

Example

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

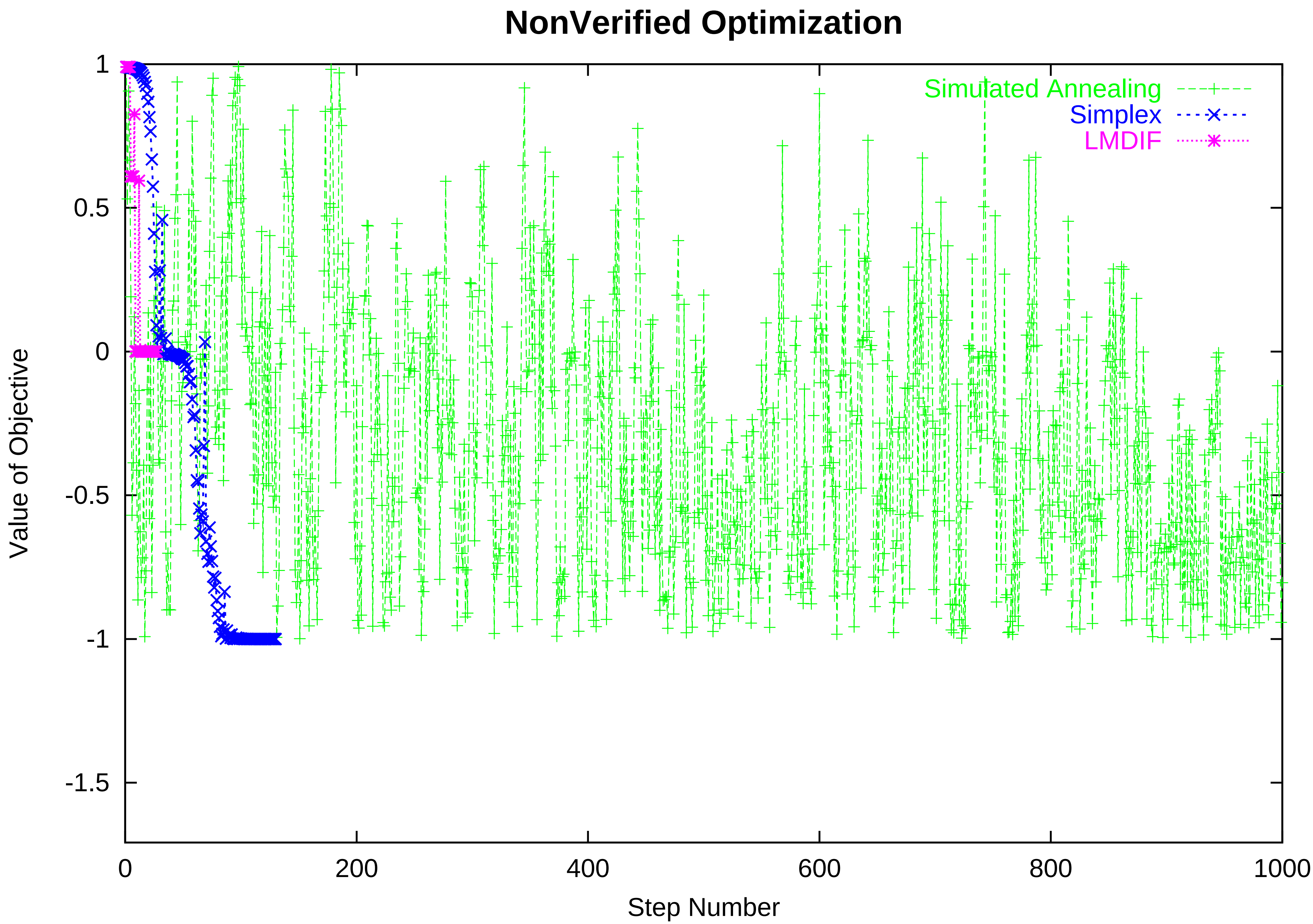
$$f(x,y) = \cos x \cos y - 2 \exp\left[-500 \cdot \left((x-1)^2 + (y-1)^2\right)\right]$$

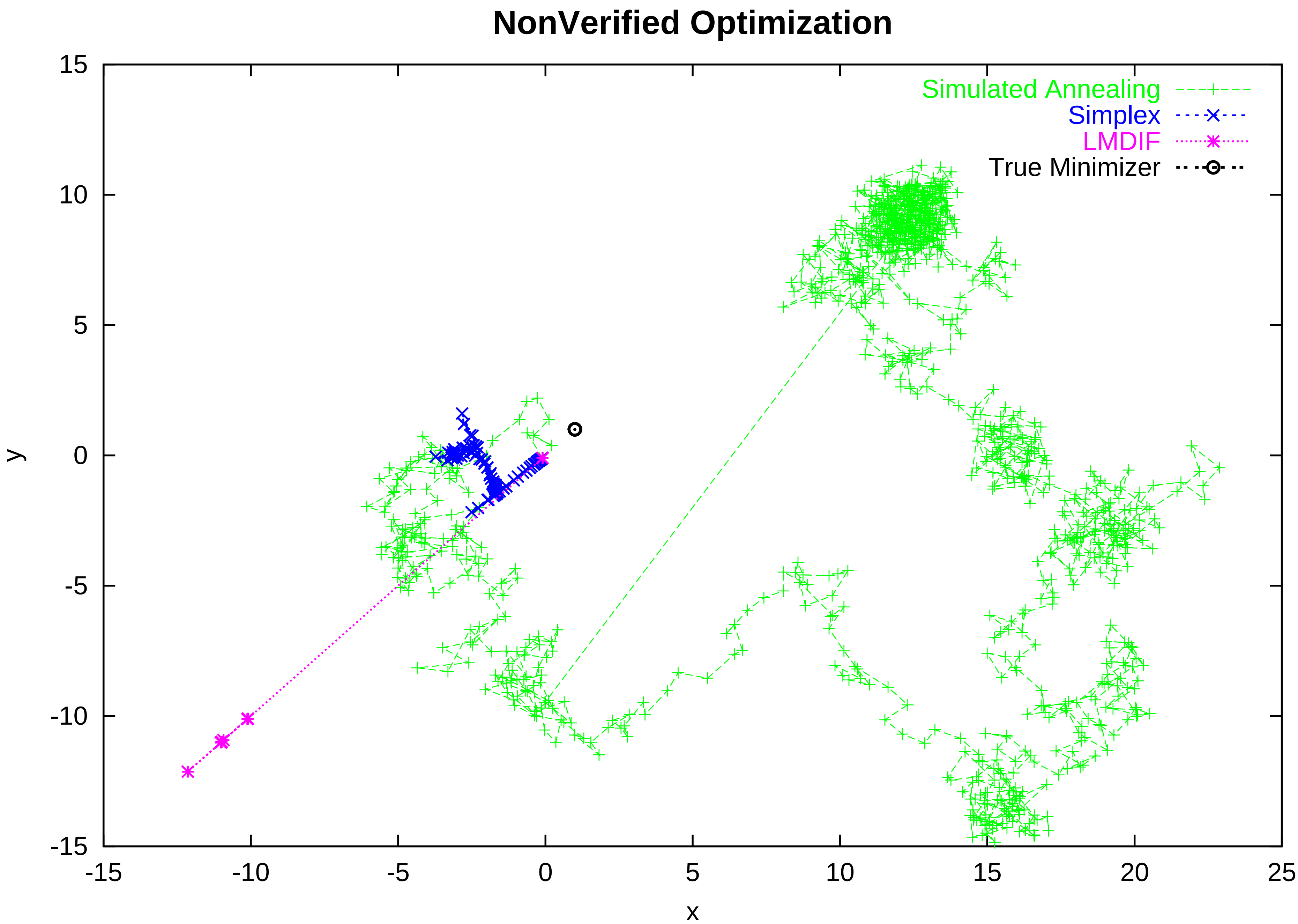
- A spike on top of the periodic cos behavior.

$$\min = -1.70817675216 07^{26}_{39} \quad \text{at } x = y = 1.0002272^{92}_{85}.$$

- Use various local optimization algorithms

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


NonVerified Optimization
Simulated Annealing
Simplex
LMDIF
Value of Objective
Step Number
1
0.5
0
-0.5
-1
-1.5
0
200
400
600
800
1000



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

Example

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

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

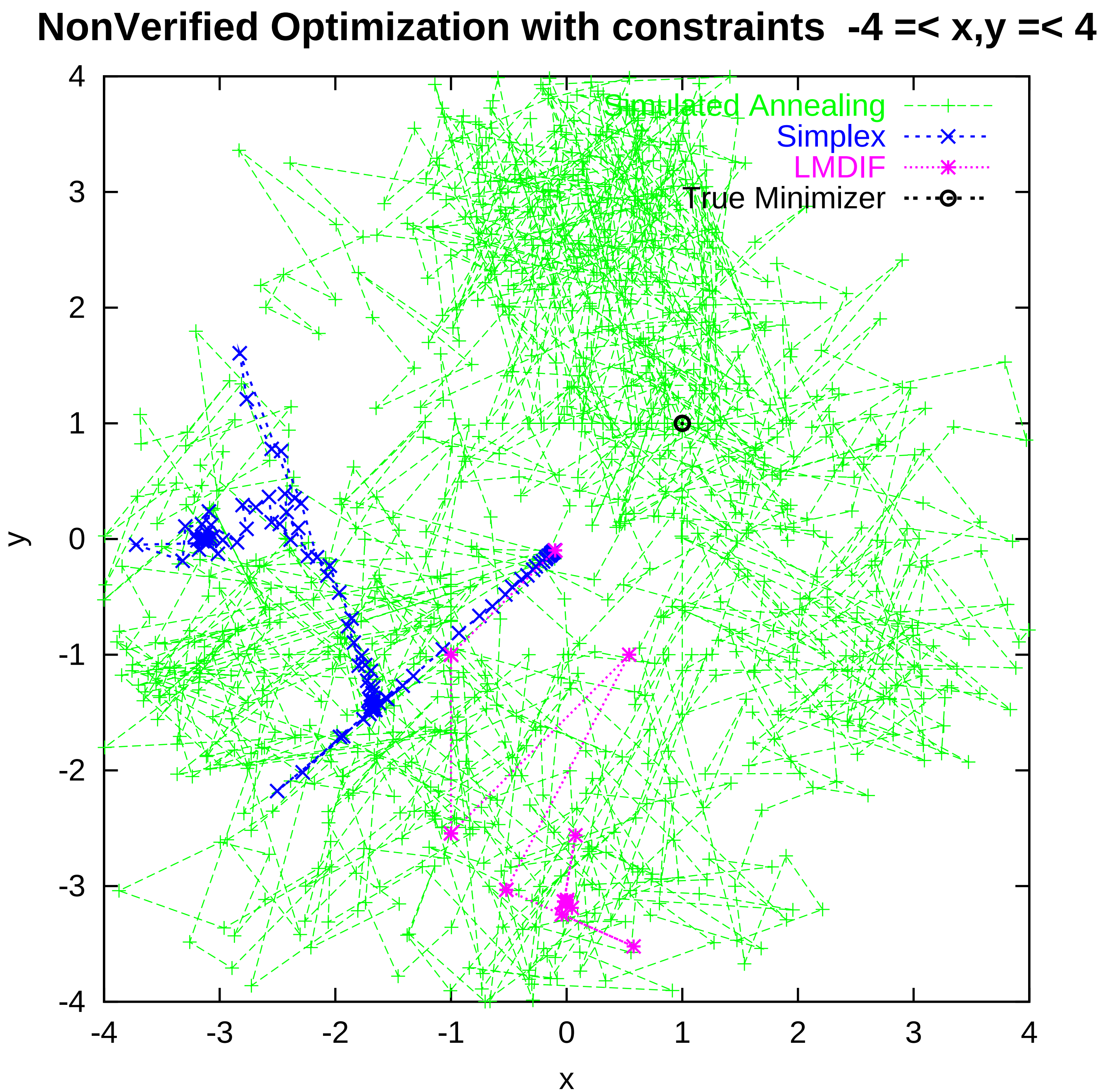
- A spike on top of the periodic cos behavior.

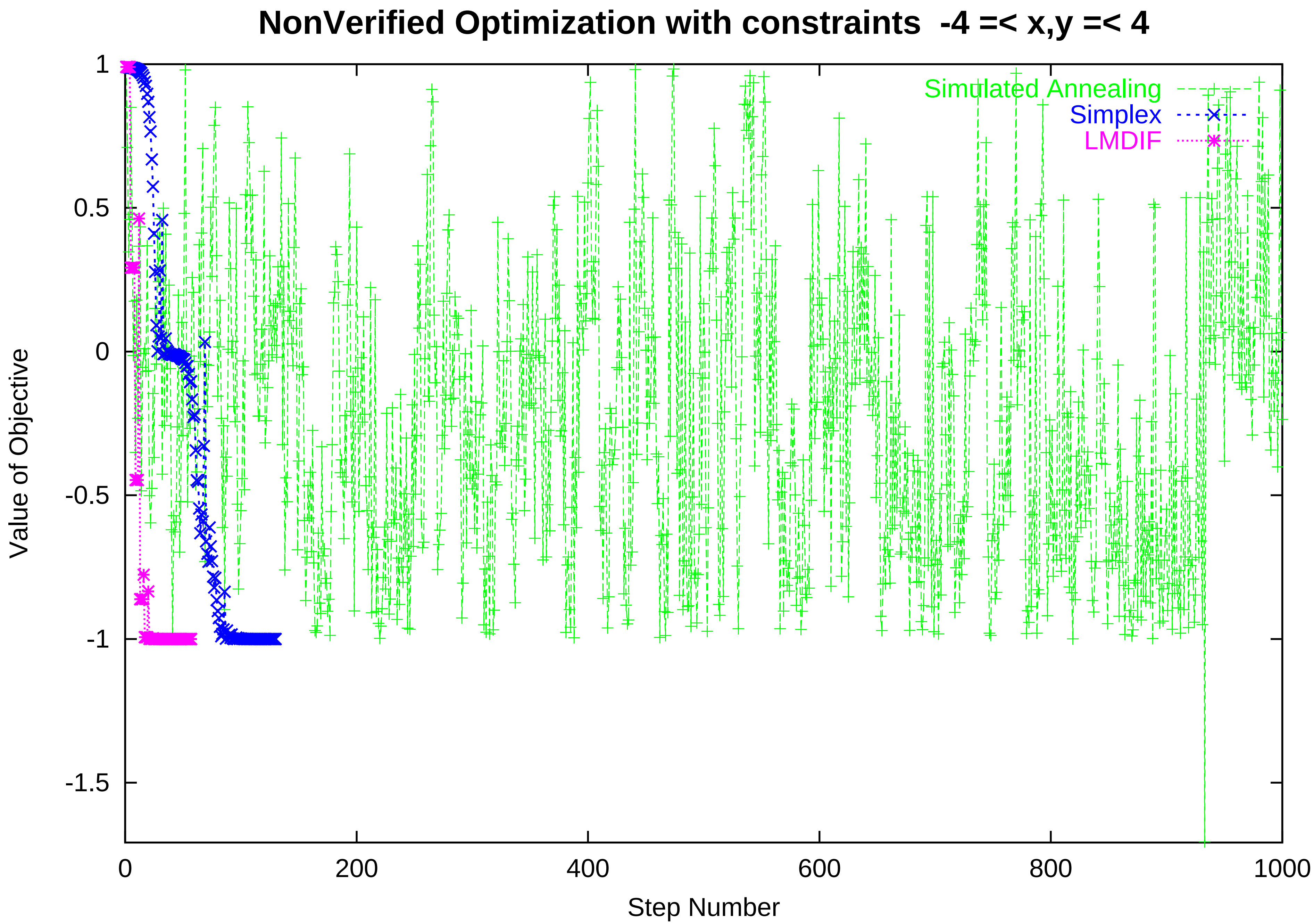
$$\min = -1.70817675216607^{26}_{39} \quad \text{at } x = y = 1.0002272^{92}_{85}.$$

- Use various local optimization algorithms

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

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


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



NonVerified Optimization with constraints -4 =< x,y =< 4
Value of Objective
Step Number
Simulated Annealing
Simplex
LMDIF
1
0.5
0
-0.5
-1
-1.5
0
200
400
600
800
1000

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 $\lvert x\rvert, \lvert y\rvert \leq 4$	
	Steps	min	Steps	min
Simulated Annealing	1000	~ -1	1000	10^{-2} accurate
Simplex	130	~ -1	130	~ -1
LMDIF	27	~ 0	57	~ -1

Example

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

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

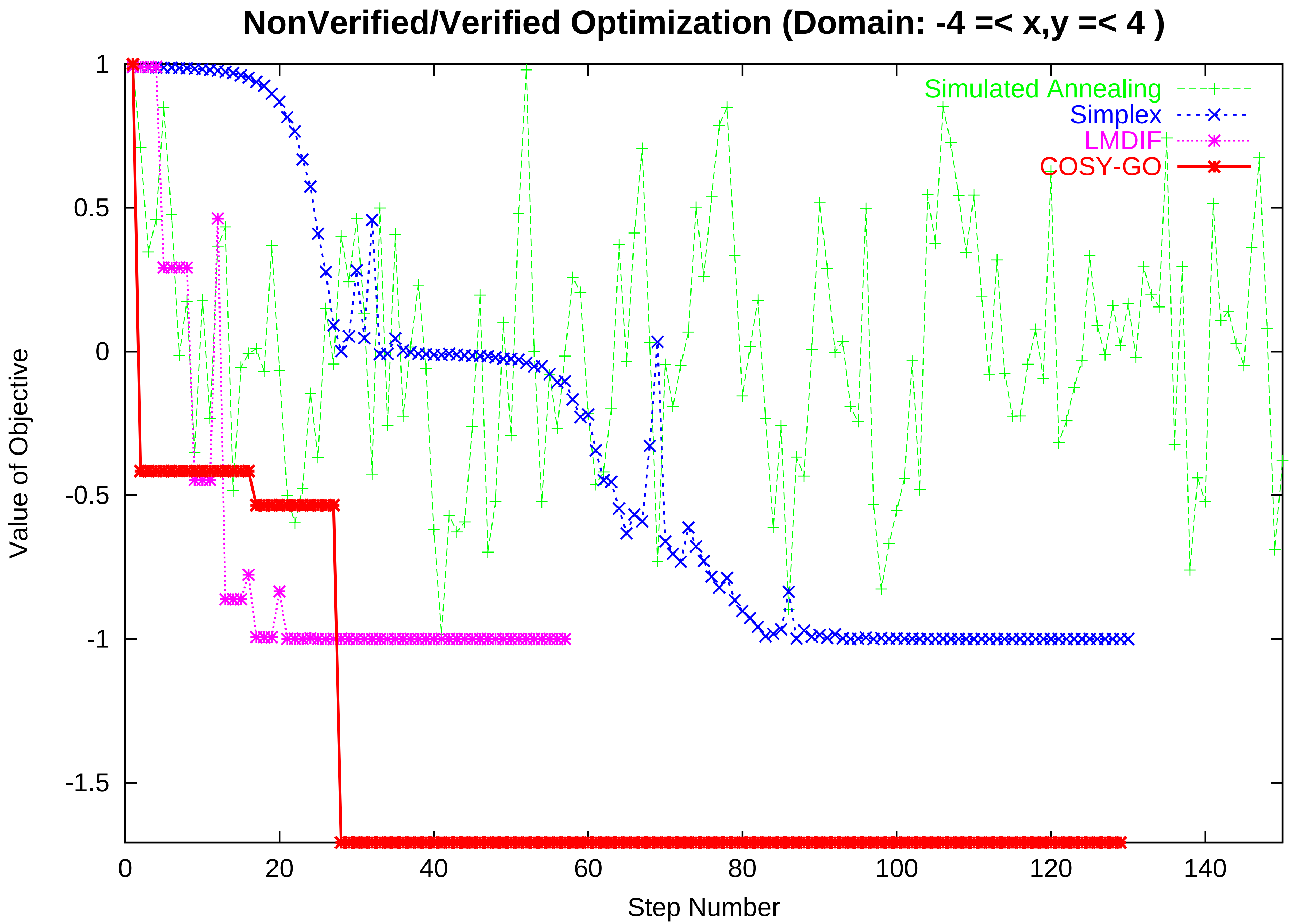
- A spike on top of the periodic cos behavior.

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

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

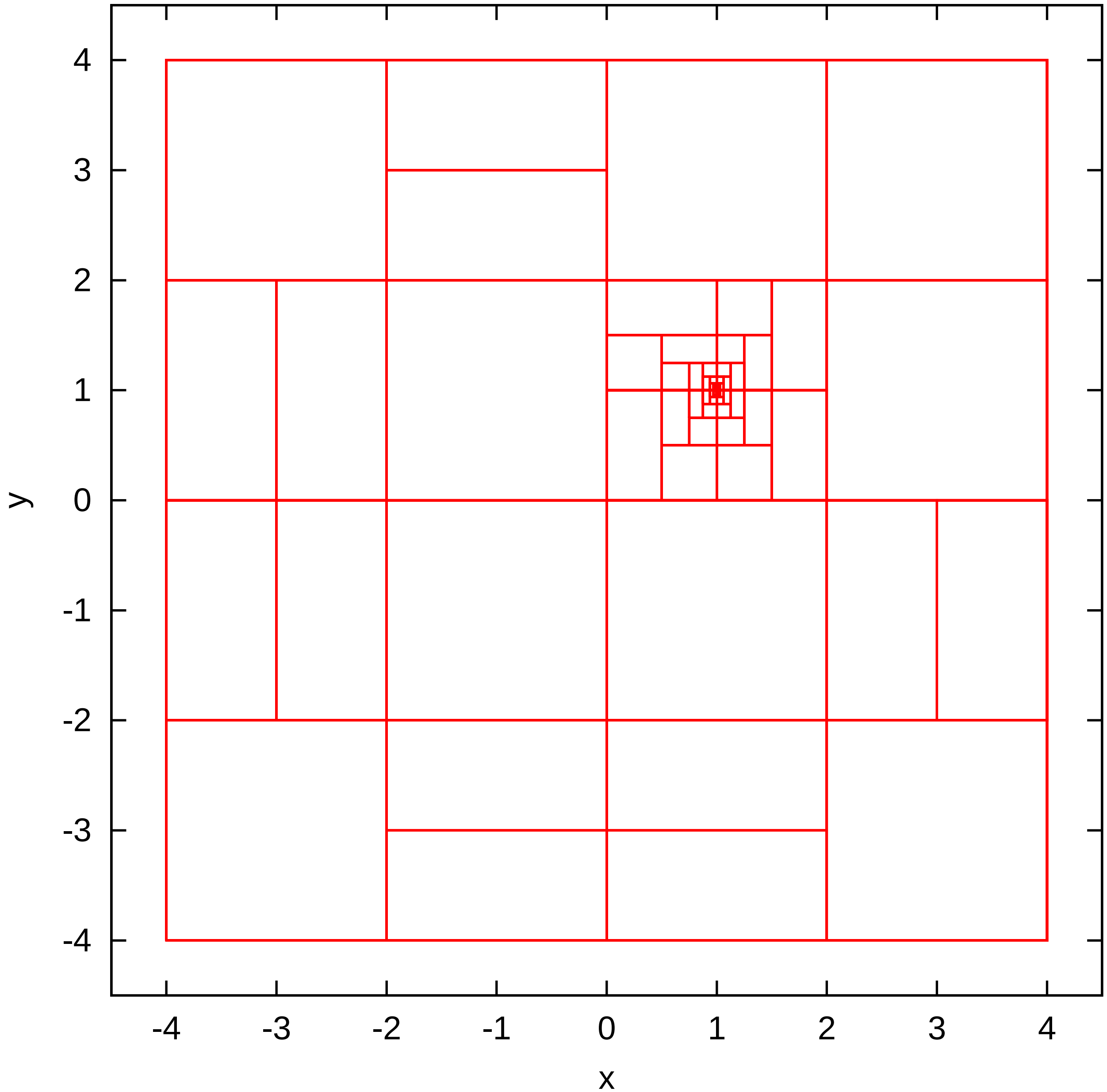
Algorithm	Without Constraint		Constraint $\|x\|, \|y\| \leq 4$	
	Steps	min	Steps	min
Simulated Annealing	1000	~ -1	1000	10^{-2} accurate
Simplex	130	~ -1	130	~ -1
LMDIF	27	~ 0	57	~ -1

- 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)



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 $\|x\|, \|y\| \leq 4$	
	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 $5 \cdot 10^{-17}$.

Design Parameter Optimization Example

Determine magnet parameters to eliminate several nonlinear aberrations.
Using local optimizers:

	Sim. Annealing	Simplex	LMDIF
S1	-4.220040774	-.1752976958	-.4129845844E-4
S2	-2.468406098	-.1270744624	-.1800819610E-3
S3	5.268669445	-.6394106627E-1	-.2079220929E-3
S4	-7.144285915	0.1167017471	-.9007439665E-4
S5	1.174139460	0.2231863960	0.2139893916
S6	-8.121172250	-.7800825893E-1	-.2800618129E-3
S7	9.135056510	-.2296525059	-.3195229128
S8	-2.243273322	0.9506890185E-1	0.1165687798
Obj	0.9348125285E-3	0.5486180417E-7	0.1003910906E-4
Steps	10000	691	10000

Definitions - Taylor Models and Operations

We begin with a review of the definitions of the basic operations.

Definition (Taylor Model) Let $f : D \subset R^v \to R$ be a function that is $(n+1)$ times continuously partially differentiable on an open set containing the domain v-dimensional domain D. Let x_0 be a point in D and P the n-th order Taylor polynomial of f around x_0. Let I be an interval such that

$$f(x) \in P(x - x_0) + I \text{ for all } x \in D.$$

Then we call the pair (P, I) an n-th order Taylor model of f around x_0 on D.

Definition (Addition and Multiplication) Let $T_{1,2} = (P_{1,2}, I_{1,2})$ be n-th order Taylor models around x_0 over the domain D. We define

$$T_1 + T_2 = (P_1 + P_2, I_1 + I_2)$$
$$T_1 \cdot T_2 = (P_{1\cdot 2}, I_{1\cdot 2})$$

where $P_{1\cdot 2}$ is the part of the polynomial $P_1 \cdot P_2$ up to order n and

$$I_{1\cdot 2} = B(P_e) + B(P_1) \cdot I_2 + B(P_2) \cdot I_1 + I_1 \cdot I_2$$

where P_e is the part of the polynomial $P_1 \cdot P_2$ of orders $(n + 1)$ to $2n$, and $B(P)$ denotes a bound of P on the domain D. We demand that $B(P)$ is at least as sharp as direct interval evaluation of $P(x - x_0)$ on D.

Definitions - Taylor Model Intrinsics

Definition (Intrinsic Functions of Taylor Models) Let $T = (P, I)$ be a Taylor model of order n over the v-dimensional domain $D = [a, b]$ around the point x_0. We define intrinsic functions for the Taylor models by performing various manipulations that will allow the computation of Taylor models for the intrinsics from those of the arguments. In the following, let $f(x) \in P(x - x_0) + I$ be any function in the Taylor model, and let $c_f = f(x_0)$, and $\bar{f}$ be defined by $\bar{f}(x) = f(x) - c_f$. Likewise we define $\bar{P}$ by $\bar{P}(x - x_0) = P(x - x_0) - c_f$, so that $(\bar{P}, I)$ is a Taylor model for $\bar{f}$. For the various intrinsics, we proceed as follows.

Exponential. We first write

$$\begin{aligned}\exp(f(x)) &= \exp\left(c_f + \bar{f}(x)\right) = \exp(c_f) \cdot \exp\left(\bar{f}(x)\right) \\ &= \exp(c_f) \cdot \Bigg\{ 1 + \bar{f}(x) + \frac{1}{2!}(\bar{f}(x))^2 + \cdots + \frac{1}{k!}(\bar{f}(x))^k \\ &\qquad + \frac{1}{(k+1)!}(\bar{f}(x))^{k+1} \exp\left(\theta \cdot \bar{f}(x)\right) \Bigg\},\end{aligned}$$

where $0 < \theta < 1$.

Definitions - Taylor Model Exponential, cont.

Taking $k \geq n$, the part

$$\exp(c_f) \cdot \left\{ 1 + \bar{f}(x) + \frac{1}{2!}(\bar{f}(x))^2 + \cdots + \frac{1}{n!}(\bar{f}(x))^n \right\}$$

is merely a polynomial of $\bar{f}$, of which we can obtain the Taylor model via Taylor model addition and multiplication. The remainder part of $\exp(f(x))$, the expression

$$\exp(c_f) \cdot \left\{ \frac{1}{(n+1)!}(\bar{f}(x))^{n+1} + \cdots + \frac{1}{(k+1)!}(\bar{f}(x))^{k+1} \exp\left(\theta \cdot \bar{f}(x)\right) \right\},$$

will be bounded by an interval. First observe that since the Taylor polynomial of $\bar{f}$ does not have a constant part, the $(n+1)$-st through $(k+1)$-st powers of the Taylor model $(\bar{P}, I)$ of $\bar{f}$ will have vanishing polynomial part, and thus so does the entire remainder part. The remainder bound interval for the Lagrange remainder term

The Henon Map

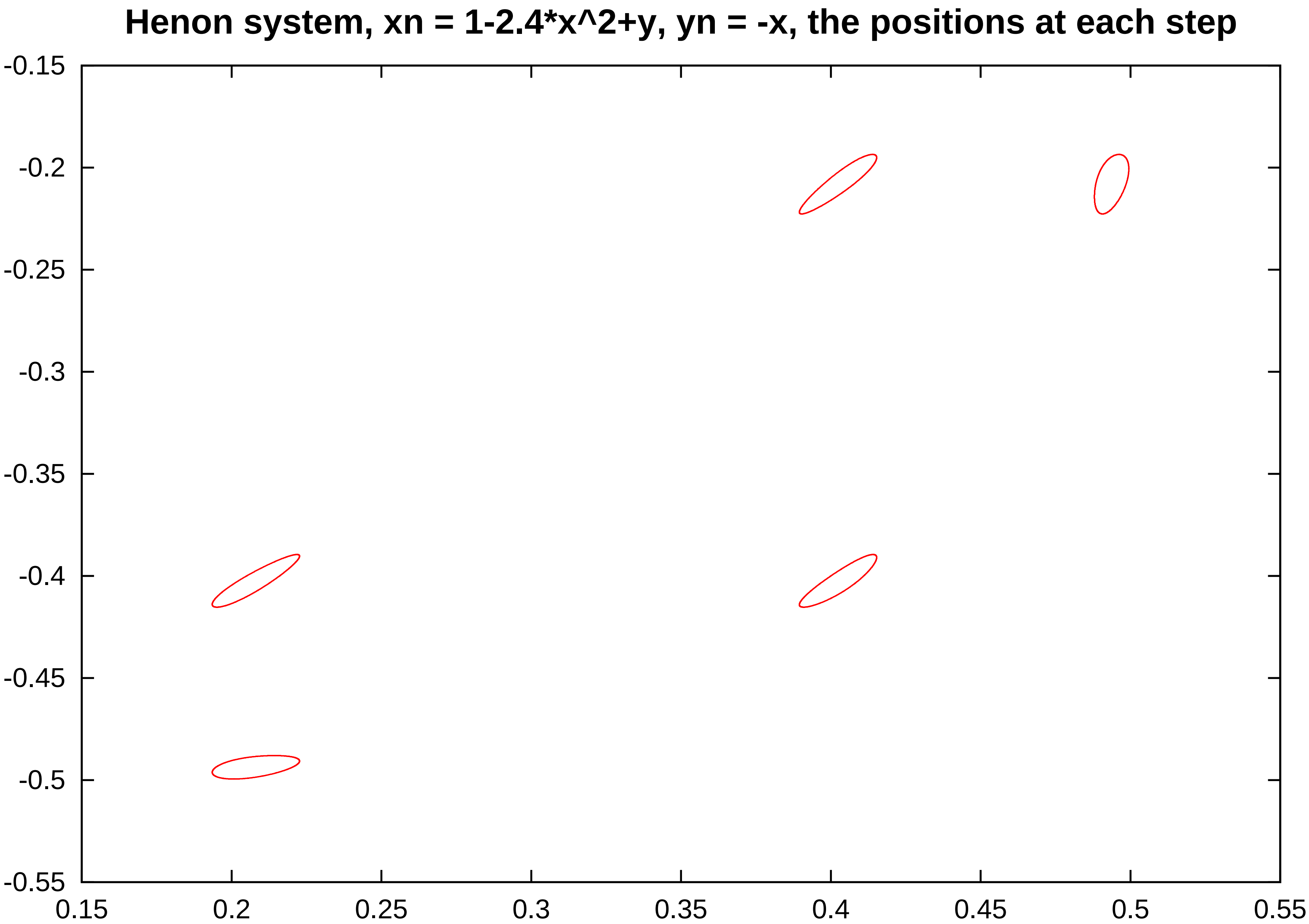
Henon Map: frequently used elementary example that exhibits many of the well-known effects of nonlinear dynamics, including chaos, periodic fixed points, islands and symplectic motion. The dynamics is two-dimensional, and given by

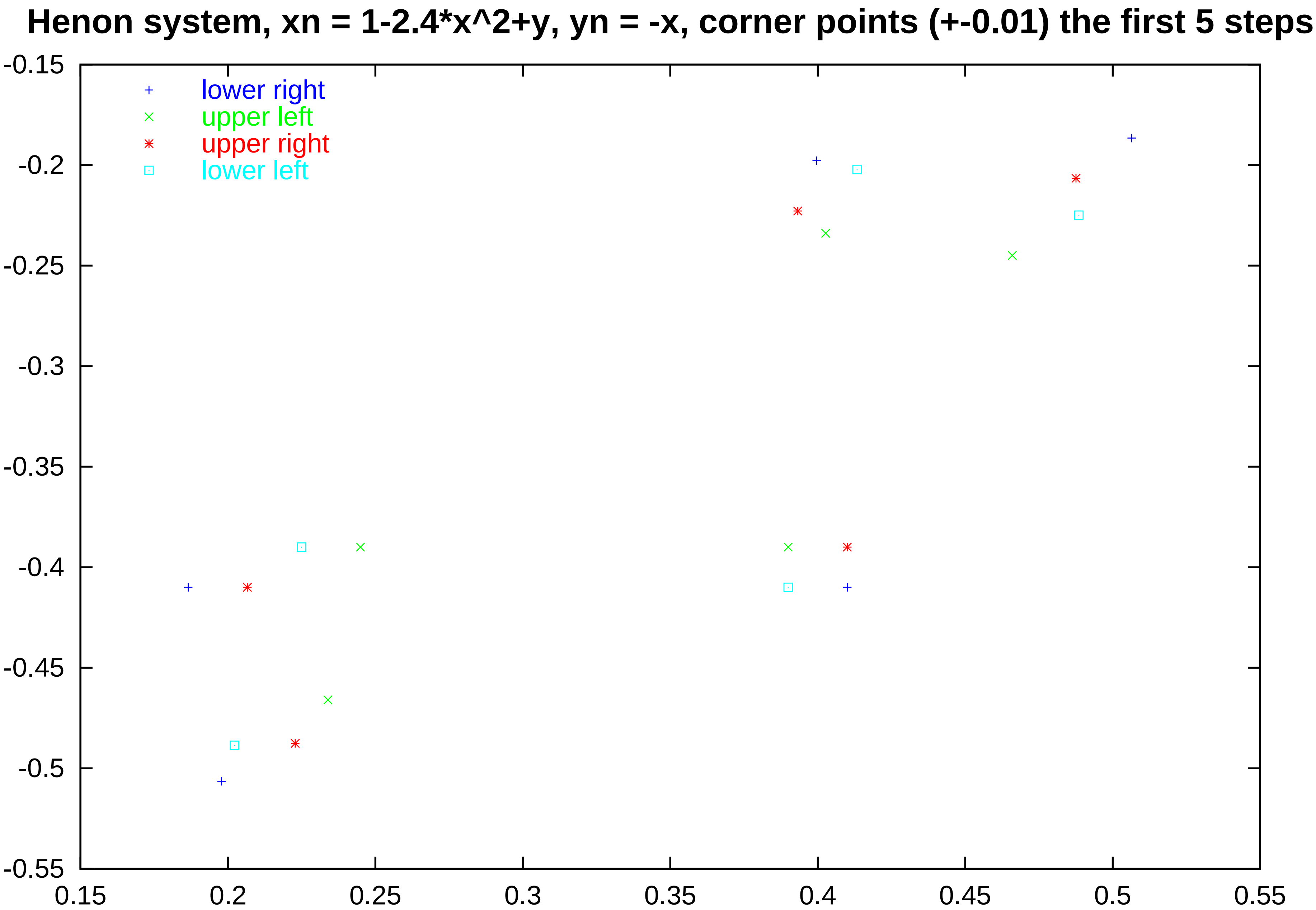
$$x_{n+1} = 1 - \alpha x_n^2 + y_n$$
$$y_{n+1} = \beta x_n.$$

It can easily be seen that the motion is area preserving for $|\beta| = 1$.We consider

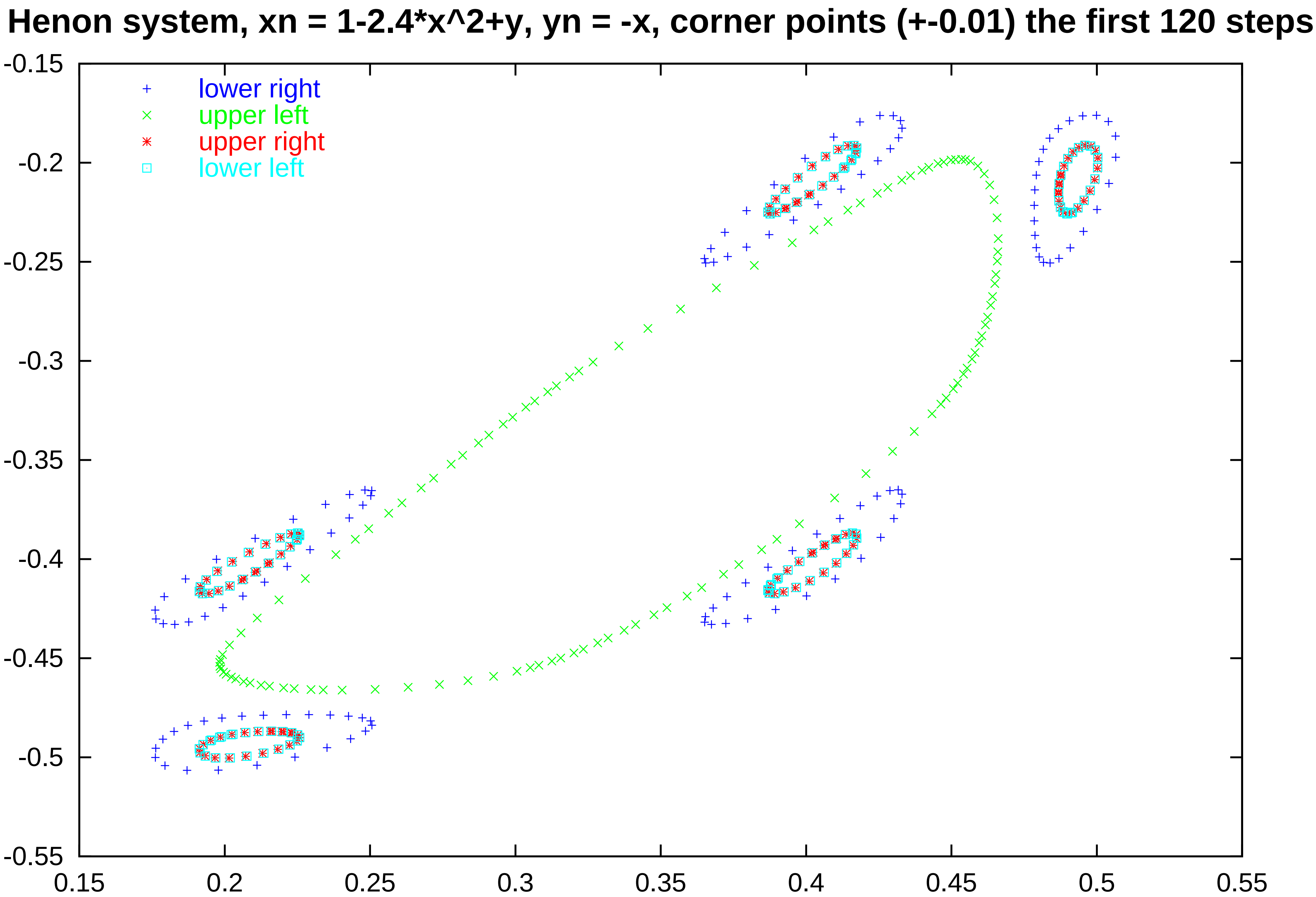
$$\alpha = 2.4 \text{ and } \beta = -1,$$

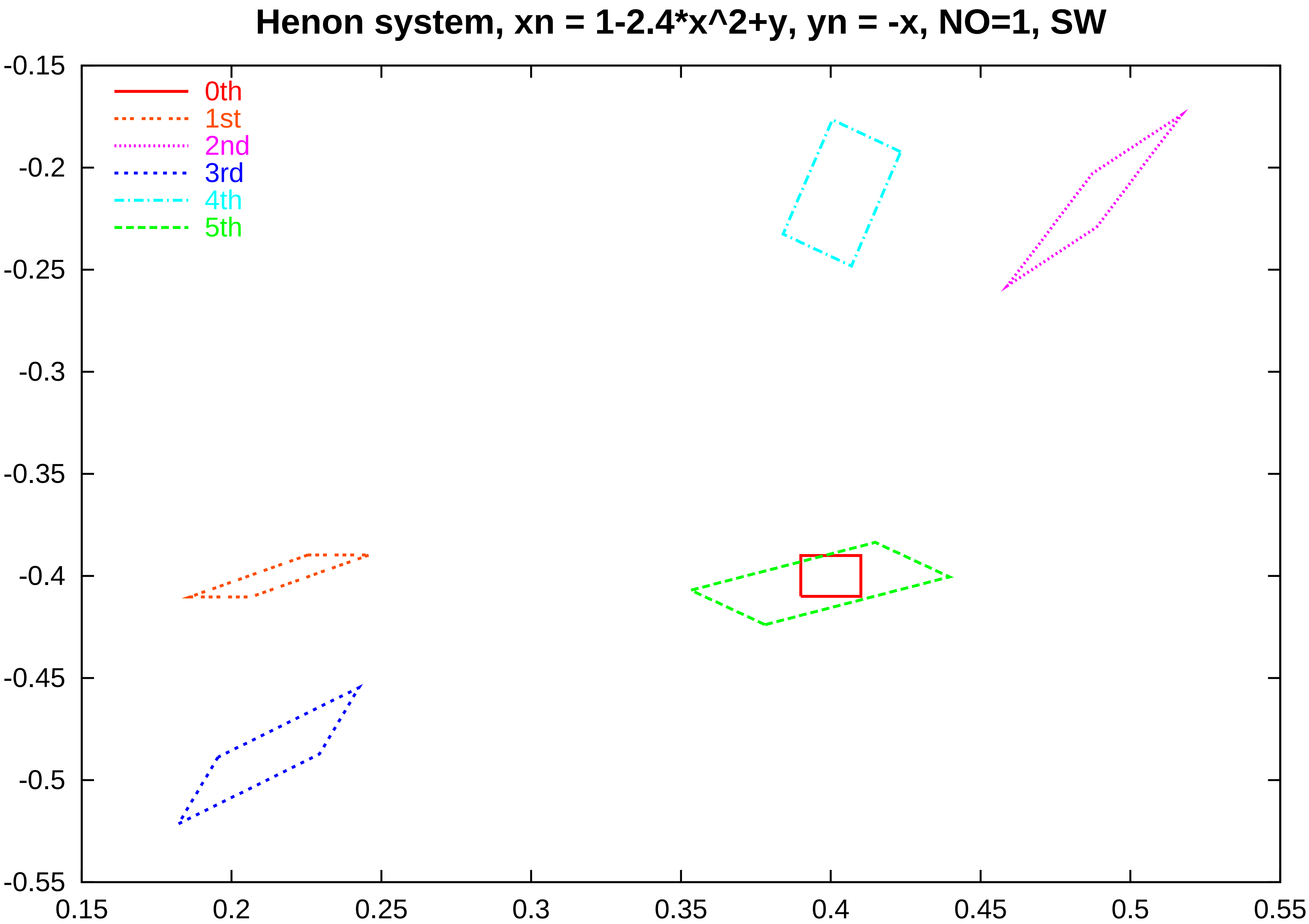
and concentrate on initial boxes of the from $(x_0, y_0) \in (0.4, -0.4) + [-d, d]^2$.

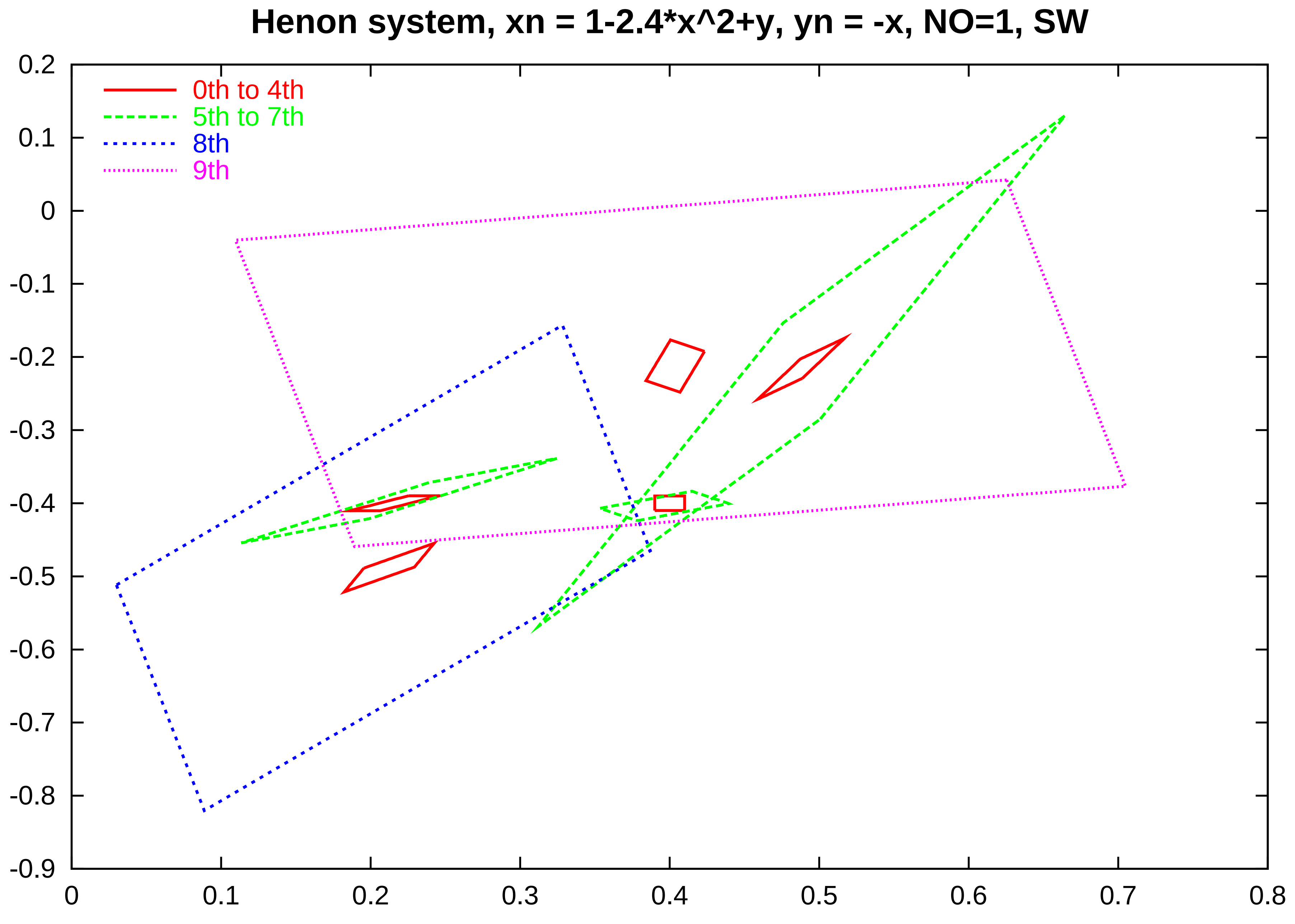

Henon system, xn = 1-2.4*x^2+y, yn = -x, the positions at each step
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4
-0.45
-0.5
-0.55
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55



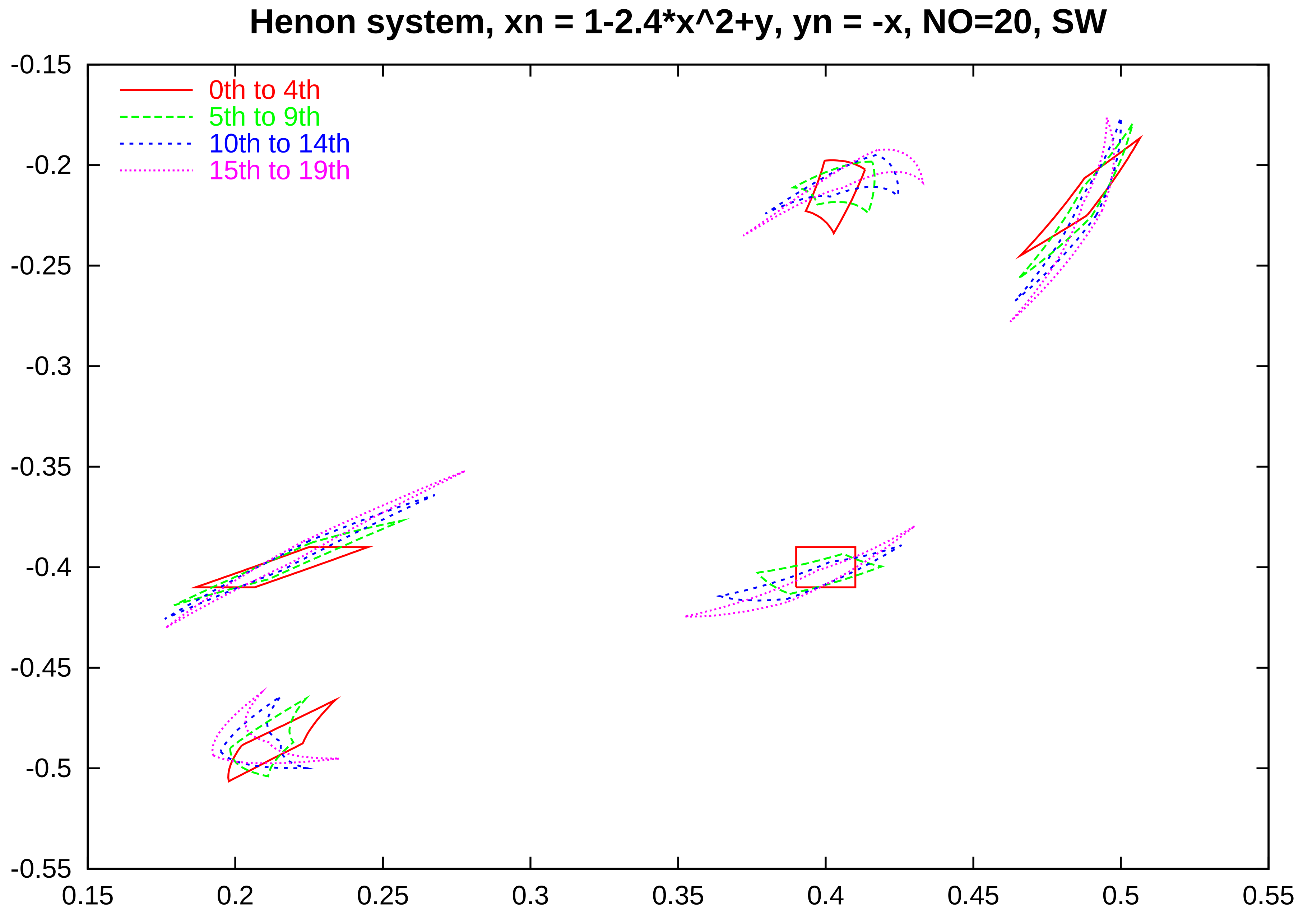
Henon system, xn = 1-2.4*x^2+y, yn = -x, corner points (+-0.01) the first 5 steps
lower right
upper left
upper right
lower left
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4
-0.45
-0.5
-0.55
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55

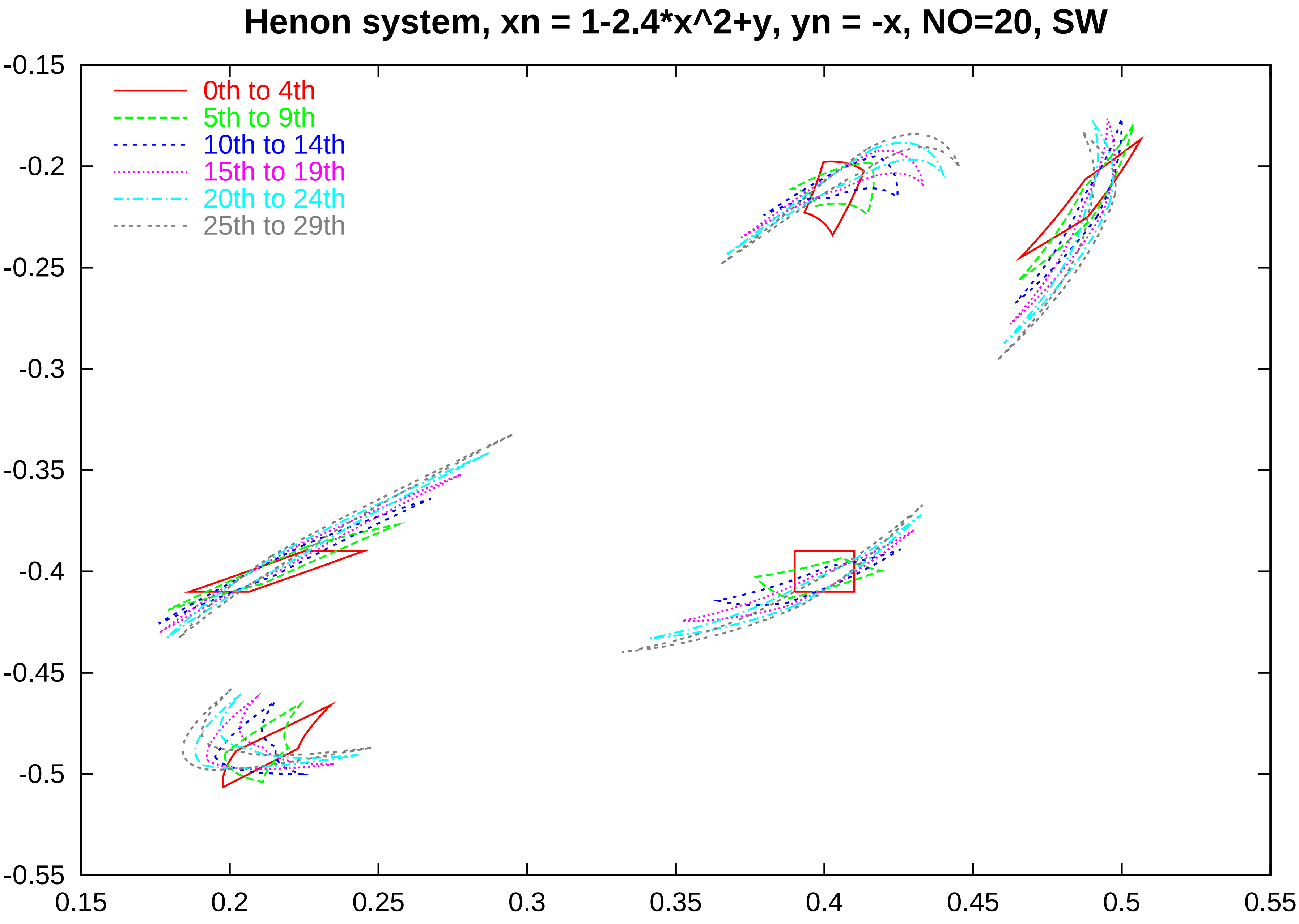

Henon system, xn = 1-2.4*x^2+y, yn = -x, corner points (+-0.01) the first 120 steps
lower right
upper left
upper right
lower left
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4
-0.45
-0.5
-0.55
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55


Henon system, xn = 1-2.4*x^2+y, yn = -x, NO=1, SW
0th
1st
2nd
3rd
4th
5th
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4
-0.45
-0.5
-0.55
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55



Henon system, xn = 1-2.4*x^2+y, yn = -x, NO=1, SW
0th to 4th
5th to 7th
8th
9th
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8


Henon system, xn = 1-2.4*x^2+y, yn = -x, NO=20, SW
0th to 4th
5th to 9th
10th to 14th
15th to 19th
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4
-0.45
-0.5
-0.55
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55



Henon system, xn = 1-2.4*x^2+y, yn = -x, NO=20, SW
0th to 4th
5th to 9th
10th to 14th
15th to 19th
20th to 24th
25th to 29th
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4
-0.45
-0.5
-0.55
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55

Key Features and Algorithms of COSY-GO

- List management of boxes not yet determined to not contain the global minimizer. Loading a new box. Discarding a box with range above the current threshold value. Splitting a box with range not above the threshold value for further analysis. Storing a box smaller than the specified size.
- Application of a series of bounding schemes, starting from mere interval arithmetic to naive Taylor model bounding, LDB, then QFB. A higher bounding scheme is executed only if all the lower schemes fail.
- Update of the threshold cutoff value via various schemes. It includes upper bound estimates of the local minimum by corresponding bounding schemes, the mid point estimate, global estimates based on local behavior of function using gradient line search and convex quadratic form.
- Box size reduction using LDB QPB.
- Resulting data is available in various levels including graphics output.

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 pre-determined 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

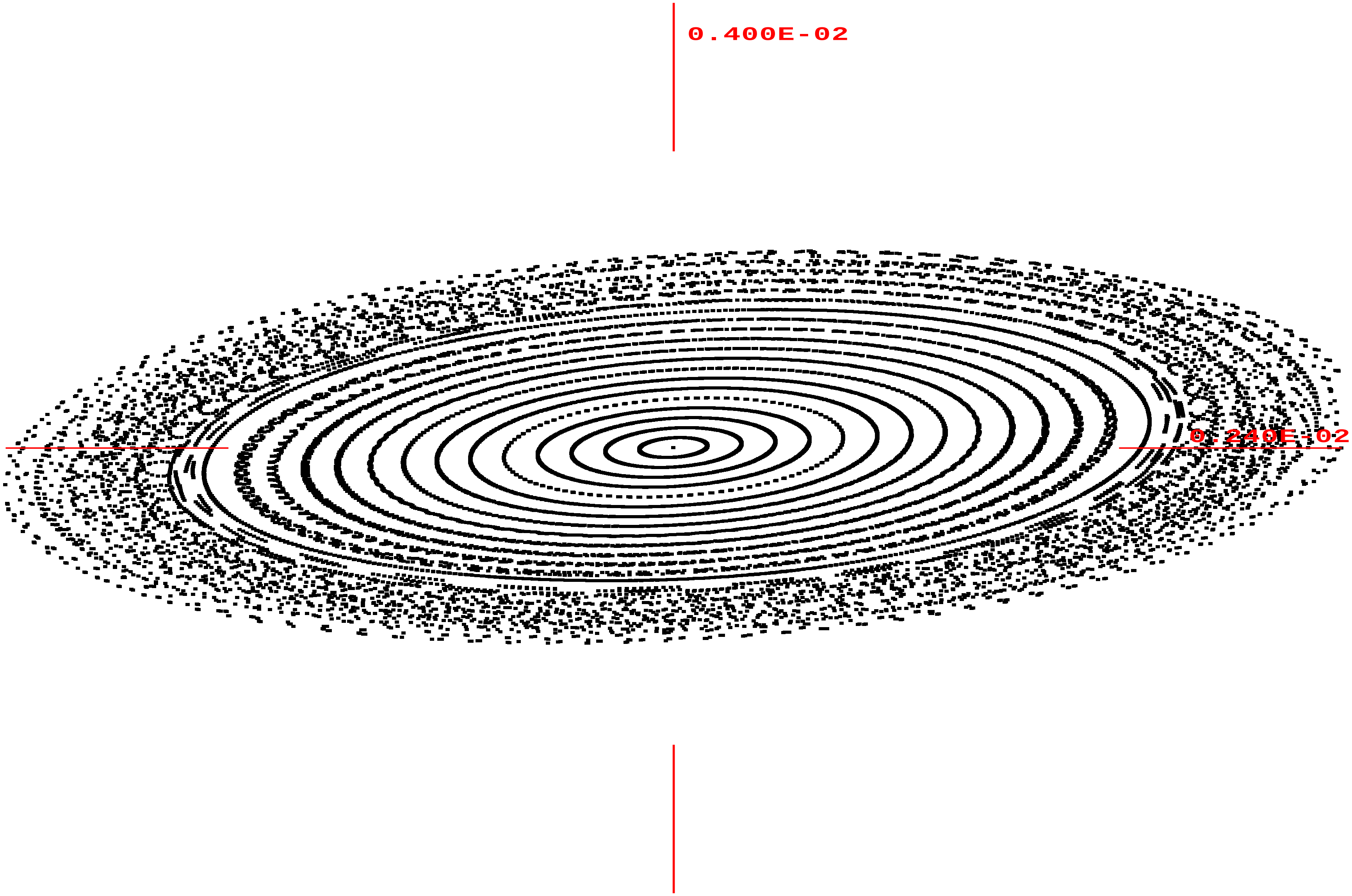
What Happens Between Meetings

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.

Design Parameter Optimization Example

Determine magnet parameters to eliminate several nonlinear aberrations.
Using nonverified local optimizers and a verified global optimizer:

	Sim. Annealing	Simplex	LMDIF	COSY-GO
S1	-4.220040774	-.1752976958	-.4129845844E-4	0.3531413928E-1
S2	-2.468406098	-.1270744624	-.1800819610E-3	0.1920157459E-1
S3	5.268669445	-.6394106627E-1	-.2079220929E-3	-.2907299679
S4	-7.144285915	0.1167017471	-.9007439665E-4	0.2478068398
S5	1.174139460	0.2231863960	0.2139893916	0.2478068398
S6	-8.121172250	-.7800825893E-1	-.2800618129E-3	-.2907299679
S7	9.135056510	-.2296525059	-.3195229128	0.1920157430E-1
S8	-2.243273322	0.9506890185E-1	0.1165687798	0.3531414016E-1
Obj	0.9348125285E-3	0.5486180417E-7	0.1003910906E-4	0.7644984272E-26
Steps	10000	691	10000	


0.400E-02
0.240E-02

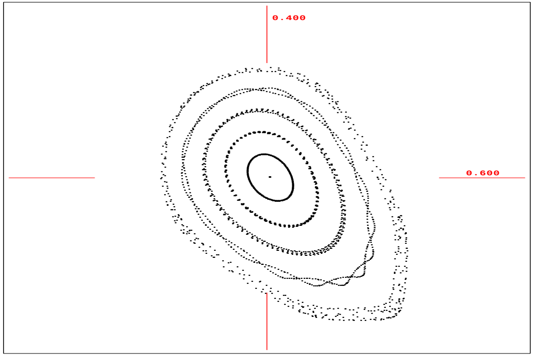


Figure 3. Motion of five particles of a dynamical system. Different from the system shown in figure 1, here the linearized motion has eienvalues of unity modulus.

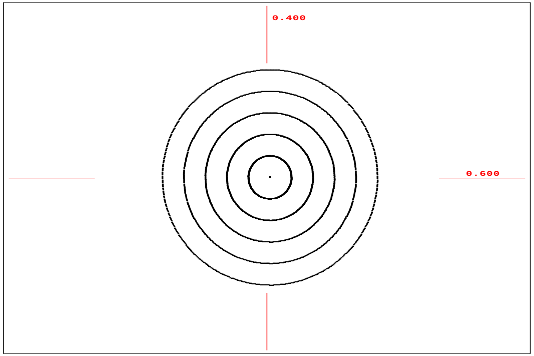
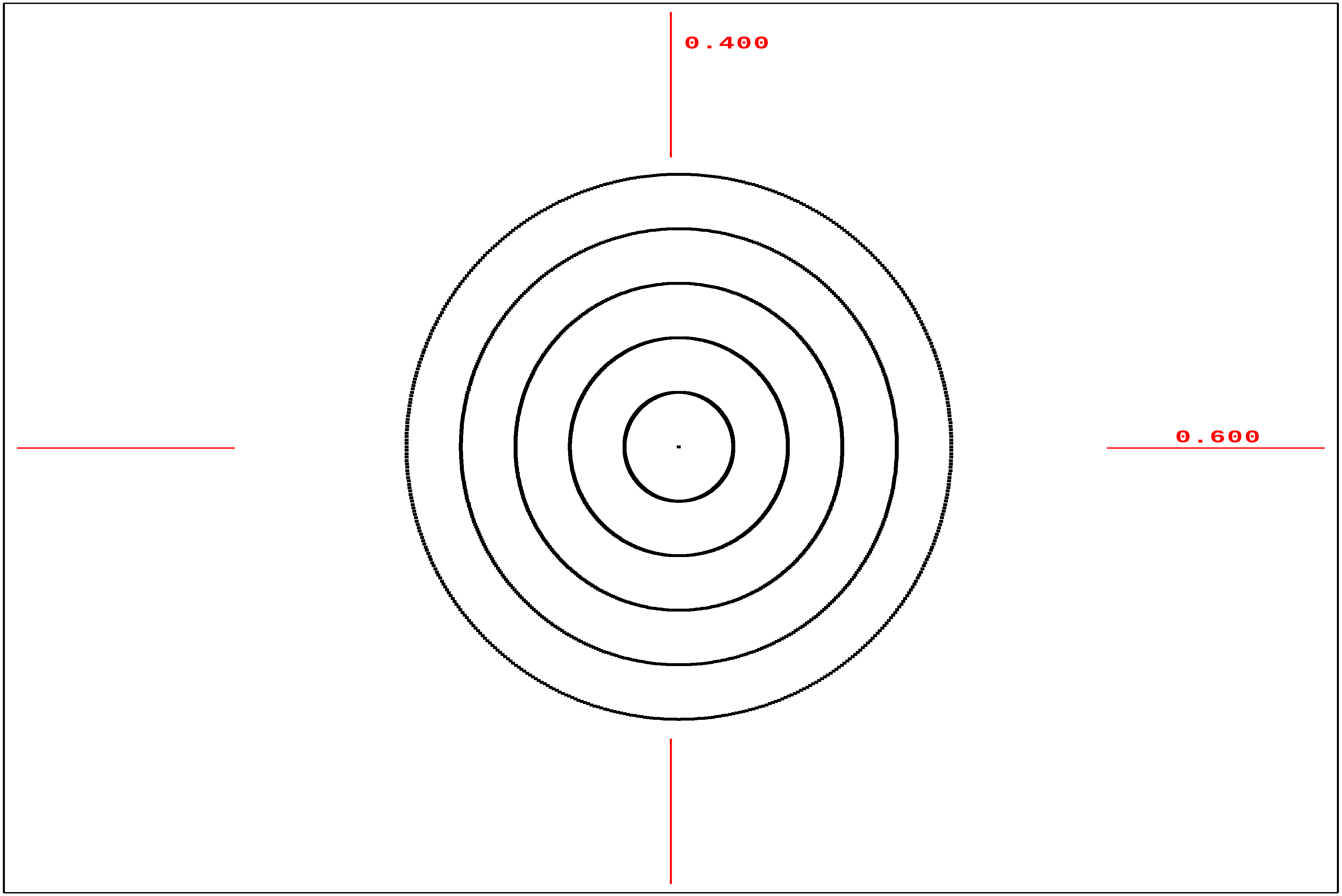
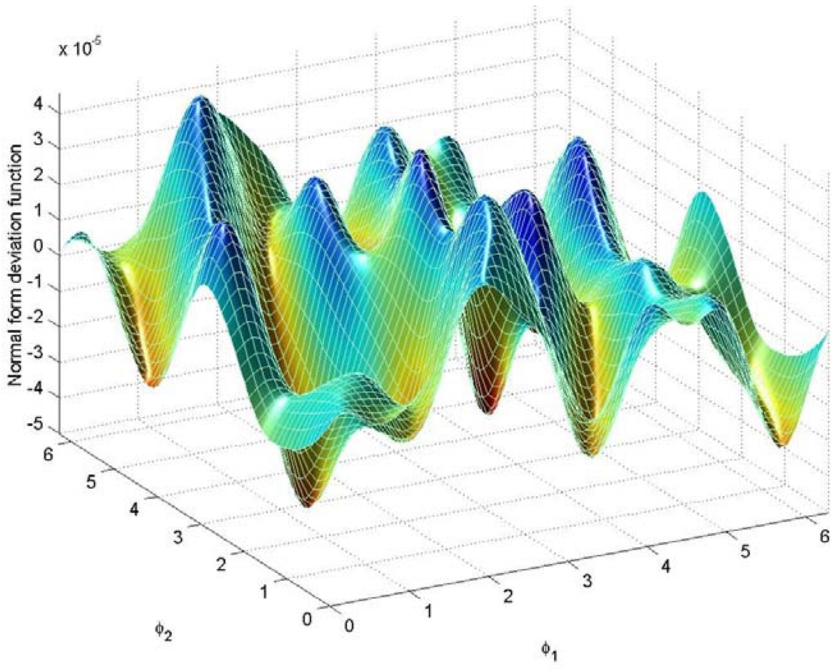


Figure 4. Motion of five particles of the system from Figure 3, displayed in normal form coordinates. The distance to the origin represents a perturbative first integral of the motion.


0.400
0.600


x 10-5
Normal form deviation function
4
3
2
1
0
-1
-2
-3
-4
-5
6
5
4
3
2
1
0
ϕ2
0
1
2
3
4
5
6
ϕ1

The Normal Form Invariant Defect Function

- **Extreme cancellation**; one of the reasons TM methods were invented
- Six-dimensional problem from dynamical systems theory
- Describes invariance defects of a particle accelerator
- Essentially composition of three tenth order polynomials
- The function vanishes identically to order ten
- Study for $a \cdot (1,1,1,1,1,1)$ for $a = .1$ and $a = .2$
- Interesting **Speed observation**: on same machine,
  * one CF in INTLAB takes 45 minutes (Version 3.1 under Matlab V.6)
  * one TM of order 7 takes 10 seconds

$$f(x_1, .., x_6) = \sum_{i=1}^{3} \left(\sqrt{y_{2i-1}^2 + y_{2i}^2} - \sqrt{x_{2i-1}^2 + x_{2i}^2} \right)^2$$

$$\text{where } \vec{y} = \vec{P}_1 \left(\vec{P}_2 \left(\vec{P}_3(\vec{x}) \right) \right)$$

The Tevatron NF Invariant Defect Function

Estimate bound of the defect function over the Tevatron actual emittance (radius r) by **global optimization**. Make the stability estimate via Nekhoroshev-type for $2 \cdot r$.

The Tevatron NF Defect Function - GlobSol Results

For the computations, GlobSol's maximum list size was changed to 10^6, and the CPU limit was set to 10 days. All other parameters affecting the performance of GlobSol were left at their default values.

Dimension	CPU-time needed	Max list	Total # of Boxes
2	18810 sec		4733
3	>562896 sec (not finished yet)		
4	>259200 sec (could not finish)		63446 (remaining)
5	> 86400 sec (could not finish)		21306 (remaining)
6	not attempted		

We observe that in this example, COSY outperforms GlobSol by many orders of magnitude. However, we are not completely sure if a different choice of parameters for GlobSol could result in better performance.

The Tevatron NF Defect Function - COSY-GO Results

Tolerance on the sharpness of the resulting minimum is 10^{-10}.For the evaluation of the objective function, Taylor models of order 5 were used. For the range bounding of the Taylor models, LDB with domain reduction was being used.

Dimension	CPU-time needed	Max list	Total # of Boxes
2	5.747071 sec	11	31
3	38.48828 sec	44	172
4	346.8604 sec	357	989
5	3970.746 sec	2248	6641
6	57841.94 sec	17241	49821

The Tevatron NF Invariant Defect Function

Estimate bound of the defect function over the Tevatron actual emittance (radius r) by **global optimization**. Make the stability estimate via Nekhoroshev-type for $2 \cdot r$.

- The result was very much limited by floating point floor.
- Can guarantee stability for 10^7 turns (emittance: $1.2\pi \cdot 10^{-4}$mm mrad, normal form radius $R_{NF} = 10^{-5}$).

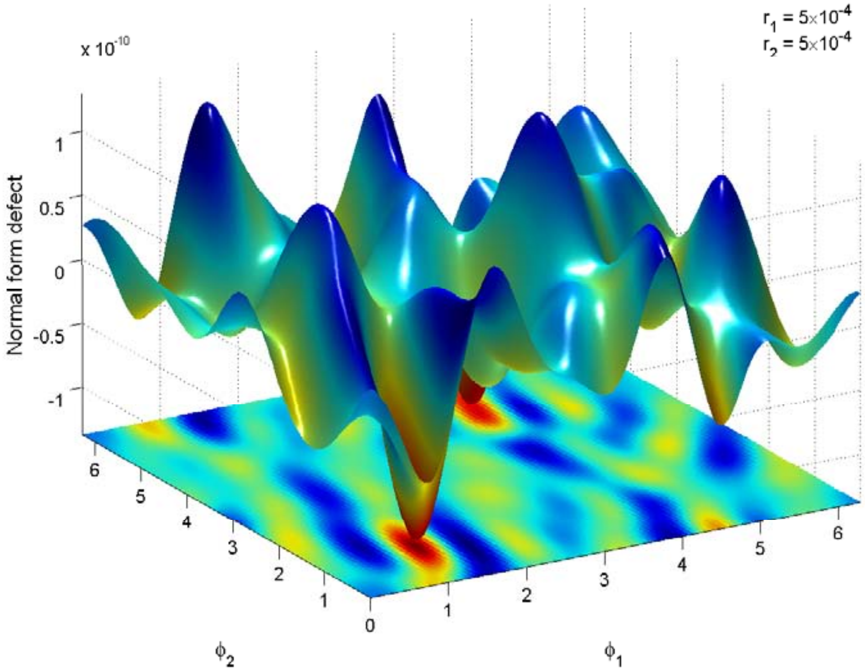


Fig. 9. Projection of the normal form defect function. Dependence on two angle variables for the fixed radii $r_1 = r_2 = 5 \cdot 10^{-4}$

Region	Boxes studied	CPU-time	Bound	Transversal Iterations
$[0.2, 0.4] \cdot 10^{-4}$	82,930	30,603 sec	$0.859 \cdot 10^{-13}$	$2.3283 \cdot 10^8$
$[0.4, 0.6] \cdot 10^{-4}$	82,626	30,603 sec	$0.587 \cdot 10^{-12}$	$3.4072 \cdot 10^7$
$[0.6, 0.9] \cdot 10^{-4}$	64,131	14,441 sec	$0.616 \cdot 10^{-11}$	$4.8701 \cdot 10^6$
$[0.9, 1.2] \cdot 10^{-4}$	73,701	13,501 sec	$0.372 \cdot 10^{-10}$	$8.0645 \cdot 10^5$
$[1.2, 1.5] \cdot 10^{-4}$	106,929	24,304 sec	$0.144 \cdot 10^{-9}$	$2.0833 \cdot 10^5$
$[1.5, 1.8] \cdot 10^{-4}$	111,391	26,103 sec	$0.314 \cdot 10^{-9}$	$0.95541 \cdot 10^5$

Table 8
Global bounds obtained for six radial regions in normal form space for the Tevatron. Also computed are the guaranteed minimum transversal iterations.


Fourth International Workshop
on Taylor Methods
Boca Raton, Florida
December 16-19, 2006
http://bt.pa.msu.edu/TM/BocaRaton2006/
Topics:
High-Order Methods
Verification & Taylor Models
Automatic Differentiation
Differential Algebraic Tools
and their use for:
ODE and PDE Solvers
Global Optimization
Constraint Satisfaction
Dynamical Systems
Beam Physics
Support: Department of Energy
Michigan State University