Preventing Premature Convergence and Proving the Optimality in Evolutionary Algorithms

Charlie Vanaret
vanaret@cena.fr
Jean-Baptiste Gotteland   Nicolas Durand

Journée de présentation des doctorants, Equipe APO
13 septembre 2013

Objectives

- solve **nonlinear multimodal problems**
- find and **bound** a **global minimum** $(x^*,\ f^*)$

Hard task to find global optimum of numerical (un)constrained problems

- deterministic methods do not converge within reasonable time
- stochastic methods do not guarantee optimality

Considered approach: combine

- a **stochastic** research (Evolutionary Algorithms)
- a **deterministic** research (Interval-based algorithms)

1. Combining Evolutionary Algorithms and Interval-based methods

2. Speeding up the convergence

3. Experimental results

Evolutionary Algorithms (EA)

Evolutionary algorithms $\subset$ metaheuristics: random walk guided by heuristics

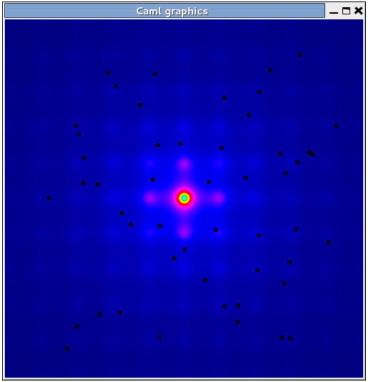
- do not require specific properties on f (black box)
- ideal for **difficult problems**: mixed variables, constraints
- costly methods, **no guarantee of optimality**

Stochastic global optimization algorithms

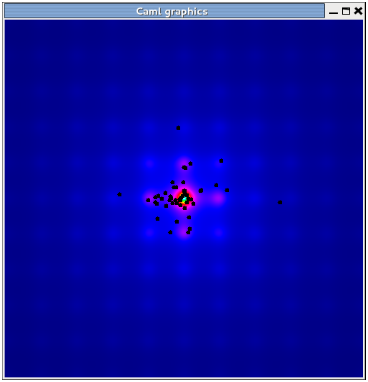
- based on the **theory of evolution**
- selection, crossover, mutation operators
- iterative improvement of a **population of individuals** x_i
- adaptation criterion (fitness) $f(x_i)$

Evolutionary Algorithms (EA)

- GA [Holland, 1975], DE [Storn and Price, 1997], PSO, ES, Genetic programming



Initial population



After 50 iterations

Figure: Griewank function ($n = 2$), 50 individuals

Interval Analysis (IA)

Numerical analysis method to bound round-off errors [Moore, 1966]

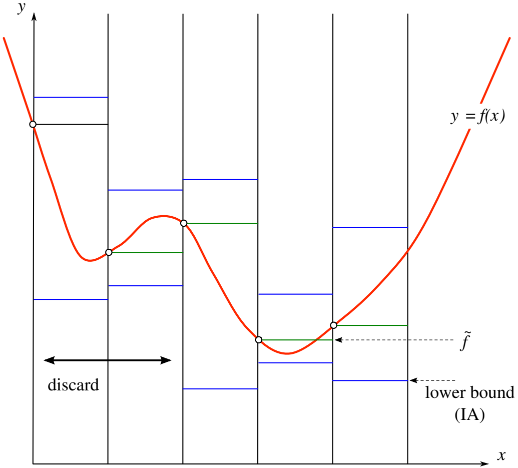
$$\text{Interval } X = [\underline{X}, \overline{X}] = \{x \in \mathbb{R} \mid \underline{X} \leq x \leq \overline{X}\}$$
$$\text{Box } \mathbf{X} = \text{interval vector}$$

Interval arithmetic

- computing with **sets** $\neq$ computing with points
- **rigorous enclosure** (outward rounding)
- **inclusion function** $F: f(X) = \{f(x) \mid x \in X\} \subset F(X)$ (natural, centered, mean value, etc.)

Branch and Bound framework (B&B)

Computation of **guaranteed lower bounds** in B&B [Hansen, 1992]



- If $\tilde{f} \leq \underline{F(X)}$, discard X
- otherwise store in priority queue $\mathcal{L}$

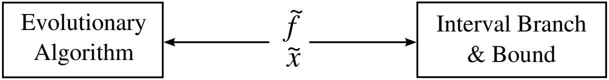
In n dimensions:

- dichotomy variable after variable
- iterate until required precision ϵ on image
- **exponential** complexity $\Rightarrow n$ small

Cooperative hybrid algorithm

Hybridization between **efficiency** of EA and **reliability** of IA

- integrative approach [Sotiropoulos et al., 1997, Zhang and Liu, 2007]
- **cooperative** approach: [Alliot et al., 2012, Vanaret et al., 2013]
  - independent processes exchange bounds and solutions (MPI)
  - **fully reliable** $\neq$ complete (BARON, Couenne) or local (Minos, Ipopt)



- DE: Update $(\tilde{x}, \tilde{f})$ with $(x_{best}, \overline{F(x_{best})})$
- IB&B: Whenever the evaluation of $mid(\mathbf{X})$ improves $\tilde{f}$, add $mid(\mathbf{X})$ to DE population

Improvements

Preliminary results showed that

- on some problems, the EA quickly finds the global minimum
- it takes the IB&B ages to prove the optimality
- how to speed up convergence?

Issues to tackle

- problem inherent to Interval Analysis: large overestimation
- exploit the constraints: propagation of $g_i \leq 0$, $f \leq \tilde{f}$ and $\nabla f = 0$
- enhance the cooperation

The dependency problem

Quality of inclusion depends on syntactic form

- $f(x) = x^2 - 2x$, $F_N([1,4]) = [-7, 14]$
- $g(x) = x(x-2)$, $G_N([1,4]) = [-4, 8]$,
- $h(x) = (x-1)^2 - 1$, $H_N([1,4]) = [-1, 8] = f([1,4])$

Dependency problem = main source of **overestimation**

- multiple occurrences of a variable are decorrelated $(X - X = \{x - y \mid x \in X, y \in X\} \ni 0)$
- optimal range if single occurrence of each variable
- overerstimation error reduces
  - linearly for natural extension
  - quadratically for mean value form $F(X) = F(c) + (X - c)F'(X)$

How does dependency affect the convergence?

Rana's function, $x_i \in [-512, 512]$

- first syntax:

$$f(x) = \sum_{i=1}^{n-1} (x_i \cos u \sin v + (1 + x_{i+1}) \sin u \cos v)$$

where $u = \sqrt{|x_{i+1} + x_i + 1|}$ and $v = \sqrt{|x_{i+1} - x_i + 1|}$

- second syntax: rewriting using $\cos x \sin y = \frac{1}{2}(\sin(x+y) - \sin(x-y))$

Table: CPU times (s) of convergence ($\epsilon = 10^{-6}$, $NP = 70$, $W = 0.7$, $CR = 0.5$)

n	First syntax	Second syntax
2	0.25	0.009
3	6.5	0.12
4	254	1.45
5	∞	18.5
6	∞	244
7	∞	3300

Interval Constraint Programming (ICP)

Constraint propagation may

- **contract** (narrow bounds) a box without loss of solutions
- prove that a box **cannot contain** a global minimizer

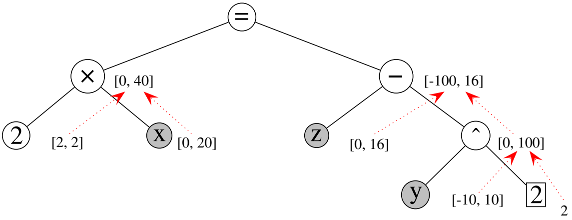
Filtering/Contraction algorithms [Chabert and Jaulin, 2009]

- `HC4` (`HC4Revise`) [Benhamou et al., 1999] for (in)equalities
- Interval Newton for equalities
- Mohc: monotonic `HC4Revise` + Newton [Araya et al., 2010]

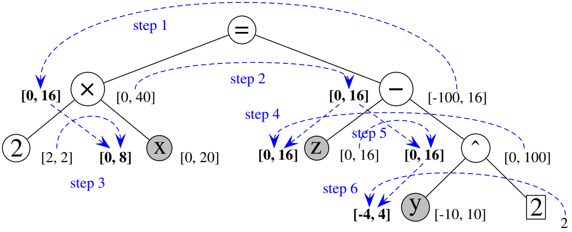
`Charibde` [Vanaret et al., 2013] = DE + Interval Branch and Contract (IB&C)
(IB&B + `HC4Revise` + Mohc + gradient info)

HC4Revise algorithm (w.r.t. $2x = z - y^2$)

- bottom-up: evaluate each node using IA



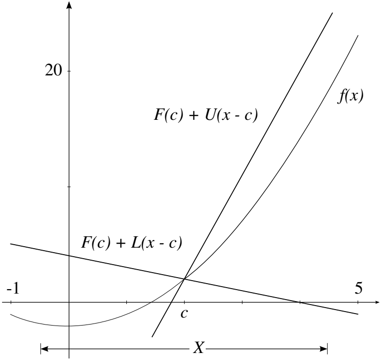
- top-down: constraint propagation using inverse functions



Interval Newton algorithm

Equality constraint $f = 0$

- Newton operator: $N(X, c) = c - \frac{F(c)}{F'(X)}$
- **Any zero** of f in X lies in $X \cap N(X, c)$
- If $N(X, c) \subset int(X)$, then **existence of a unique zero** in X



$$f(x) = x^2 - 2, x \in X = [-\tfrac{1}{2}, \tfrac{9}{2}], c = 2$$

$$N(X, c) = 2 - \frac{2^2 - 2}{2[-\frac{1}{2}, \frac{9}{2}]}$$
$$= [-\infty, \frac{16}{9}] \cup [4, +\infty]$$

(separation of possible zeros)

$$X \cap N(X, x) = [-\frac{1}{2}, \frac{16}{9}] \cup [4, \frac{9}{2}]$$

Computing an upper bound

Computation of good upper bound $\tilde{f}$ crucial in B&B

- deterministic local solvers (MINOS, Ipopt) start from a feasible point
- if x is a feasible point, then $\overline{F(x)}$ is an upper bound of f^*

Identification of feasible solutions

- Evolutionary Algorithms
  - find feasible points from random starting points
  - exploration generally not limited to a neighborhood
- Interval-based methods
  - contraction w.r.t. to $g_i \leq 0$ discards inconsistent values
  - contraction w.r.t. to $g_i \geq 0$ discards **solutions**

New adaptive heuristic of box selection

Box selection from the priority queue determines the order of exploration

- depth first, breadth first, most promising box (smallest lower bound)

Approach that takes advantage of $\tilde{f}$ AND $\tilde{x}$

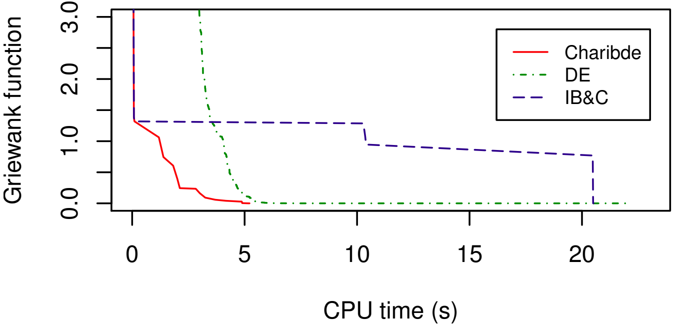
- idea: carry out research in the boxes that are **farthest** from $\tilde{x}$
  - either $\tilde{x}$ close to x^*: fast local exploration by EA
  - or $\tilde{x}$ far from x^*: might as well search elsewhere
- postpone tedious interval search around $\tilde{x}$, waiting for the best possible $\tilde{f}$
- **adaptive** heuristic: recompute priority when $\tilde{x}$ is updated

Experimental results

- generally more efficient than classical heuristics
- maintains a low size of the priority queue

Performance comparison

Griewank function $f(x) = 1 + \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\frac{x_i}{\sqrt{i}}$, $n = 200$



- DE: local optimum close to 0 (22s)
- IB&C: 20.5s
- `Charibde`: 5.2s

Certified optima for a benchmark of functions

Table: Test functions with best known and certified minima

	n	Type	Best known f^*	$\tilde{f}$ by `Charibde`	CPU time (s)
Bound-Constrained Problems					
Rastrigin	50	NL	0	0	0.3
Schwefel	10	NL	-4189.82887	-4189.82887	2.3
Rosenbrock	50	Q	0	0	3.3
Griewank	200	NL	0	0	12
Michalewicz	75	NL	-	**-74.6218112**	138
Egg Holder	10	NL	-8247	**-8291.2400675**	768
Rana	7	NL	-	**-3070.2475210**	3300
Inequality-Constrained Problems					
Himmelblau	5	Q	-31025.56024	-31025.56024	0.07
Beam	4	NL	1.7248523	1.7248523	2.2
Tension	3	P	0.01266523	0.01266523	3.8
Keane	5	NL	-0.63445	-0.63445	472

Lennard-Jones clusters [Jones, 1924]

Find the most stable configuration of a cluster of n atoms

- potential describes pairwise interactions between atoms

$$f(x) = \sum_{i<j}^{n} v(r_{ij}) = 4\sum_{i<j}^{n} \left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}} \right) = 4\sum_{i<j}^{n} \left[\left(\frac{1}{r_{ij}^{6}} - \frac{1}{2} \right)^2 - 1 \right]$$

where $r_{ij}^2 = (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2$

- non-convex and highly combinatorial ($O(e^n)$ local minima)
- open problem for $n \geq 5$
- putative minimum for $n = 5$: -9.103852415708 (triangular bipyramid)

	BARON [Sahinidis, 1996]	Charibde
Optimum	-9.1039	$\tilde{f} = -9.10385241570755$2
CPU (s)	0.002	0.12 + 1000
Status	"locally optimal"	certified ($\epsilon = 10^{-6}$)

Conclusion

`Charibde` algorithm

- prevents premature convergence toward local minima
- **proves the optimality** of the solution
- achieved new optimal results for difficult benchmark functions

Perspectives

- applications to aeronautics (conflict resolution), 3D computer graphics, geometry, optimization of neural networks
- parallelization of IB&B

Perspectives

Perform DE restarts

- add diversity in the search process
- possibly on a reduced search-space
  - IB&B contains list $\mathcal{L}$ of remaining boxes
  - our selection heuristic "peels" the search-space from outside
  - reduced search-space = convex hull (enclosing box) of $\mathcal{L}$
  - crude enclosure...

Table: Lennard-Jones clusters ($n = 5$)

DE iteration	Search-space retrieved from IB&B
0	$[0, 1.2] \times [0, 1.2]^5 \times [-1.2, 1.2] \times [-1.2, 1.2] \times [0, 1.2]$
50000	$[0.94, 1.2] \times [0, 1.2]^5 \times [-0.47, 1.2] \times [-1.2, 1.11] \times [0.24, 1.2]$
60000	$[0.94, 1.2] \times [0, 1.2]^5 \times [-0.47, 1.2] \times [-1.2, 1.11] \times [0.52, 1.2]$

References I

Alliot, J.-M., Durand, N., Gianazza, D., and Gotteland, J.-B. (2012).
Finding and proving the optimum: Cooperative stochastic and deterministic search.
20th European Conference on Artificial Intelligence.

Araya, I., Trombettoni, G., and Neveu, B. (2010).
Exploiting monotonicity in interval constraint propagation.
In *Proc. AAAI*, pages 9–14.

Benhamou, F., Goualard, F., Granvilliers, L., and Puget, J.-F. (1999).
Revising hull and box consistency.
In *International Conference on Logic Programming*, pages 230–244. MIT press.

Chabert, G. and Jaulin, L. (2009).
Contractor programming.
Artificial Intelligence, 173:1079–1100.

References II

Hansen, E. (1992).
Global optimization using interval analysis.
Dekker.

Holland, J. H. (1975).
Adaptation in Natural and Artificial Systems.
University of Michigan Press.

Jones, J. E. (1924).
On the determination of molecular fields. i. from the variation of the viscosity of a gas with temperature.
Proceedings of the Royal Society of London. Series A, 106(738):441–462.

Moore, R. E. (1966).
Interval Analysis.
Prentice-Hall.

References III

Sahinidis, N. V. (1996).
Baron: A general purpose global optimization software package.
Journal of Global Optimization, 8(2):201–205.

Sotiropoulos, G. D., Stavropoulos, C. E., and Vrahatis, N. M. (1997).
A new hybrid genetic algorithm for global optimization.
In *Proceedings of second world congress on Nonlinear analysts*, pages 4529–4538. Elsevier Science Publishers Ltd.

Storn, R. and Price, K. (1997).
Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces.
Journal of Global Optimization, pages 341–359.

Vanaret, C., Gotteland, J.-B., Durand, N., and Alliot, J.-M. (2013).
Preventing premature convergence and proving the optimality in evolutionary algorithms.
Lecture Notes in Computer Science.

References IV

Zhang, X. and Liu, S. (2007).
A new interval-genetic algorithm.
International Conference on Natural Computation, 4:193–197.

Preventing Premature Convergence and Proving the Optimality in Evolutionary Algorithms

Charlie Vanaret
vanaret@cena.fr
Jean-Baptiste Gotteland Nicolas Durand

Journée de présentation des doctorants, Equipe APO
13 septembre 2013