Stochastic optimization algorithms
Lecture 17, 20181016

Performance comparison

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

CHALMERS

Today's learning goals

- After this lecture you should be able to
  - Define suitable benchmark functions for comparing and testing optimization algorithms.
  - Define a suitable setup for performance comparison.
  - Select a suitable stochastic optimization algorithm for different kinds of problems.

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Benchmark functions

- Good benchmark functions (Appendix D) should…
  - … have many local optima or a difficult approach (e.g. very small gradient) to the global optimum
  - … have uneven spacing of local optima
  - … not be separable, i.e. it should *not* be possible to write the objective function $f(x_1, x_2, \ldots x_n)$ as $\sum_{i=1}^{n} f_i(x_i)$.
  - … have many local optima along any coordinate axis

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Benchmark functions

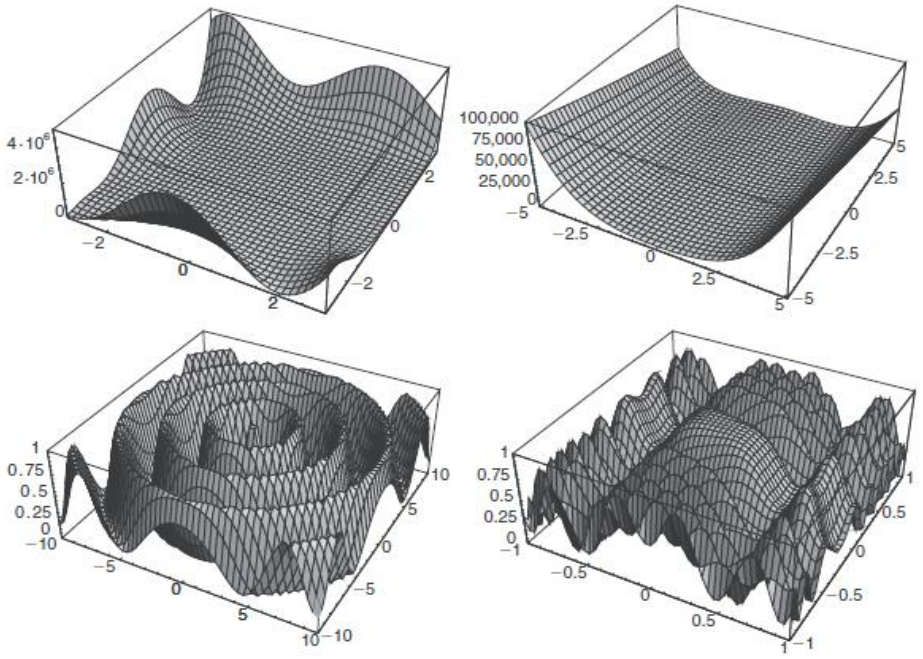

Figure D.1: *Four of the five benchmark functions defined in the text. Top left panel: the Goldstein–Price function; top right panel: the Rosenbrock function, with $n = 2$; bottom left panel: the Sine square function; bottom right panel: the benchmark function $\Psi_5^{[2]}$, with $a = 0.05$ and $b = 10$.*

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

CHALMERS

Today's learning goals

- After this lecture you should be able to
  - Define suitable benchmark functions for comparing and testing optimization algorithms.
  - Define a suitable setup for performance comparison.
  - Select a suitable stochastic optimization algorithm for different kinds of problems.

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Performance comparison

- Two approaches
  1. Running the algorithm until a given value of the objective function is reached,
  2. Running the algorithm for a given number of evaluations (e.g. individuals, in an EA).
- In either case, averages should be taken over many runs.
- Problem with Method 1: In some cases the algorithm may get stuck, distorting the averages. Method 2 is generally preferred.

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Today's learning goals

- After this lecture you should be able to
  - Define suitable benchmark functions for comparing and testing optimization algorithms. ✓
  - Define a suitable setup for performance comparison. ✓
  - Select a suitable stochastic optimization algorithm for different kinds of problems.

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Unconstrained function optimization

- Classical methods preferred when applicable, i.e. typically in problems with (one or more of…)
  - Differentiable objective functions
  - Few variables (or at least not thousands…)
  - Convex objective functions
- Otherwise, use EAs or PSOs.
- Examples, see Table 6.1.
  - $\Psi_2^{[5]}$ has a minimum value of 0.
  - $\Psi_5^{[10]}$ has a maximum value of 1.

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Unconstrained function optimization

$\Psi_2^{[5]}(x_1, \ldots, x_5)$					
Method	Settings	N	I	Avg.	S.D.
GA	Binary	40	250	19.06	40.64
GA	Binary	100	100	3.849	7.138
GA	Binary	250	40	2.401	3.403
GA	RN, $p_{cr} = 0.8$, $C_r = 0.2$	250	40	5.404	1.682
GA	RN, $p_{cr} = 0.8$, $C_r = 0.02$	250	40	1.829	1.244
PSO	–	20	500	0.727	1.161
PSO	–	40	250	0.689	0.845
PSO	–	100	100	2.600	1.325
PSO	–	250	40	55.90	31.64
RS	–	1	10,000	95.19	48.74

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Unconstrained function optimization

$\Psi_5^{[10]}(x_1, \ldots, x_{10})$					
Method	Settings	N	I	Avg.	S.D.
GA	Binary	40	250	0.817	0.058
GA	Binary	100	100	0.875	0.048
GA	Binary	250	40	0.875	0.038
GA	RN, $p_{cr} = 0.8$, $C_r = 0.2$	250	40	0.747	0.034
GA	RN, $p_{cr} = 0.8$, $C_r = 0.02$	250	40	0.851	0.044
PSO	–	20	500	0.898	0.049
PSO	–	40	250	0.881	0.049
PSO	–	100	100	0.786	0.035
PSO	–	250	40	0.684	0.030
RS	–	1	10,000	0.687	0.026

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Unconstrained function optimization

- Note:
  - The SOAs generally outperform random search,
  - No systematic difference between binary and real-number encoding,
  - PSO does very well, particularly for small swarm sizes,
  - Parameter setting is important: PSO is both best and worst (with different parameter settings) for $\Psi_5^{[10]}$!
  - The runs in Tables 6.1 and 6.2 were quite short. For longer runs, the SOAs would almost always outperform RS by an even bigger margin.

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Constrained function optimization

- Several methods available in classical optimization.
- Penalty method: Can also be used in connection with SOAs:
  - Define a penalty function
  - Solve the unconstrained problem using EA, PSO etc.
- Two specific examples given in Tables 6.2 and 6.3.

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Constrained function optimization

Method	Settings	N	I	$r\,(\mu = 100)$	$r\,(\mu = 1{,}000)$
GA	Binary	100	200	0.64	0.35
GA	Binary	250	80	0.95	0.75
GA	RN, $p_{cr} = 0.8$, $C_r = 0.10$	250	80	0.76	0.57
PSO	–	40	500	1.00	1.00
PSO	–	100	200	1.00	1.00

$\min \quad (x_1 - 2)^2 + (x_2 - 1)^2$ subject to

$x_1 + x_2 - 2 \leq 0$

$x_1^2 - x_2 \leq 0$

(minimum value = 1)

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Constrained function optimization

- Note:
  - PSO did better in this case, but…
  - …the PSO often got stuck *near* the optimum, whereas the EA (with real-number encoding) reached more or less the exact minimum *in its successful runs*.

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Constrained function optimization

Method	Settings	N	I	$\overline{f}_2, \overline{p}$ $(\mu = 10^5)$	$\overline{f}_2, \overline{p}$ $(\mu = 10^6)$
GA	RN	100	200	−6,968.73, 37.31	−6,735.30, 8.09
GA	RN	250	80	−6,598.83, 65.63	−6,671.18, 4.73
PSO	–	40	500	−6,968.59, 6.76	−6,962.57, 0.68
PSO	–	100	200	−6,968.47, 6.52	−6,959.69, 0.56

$\min \quad (x_1 - 10)^3 + (x_2 - 20)^3$ subject to

$100 - (x_1 - 5)^2 - (x_2 - 5)^2 \leq 0$

$-82.81 + (x_1 - 6)^2 + (x_2 - 5)^2 \leq 0$

x_1 in [13, 100] and x_2 in [0, 100]

(Constrained) minimum = -6961.813

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Neural network optimization

- GA, PSO, and backpropagation (gradient descent) were applied to the three-parity problem.
- N-parity problem:
  - If an even number of inputs are equal to 1, give a 0 as output.
  - If an odd number of inputs are equal to 1, give a 1 as output.
- A 3-3-1 FFNN was used.
- Training: 20000 epochs for backpropagation, 20000 evaluations for GA and PSO.

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Neural network optimization

Method	Settings	N	I	r	Avg.	S.D.
BP	$\eta = 0.15$	1	20,000	0.87	0.0108	0.0012
BP	$\eta = 0.30$	1	20,000	0.93	0.0068	0.0004
GA	RN	100	200	1.00	0.0008	0.0004
GA	Binary	100	200	0.67	0.0042	0.0004
PSO	–	40	500	0.90	0.0002	0.0012
PSO		100	200	0.96	0.0000	0.0000

BP = backpropagation. I *denotes the number of iterations (i.e. the number of generations, in the case of the GA).*

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Neural network optimization

- Note:
  - The SOAs outperformed backpropagation for this problem.
  - The PSO generally got the best results…
  - …but the GA (with real-number encoding, in this case) was slightly better at finding acceptable solutions (r = 1.00 vs. r = 0.96 for the best PSO).

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Travelling salesman problem

Method	Settings	N	I	Avg.	S.D.
AS	$\alpha = 1, \beta = 2, \tau_0 = 0.454$	100	100	179.68	2.81
AS	$\alpha = 1, \beta = 5, \tau_0 = 0.454$	100	100	173.84	1.89
MMAS	$\alpha = 1, \beta = 5, \tau_0 = 0.00909$	100	100	193.58	2.67
GA-R	$p_c = 0.00, p_{mut} = 0.01$	50	200	577.18	16.94
GA-R	$p_c = 0.00, p_{mut} = 0.03$	50	200	658.40	25.45
GA-R	$p_c = 0.20, p_{mut} = 0.01$	100	100	647.57	18.73
GA-NN	$p_c = 0.00, p_{mut} = 0.01$	50	200	217.89	16.04
GA-NN	$p_c = 0.00, p_{mut} = 0.01$	100	100	211.67	12.62
GA-NN	$p_c = 0.20, p_{mut} = 0.01$	100	100	216.68	10.29
GA-NN	$p_c = 0.00, p_{mut} = 0.01$	200	50	206.57	10.13

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

Travelling salesman problem

- Note:
  - A naïvely applied GA did not do very well, but…
  - …a GA started from the nearest-neighbor path did much better than a GA starting from random paths.
  - Still, ACO outperformed all GAs for this problem.
  - AS did better than MMAS in this case.
  - If the running time is included in the comparison, the performance difference between ACO and GA becomes smaller!

Today's learning goals

- After this lecture you should be able to
  - Define suitable benchmark functions for comparing and testing optimization algorithms.
  - Define a suitable setup for performance comparison.
  - Select a suitable stochastic optimization algorithm for different kinds of problems.

Mattias Wahde, PhD, Professor, Chalmers University of Technology
e-mail: mattias.wahde@chalmers.se, http://www.me.chalmers.se/~mwahde

CHALMERS