PSO & Parameter Selection

Christoph Strößner

September 26, 2014

How nerds find the moutain top

How nerds find the moutain top

smartphone

How nerds find the moutain top

smartphone

gps

facebook

How nerds find the moutain top

smartphone

gps

facebook

NSA

Particle Swarm Optimization

smartphone	particles
gps	position
facebook	velocity
NSA	result

Problems

function: $\mathcal{S} \to \mathbb{R}$

possible additional constraints

"almost nonsolvable"

Algorithm

```
for each particle i(i=1,...,m) do {particle initial.}
    initialize position and velocity
end for
initialize neighborhood structure
while (termination criteria not met)
    for each particle do {particle movement}
        velocity update
        position update
    end for
for each particle do {best update}
    if success then update
end for
end while
```

Initialization of Particles

number of particles

- 20-50
- $10 + 2\sqrt{2d}$ where d = dimension

Initialization of Particles

number of particles

- 20-50
- $10 + 2\sqrt{2d}$ where d = dimension

position

- uniformly at random in search space

Initialization of Particles

number of particles

- 20-50
- $10 + 2\sqrt{2d}$ where d = dimension

position

- uniformly at random in search space

velocity

- uniformly random from search space
- zero
- half-diff: middle between random vector in search space

Neighborhood

- graph representation
  - ring
  - fully connected
  - ...

Neighborhood

- graph representation
  - ring
  - fully connected
  - ...

now: always fully connected

Algorithm

```
for each particle i(i=1,...,m) do {particle initial.}
    initialize position and velocity
end for
initialize neighborhood structure
while (termination criteria not met)
    for each particle do {particle movement}
        velocity update
        position update
    end for
for each particle do {best update}
    if success then update
end for
end while
```

Formal Definition

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 r_{1,i}^d(t)(P_i^d(t) - X_i^d(t)) + c_2 r_{2,i}^d(t)(P_g^d(t) - X_i^d(t))$$

$$X_i^d(t+1) = aX_i^d(t) + bV_i^d(t+1)$$

Formal Definition

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 r_{1,i}^d(t)(P_i^d(t) - X_i^d(t)) + c_2 r_{2,i}^d(t)(P_g^d(t) - X_i^d(t))$$

$$X_i^d(t+1) = aX_i^d(t) + bV_i^d(t+1)$$

t evolutionary step

i particle index

$V_i(t)$ particle's velocity

$X_i(t)$ particle's position

Formal Definition

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 r_{1,i}^d(t)(P_i^d(t) - X_i^d(t)) + c_2 r_{2,i}^d(t)(P_g^d(t) - X_i^d(t))$$

$$X_i^d(t+1) = aX_i^d(t) + bV_i^d(t+1)$$

t evolutionary step

i particle index

$V_i(t)$ particle's velocity

$X_i(t)$ particle's position

d particle dimension

P_i particle's best position

P_s swarm's best position

Formal Definition

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 r_{1,i}^d(t)(P_i^d(t) - X_i^d(t)) + c_2 r_{2,i}^d(t)(P_g^d(t) - X_i^d(t))$$

$$X_i^d(t+1) = aX_i^d(t) + bV_i^d(t+1)$$

t evolutionary step

i particle index

$V_i(t)$ particle's velocity

$X_i(t)$ particle's position

d particle dimension

P_i particle's best position

P_s swarm's best position

$r_{1,i}, r_{2,i}$ independent uniform random number in [0,1]

ω inertia weight

$\{\omega, c_1, c_2, a, b\}$ parameter tuple

Goal

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 r_{1,i}^d(t)(P_i^d(t) - X_i^d(t)) + c_2 r_{2,i}^d(t)(P_g^d(t) - X_i^d(t))$$

$$X_i^d(t+1) = aX_i^d(t) + bV_i^d(t+1)$$

find "good" $\{\omega, c_1, c_2, a, b\}$

$a = b = 1$

$c_1 = c_2 = \frac{(\omega+1)^2}{2}$

$\frac{2c_1}{\omega} = \frac{2c_2}{\omega} < 4$

Strategy

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 r_{1,i}^d(t)(P_i^d(t) - X_i^d(t)) + c_2 r_{2,i}^d(t)(P_g^d(t) - X_i^d(t))$$

$$X_i^d(t+1) = aX_i^d(t) + bV_i^d(t+1)$$

find "good" $\{\omega, c_1, c_2, a, b\}$

assume stagnation (no updates to P_i or P_s)

particle's position regarded as stochastical vector

expected value and variance

aim: find convergence criteria for stagnating particle swarm

aim: find convergence criteria for stagnating particle swarm

a single one-dimensional particle

$\Rightarrow P_i = P_s = P$

no inertia weight $\omega = 1$ & $a = b = 1$

no stochastic component $\phi_1 = c_1, \phi_2 = c_2$

$$V(t+1) = V(t) + \phi_1(P(t) - X(t)) + \phi_2(P(t) - X(t))$$

aim: find convergence criteria for stagnating particle swarm

a single one-dimensional particle

$\Rightarrow P_i = P_s = P$

no inertia weight $\omega = 1$ & $a = b = 1$

no stochastic component $\phi_1 = c_1, \phi_2 = c_2$

$$V(t+1) = V(t) + \phi_1(P(t) - X(t)) + \phi_2(P(t) - X(t))$$

generalized: multiple j-dimensional particles (Ozcan & Mohan)

$$X_{ij}(t) - (2 - \phi_1 - \phi_2)X_{ij}(t-1) + X_{ij}(t-2) = \phi_1 P_{ij} + \phi_2 P_{ij}$$

Clerc & Kennedy

use the same simple PSO but P_i and P_g

stable Point $P = \frac{c_1 P_i + c_2 P_g}{c_1 + c_2}$

convergence ensured by:

$$V_{ij}(t+1) = \chi[V_{ij}(t) + \phi_{1j}(t)(P_{ij}(t) - X_{ij}(t)) + \phi_{2j}(t)(P_{gj}(t) - X_{ij}(t)]$$

where

$$\chi = \frac{2\kappa}{2 - \phi - \sqrt{\phi^2 - 4\phi}} \text{ with } \phi = \phi_1 + \phi_2 \geq 4 \text{ and } \kappa \in [0,1]$$

κ small $\Leftrightarrow$ fast convergence to stable point

van den Bergh

uses the same simple PSO and inertia weight

$$X_{i+1} = (1 + \omega - \phi_1 - \phi_2)X_i - \omega X_{i-1} + \phi_1 P_i + \phi_2 P_g$$

characteristic equation corresponding to the recurrence relation

$$\lambda^2 - (1 + \omega - \phi_1 - \phi_2)\lambda + \omega = 0$$

eigenvalues

$$\lambda_{1/2} = \frac{1 + \omega - \phi_1 - \phi_2 \pm \sqrt{(1 + \omega - \phi_1 - \phi_2)^2 - 4\omega}}{2}$$

van den Bergh

new equation

$$X_t = k_1 + k_2\lambda_1^t + k_3\lambda_2^t$$

k_i can be calculated given $\phi_1, \phi_2, X, \lambda, \omega$

equation remains valid until a better position is discovered
can always be updated with new positions

van den Bergh

new equation

$$X_t = k_1 + k_2\lambda_1^t + k_3\lambda_2^t$$

k_i can be calculated given $\phi_1, \phi_2, X, \lambda, \omega$

equation remains valid until a better position is discovered
can always be updated with new positions

assuming $(1 + \omega - \phi_1 - \phi_2)^2 < 4\omega$ and $max\{||\lambda_1||, ||\lambda_2||\} < 1$

$$\Rightarrow \lim_{t \to +\infty} X_t = \frac{\phi_1 P_1 + \phi_2 P_s}{\phi_1 + \phi_2}$$

Summary

found criteria for convergence of non-stochastical PSO

Original Version

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 r_{1,i}^d(t)(P_i^d(t) - X_i^d(t)) + c_2 r_{2,i}^d(t)(P_g^d(t) - X_i^d(t))$$

$$X_i^d(t+1) = aX_i^d(t) + bV_i^d(t+1)$$

find useful $\{\omega, c_1, c_2, a, b\}$

Original Version

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 r_{1,i}^d(t)(P_i^d(t) - X_i^d(t)) + c_2 r_{2,i}^d(t)(P_g^d(t) - X_i^d(t))$$

$$X_i^d(t+1) = aX_i^d(t) + bV_i^d(t+1)$$

find useful $\{\omega, c_1, c_2, a, b\}$

assuming stagnation:

all particles evolve independently
$\Rightarrow$ only arbitrary particle i
dimensions update independently
$\Rightarrow$ only one-dimension

Original Version

$$V_i^d(t+1) = \omega V_i^d(t) + c_1 r_{1,i}^d(t)(P_i^d(t) - X_i^d(t)) + c_2 r_{2,i}^d(t)(P_g^d(t) - X_i^d(t))$$

$$X_i^d(t+1) = aX_i^d(t) + bV_i^d(t+1)$$

find useful $\{\omega, c_1, c_2, a, b\}$

assuming stagnation:

- all particles evolve independently
- $\Rightarrow$ only arbitrary particle i
- dimensions update independently
- $\Rightarrow$ only one-dimension

$$V_{t+1} = \omega V_t + c_1 r_{1,t}(P_i - X_t) + c_2 r_{2,t}(P_g - X_t)$$

$$X_{t+1} = aX_t + bV_{t+1}$$

$b = 1$

$$V_{t+1} = \omega V_t + c_1 r_1, t(P_i - X_i) + c_2 r_{2,t}(P_g - X_t)$$
$$X_{t+1} = aX_t + bV_{t+1}$$

$$X_{t+1} = (a + \omega - c_1 br_{1,t} - c_2 br_{2,t})X_t - a\omega X_{t-1} + c_1 br_{1,t}P_i + c_2 br_{2,t}P_g$$

only $c_1 b$ and $c_2 b$ relevant
without loss of generality $b = 0$ or $b = 1$

$b = 1$

let b=0 $\Rightarrow$

$$X_i^d(t+1) = a^{t+1} X_i^d(0)$$

trivial convergence criterium $a < 1$

$b = 1$

let b=0 $\Rightarrow$

$$X_i^d(t+1) = a^{t+1} X_i^d(0)$$

trivial convergence criterium $a < 1$

let b=1 $\Rightarrow$

$$X_{t+1} = (a + \omega - c_1 r_{1,t} - c_2 r_{2,t}) X_t - a\omega X_{t-1} + c_1 r_{1,t} P_i + c_2 r_{2,t} P_g$$

regard X_t as random variable and $\{X_t\}$ as stochastic process

Analysis of Expectation

$$X_{t+1} = (a + \omega - c_1 r_{1,t} - c_2 r_{2,t})X_t - a\omega X_{t-1} + c_1 r_{1,t} P_i + c_2 r_{2,t} P_g$$

calculate expectation EX_t of X_t

$$EX_{t+1} = (a + \omega - \frac{c_1 + c_2}{2})EX_t - a\omega EX_{t-1} + \frac{c_1 P_i + c_2 P_g}{2}$$

characteristic equation

$$\lambda^2 - (a + \omega - \frac{c_1 + c_2}{2})\lambda + a\omega$$

Analysis of Expectation

Theorem:
If and only if conditions $-1 < a\omega < 1$ and $2(1-\omega)(a-1) < c_1 + c_2 < 2(1+\omega)(1+a)$ are satisfied together then $\{EX_t\}$ is guaranteed to converge to $EX = \frac{c_1 P_i + c_2 P_g}{2(1-\omega)(1-a)+c_1+c_2}$

Analysis of Variance

find equation for DX_t and analyze convergent condition of $\{DX_t\}$

let: $\overline{\omega} = a\omega$ $\quad v = (c_1 + c_2)/2$ $\quad \psi = a + \omega - v$
$R_t = c_1 r_{1,t} + c_2 r_{2,t} - v$ $\quad Q_t = c_1 r_{1,t} P_i + c_2 r_{2,t} P_g$

$$X_{t+1} = (\psi - R_t)X_t - \overline{\omega}X_{t-1} + Q_t$$

Analysis of Variance

find equation for DX_t and analyze convergent condition of $\{DX_t\}$

let: $\overline{\omega} = a\omega$ $\quad v = (c_1 + c_2)/2$ $\quad \psi = a + \omega - v$

$R_t = c_1 r_{1,t} + c_2 r_{2,t} - v$ $\quad Q_t = c_1 r_{1,t} P_i + c_2 r_{2,t} P_g$

$$X_{t+1} = (\psi - R_t)X_t - \overline{\omega}X_{t-1} + Q_t$$

$$DX_{t+2} = (\psi^2 + DR_t - \overline{\omega})DX_{t+1} - \overline{\omega}(\psi^2 - DR_t - \overline{\omega})DX_t + \overline{\omega}^3 DX_{t-1} + R[(EX_{t+1})^2 + \overline{\omega}(EX_t)^2] - 2E(R_t Q_t)(EX_{t-1} + \overline{\omega}EX_t) + DQ_t(1 + \overline{\omega})$$

Analysis of Variance

characteristic equation

$$\chi(\lambda) = \lambda^3 - (\psi^2 + DR_t - \overline{\omega})\lambda^2 + \overline{\omega}(\psi^2 - DR_t - \overline{\omega})\lambda - \overline{\omega}^3$$

3 eigenvalues (assuming $c_1 \neq 0$ or $c_2 \neq 0$)

$$1 > \lambda_{maxD} = \lambda_{D1} > |\overline{\omega}| \geq \|\lambda_{D2}\| \geq \|\lambda_{D3}\|$$

$$\lambda_{maxD} > 0 \ \in \mathbb{R}$$

Analysis of Variance

Theorem:
Given $c_1 \neq 0$ or $c_2 \neq 0$, if and only if $-1 < \overline{\omega} < 1$ and $\chi(1) > 0$ are satisfied together,
$\lambda_{maxD} < 1$ and $\{DX_t\}$ is guaranteed to converge to

$$DX = \frac{\frac{1}{12}(1+\overline{\omega})}{\chi(1)[2(1-a)(1-\omega)+c_1+c_2]^2}\{2c_1^2c_2^2(P_g-P_i)^2+$$

$$4(a-1)(1-\omega)c_1c_2(c_2P_g-c_1P_i)(P_g-P_i)+(c_1^2P_i^2+c_2^2P_g^2)[2(1-a)(1-\omega)]^2\}$$

Analysis of Variance

Theorem:
Given $c_1 \neq 0$ or $c_2 \neq 0$, if and only if $-1 < \overline{\omega} < 1$ and $\chi(1) > 0$ are satisfied together,
$\lambda_{maxD} < 1$ and $\{DX_t\}$ is guaranteed to converge to

$$DX = \frac{\frac{1}{12}(1+\overline{\omega})}{\chi(1)[2(1-a)(1-\omega)+c_1+c_2]^2}\{2c_1^2c_2^2(P_g-P_i)^2+$$

$$4(a-1)(1-\omega)c_1c_2(c_2P_g-c_1P_i)(P_g-P_i)+(c_1^2P_i^2+c_2^2P_g^2)[2(1-a)(1-\omega)]^2\}$$

DX is generally not zero.

Summary

under certain conditions:

$$EX = \frac{c_1 P_i + c_2 P_g}{2(1-\omega)(1-a) + c_1 + c_2}$$

$$DX = \frac{\frac{1}{12}(1+\overline{\omega})}{\chi(1)[2(1-a)(1-\omega) + c_1 + c_2]^2}\{2c_1^2 c_2^2 (P_g - P_i)^2 +$$

$$4(a-1)(1-\omega)c_1 c_2 (c_2 P_g - c_1 P_i)(P_g - P_i) + (c_1^2 P_i^2 + c_2^2 P_g^2)[2(1-a)(1-\omega)]^2\}$$

Summary

- $(b = 1)$
- $-1 < a\omega < 1$
- $2(1-\omega)(a-1) < c_1 + c_2 < 2(1+\omega)(1+a)$
- $c_1 \neq 0$ or $c_2 \neq 0$
- $\chi(1) > 0$

important λ_{maxD}

Holland

balance between exploration and exploitation

- exploration: new regions
- exploitation: focused

Holland

balance between exploration and exploitation

- exploration: new regions
- exploitation: focused

for PSO: the larger λ_{maxD} the stronger exploration ability

Enhance local Search Ability

$(1-a)(1-\omega) = 0$ then $\exists i$ with $P_i = P_g$
converges to P_g with variance 0
$\Rightarrow a = 1$ without loss of generality

Enhance local Search Ability

$(1-a)(1-\omega) = 0$ then $\exists i$ with $P_i = P_g$
converges to P_g with variance 0
$\Rightarrow a = 1$ without loss of generality

three-dimensional parameter tuple $\{\omega, c_1, c_2\}$

$-1 < \omega < 1$

$0 < c_1 + c_2 < 4(1+\omega)$

$\chi(1) > 0$

Enhance local Search Ability

$(1-a)(1-\omega)=0$ then $\exists i$ with $P_i = P_g$
converges to P_g with variance 0
$\Rightarrow a = 1$ without loss of generality

three-dimensional parameter tuple $\{\omega, c_1, c_2\}$

$-1 < \omega < 1$

$0 < c_1 + c_2 < 4(1+\omega)$

$\chi(1) > 0$

$\Rightarrow EX = (c_1 P_i + c_2 P_g)/(c_1 + c_2)$ and
$DX = (c_1 c_2)^2 (P_g - P_i)^2 (1+\omega)/(c_1+c_2)^2/\chi(1)/6$

Parameter Selection

best search ability: $\lambda_{maxD} \approx 1$
criteria: $-1 < \omega < 1$ & $0 < c_1 + c_2 < 4(1+\omega)$ & $\chi(1) > 0$

Parameter Selection

best search ability: $\lambda_{maxD} \approx 1$
criteria: $-1 < \omega < 1$ & $0 < c_1 + c_2 < 4(1 + \omega)$ & $\chi(1) > 0$

	ω	c_1	c_2	λ_{maxD}
Clerc & Kennedy	0.729	1.494	1.494	0.942
Trelea	0.6	1.7	1.7	0.889
Carlisle & Dozier	0.729	2.041	0.948	0.975
Jiang & Luo & Yang	0.715	1.7	1.7	0.995

Benchmark Functions

function	formula	optimal position
sphere	$\Sigma_{i=1}^{d} x_i^2$	0
griewank	$\frac{1}{4000}\Sigma_{i=1}^{d} x_i^2 - \prod_{i=1}^{d} \cos(\frac{x_i}{\sqrt{i}}) + 1$	0
rastrigin	$\Sigma_{i=i}^{d}(x_i^2 - 10\cos(2\pi x_i) + 10)$	0
rosenbrock	$\Sigma_{i=i}^{d-1}(100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2)$	(1,...,1)

Visualization in Matlab

personal emergency guideline:

1. start Matlab as administrator
2. wait until command window is ready
3. type "Project" and press enter
4. velocity scale = 1

code from https://www.youtube.com/watch?v=bbbUwyMr1W8

Experimental Results (Jiang, Luo, Yang)

	ω	c_1	c_2	λ_{maxD}
Clerc & Kennedy	0.729	1.494	1.494	0.942
Carlisle & Dozier	0.729	2.041	0.948	0.975
Trelea	0.6	1.7	1.7	0.889
Jiang & Luo & Yang	0.715	1.7	1.7	0.995

dimensions: 10, 30, 50 particles: 20, 40, 100
each set run 100 times $\Rightarrow$ statistical results

Dimension 10

		Clerc	Carlisle	Trelea	Jiang
Succes Rate	Ackley	0.92	0.90	0.86	0.96
	Griewank	0.64	0.90	0.67	0.72
	Rastrigin	0.68	1.00	0.96	0.88
	Rosenbrock	1.00	0.99	1.00	1.00
	Sphere	1.00	1.00	1.00	1.00
Success Iteration	Ackley	84	72.5	63	108
	Griewank	265	192	160.5	426
	Rastrigin	179	161	141	201
	Rosenbrock	333.5	319.5	280.5	538.5
	Sphere	186	162	139	259

Dimension 30

		Clerc	Carlisle	Trelea	Jiang
Succes Rate	Ackley	0.82	0.78	0.70	1.00
	Griewank	0.91	0.86	0.90	0.98
	Rastrigin	0.61	0.94	0.59	0.94
	Rosenbrock	0.99	0.90	0.97	0.84
	Sphere	1.00	1.00	1.00	1.00
Success Iteration	Ackley	190	142.5	150.5	280
	Griewank	285	197	237	503.5
	Rastrigin	287.5	182.5	179.5	345.5
	Rosenbrock	1832	2051	1689	5323
	Sphere	435.5	295	363	772

Dimension 50

		Clerc	Carlisle	Trelea	Jiang
Succes Rate	Ackley	0.77	0.60	0.82	1.00
	Griewank	0.94	0.94	0.93	0.98
	Rastrigin	0.25	0.80	0.27	0.97
	Rosenbrock	0.80	0.60	0.91	0.61
	Sphere	1.00	1.00	1.00	1.00
Success Iteration	Ackley	257	214	214	471
	Griewank	383	214	338	787
	Rastrigin	∞	262	∞	597
	Rosenbrock	2188	3981	2205	5474
	Sphere	603	332	531	1281

Dimension Problem

$\lim_{dim \to \infty}$ "volume of circle" / "volume of square" $= 0$

Other Versions

- random numbers instead of random vectors
- several social groups
- several swarms
- boundary handling
- discrete version
- ...

PSO often does not find the "real" global maxima

only good on some function with given parameters

Sources

M. Jiang, Y. P. Luo and S. Y. Yang (2007). Particle Swarm Optimization - Stochastic Trajectory Analysis and Parameter Selection, Swarm Intelligence, Focus on Ant and Particle Swarm Optimization, FelixT.S.Chan and Manoj KumarTiwari (Ed.), ISBN: 978-3-902613-09-7, InTech, Available from: http:/ www.intechopen.com/books/swarm_intelligence_focus_on_ant _and_particle_swarm_optimization// _stochastic_trajectory _analysis_and_parameter_selection

S. Helwig (2010): Particle Swarms for Constrained Optimization, Partikelschwärme für Optimierungsprobleme mit Nebenbedingungen. Universität Erlangen-Nürnberg.