Help me find it!

Tim Blackwell
Goldsmiths
November 2010

Outline

1. PSO from above
2. Focus, spread and stability
3. Bare Bones
4. Study of collapse in BB

PSO from above

I've lost it.

It should be here, or at least somewhere close to here.

Can you help me? Could your friends help me as well?

How do we share information, and what do we do with it?

My current position x_i.

My best, p_i; my helpers bests, p_j; informer neighbourhood $\mathcal{N}_i$.

PSO as second order stochastic difference equation

for each particle $i = 1 \dots N$

for each dimension $d = 1 \dots D$

$$x_{t+1,id} = -a_t x_{t,id} - b_t x_{t-1,id} + c_t(\mathcal{N}_i)$$

end for

$\vec{p}_{t+1} = \text{BEST}(\vec{x}_{t+1}, \vec{p}_t)$

end for

Underlying assumption: BEST has some structure (nearer is better).

Examples: Clerc-Kennedy

The Clerc-Kennedy formulation has become the *de facto* standard PSO:

$$\begin{cases} a_t = -(1+w) + \frac{1}{2}(\Phi_1 + \Phi_2) \\ b_t = w \\ c_t(p) = \frac{1}{2}(\Phi_1 p_1 + \Phi_2 p_2) \end{cases}$$

where $\Phi \sim U[0, \phi]$, $p_1 = p_i$, and p_2 is the best informer in $\mathcal{N}_i$ (the same Φ_k appear in a and b).

It is written more conventionally as

$$x_{t+1} = x_t + w v_t + \frac{\Phi_1}{2}(p_1 - x_t) + \frac{\Phi_2}{2}(p_2 - x_t).$$

Examples: Discrete Recombinant

Peña's Discrete Recombinant PSO has an update rule:

$$\begin{cases} a_t = -(1+w) + \frac{1}{K}\sum_{k=1}^{K} \phi_k \\ b_t = w \\ c_t(p) = \frac{1}{K}\sum_{k=1}^{K} \phi_k \widehat{P}_k(p) \end{cases}$$

where ϕ_k are real constants and $\widehat{P}_k$ is a selection operator over K informers p.

Examples: Discrete Recombinant Model 3

Various other recombinant PSOs were studied by Bratton and Blackwell including a reduced version known as Model 3,

$$\begin{cases} a = -1 + \phi \\ b = 0 \\ c = \phi \; U\{p_1, p_2\} \end{cases}$$

which is a first order SDE (i.e. a particle update without velocity). Denoting the d'th component of the recombinant informer as r ($= p_{1d} \; or \; p_{2d}$), Model 3 is simply written as

$$x_{t+1} = x_t + \phi(r - x_t).$$

Examples: Bare Bones

Bare Bones PSO, originally formulated by Kennedy:

$$\begin{cases} a_t = 0 \\ b_t = 0 \\ c_t(p) = N(\mu(p), \sigma^2(p)) \end{cases}$$

where N is the Normal distribution.

In Kennedy's formulation, N has mean $\mu = \frac{p_1+p_2}{2}$ and variance $\sigma^2 = (p_1 - p_2)^2$.

Focus

$$\langle x_{t+1} \rangle + \langle a_t \rangle \langle x_t \rangle + \langle b_t \rangle \langle x_{t-1} \rangle = \langle c_t \rangle$$

where the random variables a, b, c are independent of x.

Order-1 stability condition is found by solving the homogeneous equation for $x_t = \lambda^t, |\lambda| < 1$. The conditions for real and imaginary roots within the unit circle are

$$-1 - \langle b \rangle < \ \langle a \rangle < 1 + \langle b \rangle$$
$$(real\ roots, a^2 > 4b)$$

and

$$\frac{\langle a \rangle^2}{4} \quad < \quad \langle b \rangle \quad < \quad 1$$
$$(imaginary\ roots, a^2 < 4b)$$

with fixed point

$$\langle x \rangle = \frac{\langle c \rangle}{1 + \langle a \rangle + \langle b \rangle}.$$

$\langle x \rangle$ is the mean position generated by iterating the SDE; it a focus of the search at fixed p.

Focus - examples

CK, DR:

$$\langle x \rangle = \frac{1}{K} \langle \sum P \rangle = \langle \bar{P} \rangle,$$

demonstrating that the search (at stagnation) focuses around the centroid of the neighbouring attractors $\bar{P}$ (CK) and around the expectation value of the centroid (DR).

BB:

$$\langle x \rangle = \langle N \rangle = \mu.$$

Spread

The variance in x is obtained from $\langle \delta x^2 \rangle = \langle (x - \langle x \rangle)^2 \rangle$.

$$\langle \delta x^2 \rangle = \frac{\langle d^2 \rangle}{1 - \langle a^2 \rangle - \langle b^2 \rangle + \left(\frac{2\langle ab \rangle \langle a \rangle}{1 + \langle b \rangle} \right)}.$$

This equation gives the standard deviation of the general PSO, when order-2 stable, in terms of averages over the random variables a, b and c.

Spread - examples

CK, DR:

$$\sqrt{\langle \delta x^2 \rangle} = \gamma \sqrt{\langle \delta \widehat{P}^2 \rangle}$$

where

$$\begin{aligned} \gamma &= \sqrt{\frac{\langle \sum \Phi_j \rangle^2}{C}} \\ C &= 2(1-w)\langle \sum \Phi \rangle \\ &\quad - \langle (\sum \Phi)^2 \rangle + \frac{2w}{1+w} \langle \sum \Phi \rangle^2. \end{aligned}$$

and $\delta \widehat{P}_j = \widehat{P}_j - \langle x \rangle$

$\sqrt{\langle \delta \widehat{P}^2 \rangle}$ is a measure of the spread of the informer group.

BB:

$$\langle \delta x^2 \rangle = \sigma^2$$

Spread - examples - summary

The standard deviations for CK, DR and BB follow a common form,

$$\sqrt{\langle \delta x^2 \rangle} = \alpha |p_1 - p_2|$$

with

$$\alpha_{CK} = 1.042$$
$$\alpha_{DR} = 0.612$$
$$\alpha_{BB-Kennedy} = 1.0$$

Stability

$$\begin{cases} 1+ < a > + < b > \neq 0 & \text{Order 1} \\ 1 - \langle a^2 \rangle - \langle b^2 \rangle + \left(\frac{2\langle ab \rangle \langle a \rangle}{1+\langle b \rangle} \right) > 0 & \text{Order 2.} \end{cases}$$

CK: $2K(1 - w^2) - \frac{7}{6}\phi + \frac{5}{6}w\phi \geq 0$

DR (model 3): $0 < \phi < 2$.

BB: Since $a = b = 0$, stability is immediately satisfied.

General Bare Bones

Bare bones is simplest PSO in the sense that $a = b = 0$.

Not a difference equation at all; unsuccessful trials x are ignored.

Kennedy: $\mu = \frac{p_1 + p_2}{2}$, $\sigma = |p_1 - p_2|$.

Hidden parameter α: $\sigma = \alpha\ |p_1 - p_2|$.

In general, mean and informer separation can be chosen from the neighbourhood informers:

$$x = \mu(p) + \alpha\delta(p)N(0, 1).$$

A problem - collapse

First (DR) and second order (CK) PSOs have stability conditions that help us chose parameters ϕ and w.

The bare bones swarm cannot become unstable, but it may *collapse*.

Collapse, which is undesirable, is to be contrasted to *convergence*.

In arbitrary precision arithmetic, convergence means that the swarm best informer, p_g, approaches, but does not reach, a limit point x^*.

Suppose the swarm is stable and the best informer g is approaching x^*.

The dimensionless variable $\bar{\sigma} = \frac{\sigma}{|g - x^*|}$ measures the standard deviation of the sampling distribution in units of the separation from the optimum.

There are two scenarios.

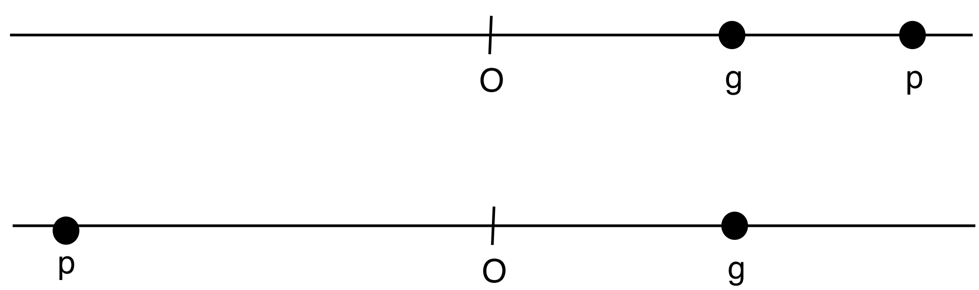
(1) $\bar{\sigma} \to 0$ with $\sigma \to 0$ faster than $g \to x^*$ and the swarm collapses and progress towards x^* slows until the swarm stagnates at a finite distance from x^*.

(2) $\bar{\sigma} \to \ const$ and the swarm converges on x^*.

This is the most desirable scenario; without the constraints of numerical precision, the g will become as close to x^* as we care to specify.

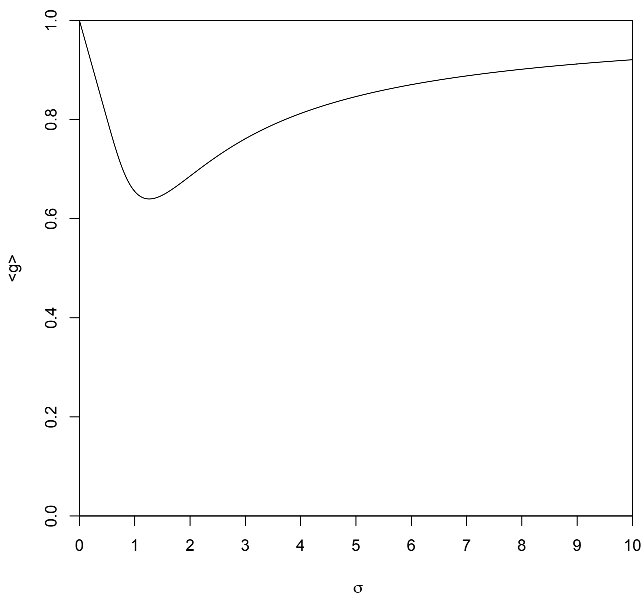
A consideration of collapse must, unlike the stability analysis mentioned above, consider informer movement.

Analysis of simple model with informer movement



Two possible configurations for a Bare Bones particle interacting with an effective particle. The effective particle represents the effects that $N-1$ particles have on the single particle. The informers are placed at g and p; either p or g can be regarded as the effective informer. The optimum is at O and g, which is closer to O, is the better informer.

$$\langle g \rangle = g + \int_{-g}^{g} (x - g)\rho_{g,\sigma^2} dx$$
$$= g - \frac{\sigma}{\sqrt{2\pi}}(1 - e^{-\frac{2g^2}{\sigma^2}}).$$



Expected value of g after a single update from $g = 1$, plotted as a function of standard deviation $\sigma = \alpha|\delta|$. The minimum of $\langle g \rangle$ is 0.64 at $\sigma = 1.26$.

$$\begin{aligned}\langle\delta\rangle &= \delta + \left(\int_{-|p|}^{-g} + \int_{g}^{|p|}\right)(x-p)\rho_{g,\sigma^2}dx \\ &- \int_{-g}^{g}(x-g)\rho_{g,\sigma^2}dx \\ &= \delta A + \sigma B\end{aligned}$$

where

$$\begin{aligned}A &= 1 - \int_a^b \rho_{0,1}dx - \int_0^c \rho_{0,1}dx \\ B &= \frac{1}{\sqrt{2\pi}}\left(2 + e^{-\frac{1}{2}a^2} - e^{-\frac{1}{2}c^2}\right) \\ &+ \frac{1}{\sqrt{2\pi}}\left(-2e^{-\frac{1}{2}b^2}\right)\end{aligned}$$

and

$$\begin{aligned}a &= \frac{-|p| - g}{\sigma} \\ b &= -\frac{2g}{\sigma} \\ c &= \frac{|p| - g}{\sigma}.\end{aligned}$$

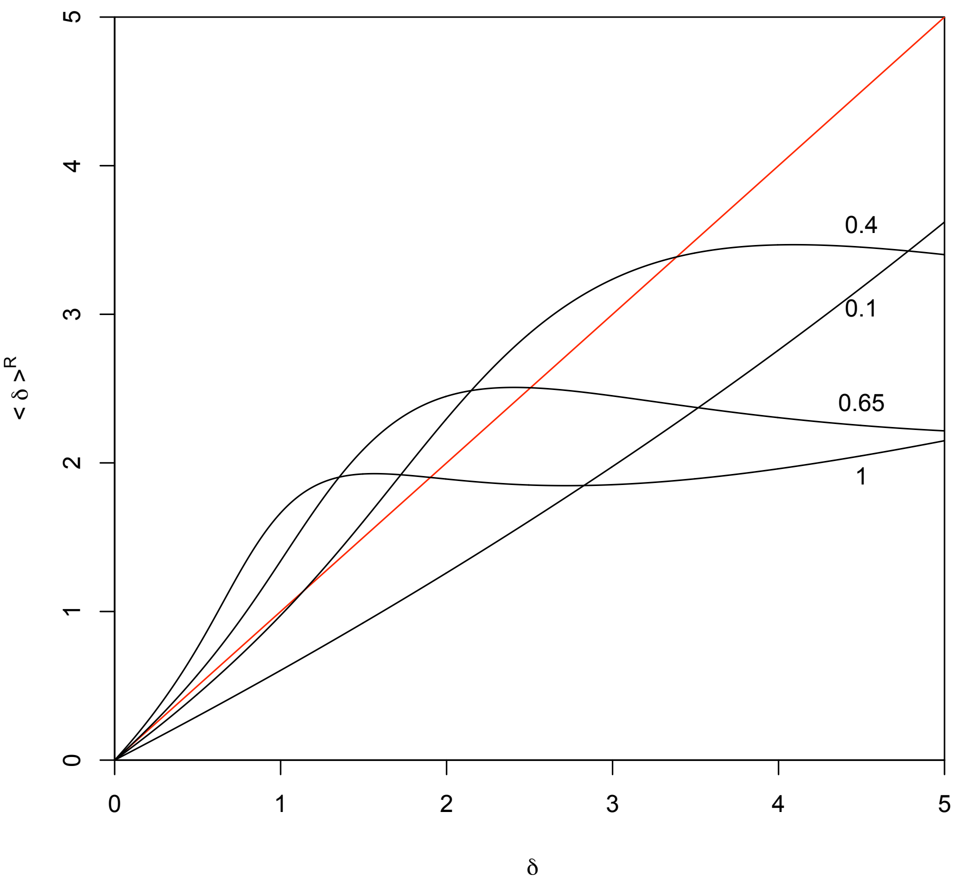
$$\begin{cases} \langle g \rangle_R = \frac{g}{\langle g \rangle} \langle g \rangle = 1 \\ \langle \delta \rangle_R = \frac{g}{\langle g \rangle} \langle \delta \rangle = \frac{\langle \delta \rangle}{\langle g \rangle}. \end{cases}$$

The rescaled system can be viewed as a dynamical system. Since $g = 1$, there is a single state $\delta_R \equiv \langle \delta \rangle^R(t)$ with dynamics

$\delta_R(t+1) = \frac{\langle \delta_R(t) \rangle}{\langle g \rangle} \equiv F(\delta_R(t))$

Self consistent condition (fixed points of F):

$$\langle \delta \rangle_R = \delta.$$

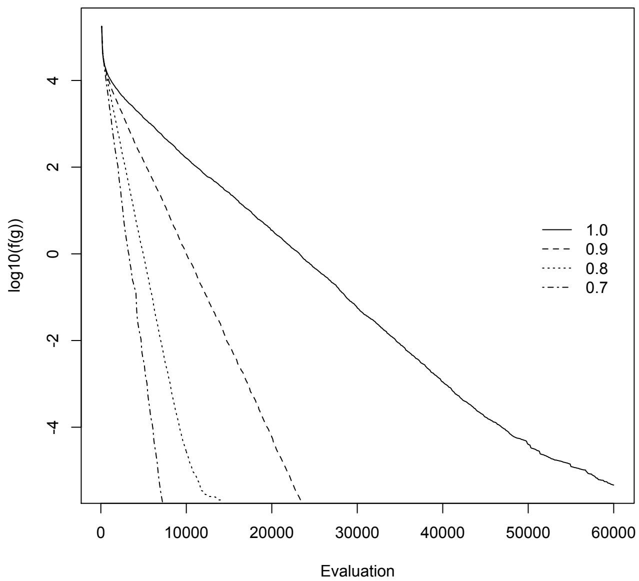


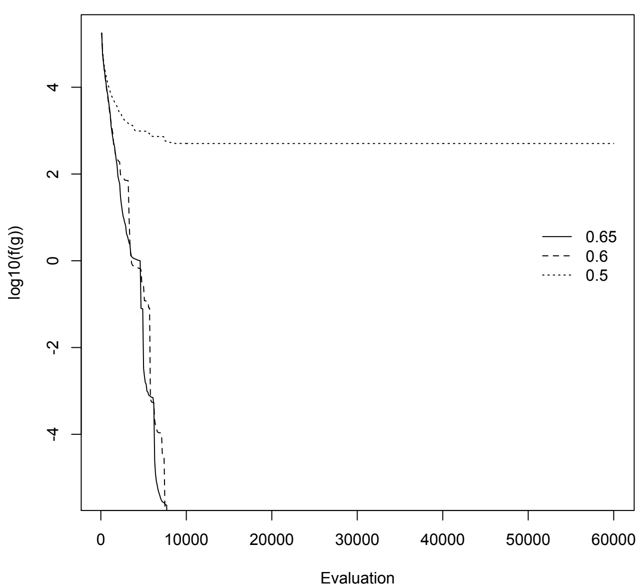
Expected value of δ after rescaling. The straight line is drawn at $<\delta>^R = \delta$.

$\alpha \geq 0.65$: There are two attractors, $\delta_b^* > 0$ and $\delta_-^* < -2$. Repeller at 0. States close to 0 are driven further away; the system resists collapse.

$\alpha < 0.65$: Attractor at 0. States close to 0 are driven towards 0 and the systems collapses.

Empirical test of $\alpha = 0.65$





Conclusions - BB

In tests over Yao et al and CEC2005 benchmarks, global and local focus BB at $\alpha = 0.65$ performs as well as PSO-CK and DR-Model 3 at their standard parameter setting.

All PSO's use information sharing to guide exploration. The focus and spread are determined by the dynamics, i.e. by the 2nd order SDE.

How important are the dynamics?

Second order SDE's with multiplicative stochasticity have bursts, but not first or zero order SDE's (Blackwell and Bratton). Bursts enable exploration of the whole search space at any stage. A simple jump mechanism improves BB performance in some cases.

The shape of the distribution itself may have a small effect, but since the distribution scales with the swarm, it does not allow distant exploration.

Conclusions - PSO

An understanding of the general properties of Particle Swarm Optimisation would help integrate the knowledge gained by the seemingly limitless exploration of new models. Search focus and spread, swarm stability and collapse and feasibility can be explored using simplified algorithms, and with the adoption of new techniques such as the mean field approximation, permitting analysis of informer, as well as of particle, movement. It is hoped that a theory of the particle swarm paradigm as a whole will emerge from these studies.