

# The Sampling Distribution of Particle Swarm Optimisers and their Stability

Riccardo Poli  
Department of Computer Science  
University of Essex, UK  
rpoli@essex.ac.uk

Department of Computer Science  
University of Essex  
Technical Report CSM-465  
ISSN: 1744-8050  
March 2007

## Abstract

Several theoretical analyses of the dynamics of particle swarms have been offered in the literature over the last decade. Virtually all rely on substantial simplifications, often including the assumption that the particles are deterministic. This has prevented the exact characterisation of the sampling distribution of the PSO. In this paper we introduce a novel method that allows us to exactly determine all the characteristics of a PSO's sampling distribution and explain how it changes over any number of generations, in the presence stochasticity. The only assumption we make is stagnation, i.e., we study the sampling distribution produced by particles in search for a better personal best. We apply the analysis to the PSO with inertia weight, but the analysis is also valid for the PSO with constriction and other forms of PSO.

## 1 Introduction

Let us consider the basic form of PSO with inertia weight shown in Algorithm 1. Despite its apparent simplicity, this PSO has presented formidable challenges to those interested in swarm intelligence theory. Firstly, the PSO is made up of a large number of interacting elements (the particles). Although the nature of the elements and of the interactions is simple, understanding the dynamics of the whole is non-trivial. Secondly, the particles are provided with memory and (albeit limited) intelligence, which mean that from one iteration to the next a particle may be attracted towards a new  $y_i$  or a new  $\hat{y}$  or both. Thirdly, forces are stochastic. This prevents the use of standard mathematical tools used in the analysis of deterministic dynamical systems. Fourthly, the behaviour of the PSO depends crucially on the structure of the fitness function. However, PSOs have been used on such a wide range of fitness functions that it is difficult to characterise a useful function space in which to study the role of the fitness function, and so it is hard to find general results. Nonetheless some progress has been made, by considering simplifying assumptions such as isolated single individuals, search stagnation (i.e., no improved solutions are found) and, crucially, *absence of randomness*.

For example, Ozcan and Mohan [3] studied the behaviour of one particle, in isolation, in one dimension, in the absence of stochasticity and during stagnation. Also,  $y$  and  $\hat{y}$  were assumed to coincide, as is the case for the best particle in a neighbourhood. The work was extended in [4] where multiple multi-dimensional particles were covered. Similar assumptions were used by Clerc and Kennedy's model [5]: one particle, one dimension, deterministic behaviour and stagnation. Under these conditions the swarm is a discrete-time linear dynamical system. The dynamics of the state (position and velocity) of a particle can be determined by finding the eigenvalues and eigenvectors of the state transition matrix. The model, therefore, predicts that the particle will converge to equilibrium if the magnitude of the eigenvalues is smaller than 1.

---

**Algorithm 1** Classical PSO.

---

- 1: Initialize a population array of particles with random positions and velocities on  $D$  dimensions in the problem space.
- 2: **loop**
- 3: For each particle, evaluate the desired optimization fitness function in  $D$  variables.
- 4: Compare particle's fitness evaluation with its personal best fitness  $pbest^i$ . If current value is better than  $pbest^i$ , then set  $pbest^i$  equal to the current value, and  $y^i$  equal to the current location  $x^i$  in  $D$ -dimensional space.
- 5: Identify the particle in the neighbourhood with the best success so far, and assign its position to the variable  $\hat{y}$ .
- 6: Change the velocity and position of the particle according to the following equations:

$$v_{t+1}^i = wv_t^i + \phi_1 \otimes (y^i - x_t^i) + \phi_2 \otimes (\hat{y} - x_t^i) \quad (1)$$

$$x_{t+1}^i = x_t^i + v_{t+1}^i \quad (2)$$

- 7: If a criterion is met, exit loop.

8: **end loop**

Note:  $\phi_i$  represents a vector of random numbers uniformly distributed in  $[0, c_i]$  and  $\otimes$  is component-wise multiplication.

---

A similar approach was used by van den Bergh [6] (see also [2]), who, again, modelled one particle, with no randomness and during stagnation. As in previous work, van den Bergh provided an explicit solution for the trajectory of the particle. He showed that the particle is attracted towards a fixed point. He also argued that the analysis would be valid also in the presence of stochasticity. [6] also suggested the possibility that particles may converge on a point that is neither the global optimum nor indeed a local optimum. This implies that a PSO is not guaranteed to be an optimiser.

A simplified model of particle was also studied by Yasuda *et al.* [7]. The assumptions were: one one-dimensional particle, stagnation and absence of stochasticity. Inertia was included in the model. Again an eigenvalue analysis of the resulting dynamical system was performed with the aim of determining for what parameter settings the system is stable and what classes of behaviours are possible for a particle. Conditions for cyclic behaviour were analysed in detail.

Blackwell [20] investigated how the spatial extent of a particle swarm varies over time. A simplified swarm model was adopted which is an extension of the one by Clerc and Kennedy where more than one particle and more than one dimensions are allowed. This allowed particles to interact, in the sense that they could change their personal best. Constriction was included but not stochasticity. [20] suggested that spatial extent decreases exponentially with time.

Brandstätter and Baumgartner [10] drew an analogy between Clerc and Kennedy's model [5] and a damped mass-spring oscillator, making it possible to rewrite the model using the notions of damping factor and natural vibrational frequency. Like the original model, this model assumes one particle, one dimension, no randomness and stagnation.

Under the same assumptions as [5] and following a similar approach, Trelea [11] performed a lucid analysis of a 4-parameter family of particle models and identified regions in the parameter space where the model exhibits qualitatively different behaviours (either stability, harmonic oscillations or zigzagging behaviour).

The dynamical system approach proposed by Clerc and Kennedy has recently been extended by Campana *et al.* [12, 13] who studied an extended PSO. Under the assumption that no randomness is present, the resulting model is a discrete, linear and stationary dynamical system, for which [12, 13] formally expressed the free and forced responses. However, since the forced response depends inextricably on the specific details of the fitness function, they were able to study in detail only the free response.

To better understand the behaviour of the PSO during phases of stagnation, Clerc [16] analysed the distribution of velocities of one particle controlled by the standard PSO update rule with inertia and *stochastic forces*. In particular, he was able to show that a particle's new velocity is the sum of three components: a forward force, a backward force and noise. Clerc studied the distributions of these forces.

Kadirkamanathan *et al.* [17] were able to study the stability of particles *in the presence of stochasticity* by using Lyapunov stability analysis. They considered the behaviour of a single particle – the swarm best – with inertia and during stagnation. By representing the particle as a non-linear feedback system, they were

able to apply a large body of knowledge from control theory. E.g., they found sufficient conditions on the PSO parameters to guarantee convergence. Since Lyapunov theory is very conservative, the conditions found are very restrictive, effectively forcing the PSO to have little oscillatory behaviour.

In summary, with very few exceptions all mathematical models of PSO behaviour have been obtained under rather unrealistic assumptions. In particular, very little is known regarding how the sampling distribution of particles changes over time. In this paper we introduce a novel method, which allows one to exactly determine the moments of a PSO's sampling distribution and explain how they change over any number of generations. The only assumption we make is stagnation, i.e., we study the sampling distribution produced by particles in search for a better personal best.

To start with, we will apply the analysis to the PSO with inertia weight (Algorithm 1). However, we should note that a PSO with constriction (see [5]) is algebraically equivalent to a PSO with inertia. Indeed, in this PSO, particles are controlled by the equation

$$v_{t+1}^i = \chi \left( v_t^i + \tilde{\phi}_1 \otimes (y^i - x_t^i) + \tilde{\phi}_2 \otimes (\hat{y} - x_t^i) \right) \quad (3)$$

which can be transformed into Equation (1) via the mapping  $\chi \rightarrow w$  and  $\chi\tilde{\phi}_i \rightarrow \phi_i$ . So, the theory applies to the PSO with constriction as well.

The paper is organised as follows. In Section 2 we derive recursions for the dynamics of first and second order statistics of the sampling distribution of a PSO's particle during stagnation. We study the fixed-points for these quantities and the PSO's stability in Section 3. In Section 4 we show the results of numerically integrating the dynamic equations for the distribution's statistics. We show how moments of higher order can be studied in Section 6. We explicitly study the stability of order-3 and order-4 moments in Section 7. We extend the model to include a variety of other forms of PSO, and we explicitly compare three of such forms – a purely social PSO, the canonical PSO and a fully informed PSO [14, 15] – in Section 8. We show how we can approximately reconstruct the sampling distribution of PSOs during stagnation from their moments in Section 9. Finally, we provide some discussion, indications for future work and our conclusions in Section 10.

## 2 Dynamics of first and second moments of the PSO sampling distribution

If the PSO is in a stagnation phase (i.e., there are no fitness improvements), each particle effectively behaves independently. Also, each dimension is treated independently. So, we can analyse each particle's behaviour in isolation. Dropping the superscript  $i$  in Equations (1) and (2), we can rewrite them as a single (second order) difference equations, as was done by other researchers (e.g., in [6]), by making use of the relation  $v_t = x_t - x_{t-1}$ . We obtain

$$x_{t+1} = x_t(1 + w) - x_t(\phi_1 + \phi_2) - wx_{t-1} + \phi_1 y + \phi_2 \hat{y}. \quad (4)$$

### 2.1 Dynamics of $E[x_t]$

Unlike previous research, we will not make the simplifying assumption that  $\phi_1$  and  $\phi_2$  are constant in Equation (4). Instead, we treat them for what they are, i.e., uniformly distributed stochastic variables, and we apply the expectation operator to both sides of the equation obtaining

$$E[x_{t+1}] = E[x_t](1 + w) - E[x_t](E[\phi_1] + E[\phi_2]) - wE[x_{t-1}] + E[\phi_1]y + E[\phi_2]\hat{y} \quad (5)$$

where we performed the substitution  $E[x_t\phi_i] = E[x_t]E[\phi_i]$  because of the statistical independence between  $\phi_i$  and  $x_t$ . Because  $\phi_i$  is uniformly distributed in  $[0, c_i]$  we have

$$E[\phi_1] = \frac{c_1}{2} \quad E[\phi_2] = \frac{c_2}{2} \quad (6)$$

and, so,

$$E[x_{t+1}] = E[x_t] \left( 1 + w - \frac{c_1 + c_2}{2} \right) - wE[x_{t-1}] + \frac{c_1}{2}y + \frac{c_2}{2}\hat{y} \quad (7)$$

Let  $p$  be a fixed point for this equation. This requires

$$p = \frac{c_1 y + c_2 \hat{y}}{c_1 + c_2} \quad (8)$$

For the sake of simplicity let us now restrict our attention to the case  $c_1 = c_2 = c$  (we will remove this restriction in Section 8). Furthermore, let us rename  $(1 + w) = w'$ . So

$$x_{t+1} = x_t w' - x_t \phi_1 - x_t \phi_2 - w x_{t-1} + \phi_1 y + \phi_2 \hat{y} \quad (9)$$

and

$$E[x_{t+1}] = E[x_t](w' - c) - wE[x_{t-1}] + c \frac{y + \hat{y}}{2} \quad (10)$$

Naturally, the stability of this equation is determined by the magnitude of the roots of the associated characteristic polynomial, or of the eigenvalues of the associated first-order vectorial difference equation. Figure 1 plots the magnitude of the largest eigenvalue of the equation for  $c = 0.01, 0.02, \dots, 4.00$  and  $w = 0.01, 0.02, \dots, 1.0$ . The straight line on the surface limits the stable region. We will say that a PSO for which  $E[x_t]$  has a stable fixed-point is *order-1 stable*.

Note that if we assumed that  $\phi_1$  and  $\phi_2$  are constant and equal to their maximum value,  $c$ , Equation (9) would become

$$x_{t+1} = x_t(w' - 2c) - w x_{t-1} + c(y + \hat{y}) \quad (11)$$

This equation has been studied extensively in previous research and has exactly the same form as Equation (10), except that here we have  $2c$  instead of  $c$  and the magnitude of the forcing term,  $c(y + \hat{y})$ , is doubled. So, the stability of Equation (10) has effectively been studied in previous research (e.g., [11], [6] and [5]; see also [2] for an extensive review). Indeed, the stable region depicted in Figure 1 is exactly the same as reported in [11, Figure 1(a)], and the explicit dynamics of  $E[x_t]$  is explicitly given in previous work (e.g., [6]) if parameters are appropriately rescaled.

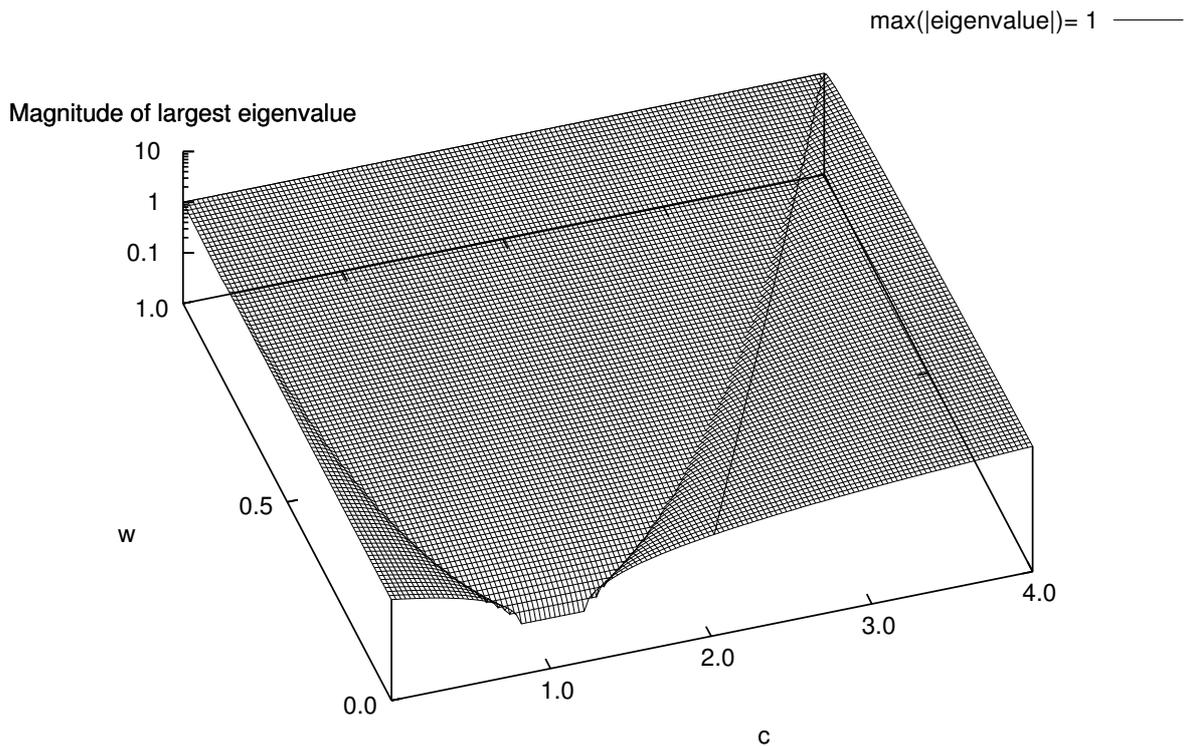


Figure 1: Stability analysis based on the difference equation for  $E[x_t]$  as a function of the parameters  $w$  and  $c$ . The straight line on the surface bounds the order-1 stability region.

## 2.2 Dynamics of $E[x_t^2]$ , $E[x_t x_{t-1}]$ and $StdDev[x_t]$

Let us now compute  $x_{t+1}^2$ . From Equation (9) we obtain:

$$\begin{aligned}
x_{t+1}^2 &= (x_t w' - x_t \phi_1 - x_t \phi_2 - w x_{t-1} + \phi_1 y + \phi_2 \hat{y})^2 \\
&= x_t^2 w'^2 - x_t^2 \phi_1 w' - x_t^2 \phi_2 w' - w x_{t-1} x_t w' + \phi_1 y x_t w' \\
&+ \phi_2 \hat{y} x_t w' - x_t^2 w' \phi_1 + x_t^2 \phi_1^2 + x_t^2 \phi_2 \phi_1 + w x_{t-1} x_t \phi_1 \\
&- \phi_1^2 y x_t - \phi_2 \hat{y} x_t \phi_1 - x_t^2 w' \phi_2 + x_t^2 \phi_1 \phi_2 + x_t^2 \phi_2^2 \\
&+ w x_{t-1} x_t \phi_2 - \phi_1 y x_t \phi_2 - \phi_2^2 \hat{y} x_t - x_t w' w x_{t-1} \\
&+ x_t \phi_1 w x_{t-1} + x_t \phi_2 w x_{t-1} + w^2 x_{t-1}^2 - \phi_1 y w x_{t-1} \\
&- \phi_2 \hat{y} w x_{t-1} + x_t w' \phi_1 y - x_t \phi_1^2 y - x_t \phi_2 \phi_1 y - w x_{t-1} \phi_1 y \\
&+ \phi_1^2 y^2 + \phi_2 \hat{y} y \phi_1 + x_t w' \phi_2 \hat{y} - x_t \phi_1 \phi_2 \hat{y} - x_t \phi_2^2 \hat{y} \\
&- w x_{t-1} \phi_2 \hat{y} + \phi_1 y \phi_2 \hat{y} + \phi_2^2 \hat{y}^2
\end{aligned} \tag{12}$$

Again we apply the expectation operator to both sides of the equation, obtaining

$$\begin{aligned}
E[x_{t+1}^2] &= E[x_t^2] (w'^2 - 4\mu w' + 2\nu + 2\mu^2) \\
&+ E[x_{t-1} x_t] (-2w w' + 4w\mu) \\
&+ E[x_{t-1}^2] (w^2) \\
&+ E[x_t] (2\mu y w' + 2\mu \hat{y} w' - 2\nu y - 2\mu^2 \hat{y} - 2\mu^2 y - 2\nu \hat{y}) \\
&+ E[x_{t-1}] (-2\mu y w - 2\mu \hat{y} w) \\
&+ \nu y^2 + 2\mu^2 y \hat{y} + \nu \hat{y}^2
\end{aligned} \tag{13}$$

where we set  $\mu = E[\phi_i] = c/2$  and  $\nu = E[\phi_i^2] = c^2/3$ , for brevity.

As we discussed in Section 2.1, we have a recursion (and in fact an explicit solution) for  $E[x_t]$ , so the recursion in Equation (19) could be solved if we had a recursion for  $E[x_t x_{t-1}]$ . Let us obtain such a recursion.

We multiply both sides of Equation (9) by  $x_t$ , obtaining

$$x_{t+1} x_t = x_t^2 w' - x_t^2 (\phi_1 + \phi_2) - w x_t x_{t-1} + x_t \phi_1 y + x_t \phi_2 \hat{y} \tag{14}$$

thereby

$$E[x_{t+1} x_t] = E[x_t^2] (w' - c) - w E[x_t x_{t-1}] + E[x_t] c \frac{y + \hat{y}}{2}. \tag{15}$$

With this additional equation we are now in a position to determine the dynamics of  $E[x_t^2]$  and  $E[x_t x_{t-1}]$ , in addition to the dynamics of  $E[x_t]$  we derived in Section 2.1.

The recursions for  $E[x_t]$ ,  $E[x_t^2]$  and  $E[x_t x_{t-1}]$  form the following set of coupled difference equations

$$\left\{ \begin{array}{l}
E[x_{t+1}] = E[x_t] (w' - c) - w E[x_{t-1}] + c \frac{y + \hat{y}}{2} \\
E[x_{t+1}^2] = E[x_t^2] (w'^2 - 4\mu w' + 2\nu + 2\mu^2) + \\
\quad E[x_{t-1} x_t] (-2w w' + 4w\mu) + \\
\quad E[x_{t-1}^2] (w^2) + \\
\quad 2E[x_t] (y + \hat{y}) (\mu w' - \nu - \mu^2 y) - \\
\quad 2w\mu E[x_{t-1}] (y + \hat{y}) + \nu y^2 + 2\mu^2 y \hat{y} + \nu \hat{y}^2 \\
E[x_{t+1} x_t] = E[x_t^2] (w' - c) - w E[x_t x_{t-1}] + E[x_t] c \frac{y + \hat{y}}{2}
\end{array} \right. \tag{16}$$

These can be integrated either symbolically or numerically. By using the relation

$$StdDev[x_t] = \sqrt{E[x_t^2] - (E[x_t])^2} \tag{17}$$

one can also derive the dynamics for the standard deviation of the sampling distribution of a PSO during stagnation.

### 2.3 Initial conditions

Let us evaluate the initial conditions for Equations (16). To do so, we must specify how we perform the initialisation of the particle swarm. As an example, let us consider the following very typical conditions: a) a particle's initial position,  $x_0$ , is chosen uniformly at random in a symmetric range  $[-\Omega, \Omega]$ , b) a particle's initial velocity,  $v_0$ , is also chosen uniformly at random in the same range.

In these conditions, clearly,  $E[x_0] = 0$  and  $E[v_0] = 0$ . So,  $E[x_1] = E[x_0 + v_1] = E[x_0] + E[v_1] = E[v_1]$ . Let us compute  $E[v_1]$ . We have that

$$\begin{aligned} E[v_1] &= E[wv_0 + \phi_1(y - x_0) + \phi_2(\hat{y} - x_0)] \\ &= wE[v_0] + E[\phi_1](y - E[x_0]) + E[\phi_2](\hat{y} - E[x_0]) \\ &= c \frac{y + \hat{y}}{2}. \end{aligned}$$

So,  $E[x_1] = c \frac{y + \hat{y}}{2}$ . We also have that  $E[x_0^2] = E[v_0^2] = \frac{\Omega^2}{3}$ , while  $E[x_1x_0] = E[(x_0 + v_1)x_0] = \frac{\Omega^2}{3} + E[v_1x_0]$ , the second term of which is given by

$$\begin{aligned} E[v_1x_0] &= E[wv_0x_0 + \phi_1(yx_0 - x_0^2) + \phi_2(\hat{y}x_0 - x_0^2)] \\ &= wE[v_0]E[x_0] + E[\phi_1](yE[x_0] - E[x_0^2]) + E[\phi_2](\hat{y}E[x_0] - E[x_0^2]) \\ &= -cE[x_0^2] \\ &= -c \frac{\Omega^2}{3}, \end{aligned}$$

resulting in  $E[x_1x_0] = (1 - c) \frac{\Omega^2}{3}$ .

The only remaining initial condition we need is  $E[x_1^2] = E[(x_0 + v_1)^2] = E[x_0^2] + 2E[v_1x_0] + E[v_1^2] = (1 - 2c) \frac{\Omega^2}{3} + E[v_1^2]$ , which, after similar additional calculations leads to  $E[x_1^2] = \frac{7c^2 - 12c + 6w^2 + 6}{18} \Omega^2 + c^2 \frac{(y + \hat{y})^2}{3}$ .

### 3 Order-1 and -2 stability analysis for particles with randomness

The system of equations (16) can be written in matrix notation as the extended first order system

$$\mathbf{z}(t + 1) = M\mathbf{z}(t) + \mathbf{b} \quad (18)$$

where

$$\mathbf{z}(t) = (E[x_t] \ E[x_{t-1}] \ E[x_t^2] \ E[x_t x_{t-1}] \ E[x_{t-1}^2])^T$$

and

$$M = \begin{pmatrix} w' - c & -w & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 4p(\mu w' - \nu - \mu^2) & -4\mu w p & w'^2 - 4\mu w' + 2\nu + 2\mu^2 & 2w(2\mu - w') & w^2 \\ cp & 0 & w' - c & -w & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} cp \\ 0 \\ \nu y^2 + 2\mu^2 y \hat{y} + \nu \hat{y}^2 \\ 0 \\ 0 \end{pmatrix}$$

It is then trivial to verify under what conditions  $E[x_t]$ ,  $E[x_t^2]$  and  $E[x_t x_{t-1}]$  (thereby also  $StdDev[x_t]$ ) will converge to stable fixed-points. We need to have that all eigenvalues of  $M$  must be within the unit circle, i.e.  $\Lambda_m = \max_i |\lambda_i| < 1$ . When this happens we will say that the PSO is *order-2 stable*.

The analysis of the stability of the system can be done easily. Any good computer algebra system can provide the eigenvalues of  $M$  in symbolic form. Two of them are simply:

$$\frac{1 + w - c \pm \sqrt{(w - c)^2 - 2c - 2w + 1}}{2}$$

The expressions for remaining three, however, are too big to report in this paper. The analysis reveals that none of the eigenvalues depends on either  $y$  or  $\hat{y}$  (nor  $p$ ). That is, whether or not the system is order-2 stable does not depend on where personal best and swarm best are located in the search space.

Naturally, when  $\Lambda_m < 1$ , in principle we could symbolically derive the fixed-point for the system, which we will denote as  $\mathbf{z}^*$ . This would be simply given by

$$\mathbf{z}^* = (I - M)^{-1} \mathbf{b}$$

For simplicity, below we will find explicit expressions for some components of  $\mathbf{z}^*$  by other means.

When the system is order-2 stable, by the simple change of variables  $\tilde{\mathbf{z}}(t) = \mathbf{z}(t) - \mathbf{z}^*$  can then represent the dynamics of the system via following linear homogeneous equation

$$\tilde{\mathbf{z}}(t+1) = M\tilde{\mathbf{z}}(t)$$

which can trivially be integrated to obtain the explicit solution

$$\tilde{\mathbf{z}}(t) = M^t\tilde{\mathbf{z}}(0).$$

Naturally, all these operations can be performed numerically once  $c$  and  $w$  are fixed. For example, Figure 2 shows a plot of  $\Lambda_m$  as a function of  $c$  and  $w$  for  $y = -1$  and  $\hat{y} = 1$ . The plot also shows a line where  $\Lambda_m = 1$ . As we explained earlier, although in order to compute  $M$  we have to specify  $y$  and  $\hat{y}$  as well as  $c$  and  $w$ ,  $\Lambda_m$  is not affected by what values  $y$  and  $\hat{y}$  have. So, one obtains exactly the same plot, for example, for  $y = 9$  and  $\hat{y} = 10$  (same distance as  $y = -1$  and  $\hat{y} = 1$ , but different  $p$ ) or  $y = -10$  and  $\hat{y} = 10$  (different distance, but same  $p$  as for  $y = -1$  and  $\hat{y} = 1$ ).

Naturally, knowing the region where the system is order-2 stable allows one to perform an informed choice of the parameters of the PSO. We should note that in this respect the region of order-1 stability provided by the analysis of the  $E[x_t]$  alone, as it has effectively been done in previous research, does not provide enough information to guarantee convergence of the particles. It only guarantees convergence of the mean. Compare, for example Figure 2 with Figure 1. Note how the actual region of order-2 stability shown in Figure 2 lays completely inside the region of order-1 stability obtained by analysing  $E[x_t]$  only (Figure 1). Interestingly, by choosing parameters between the two curves, one obtains PSOs where  $E[x_t] \rightarrow p$ , but  $StdDev[x_t]$  drifts (perhaps slowly) to infinity, which might be a desirable property if one wants PSOs capable of escaping from local optima. Note also, that choosing parameters  $c$  and  $w$  within the region of convergence *does not* imply that  $StdDev[x_t] \rightarrow 0$ . In the following we clarify when this is the case.

Simple inspection of the equations in Equation (16), reveals that the dynamics of  $E[x_t]$  is independent from those of  $E[x_t^2]$  and  $E[x_t x_{t-1}]$ , while the converse is not true. This means that  $E[x_t^2]$  and  $E[x_t x_{t-1}]$  cannot be at a fixed-point unless also  $E[x_t]$  is. Let us assume that  $(c, w)$  is in the region of convergence for  $E[x_t]$ . Then, for sufficiently large  $t$ ,  $E[x_t]$  becomes almost indistinguishable from the fixed-point  $p = \frac{w+\hat{y}}{2}$ . In these conditions, for the purpose of finding fixed-points for  $E[x_t^2]$  and  $E[x_t x_{t-1}]$ , we can replace  $E[x_t]$  and  $E[x_{t-1}]$  with  $p$  in the second and third equations of Equation (16), obtaining

$$E[x_{t+1}^2] = E[x_t^2] (w'^2 - 4\mu w' + 2\nu + 2\mu^2) + E[x_{t-1}x_t] (-2ww' + 4w\mu) + E[x_{t-1}^2]w^2 + 4p^2 (\mu - \nu - \mu^2) + \nu y^2 + 2\mu^2 y \hat{y} + \nu \hat{y}^2 \quad (19)$$

$$E[x_{t+1}x_t] = E[x_t^2](w' - c) - wE[x_t x_{t-1}] + p^2 c \quad (20)$$

We know that if  $(c, w)$  are additionally within the convergence region for the system, shown in Figure 2, then also  $E[x_t^2]$  and  $E[x_{t-1}x_t]$  will tend to a fixed-point. Let us find such fixed points. To do so we will assume we are at those fixed points, which we call  $p_{x^2}$  and  $p_{xx}$ , respectively. We substitute these into Equations (19) and (20) to obtain

$$p_{x^2} = p_{x^2} (w'^2 - 4\mu w' + 2\nu + 2\mu^2) + p_{xx} (-2ww' + 4w\mu) + p_{x^2} w^2 + 4p^2 (\mu - \nu - \mu^2) + \nu y^2 + 2\mu^2 y \hat{y} + \nu \hat{y}^2 \quad (21)$$

$$p_{xx} = p_{x^2}(w' - c) - wp_{xx} + p^2 c \quad (22)$$

The second equation allows us to compute

$$p_{xx} = p_{x^2} \left(1 - \frac{c}{w'}\right) + p^2 \frac{c}{w'} \quad (23)$$

Substitution of this in the first equation in (19) gives the following fixed-points:

$$p_{x^2} = \frac{4p^2 (\mu - \nu - \mu^2) + \nu y^2 + 2\mu^2 y \hat{y} + \nu \hat{y}^2 + p^2 \frac{c}{w'} 2w (2\mu - w')}{\Delta} \quad (24)$$

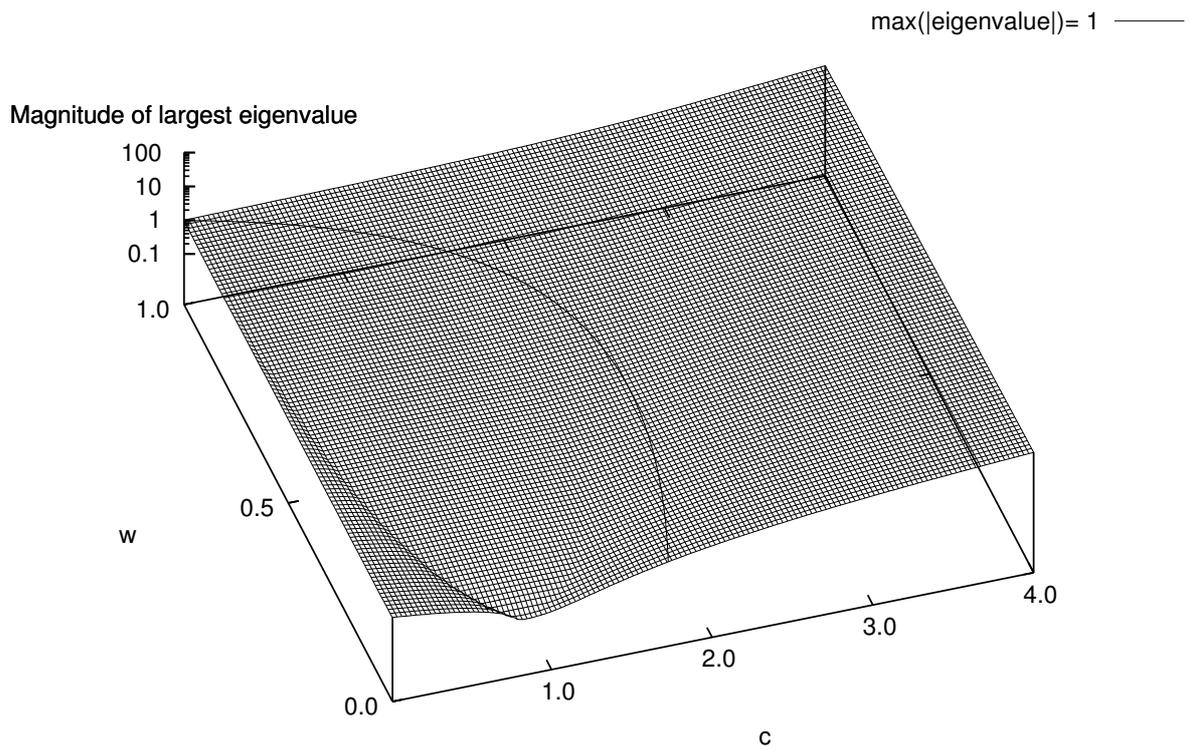


Figure 2: Magnitude of the largest eigenvalue of  $M$  as a function of the parameters  $w$  and  $c$  when  $y = -1$  and  $\hat{y} = 1$ . The curved line on the surface encloses the order-2 stable region.

$$p_{xx} = \frac{4p^2(\mu - \nu - \mu^2) + \nu y^2 + 2\mu^2 y \hat{y} + \nu \hat{y}^2 + p^2 \frac{c}{w'} 2w(2\mu - w')}{\Delta} \left(1 - \frac{c}{w'}\right) + p^2 \frac{c}{w'} \quad (25)$$

where

$$\Delta = 1 - \left[ (w'^2 - 4\mu w' + 2\nu + 2\mu^2) + \left(1 - \frac{c}{w'}\right) 2w(2\mu - w') + w^2 \right] \quad (26)$$

$$= c \times \frac{c \times (5w - 7) + 12(1 - w^2)}{6(w + 1)}. \quad (27)$$

In order for a particle to converge, i.e.,  $\lim_{t \rightarrow \infty} x_t = p$ , it is not enough to have  $\lim_{t \rightarrow \infty} E[x_t] = p$ : we must also have  $\lim_{t \rightarrow \infty} StdDev[x_t] = 0$ . This in turn requires  $\lim_{t \rightarrow \infty} E[x_t^2] = p^2$ . That is, we require  $p_{x^2} = p^2$ . To see when this can be the case, let us analyse Equation (24) in more detail.

With little algebra one can see that

$$p_{x^2} = \left( \frac{\Delta - 2(\mu^2 - \nu)}{\Delta} \right) p^2 + \left( \frac{4y(\mu^2 - \nu)}{\Delta} \right) p + \left( \frac{2y^2(\nu - \mu^2)}{\Delta} \right) \quad (28)$$

where we used the substitution  $\hat{y} = 2p - y$ .

So, in general  $p_{x^2} \neq p^2$  except if  $y = p$ , i.e.,  $\hat{y} = y$ . Then  $p_{x^2} = p^2$ . So, except for the best particle in the swarm, the standard deviation of the sampling distribution,  $StdDev[x_t]$ , does not converge to 0, but to

$$\begin{aligned} p_{sd} &= \sqrt{\left( \frac{-2(\mu^2 - \nu)}{\Delta} \right) p^2 + \left( \frac{4y(\mu^2 - \nu)}{\Delta} \right) p + \left( \frac{2y^2(\nu - \mu^2)}{\Delta} \right)} \\ &= \sqrt{2 \frac{(\nu - \mu^2)}{\Delta} (p^2 - 2yp + y^2)} \\ &= \sqrt{2 \frac{(\nu - \mu^2)}{\Delta} (p - y)^2} \\ &= \sqrt{2 \frac{(\nu - \mu^2)}{\Delta}} \cdot |p - y| \end{aligned}$$

which we can finally rewrite as

$$p_{sd} = \frac{1}{2} \sqrt{2 \frac{(\nu - \mu^2)}{\Delta}} \cdot |\hat{y} - y| \quad (29)$$

$$= \frac{1}{2} \sqrt{\frac{c \times (w + 1)}{c \times (5w - 7) - 12w^2 + 12}} \cdot |\hat{y} - y| \quad (30)$$

Hence the search continues unless  $y = \hat{y}$ .

It is interesting to note that the observation that led to the definition of the bare-bones PSO [19] — that the standard deviation of the search distribution is proportional to  $|\hat{y} - y|$  — was fundamentally correct. There is, however, a multiplicative factor,  $\frac{1}{2} \sqrt{2(\nu - \mu^2)/\Delta}$ , in Equation (29) which depends on the parameters  $c$  and  $w$  and that was not previously detected. This factor may explain part of the differences in performance observed when comparing the bare-bones PSO and the classical algorithm. The factor is 1 only when

$$c = \frac{48(w^2 - 1)}{19w - 29}$$

As one can see from Figure 3, which graphs all  $(c, w)$  pairs for which  $p_{sd} = |\hat{y} - y|$ , the typical parameter values used in the PSO literature,  $c = 1.49618$  and  $w = 0.7298$ , are not one such pair. More precisely, for these parameter values  $p_{sd} \approx 1.0428 \times |\hat{y} - y|$ .

By setting  $\Delta = 0$  and solving for  $c$  one obtains

$$c = \frac{12(w^2 - 1)}{5w - 7}$$

which represents the line where the magnitude of the largest eigenvalue of  $M$  is 1 (also shown as contour in Figure 2). This is also plotted in Figure 3. As one can see, in this case the pair  $(c, w) = (1.49618, 0.7298)$  lays below the curve, indicating that the variance of the sampling distribution of a PSO with standard parameters has a fixed-point, i.e., the PSO is order-2 stable.

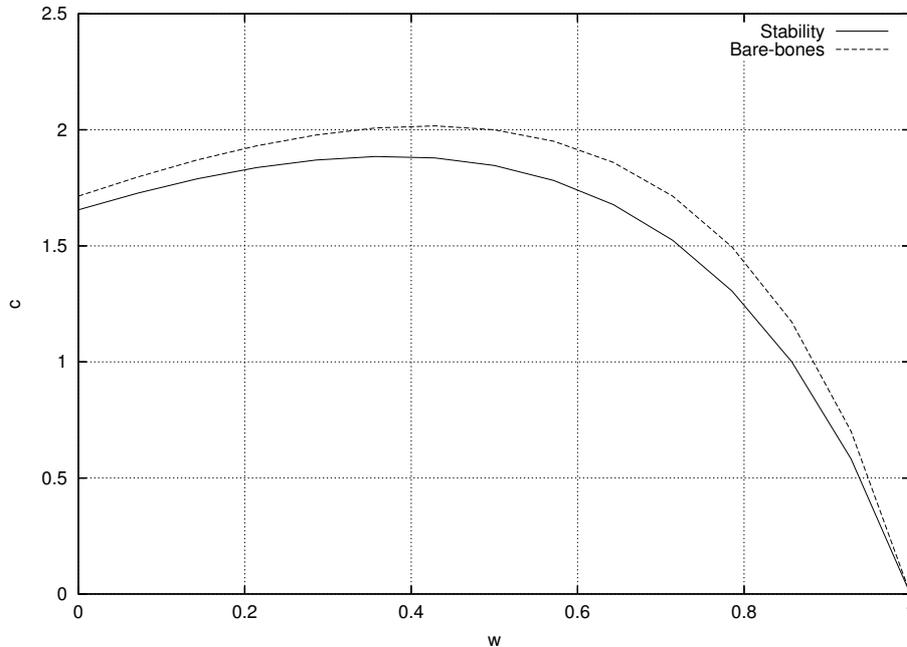


Figure 3: Parameter values for which  $p_{sd} = |\hat{y} - y|$  as predicted by the bare-bones model (“bare-bones” curve) and line below which the variance of the sampling distribution has a fixed point (“stability” curve).

## 4 Numerical integration of order-1 and -2 moments

In this section we report the results of numerical integration of the dynamic equations for  $E[x_t]$ ,  $E[x_t^2]$  and  $E[x_t x_{t-1}]$  (Equation (16)).

We start (Figure 4) by considering the case  $c = 1.49618$  and  $w = 0.7298$  which corresponds to the parameter values recommended in [5] for the PSO with constriction. Note how, while  $E[x_t]$  converges to  $p = 3$  within 30 generations,  $StdDev[x_t]$  converges onto a value of just over 2.0 (more precisely,  $1.0428 \times |\hat{y} - y| = 2.0856$ ) within about 70 generations. The picture is very different, however, if  $y = \hat{y}$ , as shown in Figure 5. In this case,  $E[x_t^2]$  and  $E[x_t x_{t-1}]$  converge to  $p^2$ . As a result,  $StdDev[x_t]$  decreases to zero. The decrease is exponential, corroborating Blackwell’s analysis of how the spatial extent of a particle swarm varies over time [20] (see Section 1).

Examples of configurations where the mean converges to its fixed-point while  $StdDev[x_t]$  does not converge are shown in Figure 6. Note that this is not necessarily an undesirable behaviour. In some situations having a sampling distribution that progressively widens if improvements cannot be found might be exactly what one needs. What is important is to be able to control whether or not there is growth of  $StdDev[x_t]$  and a what rate. This is exactly what our model allows one to do.

By an appropriate setting of parameters we can even achieve a self-limiting growth in  $StdDev[x_t]$ , and, furthermore, we can fix its asymptote by design. A way to achieve this is to note that if  $c = w' = 1 + w$ , the fixed-point for  $E[x_t x_{t-1}]$  in Equation (23) simplifies to  $p_{xx} = p^2$ . Then we have

$$p_{x^2} = p_{x^2} (w'^2 - 4\mu w' + 2\nu + 2\mu^2) + p^2 (-2ww' + 4w\mu) + p_{x^2} w^2 + 4p^2 (\mu - \nu - \mu^2) + \nu y^2 + 2\mu^2 y \hat{y} + \nu \hat{y}^2 \quad (31)$$

which can be solved for  $p_{x^2}$ , obtaining, after simplification, the following fixed-point:

$$p_{x^2} = \frac{(w+1)^2(2p-y)^2/3 + 0.5y(2p-y)(w+1)^2 + y^2(w+1)^2/3 + 2(w - \frac{7}{6}(w+1)^2 + 1)p^2}{-\frac{1}{6}(w+1)^2 - w^2 + 1} \quad (32)$$

With this in hand one can then compute  $p_{sd} = \sqrt{p_{x^2} - p^2}$ . For example, for  $w = 0.5$ ,  $y = -3$  and  $\hat{y} = 9$ , we obtain  $p_{x^2} = 45$  and  $p_{sd} = \sqrt{45 - 9} = 6$ , while for  $w = 0.7$  one obtains  $p_{x^2} = 621$  and  $p_{sd} = \sqrt{621 - 9} = 24.739$ . As one can see in Figure 7, there are indeed the asymptotes to which the system converges.

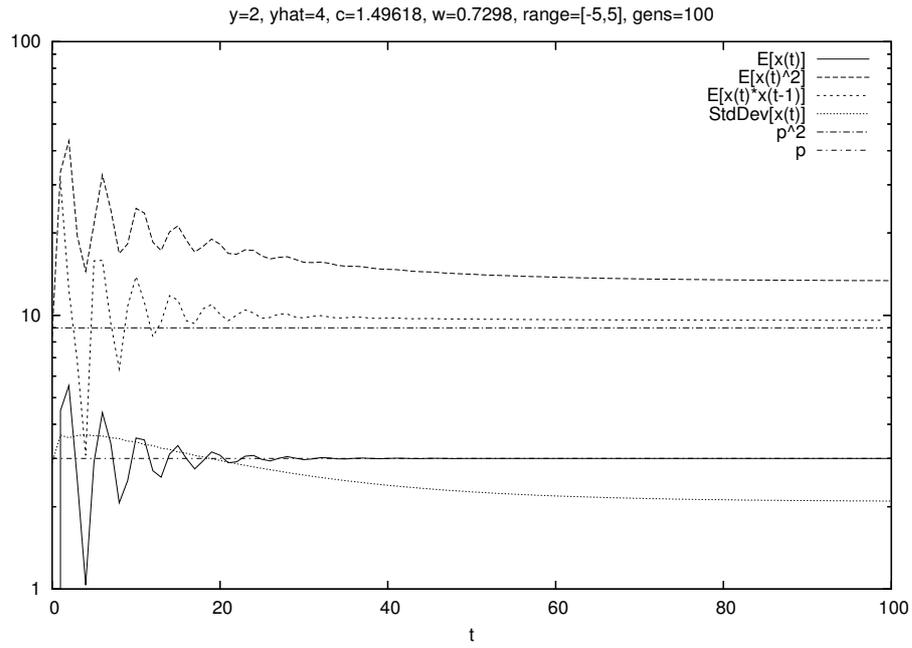


Figure 4: Numerical integration of Equations (16) for  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = 2$ ,  $\hat{y} = 4$  and  $\Omega = 5$ .

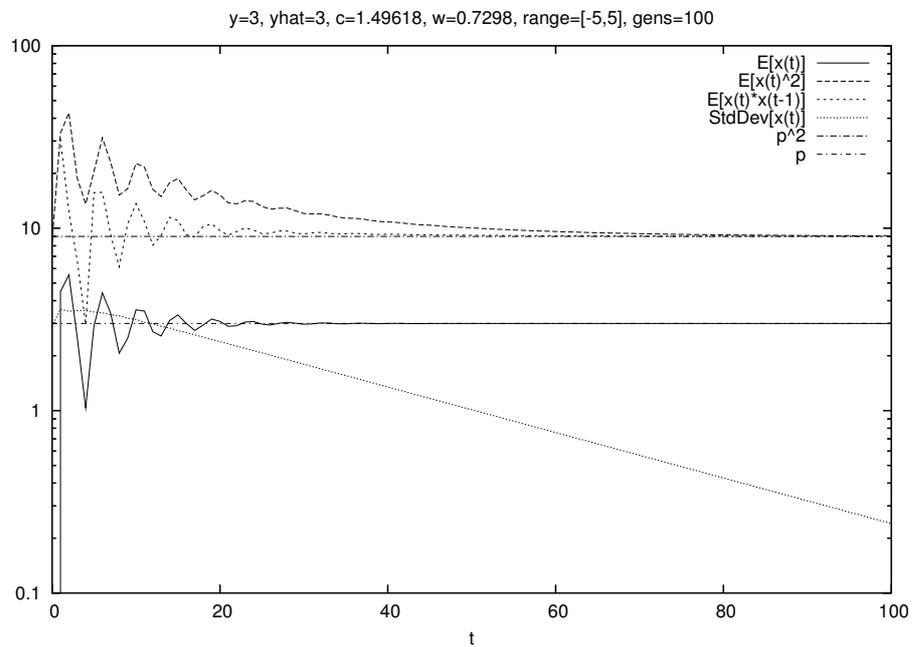


Figure 5: Numerical integration of Equations (16) for  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = \hat{y} = 3$  and  $\Omega = 5$ .

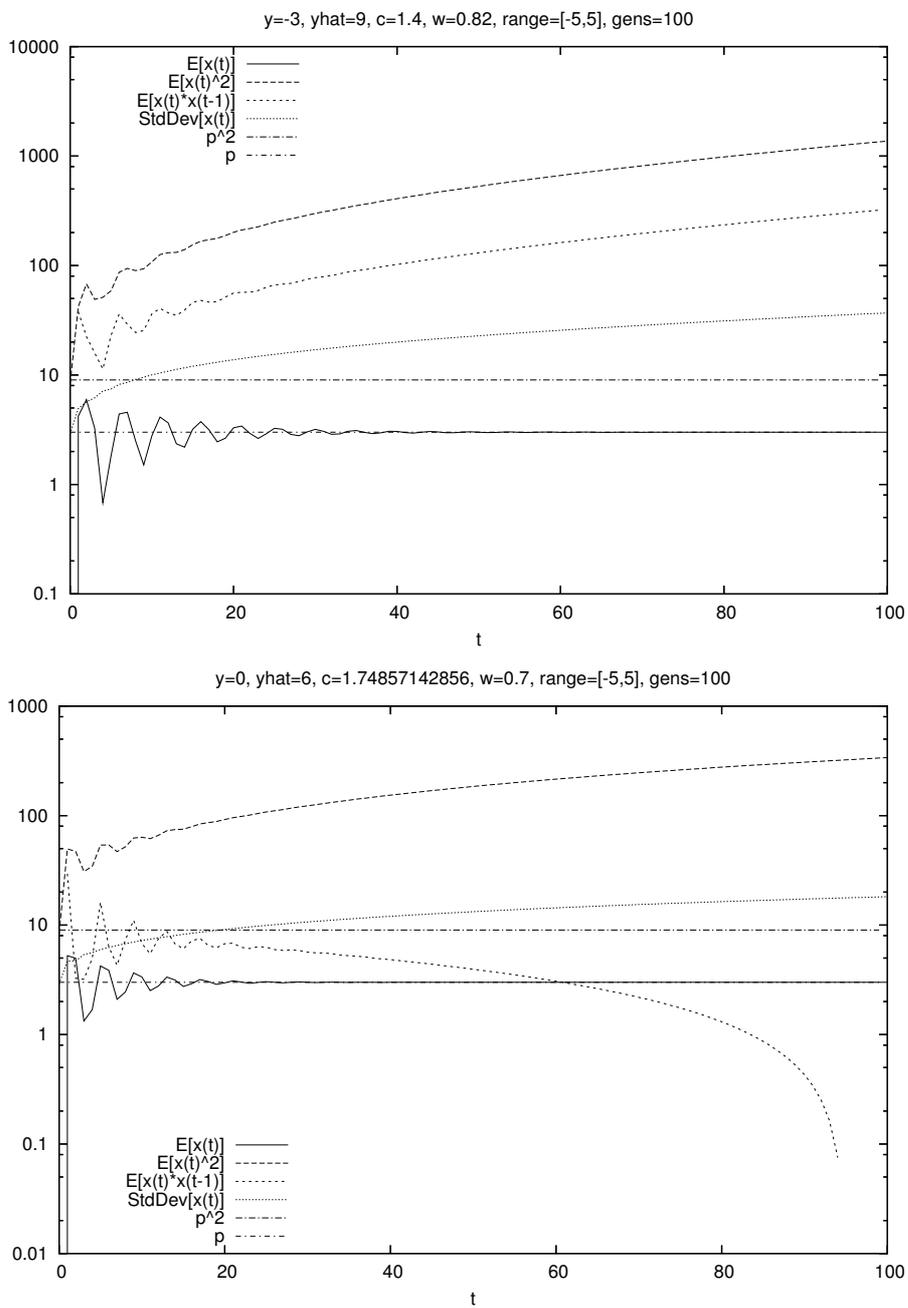


Figure 6: Numerical integration of Equations (16) for values of  $c$  and  $w$  where  $E[x_t]$  is convergent, but  $StdDev[x_t]$  is not.

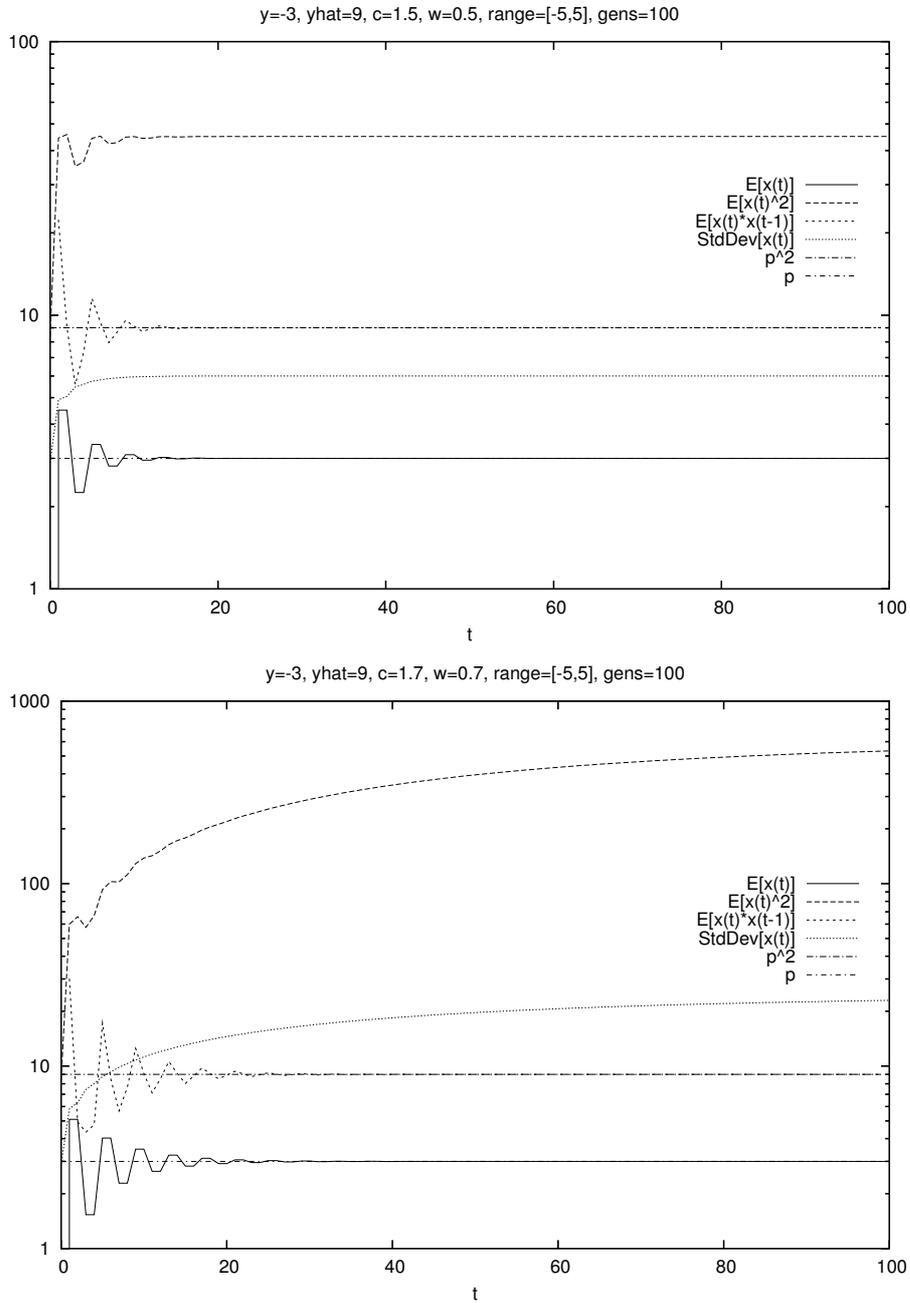


Figure 7: Numerical integration of Equations (16) for values of  $c$  and  $w$  where  $E[x_t]$  is convergent to  $p$ ,  $E[x_t x_{t-1}]$  is convergent to  $p^2$  and  $E[x_t^2]$  is convergent to the value given in Equation (32).

## 5 Empirical validation of order-1 and -2 equations

A key question one needs to answer for any theoretical model, including the one proposed in this paper, is “how accurate is the model?”. To answer this question we compared the behaviour of the model with the behaviour observed in actual runs.

Note that Equation (4) defines a chain of stochastic variables,  $\{x_t\}_{t=0}^{\infty}$ , where each variable depends on the previous ones and on the stochastic variables  $\phi_i$ . For example

$$x_2 = x_1(1 + w - \phi_1^{(1)} - \phi_2^{(1)}) - wx_0 + \phi_1^{(1)}y + \phi_2^{(1)}\hat{y}$$

and

$$\begin{aligned} x_3 &= x_2(1 + w - \phi_1^{(2)} - \phi_2^{(2)}) - wx_1 + \phi_1^{(2)}y + \phi_2^{(2)}\hat{y} \\ &= \left(x_1(1 + w - \phi_1^{(1)} - \phi_2^{(1)}) - wx_0 + \phi_1^{(1)}y + \phi_2^{(1)}\hat{y}\right) (1 + w - \phi_1^{(2)} - \phi_2^{(2)}) - wx_1 + \phi_1^{(2)}y + \phi_2^{(2)}\hat{y} \end{aligned}$$

where  $\phi_i^{(t)}$  are stochastic variables uniformly distributed in the ranges  $[0, c]$  (we used the superscripts  $(1)$  and  $(2)$  to distinguish them in different generations). The number of variables involved in the determination of each new variable grows rapidly with  $t$ . The situation is even worse for the chains of variables  $\{x_t\}_{t=0}^{\infty}$  and  $\{x_t x_{t-1}\}_{t=1}^{\infty}$  and Equations (12) and (14).

Naturally, with so many stochastic variables determining the future of the sampling distribution we should expect to need large numbers of runs to corroborate our results. We should also expect that as  $t$  grows, eventually the experimental results will become increasingly affected by stochastic noise, making the comparison between model and real system harder.

To limit these problems, we present statistics based on large numbers of independent runs of one one-dimensional particle in stagnation. Because no fitness evaluation is needed, these can be done relatively quickly on an ordinary computer. Figure 9(top) shows a comparison between the values of  $\mu_t = E[x_t]$  computed using our model and the average positions of the particle recorded in one billion (1,000,000,000) real runs in the first 30 iterations for the case  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = 0$ ,  $\hat{y} = 1$  and  $\Omega = 5$ . As one can see there is a perfect match between the model’s predictions and the stagnation behaviour of particles in real runs. As shown in Figure 9(bottom) the model also predicts exactly (within experimental errors) the behaviour of the variance,  $\sigma_t^2 = E[(x_t - \mu_t)^2]$ , of the sampling distribution.

The oscillations in the mean and the progressive reduction of the variance towards its (non-zero) fixed-point present is a PSO predicted by our model when  $c = 1.49618$  and  $w = 0.7298$  are also clearly visible in histograms of  $x_t$  observed in actual runs. Figure 8, for example, shows the histogram of the sampling distribution obtained in 1,000,000 independent runs for the case  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = 0$ ,  $\hat{y} = 10$  and  $\Omega = 5$ .

Perfect matches between model and real runs are obtained for all choices of parameters. This further confirms that our model is an exact characterisation of the behaviour of the sampling distribution of a PSO, with all its stochasticity, during stagnation. Naturally, the first and second order moments of such a distribution do not *fully* describe the distribution (there are infinitely many distributions with a given mean and variance). However, as we will discuss in the next section, the method is general and can, in principle, be applied to compute all the moments of the distribution.

## 6 Higher-order moments

In the previous sections we obtained and used recursions which describe the dynamics of first and second order moments of the sampling distribution of a standard PSO during stagnation. One may then wonder whether it would be possible to follow a similar approach to study the dynamics of higher-order moments.

The fundamental question in: what quantities would we have to deal with if we took higher powers of both sides of Equation (4) as we did to derive Equation (12)? Generally, the r.h.s. we would be a sum of terms of the form

$$a_0 x_t^{a_1} x_{t-1}^{a_2} w^{a_3} \phi_1^{a_4} \phi_2^{a_5} y^{a_6} \hat{y}^{a_7} \quad (33)$$

where  $a_k$  are suitable constants. Naturally, taking powers of Equation (4) and then multiplying both sides by some power of  $x_t$ , as we did to derive Equation (14), would also lead to equations involving terms such as

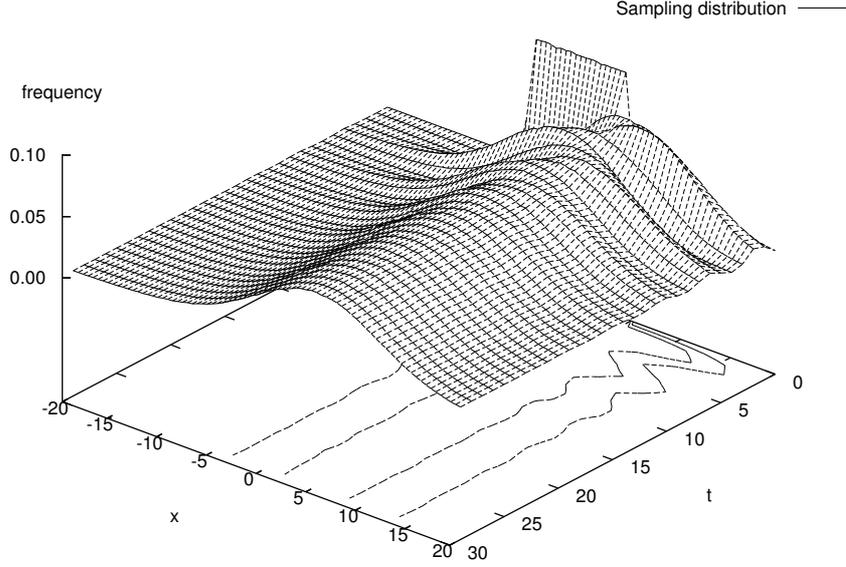


Figure 8: Sampling distribution observed in real runs when  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = 0$ ,  $\hat{y} = 10$  and  $\Omega = 5$ . Histograms based on 1,000,000 runs. Contours correspond to frequencies of 0.025 and 0.05.

those in Equation (33). That is, for any choice of  $b_1 \in \mathbb{N}$  and  $b_2 \in \mathbb{N}$ ,

$$x_{t+1}^{b_1} x_t^{b_2} = \sum_i a_{0_i} x_t^{a_{1_i}} x_{t-1}^{a_{2_i}} w^{a_{3_i}} \phi_1^{a_{4_i}} \phi_2^{a_{5_i}} y^{a_{6_i}} \hat{y}^{a_{7_i}} \quad (34)$$

where  $a_{k_i}$  are suitable constants. If we then take expectations for both sides we obtain

$$E[x_{t+1}^{b_1} x_t^{b_2}] = \sum_i a_{0_i} w^{a_{3_i}} y^{a_{6_i}} \hat{y}^{a_{7_i}} E[\phi_1^{a_{4_i}}] E[\phi_2^{a_{5_i}}] E[x_t^{a_{1_i}} x_{t-1}^{a_{2_i}}] \quad (35)$$

where we used the independence of  $\phi_1$ ,  $\phi_2$ ,  $x_t x_{t-1}$  and, of course, their powers. Because  $\phi_j$  is uniformly distributed in the range  $[0, c]$ , it is easy to verify that

$$E[\phi_j^n] = \frac{c^n}{n+1}. \quad (36)$$

So,

$$E[x_{t+1}^{b_1} x_t^{b_2}] = \sum_i \omega_i E[x_t^{a_{1_i}} x_{t-1}^{a_{2_i}}] \quad (37)$$

where

$$\omega_i = \left( \frac{a_{0_i} w^{a_{3_i}} c^{a_{4_i} + a_{5_i}} y^{a_{6_i}} \hat{y}^{a_{7_i}}}{(1 + a_{4_i})(1 + a_{5_i})} \right). \quad (38)$$

It is important to note here that, because Equation (4) is linear in  $x_t$  and  $x_{t-1}$ , all the terms on the r.h.s. of Equations (34) and (37) respect the relation  $a_{1_i} + a_{2_i} \leq b_1 + b_2$ . This implies that *it is possible to construct recursions for moments of arbitrary order*.

For example, if one wanted to push the analysis up to order 3, one would need to instantiate Equation (37) for  $E[x_{t+1}^3]$ ,  $E[x_{t+1}^2 x_t]$ ,  $E[x_{t+1} x_t^2]$  and add the resulting equations to the three in Equation (16). If one wanted to go to order four, an additional set of four equations (for  $E[x_{t+1}^4]$ ,  $E[x_{t+1}^3 x_t]$ ,  $E[x_{t+1}^2 x_t^2]$  and  $E[x_{t+1} x_t^3]$ ) would be needed, bringing the total to 10.

More generally, in order to compute statistics of order  $n$  one needs to construct and iterate

$$Q(n) = \frac{n \times (n+1)}{2} \quad (39)$$

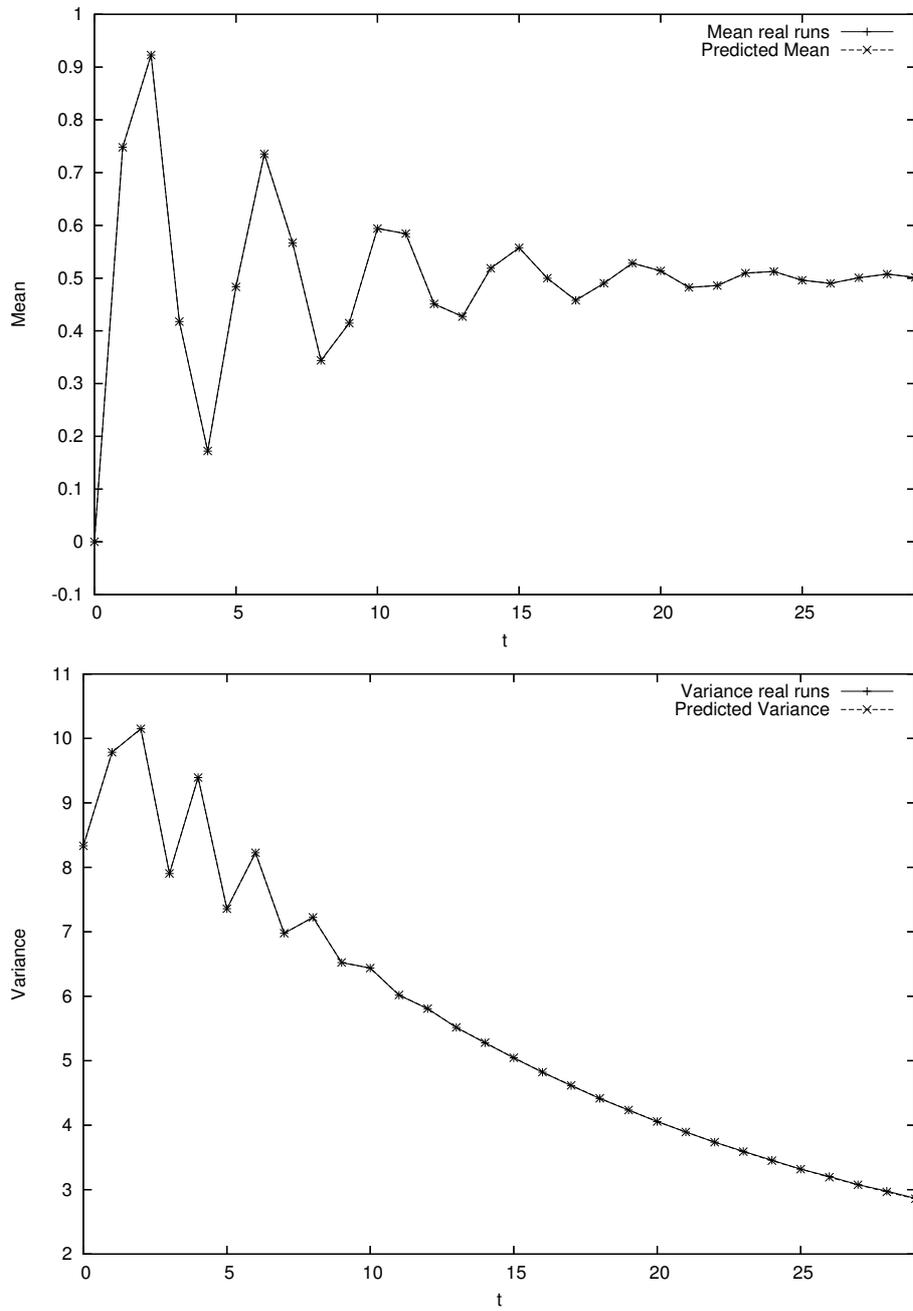


Figure 9: Comparison between predicted and experimental means and variances of the PSO sampling distribution for  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = 0$ ,  $\hat{y} = 1$  and  $\Omega = 5$ .

second order difference equations. Since, after expansion, the r.h.s. of Equation (4) contains 7 atomic terms of the form in Equation (33), the r.h.s. of Equation (37) contains  $7^{b_1}$  terms.<sup>1</sup> So, the total number of terms one needs to compute to construct the equations for order- $n$  statistics is: 7 (for the order 1 equations) plus  $(7^2 + 7)$  (for the order 2 equations) plus  $(7^3 + 7^2 + 7)$  (for the order 3 equations) etc., which gives us a total of

$$T(n) = \sum_{i=1}^n (n - i + 1) \times 7^i$$

terms. E.g.,  $T(1) = 7$ ,  $T(2) = 63$ ,  $T(3) = 462$ ,  $T(4) = 3262$  and  $T(5) = 22869$ . Note that  $T(n)$  grows exponentially approximately as  $1.36 \times 7^n$ . So, although the number of equations one needs to deal with grows quadratically, the computational effort required to instantiate them is exponential. The growth in number of terms can be reduced if one makes explicit use of  $w'$  (i.e., by adding the factor  $w'^{a_8}$  in Equation (33)). Then  $T(n) = O(6^n)$ . Either way, manually deriving equations for moments of order 3 is already vary laborious. The process, however, is clearly mechanisable. This can be using computer algebra systems, or by explicitly representing and manipulating the  $\omega_i$ 's for each equation (this is what we did). As a result of mechanisation, computing the equations for up to order 6 or 7 is feasible with an ordinary personal computer.

Some of the  $\omega_i$ 's in Equation (37) present the same pattern of exponents for  $w$ ,  $c$ ,  $y$  and  $\hat{y}$ , so terms can be collected leading to more compact equations (for example, compare Equations (12) and (13)). Also, given their size, one will normally want to study (e.g., integrate) Equations (37) numerically. In this case  $w$ ,  $c$ ,  $y$  and  $\hat{y}$  are all numerical parameters. So, the  $\omega_i$ 's become constants and, after collecting terms, each equation contains at most  $Q(n)$  terms, which, as we know, is quadratic in the order  $n$ . As a result, although the complexity of the construction of the motion equations for the moments is exponential in the order of the moments, their numerical integration is only of order  $O(n^4)$ .

Naturally, the system of  $Q(n)$  second order difference equations necessary to predict the dynamics of moments of order 1 to  $n$  can be turned into a system of order 1 of the form in Equation (18), via the choice

$$\mathbf{z}(t) = (E[x_t] \ E[x_{t-1}] \ E[x_t^2] \ E[x_t x_{t-1}] \ E[x_{t-1}^2] \ E[x_t^3] \ E[x_t^2 x_{t-1}] \ E[x_t x_{t-1}^2] \ E[x_{t-1}^3] \ \dots \ E[x_{t-1}^n])^T$$

This effectively means adding artificial update equations of the form  $E[x_t^k] = E[x_t^k]$  for  $k = 1, \dots, n$ , bringing the total to  $Q'(n) = Q(n) + n$ . The transition matrix for the system is therefore of size  $Q'(n) \times Q'(n)$ . We will denote this with  $M_n$ . For example, for  $n = 4$ , which would allow one to study the mean, variance, skewness and kurtosis of the sampling distribution as a function of  $t$ ,  $M_4$  is merely a  $14 \times 14$  matrix.

Interestingly,  $Q'(n)$  grows so slowly that one can perform an eigenvalue analysis for any  $M_n$  that one is able to compute. That is, the expensive part of the process is the construction of  $M_n$ . Once this is done, iterating the system, establishing its stability or finding its fixed-points is a trivial matter.

In the next section we provide results for statistics of order 3 and 4, i.e.,  $n = 4$ , a value of  $n$  for which computing  $M_n$  takes only a few seconds. However, before we do this, we need to consider the initial conditions for the system. In particular we need to compute  $E[x_0^k]$  and  $E[x_1^k x_0^l]$  for generic  $k > 0$  and  $l \geq 0$ .

Under the assumption that a particle's initial position,  $x_0$ , is chosen uniformly at random in a symmetric range  $[-\Omega, \Omega]$ , we have

$$E[x_0^k] = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ \frac{\Omega^k}{k+1} & \text{otherwise.} \end{cases} \quad (40)$$

In order to compute  $E[x_1^k x_0^l]$  we need to consider the equation

$$x_1 = x_0 + wv_0 - x_0(\phi_1 + \phi_2) + \phi_1 y + \phi_2 \hat{y} \quad (41)$$

where a particle's initial velocity,  $v_0$ , is a stochastic variable uniformly distributed the range  $[-\Omega_v, \Omega_v]$  (often  $\Omega_v = \Omega$ ). By taking the  $k$ -th power of both sides of the equation, multiplying by  $x_0^l$ , and taking expectations, as we did to construct Equation (37), one obtains the desired expressions for  $E[x_1^k x_0^l]$ . Like for Equation (37), these expressions contain a number of terms that grows exponentially for with  $n$ . However, this process, too, can be trivially mechanised.

---

<sup>1</sup>The exponent  $b_2$  does not influence the number of terms. This is because the recursion for  $E[x_{t+1}^{b_1} x_t^{b_2}]$  in obtained as follows: a) we compute  $x_{t+1}^{b_1}$ , which is given by an expression containing  $7^{b_1}$  terms; b) we multiply each term by  $x_t^{b_2}$ , which changes the exponents  $a_{1_i}$  but does not alter the number of terms; c) we apply the expectation operator, which again does not modify the number of terms.

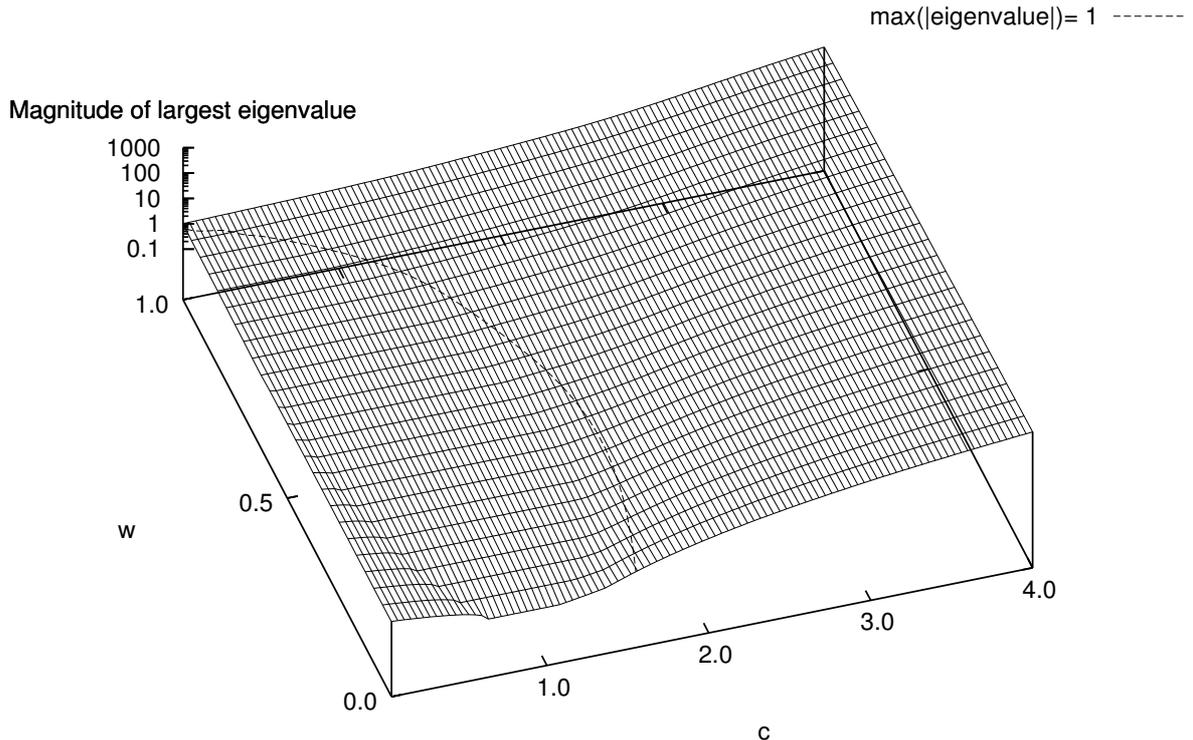


Figure 10: Magnitude of the largest eigenvalue of  $M_4$  as a function of the parameters  $w$  and  $c$ . The curved line on the surface encloses the order-4 stable region.

## 7 Skewness and kurtosis of the PSO's sampling distribution

We constructed the recursions for moments of up to order 4 as described in the previous section for the canonical PSO. In principle, we could do for these exactly the same type of analysis we did for the mean and standard deviation of the sampling distribution.

For example, Figure 10 shows the magnitude of the largest eigenvalue,  $\Lambda_m$ , of  $M_4$  as a function of the parameters  $w$  and  $c$ . The system is order-4 stable, i.e., mean, variance, skewness and kurtosis have a stable fixed point, whenever  $\Lambda_m < 1$ , i.e., within the curved region enclosed by the contour shown in the figure. Note how this contour is qualitatively similar to that in Figures 2 and 3, but it encloses a smaller region.

For easier comparison we show the lines where  $\Lambda_m = 1$  for  $M_1$ ,  $M_2$ ,  $M_3$  and  $M_4$  in Figure 11 (ordered from top to bottom, respectively). The regions of order-1, -2, -3 and -4 stability are nested. Note how the  $\Lambda_m = 1$  lines for  $M_2$  and  $M_3$  coincide for many values of  $w$ . Note also that the standard setting,  $c = 1.49618$  and  $w = 0.7298$ , lays within the narrow region of order-3 stability. This implies that while mean, variance and skewness of the standard PSO tend to a fixed-point, kurtosis is unstable and will tend to grow indefinitely. Interestingly, a growth in the kurtosis of samples was observed by Kennedy[19], although this was effectively computed under the assumption that the sampling distribution is time-independent. So, the values of  $x_t$  recorded in a run at  $t$  grows were treated as different samples from the same distribution, while we know this may be incorrect.

In order to corroborate these findings, one needs to cover the  $(c, w)$  plane with a regular lattice of points. We used a grid side of 0.04, giving us 25 divisions for the  $w$  axis and 100 divisions for the  $c$  axis, for a total of 25,000 points. For each such point one then needs to perform a large set of runs of the PSO, computing mean, variance, skewness and kurtosis at each time step. We used 1,000,000 runs for each  $(w, c)$  pair. Runs lasted 100 iterations. Finally one needs to analyse the dynamics of each of these statistics to determine whether the system is stable. This last step is particularly difficult since no empirical criterion can really do

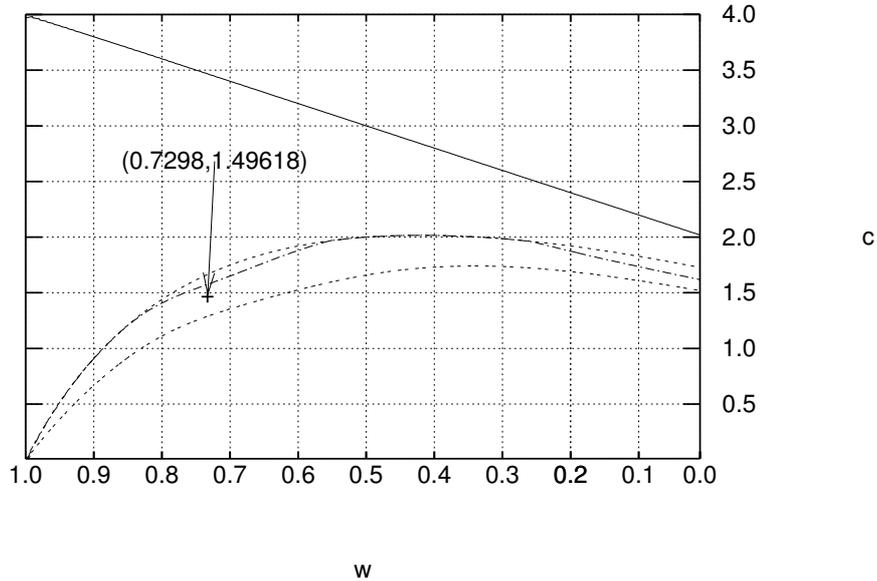


Figure 11: Plot of the regions of order-1, -2, -3 and -4 stability for the canonical PSO.

what eigenvalue analysis does. However, it is clear that, after some transient, moments that have fixed points will tend to have less energy in their power spectrum than those that are unstable (excluding, of course, the continuous component of the spectrum). So, we computed the power spectrum for each moment and each  $(c, w)$  and then applied an empirically chosen threshold to classify the dynamics. The result of this process is shown in Figure 12 (B-splines were used to smooth the contours of the regions of stability). As one can see there is a very close resemblance between these regions and those shown in Figure 11 except for the case of the mean where ample oscillations were often still present after 100 iterations for values of  $c > 2$  leading to large energy components being still present in the power spectrum with consequent misclassification. Much longer runs would be required to more closely match the profile of the order-1 stability region.

That the predictions of the model are exact is also confirmed by the comparison of the dynamics of predicted and recorded higher order moments. Figure 13(top) shows a comparison between the skewness  $E[(x_t - \mu_t)^3]/\sigma_t^3$  computed using our model and the average positions of the particle recorded in one billion (1,000,000,000) real runs in the first 30 iterations for the case  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = 0$ ,  $\hat{y} = 1$  and  $\Omega = 5$ . As one can see there is a very good match between the model's predictions and the stagnation behaviour of particles in real runs. Only after about 20-25 generations the sampling errors start accumulating enough to show significant differences. As shown in Figure 13(bottom) the model also predicts very well the behaviour of the (excess) kurtosis  $E[(x_t - \mu_t)^4]/\sigma_t^2 - 3$  of the sampling distribution.<sup>2</sup> Note that for  $c = 1.49618$  and  $w = 0.7298$  the system is order-3 stable, and so, although the oscillations of the skewness shown in Figure 13(top) appear to grow bigger and bigger, suggesting instability, this is actually only a transient effect, as shown in Figure 14 where we integrate the equations over 200 generations instead of 30.

## 8 Comparison between different PSOs

In the previous sections we studied the canonical PSO with the restriction that the acceleration coefficients,  $c_1$  and  $c_2$ , were identical:  $c_1 = c_2 = c$ . One may wonder, however, whether allowing such coefficients to

<sup>2</sup>Following standard practice, in this paper whenever we use the term "kurtosis" we will refer to the excess kurtosis  $E[(x_t - \mu_t)^4]/\sigma_t^2 - 3$ . The excess kurtosis of the normal distribution is 0.

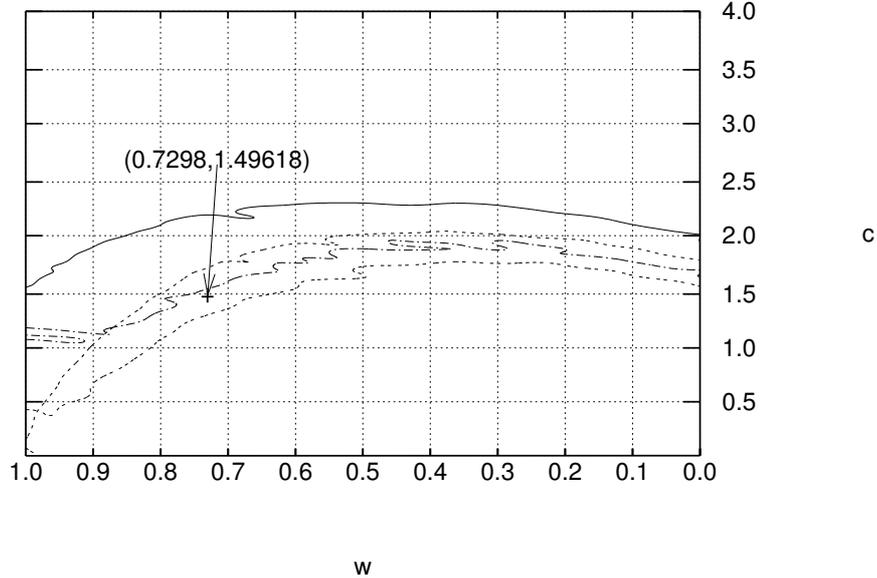


Figure 12: Plot of the regions of order-1, -2, -3 and -4 stability for the canonical PSO estimated in real runs.

differ would produce qualitatively very different dynamics. For example, what would happen if we set one of the  $c_i$  to zero as in a purely cognitive or purely social PSO model? This effectively would reduce to one the sources of random influences on a particle's dynamics. Conversely, one might wonder what would happen if we increased such sources of influence, as is done, for example, in the Fully Informed Particle Swarm (FIPS) [14, 15].

To answer these (and other) important questions on the sampling distribution of different PSO models we adopt a FIPS-like general class of PSOs described by the following difference equation

$$x_{t+1} = x_t + wv_t + \sum_{i=1}^m \phi_i(\hat{y}_i - x_t) \quad (42)$$

where the  $\phi_i$ 's are stochastic variables uniformly distributed in the range  $[0, c_i]$ ,  $c_i$  being constants, and the  $\hat{y}_i$ 's are the personal best positions of neighbours of the particle (the particle itself may be included in its own neighbourhood). Naturally, this equation can be converted into the following

$$x_{t+1} = x_t(1 + w) - wx_{t-1} - \sum_{i=1}^m \phi_i x_t + \sum_i \phi_i \hat{y}_i \quad (43)$$

which is a generalisation of Equation (4). All of the steps we performed in Section 6 can be repeated for Equation (43). These lead to recursion of the form in Equation (37) with the only difference that the coefficients  $\omega_i$  take the more general form

$$\omega_i = \left( \frac{a_{0_i} w^{a_{w_i}} c_1^{a_{c1_i}} \dots c_m^{a_{cm_i}} \hat{y}_1^{a_{y1_i}} \dots \hat{y}_m^{a_{ym_i}}}{\prod_{j=1}^m (1 + a_{cj_i})} \right) \quad (44)$$

where  $a_{0_i}$ ,  $a_{w_i}$ ,  $a_{c1_i}, \dots, a_{cm_i}$ ,  $a_{y1_i}, \dots, a_{ym_i}$  are appropriate constants.

Because Equation (43) contains  $3 + 2 \times m$  terms, the complexity of the expansion now grows exponentially as  $O((3 + 2 \times m)^n)$ , where  $n$  is the order of the moments we are interested in. So, the larger  $m$ , the heavier the computation load required to compute  $M_n$ . Once the transition matrices  $M_n$  are computed, however,

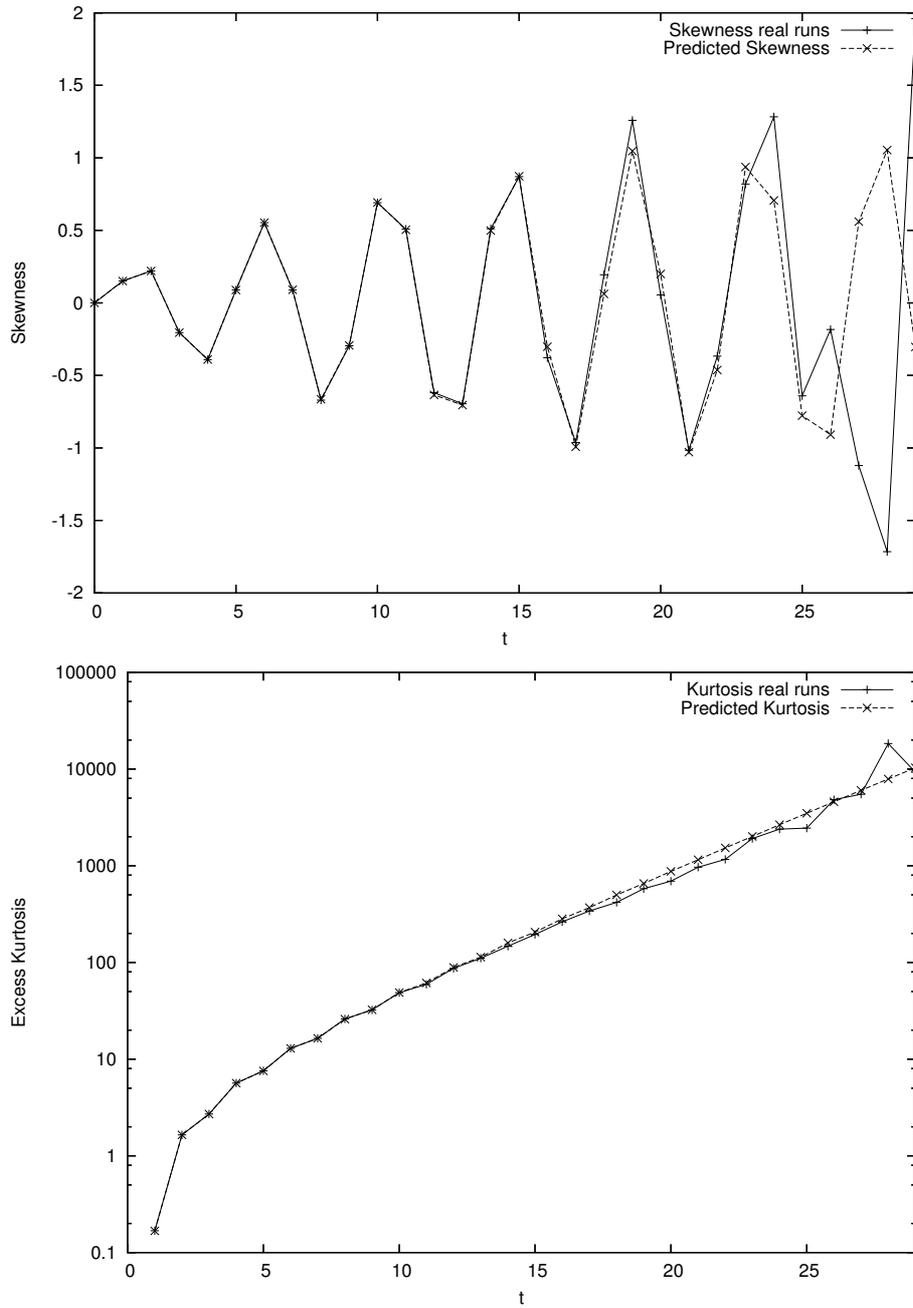


Figure 13: Comparison between predicted and experimental skewness and (excess) kurtosis of the PSO sampling distribution for  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = 0$ ,  $\hat{y} = 1$  and  $\Omega = 5$ . Kurtosis grows exponentially and, so, it is plotted on a logarithmic scale. The first point is not plotted because the excess kurtosis was negative (-1.2).

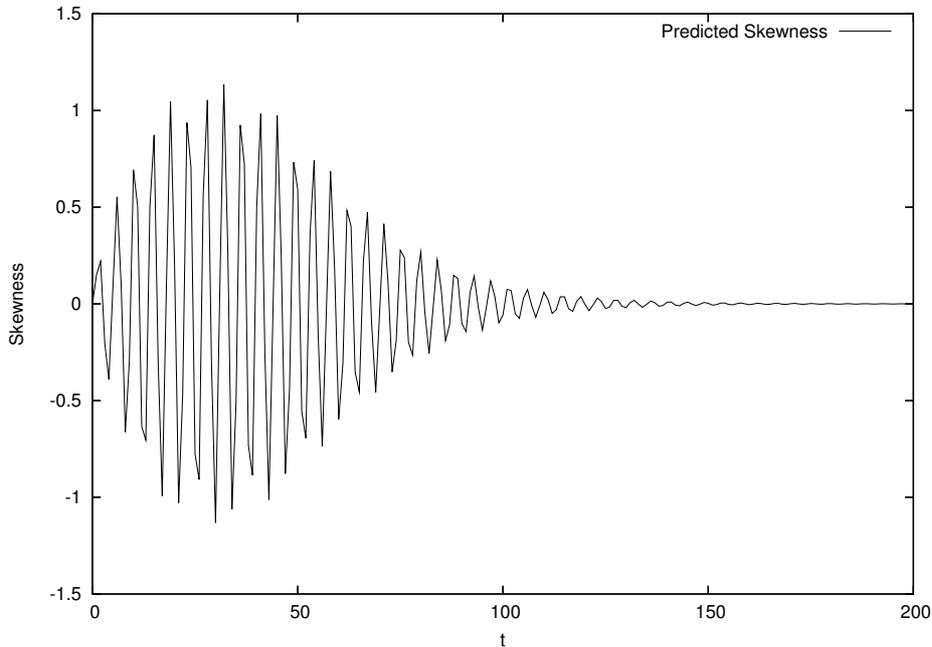


Figure 14: Predicted dynamics of the skewness of the PSO sampling distribution over 200 generations for  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = 0$ ,  $\hat{y} = 1$  and  $\Omega = 5$ .

they are exactly of the same size for all PSO models within the class defined by Equation (42). Initial conditions can be found following the approach described in Section 6. Calculations are expensive but can be mechanised. We did this for the examples described below.

An extensive comparison of different PSOs is beyond the scope of this paper. However, as an example of the kind of comparisons one can make using our approach, we considered the PSOs in Equation (42) with  $N = 3$ . Within this class of PSOs we considered three variants:

- a) a purely social variant of PSO, which we will call *social PSO* for brevity, where  $c_1 > 0$  and  $c_2 = c_3 = 0$  (due to symmetries, the behaviour of a purely cognitive PSO where  $c_2 > 0$  and  $c_1 = c_3 = 0$  is effectively identical to that of this social PSO);
- b) the *canonical PSO* we have studied so far in the paper, which is obtained by setting  $c_1 = c_2 > 0$  and  $c_3 = 0$ ; and
- c) the simplest version of FIPS with a neighbourhood of three individuals, e.g., obtained using an *lbest* topology and an interaction radius of 1, where  $c_1 = c_2 = c_3 > 0$ . We will call this version *FIPS3*.

In order to perform a fair comparison of the stability properties of these PSO variants, we study them in conditions where the sum of the amplitudes of the random components,  $\phi_i$ , is identical across models. That is, we set  $c = \sum_i c_i$ , we compare models with the same  $c$  value. Again we analyse eigenvalues. Figures 15–18 show the lines in the  $(w, c)$  plane where the magnitude of the largest eigenvalue of  $M_n$ ,  $\Lambda_m$ , is 1 for  $n = 1, 2, 3, 4$  and for the three PSO variants mentioned above. Let us analyse these figures in detail.

Firstly, we should note that the regions of order-1 stability for the three models are identical. This is because the dynamics of the mean of the three models is governed by equations of the same form, namely:

$$E[x_{t+1}] = E[x_t] \left(1 + w - \frac{c}{2}\right) - wE[x_{t-1}] + \text{constant}, \quad (45)$$

where the constant term may differ in different PSO variants.<sup>3</sup> Note also that the rightmost point in each plot is an artifact due to the fact that, at  $w = 1$ ,  $\Lambda_m = 1$  for  $M_1$  irrespective of the value of  $c$ .

The regions where the variance is stable for the three models, instead, are different, with FIPS3 having the largest region of order-2 stability, followed by the canonical PSO, and, finally, by the social PSO. Exactly

<sup>3</sup>This is irrelevant for the stability of the system, since stability is determined by the homogeneous part of the equation.

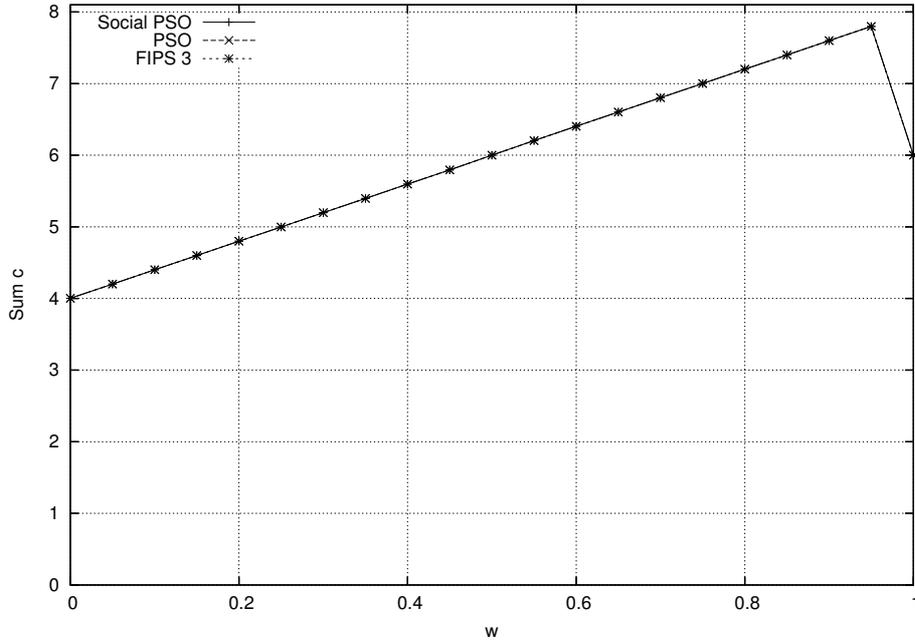


Figure 15: Lines below which the mean of the sampling distributions for a social PSO, a canonical PSO and FIPS with a neighbourhood of three individuals have a fixed point (order-1 stability). Note: the three lines coincide.

the same happens with skewness (Figure 17) and kurtosis (Figure 18), with the order-3 stability region largely coinciding with the order-2 ones also for FIPS3 and the social PSO. These results are counter intuitive. One would expect that the more sources of randomness, the  $\phi_i$ 's, there are, the more a PSO should be unstable. However, the exact opposite happens. The social PSO, where the only influence is  $\phi_1$ , is the least stable of all models, while FIPS3, which has three sources of randomness, is the most stable. What are the reasons for this behaviour?

We can understand this by rewriting Equation (43) as follows

$$x_{t+1} = x_t(1 + w) - wx_{t-1} - x_t\Phi_m + \Psi_m \quad (46)$$

where  $\Phi_m = \sum_{i=1}^m \phi_i$  and  $\Psi_m = \sum_i \phi_i \hat{y}_i$ . Both  $\Phi_m$  and  $\Psi_m$  are the sum of independent and uniformly distributed variables: the variables  $\phi_i$  in the case of  $\Phi_m$  and the variables  $\hat{y}_i \phi_i$  in the case of  $\Psi_m$ .

We know that  $\sum_i c_i = c$ . To simplify our treatment, let us further assume that the  $\phi_i$ 's are i.i.d., i.e., that all  $c_i$  are identical, and, so,  $c_i = c/m$ . We can then apply the central limit theorem to  $\Phi_m$ . This predicts that for sufficiently large  $m$ , the distribution of  $\Phi_m$  is approximately Gaussian with mean  $\sum_i c_i/2 = c/2$  and variance  $\sum_i (c_i^2/3 - (c_i/2)^2) = \sum_i c_i^2/12 = c^2/(12m)$ . So, the larger  $m$ , the smaller the variance of  $\Phi_m$ , and, consequently the less the stochasticity of Equation (46).

In the case of the stochastic variable  $\Psi_m$ , the quantities  $\phi_i \hat{y}_i$  are not identically distributed even if all  $c_i$  are identical. This is because, in principle, each  $\hat{y}_i$  may be different. This prevents the use of the standard central limit theorem. We can, however, apply Lyapunov's central limit theorem to  $\Psi_m$ . The conditions for its application are:

1. the variables  $\phi_i \hat{y}_i$  must have finite mean, which is the case since  $\mu_i = E[\phi_i \hat{y}_i] = c_i \hat{y}_i/2 = c \hat{y}_i/(2m)$ ,
2. the  $\phi_i \hat{y}_i$  must have finite variance, which, again, is the case since  $\sigma_i^2 = E[(\phi_i \hat{y}_i - \mu_i)^2] = (c_i \hat{y}_i)^2/12 = (c \hat{y}_i/m)^2/12$ ,
3.  $\phi_i \hat{y}_i$  must have finite third central moment, which is satisfied since  $r_i^3 = E[(\phi_i \hat{y}_i - \mu_i)^3] = 0$ , and, finally,
4. the Lyapunov condition,  $\lim_{m \rightarrow \infty} \frac{(\sum_{i=1}^m r_i^3)^{1/3}}{(\sum_{i=1}^m \sigma_i^2)^{1/2}} = 0$ , must be satisfied, which, again, is the case since all  $r_i^3 = 0$ .

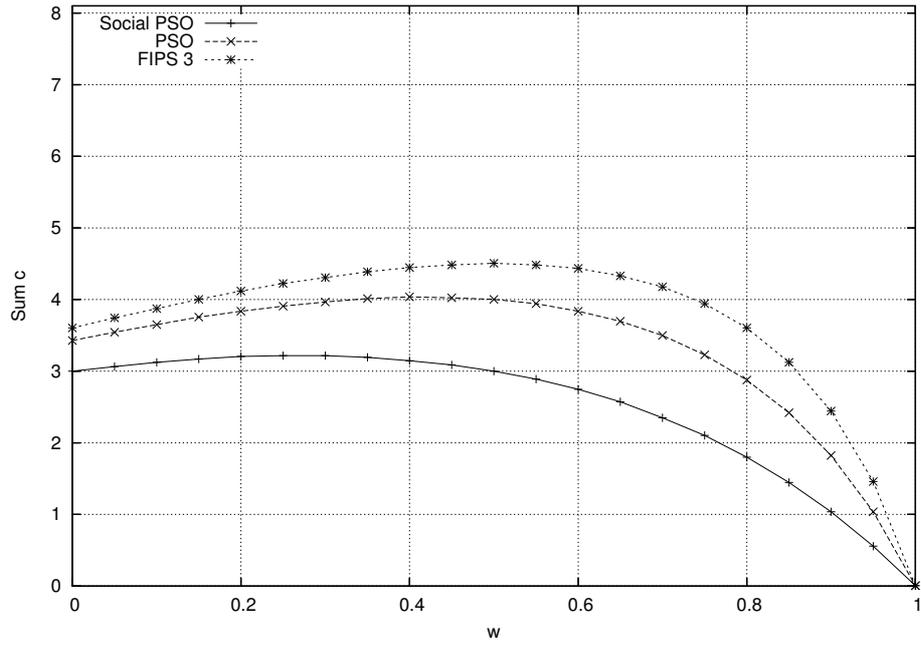


Figure 16: Lines below which the variance of the sampling distributions for a social PSO (bottom), a canonical PSO (middle) and FIPS with a neighbourhood of three individuals (top) have a fixed point (order-2 stability).

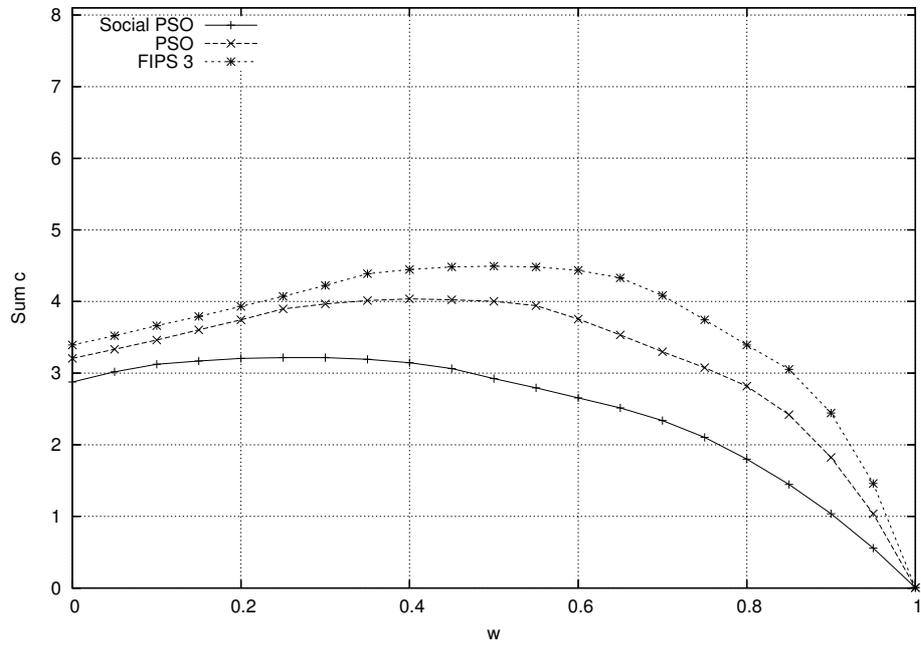


Figure 17: Lines below which a social PSO, a canonical PSO and FIPS3 are order-3 stable (from bottom to top, respectively).

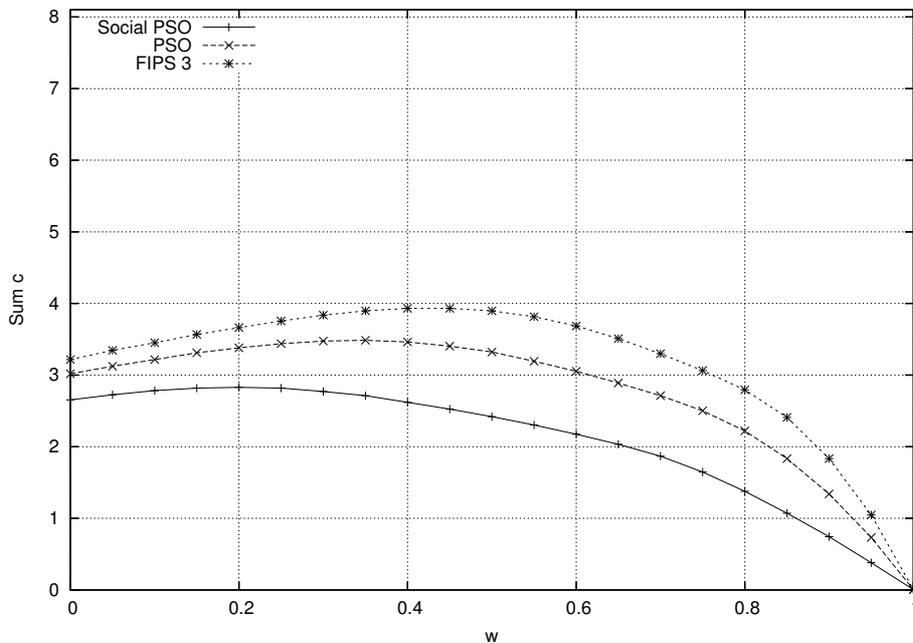


Figure 18: Lines below which the kurtosis of the sampling distributions for a social PSO (bottom), a canonical PSO (middle) and FIPS3 (top) have a fixed point (order-4 stability).

Then for sufficiently large  $m$ , also the distribution of  $\Psi_m$  is approximately Gaussian with mean  $\sum_i c\hat{y}_i/(2m) = \frac{c}{2} \times \left(\frac{\sum_i \hat{y}_i}{m}\right)$  and variance  $\frac{c^2}{12m} \times \left(\frac{\sum_i \hat{y}_i^2}{m}\right)$ . Note that  $\frac{\sum_i \hat{y}_i}{m}$  and  $\frac{\sum_i \hat{y}_i^2}{m}$  are the mean  $\hat{y}_i$  and the mean  $\hat{y}_i^2$ , respectively. So, these are finite quantities if, as is normally the case, all  $\hat{y}_i$  are finite.<sup>4</sup> So, like for  $\Phi_m$ , the larger  $m$ , the smaller the variance of  $\Psi_m$ , and, consequently the less the stochasticity of Equation (46).

Effectively the larger  $m$  the more  $\Phi_m$  and  $\Psi_m$  become deterministic and approach constant values. This explains why adding more and more sources of randomness – while keeping  $c$  constant – produces progressively more and more stable PSOs.

## 9 The density function of the PSO sampling distribution

The technique described in this paper, in principle, would allow one to determine all the moments of the sampling distribution of the PSO at all times. The question then is, could we derive the PSO sampling distribution itself? The answer is of course in the positive since knowing all the moments of a distribution implies knowing its moment generating function. This, in turn, allows one to obtain the density function of the distribution via inverse Laplace transform.

In practice, however, it is impossible to compute all the moments of the PSO sampling distribution. This is for two reasons. Firstly, there are infinitely many such moments. Secondly, as we have seen in the previous sections, the cost of computing moments is exponential in the order of the moments. The next question is then, to what extent can we still reconstruct the PSO's density function from a finite number of moments? This is an instance of the well-known *truncated moment problem*, a difficult, inverse ill-posed problem for which many approaches have been proposed. Here we consider only one such approach.

A particularly simple idea is to consider a family of density functions  $f(x; \lambda_1, \lambda_2, \dots)$  with parameters  $\lambda_1, \lambda_2$ , etc. with sufficient expressive power to represent distributions with widely different shapes, with more or less asymmetries, with tails of different characteristics, etc.. Then one can use optimisation techniques to identify the parameters of the distribution  $f$  which minimise the difference between the moments of  $f$  and the moments of the PSO's sampling distribution. This is called the *moment matching method*. Once the parameters  $\lambda_1, \lambda_2$ , etc., are identified,  $f$  can be used as an approximation of the true PSO sampling distribution. This approach to reconstructing probability distributions from moments was proposed [23] (see

<sup>4</sup>PSO search is normally confined to a pre-fixed, finite region of  $\mathbb{R}^N$ , and so, all  $\hat{y}_i$  must be finite.

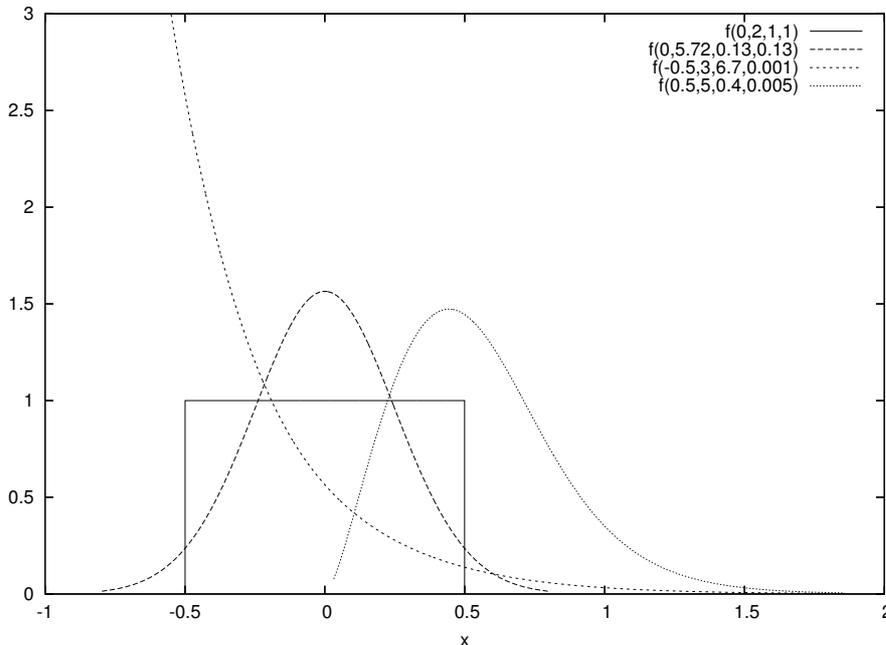


Figure 19: Examples of GLD probability density functions.

also [22, 21]) where a *Generalised Lambda Distribution* (GLD) was used. We adopt this same approach here.

GLD is a four-parameter distribution defined via its quantile function:

$$Q(u) = \lambda_1 + \frac{1}{\lambda_2} \left( \frac{u^{\lambda_3} - 1}{\lambda_3} + \frac{1 - (1 - u)^{\lambda_4}}{\lambda_4} \right) \quad (47)$$

where  $u \in [0, 1]$ . Its density function can be obtained via the relation

$$f(x; \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \left( \frac{dQ(u)}{du} \right)^{-1} = \frac{\lambda_2}{u^{(\lambda_3-1)} + (1-u)^{(\lambda_4-1)}}. \quad (48)$$

where  $u = Q^{-1}(x)$ .

The GLD is enormously flexible in terms of the shape of the distribution. For example, as shown in Figure 19, the uniform, Gaussian, exponential and Gamma distribution are all special cases of GLD. Effectively  $\lambda_1$  determines the location of the distribution,  $\lambda_2$  determines its scale, while  $\lambda_3$  and  $\lambda_4$  determine other shape characteristics. In particular, only if  $\lambda_3 = \lambda_4$  the distribution is symmetric.

Because GLD has 4 parameters, all we need are four moments – the mean, variance, skewness and kurtosis – of the PSO's sampling distribution in order to identify such parameters with the moment-matching method described above.

As an illustration, we apply this technique to reconstruct the sampling distribution during stagnation of a canonical PSO with parameters  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = 0$ ,  $\hat{y} = 10$  and  $\Omega = 5$  (the same parameters as in Figure 8). In Figure 20 we show snapshots at times  $t = 0, 2, 4, 12$  and  $24$  of the theoretical sampling distribution together with estimates of the distribution based on 1,000,000 actual runs. In all cases the match between the moments of the GLD and those of the PSO sampling distribution was exact (within experimental errors). Also, there is considerable agreement between the theoretical lines and histograms obtained in real runs. Note how widely the mean of the density function oscillates in the first few generations. Also note the asymmetry in the distributions due to the oscillations of the skewness.

## 10 Discussion, future work and conclusions

Several theoretical analyses of the dynamics of particle swarms have been offered in the literature over the last decade. These have been very illuminating. However, virtually all have relied on substantial simplifications,

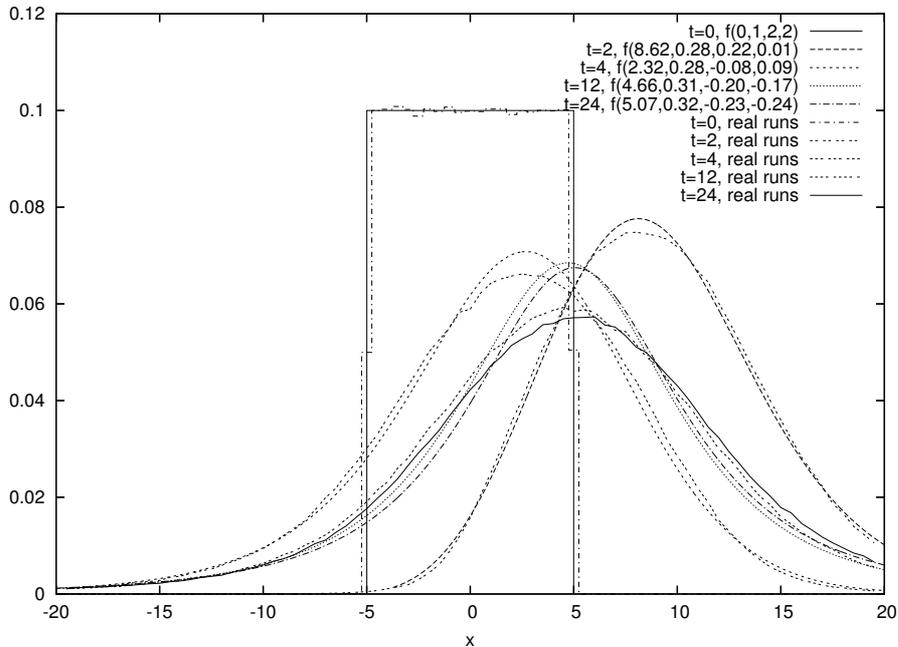


Figure 20: Estimates of the sampling distribution of a canonical PSO with parameters  $c = 1.49618$ ,  $w = 0.7298$ ,  $y = 0$ ,  $\hat{y} = 10$  and  $\Omega = 5$  during stagnation, reconstructed via GLD best fitting vs. histograms over 1,000,000 real runs. Snapshots at times  $t = 0, 2, 4, 12$  and  $24$  are shown. For each theoretical sampling distribution we report the parameters of the corresponding GLD rounded to the second decimal figure.

and on the assumption that the particles are deterministic. Naturally, these simplifications make it impossible to derive an exact characterisation of the sampling distribution of the PSO. This distribution has, therefore, remained, so far, the “holy grail” of PSO research.

By using of surprisingly simple techniques, in this paper we started by exactly determining perhaps the most important characteristic of a PSO’s sampling distribution, its variance, and we have been able to explain how it changes change over any number of generations. The only assumption we made is stagnation, so our characterisation is valid for as long as a particle searches for a better personal best.

Knowing the dynamics of the variance of the PSO’s sampling distribution and being able to control it, as, for example, we illustrated in Section 4, is very important because it allows one to understand the search behaviour of the PSO and adapt it to a problem at hand.

The dynamics of the variance of the PSO’s sampling distribution is also important from a theoretical standpoint. In order for a particle to converge, it is not enough to require  $\lim_{t \rightarrow \infty} E[x_t] = p$ : we must also have  $\lim_{t \rightarrow \infty} StdDev[x_t] = 0$ . In the absence of accurate information on  $StdDev[x_t]$ , previous research has effectively assumed that  $\lim_{t \rightarrow \infty} E[x_t] = p$  would eventually drive  $StdDev[x_t]$  to zero. This assumption has, for example, been used in the proof provided in [6] and [2] that the PSO is not guaranteed to be an optimiser. However, as we have shown in this work,  $\lim_{t \rightarrow \infty} StdDev[x_t] = 0$  only if  $y = \hat{y}$ , and so whether or not the PSO is an optimiser is still effectively a conjecture. How could our results help obtain a formal proof of convergence? The stagnation assumption essentially removes the dependence on the details of the fitness function. Our results can be used to prove convergence when stagnation has occurred. So, a proof of convergence for the PSO would require finding under which conditions and for what fitness functions the system stagnates. We will pursue this line of attack in future research.

After applying our analysis to first and second order statistics of the sampling distribution, we then moved on to study higher order statistics (Sections 5–7). In particular we analysed in detail the skewness and kurtosis of the distribution. Because of the complexity of the calculations involved, this required mechanising the derivation of the recursions for these moments.

We applied the analysis to the PSO with inertia weight, but, as we explained in Section 1, the analysis is also valid for the PSO with constriction via a simple parameter mapping.

We also generalised our model so as to include FIPS. This made it possible to explicitly compare the stability of different forms of PSO, leading to a deeper understanding of their properties. In particular, we

showed that, while FIPS and standard forms of PSO present exactly the same order-1 stability, in FIPS higher order moments are more stable than in the other PSOs, and we were able to explain why this is the case using two forms of central limit theorem (Section 8).

Finally, with all these tools in hand, we went in search for the “holy grail” – the actual PSO sampling density function. We treated the problem as an ill-posed inverse problem, which we regularised thanks to the use (and best fit) of a family of distributions – the Generalised Lambda Distribution (GLD). All empirical evidence we have suggests that this distribution approximates very closely the sampling behaviour of PSOs. So, much so that one would be tempted to try to prove that, indeed, the PSO’s sampling distribution is always and exactly a GLD, albeit, naturally, with parameters that are functions of time, i.e.,  $\lambda_i = \lambda_i(t)$ . We will explore this issue in future research.

Whether or not GLD is the exact PSO sampling distribution or just a very good approximation, if one could determine (again either exactly or approximately) how the  $\lambda_i(t)$ ’s are affected by the parameters  $c$ ,  $w$ ,  $y$  and  $\hat{y}$  and by the initial conditions  $x_0$  and  $v_0$ , it would then be possible to accurately simulate the behaviour of the PSO by sampling from  $f(x; \lambda_1(t), \lambda_2(t), \lambda_3(t), \lambda_4(t))$ . This is easily done since GLD deviates can trivially be produced by picking  $u$  uniformly at random in the interval  $[0, 1]$  and applying Equation (47), i.e.,  $Q(U[0, 1])$  is Generalised Lambda distributed. We will study this more sophisticated form of bare-bones PSO in future research.

## Acknowledgements

The author would like to acknowledge support by EPSRC XPS project (GR/T11234/01) and help and comments from Maurice Clerc and David Broomhead. Susanna Wau Men Au-Yeung (Department of Computing, Imperial College) and Roberto Reno’ (Dipartimento di Economia Politica, Universita’ di Siena) for source code and help with fitting GLD to data.

## References

- [1] Kennedy, J. (1998). The behavior of particles. In V. W. Porto, N. Saravanan, D. Waagen, and A. E. Eiben, Eds. *Evolutionary Programming VII: Proc. 7<sup>th</sup> Ann. Conf. on Evolutionary Programming Conf.*, San Diego, CA, 581–589. Berlin: Springer-Verlag.
- [2] Engelbrecht, A. P. (2005) *Fundamentals of Computational Swarm Intelligence*. Wiley, November 2005.
- [3] Ozcan, E. and Mohan, C. K. (1998). Analysis of a simple particle swarm optimization system. *Intelligent Engineering Systems Through Artificial Neural Networks* Vol. 8, pp. 253–258.
- [4] Ozcan, E., and Mohan, C. (1999). Particle swarm optimization: surfing the waves. *Proc. 1999 Congress on Evolutionary Computation*, 1939–1944. Piscataway, NJ: IEEE Service Center.
- [5] Clerc, M., and Kennedy, J. (2002) The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Transaction on Evolutionary Computation*, 6(1):58–73, February 2002.
- [6] van den Bergh, F. (2002) *An Analysis of Particle Swarm Optimizers*. PhD thesis, Department of Computer Science, University of Pretoria, Pretoria, South Africa.
- [7] K. Yasuda, A. Ide, and N. Iwasaki. Adaptive particle swarm optimization. *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, 2:1554–1559, 2003.
- [8] Iwasaki, N., and Yasuda, K. (2005) Adaptive particle swarm optimization using velocity feedback. *International Journal of Innovative Computing, Information and Control*, 1(3):369–380, September.
- [9] Blackwell, T. M. (2003) Particle swarms and population diversity I : Analysis. In Alwyn M. Barry, editor, *GECCO 2003: Proceedings of the Bird of a Feather Workshops, Genetic and Evolutionary Computation Conference*, pages 103–107, Chigaco. AAAI.
- [10] Brandstatter, B., and Baumgartner, U. (2002) Particle swarm optimization-mass-spring system analogon. *Magnetics, IEEE Transactions on*, 38(2):997–1000, 2002.

- [11] Trelea, I. C. (2003) The particle swarm optimization algorithm: convergence analysis and parameter selection. *Information Processing Letters*, 85(6):317–325.
- [12] Campana, E.F., Fasano, G., and Pinto, A. (2006) Dynamic system analysis and initial particles position in particle swarm optimization. In *IEEE Swarm Intelligence Symposium*, Indianapolis.
- [13] Campana, E.F., Fasano, G., Peri, D., and Pinto, A. (2006) Particle swarm optimization: Efficient globally convergent modifications. In C. A. Mota Soares et al., editor, *Proceedings of the III European Conference on Computational Mechanics, Solids, Structures and Coupled Problems in Engineering*, Lisbon, Portugal.
- [14] Kennedy, J. and Mendes, R. (2002). Population structure and particle swarm performance. *IEEE Congress on Evolutionary Computation, 2002* Honolulu, Hawaii USA.
- [15] Mendes, R. (2004). Population Topologies and Their Influence in Particle Swarm Performance. PhD thesis, Departamento de Informatica, Escola de Engenharia, Universidade do Minho, 2004.
- [16] Clerc, M. (2006) Stagnation analysis in particle swarm optimisation or what happens when nothing happens. Technical Report CSM-460, Department of Computer Science, University of Essex, August 2006. Edited by Riccardo Poli. Also available from <http://clerc.maurice.free.fr/pso/>
- [17] Kadirkamanathan, V., Selvarajah, K., and Fleming, P. J. (2006) Stability analysis of the particle dynamics in particle swarm optimizer. *IEEE Trans. Evolutionary Computation*, 10(3):245–255.
- [18] Poli, R., Langdon, W. B., Clerc, M., and Stephens, C. R. (2007) Continuous optimisation theory made easy? Finite-element models of evolutionary strategies, genetic algorithms and particle swarm optimizers. Proceedings of the Foundations of Genetic Algorithms (FOGA) workshop. (Also available as Technical Report CSM-463, Department of Computer Science, University of Essex.)
- [19] Kennedy, J. (2003) Bare bones particle swarms. *Proceedings of the IEEE Swarm Intelligence Symposium*, 80–87. Indianapolis, IN.
- [20] T. M. Blackwell. Particle swarms and population diversity. *Soft Computing*, 9:793–802, 2005.
- [21] AuYeung, S. W. M. (2003) Finding Probability Distributions from Moments, Master’s Thesis, Imperial College, London.
- [22] Lakhany, A. and H. Mausser (2000), Estimating the Parameters of the Generalized Lambda Distribution, *Algo Research Quarterly*, 3(3): 47- 58.
- [23] Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R. and Mykytka, E. F. (1979), A probability distribution and its uses in fitting data, *Technometrics*, 21(2): 201–214.