On the Moments of the Sampling Distribution of Particle Swarm Optimisers

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

ABSTRACT

A method is presented that allows one to exactly determine all the characteristics of a PSO's sampling distribution and explain how it changes over time, in the presence stochasticity. The only assumption made is stagnation (particles are in search for a better personal best).

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search

General Terms

Algorithms, Performance

Keywords

Particle Swarm Optimisation, Theory

1. INTRODUCTION

We start by considering the basic form of PSO with inertia weight, which is controlled by the following two equations

$$v^i_{t+1} = wv^i_t + \phi_1 \otimes (y^i - x^i_t) + \phi_2 \otimes (\hat{y} - x^i_t) \quad (1)$$

$$x^i_{t+1} = x^i_t + v^i_{t+1} \quad (2)$$

where ϕ_i represents a vector of random numbers uniformly distributed in $[0, c_i]$ and $\otimes$ is component-wise multiplication. Traditionally this model (and the related PSO with constriction) has been studied theoretically under strong simplifying assumptions such as isolated single individuals, search stagnation (i.e., no improved solutions are found) and absence of randomness [3, 4, 5, 6, 2, 7, 20, 10, 11, 12, 13].

Only very few attempts to understand the behaviour of the PSO in the presence of stochasticity have been made. For example, Clerc [16] analysed the distribution of velocities of one particle controlled by the standard PSO update rule with inertia and stochastic forces and was able to show that a particle's new velocity is the sum of three components: a forward force, a backward force and noise. Kadirkamanathan *et al.* [17] were able to study the stability of particles in the presence of stochasticity by using Lyapunov stability analysis. However, it is fair to say that for a very long time very little has been known regarding the shape of the sampling distribution of particles and its changes over time.

In recent work[24] we introduced a novel method, which allowed us to exactly determine the mean and standard deviation of the sampling distribution of the canonical PSO as well as their changes over any number of generations. The only assumption we made was stagnation, i.e., we studied the sampling distribution produced by particles in search for a better personal best.

In this paper we generalise the technique and apply it to determine higher order moments of the sampling distribution of PSOs.

We start with a summary of the results presented in[24].

2. FIRST AND SECOND MOMENTS OF THE PSO SAMPLING DISTRIBUTION

During stagnation each particle 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 rewrite them as

$$x_{t+1} = x_t(1 + w) - x_t(\phi_1 + \phi_2) - wx_{t-1} + \phi_1 y + \phi_2 \hat{y} \quad (3)$$

where we used of the relation $v_t = x_t - x_{t-1}$. We apply the expectation operator to both sides of the equation obtaining

$$\begin{aligned} E[x_{t+1}] &= E[x_t](1 + w) - E[x_t](E[\phi_1] + E[\phi_2]) \\ &- wE[x_{t-1}] + E[\phi_1]y + E[\phi_2]\hat{y} \end{aligned} \quad (4)$$

where we performed the substitution $E[x_t\phi_i] = E[x_t]E[\phi_i]$ because of the statistical independence between ϕ_i and x_t. Assuming that the ϕ_i's are uniformly distributed in $[0, c]$ we have $E[\phi_i] = \frac{c}{2}$ and, so,

$$E[x_{t+1}] = E[x_t](w' - c) - wE[x_{t-1}] + c\frac{y + \hat{y}}{2} \quad (5)$$

where we renamed $(1 + w) = w'$.

Let us now compute x^2_{t+1}. From Equation (3) we obtain:

$$x^2_{t+1} = (x_t w' - x_t\phi_1 - x_t\phi_2 - wx_{t-1} + \phi_1 y + \phi_2 \hat{y})^2 \quad (6)$$

Expanding and applying the expectation operator to both

GECCO'07, July 7–11, 2007, London, England, United Kingdom.

sides of the equation, one obtains

$$\begin{aligned} & E[x_{t+1}^2] \\ &= E[x_t^2]\left(w'^2 - 4\mu w' + 2\nu + 2\mu^2\right) \\ &+ E[x_{t-1}x_t]\left(-2ww' + 4w\mu\right) \\ &+ E[x_{t-1}^2]\left(w^2\right) \\ &+ E[x_t]\left(2\mu y w' + 2\mu\hat{y}w' - 2\nu y - 2\mu^2\hat{y} - 2\mu^2 y - 2\nu\hat{y}\right) \\ &+ E[x_{t-1}]\left(-2\mu y w - 2\mu\hat{y}w\right) \\ &+ \nu y^2 + 2\mu^2 y\hat{y} + \nu\hat{y}^2 \end{aligned} \quad (7)$$

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

If we multiply both sides of Equation (3) by x_t and apply the expectation operator we obtain

$$E[x_{t+1}x_t] = E[x_t^2](w' - c) - wE[x_t x_{t-1}] + E[x_t]c\frac{y+\hat{y}}{2}. \quad (8)$$

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

$$\begin{cases} E[x_{t+1}] &= E[x_t](w' - c) - wE[x_{t-1}] + c\frac{y+\hat{y}}{2} \\ E[x_{t+1}^2] &= E[x_t^2]\left(w'^2 - 4\mu w' + 2\nu + 2\mu^2\right) + \\ & \quad E[x_{t-1}x_t]\left(-2ww' + 4w\mu\right) + \\ & \quad E[x_{t-1}^2]\left(w^2\right) + \\ & \quad 2E[x_t](y+\hat{y})\left(\mu w' - \nu - \mu^2 y\right) - \\ & \quad 2w\mu E[x_{t-1}](y+\hat{y}) + \nu y^2 + 2\mu^2 y\hat{y} + \nu\hat{y}^2 \\ E[x_{t+1}x_t] &= E[x_t^2](w' - c) - wE[x_t x_{t-1}] + E[x_t]c\frac{y+\hat{y}}{2} \end{cases} \quad (9)$$

This system of equations can be written in matrix notation as an extended first order system obtaining

$$\mathbf{z}(t+1) = M\mathbf{z}(t) + \mathbf{b} \quad (10)$$

where

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

and the matrix M and the forcing vector $\mathbf{b}$ are given in Figure 1.

It is then trivial to verify under what conditions $E[x_t]$, $E[x_t^2]$ and $E[x_t x_{t-1}]$ 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*.

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

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

When the system is order-2 stable, by the simple change of variables $\tilde{\mathbf{z}}(t) = \mathbf{z}(t) - \mathbf{z}^*$, we can 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

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

3. HIGHER-ORDER MOMENTS

In the previous section we obtained 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 (3) as we did to derive Equation (6)? Generally, the r.h.s. we would be a sum of terms of the form

$$a_0\, x_t^{a_1}\, x_{t-1}^{a_2}\, w^{a_3}\, \phi_1^{a_4}\, \phi_2^{a_5}\, y^{a_6}\, \hat{y}^{a_7} \quad (11)$$

where a_k are suitable constants. Naturally, taking powers of Equation (3) and then multiplying both sides by some power of x_t would also lead to equations involving terms such as those in Equation (11). That is, for any choice of $b_1 \in \mathbb{N}$ and $b_2 \in \mathbb{N}$,

$$x_{t+1}^{b_1} x_t^{b_2} = \sum_i a_{0_i}\, x_t^{a_{1_i}}\, x_{t-1}^{a_{2_i}}\, w^{a_{3_i}}\, \phi_1^{a_{4_i}}\, \phi_2^{a_{5_i}}\, y^{a_{6_i}}\, \hat{y}^{a_{7_i}} \quad (12)$$

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

$$E[x_{t+1}^{b_1} x_t^{b_2}] = \sum_i a_{0_i}\, w^{a_{3_i}} y^{a_{6_i}}\, \hat{y}^{a_{7_i}}\, E[\phi_1^{a_{4_i}}]\, E[\phi_2^{a_{5_i}}]\, E[x_t^{a_{1_i}}\, x_{t-1}^{a_{2_i}}] \quad (13)$$

where we used the independence of ϕ_1, ϕ_2, $x_t x_{t-1}$ and, of course, their powers. Because ϕ_j is uniformly distributed in the range $[0, c]$, it is easy to verify that

$$E[\phi_j^n] = \frac{c^n}{n+1}. \quad (14)$$

So,

$$E[x_{t+1}^{b_1} x_t^{b_2}] = \sum_i \omega_i\, E[x_t^{a_{1_i}}\, x_{t-1}^{a_{2_i}}] \quad (15)$$

where

$$\omega_i = \left(\frac{a_{0_i}\, w^{a_{3_i}}\, c^{a_{4_i}+a_{5_i}}\, y^{a_{6_i}}\, \hat{y}^{a_{7_i}}}{(1+a_{4_i})(1+a_{5_i})}\right). \quad (16)$$

It is important to note here that, because Equation (3) is linear in x_t and x_{t-1}, all the terms on the r.h.s. of Equations (12) and (15) 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 (15) 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 (9). 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

$$Q(n) = \frac{n \times (n+1)}{2} \quad (17)$$

second order difference equations. Since, after expansion, the r.h.s. of Equation (3) contains 7 atomic terms of the form in Equation (11), the r.h.s. of Equation (15) contains 7^{b_1} terms.[1] So, the total number of terms one needs to

[1]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.

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

Figure 1: State update matrix M and forcing vector b (see Equation (10)).

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 1.36×7^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 (11). 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 ω_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 ω_i's in Equation (15) 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 (6) and (7)). Also, given their size, one will normally want to study (e.g., integrate) Equations (15) numerically. In this case w, c, y and $\hat{y}$ are all numerical parameters. So, the ω_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 (10), via the choice in Figure 2.

This effectively means adding artificial update equations of the form $E[x_t^k] = E[x_t^k]$ for $k = 1,\cdots,n$, 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 14×14 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

$$E[x_0^k] = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ \frac{\Omega^k}{k+1} & \text{otherwise.} \end{cases} \tag{18}$$

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

$$x_1 = x_0 + wv_0 - x_0(\phi_1+\phi_2) + \phi_1 y + \phi_2\hat{y} \tag{19}$$

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 (15), one obtains the desired expressions for $E[x_1^k x_0^l]$. Like for Equation (15), these expressions contain a number of terms that grows exponentially for with n. However, this process, too, can be trivially mechanised.

4. 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 easier comparison we show the lines where $\Lambda_m = 1$ for M_1, M_2, M_3 and M_4 in Figure 3 (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.

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 4(top) shows a comparison between the skewness $E[(x_t-\mu_t)^3]/\sigma_t^3$ computed

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

Figure 2: Extended state variable for analysis of higher order moments.

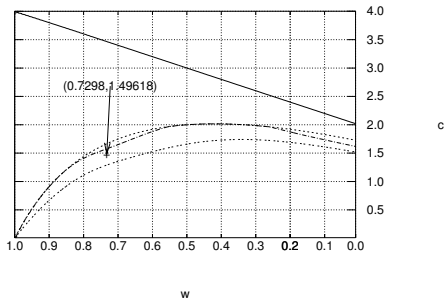

Figure 3: Plot of the regions of order-1, -2, -3 and -4 stability for the canonical PSO.

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 4(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.[2] 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 4(top) appear to grow bigger and bigger, suggesting instability, this is actually only a transient effect, as shown in Figure 5 where we integrate the equations over 200 generations instead of 30.

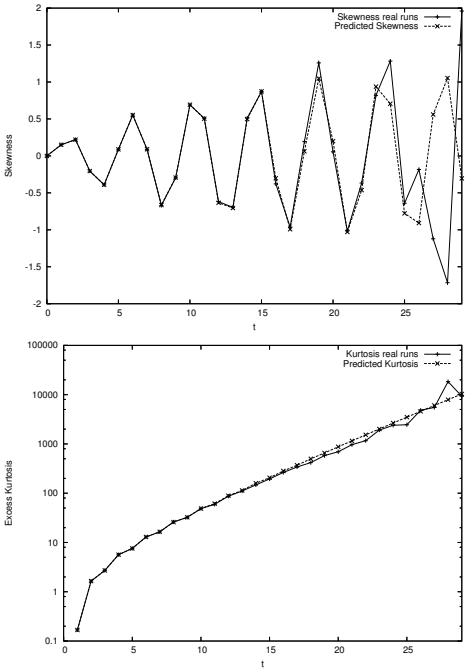

Figure 4: Comparison between predicted and experimental skewness and (excess) kurtosis of the PSO sampling distribution for $c = 1.49618$, $w = 0.7298$, $y = 0$, $\hat{y} = 1$ and $\Omega = 5$. Kurtosis grows exponentially and, so, it is plotted on a logarithmic scale. The first point is not plotted because the excess kurtosis was negative (-1.2).

5. COMPARISON BETWEEN 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 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 wander 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

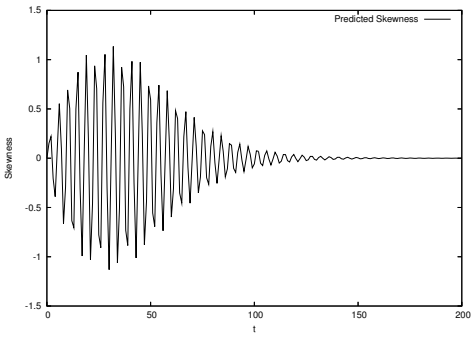

Figure 5: Predicted dynamics of the skewness of the PSO sampling distribution over 200 generations for $c = 1.49618$, $w = 0.7298$, $y = 0$, $\hat{y} = 1$ and $\Omega = 5$.

[2]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 to 0.

FIPS-like general class of PSOs described by the following difference equation

$$x_{t+1} = x_t + wv_t + \sum_{i=1}^{m} \phi_i(\hat{y}_i - x_t) \tag{20}$$

where the ϕ_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

$$x_{t+1} = x_t(1 + w) - wx_{t-1} - \sum_{i=1}^{m} \phi_i x_t + \sum_i \phi_i \hat{y}_i \tag{21}$$

which is a generalisation of Equation (3). All of the steps we performed in Section 3 can be repeated for Equation (21). These lead to recursion of the form in Equation (15) with the only difference that the coefficients ω_i take the more general form

$$\omega_i = \left(\frac{a_{0_i}\, w^{a_{w_i}} c_1^{a_{c1_i}} \cdots c_m^{a_{cm_i}}\, \hat{y}_1^{a_{y1_i}} \cdots \hat{y}_m^{a_{ym_i}}}{\prod_{j=1}^{m}(1 + a_{cj_i})} \right) \tag{22}$$

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

Because Equation (21) contains $3 + 2 \times m$ terms, the complexity of the expansion now grows exponentially as $O\left((3 + 2 \times m)^n\right)$, 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, they are exactly of the same size for all PSO models within the class defined by Equation (20). Initial conditions can be found following the approach described in Section 3. 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 (20) 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 $c_3 = 0$;

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, ϕ_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 6–9 show the lines in the (w, c) plane where the magnitude of the largest eigenvalue of M_n, Λ_m, is 1 for $n = 1, 2, 3, 4$ and for the three PSO variants mentioned above. Let us analyse these figures in detail.

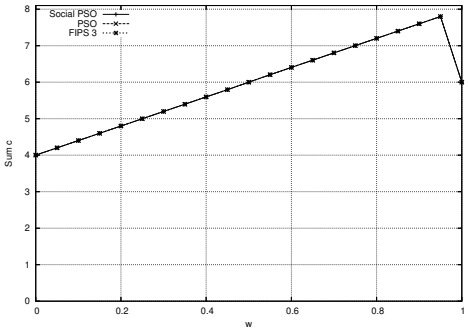

Figure 6: Lines below which the mean of the sampling distributions for a social PSO, a canonical PSO and FIPS with a neighbourhood of three individuals have a fixed point (order-1 stability). Note: the three lines coincide.

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:

$$E[x_{t+1}] = E[x_t]\left(1 + w - \frac{c}{2}\right) - wE[x_{t-1}] + \text{constant}, \tag{23}$$

where the constant term may differ in different PSO variants.[3] 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 the same happens with skewness (Figure 8) and kurtosis (Figure 9), 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 ϕ_i's, there are, the more a PSO should be unstable. However, the exact opposite happens. The social PSO, where the only influence is ϕ_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 (21) as follows

$$x_{t+1} = x_t(1 + w) - wx_{t-1} - x_t\Phi_m + \Psi_m \tag{24}$$

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

We know that $\sum_i c_i = c$. To simplify our treatment, let us further assume that the ϕ_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 Φ_m. This predicts that for sufficiently large m, the distribution of Φ_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 Φ_m and the stochasticity of Equation (24).

[3]This is irrelevant for the stability of the system, since stability is determined by the homogeneous part of the equation.

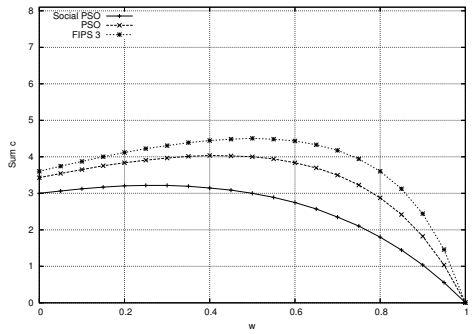

Figure 7: Lines below which the variance of the sampling distributions for a social PSO (bottom), a canonical PSO (middle) and FIPS with a neighbourhood of three individuals (top) have a fixed point (order-2 stability).

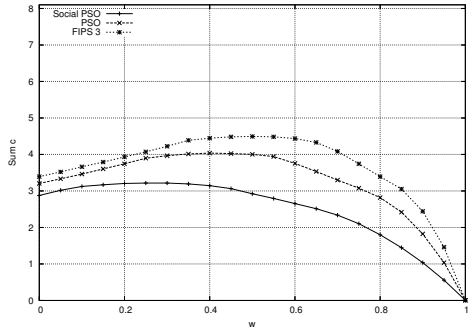

Figure 8: Lines below which a social PSO, a canonical PSO and FIPS3 are order-3 stable (from bottom to top, respectively).

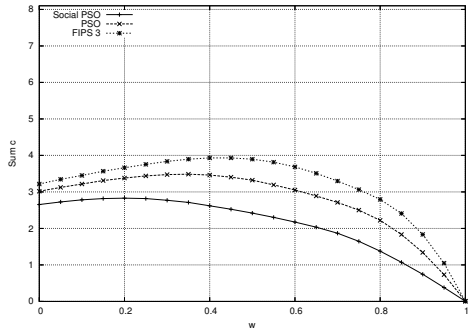

Figure 9: Lines below which the kurtosis of the sampling distributions for a social PSO (bottom), a canonical PSO (middle) and FIPS3 (top) have a fixed point (order-4 stability).

In the case of the stochastic variable Ψ_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 Ψ_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, $\lim_{m\to\infty} \frac{(\sum_{i=1}^m r_i^3)^{1/3}}{(\sum_{i=1}^m \sigma_i^2)^{1/2}} = 0$, must be satisfied, which, again, is the case since all $r_i^3 = 0$.

Then for sufficiently large m, also the distribution of Ψ_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 $\frac{c^2}{12\,m} \times \left(\frac{\sum_i \hat{y}_i^2}{m}\right)$. 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.[4] So, like for Φ_m, the larger m, the smaller the variance of Ψ_m, and, consequently the less the stochasticity of Equation (24).

Effectively the larger m the more Φ_m and Ψ_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.

6. 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 $f(x; \lambda_1, \lambda_2, ...)$ with parameters λ_1, λ_2, etc. with sufficient expressive power to represent distributions

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

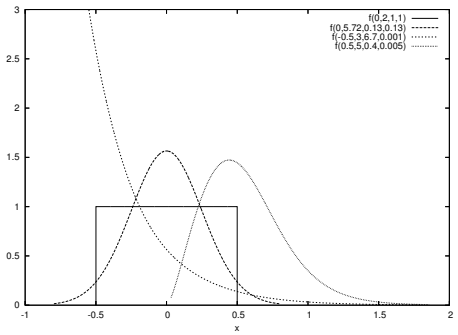

Figure 10: Examples of GLD density functions.

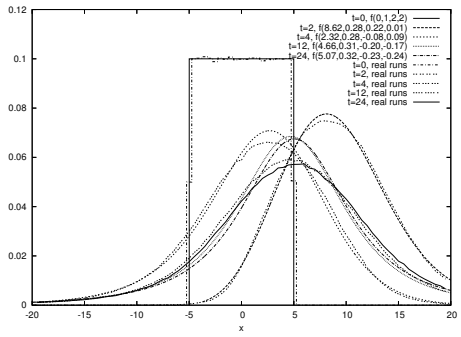

Figure 11: Estimates of the sampling distribution of a canonical PSO with parameters $c = 1.49618$, $w = 0.7298$, $y = 0$, $\hat{y} = 10$ and $\Omega = 5$ during stagnation, reconstructed via GLD best fitting vs. histograms over 1,000,000 real runs. Snapshots at times $t = 0$, 2, 4, 12 and 24 are shown. For each theoretical sampling distribution we report the parameters of the corresponding GLD.

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 λ_1, λ_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 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:

$$Q(u) = \lambda_1 + \frac{1}{\lambda_2}\left(\frac{u^{\lambda_3} - 1}{\lambda_3} + \frac{1 - (1-u)^{\lambda_4}}{\lambda_4}\right) \qquad (25)$$

where $u \in [0, 1]$. Its density function is given by

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

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

The GLD is enormously flexible in terms of the shape of the distribution. For example, as shown in Figure 10, the uniform, Gaussian, exponential and Gamma distribution are all special cases of GLD. Effectively λ_1 is determines the location of the distribution, λ_2 determines its scale, while λ_3 and λ_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 $\Omega = 5$. In Figure 11 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.

7. DISCUSSION 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, 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.

In previous work[24], by using of surprisingly simple techniques, 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. Here we extended this technique to the study of higher order statistics. 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 the analysis is also valid for the PSO with constriction, because of the well-known equivalence of these two models (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 5).

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 (25), 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

I would like to acknowledge support by EPSRC (GR/T11234/01) and help and comments from Maurice Clerc, David Broomhead, Susanna Wau Men Au-Yeung and Roberto Renó.

8. 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^{th} 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.

[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.

[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.

[24] Poli, R. and Broomhead, D. (2007), Exact Analysis of the Sampling Distribution for the Canonical Particle Swarm Optimiser and its Convergence during Stagnation, Genetic and Evolutionary Computation Conference (GECCO), London (accepted).