Mean and Variance of the Sampling Distribution of Particle Swarm Optimizers During Stagnation

Riccardo Poli

***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 characterization of the sampling distribution of the particle swarm optimizer (PSO). In this paper we introduce a novel method that allows us to exactly determine all the characteristics of a PSO 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.**

***Index Terms*—Particle swarm optimization, PSO theory, sampling distribution, stagnation.**

I. Introduction

LET US CONSIDER the basic form of particle swarm optimizer (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 nontrivial. 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 toward a new personal best position $\mathbf{y}^i$, or a new neighborhood best position $\hat{\mathbf{y}}$, or both. Thirdly, forces are stochastic. This prevents the use of standard mathematical tools used in the analysis of deterministic dynamical systems. Fourthly, the behavior 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 characterize 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., when no improved solutions are found) and, crucially, *absence of randomness*.

Manuscript received March 5, 2007; revised May 17, 2008 and September 16, 2008; accepted November 13, 2008. First version published July 28, 2009; current version published August 14, 2009. This work was supported by the EPRC XPS project GR/T1/234/01.

The author is with the School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, U.K. (e-mail: rpoli@essex.ac.uk).

Digital Object Identifier 10.1109/TEVC.2008.2011744

Algorithm 1 Classical PSO

Initialize a population array of particles with random positions and velocities on D dimensions in the problem space.
loop
  For each particle, evaluate the desired optimization fitness function in D variables.
  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 $\mathbf{y}^i$ equal to the current location $\mathbf{x}^i$ in D–dimensional space.
  Identify the particle in the neighborhood with the best success so far, and assign its position to the variable $\hat{\mathbf{y}}$.
  Change the velocity and position of the particle according to the following equations

$$\mathbf{v}^i_{t+1} = \omega \mathbf{v}^i_t + \boldsymbol{\phi}_{1,t} \odot (\mathbf{y}^i - \mathbf{x}^i_t) + \boldsymbol{\phi}_{2,t} \odot (\hat{\mathbf{y}} - \mathbf{x}^i_t) \quad (1)$$

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

  If a criterion is met, exit loop.
end loop

Note that $\boldsymbol{\phi}_{1,t}$ and $\boldsymbol{\phi}_{2,t}$ are vectors whose elements are random numbers uniformly distributed in $[0, c_i]$. New random vectors are drawn for each particle i and iteration t. The symbol $\odot$ represents componentwise multiplication.

For example, Ozcan and Mohan [3] studied the behavior of one particle, in isolation, in one dimension, in the absence of stochasticity and during stagnation. Also, $\mathbf{y}^i$ and $\hat{\mathbf{y}}$ were assumed to coincide, as is the case for the best particle in a neighborhood. The work was extended in [4] where multiple multidimensional particles were covered. Similar assumptions were used in Clerc and Kennedy's model [5]: one particle, one dimension, deterministic behavior, 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.

A similar approach was used by van den Bergh and Engelbrecht [6], [7] (see also [2]), who, again, modeled one particle, with no randomness and during stagnation. Inertia was also included in the model. As in previous work, van den Bergh provided an explicit solution for the trajectory of the particle. He showed that the particle is attracted toward a fixed point. [6] also suggested the possibility that particles may converge on a point that is neither the global optimum nor indeed a local optimum. This would imply that a PSO is not guaranteed to be an optimizer.

A simplified model of a particle was also studied by Yasuda *et al.* [8]. The assumptions were one 1-D particle, stagnation, absence of stochasticity, and presence of an inertia term. 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 behaviors are possible for a particle. Conditions for cyclic behavior were analyzed in detail.

Blackwell [21] 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 permitted. This allowed particles to interact, in the sense that they could change their personal best. Constriction was included but not stochasticity. [21] suggested that spatial extent decreases exponentially with time.

Brandstätter and Baumgartner [11] 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 [12] performed a lucid analysis of a four-parameter family of particle models and identified regions in the parameter space where the model exhibits qualitatively different behaviors (either stability, harmonic oscillations, or zigzagging behavior). Similar behavior maps had also been obtained in [6].

The dynamical system approach proposed by Clerc and Kennedy has recently been extended by Campana *et al.* [13], [14] who studied an extended PSO. The model assumes that no randomness is present.[1] In these conditions, the model is a discrete, linear, and stationary dynamical system, for which [13] and [14] 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 behavior of the PSO during phases of stagnation, Clerc [17] analyzed 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.* [18] were able to study the stability of particles *in the presence of stochasticity* by using Lyapunov stability analysis. They considered the behavior of a single particle—the swarm best—with inertia and during stagnation. By representing the particle as a nonlinear 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 behavior.

In summary, with very few exceptions, all mathematical models of PSO behavior 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 that allows one to exactly determine the moments of a PSO 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 of 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

$$\mathbf{v}^i_{t+1} = \chi \left(\mathbf{v}^i_t + \tilde{\boldsymbol{\phi}}_{1,t} \odot \left(\mathbf{y}^i - \mathbf{x}^i_t\right) + \tilde{\boldsymbol{\phi}}_{2,t} \odot \left(\hat{\mathbf{y}} - \mathbf{x}^i_t\right)\right) \quad (3)$$

which can be transformed into (1) via the mapping $\chi \to w$ and $\chi\tilde{\boldsymbol{\phi}}_{i,t} \to \boldsymbol{\phi}_{i,t}$. So, the theory applies to the PSO with constriction as well.

The paper is organized as follows. In Section II we derive recursions for the dynamics of first- and second-order statistics of the sampling distribution of a PSO particle during stagnation. We study the fixed points for these quantities and the PSO stability in Section III. In Section IV we show the results of numerically integrating the dynamic equations for the distribution's statistics. Finally, we provide some discussion, indications for future work, and our conclusions in Section VI.

II. 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. So, we can analyze each particle's behavior in isolation. Also, during stagnation, each dimension is independent. So, we can drop the superscripts i and the boldfaces (used to represent vectorial qualities) in (1) and (2). We then rewrite them as a single (second-order) difference equation, 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

$$x_{t+1} = x_t(1+w) - x_t(\phi_{1,t}+\phi_{2,t}) - wx_{t-1} + \phi_{1,t}y + \phi_{2,t}\hat{y}. \quad (4)$$

A. Dynamics of $E[x_t]$

Unlike previous research, we will not make the simplifying assumption that $\phi_{1,t}$ and $\phi_{2,t}$ are constant in (4). Instead, we treat them for what they are, i.e., uniformly distributed stochastic variables.[2] If we apply the expectation operator to

[1] Campana *et al.* introduce a general PSO which includes randomness [specifically, the variables $r^k_{h,j}$ in (6) in [13]]. However, they then go on to study a system where randomness is absent (they set $r^k_{h,j} = r_h$). That is, they assume that the random accelerations are identical for all particles (dropping the index j) and are constant for all iterations (dropping k).

[2] Although the variables $\phi_{1,t}$ for $t = 0, 1, \ldots$ are independent and identically distributed, they are distinct variables. The same applies for $\phi_{2,t}$. So, care must be taken when evaluating moments involving products and powers of such variables.

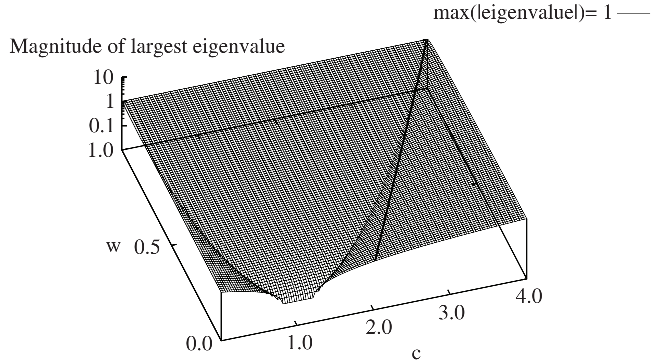

Fig. 1. Stability analysis based on the difference equation for $E[x_t]$ as a function of the parameters w and c. The straight line on the surface bounds the order-1 stability region.

both sides of the equation, we obtain

$$E[x_{t+1}] = E[x_t](1+w) - E[x_t](E[\phi_{1,t}] + E[\phi_{2,t}]) - wE[x_{t-1}] + E[\phi_{1,t}]y + E[\phi_{2,t}]\hat{y} \tag{5}$$

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

$$E[\phi_{1,t}] = \frac{c_1}{2} \qquad E[\phi_{2,t}] = \frac{c_2}{2} \tag{6}$$

and, so

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

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

$$p = \frac{c_1 y + c_2\hat{y}}{c_1+c_2}. \tag{8}$$

For the sake of simplicity let us now restrict our attention to the case $c_1 = c_2 = c$. Furthermore, let us rename $(1+w) = w'$. So

$$x_{t+1} = x_t w' - x_t\phi_{1,t} - x_t\phi_{2,t} - wx_{t-1} + \phi_{1,t}y + \phi_{2,t}\hat{y} \tag{9}$$

and

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

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. Fig. 1 plots the magnitude of the largest eigenvalue of the equation for $c = 0.01, 0.02, ..., 4.00$ and $w = 0.01, 0.02, \ldots, 1.0$. 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 $\phi_{1,t}$ and $\phi_{2,t}$ are constant and equal to their maximum value c, (9) would become

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

This equation has been studied extensively in previous research and has exactly the same form as (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 (10) has effectively been studied in previous research (e.g., [12], [6], and [5]; see also [2] for an extensive review). Indeed, the stable region depicted in Fig. 1 is exactly the same as reported in [12, Fig. 1(a)], and the explicit dynamics of $E[x_t]$ is explicitly given in previous work (e.g., [6]) if parameters are appropriately rescaled.

B. 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 (9) we obtain

$$\begin{aligned} x_{t+1}^2 &= (x_t w' - x_t\phi_{1,t} - x_t\phi_{2,t} - wx_{t-1} + \phi_{1,t}y + \phi_{2,t}\hat{y})^2 \\ &= x_t^2 w'^2 - x_t^2\phi_{1,t}w' - x_t^2\phi_{2,t}w' - wx_{t-1}x_t w' \\ &\quad + \phi_{1,t}yx_t w' + \phi_{2,t}\hat{y}x_t w' - x_t^2 w'\phi_{1,t} + x_t^2\phi_{1,t}^2 \\ &\quad + x_t^2\phi_{2,t}\phi_{1,t} + wx_{t-1}x_t\phi_{1,t} - \phi_{1,t}^2 yx_t \\ &\quad - \phi_{2,t}\hat{y}x_t\phi_{1,t} - x_t^2 w'\phi_{2,t} + x_t^2\phi_{1,t}\phi_{2,t} + x_t^2\phi_{2,t}^2 \\ &\quad + wx_{t-1}x_t\phi_{2,t} - \phi_{1,t}yx_t\phi_{2,t} - \phi_{2,t}^2\hat{y}x_t \\ &\quad - x_t w' wx_{t-1} + x_t\phi_{1,t}wx_{t-1} + x_t\phi_{2,t}wx_{t-1} \\ &\quad + w^2x_{t-1}^2 - \phi_{1,t}ywx_{t-1} - \phi_{2,t}\hat{y}wx_{t-1} \\ &\quad + x_t w'\phi_{1,t}y - x_t\phi_{1,t}^2 y - x_t\phi_{2,t}\phi_{1,t}y - wx_{t-1}\phi_{1,t}y \\ &\quad + \phi_{1,t}^2 y^2 + \phi_{2,t}\hat{y}y\phi_{1,t} + x_t w'\phi_{2,t}\hat{y} - x_t\phi_{1,t}\phi_{2,t}\hat{y} \\ &\quad - x_t\phi_{2,t}^2\hat{y} - wx_{t-1}\phi_{2,t}\hat{y} + \phi_{1,t}y\phi_{2,t}\hat{y} + \phi_{2,t}^2\hat{y}^2. \end{aligned} \tag{12}$$

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

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

where we set $\mu = E[\phi_{i,t}] = c/2$ and $\nu = E[\phi_{i,t}^2] = c^2/3$, for brevity.

As we discussed in Section II-A, we have a recursion (and in fact an explicit solution) for $E[x_t]$, so the recursion in (13) 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 (9) by x_t, obtaining

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

thereby

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

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 II-A.

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

$$\begin{cases} E[x_{t+1}] = E[x_t](w' - c) - wE[x_{t-1}] \\ \qquad + c\dfrac{y+\hat{y}}{2} \\ E[x_{t+1}^2] = E[x_t^2]\left(w'^2 - 4\mu w' + 2\nu + 2\mu^2\right) \\ \qquad + E[x_{t-1}x_t]\left(-2ww' + 4w\mu\right) \\ \qquad + E[x_{t-1}^2]\left(w^2\right) \\ \qquad + 2E[x_t](y+\hat{y})\left(\mu w' - \nu - \mu^2 y\right) \\ \qquad - 2w\mu E[x_{t-1}](y+\hat{y}) + \nu y^2 \\ \qquad + 2\mu^2 y\hat{y} + \nu\hat{y}^2 \\ E[x_{t+1}x_t] = E[x_t^2](w' - c) - wE[x_t x_{t-1}] \\ \qquad + E[x_t]c\dfrac{y+\hat{y}}{2}. \end{cases} \quad (16)$$

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

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

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

C. Initial Conditions

Let us evaluate the initial conditions for (16). To do so, we must specify how we perform the initialization of the particle swarm. As an example, let us consider the following very typical conditions: 1) a particle's initial position x_0 is chosen uniformly at random in a symmetric range $[-\Omega, \Omega]$; 2) 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

$$\begin{aligned} E[v_1] &= E[wv_0 + \phi_{1,t}(y - x_0) + \phi_{2,t}(\hat{y} - x_0)] \\ &= wE[v_0] + E[\phi_{1,t}](y - E[x_0]) \\ &\quad + E[\phi_{2,t}](\hat{y} - E[x_0]) \\ &= c\frac{y+\hat{y}}{2}. \end{aligned}$$

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

$$\begin{aligned} E[v_1 x_0] &= E[wv_0 x_0 + \phi_{1,t}(yx_0 - x_0^2) + \phi_{2,t}(\hat{y}x_0 - x_0^2)] \\ &= wE[v_0]E[x_0] + E[\phi_{1,t}](yE[x_0] - E[x_0^2]) \\ &\quad + E[\phi_{2,t}](\hat{y}E[x_0] - E[x_0^2]) \\ &= -cE[x_0^2] \\ &= -c\frac{\Omega^2}{3} \end{aligned}$$

resulting in $E[x_1 x_0] = (1 - c)(\Omega^2/3)$.

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

III. 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

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

and

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

$$\mathbf{b} = \begin{pmatrix} cp \\ 0 \\ \nu y^2 + 2\mu^2 y\hat{y} + \nu\hat{y}^2 \\ 0 \\ 0 \end{pmatrix}.$$

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 the 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

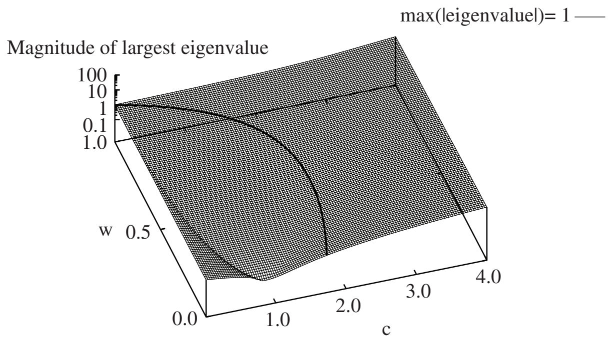

Fig. 2. Magnitude of the largest eigenvalue of M as a function of the parameters w and c when $y = -1$ and $\hat{y} = 1$. The curved line on the surface encloses the order-2 stable region.

$\mathbf{z}^*$. 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 $\tilde{\mathbf{z}}(t) = \mathbf{z}(t) - \mathbf{z}^*$, we can then represent the dynamics of the system via the 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, Fig. 2 shows a plot of Λ_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, Λ_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 Fig. 2, with Fig. 1. Note how the actual region of order-2 stability shown in Fig. 2 lays completely inside the region of order-1 stability obtained by analyzing $E[x_t]$ only (Fig. 1). Interestingly, by choosing parameters between the two curves, one obtains PSOs where $E[x_t] \to p$, but $StdDev[x_t]$ drifts to infinity (at a rate that depends on the magnitude of the eigenvalues; see Section IV for more on this), 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 $StdDev[x_t] \to 0$. In the following we clarify when this is the case.

Simple inspection of the equations in (16) reveals that the dynamics of $E[x_t]$ is independent of 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 $p = (y + \hat{y})/2$. 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 (16), obtaining

$$\begin{aligned} E[x_{t+1}^2] = {} & E[x_t^2]\left(w'^2 - 4\mu w' + 2\nu + 2\mu^2\right) \\ & + E[x_{t-1}x_t]\left(-2ww' + 4w\mu\right) \\ & + E[x_{t-1}^2]w^2 + 4p^2\left(\mu - \nu - \mu^2\right) \\ & + \nu y^2 + 2\mu^2 y\hat{y} + \nu\hat{y}^2 \end{aligned} \tag{19}$$

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

We know that if (c, w) are additionally within the convergence region for the system, shown in Fig. 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 (19) and (20) to obtain

$$\begin{aligned} p_{x^2} = {} & p_{x^2}\left(w'^2 - 4\mu w' + 2\nu + 2\mu^2\right) \\ & + p_{xx}\left(-2ww' + 4w\mu\right) \\ & + p_{x^2}w^2 + 4p^2\left(\mu - \nu - \mu^2\right) \\ & + \nu y^2 + 2\mu^2 y\hat{y} + \nu\hat{y}^2 \end{aligned} \tag{21}$$

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

The second equation allows us to compute

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

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

$$p_{x^2} = \frac{\begin{pmatrix} 4p^2\left(\mu - \nu - \mu^2\right) + \nu y^2 + 2\mu^2 y\hat{y} \\ +\nu\hat{y}^2 + p^2\frac{c}{w'}2w\left(2\mu - w'\right) \end{pmatrix}}{\Delta} \tag{24}$$

$$\begin{aligned} p_{xx} = {} & \left[\frac{\begin{pmatrix} 4p^2\left(\mu - \nu - \mu^2\right) + \nu y^2 + 2\mu^2 y\hat{y} \\ +\nu\hat{y}^2 + p^2\frac{c}{w'}2w\left(2\mu - w'\right) \end{pmatrix}}{\Delta}\right] \\ & \times \left(1 - \frac{c}{w'}\right) + p^2\frac{c}{w'} \end{aligned} \tag{25}$$

where

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

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

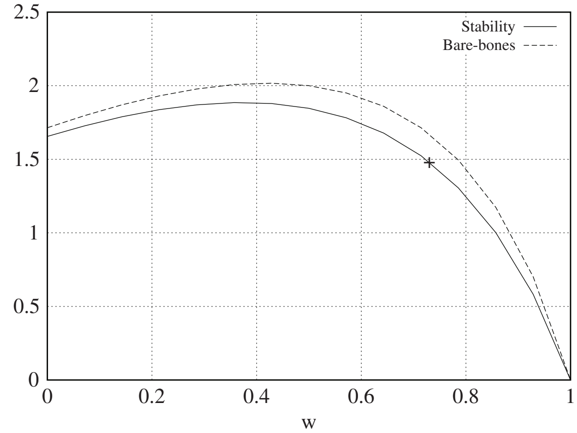

Fig. 3. Parameter values for which $p_{sd} = \left|\hat{y} - y\right|$ as predicted by the bare-bones model ("bare-bones" curve) and line below which the variance of the sampling distribution has a fixed point ("stability" curve).

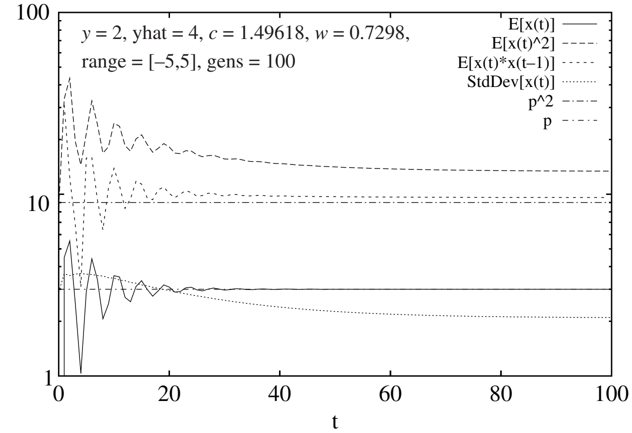

Fig. 4. Numerical integration of (16) for $c = 1.49618$, $w = 0.7298$, $y = 2$, $\hat{y} = 4$, and $\Omega = 5$.

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

With little algebra one can see that

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

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

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

which we can finally rewrite as

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

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

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 [20]—that the standard deviation of the search distribution is proportional to $\left|\hat{y} - y\right|$—was fundamentally correct. There is, however, a multiplicative factor $(1/2)\sqrt{2(\nu - \mu^2)/\Delta}$ in (29) that 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 one only when

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

As one can see from Fig. 3, which graphs all (c, w) pairs for which $p_{sd} = \left|\hat{y} - y\right|$, which are 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 $p_{sd} \approx 1.0428 \times \left|\hat{y} - y\right|$.

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 one (also shown as contour in Fig. 2). This is also plotted in Fig. 3. In this case the pair $(c, w) = (1.49618, 0.7298)$ lies 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.

IV. 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 $E[x_t x_{t-1}]$ (16).

We start (Fig. 4) by considering the case $c = 1.49618$ and $w = 0.7298$, which correspond 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, $1.0428 \times \left|\hat{y} - y\right| = 2.0856$) within about 70 generations. The picture is very different, however, if $y = \hat{y}$, as shown in Fig. 5. In this case, $E[x_t^2]$ and $E[x_t x_{t-1}]$ converge to p^2. As a result,

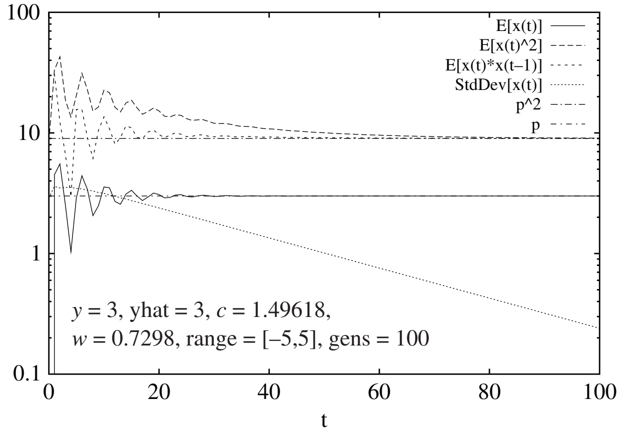

Fig. 5. Numerical integration of (16) for $c = 1.49618$, $w = 0.7298$, $y = \hat{y} = 3$, and $\Omega = 5$.

$StdDev[x_t]$ decreases to zero. The decrease is exponential, corroborating Blackwell's analysis of how the spatial extent of a particle swarm varies over time [21] (see Section I).

Examples of configurations where the mean converges to its fixed point while $StdDev[x_t]$ does not converge are shown in Fig. 6. Note that this is not necessarily an undesirable behavior. 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 at 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 (23) simplifies to $p_{xx} = p^2$. Then we have

$$\begin{aligned} p_{x^2} = {} & p_{x^2}\left(w'^2 - 4\mu w' + 2\nu + 2\mu^2\right) \\ & + p^2\left(-2ww' + 4w\mu\right) + p_{x^2}w^2 \\ & + 4p^2\left(\mu - \nu - \mu^2\right) + \nu y^2 + 2\mu^2 y\hat{y} + \nu\hat{y}^2 \end{aligned} \tag{31}$$

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

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

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 $\hat{y} = 9$, one obtains $p_{x^2} = 45$ 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 Fig. 7, there are indeed the asymptotes to which the system converges.

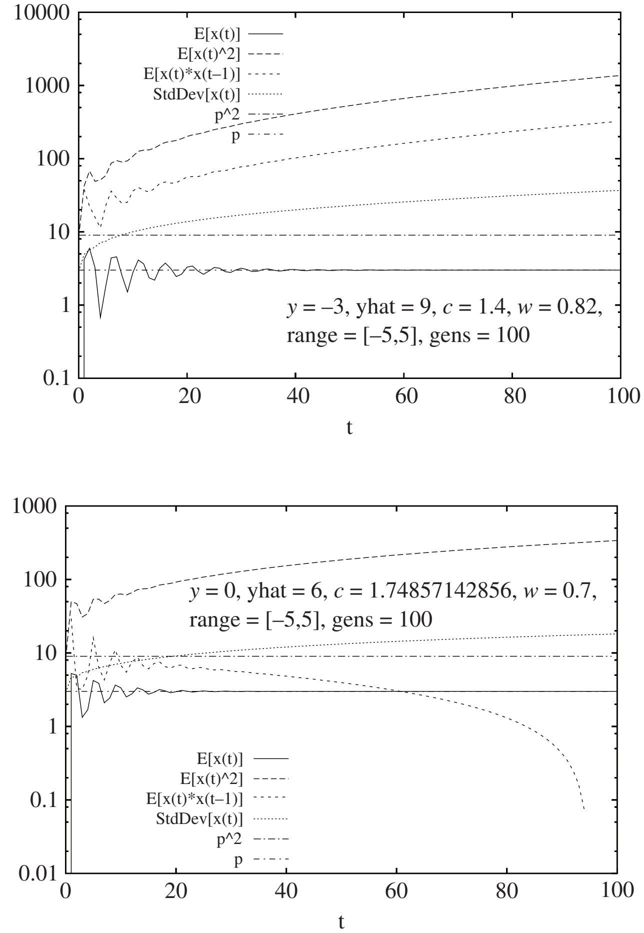

Fig. 6. Numerical integration of (16) for values of c and w where $E[x_t]$ is convergent but $StdDev[x_t]$ is not.

V. 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 behavior of the model with that observed in actual runs.

Note that (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 $\phi_{i,t}$. 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

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

where $\phi_{i,t}$ are stochastic variables uniformly distributed in the range $[0, c]$. 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 $\{x_t^2\}_{t=0}^{\infty}$ and $\{x_t x_{t-1}\}_{t=1}^{\infty}$ and (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

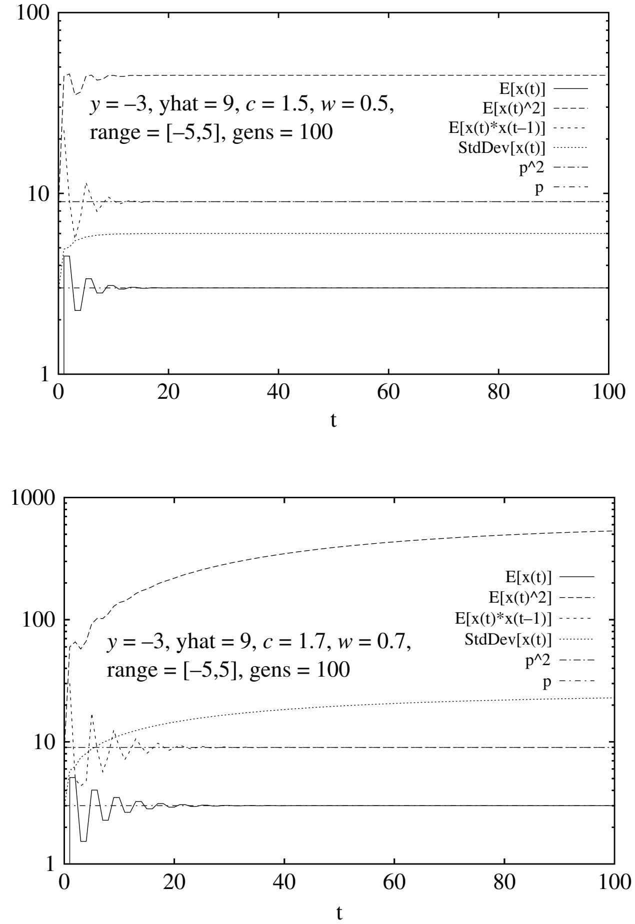

Fig. 7. Numerical integration of (16) for values of c and w where $E[x_t]$ is convergent to p, $E[x_t x_{t-1}]$ is convergent to p^2, and $E[x_t^2]$ is convergent to the value given in (32).

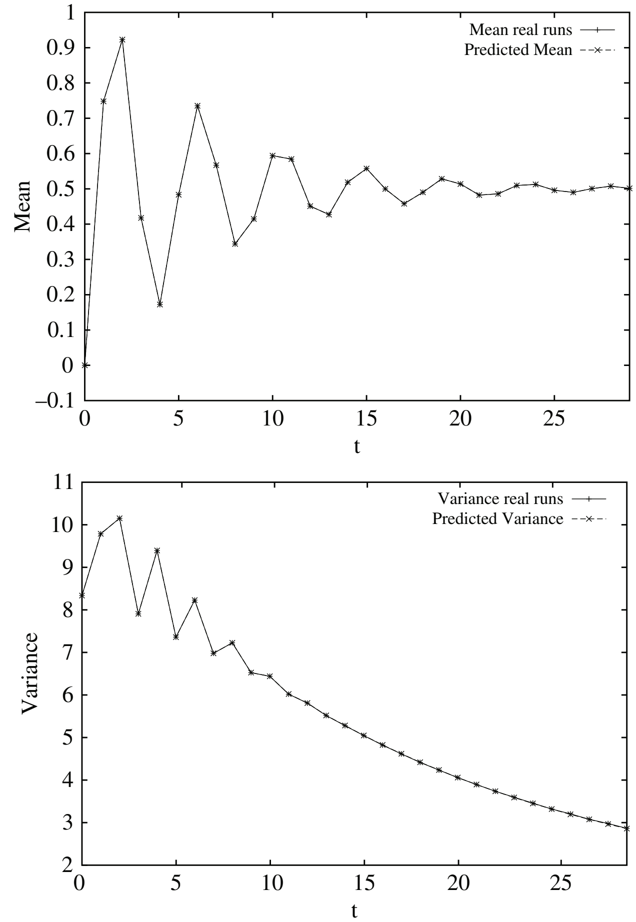

Fig. 9. Comparison between predicted and experimental means and variances of the PSO sampling distribution for $c = 1.49618$, $w = 0.7298$, $y = 0$, $\hat{y} = 1$, and $\Omega = 5$.

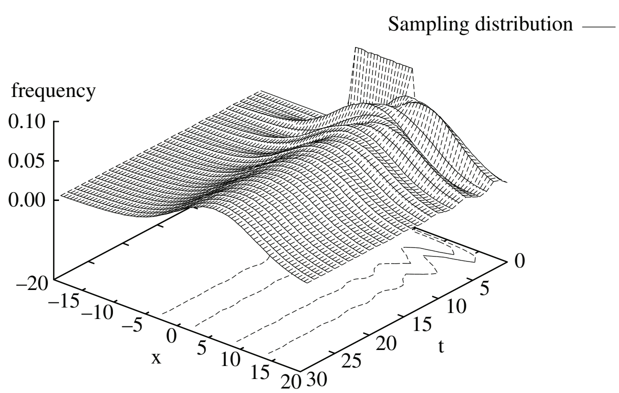

Fig. 8. Sampling distribution observed in real runs when $c = 1.49618$, $w = 0.7298$, $y = 0$, $\hat{y} = 10$, and $\Omega = 5$. Histograms based on 1 000 000 runs. Contours correspond to frequencies of 0.025 and 0.05.

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 1-D particle in stagnation. Because no fitness evaluation is needed, these can be done relatively quickly on an ordinary computer. Fig. 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 behavior of particles in real runs. As shown in Fig. 9 (bottom), the model also predicts exactly (within experimental errors) the behavior 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 toward its (nonzero) 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. Fig. 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 characterization of the behavior 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.

VI. Discussion, Future Work, and Conclusion

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 characterization of the sampling distribution of the PSO. This distribution has, therefore, remained, so far, one of the most important open questions of PSO research.

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

Knowing the dynamics of the variance of the PSOs sampling distribution and being able to control it, as, for example, we illustrated in Section IV, are very important because they allow one to understand the search behavior of the PSO and adapt it to a problem at hand.

The dynamics of the variance of the PSO sampling distribution is also important from a theoretical standpoint. In order for a particle to converge, it is not enough to require $\lim_{t\to\infty} E[x_t] = p$: we must also have $\lim_{t\to\infty} StdDev[x_t] = 0$. In the absence of accurate information on $StdDev[x_t]$, previous research has effectively assumed that $\lim_{t\to\infty} E[x_t] = p$ 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 optimizer. However, as we have shown in this paper, $\lim_{t\to\infty} StdDev[x_t] = 0$ only if $y = \hat{y}$, and so whether or not the PSO is an optimizer 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.

In this paper, we have clarified how the sampling distribution of a PSO changes over time. With the typical parameter settings used in many practical PSO applications, the system is under-dumped and the distribution presents oscillations. At the first oscillation of the mean, for example, the PSO effectively overshoots its target, which is the middle point between $\hat{y}$ and y. When this happens, essentially the PSO allocates new samples by performing some sort of extrapolation (similar to that performed in certain forms of crossover in real-valued genetic algorithms) based on direction of movement and two "good" points the PSO knows already, $\hat{y}$ and y. Often, the oscillation of the mean is accompanied by an oscillation of the skewness, which leads to an even more markedly extrapolating distribution. If an improved point is found, then the strategy pays off. However, the overshot is temporary. So, if an improvement is not found, the distribution will then swing back and the PSO will extrapolate in the opposite direction. This swinging process continues for a some time, but eventually the PSO settles onto a sample distribution that favors the region between $\hat{y}$ and y.

Clearly, the duration of the oscillatory phase, as well as the characteristics of the sampling distribution, particularly its variance, will influence performance. Whether it is better to try to extrapolate before giving up and for how long depend on the fitness function. Likewise, whether it is better to have a wide or a narrow sampling distribution depends on the fitness function. So, no conclusions can be drawn on the superiority of one mode of exploration over another.

The work presented in this paper allows one to set the parameters of a PSO in such a way to achieve the desired transient and long-term sampling behavior.

References

[1] J. Kennedy, "The behavior of particles," in *Proc. 7th Ann. Conf. Evol. Program.*, San Diego, CA, 1998, pp. 581–589.

[2] A. P. Engelbrecht, *Fundamentals Comput. Swarm Intell.*. New York: Wiley, 2005.

[3] E. Ozcan and C. K. Mohan, "Analysis of a simple particle swarm optimization system," in *Proc. Intell. Eng. Syst. Through Artificial. Neural Netw. (ANNIE '98)*, vol. 8. Oct. 1998, pp. 253–258.

[4] E. Ozcan and C. Mohan, "Particle swarm optimization: surfing the waves," in *Proc. 1999 Congr. Evol. Comput.*, Piscataway, NJ: IEEE Service Center, pp. 1939–1944.

[5] M. Clerc and J. Kennedy, "The particle swarm explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, Feb. 2002.

[6] F. van den Bergh, "Analysis of Particle Swarm Optimizers," Ph.D. dissertation, Dept. Comput. Sci., Univ. of Pretoria, Pretoria, South Africa, 2002.

[7] F. van den Bergh and A. P. Engelbrecht "A study of particle swarm optimization particle trajectories," *Inform. Sci.*, vol. 176, no. 8, pp. 937–971, Apr. 2006.

[8] K. Yasuda, A. Ide, and N. Iwasaki "Adaptive particle swarm optimization," in *Proc. IEEE Int. Conf. Sys., Man Cybern. 2003.*, vol. 2. Washington, D.C., pp. 1554–1559.

[9] N. Iwasaki and K. Yasuda, "Adaptive particle swarm optimization using velocity feedback," *Int. J. Innovative Comput., Inform. and Control*, vol. 1, no. 3, pp. 369–380, Sep. 2005.

[10] T. M. Blackwell, "Particle swarms and population diversity I: Analysis," in *Proc. Bird Feather Workshops, Genetic Evol. Comput. Conf. 2003*, Chigaco, IL: American Association for Artificial Intelligence, pp. 103–107.

[11] B. Brandstatter and U. Baumgartner, "Particle swarm optimization-mass-spring system analogon," *IEEE Trans. Magnetics*, vol. 38, no. 2, pp. 997–1000, Mar. 2002.

[12] I. C. Trelea, "The particle swarm optimization algorithm: Convergence analysis and parameter selection," *Inform. Process. Lett.*, vol. 85, no. 6, pp. 317–325, Mar. 2003.

[13] E. F. Campana, G. Fasano, and A. Pinto, "Dynamic system analysis and initial particles position in particle swarm optimization," in *Proc. IEEE Swarm Intell. Symp.*, Indianapolis, IN, 2006.

[14] E. F. Campana, G. Fasano, D. Peri, and A. Pinto, "Particle swarm optimization: Efficient globally convergent modifications," in *Proc. III Eur. Conf. Comput. Mech., Solids, Structures Coupled Problems Eng.*, Lisbon, Portugal, 2006.

[15] J. Kennedy and R. Mendes, "Population structure and particle swarm performance," in *Proc. IEEE Congr. Evol. Comput. 2002*, Honolulu, HI, pp. 1671–1676.

[16] R. Mendes, "Population topologies and their influence particle swarm performance," Ph.D. dissertation, Departamento de Informatica, Escola de Engenharia, Universidade do Minho, Graga, Portugal, 2004.

[17] M. Clerc. (2006, Aug.). Stagnation Analysis Particle Swarm Optimisation or What Happens When Nothing Happens. Dept. of Comput. Sci., Univ. of Essex, Colchester, U.K. Tech. Rep. CSM-460. [Online]. Available: http://clerc.maurice.free.fr/pso/

[18] V. Kadirkamanathan, K. Selvarajah, and P. J. Fleming, "Stability analysis of the particle dynamics in particle swarm optimizer," *IEEE Trans. Evol. Comput.,* vol. 10, no. 3, pp. 245–255, May 2006.

[19] R. Poli, W. B. Langdon, M. Clerc, and C. R. Stephens, "Continuous optimisation theory made easy? Finite-element models of evolutionary strategies, genetic algorithms and particle swarm optimizers," in *Proc. Foundations Genetic Algorithms Workshop*, Mexico City, Mexico, 2007, pp. 165–193.

[20] J. Kennedy, "Bare bones particle swarms," in *Proc. IEEE Swarm Intell. Symp.*, Indianapolis, IN, 2003, pp. 80–87.

[21] T. M. Blackwell, "Particle swarms and population diversity," *Soft Computing*, vol. 9, no. 11, pp. 793–802, Nov. 2005.

[22] S. W. M. AuYeung, "Finding probability distributions from moments," M.S. thesis, Imperial College of Science, Technology and Medicine, Department of Computing, London, U.K. 2003.

[23] A. Lakhany and H. Mausser, "Estimating the Parameters of the Generalized Lambda Distribution," *Algo Research Quarterly*, vol. 3, no. 3, pp. 47–58, Dec. 2000.

[24] J. S. Ramberg, E. J. Dudewicz, P. R. Tadikamalla, and E. F. Mykytka, "A probability distribution and its uses in fitting data," *Technometrics*, vol. 21, no. 2, pp. 201–214, May 1979.

[25] S. Zahl, "Bounds for the central limit theorem error," *SIAM J. Appl. Math.*, vol. 14, no. 6, pp. 1225–1245, Nov. 1966.

Riccardo Poli received the Ph.D. degree in biomedical image analysis.

He is a Professor in the School of Computer Science and Electronic Engineering, Essex. He started his academic career as an Electronic Engineer and later became an expert in the field of evolutionary computation. He has published around 250 refereed papers and two books on the theory and applications of genetic programming, evolutionary algorithms, particle swarm optimization, biomedical engineering, brain-computer interfaces, neural networks, image/signal processing, biology and psychology.

Prof. Poli is a Fellow of the International Society for Genetic and Evolutionary Computation since 2003, a recipient of the EvoStar award for outstanding contributions to this field in 2007, and an ACM SIGEVO executive board member from 2007 to 2013. He was co-founder and co-chair of the European Conference on Genetic Programming from 1998 to 2000, and 2003. He was general chair in 2004, track chair in 2002, 2007, and 2009, business committee member in 2005, and competition chair in 2006 of ACM's Genetic and Evolutionary Computation Conference, co-chair of the Foundations of Genetic Algorithms Workshop in 2002 and technical chair of the International Workshop on Ant Colony Optimization and Swarm Intelligence in 2006. He is an Associate Editor of *Genetic Programming and Evolvable Machines* and the *International Journal of Computational Intelligence Research*. He is an advisory board member of the *Evolutionary Computation journal* and the *Journal on Artificial Evolution and Applications* and an editorial board member of Swarm Intelligence. He is a member of the EPSRC Peer Review College, an EU expert evaluator and a grant-proposal referee for Irish, Swiss and Italian funding bodies.