Origin of Bursts

Tim Blackwell
Dept Computing
Goldsmiths, University of London
New Cross
London SE14 6NW, UK
tim.blackwell@gold.ac.uk

Daniel Bratton
Dept Computing
Goldsmiths, University of London
New Cross
London SE14 6NW, UK
dbratton@gmail.com

ABSTRACT

The phenomenon of particle bursts, a well-known feature of PSO is investigated. Their origin is concluded to lie in multiplicative stochasticity, previously encountered in the study of first order stochastic difference equations. The work here demonstrates that bursts contribute to fattening of the tail of the particle position distribution and that these tails are well described by power laws. It is argued that recombinant PSO, a competitive PSO variant without multiplicative randomness, is burst-free.

Categories and Subject Descriptors

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

General Terms

Algorithms

Keywords

Particle Swarm Optimisation, Theory

1. INTRODUCTION

The particles in a Particle Swarm Optimisation (PSO) [4] simulation undergo a guided flight through a search space of solutions. The aim is a successful optimisation of a given objective function f. Each particle i in the swarm has dynamic variables position and velocity, $\vec{x}_i$ and $\vec{v}_i$, and a memory $\vec{p}_i$ of a past position visited. Furthermore, each particle is embedded in a social, rather than spatial, neighbourhood N_i of informers. The algorithm is made of two essential steps: particle movement and memory update. Particles guide each other by means of the accelerations that govern position update. These accelerations are towards the best memory(s) of informers within N_i. Interaction between the particles is therefore mediated by the network, and not by direct interaction between the particles themselves. Despite

the simplicity of the scheme, the algorithm is effective over the standard benchmark problems and has increasing numbers of real-world applications. However, little is known about how a PSO achieves its results.

Algorithm 1 canonical PSO

---

0. initialise swarm
FOR EACH particle i
    randomise $\vec{v}_i, \vec{x}_i$, set $\vec{p}_i = \vec{x}_i$
FOR EACH particle i
   1. find neighbourhood best
     $\vec{p_{g(i)}} = arg\ min(f(p_j), j \in N_i)$
   FOR EACH dimension d
     2. move particle
       $F(p) : x_{id} \leftarrow x_{id}$
     3. update memory
       IF $f(\vec{x}_i) < f(\vec{p}_i)$ THEN
         $\vec{p}_i \leftarrow \vec{x}_i$
   END
END

PSO is summarised in Algorithm 1. The dynamic rule $F(p)$ for a single particle is:

$$\begin{cases} v_{id}(t+1) = wv_{id}(t) + \sum_{j=1}^{K} \Phi_j(t)(p_{jd} - x_{id}(t)) \\ x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \end{cases} \quad (1)$$

The sum in Eq. 1 is over K informers p_i and for each each dimension d. In standard PSO [2], $K = 2$ and p_i and p_{g_i} are the personal and neighbourbood best positions and

$$\Phi_{1,2} = \phi_{1,2}u_{1,2} \quad (2)$$

where $\phi_{1,2}$ are 'acceleration' constants and $u_{1,2} \sim U(0,1)$ are random numbers drawn from the uniform distribution on the unit interval. This 'inertia weight' (IW) formulation of PSO [18] derives from the parameter w which imitates inertia in the sense that it weights the tendency to move in a straight line at constant speed ($w = 1$) to the tendency to move erratically about the attractors Φ. Other formulations of PSO include Kennedy and Mendes' [8] fully informed particle swarm (FIPS), wherein a particle is influenced by $K > 2$ neighbours, $\Phi_j = \frac{1}{K}\phi_j u_j$ and the 'constricted' Clerc-Kennedy (CK) swarm [3] which is equivalent to Eq. 1 with the identification $\chi = w$, $\phi_{1,2} = \chi\phi_{1,2}^{CK}$

As mentioned, surprisingly little is known about how the method actually achieves its results. The complete model

is very difficult to analyse due to the stochasticity of the particle dynamics, and the relationship between this dynamics, the social network, and the underlying objective function. Progress can, however, be made with simplified models. Two stages of approximation may be considered: particle decoupling and the removal of stochasticity.

Particle decoupling renders the particles non-interacting, so that no memory update happens at step 3 of the algorithm. This state of affairs can effectively occur in the full interacting PSO when the swarm reaches a 'stagnant' configuration whereby further improvement of any p_i is impossible, or at least very unlikely. In this case the dynamics simplify to d independent update rules, so only the one-dimensional case needs to be considered. However, even the decoupled system is difficult to study. This is due to the stochastic nature of $F(p)$.

However, an even simpler model, a 'bare bones' appproximation to the decoupled system can be employed [6]. This replaces $F(p)$ by $x(t+1) = N(p, x(t))$ where N is a given probability distribution, for example, a Gaussian, or a Levy. Indeed, a bare bones, velocity-free, formulation can be considered as an effective PSO in its own right: comparable performance to canonical PSO over a standard suite of functions has been demonstrated for the Levy distribution [16]. A recent theoretical analysis of bare bones PSO has been given in [17].

Another approximation after decoupling, but retaining velocity, can be considered by the replacement of the random numbers u_j by fixed constants u. This removes stochasticity and renders the rule $F(p)$ amenable to stability analysis using established techniques for deterministic equations. In fact this is the chief theoretical result on PSO dynamics to date.

A summary of stability conditions for various formulations of PSO under the fixed-u approximation is presented in the following section. These results are important for the subsequent analysis developed in this paper. Section 3 presents a series of experiments which reveal power law tails of particle position in decoupled PSO. Section 4 introduces multiplicative stochasticity and power law tails in first order processes. These results are generalised to second order SDEs and applied to PSO. Section 5 discovers that recombinant PSO will not be subject to bursting since it only has additive stochasticity. This is intriguing because PSO-DR is remarkably competitive.

2. ANALYSIS OF NON-STOCHASTIC PSO

After decoupling, we only need to consider a single particle in one dimension, so particle labels i will be dropped. By virtue of $v(t) = x(t) - x(t-1)$, Eq. 1 can be rewritten as a stochastic second order difference equation (DE):

$$x(t+1) + a(t)x(t) + bx(t-1) = c(t) \qquad (3)$$

with

$$\begin{cases} a(t) = \sum_j \Phi_j(t) - w - 1 \\ b = w \\ c(t) = \sum_j \Phi_j(t)p_j \end{cases} \qquad (4)$$

where the parameters a and c are stochastic variables because of the presence of random numbers in Eq. 2. (However they are not independent because the same random numbers u_1 and u_2 appear in a and c).

First order stochastic difference equations (SDEs) and constant parameter first and second order DE's have been studied by many authors in a number of domains. A constant parameter second order difference equation, $a(t) = a$, $c(t) = a$, $\Phi(t) = \Phi$ is obtained by replacing the random variables $u_{1,2}$ by a constant u. Stability conditions can then be found by substituting the trial solution $x = \lambda^t$ into Eq. 3 and considering roots of the resulting characteristic equation $\lambda^2 + a\lambda + b = c$ (see for example, [5]). Stability then requires $|\lambda| < 1$. Complex, and therefore oscillatory, solutions are found if

$$\begin{cases} 0 < b < 1 \\ a^2 < 4b \end{cases} \qquad (5)$$

and stable real solutions for

$$\begin{cases} |a| < 1 + b \\ a^2 \geq 4b \end{cases} \qquad (6)$$

The stability conditions can be combined:

$$\begin{cases} |b| < 1 \\ |a| < 1 + b \end{cases} \qquad (7)$$

In terms of PSO parameters w and Φ (assumed constant) this gives

$$\begin{cases} |w| < 1 \\ 0 < \sum_j \Phi_j < 2(1 + w) \end{cases} \qquad (8)$$

In order to relate these stability conditions to the stochastic model Eqs 3 and 4, a number of authors have suggested replacing the random variable $u_{1,2}$ by its maximum, i.e. $u = 1$ [11], [22] [3]. Certainly this should ensure that a sequence of large u_i's would not constitute temporary instability, although it cannot guarantee very large fluctuations (see below). An alternative approach due to Poli [14] is to replace $u_{1,2}$ by the expected value, $< u > = \frac{1}{2}$, and consider expectations of first and second moments, leading to a stability relation for the mean and variance of x. The upper bound for $u = \frac{1}{2}$ on the acceleration parameters ϕ is twice the $u = 1$ bound.

Standard PSO is implemented with equal acceleration parameters and $K = 2$ so that $\phi_1 = \phi_2$. Defining $\phi_1 + \phi_2 = \phi$, the dynamics for the inertia weight (IW) and Clerc-Kennedy (CK) formulations is

$$IW : v(t+1) = wv(t) + \frac{\phi^{IW}}{2}[u_1(p_i - x(t)) + u_2(p_2 - x(t))] \qquad (9)$$

$$CK : v(t+1) = \chi v(t) + \chi \frac{\phi^{CK}}{2}[u_1(p_i - x(t) + u_2(p_2 - x(t))] \qquad (10)$$

Eq. 8 gives, at $u = 1$,

$$0 < \phi^{IW} < 2(1 + w) \qquad (11)$$

$$0 < \phi^{CK} < \frac{2(1 + \chi)}{\chi} \qquad (12)$$

The CK condition for complex eigenvalues and oscillation, $a^2 < 4b$ becomes, $a = \chi\phi$, $b = \chi$, becomes

$$\begin{cases} |\chi| < 1 \\ 1 + \frac{1}{\chi} - \frac{2}{\sqrt{\chi}} < \phi^{CK} < 1 + \frac{1}{\chi} - \frac{2}{\sqrt{\chi}} \end{cases} \quad (13)$$

In order to simplify the choice of χ and ϕ^{CK}, Clerc and Kennedy suggest a single relation $\phi^{CK} = \phi^{CK}(\chi)$,

$$\phi^{CK} = \frac{1}{\chi} + \chi + 2 \quad (14)$$

which can be easily seen to satisfy Eq. 13. This relation is usually inverted in the literature,

$$\chi^{CK} = \frac{2}{\phi - 2 + \sqrt{\phi^2 - 4\phi}}, \phi > 4 \quad (15)$$

and a common choice is $\phi^{CK} = 4.1$, $\chi \approx 0.73$. As $\phi^{CK} \rightarrow 4$, $\chi \rightarrow 1$ and the system becomes unstable, and as ϕ^{CK} grows from 4, χ decreases from 1 and the system is increasingly damped. In terms of the inertia weight formulation, these parameters correspond to $w \approx 0.73$ and $\phi_{IW} \approx 3.0$.

Many trials of PSO have found that best performance over a suite of test functions is attained at ϕ close to $u = 1$ instability. The reason behind this can be elucidated by considering the statistical distribution of $x(t)$. This is the subject of the next section.

3. PARTICLE DISTRIBUTION

It is known that decoupled PSO exhibits bursts of outliers [7]. These are temporary excursions of the particle to large distances from the attractors. A burst will typically grow to a maximum and then return through a number of damped oscillations to the region of the attractors. Figure 1 shows the development of a spectacular burst for the IW system defined by Eq. 9 at $w = 0.75$, $\phi = 3.0$. (Henceforth, unless specified otherwise, only the IW formulation will be considered, and the suffix IW will be dropped.) The particle is close to the $u = 1$ instability condition since, from Eq. 11, $\phi_{max} = 3.5$. In this simulation and for all others described in this section, $x(1)$ and $x(2)$ are random starting positions between the two fixed attractors, $p_1 = -0.5$ and $p_2 = 0.5$. Figure 1 shows a burst of two orders of magnitude, as measured in units of the intrinsic scale $|p_1 - p_2|$.

Figure 2 shows the frequency N of particle distance $r = |x|$ for the same system as Fig. 1 for a run of 10^6 iterations. A logarithmic scale (all logs in this paper are to base 10) has been used for the y-axis so that the infrequent but large bursts are visible on the plot. For this single run, the mean distance was 0.747 (standard deviation 1.05) and all distances are in the interval $[1.01 \times 10^{-6}, 105]$. Many updates are therefore over very small distances from the origin, which is the fixed point $\frac{<c>}{1 + <a> + b}$ of the constant parameter DE. Although the standard deviation is of the order of the attractor separation, r can range over 8 orders of magnitude.

Bursts would be expected to fatten the tail of the particle distance distribution $p(r)$ when compared to distributions with exponential fall-offs such as a Gaussian. Evidence for possible power law fattening of the distribution tail, $p(r) \sim r^{-\alpha}$, where $p(r)dr$ is the probability of a particle at distance r would be revealed in a plot of the logarithm of the cumulative distribution function, $P(r)$, where
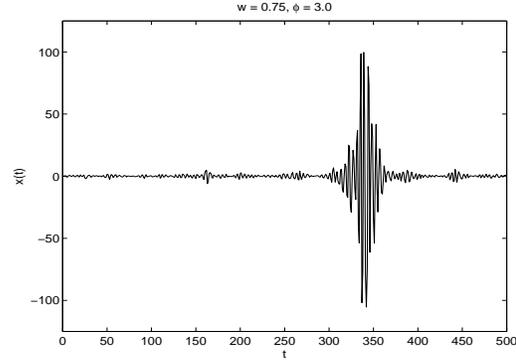

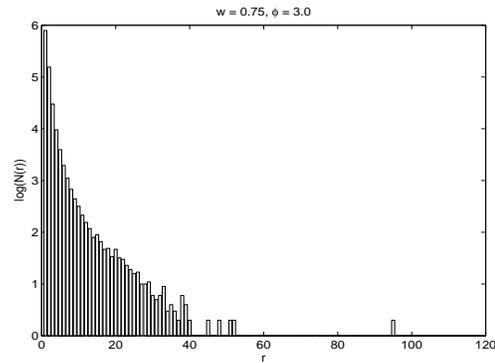**Figure 1: A burst of outliers in decoupled PSO**

Figure 2: Frequency N of particle distances r from the origin for 10^6 iterations of decoupled PSO
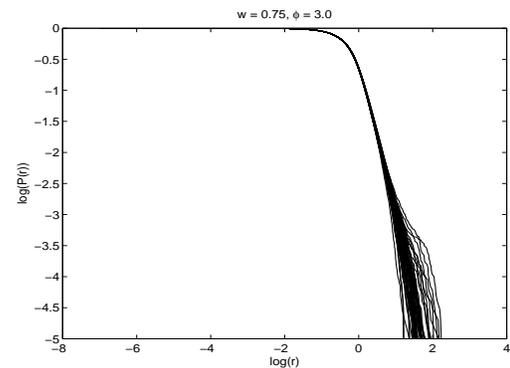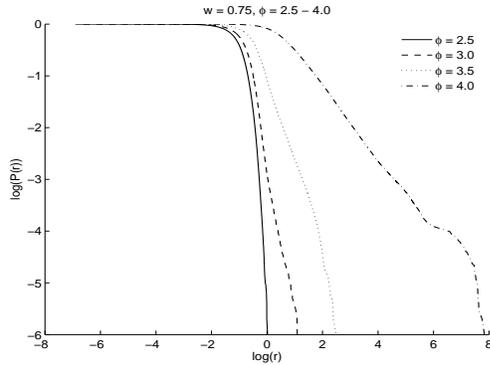

Figure 3: Cumulative probability distribution P versus particle distance r from the origin. The plot shows 50 runs.

Figure 4: **Cumulative probability distribution of particle distances** r **for various values of acceleration parameter** ϕ

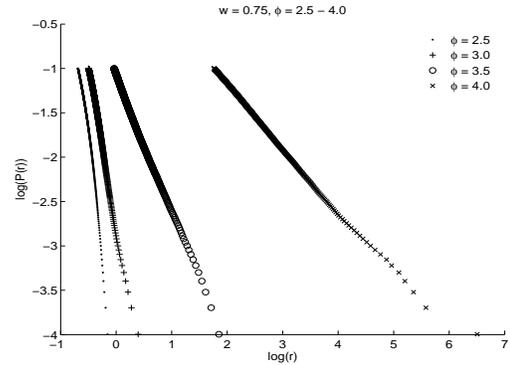Figure 5: **Power law tails at various values of acceleration parameter** ϕ

$P(R) = prob(r > R)$. A cumulative plot (also known as a rank/frequency plot [23]) reduces sampling errors in the tail of the plot, even with logarithmic binning [9]. A relation $p(r) \sim r^{-\alpha}$ corresponds to $P(R) \sim r^{1-\alpha}$ and a plot of $log(P)$ against $log(r)$ would show a straight line with gradient $1-\alpha$.

Figure 3 shows such cumulative distributions for 50 runs of 10^5 iterations, once more for the decoupled PSO defined by $w = 0.75$, $\phi = 3.0$, $p_{1,2} = \pm 0.5$, $x(1), x(2) = U(p_1, p_2)$. All runs have been plotted on this figure to give an idea of the deviations between runs. The straight portion in Fig. 3 is evidence for a power law.

Figure 4 shows cumulative probability distribution plots for four values of ϕ. This data is collected over 50 runs of 10^6 iterations for each value of ϕ. Each line shows a straight central part. The lines curve inwards towards the end of the sample where probabilities are small ($< 10^{-5}$) and there are just a few events. Once more a large part of the distribution is concentrated in the region between the attractors, $r < 1$. The power laws become established by $r \approx 1.0$, the separation of the attractors. At $\phi = 4.0$ the power law is evident for some 4 orders of magnitude.

Putative power laws as revealed by log-log plots are hardly distinguished from log-normal laws $p(x) \sim exp(-ln(x - \mu)^2)$ over four or less orders of magnitude [9]. These plots therefore only show that the tails *might* be modelled by a power-law distribution. The underlying distribution could be power-law or another distribution, such as the log-normal, whose tail can be approximated by a power law over some range.

Figure 5 plots the same data as Fig. 4 but over the interval $10^{-4} < P(r) < 10^{-1}$ where the power laws are becoming established. For clarity, every 1000th r in this range has been plotted. The gradients and correlation coefficients of the four lines are -3.94(-0.987), -3.74(-0.998), -1.08(-0.999) and -0.73(-0.999) for $\phi = 2.5, 3.0, 3.5, 4.0$ respectively. (A correlation coefficient of -1.0 indicates perfect negative correlation.) At the edge of $u = 1$ instability, $\phi = 3.5$ and $p \sim r^{-2.08}$. Interestingly, this is very close to the condition for a finite mean: $\int_{r_{min}}^{\infty} r p(r) dr$ is finite only for $\alpha > 2$, where $p(r)$ is power-law, $p(r) \sim r^{-\alpha}$ for $r > r_{min}$. This indicates the $u = 1$ stability condition corresponds to a well-defined mean particle displacement. Lower values of u, and hence higher values of ϕ, lead to systems whose em-

pirical mean over a finite number of iterations will be finite but will nevertheless vary enormously, sometimes taking on large values, in order to respect the formal divergence of the mathematical mean.

From the condition $P(r) = 0.1$, Figure 5 shows that 10 per cent of the particle positions are at distances greater than 0.21, 0.32, 0.92 and 56.8 from the origin for $\phi = 2.5$, 3.0, 3.5 and 4.0 respectively. This indicates that good coverage of the region $p_1 < r < p_2$ is attained for $\phi = 2.5$, 3.0 and 3.5. On the other hand, there is no coverage outside this interval for $\phi = 2.5$, showing that this PSO concentrates all its search between the attractors. At $\phi = 3.0$, particles will move outside this region, enhancing exploration away from the fixed points in the interactive model (where $p - 1$ and $p - 2$ can be updated). At $\phi = 3.5, 4.0$ the frequent bursts often take r far from the attractors. These large bursts cannot help with convergence, but they might help diversify the fully interacting system.

An analysis of the data for the 50 runs at $\phi = 4.5$, $w = 0.75$ found a probability of 10 per cent for positions of at least 10^{40}. $\phi = 4.5$ is between the $u = 1$ and $u = 0.5$ instability conditions (3.5 and 7.0 respectively). Although none of these runs exploded, bursts of extremely high amplitude were common. The inference from these experiments is that the $u = 1$ stability condition corresponds to power law tails with bounded mean. Moving ϕ beyond the $u = 1$ condition leads to unbounded mean displacements and little exploration of the region between the attractors.

4. MULTIPLICATIVE STOCHASTICITY

The previous section presented evidence that non-interacting PSO shows power-law behaviour when close to instability. However outlier particles are not isolated; rather, large excursions exist in bursts, or sequences of increasing and then decreasing amplitudes away from the origin. This is in contrast to a bare bones formulations which replace velocity with sampling from a probability distribution N. N might itself have fat tails but outliers would not be correlated. The bursts are a peculiar feature of the PSO dynamics.

Power law tails are found in many natural systems. Well known examples include the distribution of earthquake magnitudes, frequency of words in a language, wealth of the richest people and physical quantities close to a phase transition. Although power-laws have been regarded as an indicator of

self-organisation (e.g. [12]), this explanation is not necessary for power law behaviour [9], [20].

Another explanation for PSO bursts might lie in resonance. Certainly Eq. 3 has a driving term $c(t)$ and a spring-like term $\Phi(t)(p_i - x(t))$, Eq. 1, and might be expected to resonate. The system does not though have a well defined resonant frequency because the spring constants Φ are themselves random. Other explanations might lie in intermittency, which is a property of some chaotic systems. Intermittent systems show periods of constant amplitude punctuated by erratic bursts [10]. However, decoupled PSO is not chaotic in the stable regime. Another explanation for power laws can be found in the theory of random multiplicative processes [19].

4.1 First order SDE

Considering the first order SDE

$$x(t + 1) = -a(t)x(t), \qquad (16)$$

then $x(t) = (-)^t a(t-1)a(t-2)...a(0)x(0)$. The distribution of x is therefore given by the distribution of products of random numbers. The logarithm of $x(t)$ is therefore equal to a sum of log of random numbers, and by the central limit theorem, the distribution of $log(x)$ will be normal. The distribution of $x(t)$ is therefore log-normal, and log-normal laws are well approximated by power laws over intervals of four or less orders of magnitude [9]. This simple argument shows that fat, power-law tails can emerge from multiplicative processes. However, Eq. 16 is a very poor approximation to PSO. The second order SE of Eq. 3 reduces to first order for $b = 0$, $c(t) = 0$, corresponding to a PSO with $w = 0$ and $\sum \Phi_i p_i = 0$. This implies $u_1 = u_2$ and $p_1 + p_2 = 0$ giving a PSO,

$$x(t+1) = x(t) + \frac{\phi}{2}[(p_1 - u_3 x(t)) + (p_2 - u_3 x(t))]r \quad (17)$$

where u_3 is a random value. Eq. 17 was tested over a suite of 14 objective functions, duplicating the test conditions of [2], with very poor results.

The first order SDE with additive noise,

$$x(t+1) + a(t)x(t) = c(t) \qquad (18)$$

has been studied by Sornette and other workers (see, for example [19] for $a(t) < 0$). This system contains both multiplicative - $a(t)$ - and additive - $c(t)$ - stochasticity. Decoupled PSO reduces to Eq. 18 if the inertia weight is set to zero, $w = 0$,

$$x(t+1) = x(t) + \frac{\phi}{2}[u_1(p_1 - x(t)) + u_2(p_2 - x(t))] \quad (19)$$

Once more, performance of the fully coupled version of Eq. 19 is very poor. Without velocity, these PSO's cannot move through the search space and are doomed to local exploration around the initial swarm configuration.

Eq 18 exhibits a regime of power law behaviour. With $c(t) = 0$ we recover model Eq. 16 which is log-normal in its central part [15]. For c finite, iterating Eq. 17 gives the solution of Eq. 17 as

$$x(n) = (\prod_{l=0}^{n-1} a(l))x(0) + (\sum_{l=0}^{n-2} c(l) \prod_{m=l+1}^{n-1} a(m)) + c(n-1) \quad (20)$$

which shows that the fate of x is determined by the multiplications over a. The surprising feature is that Eq. 19 exhibits interesting behaviour in the *stable* regime $\langle a \rangle < 1$ [19]. This behaviour, namely intermittent bursts and power law tails to the distribution of x is contingent on $max(a(t)) > 1$ so that amplification is possible, and upon the injection of noise, $c \neq 0$ so that convergence to the fixed point is prevented.

Rewriting the $w = 0$ PSO as

$$x(t+1) + [\frac{\phi}{2}(u_1 + u_2) - 1]x(t) = \frac{\phi}{2}(u_1 p_1 + u_2 p_2) \quad (21)$$

facilitates comparison to Eq. 18. The fixed-u stability condition is $0 < \phi u < 2$. Without loss of generality we can place $p_1 = 1.0$, $p_2 = 0$ so that $c(t) = \frac{\phi}{2}u_1$. From the $u = 1$ stability condition, $\phi < 2$ so $c(t) \sim U(0, c_{max})$, $c_{max} < 1$. Furthermore, $a(t) \in [-1, \phi]$, although the distribution within this interval is triangular rather than uniform. This means that the $w = 0$ PSO differs from Eq. 18 in two respects: $a(t)$ can become positive and c and a are not independent. Indeed, $a(t) = [c(t) + \frac{\phi}{2}u_2 - 1]$.

These changes were investigated by trials on Eq. 17 with $a \sim U(a_{min}, a_{max})$ and $c(t) \sim U(0, 1)$. The results are shown in Figure 6. The plots depict average distances $r = |x|$ from the origin over 50 runs of 10^6 iterations and show results for four uniform distributions of a, each with $|\langle a \rangle| < 1.0$ and $max(|a| > 1)$. Line (ii) $a \sim U(-1.48, -0.48)$ corresponds to the system previously studied by Sornette and Cont [21]. Line (i) $a \sim U(0.48, 1.48)$ is the Sornette-Cont system but with a negative multiplicative term $x(t + 1) = -|a|x(t) + c(t)$. $a \sim U(-1.5, 1.5)$ is a symmetrical distribution, and $a \sim U(-1.75, 1.25)$ has $\langle a \rangle = -0.25$.

The results indicate that burst amplitudes and frequency are reduced in the Sornette-Cont system if the sign of a is reversed. This is possibly because the additive injection term $c(t)$ and the multiplicative term are then of opposite signs, and $c(t)$ will reduce the amplitude of positions $x < 0$. The two distributions that straddle $a = 0$ show markedly quenched bursts, once more because the fluctuating sign of $a(t)$ will lead to reductions in r when a is opposite in sign to x. Burst quenching is more prominent when $\langle a \rangle = 0$; displaced distributions will give larger bursts. All four lines in Fig. 4.1 show power law regimes for some r.
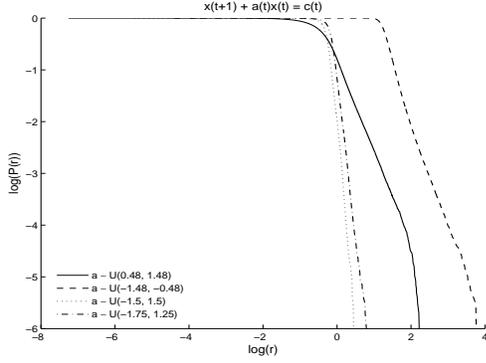
4.2 Second order SDE

The second order stochastic system with uniform distributions

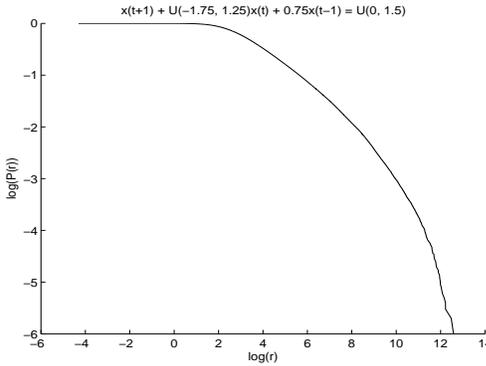$$x(t+1) + U(a_l, a_u)x(t) + bx(t-1) = U(c_l, c_u) \qquad (22)$$

has not, to our knowledge, been studied in the burst regime $max(|a|) > 1$. Eq. 22 is closely related to the decoupled PSO with $p_1 = 1$, $p_2 = 0$:

$$x(t+1) + [\frac{\phi}{2}(u_1 + u_2) - w - 1]x(t) + wx(t-1) = \frac{\phi}{2}u_1 \quad (23)$$

Replacing $u_1 + u_2$ and u_1 with uniform distributions leads to

Figure 6: Investigation of stochastic first order equation for various ranges of random variables $a(t)$

Figure 7: Investigation of a stochastic second order equation

$$x_{(t+1)} + U(-w-1, \phi-w-1)x(t) + wx(t-1) = U(0, \frac{\phi}{2}) \quad (24)$$

and in particular, for the $\phi = 3.0$, $w = 0.75$ system,

$$x(t+1) + U(-1.75, 1.25)x(t) + 0.75x(t-1) = U(0, 1.5) \quad (25)$$

Figure 7 shows the cumulative distribution $r = |x|$ for Eq. 25 for 50 runs of 10^6 iterations. The frequent and large bursts are apparently not well modeled by a power law over the entire tail, although the portion $3 < log(r) < 6$ might be modelled by a straight line.

Formally we can derive a solution for $x(t)$ as a sum over products of random variables. Rewriting Eq. 3 as

$$x_n = a_{n-1}x_{n-1} + b_{n-2}x_{n-2} + c_{n-1} \equiv a + b + c \quad (26)$$

where for convenience the sign of the random variables $a(t)$ and $b(t)$ have been reversed, $a_n = -a(t), b_n = -b(t)$ and the symbol $a+b+c$ has been introduced as a shorthand for $a_{n-1}x_{n-1} + b_{n-2}x_{n-2} + c_{n-1}$. Iterating back in time,

$$x_n = a(a+b+c) + b(a+b+c) + c \quad (27)$$

with the understanding $aa \equiv a_{n-1}a_{n-2}x_{n-2}$, $ab \equiv a_{n-1}b_{n-3}x_{n-3}$, $ba \equiv b_{n-2}a_{n-3}x_{n-3}$, $bb \equiv b_{n-2}b_{n-4}x_{n-4}$.

Hence, after m iterations,

$$x_n = (a+b)^m + (\sum_{j=0}^{m-1}(a+b)^j)c \quad (28)$$

where the symbol $a^{m-1}c \equiv a_{n-1}a_{n-2}...a_{n-m+1}c_{n-m}$ etc. and $(a+b)^0 \equiv 1$. Eq. 28 reduces to the solution of the first order difference equation, Eq. 20, for $m = n$, $b = 0$.

We take the order of expansion, m to the edge of the burst. The definition of a burst is arbitrary; when x is not being amplified, the particle will move close to the fixed point at $\frac{<c>}{1-<a>-b}$. During this normal, non-bursting activity, all x_i's appearing in Eq. 28 can be replaced by their non-burst average, $\langle x \rangle_{noburst}$. We also note that the role of c_n is to repel x away from the fixed point during normal activity. During a burst, c will be small in comparison with x and plays no part in the amplification. During bursting, we therefore replace c_n by its average value, $< c >$.

The aim of the following analysis is to use Eq. 28 to investigate the mean size of a burst of length m iterations. This can be achieved by replacing each term in the expansion of $(a+b)^m$ with the m'th moment $< a^p b^{m-p} >=< a^p > b^{m-p}$. From $\langle a^p \rangle = \frac{1}{a_u - a_l} \int_{a_l}^{a_u} a^p da = \frac{a_u^{p+1} - a_l^{p+1}}{(p+1)(a_u - a_l)}$ where u and l denote the upper and lower limits of the distribution,

$$< a^p b^{m-p} > = \frac{a_u^{p+1} - a_l^{p+1}}{(p+1)(a_u - a_l)} b^{m-p} \quad (29)$$

and

$$< (a+b)^m > = \sum_{p=0}^{m} \binom{m}{p} < a^p b^{m-p} > \quad (30)$$

From Eq. 28, the position $< x_m >_{burst}$ at the end of a burst of length m is given by

$$< x_m >_{burst} \approx < x >_{noburst} \sum_{p=0}^{m} \binom{m}{p} < a^p b^{m-p} > \quad (31)$$

$$+ < c > \sum_{j=0}^{m-1} \sum_{p=0}^{j} \binom{j}{p} < a^p b^{j-p} >$$

The final term $(p = m)$ in the expression for $< (a+b)^m >$ in Eq. 30, $< a^p >$, is the moment due to the first order process $x_{n+1} = a_n x_n + c$. The remaining terms in Eq. 30 give the contributions due to finite b, i.e. for the second order process $x_{n+1} = a_n x_n + b x_{n-1} + c$. These other terms, which alternate in sign for $a < 0$, grow in size compared to $\langle a^p \rangle$ for larger m and contribute significantly to the overall sum. Figure 8 depicts the rise in $\langle (a+b)^m \rangle$ compared to the $b = 0$ first order process $\langle a^m \rangle$ for m between 0 and 10 for the system described by Eq. 25. This Figure demonstrates that the addition of a second order term considerably increases burst amplitude.

Figure 9 charts, for the same PSO, $r_{burst} = | < x_m >_{burst} |$, using Eq. 31 where $\langle a \rangle = 0.25$, $\langle c \rangle = 0.75$ and the fixed point $x^\star = 0.5 \approx < x >_{noburst}$. The approximations leading to expression 31 are only valid after a burst has become established and $x_m >> \langle x \rangle_{noburst}$ which is true for this data for $m \geq 10$. According to the Figure, bursts of length 10 and above grow exponentially in size, $r_{burst} \sim e^{0.65m}$. This
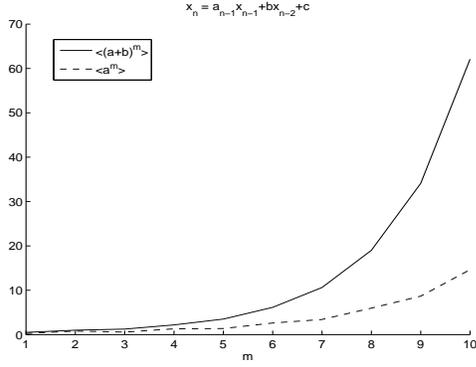
Figure 8: Comparison of moment sums for first and second order processes
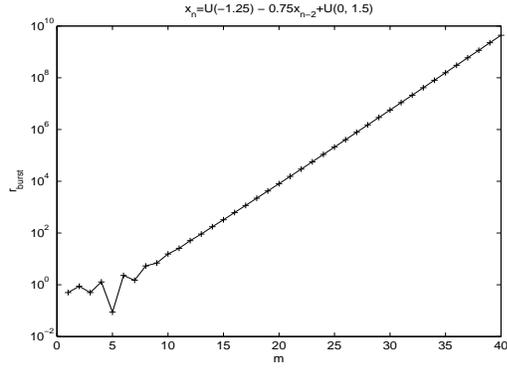

Figure 9: Average burst size r_{burst} as a function of burst length m

provides an explanation for the extremely high amplitude bursts found in the second order SDE Eq. 22.

This analysis motivates the conclusion that the tail of the second order SDE with $max|a| > 1$) is built from a sum of multiplicative processes. Each multiplicative processes will provide a log-normal distribution, as described in the previous section. The distribution of positions in the full 2nd order SDE is therefore a sum of log-normals, each of which is approximated by a power law over some part of its range.

5. ADDITIVE STOCHASTICITY

If distribution tails in SDEs are caused solely by multiplicative stochasticity, a second order SDE with only additive stochasticity i.e. $a, b = const$, $c = c(t)$ should be tail-free. Recently a novel PSO variant, recombinant PSO (DR) has been proposed with only additive stochasticity [13]. Recent work tests PSO-DR for various neighbourhoods and parameter choices with impressiveley competitive results over a suite of 15 common benchmarks [1]. PSO-DR is similar to the PSO-IW, Eq. 9, except that one of the informers is replaced by a discrete recombination of a particle's immediate neighbours in a ring topology,

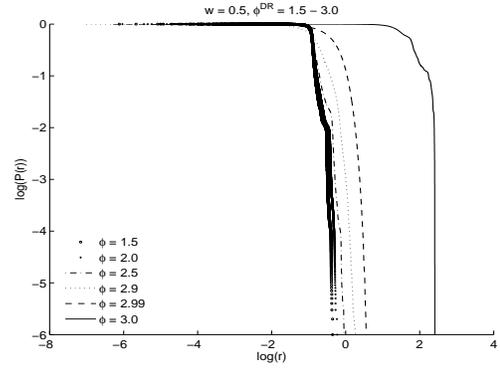$$DR : v(t+1) = wv(t) + \frac{\phi^{DR}}{2}[(p_1 - x(t)) + (p_2 - x(t))] \quad (32)$$

Figure 10: Plot of cumulative probability r for PSO-DR

where $p_2 = \eta p_l + (1 - \eta)p_r$, $\eta = U\{0, 1\}$ and p_l and p_r are left and right neighbours and p_1 is either the personal best position of particle i, or the best position in i's neighbourhood, depending on the particular formulation of PSO-DR. The stability condition from Eq. 11 is $0 < \phi^{DR} < 2(1 + w)$.

Figure 10 reports on the cumulative distribution of particle separation r for Eq. 32. The distributions for $w = 0.5$ and various ϕ up to the maximum stable value of $\phi = 3.0$ were collected for 50 runs of 10^6 iterations with $x(1), x(2) \sim U(-0.5, 0.5)$, $p_1 = -0.5, p_l = 0.5, p_r = 1.0$.

The cumulative distributions are flat for small r, and then drop vertically at a cut-off $r - c$, suggesting $p(0 \geq r \geq r_c) = U(0, r_c)$ (although the log-log plot is not sensitive enough to show variations from uniformity). The non critical systems $\phi \leq 2.9$ place the majority of the positions between the attractors. At sub-criticality, $\phi = 2.99$, the system is inclined to explore beyond the attractors, $r_c > 1$. At instability, r_c is between 50 and 100. A vertical drop off beyond r_c would be evidence that additive-SDE does not develop tails, and this is confirmed by these plots, except perhaps for $\phi = 2.99$ which appears to have a finite, but very large slope.

In fact PSO models such as Eq. 32 might produce tails from a resonance effect. This is because the spring constants are fixed and the system has a defined natural frequency. For the case of PSO-DR, setting $p_3 = \frac{p_1 + p_2}{2}$ gives a simple oscillator with force law $F = \phi(p_3 - x)$ and natural frequency $\omega = \sqrt{(\phi)}$. The periodic time, $T = \frac{2\pi}{\omega}$ for $\phi = 1$ (this is empirically a good value for interacting PSO-DR [1]) is therefore about 6 with the implication that an oscillating p_3 on the timescale of 6 iterations could drive the oscillator and amplify x. This could happen from a shifting neighbour best position p_1, or from an oscillation between p_l and p_r in the p_2 term, or by a combination of the two.

6. CONCLUSIONS

This paper has investigated the position distribution of de-coupled PSO. Particular attention has been paid to the tail of this distribution, a regime dominated by power laws. The origin of these tails lies with the phenomenon of erratic particle bursts. In order to study how these bursts might develop, de-coupled PSO has been formulated as a second order stochastic difference equation. A series of approximations has demonstrated that fat distribution tails, well modelled by power laws, arise from multiplicative stochas-

ticity, a phenomenon previously encountered in first order SDEs.

This conclusion is valid for first or second order SDE's, and for either sign of the parameters, as long as one of more of these parameters is capable of amplification i.e. can exceed unity. These trials are supported by a theoretical analysis: bursting in a second order SDE is built from a sum of multiplicative stochastic processes. Each multiplicative processes will provide a log-normal distribution, approximated by a power law over some part of its range. This result explains the observation that the particle distribution shifts as more iterations are collected. The distribution is dominated by large but rare events that are only manifest after many iterations.

A stability condition for the PSO parameters w and ϕ can be achieved by replacing the random variables in the difference equation by a constant, u. The inference from a set of experiments is that the $u = 1$ stability condition leads to power law tails with bounded mean. Moving ϕ beyond this condition leads to unbounded mean displacements. The popular $\phi = 3.0, w = 0.75$ PSO is within the stable region and has weak power law tails, which enhance exploration, yet also has good coverage of the region close to the attractors, enabling convergence.

There is a tantalising possibility that the removal of stochasticity from the dynamics might render PSO amenable to further theoretical analysis. A recombinant PSO, which is demonstrably competitive to standard PSO, almost achieves this miracle. The acceleration parameters are constant in PSO-DR, but randomness, and hence diversity regeneration, is manifest in a jiggling of the attractor components. This jiggling will persist even at times of stagnation. PSO-DR, replete with just additive stochasticity of this sort, does not, according to the theoretical and empirical arguments supplied here, enjoy bursting activity.

PSO bursts differ from the outliers generated by bare bones swarms in two respects: the outliers occur in sequence, and they are 1-dimensional. Bursting will therefore produce periods of rectilinear motion where the particle will have a large velocity parallel to a coordinate axis. Whether bursts are generally beneficial, or a hindrance, to a fully interactive PSO, and under what circumstances, is the subject of ongoing research [1].

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] D. Bratton and T. M. Blackwell. Understanding particle swarms through simplification: A study of recombinant pso. In *Submitted to Particles swarms: the second decade, workshop, GECCO 2007*, 2007.

[2] D. Bratton and J. Kennedy. Defining a standard for particle swarm optimization. In *IEEE swarm intelligence symposium*, 2007.

[3] M. Clerc and J. Kennedy. The particle swarm: explosion, stability and convergence in a multi-dimensional space. *IEEE transactions on Evolutionary Computation*, 6:158–73, 2000.

[4] A. Engelbrecht. *Fundamentals of computational swarm intelligence*. John Wiley and Sons Ltd., 2005.

[5] J. D. Hamilton. *Time series analysis*. Princeton university press, 1994.

[6] J. Kennedy. Bare bones particle swarms. In *IEEE swarm intelligence symposium*, pages 80–87, 2003.

[7] J. Kennedy. Probability and dynamics in the particle swarm. In *IEEE congress on evolutionary computation*, pages 340–347, 2004.

[8] J. Kennedy and R. Mendes. Neighbor topologies in fully-informed and best-of-neighborhood particle swarms. In *IEEE int'l workshop on soft computing in industrial applications*, pages 45–50, 2003.

[9] M. E. Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[10] E. Ott. *Chaos in dynamical systems*. Cambridge University Press, New York, 2002.

[11] E. Ozcan and C. K. Mohan. particle swarm optimization: surfing the waves. In *IEEE congress on evolutionary computation*, 1999.

[12] C. T. P. Bak and K. Wiesenfeld. Self-organized criticality: an explanation for 1/f noise. *Phys. Rev. Lett*, 59:381–384, 1987.

[13] J. Pena, A. Upegui, and E. Sanchez. Particle swarm optimization with discrete recombination: an online optimizer for evolvable hardware. In *First NASA/ESA conference on adaptive hardware and systems, AHS '06*, pages 163–170, 2007.

[14] R. Poli. Exact analysis of the sampling distribution for the canonical particle swarm optimiser and its convergence during stagnation. In *GECCO (submitted)*, 2007.

[15] S. Redner. Random multiplicative processes. *Am. J. Phys*, 58:267–273, 1990.

[16] T. Richer and T. Blackwell. The levy particle swarm. In *IEEE congress on evolutionary computation*, pages 3150–3157, 2006.

[17] R.Poli, W. Langdon, M. Clerc, and C. Stephens. Continuous optimisation made easy? finite element models of evolutionary strategies, genetic algorithms and particle swarm optimisers. In *Proc. foundations of genetic algorithms (FOGA) workshop*, 2007.

[18] Y. Shi and R. Eberhart. A modified particle swarm optimizer. In *Congress on Evolutionary Computation*, pages 69–73, 1998.

[19] D. Sornette. Multiplicative processes and power laws. *Phys. Rev. E*, 57:4811–4813, 1998.

[20] D. Sornette. Mechanism for powerlaws without self-organization. *Int. J. Mod. Phys.*, pages 133–136, 2001.

[21] D. Sornette and R. Cont. Convergent multiplicative processes repelled from zero: power laws and truncated power laws. *J. Phys. I France*, 7:431–444, 1997.

[22] F. van den Burgh. *An analysis of particle swarm optimizers*. PhD thesis, University of Pretoria, 2002.

[23] G. K. Zipf. *Human behaviour and the principle of least action*. Addison Wesley, Reading, MA, 1949.