Research Article
Examination of Particle Tails

Tim Blackwell and Dan Bratton

Department of Computing Goldsmiths, University of London, New Cross, London SE14 6NW, UK

Correspondence should be addressed to Tim Blackwell, tim.blackwell@gold.ac.uk

The tail of the particle swarm optimisation (PSO) position distribution at stagnation is shown to be describable by a power law. This tail fattening is attributed to particle bursting on all length scales. The origin of the power law is concluded to lie in multiplicative randomness, previously encountered in the study of first-order stochastic difference equations, and generalised here to second-order equations. It is argued that recombinant PSO, a competitive PSO variant without multiplicative randomness, does not experience tail fattening at stagnation.

1. INTRODUCTION

This paper focuses on the effective sampling distribution of particle swarm optimisation (PSO). PSO is a population-based global optimisation technique. The algorithm is made of two essential steps: particle movement (dynamics) through a search space of solutions and an indirect particle interaction which is mediated by information sharing within a social network. Particles place personal "attractors" in the search space at positions corresponding to the best location (as determined by the evaluation of an objective function) that they have visited. The dynamical update rule which specifies the acceleration of a particle i is a sum of an attraction towards a personal attractor and an attraction towards the best attractor amongst other particles in i's social neighbourhood [1–3]. The particle acceleration is added to particle velocity in a discretisation of particle kinematics; the velocity is added in turn to particle position and a new trial solution is achieved. Despite the simplicity of this 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.

The sampling distribution provides the probability density of particle placement between and around fixed attractors. This scenario occurs when the swarm "stagnates," the swarm is not improving, the network of attractors is fixed, and the particles are moving independently of each other, that is, the particles decouple. Stagnation can occur in the full model, or can be imposed for the purposes of theoretical analysis. In general, particles are expected to search in the immediate vicinity of the attractors, and indeed this behaviour is necessary for convergence. The central portion of the sampling distribution is therefore coincident with this region. However, the swarm must retain an ability to explore outside the attractor vicinity if premature convergence, a problem in complex environments, is to be avoided. This exploration is due to the finite tail of the sampling distribution; there is a small probability that a particle will be displaced far from the region of convergence, and this probability increases with tail fatness.

Some insights into the PSO sampling distribution have been provided by a study of "bare bones" formulations [4]. In this work, Kennedy replaced the particle update rule with sampling from a Gaussian distribution with mean at the centroid of the personal and neighbourhood best positions of each particle, and standard deviation was set to the separation between them. Performance comparisons with a few benchmark problems were disappointing (later confirmed in [5]), but the addition by hand of particle "bursts" ameliorated the situation somewhat, indicating that the tails of this Gaussian distribution function are too thin to enable escape from stagnation. Further evidence for this conclusion has been provided by a study of Lévy bare bones [5] where more extensive trials showed that fat tails, as provided by the Lévy distribution, improve bare bones performance so that it becomes effectively equivalent to standard PSO. A recent

```
0. initialise swarm
FOR EACH particle i
    randomise v⃗ᵢ, x⃗ᵢ, set p⃗ᵢ = x⃗ᵢ
FOR EACH particle i
    1. find neighbourhood best
        g⃗ᵢ = arg min (f(pⱼ), j ∈ Nᵢ)
    FOR EACH dimension d
        2. move particle
            F(p) : xᵢ_d ← xᵢ_d
        3. update memory
            IF f(x⃗ᵢ) < f(p⃗ᵢ) THEN
                p⃗ᵢ ← x⃗ᵢ
    END
END
```

ALGORITHM 1: Particle swarm opimisation.

theoretical analysis of bare bones PSO has been given in [6]. However, a Gaussian bare bones version of a "fully informed" particle swarm (FIPS) was able to deliver good performance over a small testbed of functions [7], which suggests that particle tails might not be so important, at least for FIPS.

It is known that decoupled PSO exhibits bursts of outliers [8]. Bursts are temporary excursions, along a coordinate axis, 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. The addition of bursts to Gaussian bare bones increases the chance of particle displacement away from the attractors, fattening the tail of the overall sampling distribution. Lévy distributions also produce particle outliers, but these outliers are not correlated; they do not occur in sequence along a single axis. At the moment, the circumstances, when a sequence of correlated outliers (as manifest, for example, in PSO bursts) might prove to be fortuitous, are unknown. This paper investigates the tail of the PSO sampling distribution. Our results show that this tail is described by power laws and is therefore fatter than Gaussian. This fattening is indeed due to bursting, the origin of which is concluded to lie in a process known as multiplicative randomness.

Bursting is already known to occur in first-order stochastic difference equations [9]. This paper generalises the first order results to second order difference equations with multiplicative randomness, a class of equations that includes standard PSO. Since the amount of bursting is dependent on the range of the probability distributions, the paper begins by discussing possible parameter regimes. This is accomplished by a formulation of PSO as a finite difference equation. The paper continues with a series of experiments which reveal the power law tails of the sampling distribution. Section 4 introduces multiplicative randomness and power law tails in first-order processes and extends these results to second-order difference equations. A competitive reformulation PSO without multiplicative randomness has recently been proposed (recombinant PSO, [10]). Section 5 investigates decoupled recombinant PSO and demonstrates the consequent thinness of the distribution tail. The results of this paper are drawn together in a concluding section and some open research questions are outlined.

2. PSO AND STOCHASTIC DIFFERENCE EQUATIONS

This section recasts PSO as a stochastic difference equation (SDE). Since this is an unfamiliar form of PSO, a few relations from the literature governing parameter choice are gathered together and presented here.

Each PSO 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 outlined in Algorithm 1. A general dynamic rule $F(p)$ for a single particle position update is

$$v_{id}(t+1) = wv_{id}(t) + \sum_{j=1}^{K} \Phi_j(t)\big(p_{jd} - x_{id}(t)\big),$$
$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1). \tag{1}$$

The sum in (1) is over K informers p_i and for each dimension d. In standard PSO [11], however, $K = 2$ and the two attractors p_i and p_{g_i} are the personal and neighbourbood best positions. In this case,

$$\Phi_{1,2} = \phi_{1,2} u_{1,2}, \tag{2}$$

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 [12] derives its name 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' [13] fully informed particle swarm (FIPS), wherein a particle is influenced by $K > 2$ neighbours, $\Phi_j = (1/K)\phi_j u_j$, and the "constricted" Clerc-Kennedy (CK) swarm [14] which is equivalent to (1) with the identification $\chi = w$, $\phi_{1,2} = \chi\phi_{1,2}^{CK}$.

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

$$x(t+1) + a(t)x(t) + bx(t-1) = c(t) \tag{3}$$

with

$$a(t) = \sum_j \Phi_j(t) - w - 1,$$
$$b = w, \tag{4}$$
$$c(t) = \sum_j \Phi_j(t)p_j,$$

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

Constant parameter SDEs have been studied by many authors in a number of domains (e.g., [15] gives references to population dynamics, epidemics, ARCH(1) processes, investment portfolios, immigrant populations, and internet usage). 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 (3) and considering roots of the resulting characteristic equation $\lambda^2 + a\lambda + b = c$ (see, e.g., [16]). Stability then requires $|\lambda| < 1$. Complex, and therefore oscillatory, solutions are found if

$$0 < b < 1,$$
$$a^2 < 4b, \tag{5}$$

and stable real soutions exist for

$$|a| < 1 + b,$$
$$ca^2 \geq 4b. \tag{6}$$

The stability conditions can be combined:

$$|b| < 1,$$
$$|a| < 1 + b. \tag{7}$$

In order to relate these stability conditions to the stochastic model of (3) and (4), a number of authors have suggested replacing the random variable $u_{1,2}$ by its maximum, that is, $u = 1$ [14, 17, 18]. In terms of PSO parameters w and Φ (assumed constant), this gives

$$|w| < 1,$$
$$0 < \sum_j \Phi_j < 2(1 + w). \tag{8}$$

Poli and Broomhead [19] has extended this analysis by retaining randomness and deriving stability relations for the expectations of the first and second moments of position. The relation for the expected value of x is equivalent to the above analysis with $u = 0.5$, so the upper bound on the acceleration parameters ϕ is therefore twice the $u = 1$ bound.

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

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

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

Equation (8) gives, at $u = 1$,

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

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

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

$$|\chi| < 1,$$
$$1 + \frac{1}{\chi} - \frac{2}{\sqrt{\chi}} < \phi^{\text{CK}} < 1 + \frac{1}{\chi} - \frac{2}{\sqrt{\chi}}. \tag{13}$$

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

$$\phi^{\text{CK}} = \frac{1}{\chi} + \chi + 2 \tag{14}$$

which can be simply rewritten as

$$\chi\phi^{\text{CK}} = (\chi + 1)^2 \tag{15}$$

and can be easily seen to satisfy (13). This relation is usually inverted in the literature,

$$\chi^{\text{CK}} = \frac{2}{\phi - 2 + \sqrt{\phi^2 - 4\phi}}, \quad \phi > 4, \tag{16}$$

and a common choice is $\phi^{\text{CK}} = 4.1$, $\chi \approx 0.73$.

As $\phi^{\text{CK}} \to 4$, $\chi \to 1$, 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 $\phi_{\text{IW}} \approx 3.0$. Many trials of PSO over commonly used test functions have found that best performance is attained at ϕ close to the $u = 1$ stability condition. The reason behind this can be elucidated by considering the statistical distribution of $x(t)$.

3. DISTRIBUTION TAIL

This section presents an experimental investigation of the tail of sampling distribution for decoupled PSO. In each experiment described here, $x(1)$ and $x(2)$ are random starting positions between the two fixed attractors, $p_1 = -0.5$ and $p_2 = 0.5$. Henceforth, only the IW form of PSO will be considered (the Clerc-Kennedy form is a simple parameter redefinition) and the suffix IW will be dropped for ease of notation.

Figure 1 shows the development of a spectacular burst for the system defined by (9) at $w = 0.75$, $\phi = 3.0$. The particle is close to the $u = 1$ instability condition since, from (11), $\phi_{\text{max}} = 3.5$. 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 Figure 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 $\langle c \rangle/(1 + \langle a \rangle + b)$ of the expectation value of x. Although the standard deviation is of the order of the attractor separation, r can range over 8 orders of magnitude.
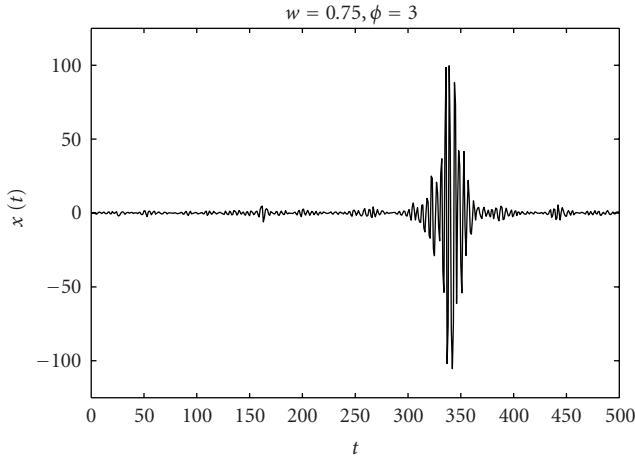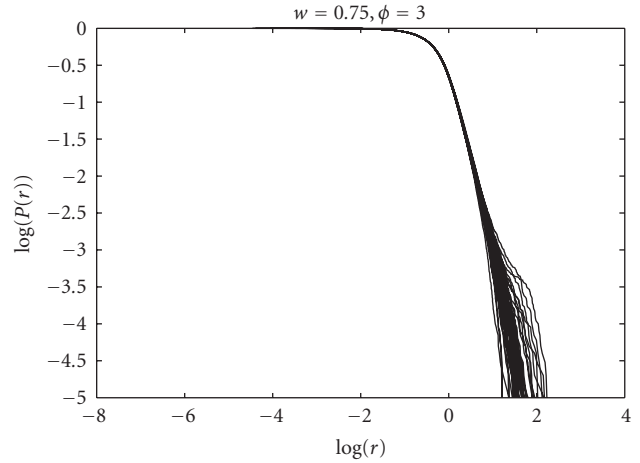
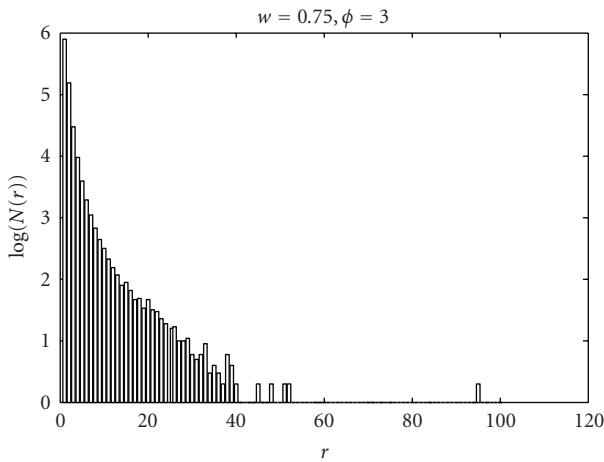FIGURE 1: A burst of outliers in decoupled PSO.

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.
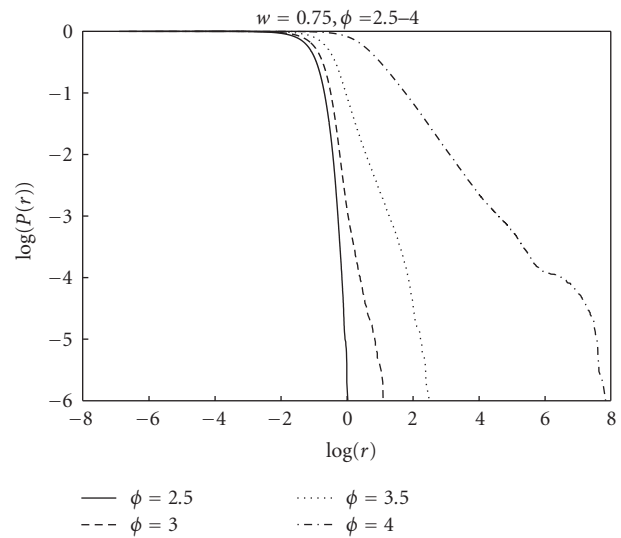

FIGURE 4: Cumulative probability distribution of particle distances r for various values of acceleration parameter ϕ.

Bursts would be expected to fatten the tail of the particle distance distribution $p(r)$ when compared to distributions with exponential falloffs 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 $P(R) = \text{prob}\,(r > R)$. A cumulative plot (also known as a rank/frequency plot [20]) reduces sampling errors in the tail of the plot, even with logarithmic binning [21]. 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 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 Figure 3 is evidence for a power law (although the underlying distribution might be log-normal; see discussion in the following paragraphs).

Figure 4 shows cumulative probability distribution plots for four values of ϕ at $w = 0.75$. 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 [21]. These plots therefore only show that the tails *might* be
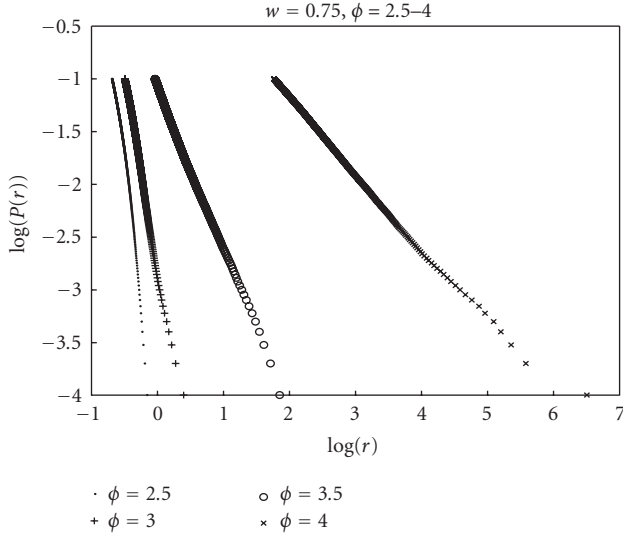
FIGURE 5: Power-law tails at various values of acceleration parameter ϕ.

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 Figure 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.)

If the sampling distribution $p(r)$ does follow a power law falloff, $p(r) \sim r^{-\alpha}$ for $r > r_{\min}$ (as these results suggest), then $\langle r \rangle$ is well defined only for $\alpha > 2$ because $\int_{r_{\min}}^{\infty} r p(r) dr$ diverges for $\alpha < 2$. According to these experiments, at $\phi = 3.5$, the tail falls off as $p \sim r^{-2.08}$, so $\langle r \rangle$ is finite at this value (and at smaller values) of ϕ, but $\langle r \rangle$ diverges at larger values. This edge is just at the $u = 1$ stability condition. Lower values of u, and hence higher values of ϕ, lead to systems whose empirical 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 coupled system.

An analysis of the data for the 50 runs at $\phi = 4.0, w = 0.75$ found a probability of 10 per cent for positions of at least 10^{40}. $\phi = 4.0$ is between the $u = 1$ and $u = 0.5$ stability conditions (3.5 and 7.0, resp.). 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. This may explain the popular empirical choices $w \approx 0.73$ and $\phi_{\text{IW}} \approx 3.0$.

4. MULTIPLICATIVE RANDOMNESS

The PSO dynamics can produce, under stagnation, outlying particles. However, these outliers particles are not isolated; rather, large excursions exist in bursts or sequences of increasing and then decreasing amplitudes away from the origin. This must be true because arbitrarily large steps $r(t + 1) - r(t)$ are prohibited;

$$
\begin{aligned}
r(t+1) &< |-a(t)x(t) - bx(t-1) + c(t)| \\
&< \max(|a|)r(t) + br(t-1) + \max(|c|).
\end{aligned}
\tag{17}
$$

(This is in contrast to a bare bones formulations which sample from a probability distribution N. N might itself has fat tails but outliers need not be correlated and arbitrarily large steps $r(t + 1) - r(t)$ are possible. Bursts are a feature of the finite difference equations.)

The previous section presented an evidence that the decoupled PSO shows power-law behaviour when close to constant-u instability. 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., [22]), this explanation is not necessary [21, 23].

Other possible explanations of bursts include resonance. Certainly, (3) has a driving term $c(t)$ and a spring-like term $\Phi(t)(p_i - x(t))$, (1), and might be expected to resonate. However, the system does not have a well-defined resonant frequency because the spring constants Φ are themselves random.

Intermittent chaotic systems show periods of constant amplitude punctuated by erratic bursts [24]. However, decoupled PSO is not chaotic in the stable regime. Another simpler explanation for power laws can be found in the theory of random multiplicative processes [9].

4.1. First-order SDE

Considering the first order SDE

$$
x(t+1) = -a(t)x(t),
\tag{18}
$$

then $x(t) = (-)^t a(t-1)a(t-2) \cdots a(0)x(0)$. The distribution of x is therefore given by the distribution of products of

random numbers. The logarithm of $x(t)$ is equal to a sum of logarithms 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 [21]. This simple argument shows that fat, power-law tails can derive from first-order multiplicative processes.

However, (18) is a very poor approximation to PSO. The second-order SDE defined by (3) reduces to the first-order SDE considered above for $b = 0$, $c(t) = 0$. This corresponds to a PSO with $w = 0$ and $\sum \Phi_i p_i = 0$. This implies that $u_1 = u_2$ and $p_1 + p_2 = 0$; giving a PSO,

$$x(t+1) = x(t) + \frac{\phi}{2}\big[(p_1 - u_3 x(t)) + (p_2 - u_3 x(t))\big]r, \tag{19}$$

where u_3 is a random value. (19) was tested over a suite of 14 objective functions, duplicating the test conditions of [11], with very poor results.

The first-order SDE with additive noise,

$$x(t+1) + a(t)x(t) = c(t), \tag{20}$$

has been studied by Sornette and other workers (e.g., see [9] for $a(t) < 0$). This system contains both multiplicative $a(t)$ and additive $c(t)$ randomness. Decoupled PSO reduces to (20) if the inertia weight is set to zero, $w = 0$,

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

Once more, performance of the fully coupled version of (21) is very poor over the above test function set. Without velocity, these PSOs cannot move through the search space and are doomed to local exploration around the initial particle configuration.

Equation (20) exhibits a regime of power-law behaviour. With $c(t) = 0$, we recover model (18) which is log normal in its central part [25]. For c finite, iterating (19) gives the solution of (19) as

$$x(n) = \left(\prod_{l=0}^{n-1} a(l)\right) x(0) + \left(\sum_{l=0}^{n-2} c(l) \prod_{m=l+1}^{n-1} a(m)\right) + c(n-1) \tag{22}$$

which shows that the fate of x is determined by the multiplications over a. The surprising feature is that (21) exhibits interesting behaviour in the *stable* regime $\langle a \rangle < 1$ [9]. 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

$$x(t+1) + \left[\frac{\phi}{2}(u_1 + u_2) - 1\right]x(t) = \frac{\phi}{2}(u_1 p_1 + u_2 p_2) \tag{23}$$
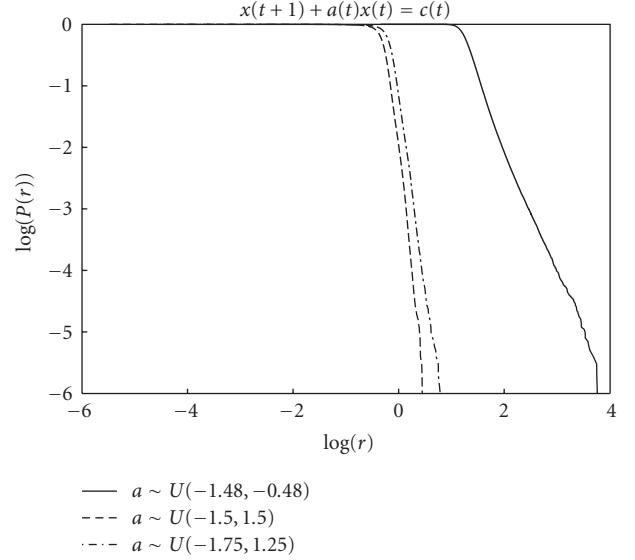

FIGURE 6: Investigation of stochastic first-order equation for various ranges of random variables $a(t)$.

facilitates the comparison with (20). 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 $c(t) = (\phi/2)u_1$. From the $u = 1$ stability condition, $\phi < 2$, $c(t) \sim U(0, c_{\max})$, $c_{\max} < 1$. 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 (20) in two respects: $a(t)$ can become positive and c and a are not independent. Indeed, $a(t) = [c(t) + (\phi/2)u_2 - 1]$.

These changes were investigated by trials on Sornette and Cont's original system, (20) 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 (i), $a \sim U(-1.48, -0.48)$, corresponds to the system previously studied by Sornette and Cont [26], line (ii), $a \sim U(-1.5, 1.5)$, is a symmetrical distribution with $\langle a \rangle = 0$, and line (iii), $a \sim U(-1.75, 1.25)$, has $\langle a \rangle = -0.25$.

The plot shows the fat power-law tail for the Sornette-Cont system. The results also show thin tails when the mean a is close to zero (lines (ii) and (iii)). This can be explained by the relative drop in probability of amplification, $|a| > 1$. In SDEs (ii) and (iii), $\mathrm{prob}(|a| > 1) = 1/3$, whereas in the Sornette-Cont system, $\mathrm{prob}(|a| > 1) = 0.48$. The second term of (22) can be ignored during a large burst $|x(n)| \gg |x(0)| \gg \max(c)$,

$$x(n) \approx \prod_{l=0}^{n-1} a(l)x(0). \tag{24}$$

Sequences of $a(l)$ with large products will occur less often in (ii) and (iii) compared to (i), and the distribution tail is quenched. A comparison of SDEs (ii) and (iii) reveals a slightly fatter tail for (iii). This is because $\max(|a|)$ is larger in this system, and amplification is increased because

products of a's of a given length n, $\prod_{l=0}^{n-1} a(l)x(0)$, can attain higher values.

In summary, power-law tail behaviour is possible in first-order SDEs, although the fatness of the tail will depend on the choice of parameters; fattening bursts are possible if $\max(|a|) > 1$ and the distribution assumes a power-law tail for finite c. Standard PSO is badly approximated by a first-order SDE (however, a novel first-order PSO variant, a simplified version of recombinant PSO, does produce good search behaviour [27]). The next section examines second-order SDEs which provide a much closer model of standard PSO.

4.2. Second-order SDE

The second-order stochastic system with uniform distributions

$$x(t+1) + U(a_l, a_u)x(t) + bx(t-1) = U(c_l, c_u) \qquad (25)$$

has not, to our knowledge, been studied in the burst regime $\max(|a|) > 1$. (25) is closely related to the decoupled PSO with $p_1 = 1$, $p_2 = 0$:

$$x(t+1) + \left[\frac{\phi}{2}(u_1 + u_2) - w - 1\right]x(t) + wx(t-1) = \frac{\phi}{2}u_1. \qquad (26)$$

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

$$x(t+1) + U(-w-1, \phi-w-1)x(t) + wx(t-1) = U\left(0, \frac{\phi}{2}\right) \qquad (27)$$

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

$$x(t+1) + U(-1.75, 1.25)x(t) + 0.75x(t-1) = U(0, 1.5). \qquad (28)$$

Figure 7 shows the cumulative distribution $r = |x|$ for (28) for 50 runs of 10^6 iterations. Comparison with the first-order SDEs of Figure 6 shows that this second-order SDE has very fat tails with 10% of all positions at distances of 4×10^5 or more from the origin. The frequent and large bursts are not well modelled by a power law.

This behaviour is in contrast to the decoupled PSO of (26), plotted in Figure 3, which shows a very much more restrained tail. The principle difference between (26) and (28) is the replacement of the triangular-shaped random number distribution by a uniform distribution. Although $\langle a \rangle$ and $\max(|a|)$ are identical, the uniform distribution increases $\text{prob}(|a| > 1)$, leading to an increased probability of large products $\prod_{l=0}^{n-1} a(l)$.

The contribution of such products to a burst can be seen by formally deriving a solution for $x(t)$ as a sum over products of random variables. Rewriting (25) as

$$x_n = (La_n - bL^2)x_n + c_n, \qquad (29)$$

where L is the lag operator $Lx_j = x_{j-1}$, $L^2 x_j = x_{j-2}$, $La_j = a_{j-1}L$, $a(t) = -a_j$, and iterating back in time,

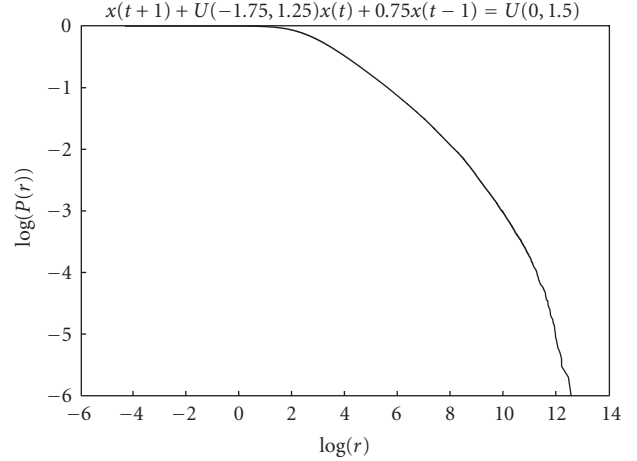$$x_n = (La - bL^2)(La - bL^2)x_n + (La - bL^2)c_n + c_n. \qquad (30)$$

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

FIGURE 7: Investigation of a stochastic second-order equation.

Hence, after m iterations,

$$x_n = (La - bL^2)^m x_n + \left(\sum_{j=0}^{m-1} (La - bL^2)^j\right)c_n. \qquad (31)$$

Equation (31) reduces to the solution of the first-order difference equation, (22), for $m = n$, $b = 0$.

During a large burst $|x_n| \gg x_{n-2m} \gg \max(c)$,

$$\begin{aligned} x_n &\approx (La - bL^2)^m x_n \\ &= ((La)^m - (La)^{m-1}b - (La)^{m-2}b(La) - \cdots)x_n \end{aligned} \qquad (32)$$

which shows that the amplification of the burst is enhanced compared to a first-order burst (24), by the terms in b. The validity of this remark depends on the relative sign of each term appearing in (32). To estimate just how big the enhancement might be, suppose that all the a's are positive and the x's alternate in sign. The probability distribution function $P_m(A)$ of the product of the random multipliers $A = a_1 a - 2 \cdots a_m$ is approximated by $P_m(A) = (p(A^{1/m})^m$ (see, e.g., [9]) for large A, where $p(a)$ is the distribution of a. Then, for large A, we can replace each random variable a in (31) by the geometric mean $a = A^{1/m}$. This gives an upper bound to x_m, $x_m < ((a + |b|)^m x_{m-1}$ where (by assumption) $x_{n-1} > x_{n-2m}$. This second order burst is boosted in comparison to a first-order burst $x_n = a^m x_{m-n}$. Although such large bursts will be rare, they will contribute to the overall probability distribution of x because, although exponentially unlikely, they are of exponential size (see, e.g., the argument of Redner [25] for the product of a binary sequence $xyxxyxyy$).

The oscillatory pattern of any large burst can be deduced by considering $x_{n+1} = a_n x_n - 0.75 x_{n-1}$ where $|x_{n+1}| > |x_n| > |x_{n-1}| \gg \max(c)$. Suppose, for the sake of argument, that x_{n-1} is positive. Then $|x_{n+1}|$ will be maximised if $a_n x_n < 0$ with the result $x_{n+1} < 0$. Hence, x_{n+1} and x_{n-1} differ in sign. x_n might be positive or negative; in either case, we expect x_{n+2} to be of the reverse sign in a large burst. The conclusion is that large bursts are likely to follow a pattern $\text{sign}(x) = \{++ --++--\}$ as demonstrated in a close-up of the burst of Figure 8.
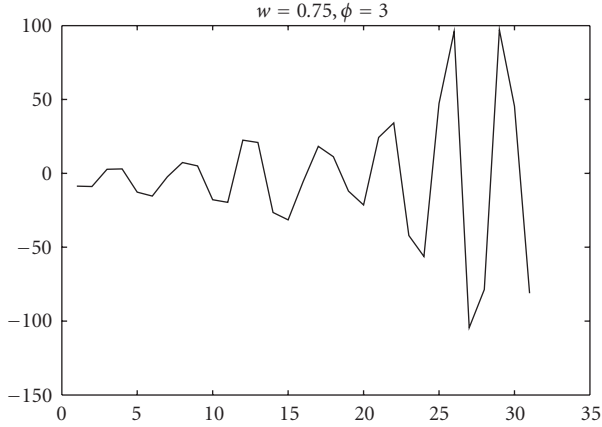
FIGURE 8: Start of a large burst, with deviations $\{++--++--\}$ about $x = 0$.

FIGURE 9: Plot of cumulative probability r for PSO-DR.

5. ADDITIVE RANDOMNESS

If distribution tails in SDEs are caused solely by multiplicative randomness, a second-order SDE with only additive randomness, that is, $a, b = $ const., $c = c(t)$ should be tail free. Recently, a novel PSO variant, recombinant PSO (DR), has been proposed with only additive randomness [10]. Recent work tests PSO-DR for various neighbourhoods and parameter choices with impressively competitive results over a suite of 14 common benchmarks [28]. PSO-DR is similar to the PSO-IW, (9), except that one of the informers is replaced by a discrete recombination of a particle's immediate neighbours in a ring topology,

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

(33)

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 (11) is $0 < \phi^{\text{DR}} < 2(1 + w)$.

Figure 9 reports on the cumulative distribution of particle separation r for (33). 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 cutoff $r - c$, 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 noncritical systems $\phi \leq 2.9$ place the majority of the positions between the attractors. At subcriticality, $\phi = 2.99$, 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 an 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 (33) might produce tails from a resonance effect. This is because the spring constants
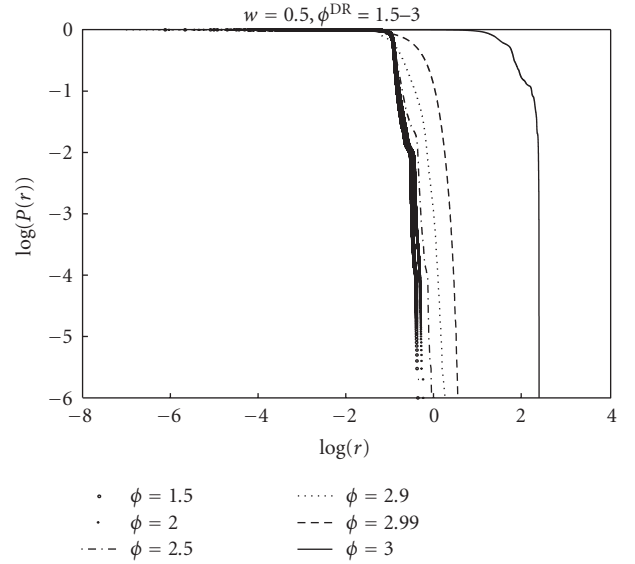
are fixed and the system has a defined natural frequency. For the case of PSO-DR, setting $p_3 = (p_1 + p_2)/2$ gives a simple oscillator with force law $F = \phi(p_3 - x)$ and natural frequency $\omega = \sqrt{(\phi)}$. The periodic time, $T = 2\pi/\omega$ for $\phi = 1$ (this is empirically a good value for interacting PSO-DR [28]), 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 decoupled 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 particle bursts. Bursts occur at all scales with decreasing probability for increasing size. The accumulation of bursts of all sizes results in distribution tail fattening. In order to study how these bursts might develop, decoupled PSO has been formulated as a second-order stochastic difference equation. Fat distribution tails, well modelled by power laws (although the underlying distribution might be log normal, or some other distribution with a power-law regime), arise from multiplicative randomness, a phenomenon previously encountered in first-order SDEs, and generalised here to second order processes.

This conclusion is valid for first- or second-order SDEs where the multiplicative random variable a is capable of amplification ($\max(|a|) > 1$), but does not permit system explosion ($\langle|a|\rangle < 1$). According to the theoretical analysis presented here, bursting in a second-order SDE is built from a sum of multiplicative stochastic processes, and the burst size is boosted by the second-order parameter b. 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 randomness of this sort, does not, according to the theoretical and empirical arguments supplied here, enjoy bursting activity. It is conjectured that this transfer of randomness from the dynamics to the information network generates enough noise to mitigate against premature convergence in the coupled model, despite the thin tails of the decoupled equations. These issues are taken up in [27, 28]. (See also [29].)

PSO bursts differ from the outliers generated by bare bones swarms in two respects: the outliers occur in sequence, and they are one 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 [28]. What does appear to be certain is that a distribution of outliers which decreases slower than a Gaussian tail is important if PSO is to escape premature convergence in difficult environments. Standard PSO achieves this through multiplicative randomness and this occurs even in the decoupled system; recombinant PSO can only, however, gain its outliers through the interaction of the particles.

ACKNOWLEDGMENT

REFERENCES

[1] A. Engelbrecht, *Fundamentals of Computational Swarm Intelligence*, John Wiley & Sons, New York, NY, USA, 2005.

[2] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1942–1948, Perth, Australia, 1995.

[3] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization: an overview," *Swarm Intelligence*, vol. 1, no. 1, pp. 33–57, 2007.

[4] J. Kennedy, "Bare bones particle swarms," in *Proceedings of the IEEE Swarm Intelligence Symposium (SIS '03)*, pp. 80–87, Indianapolis, Ind, USA, April 2003.

[5] T. Richer and T. Blackwell, "The Lévy particle swarm," in *IEEE Congress on Evolutionary Computation (CEC '06)*, pp. 3150–3157, Vancouver, Canada, July 2006.

[6] 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 *Proceedings of the Foundations of Genetic Algorithms Workshop (FOGA '07)*, Mexico City, Mexico, January 2007.

[7] J. Kennedy, "Dynamic-probabilistic particle swarms," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '05)*, pp. 201–207, Washington, DC, USA, June 2005.

[8] J. Kennedy, "Probability and dynamics in the particle swarm," in *Proceedings of the Congress on Evolutionary Computation (CEC '04)*, vol. 1, pp. 340–347, Portland, Ore, USA, June 2004.

[9] D. Sornette, "Multiplicative processes and power laws," *Physical Review E*, vol. 57, no. 4, pp. 4811–4813, 1998.

[10] J. Peña, A. Upegui, and E. Sanchez, "Particle swarm optimization with discrete recombination: an online optimizer for evolvable hardware," in *Proceedings of the1st NASA/ESA Conference on Adaptive Hardware and Systems (AHS '06)*, pp. 163–170, Istanbul, Turkey, June 2006.

[11] D. Bratton and J. Kennedy, "Defining a standard for particle swarm optimization," in *Proceedings of the IEEE Swarm Intelligence Symposium (SIS '07)*, pp. 120–127, Honolulu, Hawaii, USA, April 2007.

[12] Y. Shi and R. Eberhart, "Modified particle swarm optimizer," in *Proceedings of the IEEE Conference on Evolutionary Computation (ICEC '98)*, pp. 69–73, Anchorage, Alaska, USA, May 1998.

[13] J. Kennedy and R. Mendes, "Neighbor topologies in fully-informed and best of-neighborhood particle swarms," in *Proceedings of the IEEE International Workshop on Soft Computing in Industrial Applications (SMCia '03)*, pp. 45–50, June 2003.

[14] M. Clerc and J. Kennedy, "The particle swarm—explosion, stability and convergence in a multi-dimensional space," *IEEE transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2000.

[15] R. Sornette, "Linear stochastic dynamics with nonlinear fractal properties," *Physica A*, vol. 250, no. 1–4, pp. 295–314, 1998.

[16] J. D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, NJ, USA, 1994.

[17] E. Ozcan and C. K. Mohan, "Particle swarm optimization: surfing the waves," in *Proceedings of the Congress on Evolutionary Computation (CEC '99)*, vol. 3, p. 1944, Washington, DC, USA, July 1999.

[18] F. van den Burgh, *An analysis of particle swarm optimizers*, Ph.D. thesis, University of Pretoria, Pretoria, South Africa, 2002.

[19] R. Poli and D. Broomhead, "Exact analysis of the sampling distribution for the canonical particle swarm optimiser and its convergence during stagnation," in *Proceedings of the 9th Annual Genetic and Evolutionary Computation Conference (GECCO '07)*, pp. 134–141, London, UK, July 2007.

[20] G. K. Zipf, *Human Behaviour and the Principle of Least Action*, Addison Wesley, Reading, Mass, USA, 1949.

[21] M. E. J. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, 2005.

[22] C. T. P. Bak and K. Wiesenfeld, "Self-organized criticality: an explanation for 1/f noise," *Physical Review Letters*, vol. 59, no. 4, pp. 381–384, 1987.

[23] D. Sornette, "Mechanism for power laws without self-organization," *International Journal of Modern Physics*, vol. 13, no. 2, pp. 133–136, 2001.

[24] E. Ott, *Chaos in Dynamical Systems*, Cambridge University Press, New York, NY, USA, 2002.

[25] S. Redner, "Random multiplicative processes," *American Journal of Physics*, vol. 58, no. 3, pp. 267–273, 1990.

[26] D. Sornette and R. Cont, "Convergent multiplicative processes repelled from zero: power laws and truncated power laws," *Journal de Physique I France*, vol. 7, pp. 431–444, 1997.

[27] D. Bratton and T. Blackwell, "A simplified recombinant PSO," to appear in *Artificial Evolution and Applications*.

[28] D. Bratton and T. Blackwell, "Understanding particle swarms through simplification: a study of recombinant PSO," in *Proceedings of the 9th Annual Genetic and Evolutionary Computation Conference (GECCO '07)*, pp. 2621–2628, London, UK, July 2007.

[29] M. Clerc, *Particle Swarm Optimization*, ISTE, London, UK, 2006.