Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information Sciences 176 (2006) 937–971
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

A study of particle swarm optimization particle trajectories

F. van den Bergh, A.P. Engelbrecht *

Department of Computer Science, University of Pretoria, Roperstreet, Pretoria 0002, South Africa

Received 27 August 2004; accepted 2 February 2005

Abstract

Particle swarm optimization (PSO) has shown to be an efficient, robust and simple optimization algorithm. Most of the PSO studies are empirical, with only a few theoretical analyses that concentrate on understanding particle trajectories. These theoretical studies concentrate mainly on simplified PSO systems. This paper overviews current theoretical studies, and extend these studies to investigate particle trajectories for general swarms to include the influence of the inertia term. The paper also provides a formal proof that each particle converges to a stable point. An empirical analysis of multi-dimensional stochastic particles is also presented. Experimental results are provided to support the conclusions drawn from the theoretical findings.

Keywords: Particle swarm optimization; Particle trajectories; Equilibrium; Convergence

* Corresponding author. Tel.: +27 12 420 3578; fax: +27 12 362 5188.
E-mail address: engel@cs.up.ac.za (A.P. Engelbrecht).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2005.02.003

1. Introduction

Particle swarm optimization is a stochastic population based optimization approach, first published by Kennedy and Eberhart in 1995 [14,7]. Since its first publication, a large body of research has been done to study the performance of PSO, and to improve its performance. From these studies, much effort has been invested to obtain a better understanding of the convergence properties of PSO. These studies concentrated mostly on a better understanding of the basic PSO control parameters, namely the acceleration coefficients, inertia weight, velocity clamping, and swarm size [12,22,21,23,3,1,18]. From these empirical studies it can be concluded that the PSO is sensitive to control parameter choices, specifically the inertia weight, acceleration coefficients and velocity clamping. Wrong initialization of these parameters may lead to divergent or cyclic behaviour.

The empirical PSO studies do, however, provide some insight into the behaviour of particle swarms (PS), providing guidelines for parameter initialization. For example, Eberhart and Shi found empirically that an inertia weight of 0.7298 and acceleration coefficients of 1.49618 are good parameter choices, leading to convergent trajectories [10]. While such empirically obtained values do work well (in general), they should be considered with care, since the corresponding empirical studies are based on only a limited sample of problems. It should also be noted that PSO control parameters are usually problem dependent.

To gain a better, general understanding of the behaviour of particle swarms, indepth theoretical analyses of particle trajectories are necessary. A few theoretical studies of particle trajectories can be found, which concentrate on simplified PSO systems [19,20,6,27,26,29,31]. These studies facilitated the derivation of heuristics to select parameter values for guaranteed convergence to a stable point. This paper overviews these theoretical studies, and generalizes to more general PSO systems which includes the inertia component. The paper also provides a formal proof that particles converge to a stable point. This point is formally defined.

The remainder of the paper is organized as follows: Section 2 provides a short overview of PSO, outlining some of the problems that have been experienced, and proposed solutions. Section 3 overviews the first theoretical study of particle trajectories based on a simplified PSO system. Analysis of constricted trajectories is summarized in Section 4. Trajectory analysis is expanded in Section 5 to include the inertia weight, and a proof is provided to show that particles converge to a stable point under specific conditions. Section 5.6 generalizes to stochastic particles, with example trajectories being illustrated and discussed. Some experimental results are given in Section 6 to support the theoretical findings.

2. Particle swarm optimization

Particle swarm optimization (PSO) is a stochastic optimization approach which maintains a swarm of candidate solutions, referred to as *particles* [14,7]. Particles are "flown" through hyper-dimensional search space, with each particle being attracted towards the best solution found by the particle's neighborhood and the best solution found by the particle. The position, $\mathbf{x}_i$, of the ith particle is adjusted by a stochastic velocity $\mathbf{v}_i$ which depends on the distance that the particle is from its own best solution and that of its neighborhood. For the original PSO [14,7],

$$v_{ij}(t+1) = v_{ij}(t) + \phi_{1j}(t)(y_{ij}(t) - x_{ij}(t)) + \phi_{2j}(t)(\hat{y}_{ij}(t) - x_{ij}(t)) \tag{1}$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \tag{2}$$

for $i = 1, \ldots, s$ and $j = 1, \ldots, n$, where

$\phi_{1j}(t) = c_1 r_{1j}(t)$ and $\phi_{2j}(t) = c_2 r_{2j}(t)$,
s is the total number of particles in the swarm,
n is the dimension of the problem, i.e. the number of parameters of the function being optimized,
c_1 and c_2 are acceleration coefficients,
$r_{1j}(t), r_{2j} \sim U(0,1)$,
$\mathbf{x}_i(t)$ is the position of particle i at time step t,
$\mathbf{v}_i(t)$ is the velocity of particle i at time step t,
$\mathbf{y}_i(t)$ is the personal best solution of particle i, at time step t,
$\hat{\mathbf{y}}_i(t)$ is the best position found by the neighborhood of particle i, at time step t.

From Eq. (1), the velocity of a particle is determined by three factors:

- $\mathbf{v}_i(t)$, which serves as a momentum term to prevent excessive oscillations in search direction.
- $\phi_1(t)(\mathbf{y}_i(t) - \mathbf{x}_i(t))$, referred to as the *cognitive component*. This component represents the distance that a particle is from the best solution, $\mathbf{y}_i(t)$, found by itself. The cognitive component represents the natural tendency of individuals to return to environments where they experienced their best performance.
- $\phi_2(t)(\hat{\mathbf{y}}_i(t) - \mathbf{x}_i(t))$, referred to as the *social component*. This component represents the distance that a particle is from the best position found by its neighborhood. It represents the tendency of individuals to follow the success of other individuals.

In the social component, $\hat{\mathbf{y}}_i(t)$ represents the best solution found by the neighborhood of particle i. Neighborhood topologies are used to constrict the information flow between particles. A number of neighborhood topologies

have been investigated [13,16,15,9], of which the star, ring and Von Neumann topologies have shown to be the most popular. This paper concentrates on the star topology where the neighborhood of each particle is the entire swarm, with the resulting PSO algorithm referred to as the *gbest* PSO. In this case, $\hat{\mathbf{y}}_i(t) = \hat{\mathbf{y}}(t)$ for all particles $i = 1, \ldots, s$, where (assuming a minimization problem)

$$\hat{\mathbf{y}}(t) = \min_{i=1,\ldots,s} \{\mathbf{y}_i(t)\} \tag{3}$$

Algorithm 1 summarizes the *gbest* PSO algorithm.

Algorithm 1 (*gbest* PSO).

```
Create and initialize a n-dimensional swarm, S;
repeat
   for each particle i = 1, ... ,s
      if f(S·x_i) < f(S·y_i)
         S·y_i = S·x_i;
      end
      if f(S · y_i) < f(S · ŷ)
         S · ŷ = S · y_i;
      end
   end
   for each particle i = 1, ... ,s
   update the velocity using Eq. (1);
   update the position using Eq. (2);
   end
until stopping condition is true;
```

The remainder of this section summarizes popular parameter choices and problems experienced with the original version of PSO.

2.1. Velocity clamping

Initial PSO studies used $c_1 = c_2 = 2.0$. Although good results have been obtained, it was observed that velocities quickly exploded to large values, especially for particles far from their global best, $\hat{\mathbf{y}}$, and personal best, $\mathbf{y}_i$, positions. Consequently, particles have large position updates, with particles leaving the boundaries of their search space. To control the increase in velocity, velocities are clamped, i.e. [9]

$$v_{ij}(t+1) = \begin{cases} v'_{ij}(t+1) & \text{if } v'_{ij}(t+1) < V_{\max,j} \\ V_{\max,j} & \text{if } v'_{ij}(t+1) \geqslant V_{\max,j} \end{cases} \tag{4}$$

where $v'_{ij}(t)$ is calculated using Eq. (1), and $V_{\max,j}$ is maximum allowed velocity in dimension j. While velocity clamping does not prevent a particle from leaving the boundaries of its search space, it does limit the particle step sizes, thereby restricting divergent behaviour.

2.2. Inertia weight

The inertia weight was introduced by Shi and Eberhart to eliminate the need for velocity clamping, but to still restrict divergent behaviour [21,8]. The inertia weight, w, controls the momentum of the particle by weighing the contribution of the previous velocity–basically controlling how much memory of the previous flight direction will influence the new velocity. The velocity equation changes to

$$v_{ij}(t+1) = wv_{ij}(t) + \phi_{1j}(t)(y_{ij}(t) - x_{ij}(t)) + \phi_{2j}(t)(\hat{y}_{ij}(t) - x_{ij}(t)) \tag{5}$$

Initial empirical studies of PSO with inertia have shown that the value of w is extremely important to ensure convergent behaviour [22,10]. For $w > 1$, velocities increase over time causing divergent behaviour. Particles fail to change direction in order to move back towards promising areas. For $w < 0$, particles decelerate until their velocities reach zero (depending on the values of the acceleration coefficients, as indicated in Section 5).

Empirical results have shown that a constant inertia of $w = 0.7298$ and acceleration coefficients with $c_1 = c_2 = 1.49618$ provide good convergent behaviour [10]. While static inertia values have been used successfully, adaptive inertia values have also shown to lead to convergent behaviour [25,30,5,24,17,28].

Although the inertia term helped to ensure convergent trajectories, early empirical studies found that for certain problems velocity clamping were still necessary to prevent velocities to explode. Section 5 shows that this problem can be solved by careful selection of w, c_1 and c_2.

3. Surfing the waves

Ozcan and Mohan produced the first theoretical studies of particle trajectories [19,20]. In their first study [19] a simplified PSO system was considered with

- one particle in the swarm, i.e. $s = 1$,
- one-dimensional particles, i.e. $n = 1$,
- $y(t) = \hat{y}(t) = p$ kept constant, and
- no stochastic component, i.e. $\phi_1(t) = \phi_1$ and $\phi_2(t) = \phi_2$ for all t.

Ozcan and Mohan later generalized their findings to a PSO system with multiple, multi-dimensional particles with $\mathbf{y}_i(t) = \mathbf{y}_i$ and $\hat{\mathbf{y}}(t) = \hat{\mathbf{y}}$ not necessarily the

same point [20]. For the more general system, Ozcan and Mohan derived the following particle position update equation [20]:

$$x_{ij}(t) - (2 - \phi_1 - \phi_2)x_{ij}(t-1) + x_{ij}(t-2) = \phi_1 y_{ij} + \phi_2 \hat{y}_j \tag{6}$$

From Eq. (6), the following closed form can be obtained:

$$x_{ij}(t) = \Lambda_{ij} \sin(\theta_{ij} t) + \Gamma_{ij} \cos(\theta_{ij} t) + \kappa_{ij} \tag{7}$$

where Λ_{ij}, Γ_{ij} and κ_{ij} are constants derived from the initial conditions and the values of ϕ_1 and ϕ_2:

$$\Lambda_{ij} = \frac{2v_{0,ij} - (\phi_{1,ij} + \phi_{2,ij})x_{0,ij} + \phi_{1,ij} y_{ij} + \phi_{2,ij} \hat{y}_j}{||\gamma_{ij}||} \tag{8}$$

$$\Upsilon_{ij} = x_{0,ij} - k_{1,ij} \tag{9}$$

$$k_{1,ij} = \frac{\phi_{1,ij} y_{ij} + \phi_{2,ij} \hat{y}_j}{\phi_{1,ij} + \phi_{2,ij}} \tag{10}$$

which is valid for $\mathbf{0} < \phi_1, \phi_2 < \mathbf{2}$.

The main conclusion from their work is that particle trajectories follow periodic sinusoidal waves. An optimum is searched by randomly 'catching' another wave, and manipulating its frequency and amplitude. In addition to this general finding, Ozcan and Mohan studied the trajectories of particles under a number of special cases. The reader is referred to [19,20] for these studies and example trajectories.

4. Constricted trajectories

For the same simple PSO system as given in Section 3, Clerc and Kennedy provided a theoretical analysis of particle tajectories to ensure convergence to a stable point [6],

$$p = \frac{\phi_1 y + \phi_2 \hat{y}}{\phi_1 + \phi_2} \tag{11}$$

The main result of this work is the introduction of the constriction coefficient and different classes of constriction models. The objective of this theoretically derived constriction coefficient is to prevent the velocity to grow out of bounds, with the advantage that, theoretically, velocity clamping is no longer required. As a result of this study, the velocity equation changes to [4,6]

$$v_{ij}(t+1) = \chi[v_{ij}(t) + \phi_{1j}(t)(y_{ij}(t) - x_{ij}(t)) + \phi_{2j}(t)(\hat{y}_{ij}(t) - x_{ij}(t))] \tag{12}$$

where χ is the constriction coefficient calculated as

$$\chi = \frac{2\kappa}{|2 - \phi - \sqrt{\phi^2 - 4\phi}|} \tag{13}$$

with $\phi = \phi_1 + \phi_2 \geqslant 4$ and $\kappa \in [0, 1]$. The constant κ controls the speed of convergence. For $\kappa \approx 0$, fast convergence to a stable point is obtained, while a $\kappa \approx 1$ results in slow convergence.

The reader is referred to [6] for a more detailed derivation of the constriction coefficient and models.

5. Unconstricted trajectories

The particle trajectory analyses done by Ozcan and Mohan [19,20], and Clerc and Kennedy [6] consider a simplified PSO system without an inertia term. This section presents an analysis with the inertia weight included (the original work can be found in the thesis [27]). Section 5.1 first proves that each particle converges on a stable point, and this point is formally defined. Particle trajectory equations with inertia included is also derived in Section 5.1. A heuristic to select the best values for w, c_1 and c_2 is derived in Section 5.2. Convergence of the original PSO is discussed in Section 5.3 based on the parameter selection heuristic. Section 5.4 illustrates convergence for specific parameter choices, and example deterministic trajectories are given in Section 5.5. The analysis is then tested on stochastic particles in Section 5.6.

5.1. Particle attractor

This section proves that each particle i of a *gbest* PSO converges to a stable point, $\mathbf{p}_i$. That is, considering the sequence $\{\mathbf{x}_i(t)\}_{t=0}^{+\infty}$ of particle positions, the purpose is to show that

$$\lim_{t \to +\infty} \mathbf{x}_i(t) = \mathbf{p}_i \tag{14}$$

Nothing will be said about whether $\mathbf{p}$ represents a local or global optimum. If it can be shown that each particle converges to the point $\mathbf{p}_i$, then it can be reasoned that the entire swarm is in equilibrium.

For this purpose, consider a simplified system as defined in Section 3, but with velocity and position equations defined as

$$v(t+1) = wv(t) + \phi_1(y - x(t)) + \phi_2(\hat{y} - x(t)) \tag{15}$$

$$x(t+1) = x(t) + v(t+1) \tag{16}$$

Considering the velocity and positions of a particle at discrete time steps, by substitution of Eq. (15) into Eq. (16), the following non-homogeneous recurrence relation is obtained:

$$x_{t+1} = (1 - w - \phi_1 - \phi_2)x_t - wx_{t-1} + \phi_1 y + \phi_2 \hat{y} \tag{17}$$

This recurrence relation can be written in matrix-vector notation as the product

$$\begin{bmatrix} x_{t+1} \\ x_t \\ 1 \end{bmatrix} = \begin{bmatrix} 1 + w - \phi_1 - \phi_2 & -w & \phi_1 y + \phi_2 \hat{y} \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ x_{t-1} \\ 1 \end{bmatrix} \tag{18}$$

The characteristic polynomial of the matrix in Eq. (18) is

$$(1 - \lambda)(w - \lambda(1 + w - \phi_1 - \phi_2) + \lambda^2) \tag{19}$$

The solutions to this polynomial gives the eigenvalues

$$\lambda_1 = \frac{1 + w - \phi_1 - \phi_2 + \gamma}{2} \tag{20}$$

$$\lambda_2 = \frac{1 + w - \phi_1 - \phi_2 - \gamma}{2} \tag{21}$$

where

$$\gamma = \sqrt{(1 + w - \phi_1 - \phi_2)^2 - 4w} \tag{22}$$

For initial conditions, $x(0) = x_0$ and $x(1) = x_1$, the explicit closed form of the recurrence relation is then given by

$$x_t = k_1 + k_2\lambda_1^t + k_3\lambda_2^t \tag{23}$$

where

$$k_1 = \frac{\phi_1 y + \phi_2 \hat{y}}{\phi_1 + \phi_2}$$

$$k_2 = \frac{\lambda_2(x_0 - x_1) - x_1 + x_2}{\gamma(\lambda_1 - 1)}$$

$$k_3 = \frac{\lambda_1(x_1 - x_0) + x_1 - x_2}{\gamma(\lambda_2 - 1)}$$

and $x_3 = (1 + w - \phi_1 - \phi_2)x_1 - wx_0 + \phi_1 y + \phi_2 \hat{y}$.

Note that the above equations assume that $y(t) = y$ and $\hat{y}(t) = \hat{y}$ for all t. The closed form representation in Eq. (23) therefore remains valid until a better position x (and thus y and $\hat{y}$) is discovered. When a better position is discovered, the above equations can be used again after recalculating the coefficients k_1, k_2 and k_3.

The remainder of this section shows that

$$\lim_{t\to+\infty} x_t = k_1 = \frac{\phi_1 y + \phi_2 \hat{y}}{\phi_1 + \phi_2} \tag{24}$$

From Eq. (23), the convergence of the sequence to the stable point k_1 depends on the magnitude of the eigenvalues, λ_1 and λ_2. Both these eigenvalues depend on the value of γ. Since γ is a complex number when $(1 + w - \phi_1 - \phi_2)^2 < 4w$, the eigenvalues will also be complex. Fig. 1 depicts the range of ϕ_1, ϕ_2 and w values for which γ is a complex number.

For an arbitrary complex number, z, the magnitude is expressed as the L_2 norm,

$$||z|| = \sqrt{(\Re(z))^2 + (\Im(z))^2}$$

Any complex number, z^t, can be written as

$$z^t = (||z||e^{\imath\theta})^t \tag{25}$$

$$= ||z||^t(\cos(\theta t) + \imath \sin(\theta t)) \tag{26}$$

where $\theta = \arg(z)$. Now, the limit

$$\lim_{t\to+\infty} z^t = \lim_{t\to+\infty} (||z||^t(\cos(\theta t) + \imath \sin(\theta t)))$$

exists only when $\|z\| < 1$, in which case $\lim_{t\to+\infty} z^t = 0$. The trajectory in Eq. (23) will therefore diverge when either $\|\lambda_1\| > 1$ or $\|\lambda_2\| > 1$. In other words, when $\max\{\|\lambda_1\|, \|\lambda_2\|\} > 1$, the limit

$$\lim_{t\to+\infty} (k_1 + k_2\lambda_1^t + k_3\lambda_2^t)$$

does not exist. Furthermore, if, for any complex number z, $\|z\| = 1$, then the limit,

$$\lim_{t\to+\infty} z^t = \lim_{t\to+\infty} (1^t(\cos(\theta t) + \imath \sin(\theta t))) \tag{27}$$

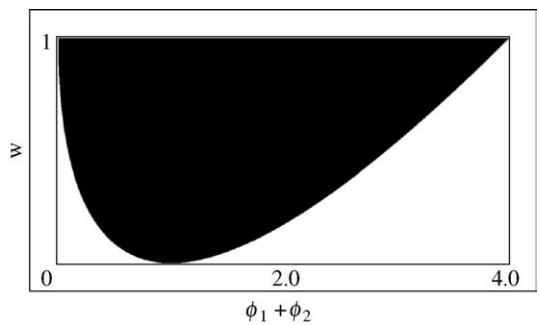

Fig. 1. Values of ϕ_1, ϕ_2 and w for which γ is complex.

does not exist. Therefore, if either $\|\lambda_1\| = 1$ or $\|\lambda_2\| = 1$, then the sequence $\{x_t\}_{t=0}^{+\infty}$ diverges.

The sequence converges when $\max\{\|\lambda_1\|, \|\lambda_2\|\} < 1$, since $\lim_{t\to+\infty}\lambda_1^t = 0$ if $\|\lambda_1\| < 1$ and $\lim_{t\to+\infty}\lambda_2^t = 0$ if $\|\lambda_2\| < 1$. In this case,

$$\lim_{t\to+\infty} x_t = k_1 = \frac{\phi_1 y + \phi_2 \hat{y}}{\phi_1 + \phi_2}$$

This means that, under the condition that $\max\{\|\lambda_1\|, \|\lambda_2\|\} < 1$, a particle converges to a weighted average of its personal best and global best positions.

In the case that ϕ_1 and ϕ_2 are stochastic, c_1 and c_2 can be considered as upper bounds for ϕ_1 and ϕ_2 respectively. The average behaviour of the system can then be observed by considering the expected values of ϕ_1 and ϕ_2 (assuming uniform distributions):

$$E[\phi_1] = c_1 \int_0^1 \frac{x}{1-0}\,\mathrm{d}x = c_1 \frac{x}{2}\Big|_0^1 = \frac{c_1}{2}$$

$$E[\phi_2] = c_2 \int_0^1 \frac{x}{1-0}\,\mathrm{d}x = c_2 \frac{x}{2}\Big|_0^1 = \frac{c_2}{2}$$

Using the expected values, the limit becomes

$$\lim_{t\to+\infty} x_t = \frac{\frac{c_1}{2} y + \frac{c_2}{2}\hat{y}}{\frac{c_1}{2} + \frac{c_2}{2}} = \frac{c_1 y + c_2 \hat{y}}{c_1 + c_2}$$

In general, for arbitrary values of c_1 and c_2,

$$\begin{aligned}\lim_{t\to+\infty} x_t &= \frac{c_1 y + c_2 \hat{y}}{c_1 + c_2} = \frac{c_1}{c_1 + c_2} y + \frac{c_2}{c_1 + c_2}\hat{y} = \left(1 - \frac{c_2}{c_1 + c2}\right) y + \frac{c_2}{c_1 + c_2}\hat{y} \\ &= (1 - a) y + a\hat{y}\end{aligned}$$

where $a = \frac{c_2}{c_1 + c_2} \in [0, 1]$.

This section has formally proven that the trajectory of a simple particle with inertia converges to a stable point, which is a weighted average of y (the personal best position) and $\hat{y}$ (the global best position).

5.2. Parameter selection heuristic

Section 5.1 proved that a particle converges to the point $\frac{c_1 y + c_2 \hat{y}}{c_1 + c_2}$ if and only if $\max\{\|\lambda_1\|, \|\lambda_2\|\} < 1$, where λ_1 and λ_2 are the eigenvalues of the matrix describing the dynamics of a simple PSO system with inertia. A question that remains to be answered is how to select the PSO control parameters, w, c_1 and c_2 to ensure that $\max\{\|\lambda_1\|, \|\lambda_2\|\} < 1$. This section derives such a heuristic.

Fig. 2 is an experimentally obtained map visualizing the ϕ_1, ϕ_2 and w values leading to convergence or divergence. The map was constructed by

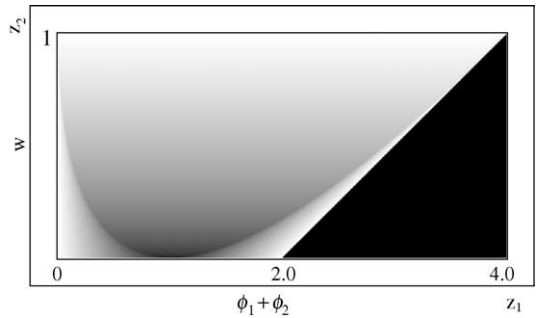

Fig. 2. The black triangle to the bottom right represents the values for which the particle strictly diverges, i.e. $\max(\|\lambda_1\|, \|\lambda_2\|) > 1$. This is the region for which $w < 0.5(\phi_1 + \phi_2) - 1$. The lighter regions represent the magnitude of $\max(\|\lambda_1\|, \|\lambda_2\|)$, with white representing magnitude 1. The darker regions (outside of the divergent region) represent values leading to more rapid convergence.

sampling the values of $\phi_1 + \phi_2$ and w on a regular grid, using 1000 horizontal samples and 500 vertical samples. The intensity of each point on the grid represents the magnitude $\max(\|\lambda_1\|, \|\lambda_2\|)$, with lighter shades representing larger magnitudes, except for the black triangular shape observed in the bottom right corner of the map. This triangle corresponds to the values of ϕ_1, ϕ_2 and w resulting in $\max(\|\lambda_1\|, \|\lambda_2\|) > 1$, which implies that the trajectory of the particle will diverge when using these values. Fig. 2 should also be compared to Fig. 1 to see the relationship between the magnitude $\max(\|\lambda_1\|, \|\lambda_2\|)$ and whether γ has a non-zero imaginary component. Note that all the parameter values leading to a divergent trajectory have real-valued γ values, since the entire divergent triangle falls inside the white area of Fig. 1. The parameters that correspond to real-valued γ values that do fall inside the convergent area of Fig. 2 have relatively large magnitudes, as can be seen from their lighter shading.

The trajectory of a particle can be guaranteed to converge if the parameters ϕ_1, ϕ_2 and w are chosen such that the corresponding point on the map in Fig. 2 always falls in the convergent region. Let z_1 represent the horizontal axis, associated with $\phi_1 + \phi_2$ and z_2 the vertical axis, associated with w. If we take into account that c_1 and c_2 represent the upper limits of ϕ_1 and ϕ_2, respectively, so that $\phi_1 \in [0, c_1]$ and $\phi_2 \in [0, c_2]$, then the range of values that $\phi_1 + \phi_2$ can assume (to ensure convergence) occur to the left of the vertical line $z_1 = c_1 + c_2$ in the figure. Searching vertically along line $z_1 = c_1 + c_2$ for the point where it exits the black divergent triangle yields the smallest w value that will result in a convergent trajectory (i.e. a point outside of the divergent triangular region). The coordinates of this intersection are $(c_1 + c_2,\ 0.5(c_1 + c_2) - 1)$. All w values larger than this critical value will also lead to convergent trajectories, so the general relation

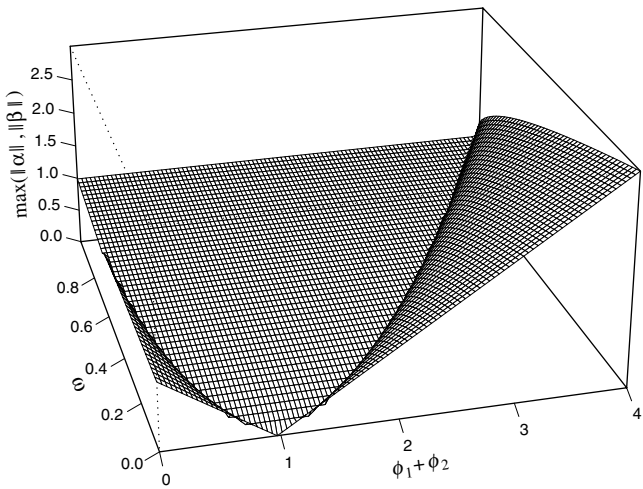

Fig. 3. A 3-dimensional representation of the value $\max(\|\lambda_1\|, \|\lambda_2\|)$. Note that the elevation of the furled right-hand bottom corner exceeds 1.0.

$$w > \tfrac{1}{2}(c_1 + c_2) - 1 \tag{28}$$

can be defined to characterize these values.

Fig. 3 is an alternate representation of Fig. 2. Note that the magnitude $\max(\|\lambda_1\|, \|\lambda_2\|)$ gradually increases from 0 to about 2.5—this is especially visible in the furled right-hand bottom corner. Keep in mind that all magnitudes greater or equal to 1.0 lead to divergent trajectories.

5.3. Original PSO convergence

The original particle swarm, with $c_1 = c_2 = 2$ and $w = 1$, is a boundary case, with $0.5(2 + 2) - 1 = 1 = w$. Further insight can be gained by directly calculating the value of $\max(\|\lambda_1\|, \|\lambda_2\|)$ using explicit formulae (20) and (21). This is achieved by setting $\phi_1 = \phi_2 = 2$, since the maximum of these two values are determined by their respective upper bounds with values $c_1 = c_2 = 2$, yielding $\|\lambda_1\| = \|\lambda_2\| = 1$. This would seem to imply that the original PSO equations resulted in divergent trajectories, according to Eq. (27). This verdict does not take into account the stochastic component, though. Now, considering the stochastic component with $\phi_1 = r_1(t)c_1$ and $\phi_2 = r_2(t)c_2$, where $r_1(t), r_2(t) \sim U(0,1)$, it is clear that $0 < \phi_1, \phi_2 < 2$ when $c_1 = c_2 = 2$. Substituting $\phi = \phi_1 + \phi_2$ into Eq. (22) yields

$$\gamma = \sqrt{(2-\phi)^2 - 4} = \imath\sqrt{4\phi - \phi^2}$$

which in turn yields

$$||\lambda_1|| = \sqrt{\frac{(2-\phi)^2}{4} + \frac{4\phi - \phi^2}{4}} = \sqrt{\frac{4 - 4\phi + \phi^2}{4} + \frac{4\phi - \phi^2}{4}} = 1 \tag{29}$$

Because λ_1 and λ_2 are complex conjugates when γ is complex, this implies that $\|\lambda_2\| = 1$ as well, so that $\max(\|\lambda_1\|, \|\lambda_2\|) = 1$. This means that the trajectory of the particle will be divergent regardless of the value of ϕ, which explains why the original PSO algorithm had to clamp the velocities to the range $[-V_{\max}, V_{\max}]$ to prevent the system from diverging.

Although the intuitive understanding of the concept of a divergent trajectory calls to mind the image of a sequence that grows without bound, a divergent trajectory need not be unbounded. A sequence may oscillate through a set of values without ever converging. This is exactly what happens in the case of the original PSO. Consider the value of λ_1 when $c_1 = c_2 = 2$, that is,

$$\lambda_1 = \frac{(2-\phi)^2 + \imath\sqrt{4\phi - \phi^2}}{2}$$

Recall that a complex number z^t can be represented in exponential form, so that

$$z^t = ||z||^t(\cos(\theta t) + \imath \sin(\theta t)) \tag{30}$$

where $\theta = \arg(z)$. Since $\|\lambda_1\| = 1$, as shown in Eq. (29), Eq. (30) can be reduced to (after substituting λ_1 for z)

$$\lambda_1^t = \cos(\theta t) + \imath \sin(\theta t)$$

This implies that the trajectory of the particles is described by

$$\begin{aligned} x_t &= k_1 + k_2\lambda_1^t + k_3\lambda_2^t \\ &= k_1 + k_2(\cos(\theta t) + \imath \sin(\theta t)) + k_3(\cos(\theta t) - \imath \sin(\theta t)) \\ &= k_1 + (k_2 + k_3)\cos(\theta t) + \imath(k_2 - k_3)\sin(\theta t) \end{aligned}$$

where $\theta = \arg(\lambda_1)$, and $\arg(\lambda_2) = -\theta$, since λ_1 and λ_2 are complex conjugates. The imaginary components cancel for integral values of t. The trajectory of a particle using the original PSO parameter settings thus traces out a superposition of two sinusoidal waves; their amplitudes and frequencies depend on the initial position and velocity of the particle. This is consistent with the findings of Ozcan and Mohan for equivalent parameter settings [19,20].

This clearly shows that the original PSO parameters led to divergent particle trajectories. The next section investigates the characteristics of trajectories obtained using parameter settings from the convergent region indicated in Fig. 2.

5.4. Convergent PSO parameters

Above it was shown that the parameter settings of the original PSO would cause the trajectories of its particles to diverge, were it not for the effect of the $V_{\max}$ clamping strategy. An infinite number of parameter settings exist that do ensure a convergent trajectory, so more information is needed to decide on a particular choice. A brief example will now show that certain parameter choices leads to convergent behaviour without having to clamp the velocities to the range $[-V_{\max}, V_{\max}]$.

One popular choice of parameters is $c_1 = c_2 = 1.49618$ and $w = 0.7298$ [10]. First, note that it satisfies relation (28) since $0.5(1.49618 + 1.49618) - 1 = 0.49618 < 0.7298$. The stochastic behaviour can also be predicted using this relation, so that $0.5(\phi_1 + \phi_2) - 1 < 0.7298$, which is trivially true for $\phi_1 + \phi_2 \in (0, 2 \times 1.49618)$. Substitution into the explicit formulae for λ_1 and λ_2, with $\phi = \phi_1 + \phi_2$, confirms this. Two sets of calculations follow: one for the real-valued γ values, and another set for complex γ values. When $\phi \in [0, 0.02122]$, implying a real-valued γ, then

$$\gamma = \sqrt{(1 + w - \phi)^2 - 4w} \approx \sqrt{0.073 - 3.4596\phi + \phi^2}$$

$$\max(||\lambda_1||, ||\lambda_2||) \approx \frac{1.7298 - \phi + \sqrt{0.073 - 3.4596\phi + \phi^2}}{2} < 1$$

Otherwise, when $\phi \in (0.02122, 2.992]$, resulting in complex γ values,

$$\gamma = \sqrt{(1 + w - \phi)^2 - 4w} \approx \imath\sqrt{-0.073 + 3.4596\phi - \phi^2}$$

$$||\lambda_1|| = ||\lambda_2|| = \sqrt{\frac{(1 + w - \phi)^2}{4} + \frac{-0.073 + 3.4596\phi - \phi^2}{4}} \approx 0.8754$$

Again, note that $\|\lambda_2\| = \|\lambda_1\|$when γ is complex, since they are complex conjugates. Fig. 4 is a cross-section of Fig. 3 along the line $w = 0.7298$. The figure clearly confirms the values derived above. Note that the figure indicates that the maximum value of ϕ can be increased to approximately 3.45 without causing the trajectory to diverge. The benefit of increasing ϕ will be discussed below. To summarize, this example shows that a popular choice of parameter settings leads to a convergent trajectory without having to clamp the velocities of the particles.

To ensure convergence, the value for w should thus be chosen so that it satisfies relation (28). This relation can also be reversed to calculate the values of c_1 and c_2 once a suitable $w < 1$ has been decided on. Note, however, that there is an infinite number of ϕ_1 and ϕ_2 values that satisfy $\phi = \phi_1 + \phi_2$, all exhibiting

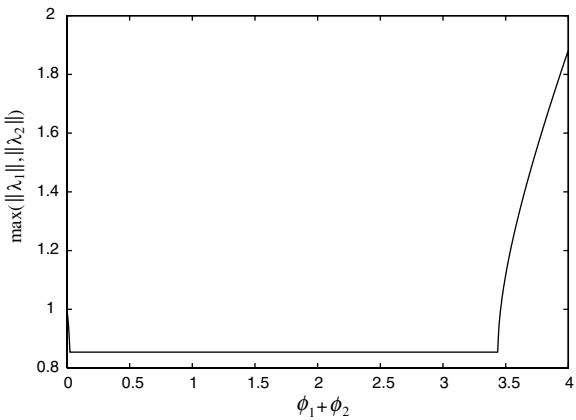

Fig. 4. The magnitude $\max(\|\lambda_1\|, \|\lambda_2\|)$ for $w = 0.7298$ and $\phi_1 + \phi_2 \in (0, 4)$.

the exact same convergence behaviour under the assumptions of this analysis, so it is customary to set $\phi_1 = \phi_2$.

The trajectory of a particle obtained by using a set of parameters leading to convergent behaviour can be associated with a physical phenomenon. First, Eq. (23) is re-written using the alternate representational form for complex numbers to yield

$$\begin{aligned} x(t) &= k_1 + k_2\lambda_1^t + k_3\lambda_2^t \\ &= k_1 + k_2||\lambda_1||^t(\cos(\theta_1 t) + \imath\sin(\theta_1 t)) + k_3||\lambda_2||^t(\cos(\theta_2 t) + \imath\sin(\theta_2 t)) \end{aligned}$$

where $\theta_1 = \arg(\lambda_1)$ and $\theta_2 = \arg(\lambda_2)$. When γ is complex, λ_1 and λ_2 will be complex conjugates. This leads to the simplified form

$$x(t) = k_1 + ||\lambda_1||^t(k_2 + k_3)\cos(\theta t) + \imath||\lambda_1||^t(k_2 - k_3)\sin(\theta t)$$

In this form it is immediately clear that the trajectory of a particle is analogous to the dampened vibrations observed in a spring-dashpot system [2]. The characteristic waveform associated with dampened vibrations is also clearly visible in Fig. 5 below.

5.5. Example deterministic trajectories

This section presents several plots of trajectories that have been obtained using Eq. (23). These trajectories were computed without any stochastic component; plots taking the stochastic component into account are presented in Section 5.6. Figs. 5–7 show examples of the three types of behaviour that the non-stochastic PSO equations can exhibit: convergent, cyclic (a special form of divergent behaviour) and divergent. All these figures have been obtained

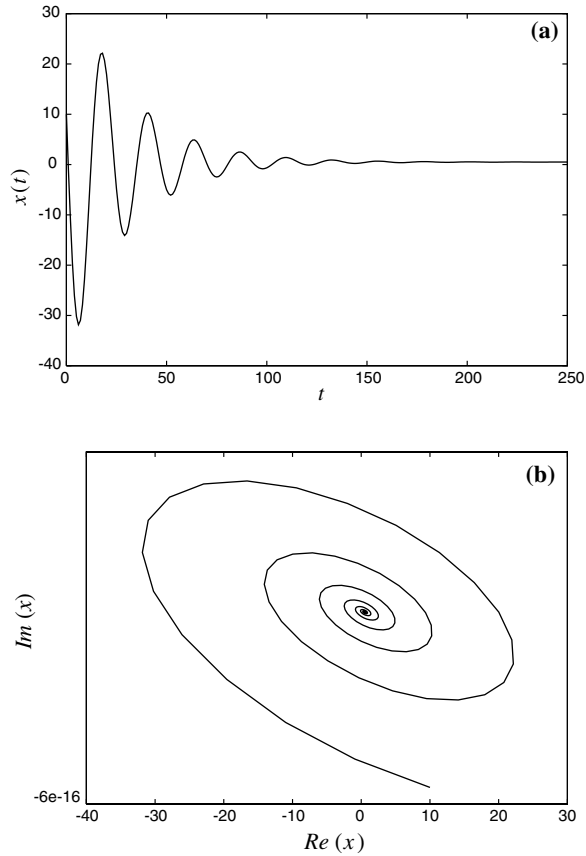

Fig. 5. A convergent particle trajectory, obtained with the parameter settings $w = 0.5$ and $\phi_1 = \phi_2 = 1.4$. (a) Plots the particle position over time; (b) shows the real and complex components of the particle trajectory over the same duration.

experimentally, using 80-bit floating point numbers, with constant values $y = 1.0$ and $\hat{y} = 0$. The initial conditions were $x(0) = 10$, and $x(1) = 10 - 9\phi_1 - 10\phi_2$, with ϕ_1 and ϕ_2 as listed for each figure.

Fig. 5 is a plot of a particle trajectory obtained with a set of parameters that leads to convergence. In Fig. 5 it is clear that the amplitude of the oscillations decays over time. This represents the radius of the search pattern of a particle in search space. Initially the particle will explore a larger area, but the amplitude decreases rapidly until the particle searches a small neighbourhood surrounding $(c_1 y + c_2 \hat{y})/(c_1 + c_2)$. In the complex representation of $x(t)$, Fig. 5(b), where t is any real number (instead of being restricted to integral values), the particle traces out a convergent spiral.

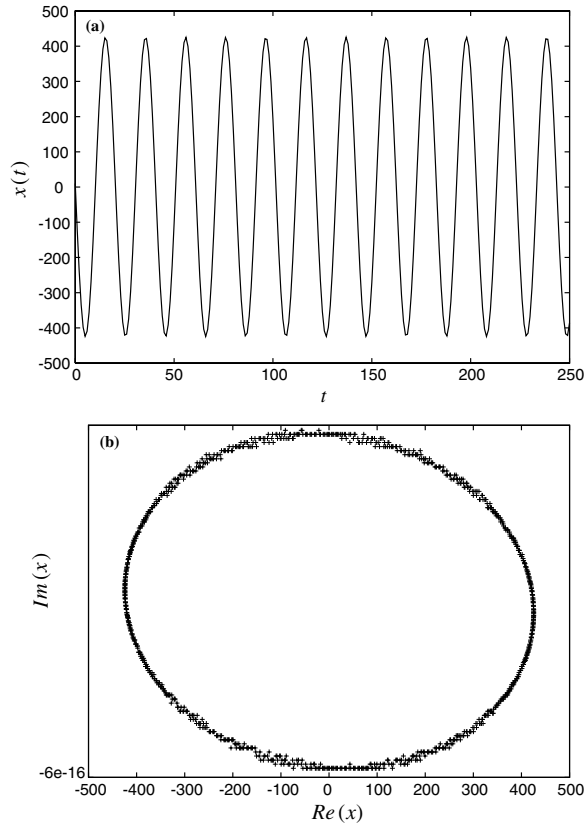

Fig. 6. A cyclic particle trajectory, obtained with the parameter settings $w = 1.0$ and $\phi_1 = \phi_2 = 1.999$. (a) Plots the particle position over time; (b) shows the real and complex components of the particle trajectory over the same duration.

Fig. 6 illustrates the second type of observed behaviour, namely that leading to cyclic trajectories. Some comments on Fig. 6(b) are in order. Ideally, the figure would be a perfect ellipse since the superimposed sine waves should trace out a smooth curve. Because the figure was obtained experimentally, however, the points are somewhat "noisy" due to numerical inaccuracies. Instead of connecting the successive points using line segments (as was done for the other two figures), it was decided, for the sake of clarity, to plot only the points. Fig. 6(a) clearly shows the non-convergent sinusoidal waveform of the particle trajectory.

Fig. 7 exhibits the classic notion of divergence: as time passes, the particle moves (or more accurately, oscillates) further and further from $\hat{y}$, the global best position of the swarm. The spiral in Fig. 7(b) is divergent, and although it looks similar to the one in Fig. 5(b), the scale of the axes clearly show the

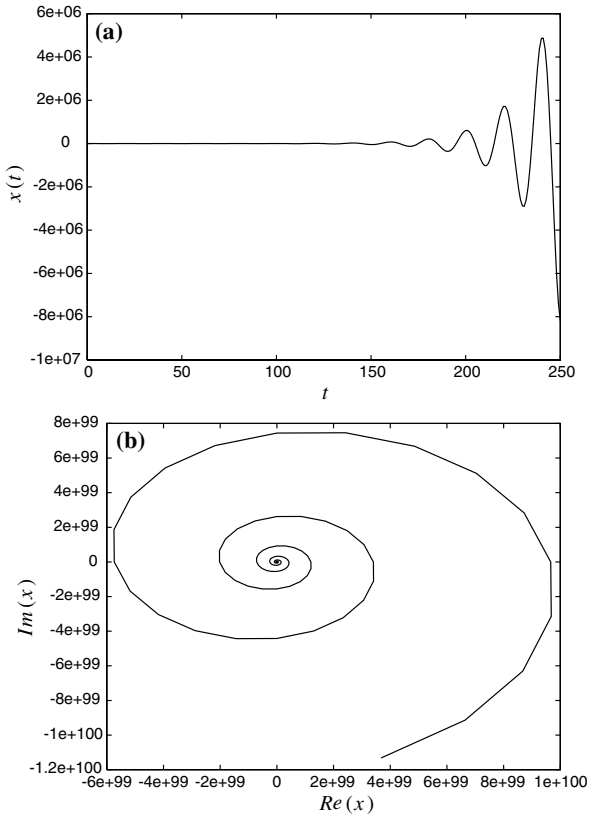

Fig. 7. A divergent particle trajectory, obtained with the parameter settings $w = 0.7$ and $\phi_1 = \phi_2 = 1.9$. (a) Plots the particle position over time; (b) shows the real and complex components of the particle trajectory over the same duration.

divergent behaviour. For a search algorithm this type of trajectory is not generally desirable, since the trajectory will rapidly exceed the numerical range of the machine.

5.6. Trajectories under stochastic influences

In the previous section the stochastic component was treated as a constant by fixing the values of ϕ_1 and ϕ_2. Although it was shown that some characteristics of the trajectory can be applied to whole ranges of ϕ values, it is still not clear what the influence of randomness will be on the trajectory. To investigate these phenomena, several sample trajectories are presented. The following parameters were used for all the experiments: $y = 1.0, \hat{y} = 0, x(0) = 10$ and

$x(1) = 10 - 9\phi_1 - 10\phi_2$, with ϕ_1, ϕ_2 and w set to the values indicated for each experiment. The stochastic values for ϕ_1 and ϕ_2 were sampled so that $0 \leqslant \phi_1 \leqslant c_1$ and $0 \leqslant \phi_2 \leqslant c_2$. Because of the stochastic component, the plots presented in this section were obtained using PSO update Eqs. (5) and (2), instead of the closed form solution offered by Eq. (23). This implies that discrete time was used, so that no phase plots were drawn. The notation x_t in this section therefore refers to the value of x at (discrete) time step t.

Fig. 8 is a plot of the trajectory of a particle using the original PSO parameters, i.e. $w = 1.0$ and $c_1 = c_2 = 2.0$. Notice how the amplitude of the oscillations increases towards the right of the graph, a clear indication of the divergent behaviour of this configuration. This is in agreement with observations in the previous section, where these parameter settings led to cyclic (i.e. divergent) behaviour. The observed increase in amplitude is caused by the randomness in the values of ϕ_1 and ϕ_2. A simple example will illustrate the nature of the problem. Assume the following (quite arbitrary) values for the parameters: $x_t = 10$, $x_{t-1} = 11$, $\phi_1 = c_1 r_1(t) = 1.9$, and $\phi_2 = c_2 r_2(t) = 1.8$. The new position of the particle can be calculated (using Eq. 17) as

$$x_{t+1} = -1.7x_t - x_{t-1} + 1.9y + 1.8\hat{y}$$

thus $x_{t+1} = -26.1$. If another iteration is executed, then

$$x_{t+2} = -1.7x_{t+1} - x_t + 1.9y + 1.8\hat{y}$$

which yields $x_{t+2} = 35.27$. If, however, different stochastic values are used so that $\phi_1 = 0.1$ and $\phi_2 = 0.2$, then

$$x_{t+2} = 1.7x_{t+1} - x_t + 0.1y + 0.2\hat{y}$$

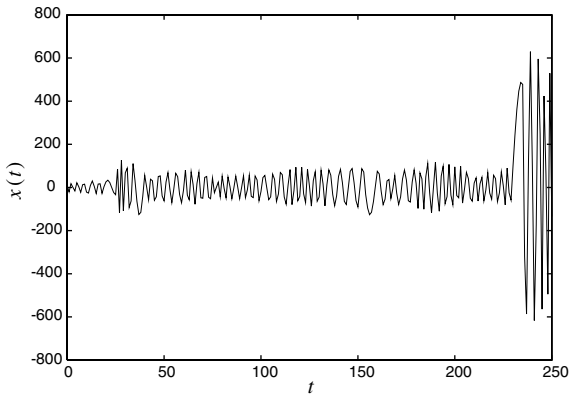

Fig. 8. Stochastic particle trajectory, obtained using $w = 1.0$ and $c_1 = c_2 = 2.0$. Note that the y-axis scale of this figure is on the order of 10^2.

resulting in $x_{t+2} = -55.27$. What this example illustrates is that alternating between large and small values for ϕ_1 and ϕ_2 may increase the distance between x_t and $(1-a)y + a\hat{y}$, instead of decreasing it (or oscillating around it). This is caused in part by the negative sign associated with the x_{t-1} term, which changes the direction that the particle moves in at every alternate time step. A large distance at time step t may result in an almost doubling of the distance in the next time step. If ϕ_1 and ϕ_2 remain constant, then the particle is able to return to its previous position, since the step size is bounded.

Fig. 9 was plotted using $w = 0.9$ and $c_1 = c_2 = 2.0$, again using stochastic values for ϕ_1 and ϕ_2. Notice how the oscillations first appear to increase in amplitude, but then gradually decrease. Near the end of the sample the amplitude increases again, but it eventually decreases at time $t = 300$ (not shown in the figure). Applying relation (28) to the parameters yields $0.5(2+2) - 1 = 2 > 0.9$, which implies that the trajectory will diverge when the upper bounds of ϕ_1 and ϕ_2 are considered. Convergent behaviour emerges when $\phi_1 + \phi_2 < 3.8$. This happens with a probability of $3.8/4 = 0.95$, since $0 < \phi_1 + \phi_2 < 4$ under a uniform distribution. In short, this implies that the trajectory of the particle will converge *most of the time*, occasionally taking divergent steps. The relative magnitude of the divergent steps versus the convergent steps must be taken into account to predict correctly whether the system will converge. Since this information is not available (because of the randomness) it is not possible to make this prediction accurately.

Fig. 10 represents the trajectory of a particle using the parameter settings $w = 0.7$ and $c_1 = c_2 = 1.4$. Applying relation (28) shows that $0.5(1.4 + 1.4) - 1 = 0.4 < 0.7$, so that the trajectory is expected to converge. This is clearly visible in the figure since the initial oscillations decay very rapidly. Minor oscillations, caused by the stochastic influence, remain present though. The parameter settings are now changed so that $w = 0.7$ and $\phi_1 = \phi_2 = 2.0$, as reflected

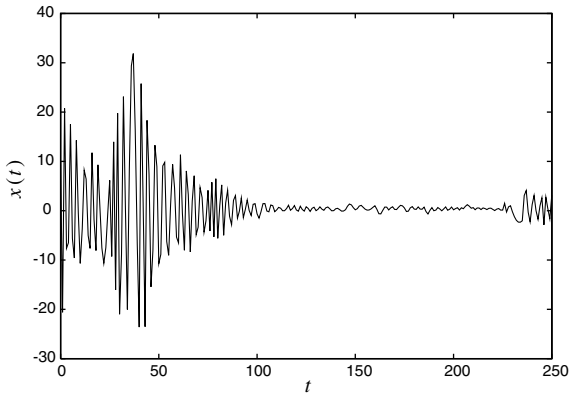

Fig. 9. Stochastic particle trajectory, obtained using $w = 0.9$ and $c_1 = c_2 = 2.0$.

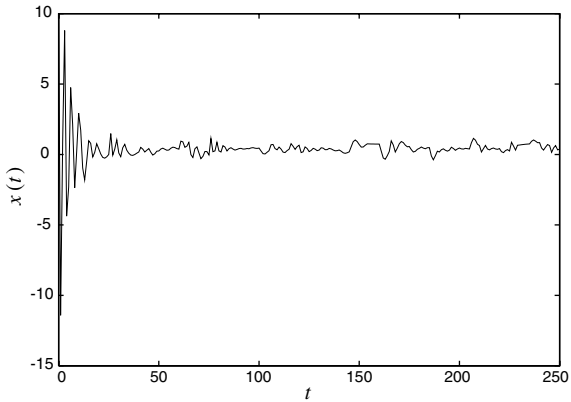

Fig. 10. Stochastic particle trajectory, obtained using $w = 0.7$ and $c_1 = c_2 = 1.4$.

in Fig. 11. Note that relation (28) dictates that the upper bound for this trajectory is divergent, since $0.5(2 + 2) - 1 = 1 > 0.7$. When $\phi_1 + \phi_2 < 3.4$, however, convergent behaviour surfaces again. This happens with probability $3.4/4 = 0.85$, a lower figure than that obtained above with parameter settings $w = 0.9$ and $\phi_1 = \phi_2 = 2$. The trajectory in Fig. 11 appears to have a faster rate of convergence than the one in Fig. 9, though. This is offset by the fact that there are more large "bumps" in Fig. 11, indicating that divergent steps occur more frequently.

These results indicate that it is not strictly necessary to choose the values of c_1 and c_2 so that relation (28) is satisfied for all values of $0 < \phi_1 + \phi_2 < c_1 + c_2$,

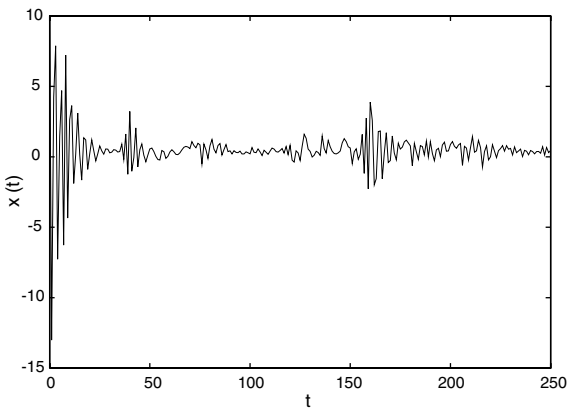

Fig. 11. Stochastic particle trajectory, obtained using $w = 0.7$ and $c_1 = c_2 = 2.0$.

for a given w value. Let ϕ_{crit} denote the largest value of $\phi_1 + \phi_2$ for which relation (28) holds. Then

$$\phi_{\text{crit}} = \sup \phi \mid 0.5\phi - 1 < w, \quad \phi \in (0, c_1 + c_2] \tag{31}$$

All values of $\phi_1 + \phi_2 \leqslant \phi_{\text{crit}}$ then satisfy relation (28). As long as the ratio

$$\phi_{\text{ratio}} = \frac{\phi_{\text{crit}}}{c_1 + c_2} \tag{32}$$

is close to 1.0, the trajectory will converge without too many disruptions. As shown above, even a ϕ_{ratio} of 0.85 resulted in a system that converges without excessively large oscillations.

Extreme cases, like $w = 0.001$ and $c_1 = c_2 = 2.0$, results in $\phi_{\text{ratio}} \approx 1/2$. This system will take divergent steps 50% of the time, but as Fig. 12 shows, the system always "recovers" after taking large divergent steps. The recovery is caused by the fact that roughly 50% of the time the particle will take a step along a *convergent* trajectory. The probabilistically divergent behaviour can have a positive influence on the diversity of the solutions that the particle will examine, thereby improving its exploration capabilities. This property is especially valuable when optimizing functions that contain many local minima.

Holland discussed the balance between *exploration* and *exploitation* that an algorithm must maintain [11]. Exploration ability is related to the algorithm's tendency to explore new regions of the search space, while exploitation is the tendency to search a smaller region more thoroughly. By choosing the PSO parameters carefully, a configuration can be found that maintains the balance reasonably well.

The rule of thumb for choosing the parameters w, ϕ_1 and ϕ_2 is that smaller w values result in faster rates of convergence. This is offset by how frequently a

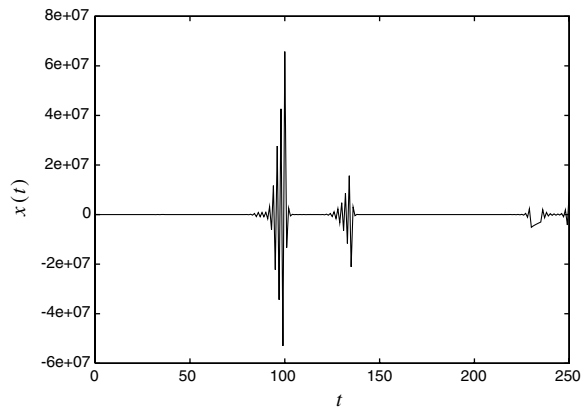

Fig. 12. Stochastic particle trajectory, obtained using $w = 0.001$ and $c_1 = c_2 = 2.0$. Note that the y-axis scale of this figure is on the order of 10^7.

Table 1
Average number of iterations to saturation and average fitness at point of saturation for different parameter choices

Problem	w	c_1, c_2	Iterations	Fitness
Ackley	0.001	2.0	82 141.8 ± 26 951.7	7.530882 ± 7.519820
	0.7298	1.49618	1190.4 ± 217.8	3.651611 ± 1.514420
	0.7	1.4	6104.6 ± 15 786.3	6.424493 ± 1.752388
	0.7	2.0	53 411.4 ± 25 644.2	2.461432 ± 6.735057
	0.9	2.0	10.0 ± 0.0	20.489068 ± 0.240307
	1.0	2.0	10.0 ± 0.0	20.489016 ± 0.237977
Bohachevsky	0.001	2.0	35.6 ± 6.4	0.009398 ± 0.066451
	0.7298	1.49618	128.6 ± 9.0	3.398e−11 ± 3.017e−11
	0.7	1.4	106.4 ± 8.0	3.429e−11 ± 3.271e−11
	0.7	2.0	202.0 ± 22.9	3.113e−11 ± 2.627e−11
	0.9	2.0	76.0 ± 73.5	1.173 ± 1.206
	1.0	2.0	21.2 ± 13.2	4.542 ± 6,807
Colville	0.001	2.0	20 432.6 ± 15 737.6	0.257 ± 0.985
	0.7298	1.49618	4205.4 ± 538.6	4.504e−11 ± 3.031e−11
	0.7	1.4	2980.6 ± 309.3	4.982e−11 ± 2.924e−11
	0.7	2.0	21 355.8 ± 2428.1	4.456e−11 ± 2.805e−11
	0.9	2.0	22.8 ± 17.8	377.614 ± 652.721
	1.0	2.0	13.2 ± 6.5	546.175 ± 805.089
Easom	0.001	2.0	55.4 ± 12.2	−1.0 ± 1.641e−11
	0.7298	1.49618	128.4 ± 14.5	−1.0 ± 3.059e−11
	0.7	1.4	105.8 ± 9.1	−1.0 ± 2.676e−11
	0.7	2.0	211.6 ± 28.0	−1.0 ± 3.032e−11
	0.9	2.0	53.2 ± 77.1	−0.271 ± 0.353
	1.0	2.0	8.0 ± 10.9	−0.045 ± 0.185
Griewank	0.001	2.0	87 157.0 ± 18739.0	101.893 ± 200.817
	0.7298	1.49618	1250.0 ± 180.1	0.052 ± 0.087
	0.7	1.4	13 141.2 ± 23 469.734	0.640 ± 1.003
	0.7	2.0	44 503.6 ± 24 367.4	25.2 ± 124.852
	0.9	2.0	10.0 ± 0.0	613.401 ± 84.197
	1.0	2.0	10.0 ± 0.0	620.5 ± 77.147
Hyperellipsoid	0.001	2.0	81 771.8 ± 26 420.3	534.927 ± 677.028
	0.7298	1.49618	1138.6 ± 163.5	5.043e−11 ± 2.838e−11
	0.7	1.4	11 831.2 ± 21 168.7	0.046 ± 0.221
	0.7	2.0	47 225 ± 29 331.9	178.207 ± 558.911
	0.9	2.0	10.0 ± 0.0	1875.976 ± 285.891
	1.0	2.0	10.0 ± 0.0	1935.705 ± 302.527

divergent step will be taken, as measured by the value ϕ_{ratio}, which is influenced by c_1 and c_2. If a w value is selected, then a truly convergent system can be constructed by choosing c_1 and c_2 so that $\phi_{\text{ratio}} = 1$. This results in a system with rapid convergence and little or no "exploration" behaviour. Choosing slightly larger c_1 and c_2 values (and keeping w fixed) results in a smaller ϕ_{ratio}. Such a

Table 2
Average number of interactions to saturation and average fitness at point of saturation for different parameter choices

Problem	w	c_1, c_2	Iterations	Fitness
Quadric	0.001	2.0	53294.2 ± 41823.3	85508.923 ± 68478.712
	0.7298	1.49618	7309.4 ± 489.6	4.499e−11 ± 2.998e−11
	0.7	1.4	11497.8 ± 17664.5	226.114 ± 367.957
	0.7	2.0	48409.6 ± 46716.9	85847.407 ± 89125.497
	0.9	2.0	10.0 ± 0.0	167363.248 ± 55771.770
	1.0	2.0	10.0 ± 0.0	160198.427 ± 52113.612
Rastrigin	0.001	2.0	69187.6 ± 38593.6	245.857 ± 159.348
	0.7298	1.49618	1280.0 ± 145.7	80.313 ± 20.493
	0.7	1.4	10194.6 ± 17891.7	83.340 ± 21.083
	0.7	2.0	78749.8 ± 33797.8	98.332 ± 152.119
	0.9	2.0	10.0 ± 0.0	456.8 ± 24.133
	1.0	2.0	10.2 ± 1.4	447.248 ± 35.139
Rosenbrock	0.001	2.0	77313.2 ± 35264.9	936.692 ± 1390.844
	0.7298	1.49618	33597.2 ± 3339.2	6.205e−5 ± 4.103e−4
	0.7	1.4	15990.0 ± 22130.2	3.817 ± 3.749
	0.7	2.0	79164.6 ± 34093.3	566.927 ± 1351.569
	0.9	2.0	10.0 ± 0.0	3299.814 ± 763.111
	1.0	2.0	10.0 ± 0.0	−3259.010 ± 942.962
Schwefel	0.001	2.0	41763.6 ± 40199.1	8947.170 ± 1752.963
	0.7298	1.49618	4025.6 ± 13508.9	4249.076 ± 577.563
	0.7	1.4	9287.2 ± 18048.7	5120.194 ± 658.799
	0.7	2.0	27177.2 ± 38277.2	8492.665 ± 2869.720
	0.9	2.0	10.0 ± 0.0	10687.778 ± 464.386
	1.0	2.0	10.0 ± 0.0	10624.172 ± 610.703
Spherical	0.001	2.0	87720.8 ± 21805.4	9490.164 ± 19173.805
	0.7298	1.49618	1231.0 ± 167.3	3.970e−11 ± 2.678
	0.7	1.4	11555.2 ± 25181.259	0.496 ± 1.007
	0.7	2.0	39793.4 ± 20331.306	2899.494 ± 14364.458
	0.9	2.0	10.0 ± 0.0	65879.727 ± 8843.272
	1.0	2.0	10.0 ± 0.0	68612.964 ± 10583.158

system will have more "exploration" behaviour, but it will have more trouble with the "exploitation" phase of the search, i.e. it will have more disruptions to its trajectory.

6. Experimental results

This section presents experimental results to illustrate the convergence behaviour of the *gbest* PSO for the different parameter combinations discussed in Section 5.6. The objective of the experiments is not to provide a PSO algorithm that produces the best possible solutions to a set of optimization problems, but to illustrate the importance of correct selection of the inertia

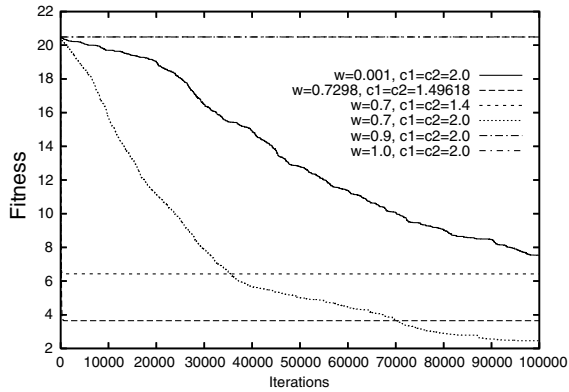

Fig. 13. Ackley.

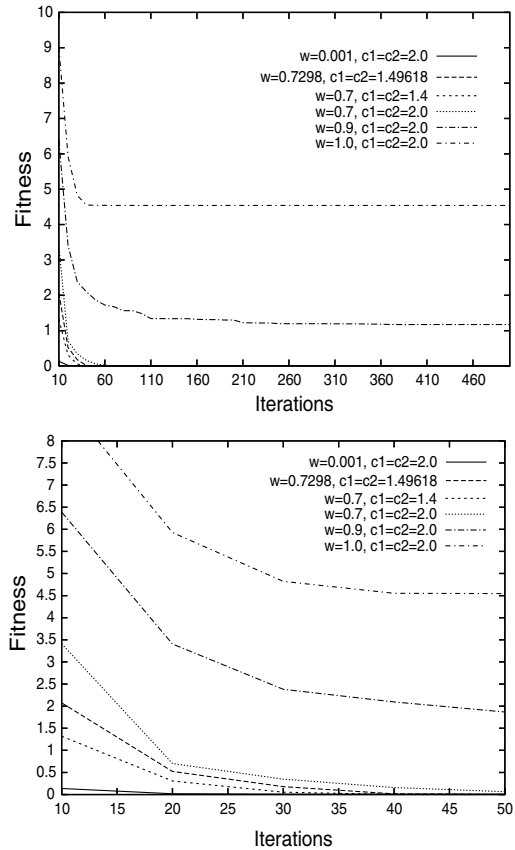

Fig. 14. Bohachevsky.

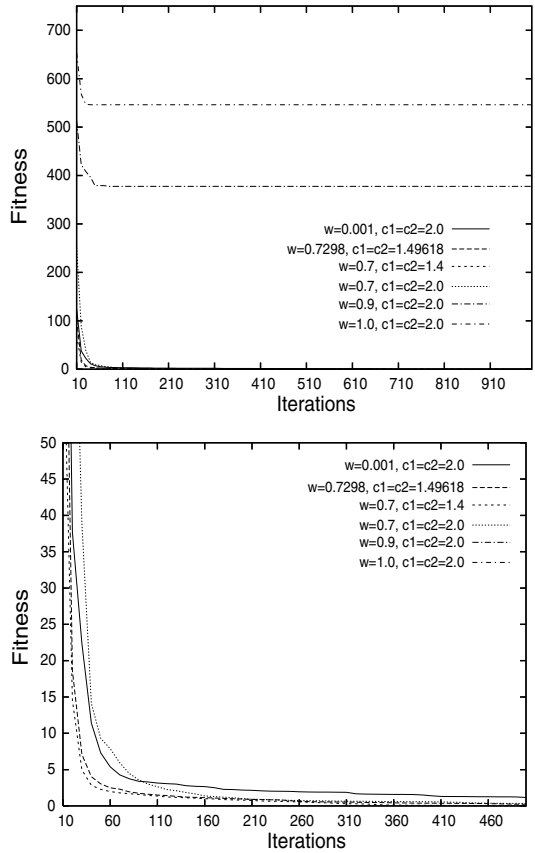

Fig. 15. Colville.

weight and acceleration coefficients. For all the experiments, 20 particles have been used, velocity clamping was not used, inertia weights were kept constant, and the PSO was executed for 100000 iterations. The CIlib (computational intelligence library)[1] PSO implementations have been used. All results reported are averages and standard deviations over 50 simulations.

The following benchmark functions have been used:

Ackley:

$$f(\mathbf{x}) = -20\mathrm{e}^{-0.2\sqrt{\frac{1}{n}\sum_{j=1}^{n} x_j^2}} - \mathrm{e}^{\frac{1}{n}\sum_{j=1}^{n}\cos(2\pi x_j)} + 20 + e$$

with $n = 30$, $x_j \in [-30, 30]$ and $f^*(\mathbf{x}) = 0.0$.

[1] http://cilib.sourceforge.net.

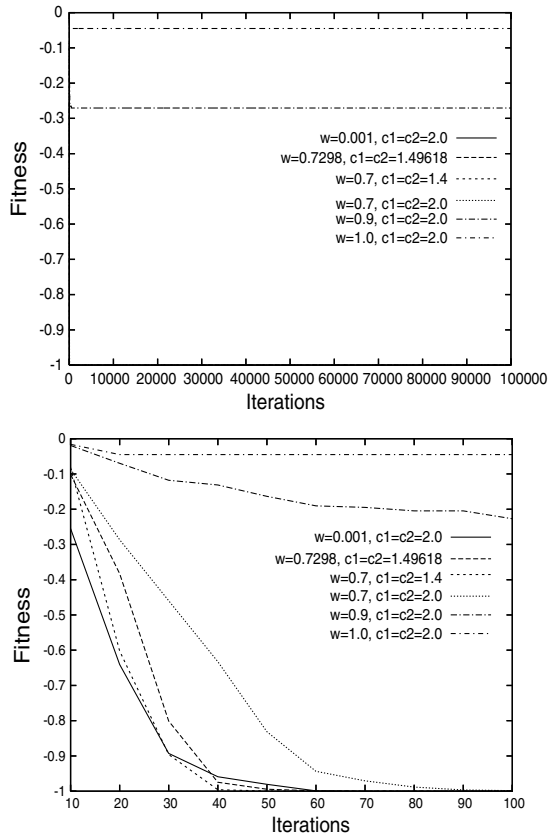

Fig. 16. Easom.

Bohachevsky 1:

$$f(x_1, x_2) = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1) - 0.4\cos(4\pi x_2) + 0.7$$

with $x_1, x_2 \in [-50, 50]$ and $f^*(x_1, x_2) = 0.0$.

Colville:

$$f(x_1, x_2, x_3, x_4) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2 + 10.1((x_2 - 1)^2 + (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1)$$

with $x_1, x_2, x_3, x_4 \in [-10, 10]$ and $f^*(x_1, x_2, x_3, x_4) = 0.0$.

Easom:

$$f(x_1, x_2) = -\cos(x_1)\cos(x_2)e^{-(x_1-\pi)^2-(x_2-\pi)^2}$$

with $x_1, x_2 \in [-100, 100]$ and $f^*(x_1, x_2) = -1.0$.

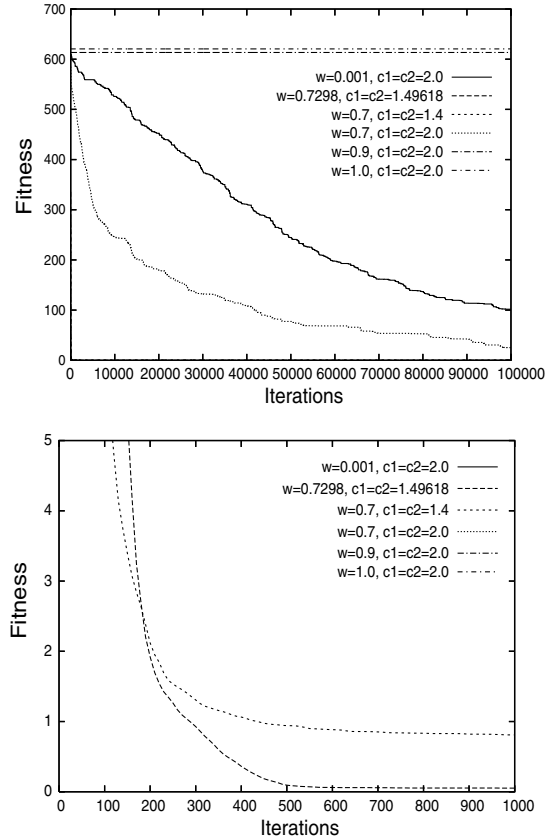

Fig. 17. Griewank.

Griewank:

$$f(\mathbf{x}) = 1 + \frac{1}{4000}\sum_{j=1}^{n} x_j^2 - \prod_{j=1}^{n} \cos\left(\frac{x_j}{\sqrt{j}}\right)$$

with $n = 30$, $x_j \in [-600, 600]$ and $f^*(\mathbf{x}) = 0.0$.

Hyperellipsoid:

$$f(\mathbf{x}) = \sum_{j=1}^{n} j^2 x_j^2$$

with $n = 30$, $x_j \in [-1, 1]$ and $f^*(\mathbf{x}) = 0.0$.

Quadric:

$$f(\mathbf{x}) = \sum_{j=1}^{n} \left(\sum_{k=1}^{j} x_j\right)^2$$

with $n = 30$, $x_j \in [-100, 100]$ and $f^*(\mathbf{x}) = 0.0$.

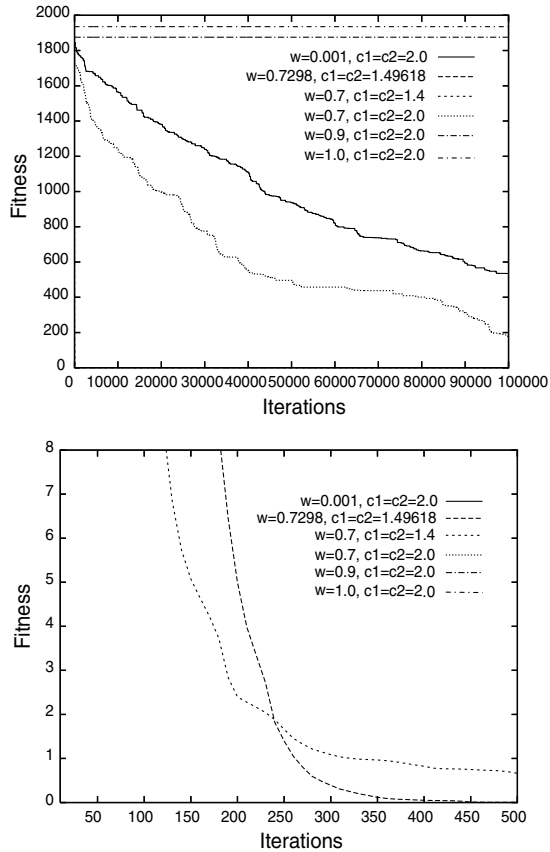

Fig. 18. Hyperellipsoid.

Rastrigin:

$$f(\mathbf{x}) = \sum_{j=1}^{n} (x_j^2 - 10\cos(2\pi x_j) + 10)$$

with $n = 30$, $x_j \in [-5.12, 5.12]$ and $f^*(\mathbf{x}) = 0.0$.

Rosenbrock:

$$f(\mathbf{x}) = \sum_{j=1}^{n/2} \left[100(x_{2j} - x_{2j-1}^2)^2 + (1 - x_{2j-1})^2 \right]$$

with $n = 30$, $x_j \in [-2.048, 2.048]$ and $f^*(\mathbf{x}) = 0.0$.

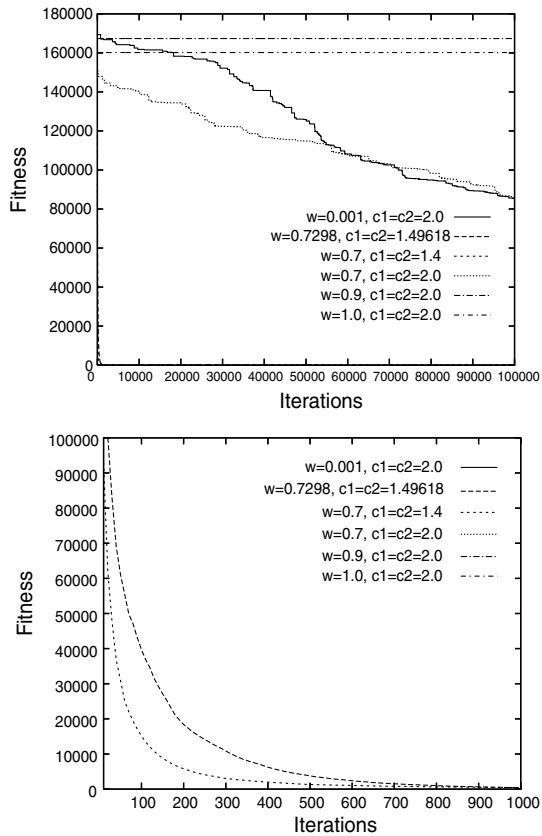

Fig. 19. Quadric.

Schwefel:

$$f(\mathbf{x}) = \sum_{j=1}^{n} x_j \sin\left(\sqrt{|x_j|}\right) + 418.9829n$$

with $n = 30$, $x_j \in [-500, 500]$ and $f^*(\mathbf{x}) = 0.0$.

Spherical:

$$f(\mathbf{x}) = \sum_{j=1}^{n} x_j^2$$

with $n = 30$, $x_j \in [-100, 100]$ and $f^*(\mathbf{x}) = 0.0$.

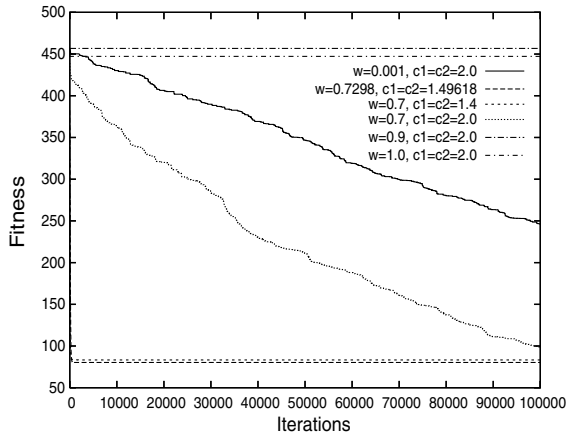

Fig. 20. Rastrigin.

For each of the above functions, the following parameter value combinations have been used (from Section 5.6):

(1) $w = 0.001$, $c_1 = c_2 = 2.0$.
(2) $w = 0.7298$, $c_1 = c_2 = 1.49618$ (from [10]).
(3) $w = 0.7$, $c_1 = c_2 = 1.4$.
(4) $w = 0.7$, $c_2 = c_2 = 2.0$.
(5) $w = 0.9$, $c_1 = c_2 = 2.0$.

Tables 1 and 2 summarize the results for the functions and each parameter combination. The average number of iterations at which fitness values improved less than 10^{-10} are given as well as the average fitness at that point. Figs. 13–23 illustrate the average fitness over the 50 simulations of the best particle as a function of iteration number. Where necessary, a blow-up is given of parts of the figures.

For all the functions, configurations 5 and 6 (with $w = 0.9$ and 1.0 respectively) failed to converge as expected. For all functions, PSO quickly stagnated at very bad fitness levels. In general, configurations 1 and 4 initially improved the fitness values, but at a very slow rate. This is expected, since both configurations violate the heuristic given in Eq. (28), but due to the stochastic component (as discussed in Section 5.6), the PSO succeeded in reducing fitness values. In these cases the fraction of convergent steps allowed the particles to move to better positions. For all the functions, except for Ackley, Bohachevsky, Colville and Easom, configuration 2 provided the best results, confirming the empirical observations of Eberhart and Shi [10]. For Easom, configuration 1–4 provided the global optimum. It is also expected that configuration 3 will

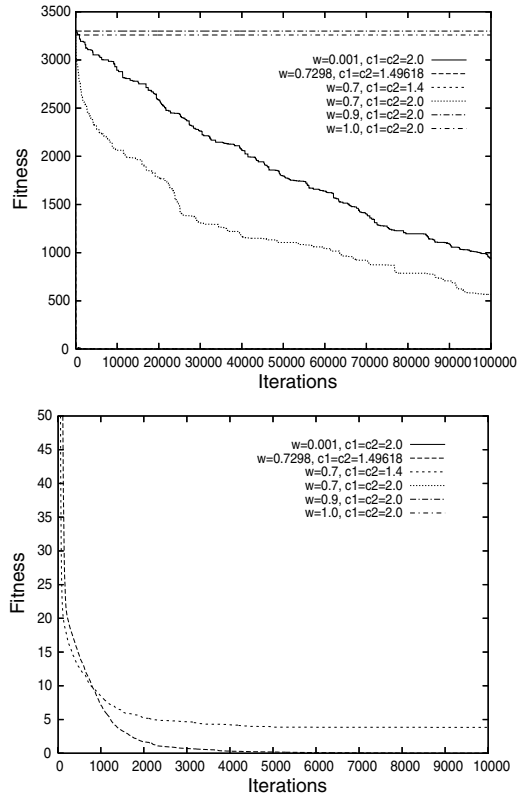

Fig. 21. Rosenbrock.

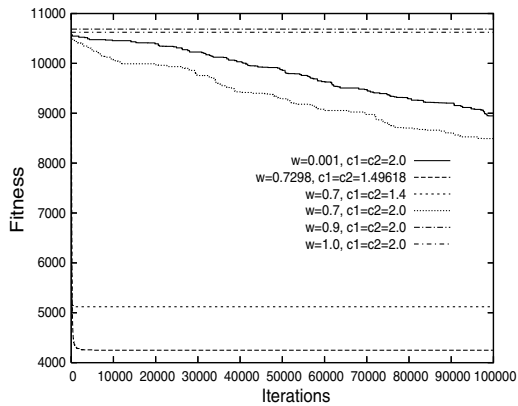

Fig. 22. Schwefel.

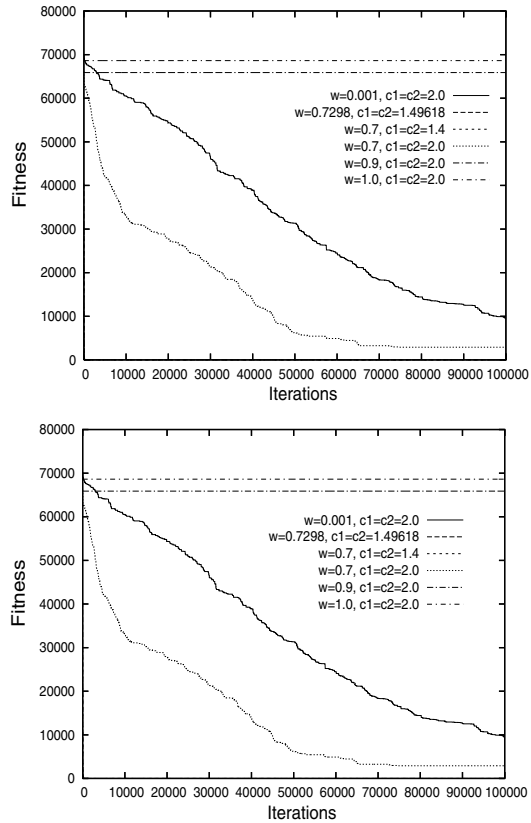

Fig. 23. Spherical.

provide convergent trajectories since the heuristic of Eq. (28) is satisfied. This happens for all the functions, although to a bad fitness for Quadric and Schwefel compared to configuration 2.

For most of the functions (9 out of 11), configurations 2 and 3 resulted in faster convergence.

In summary, the results presented confirms the expectations as derived from theoretical analyses. The basic PSO algorithm is sensitive to the values of the inertia weight and acceleration coefficients. Parameter values that satisfy the heuristic in Eq. (28) ensure convergent trajectories. As mentioned earlier, the objective of the paper was not to optimize parameters and the PSO in order to find the best possible minima. The reader is referred to the list of papers at http://www.swarmintelligence.org which provide empirical studies to optimize parameters and PSO improvements.

7. Conclusions

The main objective of this paper was to derive a heuristic for the initialization of the inertia weight and acceleration coefficient values of the PSO to guarantee convergent trajectories. This heuristic has been studied and example particle trajectories given. Experimental results confirmed that the PSO is sensitive to the inertia weight and acceleration coefficient values, and that the derived heuristic ensures convergent trajectories.

As a subobjective, the paper provided a formal analysis to prove that particles converge to a stable point, taking the inertia term in consideration. The stable point is shown to be a weighted average of the personal best and global best positions, where the weights are determined by the values of the acceleration coefficients. This is not a proof of convergence to a minimum, but states only that the swarm will reach a point of equilibrium, under certain conditions. These conditions have been formally derived.

The paper provides a better understanding of the dynamics of the PSO, and provides a valuable guideline for selecting control parameters.

References

[1] T. Beielstein, K.E. Parsopoulos, M.N. Vrahatis, Tuning pso parameters through sensitivity analysis, in: Technical Report, Reihe Computational Intelligence CI 124/02, Department of Computer Science, University of Dortmund, 2002.

[2] R.L. Burden, J.D. Faires, Numerical Analysis (Chapter 5.11), fifth ed., PWS Publishing Company, Boston, 1993, pp. 314–321.

[3] A. Carlisle, G. Dozier, An off-the-self pso, in: Proceedings of the Workshop on Particle Swarm Optimization, Indianapolis, USA, 2001.

[4] M. Clerc, The swarm and the queen: Towards a deterministic and adaptive particl e swarm optimization, in: Proceedings of the IEEE Congress on Evolutionary Computation, 1999, pp. 1951–1957.

[5] M. Clerc, Think locally, act locally: The way of life of cheap-pso, an adaptive pso, Available from <http://clerc.maurice.free.fr/pso/>, 2001.

[6] M. Clerc, J. Kennedy, The particle swarm-explosion, stability, and convergence in a multidimensional complex space, IEEE Transactions on Evolutionary Computation 6 (1) (2002) 58–73.

[7] R.C. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, in: Proceedings of the Sixth International Symposium on Micromachine and Human Science, Nagoya, Japan, 1995, pp. 39–43.

[8] R.C. Eberhart, Y. Shi, Particle swarm optimization: Developments, applications and resources, in: Proceedings of the IEEE Congress on Evolutionary Computation, IEEE Press, Seoul, Korea, 2001.

[9] R.C. Eberhart, P.K. Simpson, R.W. Dobbins, Computational Intelligence PC Tools, first ed., Academic Press Professional, 1996.

[10] R.C. Eberhart, Y. Shi, Comparing inertia weights and constriction factors in particle swarm optimization, in: Proceedings of the IEEE Congress on Evolutionary Computation, San Diego, USA, 2000, pp. 84–88.

[11] J. Holland, Adaption in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI, 1975.
[12] J. Kennedy, The particle swarm: Social adaptation of knowledge, in: Proceedings of the IEEE International Conference on Evolutionary Computation, Indianapolis, USA, 1997, pp. 303–308.
[13] J. Kennedy, Small worlds and mega-minds: Effects of neighborhood topology on particle swarm performance, in: Proceedings of the IEEE Congress on Evolutionary Computation, 1999, pp. 1931–1938.
[14] J. Kennedy, R.C. Eberhart, Particle swarm optimization, in: Proceedings of the IEEE International Joint Conference on Neural Networks, IEEE Press, 1995, pp. 1942–1948.
[15] J. Kennedy, R. Mendes, Population structure and particle performance, Proceedings of the IEEE Congress on Evolutionary Computation, IEEE Press, Honolulu, Hawaii, 2002.
[16] R. Mendes, P. Cortez, M. Rocha, J. Neves, Particle swarms for feedforward neural network training, in: Proceedings of the International Joint Conference on Neural Networks, 2002, pp. 1895–1899.
[17] S. Naka, T. Genji, T. Yura, Y. Fukuyama, Practical distribution state estimation using hybrid particle swarm optimization, in: IEEE Power Engineering Society Winter Meeting, Columbus, USA, 2001, pp. 815–820.
[18] S. Naka, T. Genji, T. Yura, Y. Fukuyama, A hybrid particle swarm optimization for distribution state estimation, IEEE Transactions on Power Systems 18 (1) (2003) 60–68.
[19] E. Ozcan, C.K. Mohan, Analysis of a simple particle swarm optimization system, in: Intelligent Engineering Systems through Artificial Neural Networks, 1998, pp. 253–258.
[20] E. Ozcan, C.K. Mohan, Particle swarm optimization: Surfing the waves, in: Proceedings of the IEEE Congress on Evolutionary Computation, Washington, DC, USA, 1999.
[21] Y. Shi, R.C. Eberhart, A modified particle swarm optimizer, in: Proceedings of the IEEE Congress on Evolutionary Computation, Piscataway, USA, 1998, pp. 69–73.
[22] Y. Shi, R.C. Eberhart, Parameter selection in particle swarm optimization, in: Proceedings of the Seventh Annual Conference on Evolutionary Programming, New York, USA, 1998, pp. 591–600.
[23] Y. Shi, R.C. Eberhart, Empirical study of particle swarm optimization, in: Proceedings of the IEEE Congress on Evolutionary Computation, IEEE Press, 1999, pp. 1945–1950.
[24] Y. Shi, R.C. Eberhart, Fuzzy adaptive particle swarm optimization, Proceedings of the IEEE Congress on Evolutionary Computation, IEEE Press, Seoul, Korea, 2001.
[25] P.N. Suganthan, Particle swarm optimiser with neighborhood operator, in: Proceedings of the IEEE Congress on Evolutionary Computation, IEEE Press, Piscataway, USA, 1999, pp. 1958–1962.
[26] I.C. Trelea, The particle swarm optimization algorithm: Convergence analysis and parameter selection, Information Processing Letters 85 (6) (2003) 317–325.
[27] F. van den Bergh, An Analysis of Particle Swarm Optimizers, PhD thesis, Department of Computer Science, University of Pretoria, Pretoria, South Africa, 2002.
[28] G. Venter, J. Sobieszczanski-Sobieski, Multidisciplinary optimization of a transport aircraft wing using particle swarm optimization, in: Ninth AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Atlanta, USA, 2002.
[29] K. Yasuda, A. Ide, N. Iwasaki, Adaptive particle swarm optimization, in: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 2003, pp. 1554–1559.
[30] H. Yoshida, Y. Fukuyama, S. Takayama, Y. Nakanishi, A particle swarm optimization for reactive power and voltage control in electric power systems considering voltage security assessment, in: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, 1999, p. 502.
[31] Y. Zheng, L. Ma, L. Zhang, J. Qian, On the convergence analysis and parameter selection in particle swarm optimization, in: Proceedings of the International Conference on Machine Learning and Cybernetics, 2003, pp. 1802–1807.