Chapter 4

Objectivity of the PSOA

4.1 Introduction

In science, most physical phenomena are invariant. It is fundamental that mathematical representation of these phenomena reflects this invariance. This fundamental requirement is known as objectivity, frame-indifference or observer independence, and is well known in classical mechanics [53]. For objectivity, the description of some quantity has to be invariant under pure translations, as well as pure rotations.

Objectivity or observer independence is also highly desirable (almost essential) in optimization procedures, to reflect the invariance of the physical processes that are optimized. Robust optimization procedures and algorithms should definitely be invariant.

In classical gradient based optimization, the gradient vector (or some conjugate direction to the gradient), indicates some direction of improvement, even if this direction is not optimal. This accounts for the reference frame; classical optimization is (usually) frame invariant.

In modern (stochastic) optimization procedures, the requirement of observer independence is equally essential. These algorithms include genetic programming [37], genetic algorithms [38], evolutionary strategies [39], differential evolution [40] and the particle swarm optimization algorithm (PSOA) [3, 4].

For the genetic algorithm (GA), Salomon [54, 55] demonstrated the lack of rotational invariance of the algorithm. He showed that the GA's performance at low mutation rates is significantly influenced by the frame of reference used to pose a problem.

The PSOA was introduced by Kennedy and Eberhart [3] as a gradient free stochastic optimization algorithm. The fundamental principle behind the PSOA is the evolutionary advantages that the sharing of information offers. This is often known as 'collaborative searching'.

The PSOA is quite simple: At first, a swarm of p particles is randomly deployed in an n-dimensional design domain. The particles then update their positions in the design domain over unit time increments using a simple stochastic rule, known as the 'velocity rule'.

The quality of each particle's position at each iteration is then evaluated using the objective or cost function. Each particle's cognitive memory allows it to remember it's own best cost function

value, with associated position, over time. Importantly, the social interaction and awareness of the particles allow them to also remember the 'best' cost function value the swarm itself found over time.

In Chapter 3 it was shown that implementation subtleties due to ambiguous notation have resulted in two distinctly different implementations of the PSOA. While this does not repute negatively on the ingenuity of the idea of Kennedy and Eberhart, discerning between these two implementations is of crucial importance. The behavior of the respective implementations is markedly different, although they only differ in the formulation of the velocity updating rule. In fact, the differences are merely due to subtle differences in the introduction of randomness into the algorithm.

In this chapter, the objectivity of the PSOA is investigated. It is shown that the first formulation PSOAF1 is objective, combined with the *disadvantage* that the particle trajectories collapse to line searches. It is then show that the second formulation PSOAF2 is not objective, although it has the *advantage* that the particle trajectories are n-dimensional space filling. A novel formulation that is *both* objective and diverse, i.e. the algorithm generates particle trajectories that are space filling, is then presented.

4.2 Notes on the investigation

The investigation into the objectivity of the PSOA is started by defining the instantaneous search domain of a particle, viz. the domain to which the search of a particle i at iteration k is restricted as discussed in Chapter 3.

From Eqs. (2.2) and (2.3), it is observed that the instantaneous search domain is composed from a deterministic contribution given by $(\boldsymbol{x}_k^i + w\boldsymbol{v}_k^i)$, and a stochastic contribution due to $\boldsymbol{\nu}_k^i$.

The stochastic domain is bounded, and has an associated probability distribution, due to the random scaling with finite scalars. In order to investigate the objectivity of the stochastic contribution $\boldsymbol{\nu}_k^i$ of the instantaneous search domain, Monte Carlo simulations [56] are used. These are conducted for different values of $\boldsymbol{p}_k^i$, $\boldsymbol{p}_k^g$ and $\boldsymbol{x}_k^i$. Scatter plots are constructed to define the domain of possible stochastic vectors $\boldsymbol{\nu}_k^i$ by generating 10^4 instances of $\boldsymbol{\nu}_k^i$. In all investigations $c_1 = c_2 = 2$.

4.3 Formulation 1 (PSOAF1)

For PSOAF1, the stochastic vector $\boldsymbol{\nu}_k^i$ is given by

$$\boldsymbol{\nu}_k^i = c_1 r_{1k}^i(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i) + c_2 r_{2k}^i(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i), \tag{4.1}$$

where r_{1k}^i and r_{2k}^i represent two uniform real random scalars between 0 and 1, which are updated at every iteration k, and for each particle i in the swarm. The random numbers r_{1k}^i and r_{2k}^i independently scale *only* the magnitudes of the cognitive and social vectors, respectively given by $c_1(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i)$ and $c_2(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i)$. The cognitive vector $c_1(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i)$ and the social vector $c_2(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i)$ can be anything from normal to parallel w.r.t. each other.

When the cognitive and social vectors are not parallel, Eq. (4.1) may be interpreted as the vector equation of a bounded plane $\boldsymbol{\mathcal{P}}_k^i$ in n-dimensional space. The bounded plane is then translated in

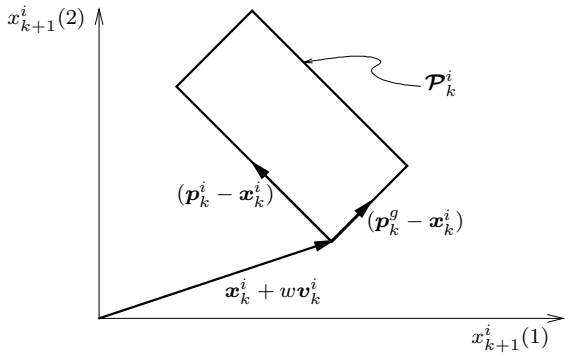


Figure 4.1: Partitioning the position vector $\boldsymbol{x}^i_{k+1}$ into a deterministic contribution $(\boldsymbol{x}^i_k + w\boldsymbol{v}^i_k)$, and a stochastic contribution $(\boldsymbol{\nu}^i_k \in \boldsymbol{\mathcal{P}}^i_k)$, for $c_1 = c_2 = 2$.

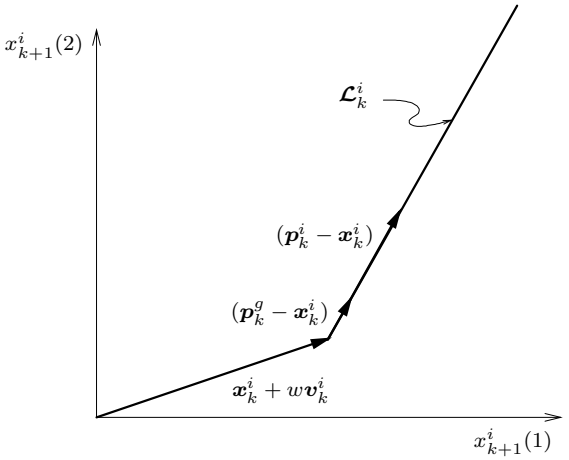


Figure 4.2: Partitioning the position vector $\boldsymbol{x}^i_{k+1}$ into a deterministic contribution $(\boldsymbol{x}^i_k + w\boldsymbol{v}^i_k)$, and a stochastic contribution $(\boldsymbol{\nu}^i_k \in \boldsymbol{\mathcal{L}}^i_k)$, for $c_1 = c_2 = 2$.

n-dimensional space by the addition of $\boldsymbol{x}^i_k$ and $w\boldsymbol{v}^i_k$, as depicted in Figure 4.1.

Whenever the cognitive and social vectors $c_1(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)$ and $c_2(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)$ are parallel, Eq. (4.1) may be interpreted as the vector equation of a bounded line $\boldsymbol{\mathcal{L}}^i_k$ in n-dimensional space. Again, the bounded line is translated in n-dimensional space by the addition of $\boldsymbol{x}^i_k$ and $w\boldsymbol{v}^i_k$, as depicted in Figure 4.2.

The intrinsic properties of a vector are its magnitude and direction; these exist independent of a reference frame [57]. In PSOAF1, only the vector magnitudes (which are invariant) are randomly scaled. Also, since the vectors $c_1(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)$ and $c_2(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)$ are constructed through the subtraction of two vectors, they are also *translationally invariant*. Both criteria for observer independence are met, hence PSOAF1 is objective.

4.3.1 PSOAF1: Investigation of the instantaneous search domain

Objectivity of PSOAF1 is now illustrated by conducting Monte Carlo simulations. Similar simulations will be conducted for the algorithmic formulations in sections to come.

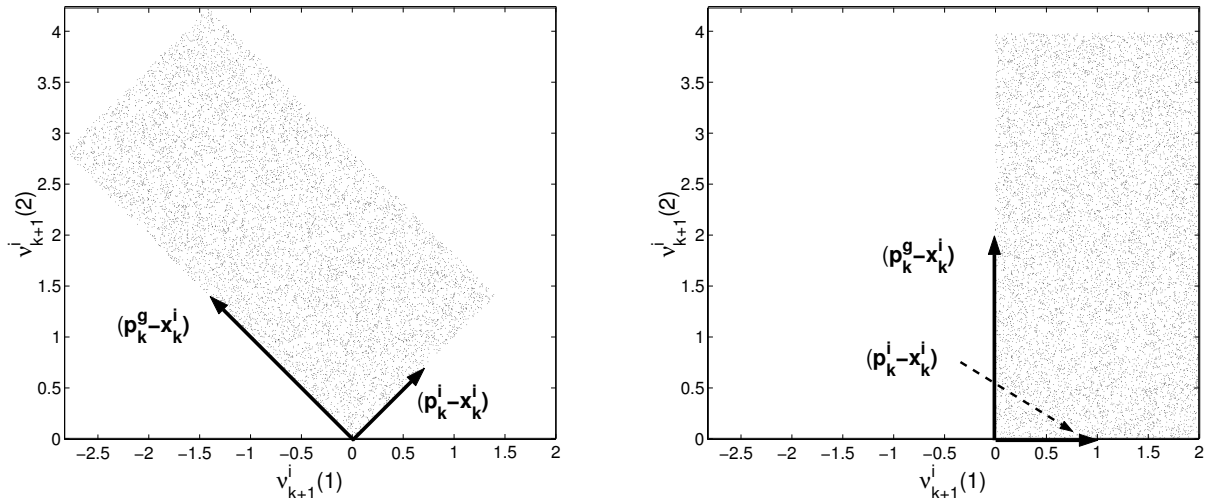


Figure 4.3: PSOAF1: Scatter plot of 10^4 possible stochastic vectors $\boldsymbol{\nu}^i_k$, generated using Monte Carlo simulations, with a) $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right) = [\frac{1}{\sqrt{2}}\ \frac{1}{\sqrt{2}}]$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right) = [-\sqrt{2}\ \sqrt{2}]$ and b) $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right) = [1\ \ 0]$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right) = [0\ \ 2]$. Each point represents the end point of a stochastic vector with origin at [0 0].

First, a study is conducted for non-parallel cognitive and social vectors $c_1(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)$ and $c_2(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)$. In Figure 4.3(a), the vectors $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right)$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right)$ are respectively given by $[\frac{1}{\sqrt{2}}\ \ \frac{1}{\sqrt{2}}]$ and $[-\sqrt{2}\ \sqrt{2}]$. A scatter plot yields the plane $\boldsymbol{\mathcal{P}}^i_k$, with c_1 and c_2 merely scaling $\boldsymbol{\mathcal{P}}^i_k$.

A scatter plot is then constructed after rotating the vectors $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right)$ and $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right)$ $45°$ clockwise, as depicted in Figure 4.3(b). Hence $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right)$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right)$ are respectively given by $[1\ 0]$ and $[0\ 2]$. From Figure 4.3(b), it is clear that the domain remains a bounded plane $\boldsymbol{\mathcal{P}}^i_k$, which is merely rotated $45°$ clockwise.

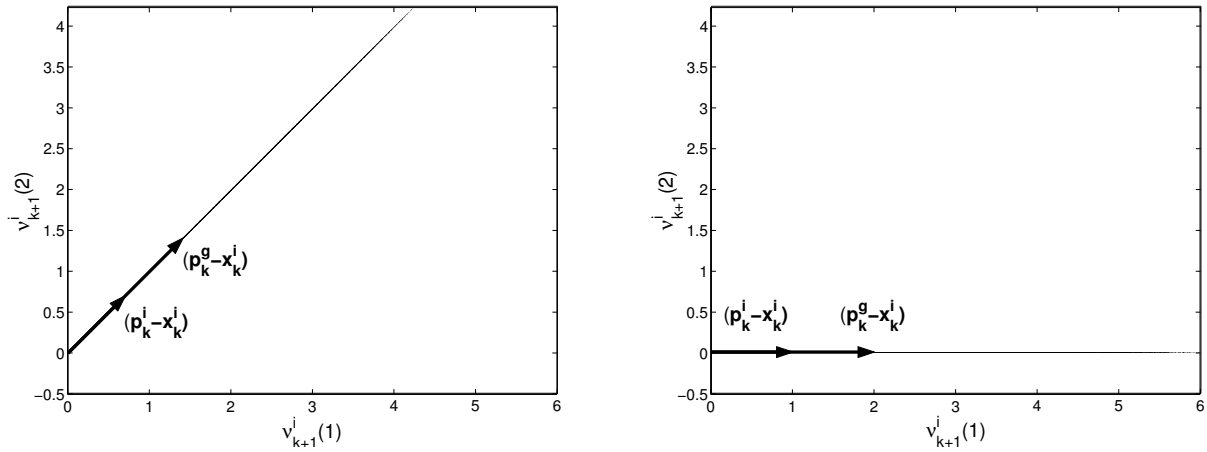
It also follows from random variable theory that the probability distribution over the domain $\boldsymbol{\mathcal{P}}^i_k$ is uniform [50], as illustrated in Figures 4.3(a) and 4.3(b).

Secondly, a similar study is conducted for parallel cognitive and social vectors $c_1(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)$ and $c_2(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)$, as depicted in Figure 4.4. In Figure 4.4(a), the parallel vectors $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right)$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right)$ are respectively given by $[\frac{1}{\sqrt{2}}\ \ \frac{1}{\sqrt{2}}]$ and $[\sqrt{2}\ \sqrt{2}]$. The domain is a bounded line $\boldsymbol{\mathcal{L}}^i_k$ with c_1 and c_2 merely scaling the length of $\boldsymbol{\mathcal{L}}^i_k$.

Again a scatter plot is constructed after rotating $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right)$ and $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right)$ $45°$ clockwise, as depicted in Figure 4.4(b). Now, $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right)$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right)$ are respectively given by $[1\ 0]$ and $[2\ 0]$. As shown in Figure 4.4(b), the bounded line $\boldsymbol{\mathcal{L}}^i_k$ is merely rotated.

It follows from random variable theory that the probability distribution over the bounded line is tri-linear [50], for different vector lengths $||\boldsymbol{p}^i_k - \boldsymbol{x}^i_k||$ and $||\boldsymbol{p}^g_k - \boldsymbol{x}^i_k||$.

As discussed earlier and graphically demonstrated here, PSOAF1 is objective. A rotation of the vectors $\boldsymbol{p}^i_k$, $\boldsymbol{p}^g_k$ and $\boldsymbol{x}^i_k$ merely results in a rotation of the stochastic domains, $\boldsymbol{\mathcal{P}}^i_k$ and $\boldsymbol{\mathcal{L}}^i_k$. This



(a) (b)

Figure 4.4: PSOAF1: Scatter plot of 10^4 possible stochastic vectors $\boldsymbol{\nu}_k^i$, generated using Monte Carlo simulations, with a) $\left(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i\right) = [\frac{1}{\sqrt{2}}\ \frac{1}{\sqrt{2}}]$ and $\left(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i\right) = [\sqrt{2}\ \sqrt{2}]$ and b) $\left(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i\right) = [1\ \ 0]$ and $\left(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i\right) = [2\ \ 0]$.

follows since only the magnitude of the cognitive and social vectors are scaled in PSOAF1.

4.4 Formulation 2 (PSOAF2)

The stochastic vector $\boldsymbol{\nu}_k^i$ of PSOAF2 is given by

$$\boldsymbol{\nu}_k^i = c_1 \boldsymbol{r}_{1k}^i \circ (\boldsymbol{p}_k^i - \boldsymbol{x}_k^i) + c_2 \boldsymbol{r}_{2k}^i \circ (\boldsymbol{p}_k^g - \boldsymbol{x}_k^i), \tag{4.2}$$

where the $\circ$ operator indicates component by component multiplication between two vectors. Hence the random vectors $\boldsymbol{r}_{mk}^i$ are given by

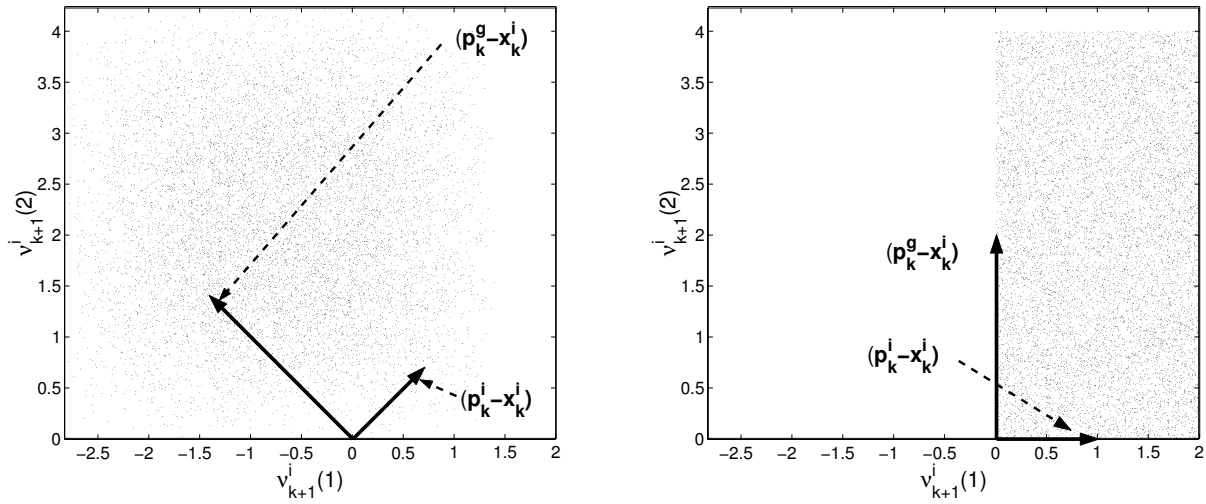
$$\boldsymbol{r}_{mk}^i = (\rho_{1k}^i,\ \rho_{2k}^i, \cdots,\ \rho_{nk}^i), \quad m = 1, 2, \tag{4.3}$$

with $\rho_{lk}^i, l = 1, 2, \cdots, n$ uniform random numbers between 0 and 1. Eq. (4.2) is no longer a vector equation of a bounded plane $\boldsymbol{\mathcal{P}}_k^i$, since every non-zero component of $(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i)$ and $(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i)$ is independently scaled. As a result, the domain of possible stochastic vectors is generalized to n-dimensional space $\boldsymbol{\mathcal{S}}_k^i$.

However, since the components of a vector are given with respect to a specific reference frame, PSOAF2 is *rotationally variant*. (Although PSOAF2 is of course translationally invariant.) Nevertheless, PSOAF2 is observer dependent, since only one of the two criteria of objectivity is met.

4.4.1 PSOAF2: Investigation of the instantaneous search domain

The observer dependence of PSOAF2 is now quantified, using Monte Carlo simulations, similar to those in Section 4.3.1.



(a) (b)

Figure 4.5: PSOAF2: Scatter plot of 10^4 possible stochastic vectors $\boldsymbol{\nu}_k^i$, generated using Monte Carlo simulations with a) $\left(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i\right) = [\frac{1}{\sqrt{2}}\ \frac{1}{\sqrt{2}}]$ and $\left(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i\right) = [-\sqrt{2}\ \sqrt{2}]$ and b) $\left(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i\right) = [1\ \ 0]$ and $\left(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i\right) = [0\ \ 2]$.

As before, the study is conducted for non-parallel cognitive and social vectors $c_1(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i)$ and $c_2(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i)$. Figure 4.5(a) depicts that the domain is an n-dimensional space $\boldsymbol{\mathcal{S}}_k^i$ (with $n = 2$ in this case), with c_1 and c_2 merely scaling $\boldsymbol{\mathcal{S}}_k^i$. It is also clear that the probability distribution over $\boldsymbol{\mathcal{S}}_k^i$ is non-uniform.

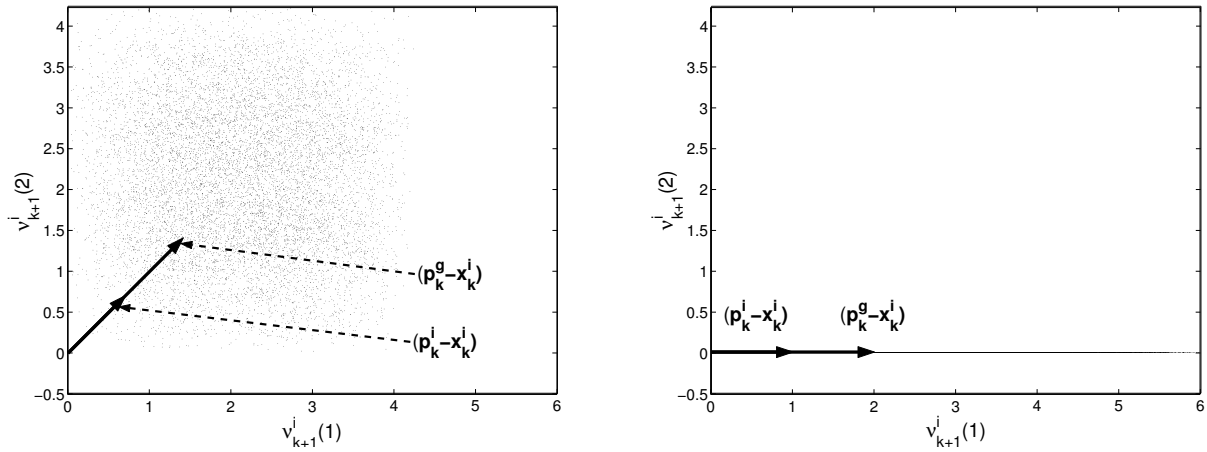
The scatter plot after rotating the vectors $\left(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i\right)$ and $\left(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i\right)$ $45°$ clockwise, is depicted in Figure 4.5(b). It is clear that the *domain changes* after rotation of the vectors. However, the domain remains an n-dimensional space $\boldsymbol{\mathcal{S}}_k^i$, but the size of, and the probability distribution over, the domain depends on the orientation w.r.t. the Cartesian coordinate axis.

The study is repeated for parallel cognitive and social vectors $c_1(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i)$ and $c_2(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i)$, as depicted in Figure 4.6(a). The domain is still generalized to n-dimensional space $\boldsymbol{\mathcal{S}}_k^i$ with c_1 and c_2 merely scaling the size of $\boldsymbol{\mathcal{S}}_k^i$.

The scatter plot after rotating $\left(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i\right)$ and $\left(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i\right)$ $45°$ clockwise, is depicted in Figure 4.6(b). It is clear that the domain changes significantly after rotation of the vectors. In fact, the domain collapses to a bounded line $\boldsymbol{\mathcal{L}}_k^i$, since both vectors are parallel to one of the Cartesian basis vectors.

As discussed earlier and graphically demonstrated here, PSOAF2 is observer *dependent*. A rotation of the vectors $(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i)$ and $(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i)$ results in the size of, and the probability distribution over, the stochastic domain to change. This follows since PSOAF2 scales the *components* of the cognitive and social vectors. Since the components of a vector are observer dependent, PSOAF2 is also observer dependent.

However, the advantage of PSOAF2 is that the particle trajectories remain space filling in n-dimensional space as shown in Chapter 3. The result is that diversity in particle trajectories are maintained.



(a) (b)

Figure 4.6: PSOAF2: Scatter plot of 10^4 possible stochastic vectors $\boldsymbol{\nu}^i_k$, generated using Monte Carlo simulations with a) $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right) = [\frac{1}{\sqrt{2}}\ \frac{1}{\sqrt{2}}]$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right) = [\sqrt{2}\ \sqrt{2}]$ and b) $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right) = [1\ \ 0]$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right) = [2\ \ 0]$.

4.5 Novel Formulation: PSOAF1*

As discussed in Section 4.3, PSOAF1 is objective, although the particle trajectories collapse to lines. (The advantage of diverse (n-dimensional) particle search trajectories are quantified in Chapter 3.) On the other hand, PSOAF2 allows for particles to have diverse search trajectories, but unfortunately this comes at the cost of sacrificing objectivity.

An implementation of the PSOA is now presented that allows for diverse particle search trajectories, while retaining objectivity. Based on PSOAF1, the novel, diverse implementation, is denoted PSOAF1*.

In PSOAF1*, the vector magnitudes are scaled, and the vector directions of $(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)$ and $(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)$ perturbed, by multiplying each of the above vectors with an independent *random rotation matrix*. The random rotation matrices are constructed anew for each particle i and for every iteration k, hence

$$\boldsymbol{\nu}^i_k = c_1 r^i_{1k} \boldsymbol{Q}^i_{1k}(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k) + c_2 r^i_{2k} \boldsymbol{Q}^i_{2k}(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k), \tag{4.4}$$

with each $\boldsymbol{Q}^i_{lk},\ l = 1, 2$, a random rotation matrix of dimension $n \times n$.

$\boldsymbol{Q}$ is a proper orthogonal matrix (with determinant 1). Numerous methods are available to construct rotation matrices (e.g. see the approach of Salomon [54]. Constructing $n \times n$ matrices using Salomon's routine is however computationally expensive, since $(n-1)(n-2)$ matrix-matrix multiplications are required.)

As a computationally viable alternative, the exponential map is used [58]. There are again numerous ways to construct exponential maps. The simple series method is selected [58]. The general

series expansion of an exponential map is given by

$$\boldsymbol{Q} = \boldsymbol{I} + \boldsymbol{W} + \frac{1}{2}\boldsymbol{W}\boldsymbol{W} + \frac{1}{6}\boldsymbol{W}\boldsymbol{W}\boldsymbol{W} + \cdots, \tag{4.5}$$

where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{W}$ is a skew matrix.

The random skew matrix $\boldsymbol{W}$ is constructed as follows:

$$\boldsymbol{W} = \frac{\alpha\pi}{180}(\boldsymbol{A} - \boldsymbol{A}^{\mathrm{T}}), \tag{4.6}$$

with $\boldsymbol{A}$ an $n \times n$ random matrix with each entry a uniform random number between -0.5 and 0.5. α is a real scaling factor and superscript T denotes the matrix transpose.

The author selects to construct the exponential map $\boldsymbol{Q}_k^i$ for "small" perturbations, using only the first two terms of a truncated series method, i.e.

$$\boldsymbol{Q}_k^i = \boldsymbol{I} + \boldsymbol{W}_k^i. \tag{4.7}$$

This is the linear approximation to a rotation matrix, and is valid for small perturbations, since the entries of the higher order terms are close to zero. The advantage of the simplification is that the number of matrix-matrix multiplications is zero.

(It is important to note that the variable bounds defining D should be normalized, such that the boundary ranges are equal.)

4.5.1 PSOAF1*: Investigation of the instantaneous search domain

As before, the objectivity of PSOAF1* is quantified using Monte Carlo simulations. In 2 dimensions, $\alpha = 15$ is selected. (Although this is not "small", this serves to clearly illustrate the proposed concept).

Again, the study for non-parallel cognitive and social vectors is conducted, as depicted in Figure 4.7(a). The domain generalizes to n-dimensional space $\boldsymbol{\mathcal{S}}_k^i$, with c_1 and c_2 scaling $\boldsymbol{\mathcal{S}}_k^i$.

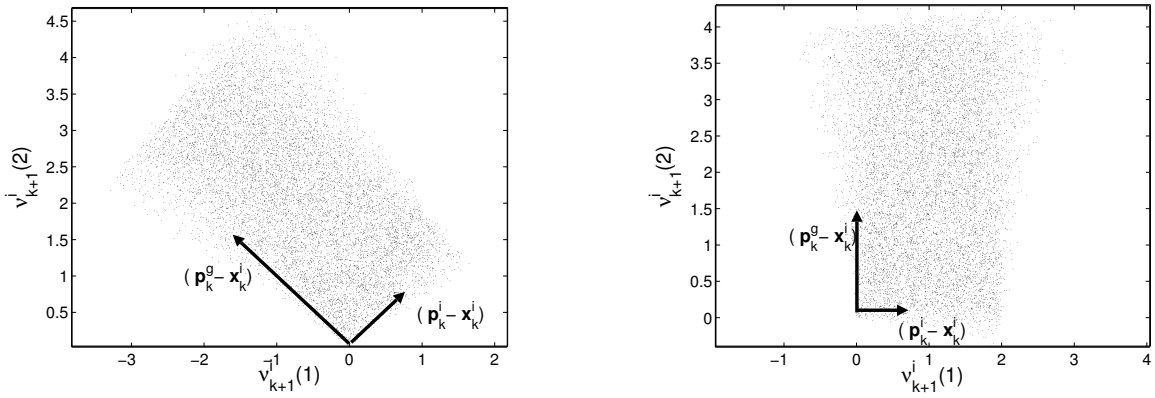
The scatter plot after rotating the vectors $\left(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i\right)$ and $\left(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i\right)$ 45° clockwise is depicted in Figure 4.7(b). Clearly, the domain remains generalized to n-dimensional space $\boldsymbol{\mathcal{S}}_k^i$, rotated 45° clockwise. The probability distribution over the domain $\boldsymbol{\mathcal{S}}_k^i$ is non-uniform.

Secondly, the study for parallel cognitive and social vectors $c_1(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i)$ and $c_2(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i)$ is conducted. Again the domain generalizes to n-dimensional space $\boldsymbol{\mathcal{S}}_k^i$, with c_1 and c_2 merely scaling the domain.

A scatter plot after rotating $\left(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i\right)$ and $\left(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i\right)$ 45° clockwise is constructed, as depicted in Figure 4.8(b). Evidently, the n-dimensional space $\boldsymbol{\mathcal{S}}_k^i$ is merely rotated, and the probability distribution over the domain is non-uniform.

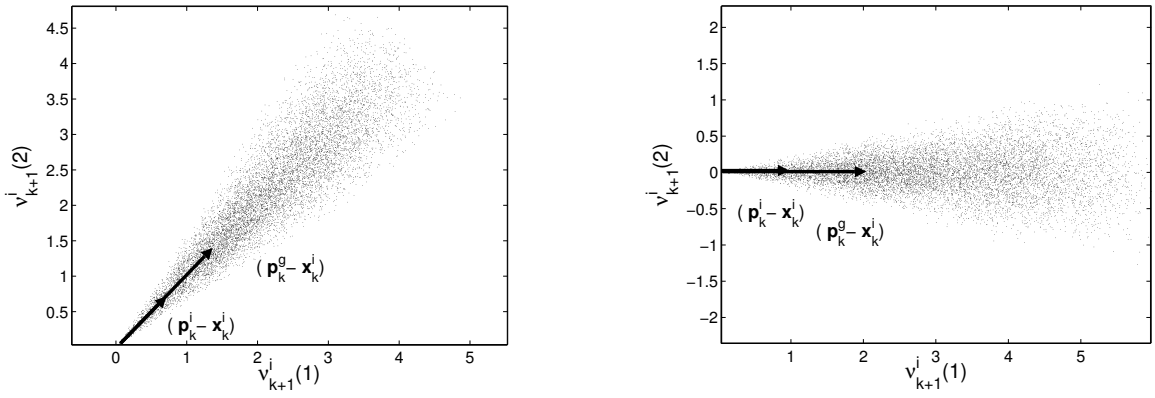
As discussed earlier and graphically demonstrated here, PSOAF1* is an objective formulation. A rotation of the vectors $(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i)$ and $(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i)$ merely results in a rotation of the stochastic domain $\boldsymbol{\mathcal{S}}_k^i$.

The drawback of PSOAF1 is overcome in PSOAF1*, where in addition to scaling the vector magnitudes, the vectors are directionally perturbed. The magnitudes and directions of $(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i)$ and



(a) (b)

Figure 4.7: PSOAF1*: Scatter plot of 10^4 possible stochastic vectors $\boldsymbol{\nu}^i_k$, generated using Monte Carlo simulations, with a) $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right) = [\frac{1}{\sqrt{2}}\ \frac{1}{\sqrt{2}}]$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right) = [-\sqrt{2}\ \sqrt{2}]$ and b) $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right) = [1\ \ 0]$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right) = [0\ \ 2]$.



(a) (b)

Figure 4.8: PSOAF1*: Scatter plot of 10^4 instances of the stochastic vectors $\boldsymbol{\nu}^i_k$, generated using Monte Carlo simulations, with a) $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right) = [\frac{1}{\sqrt{2}}\ \ \frac{1}{\sqrt{2}}]$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right) = [\sqrt{2}\ \ \sqrt{2}]$ and b) $\left(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k\right) = [1\ \ 0]$ and $\left(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k\right) = [2\ \ 0]$.

$(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i)$ are used to only indicate potential improvement, thereby placing only *some faith* in *both* direction and step size.

This is in contrast to PSOAF1, where *absolute faith* is placed in the directions prescribed by $(\boldsymbol{p}_k^i - \boldsymbol{x}_k^i)$ and $(\boldsymbol{p}_k^g - \boldsymbol{x}_k^i)$, while only *some faith* is placed on step size.

(Incidentally, termination occurs when $\boldsymbol{p}_k^i$, $\boldsymbol{p}_k^g$ and $\boldsymbol{x}_k^i$ converge on the same point in n-dimensional space, combined with $w\boldsymbol{v}_k^i \to \mathbf{0}$.)

4.6 Numerical experiments

An empirical study to quantify the (lack of) objectivity of the three discussed implementations of the PSOA is now performed. A synchronous updating method is used [31]. Real variables are implemented using double-precision floating-point arithmetic. For this study the algorithm parameters are $c_1 = c_2 = 2$, the swarm size is $p = 20$ particles and the computations are performed for various constant inertia factors w. Initial velocities are assumed to equal $\mathbf{0}$. In PSOAF1*, $\alpha = 3$ is simply selected. (The author does not seek an optimal value for α, but merely wishes to illustrate the effects of perturbing the vector directions.) Furthermore, no boundary or velocity restrictions are implemented. Each run consists of 200000 function evaluations (10000 iterations). All results presented are averaged over 100 runs.

In the study the following five test functions are used:
i) The Rosenbrock function (unimodal, f_0):

$$f_0(\boldsymbol{x}) = \sum_{i=1}^{\frac{n}{2}} \Big(100\big(x_{2i} - x_{2i-1}^2\big)^2 + \big(1 - x_{2i-1}\big)^2\Big).$$

ii) The Quadric function (unimodal, f_1):

$$f_1(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j \right)^2.$$

iii) The Ackley function (multimodal, f_2):

$$\begin{aligned} f_2(\boldsymbol{x}) = \; & -20 \exp\left(-0.2\sqrt{\tfrac{1}{n}\textstyle\sum_{i=1}^{n} x_i^2} \right) \\ & - \exp\left(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} \cos(2\pi x_i) \right) + 20 + e. \end{aligned}$$

iv) The generalized Rastrigin function (multimodal, f_3):

$$f_3(\boldsymbol{x}) = \sum_{i=1}^{n} \Big(x_i^2 - 10\cos(2\pi x_i) + 10\Big).$$

v) Finally, the generalized Griewank function (multimodal, f_4):

$$f_4(\boldsymbol{x}) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1.$$

Table 4.1: Test function parameters

Function	n	domain
f_0	30	$\pm$ 2.048
f_1	30	$\pm$ 100.0
f_2	30	$\pm$ 30.0
f_3	30	$\pm$ 5.12
f_4	30	$\pm$ 600.0

The parameters used in the study are given in Table 4.1. The *domain* column represents the range of each dimension of the design variables; the test function domains are symmetrical about 0.0 in all dimensions.

The multimodal functions f_2 and f_3 are decomposable [54], viz. the design variables are uncoupled. This implies that once an optimal value for a given design variable is obtained, it remains optimal, independent of the other design variables. This is similar to optimizing n 1-dimensional optimization problems, instead of 1 n-dimensional coupled optimization problem. The test set is therefore studied in the unrotated or decomposable reference frame $f(\boldsymbol{x})$, as well as in an arbitrary rotated reference frame $f(\boldsymbol{Rx})$, in which the design variables are coupled [55]. Here, $\boldsymbol{R}$ is a random, proper orthogonal transformation matrix, constructed as in [54]. The transformation matrix results in a pure rotation of each test function. For each of the 100 independent runs, a new random rotation matrix $\boldsymbol{R}$ is constructed, to ensure that there is no bias toward any particular reference frame.

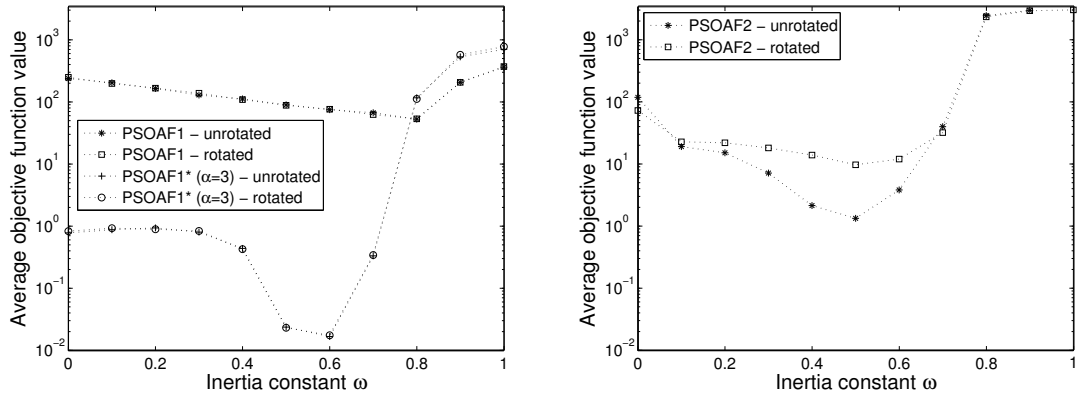
4.7 Discussion of Results

Depicted in Figures 4.9, 4.10, 4.11, 4.12 and 4.13 are the mean objective function values after 2×10^5 function evaluations (or 10000 iterations) averaged over 100 runs for both the unrotated and rotated functions.

The *rotational invariance* of PSOAF1 and PSOAF1* are evident from Figures 4.9(a), 4.10(a), 4.11(a), 4.12(a) and 4.13(a). The poor performance of PSOAF1 directly results from the particle trajectories collapsing to lines as shown in Chapter 3. There is a significant improved performance for all the test functions with PSOAF1*, due to the scaling of the vector magnitudes and perturbation of the vector directions.

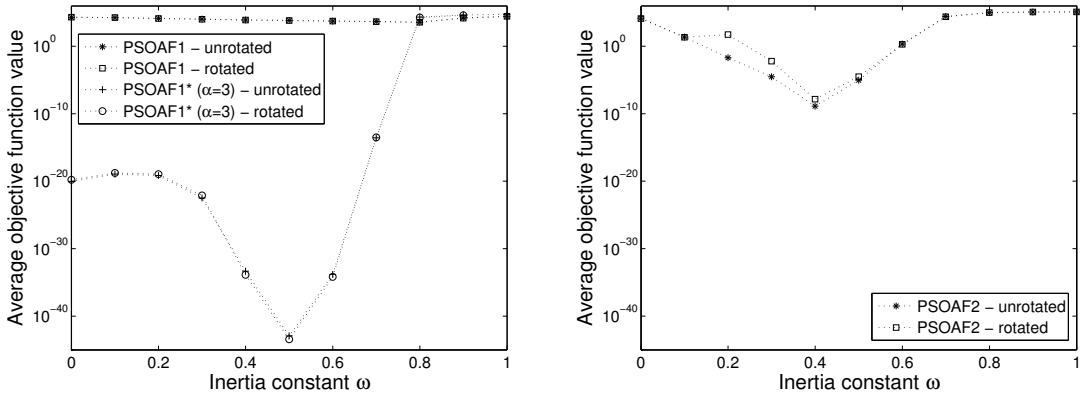
The *rotational variance* of PSOAF2 is evident from Figures 4.9(b), 4.10(b), 4.11(b), 4.12(b) and 4.13(b). There is a *severe* performance loss for some of the rotated functions compared to the unrotated functions.

Two functions result in similar performance for the rotated and unrotated functions, namely the Quadric function f_1, and the Griewank function, f_4. The Quadric and Griewank functions are almost insensitive to rotation. (The Griewank function is a spherical function on which sinusoidal "noise" is imposed. Hence this function is artificially indifferent to rotation, since many local minima appear, irrespective of rotation.)



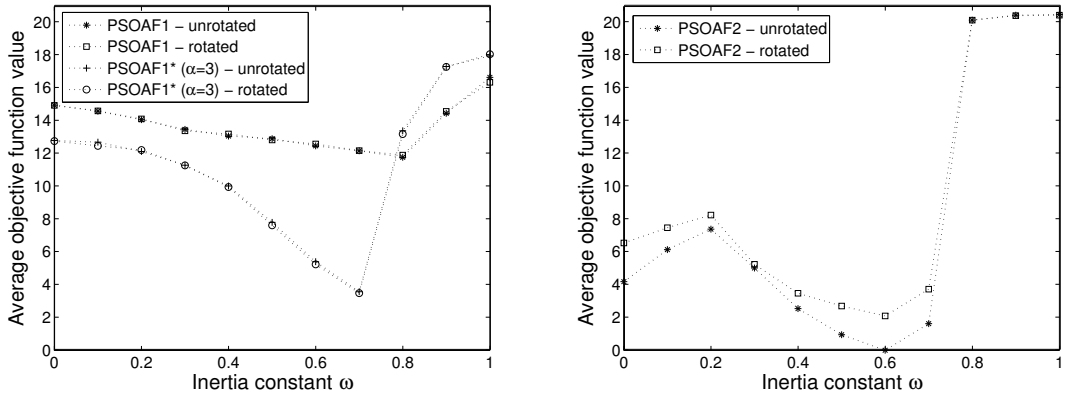
(a) (b)

Figure 4.9: Average function value obtained with a) PSOAF1 and PSOAF1*, and b) PSOAF2 after 2×10^5 function evaluations (10000 iterations) averaged over 100 runs on the rotated and unrotated Rosenbrock test function f_0.



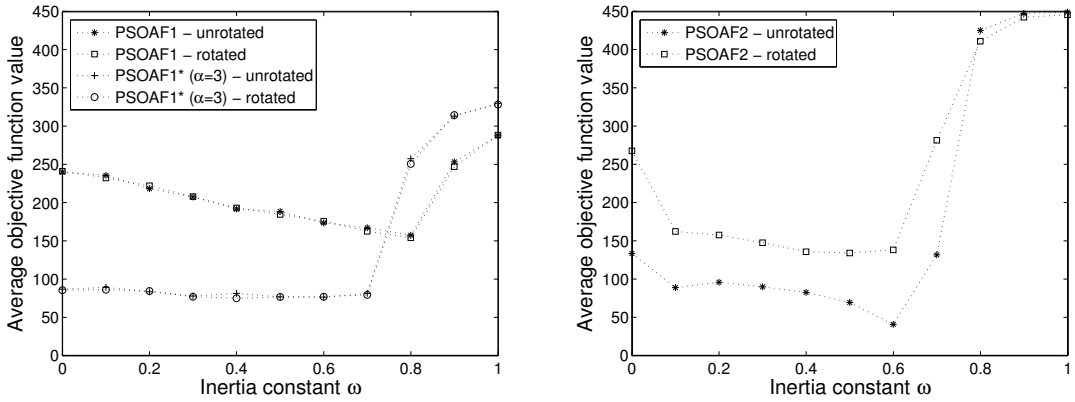
(a) (b)

Figure 4.10: Average function value obtained with a) PSOAF1 and PSOAF1*, and b) PSOAF2 after 2×10^5 function evaluations (10000 iterations) averaged over 100 runs on the rotated and unrotated Quadric test function f_1.



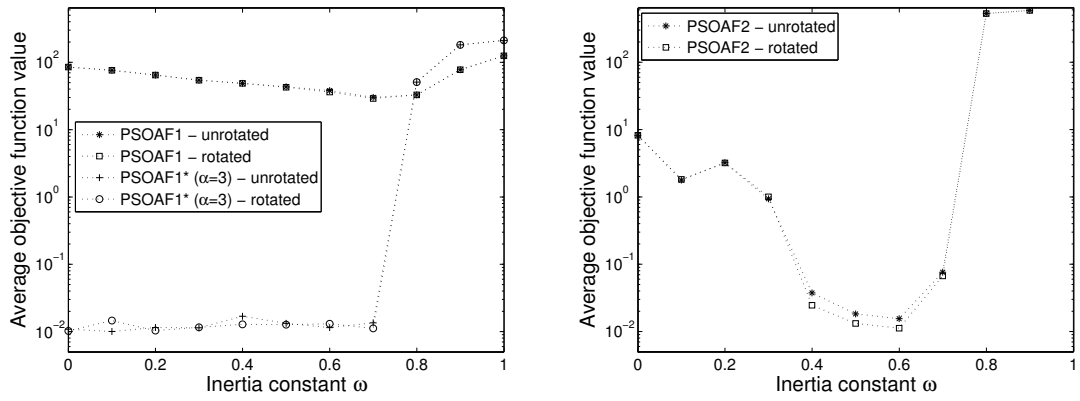
(a) (b)

Figure 4.11: Average function value obtained with a) PSOAF1 and PSOAF1*, and b) PSOAF2 after 2×10^5 function evaluations (10000 iterations) averaged over 100 runs on the rotated and unrotated Ackley test function f_2.



(a) (b)

Figure 4.12: Average function value obtained with a) PSOAF1 and PSOAF1*, and b) PSOAF2 after 2×10^5 function evaluations (10000 iterations) averaged over 100 runs on the rotated and unrotated Rastrigin test function f_3.



(a) (b)

Figure 4.13: Average function value obtained with a) PSOAF1 and PSOAF1*, and b) PSOAF2 after 2×10^5 function evaluations (10000 iterations) averaged over 100 runs on the rotated and unrotated Griewank test function f_3.

Inadvertently, this also suggest that non-spherical unimodal test functions should be used to evaluate objectivity. The two unimodal functions, namely the Rosenbrock function f_0 and the Quadric function f_1, are of some interest, since they indicate the ability of an algorithm to search within a local basin. The performance of PSOAF1* is significantly better than PSOAF2 for both functions, for both the rotated and unrotated test functions. PSOAF2 demonstrates a severe performance loss for the Rosenbrock function, for the rotated function compared to the unrotated function. (Note the scale of the graphs in Figure 4.9.)

The performance difference between PSOAF2 and PSOAF1* for the unimodal Quadric test function f_1, is depicted in Figure 4.14. Figure 4.14 depicts the mean function value convergence history of PSOAF1 (with $w = 0.8$), PSOAF2 (with $w = 0.4$) and PSOAF1* (with $w = 0.5$ and $\alpha = 3$) over 2500 iterations. The values for w are optimal for each algorithm, but no attempt was made to optimize α. Of the three formulations, it is clear that PSOAF1* is computationally the most effective on the Quadric test function.

For the multimodal functions, PSOAF2 demonstrates notable performance loss. See for example the Ackley function f_2, and the Rastrigin function f_3. In contrast, the performance of PSOAF1* is comparable to the best obtained with PSOAF2, with no performance loss due to rotation.

For the sake of clarity, an overview of the performances of PSOAF1, PSOAF2 and PSOAF1* is given in Table 4.2. The table summarizes the best function values obtained, together with the inertia factor at which the best function value is obtained after 2×10^5 function evaluations (10000 iterations). The results for both the unrotated and rotated test functions are given.

Table 4.2: Constant inertia factor at which the best average objective function value is obtained for the unrotated test functions. The accompanying average objective function value for rotated test functions is also presented.

	PSOAF1		
		$f_{\text{unrotated}}$	f_{rotated}
	w	$f_{\text{avg}}^{\text{best}}$	$f_{\text{avg}}^{\text{best}}$
f_0	0.8	54.071	54.712
f_1	0.8	4087.657	4123.801
f_2	0.8	11.791	11.925
f_3	0.8	157.656	154.368
f_4	0.7	30.924	29.731
	PSOAF2		
		$f_{\text{unrotated}}$	f_{rotated}
	w	$f_{\text{avg}}^{\text{best}}$	$f_{\text{avg}}^{\text{best}}$
f_0	0.5	1.358	9.905
f_1	0.4	1.4×10^{-9}	1.5×10^{-8}
f_2	0.6	8.6×10^{-15}	2.099
f_3	0.6	40.992	138.507
f_4	0.6	1.5×10^{-2}	1.1×10^{-2}
	PSOAF1* ($\alpha = 3$)		
		$f_{\text{unrotated}}$	f_{rotated}
	w	$f_{\text{avg}}^{\text{best}}$	$f_{\text{avg}}^{\text{best}}$
f_0	0.6	1.6×10^{-2}	1.7×10^{-2}
f_1	0.5	1.2×10^{-43}	3.8×10^{-44}
f_2	0.7	3.583	3.492
f_3	0.5	76.949	76.880
f_4	0.6	1.1×10^{-2}	1.3×10^{-2}

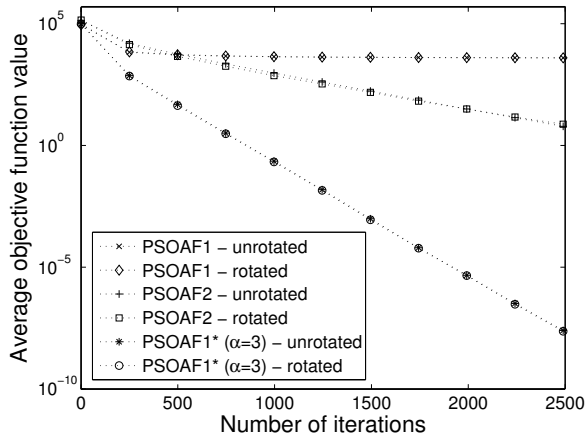


Figure 4.14: Mean function value history plot averaged over 100 runs on the rotated and unrotated Quadric test function f_3 with PSOAF1 (with $w = 0.8$), PSOAF2 (with $w = 0.4$) and PSOAF1* (with $w = 0.5$ and $\alpha = 3$).

4.8 Comments on PSOAF1*

4.8.1 On invariance

It is now shown that PSOAF1* is not strictly rotationally invariant, but only in a stochastic sense.

Consider an arbitrary vector, expressed in two different reference frames, by respectively $\boldsymbol{y}$ and $\boldsymbol{y}'$. The two reference frames are related by a pure rotation $\boldsymbol{M}$, hence

$$\boldsymbol{y}' = \boldsymbol{M}\boldsymbol{y}, \quad \boldsymbol{M} \in \text{Orth}^+, \tag{4.8}$$

where Orth^+ indicates the space of proper orthogonal matrices.

Now apply two independent directional perturbations (rotations) $\boldsymbol{Q} \in \text{Orth}^+$ and $\boldsymbol{Q}' \in \text{Orth}^+$ to $\boldsymbol{y}$ and $\boldsymbol{y}'$ respectively. The vectors $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{y}}'$ then obtained, are respectively given by

$$\hat{\boldsymbol{y}} = \boldsymbol{Q}\boldsymbol{y}, \tag{4.9}$$

and

$$\hat{\boldsymbol{y}}' = \boldsymbol{Q}'\boldsymbol{y}'. \tag{4.10}$$

Strict deterministic rotational invariance requires a one-to-one mapping of the perturbed vectors in either reference frame. Hence

$$\hat{\boldsymbol{y}}' = \boldsymbol{M}\hat{\boldsymbol{y}}, \quad \forall\, \boldsymbol{M} \in \text{Orth}^+ \tag{4.11}$$

By substituting Eqs. (4.8), (4.9) and (4.10) into Eq. (4.11), the following is obtained

$$\boldsymbol{Q}'\boldsymbol{M}\boldsymbol{y} = \boldsymbol{M}\boldsymbol{Q}\boldsymbol{y}, \quad \forall\, \boldsymbol{M} \in \text{Orth}^+. \tag{4.12}$$

Eq. (4.12) is rewritten as

$$(\boldsymbol{Q}'\boldsymbol{M} - \boldsymbol{M}\boldsymbol{Q})\boldsymbol{y} = \boldsymbol{0}, \quad \forall\, \boldsymbol{M} \in \text{Orth}^+. \tag{4.13}$$

Since Eq. (4.13) has to hold for any arbitrary vector $\boldsymbol{y}$, it follows that

$$\boldsymbol{Q}'\boldsymbol{M} = \boldsymbol{M}\boldsymbol{Q}, \quad \forall\ \boldsymbol{M} \in \text{Orth}^{+}. \tag{4.14}$$

The unique solution to Eq. (4.14) is that both $\boldsymbol{Q}'$ and $\boldsymbol{Q}$ are the second-order isotropic tensor, i.e.

$$\boldsymbol{Q}' = \boldsymbol{Q} = \boldsymbol{I}. \tag{4.15}$$

The foregoing implies that a strict enforcement of rotational invariance results in $\boldsymbol{Q}^i_{lk} = \boldsymbol{I}, \quad l = 1, 2$. In other words, PSOAF1* reduces to PSOAF1.

However, since the PSOA is a stochastic algorithm, it is adequate to satisfy Eq. (4.15) in an average sense only. In order to satisfy $\boldsymbol{Q}' = \boldsymbol{Q} = \boldsymbol{I}$ in a stochastic sense, it is sufficient to require that $\text{mean}(\boldsymbol{Q}') = \text{mean}(\boldsymbol{Q}) = \boldsymbol{I}$, if the probability distributions of $\boldsymbol{Q}'$ and $\boldsymbol{Q}$ are chosen equal over identical domains.

4.8.2 Implementational issues of PSOAF1*

Further to the implementation in Section 4.5, numerous strategies exist to achieve independent directional perturbation.

An obvious, computationally inexpensive possibility is to randomly perturb each component of the unit vectors $(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)/||(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)||$ and $(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)/||(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)||$; the vectors $(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)$ and $(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)$ are then reconstructed from the normalization of the perturbed vectors. (Although this makes a rigorous mathematical analysis of the algorithm difficult.)

In the implementation, in updating $\boldsymbol{Q}$, strategies to limit the computational expense associated with matrix multiplications and the generation of random numbers may also be implemented. For example, multiplying the sum of $c_1(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)$ and $c_2(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)$ by a single random rotation matrix, reduces the number of matrix multiplications by half.

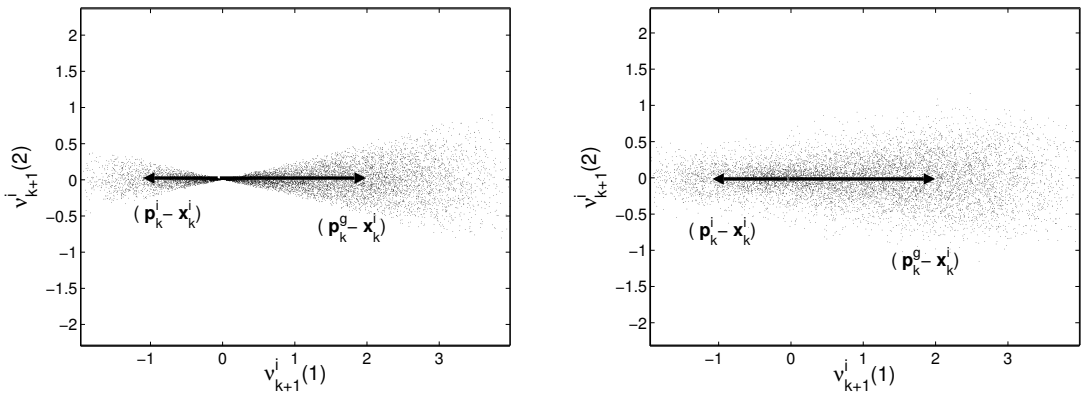
However, depicted in Figure 4.15 is the difference in instantaneous search domain that results when independent rotation matrices $\boldsymbol{Q}^i_{1k} \neq \boldsymbol{Q}^i_{2k}$ are used, as opposed to identical rotation matrices $\boldsymbol{Q}^i_{1k} = \boldsymbol{Q}^i_{2k}$.

To reduce the computational effort even further, the vectors $c_1(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)$ and $c_2(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)$ of all the particles can be directionally perturbed by the same independent rotation matrices, viz. $\boldsymbol{Q}^i_{lk} = \boldsymbol{Q}_{lk}, \ l = 1, 2$ and $i = 1, 2, \cdots, p$.

4.8.3 Alternatives to PSOAF1*

Finally, there are of course numerous methods to introduce diversity into PSOAF1, as opposed to the proposed option of independent directional perturbation.

Only a single alternative is mentioned here, namely an increase in the social awareness of the particles. In turn, this may for example be effected by increasing the number of particles that contribute to Eq. (4.1) [35, 34]. (One may of course achieve n-dimensional searches, if the number of particles $p \geq n$, unless the particle trajectories are parallel.) Additional information about the objective function is then also used in the searches of any particle i.



(a) (b)

Figure 4.15: Scatter plot of 10^4 possible stochastic vectors $\boldsymbol{\nu}^i_k$, generated using Monte Carlo simulations, with $(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k) = [-1\ 0]$ and $(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k) = [2\ 0]$ using a) identical rotation matrices $\boldsymbol{Q}^i_{1k} = \boldsymbol{Q}^i_{2k}$ and b) independent rotation matrices $\boldsymbol{Q}^i_{1k} \neq \boldsymbol{Q}^i_{2k}$.

4.9 Closure

It is shown that PSOAF1 is objective, but it demonstrates an overall poor performance, due to the particle trajectories collapsing to lines. This is a direct result of only scaling the magnitude of the cognitive and social vectors $c_1(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)$ and $c_2(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)$.

In turn, PSOAF2 is not objective, which results in severe performance loss for "rotated" functions. Nevertheless, PSOAF2 still outperforms PSOAF1 for both rotated and unrotated test functions, since the algorithm is diverse, i.e. the particle trajectories do not collapse to lines.

A novel implementation denoted PSOAF1* is proposed, which is both objective and diverse. In PSOAF1*, the magnitudes are scaled, and the directions perturbed independently, of both the cognitive and social vectors $c_1(\boldsymbol{p}^i_k - \boldsymbol{x}^i_k)$ and $c_2(\boldsymbol{p}^g_k - \boldsymbol{x}^i_k)$. (This however comes at the cost of an additional scaling factor.) PSOAF1* outperforms PSOAF2 for the unimodal functions used, for both the rotated and unrotated test functions. In addition, its performance is comparable to PSOAF2 for the multimodal functions, with the added advantage of being independent of the reference frame in which the objective function is formulated.