Adapting Particle Swarm Optimisation for Fitness Landscapes with Neutrality

The concept of ‘Neutral Networks’ in Fitness Landscapes - where amongst any rugged terrain there may also be connected regions or pathways over which solution fitness does not change - has been recently shown to be of significance to the use of Evolutionary Algorithms. To our knowledge this important aspect of the fitness landscape has not previously been examined in the context of Particle Swarm Optimisation (PSO). The standard PSO algorithm is here shown to be inadequate for optimisation tasks where such neutrality exists; we investigate modifications of a standard PSO and compare their performances on various novel fitness landscapes that contain neutrality. One simple modification to the standard PSO algorithm is shown to enable significantly improved functionality upon neutral landscapes, with no compromise to operation upon non-neutral terrains.

1. INTRODUCTION

Particle Swarm Optimisation (hereafter PSO) and the Genetic Algorithm (GA) represent population based optimisation heuristics for searching in high-dimensional spaces. Populations of potential solutions are intended to move collectively through a problem search space, under their respective algorithmic strategies, towards ‘fitter’ regions (represented by better solutions) and ideally to a solution representing the global optimum. For a merely two-dimensional search space, visualised as a map or the surface of a country area, a third dimension of ‘elevation’ can represent fitness; this leads to the metaphor of a ‘Fitness Landscape’ (FL) where the population is searching for higher peaks representing the best global fitness. The smooth or rugged nature of this landscape depends entirely upon the specific nature of the problem. This metaphor can be extended towards higher dimensional search spaces; although here ones’ intuitions based on familiar three-dimensional landscapes may prove misleading.

In the field of Evolutionary Algorithms (and particularly the GA), it has been recognised that some search problems may correspond to FLs that contain significant ‘neutral’ or flat (zero-gradient) regions, as well as slopes of varying positive or negative gradients (Fig 1). In a 2-D search space this would correspond to ridges or ‘neutral paths’ in the FL, possibly connecting two hills across what otherwise would be a valley. In higher dimensional search spaces and their associated FLs, the extra dimensions create additional scope for such neutral pathways that can form interconnected ‘neutral networks’ traversing the space [1,2]. In the context of evolutionary search algorithms, it has been shown [1,2] that where such neutral networks exist, the corresponding search dynamics are transformed; the problem of avoiding stagnation upon local optima may no longer exist, as local regions of high fitness may be connected to fitter regions via neutral networks. Instead the issue becomes one of how best to traverse these neutral networks, such that better solutions can be found. Under certain conditions, optimal evolutionary search strategies can be derived (e.g. [3]). It is thus advisable to adapt a search strategy accordingly whenever it is believed that a fitness landscape may contain significant amounts of neutrality.

We are not aware of any previous discussion of neutrality in the context of PSO, and the purpose of this paper is to draw this issue, that we believe important, to the attention of the PSO community. We demonstrate that the current ‘standard’ PSO algorithm [4] is not optimised for complex domains wherein a certain degree of neutrality exists, by developing and competitively testing modified versions of the standard PSO upon novel terrains containing tuneable degrees of neutrality. It is shown that one trivial modification to the standard algorithm not only significantly improves the operation of PSO upon terrains that include neutrality, but also does not impair performance upon terrains of the classic, non-neutral form.

This paper is organised as follows: Section 2 provides a commentary on the specific significance of neutrality to the evolutionary algorithm. Section 3 then introduces PSO, and provides an analysis of why the standard methodology is not optimised for neutral terrains. Section 4 proposes two separate modifications to the standard version enabling the exploitation of neutral landscape features, and introduces a novel landscape for testing the performance of these new heuristics. Section 5 concludes the paper in discussing a competitive experimental analysis of the performance of these new heuristics and the standard PSO.
2. NEUTRAL NETWORKS

The concept of the Fitness Landscape is derived originally from evolutionary theory (particularly Sewall Wright’s ‘adaptive landscape’ [5] and John Maynard Smith’s ‘Protein Space’ [6]), but can be extended naturally to the search spaces of PSO or the GA, where the dynamics of Darwinian selection upon a population can be visualised as traversing a FL ‘uphill’ toward peaks of highest local fitness.

Consider an artificial population within a continuous sequence space, for the moment devoid of neutrality, in which every conceivable sequence gives rise to a near unique fitness score; there are ‘hills’ and ‘valleys’ within a ‘rugged’ fitness landscape. Premature convergence might be expected upon ‘local optima’ within this environment - locally isolated sub-optimal solutions that may be very difficult to escape. Consider alterations to the evolutionary dynamic that might enable escape in such an event of local stagnation: Figure 2A (below) illustrates a population of one-dimensional ‘sequences’ converging upon the leftmost (sub-optimal) peak A. One method of escape from this ‘metastable’ optimum would require a temporary but selective reduction in individual fitness (B); but how could it be algorithmically determined that such a recursive step would actually result in a fitness increase, without an additional and possibly lengthy sequence in processing? Alternatively, macro-mutation might instead create the possibility of individuals ‘leaping’ between fitness peaks (Figure 2B). Under these circumstances the respective evolutionary dynamics would appear closer to those of a random search in sequence space however, and this is not likely to serve as a suitable dynamic in rapidly and efficiently seeking global optima.

Now imagine a population occupying an optimisation domain wherein there is some redundancy in the genotype to phenotype mapping; many different genotypes generate the same fitness score, thus there are more sequences than structures [1,2]. Series of genotypes each interconnected (by point mutations in the relevant dimension) to one or more same-fitness neighbours [1,2,7,8], form ‘ridges and bridges’ in the fitness domain. The presence of selective neutrality then creates the situation that landscape features may provide an escape from local optima; indeed, local optima may not even exist [9] (figure 3). The landscape view (as determined by the specific fitness function) has changed in such a way as to create the opportunity for continued exploration, avoiding premature convergence, if the algorithm is adequately tuned to exploit it.

For the purposes of illustration these examples are of low dimensionality. In higher dimensions however, the effects of neutrality are expected to be even more profound [1]. In some ‘real-world’ systems of high dimensionality (typically 3-d and above), neutral networks are suggested to permeate sequence space [1,2,10]. Such networks are the hyper-dimensional correlate of the previously described ‘ridges and bridges’, and are expected to connect any given sequence with, for practical purposes, untold others.

When this is so, a population may feasibly ‘explore’ sequence space over multiple generations via the explicit traversal of neutral networks [8]. This creates the possibility of innovative, higher fitness genotypes adjacent to the neutral network being discovered at each generational step [11]. Selective neutrality introduces the possibility of ‘free’ exploration in sequence space via stochastic neutral drift, if such behaviour is enabled by the evolutionary dynamic (i.e. respectively enabled within the present optimisation heuristic). Premature convergence is then unlikely, and local optima are potentially escapable, given time [3,9].

It is necessary to note that evolutionary neutrality is not solely a theoretical concept, but is observable in nature. Many original studies regarding the theme come from models of RNA ‘secondary structure’ formation.
A transfer RNA molecule, consisting of a sequence of nitrogenous bases of a limited character set, ‘folds’ under natural conditions into a three-dimensional macromolecule [10]. Sequence space maps into ‘shape’ space, which can subsequently be evaluated in terms of a scalar fitness score [1]. There exists redundancy in this mapping schema, as various sequences can be expected to fold into the same secondary structure [1,10]. Work in modelling this phenomenon by Schuster et al [1,2,10] includes the postulation of four properties of the natural RNA sequence to shape-space mapping scheme:

i) There are more sequences than structures – many individually unique sequences fold into the same secondary structure

ii) There exist few ‘common’, and many ‘rare’ structures: Most of the possible sequences form into just a few discrete structures. Consequently, there are a relatively small number of sequences that produce the remaining diverse variety of secondary structures.

iii) The distribution of sequences mapping into the same structure appears random in sequence space

iv) There exist interconnected neutral networks – ‘shape’ space is neutrally connected to such a degree that networks described as ‘percolating’ span the entire domain. All regions of the search space are attainable by the process of random neutral drift

Each postulate can be used to describe the characteristics of neutrality in generic sequence spaces, and can be referred to in investigating the existence of neutrality within both artificial and natural search spaces. Of particular note is the fourth postulate; sufficiently interconnected neutral networks may generate the possibility of an evolving population reaching the neighbourhood of any possible structure – a property termed ‘constant innovation’ [8]; the population never gets stuck on local optima.

In many ‘test functions’ designed to investigate the quality of optimisation heuristics (for example De Jong’s widely used ‘test suite’ [12]), the fitness landscape is often implicitly one of rugged hills and local peaks, leading to the legitimate expectation of premature convergence upon sub-optima. Furthermore, methodologies for the avoidance of early loss in sequence diversity are quite common in current literature, suggesting that this view is widely accepted within the field of evolutionary computation. However, the concept of neutrality presents a markedly different view of the optimisation landscape, and one that is evidently not yet widely considered or acknowledged, through lack of literature on the topic and persistence of ‘new models’ for the escape of premature convergence. Indeed, there is the possibility that neutrality may be quite common in decidedly real-world problems, such as evolutionary circuit [13] and artificial neural robotic controller design [14,15].

Recently, the significance of evolutionary neutrality to the GA was explored at length by Barnett [3,9] who performed an analysis of the dynamics of artificial evolution upon a binary landscape containing neutrality. A view of neutral networks as ‘tangled webs’ emerged, in reference to the percolation property and potential for constant innovation thereupon. A modified heuristic called the ‘Netcrawler’ [3] was subsequently proposed, which consists of a ‘population of one’ allowed to traverse neutral networks by generational mutation alone. The Netcrawler model was later deployed upon an investigation into real-world circuit evolution (Thompson and Harvey [13,16]), with some success.

Barnett [17] also proposed a modification to Kauffman’s rugged NK landscape [18] test function, suggesting a new parameter for the incorporation of a user specified level of neutrality to the landscapes generated (the NKp landscape). This forms one of only a few test functions specifically designed to incorporate dynamical testing with neutral networks (see also Newman and Engelhardt’s model [19]).

More recently still, Izquierdo-Torres [14] explored the significance of evolutionary neutrality within a GA deployed upon continuous landscapes, specifically with regards to the evolution of artificial neural robotic controllers, wherein ‘nearly neutral’ mutations were common: mutations that affect the resultant fitness so slightly as to be insignificant, and could to some level be disregarded. What significance does evolutionary neutrality have when deploying another real-numbered optimisation methodology, namely that of Particle Swarm Optimisation? Literature concerning PSO coupled with theory on neutrality appears so hard to come by, that it is assumed that the question has simply not yet been formally asked.

3. PARTICLE SWARM OPTIMISATION

James Kennedy and Russell Eberhart first proposed Particle Swarm Optimisation in the mid 1990s as a new tool for computational optimisation, inspired by the simulation of collective ‘swarming’ and ‘flocking’ of sociable animals [20]. Following a decade of growing research interest in the algorithm, an online ‘standard’ PSO was introduced, in a necessary effort to create a common yardstick for PSO development [4]. Based upon an improved version of the 1995 ‘first proposed’ model [20,21], this standard is (according to the source) validated by PSO co-creator James Kennedy, and researcher/PSO writer Maurice Clerc, and comes with suggested (though not universally optimal) parametric settings.

The Particle Swarm Optimisation methodology operates by placing a ‘swarm’ of individual agents or ‘particles’ into a continuous sequence space, wherein each particle is described by an N-dimensional vector location and a fitness score. The heuristic thus represents a population-based methodology for computational optimisation that is, in its basic form, limited to the application of real-numbered
problems. Each particle possesses a physical, $N$-dimensional velocity within this virtual hyperspace, the specific mathematical rule of which defines the ‘evolutionary’ dynamics of the swarm and generates the capacity for optimisation: iteratively applied physical dynamics guide the swarm to regions of higher fitness in solution space, by weighting individual acceleration towards the currently standing ‘global best’ or fittest yet ‘sampled’ location in solution space. Fitness is assessed upon the current location vector for each particle, and the population best updated whenever a fitter location is discovered. Stochastic factors render the methodology non-deterministic, and additional weighting towards an individual, tuneable ‘personal best’ factor maintains swarm diversity. The 2006 standard velocity update and location update rules are those given in equations (1) and (2) respectively.

**Velocity update / location update rules:**

1998 version incorporating $w$ parameter [4,21]

(1): $V_i = wV_i(t-1) + c_1r_1(pBest_i-pLoc_i) + c_2r_2(gBest_i-pLoc_i)$

(2): $pLoc_i = pLoc_i(t-1) + V_i$

**Nomenclature:**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w$</td>
<td>Inertial weight constant</td>
</tr>
<tr>
<td>$V$</td>
<td>N-d Velocity vector</td>
</tr>
<tr>
<td>$i$</td>
<td>Current particle index</td>
</tr>
<tr>
<td>$t$</td>
<td>Current iteration</td>
</tr>
<tr>
<td>$c_1$, $c_2$: Weighting constants</td>
<td></td>
</tr>
<tr>
<td>$r_1$, $r_2$: Stochastic variables $[0 &lt; r_x &lt; 1]$, drawn from a uniform distribution</td>
<td></td>
</tr>
<tr>
<td>$gBest$: N-d Global ‘population’ best solution vector</td>
<td></td>
</tr>
<tr>
<td>$pBest$: N-d Particle ‘personal’ best location vector</td>
<td></td>
</tr>
<tr>
<td>$pLoc$: N-d Particle current location</td>
<td></td>
</tr>
</tbody>
</table>

In years following its introduction, PSO gained popularity within academic circles as it was shown to perform as well as, and in some cases better than, the current favourite optimisation tool: the GA [22]. In most cases, PSO is held to be able to converge more rapidly than the GA (a consequence of the variable step size in the dynamic ‘mutation’ operator, the particle velocity) but suffers from premature convergence upon sub-optimal locations. In the past decade there has been a veritable explosion of new methodologies designed to avoid premature convergence upon local fitness peaks - a consequence of testing upon rugged ‘toy’ landscapes. In light of studies into evolutionary neutrality however, this might not remain an accurate view of the FLs inherent in real world problems.

Consider the operation of the standard PSO algorithm discussed above, upon an imagined sequence space containing evolutionary neutrality. In the standard PSO model, the current global best is updated according to the following logic:

- **If current particle’s location is of fitness greater than the current global best, make this location the new global best.**

Such behaviour would lead to the anchorage of the $gBest$ attraction point upon the first sample point assessed upon the current fittest neutral level (fig 4). Neutral levels of higher fitness might be found through basic exploration of space around the global best (motivated by the attraction towards personal best locations), but unless this is successful, the beacon attraction point $gBest$ remains attached at the point of initial discovery; neutral drift would be hindered and the swarm diversity can be expected to stagnate.

![Fig. 4: In the standard PSO model, the global best is likely to act as a point of anchorage in sequence space where a certain level of neutrality exists, stemming the exploration of neutral networks, if they exist.](image)

Projected this simple picture into a more realistic dimensionality (wherein neutral networks exist) it would appear that without drift, the potential for permanent innovation through traversal of the neutral network is foregone. From this it can be stated that the current standard PSO algorithm is not optimised for landscapes containing a certain degree of neutrality, as will be demonstrated below.

It transpires that a very simple change to the logic of the $gBest$ update rule can allow for the traversal of neutral networks. A simple but successful methodology drawn from GA research also proves to be suitably transferable into the PSO heuristic.

**4.1 NEUTRAL PSO (nPSO)**

The development of a new heuristic for terrains containing neutrality begins with a consideration of the most basic change that would be required of the standard PSO model, in order to allow the algorithm to exploit the existence of neutral networks within sequence space. It transpires that this change consists of just a single extra character in the program code, affecting the logic of the global best update rule. The current standard $gBest$ update logic (given above) can be
trivially modified to allow the algorithm a very basic explorative capacity of neutral terrains:

If current particle’s location is of fitness greater than or equal to the current global best location, make this location the new global best.

In this basic form, the most recent particle to be updated, if sharing the current global best in a given iteration, then becomes the new gBest. The nPSO methodology is expected to remove the standard methodology’s perceived issue of the gBest ‘point of anchorage’ (figure 4), by freeing the population attractor to move with the swarm on neutral zones.

4.2 EXTREMA SELECTION PSO (esPSO)

For further comparison we consider also ‘Extrema Selection PSO’ (esPSO) inspired by Terry Stewart [23], in which a modification to the basic GA was developed in order to enable accelerated evolution upon neutral networks. Stewart’s original ‘Extrema Selection’ model works by defining an average population centre (the ‘centroid’), and subsequently selecting members for replication based upon their Euclidean distance from this vector, whenever fitness levels are shared within the population. In such cases individuals further from the centroid are more likely to be selected, leading to the rapid motion of the population cluster upon neutral regions of sequence space.

This methodology can be directly transferred into the PSO heuristic: when multiple swarm members share the same fitness, the gBest location is simply decided by relative distance from the population centroid calculated in each iteration. Again this detaches the ‘anchorage’ of gBest upon neutral zones, enabling the swarm to non-deterministically drift. The subsequent stochastic dynamics already present within the standard model are believed to then motivate the swarm into non-deterministic exploration upon neutral zones; the swarm experiences neutral drift.

4.3 CONSTRUCTING A NEUTRAL TEST BED

Given time constraints, the investigation of a ‘real’ problem containing a high degree of neutrality (e.g. Harvey and Thompson’s circuit evolution [13,16]) was not possible, leading to the preference for testing the proposed new heuristics upon artificial ‘toy’ landscapes. Such neutral test landscapes exist (e.g. Barnett’s NKp landscape [17]), but having been built with the GA in mind, these are binary and thus unsuitable for the continuous methodology of PSO. For this reason it becomes necessary to create an artificial test bed of real-numbered toy neutral FLs.

‘Quantised’ classic functions: De Jong’s test suite [12] represents a group of typical, continuous problem landscapes upon which new PSO heuristics are often tested. These take the form of a mathematical formula, which maps a real-numbered N-d vector (a location in sequence space) into a one-dimensional scalar fitness value. A classic example is the ‘Sphere’ function, which calculates fitness by simply summing the square of all sequence components (equation 3):

$$f_1(d_1, d_2, ..., d_n) = \sum_{i=1}^{n} d_i^2$$

Nomenclature:

- $f_1$: Function 1 - Sphere
- $d_i$: Sequence vector components
- $i$: Component index
- $n$: Sequence vector dimensionality

The resultant landscape then appears as a ‘U’ shaped curve in one dimension, with fitness increasing as each component moves further from zero.

Neutrality can be artificially incorporated into this and similar functions, by ‘filtering’ the fitness output into a user specified quantum. This has the effect of levelling gradients into stepped plateaus: neutral regions of same-fitness space (fig 5).

‘Quantising’ the output of any function is as simple as dividing the ‘raw’ fitness score by a user-specified quantum size, and deducting the remainder from the raw value to give a ‘filtered’ fitness score. Where the current programming tool offers the ‘Mod’ functionality, this is deployed as follows:

$$\text{fitness output} = \text{raw fitness} - \text{mod(raw fitness, Q)}$$

Where Q is a user specified ‘quantum size’ that defines how much neutrality is attached to any given vector in sequence space. This filter functionality can be deployed on any continuous landscape to create tuneable neutrality, however, it does not guarantee that neutral networks with the property of constant innovation will arise; these functions represent mathematical patterns, and it cannot be assured that a given pattern gives rise to ‘realistic’ neutral networks as might be found in ‘patternless’ real-world optimisation problems.

‘Stacking plates’: Imagine a sequence of real numbers that can be deconstructed into pairs of adjacent values, wherein each pair represents the centre placement of a fixed radius circle in 2-d space (figure 6, overleaf). An optimisation scheme can be built from this model, whereby sequences...
Fig. 6: Adjacent paired components map into circle constructs in 2-d space

This optimisation scheme introduces great scope for neutrality: imagine a sequence of six numbers defining three 2-d circles, each circle is close to, but not touching any other circle. A mutation in the sequence brings two circles into overlap, and the sequence increases in fitness by one. Further mutations move the remaining untouched circle about in space but fail to bring it into overlap, whilst the two touching circles remain in contact. Each change in sequence that did not affect fitness is a neutral change, and though the former circle remains untouched, its position has drifted to a location distant from its origin.

Neutrality is tuneable with three parameters: the size of circle radius, the number of circles (the problem dimensionality halved) or the range of initial distribution in 2-d space. For example, with a large number of ‘big’ circles initialised within a small 2-d volume, some overlaps can be expected at the start with only short distances between new overlaps – there is a little neutrality. With few small circles in a large initial volume, few starting overlaps are expected, with large distances between individual circles – there is a lot of neutrality.

With a marked amount of neutrality, the landscape is expected to be very ‘smooth’, but note that the landscape dynamically changes with the motion of individual circles. This then represents an unpredictable model that might be more representative of problems in the real world.

5.1 COMPETITIVE TESTING OF HEURISTICS

The two discussed modifications to the standard PSO heuristic are here tested in competition with an unmodified standard PSO. It was decided for the purposes of testing the new heuristics to use 2 FLs: (A) an inverted, quantised N-d Sphere and (B) the plate stacking function discussed in section 4. The Sphere function follows equation 3, with inversion (multiplying the resultant fitness score by -1) giving a maximum attainable fitness score of zero. Parameter settings for each PSO methodology follows the standard values given in [4] where parameters are shared, with the only variable under test being the explicit methodological alterations discussed in section 4.

5.2 RESULTS

Results are presented here both graphically and in tabular format, given two types of data; the first is of the average fitness of gBest per iteration, and allows discrimination between the average performances of heuristics. The second is the diffusion coefficient [9], which is a measurement of the distance moved by the centroid per iteration and roughly equates to the speed of the swarm’s motion in sequence space. For each display of diffusion coefficient data presented here (graphs 2,4,6), plotted values represent an average encompassing 50 preceding iterations, in order to smooth the graphical illustration.

Graph 1: Average fitness of gBest per 500 iterations for three PSO heuristics optimising a six-dimensional inverted sphere. Quantisation of fitness values to the nearest 1000 provides discrete levels of artificial neutrality within a hypercube of initial range 10000 in each dimension. The average consists of 100 individual runs.

<table>
<thead>
<tr>
<th>Average fitness of gBest at final iteration (i = 500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSO</td>
</tr>
<tr>
<td>nPSO</td>
</tr>
<tr>
<td>esPSO</td>
</tr>
</tbody>
</table>
5.3 CONCLUSIONS

Analysis of fitness data (graphs 1,3) acquired in this study suggests that the nPSO algorithm consistently outperforms both esPSO and PSO. In all cases, nPSO also appears to maintain the fastest diffusion rate at the final iteration out of all heuristics (graphs 2,4). If it is accepted that neutral networks create the possibility for permanent innovation, wherein the neighbourhood of every possible structure can potentially be accessed through neutral drift, then the speed of network traversal becomes crucial in finding better optima. The nPSO algorithm’s ‘better’ performance upon the three test functions might seem to validate this proposition.

Further study is certainly needed to explain the difference in initial performance for the plate stacking function however (graph 4), as the similar initial diffusion rate between...
heuristics does not appear to be able to explain the observed differences in rate of discovery of better fitness solutions (graph 3). Additional investigation is considered desirable in order to identify the true cause of these performance differences.

This draws attention now to the esPSO algorithm, which was theoretically implied to have the fastest exploration rate (indeed this was the purpose of Stewarts’ original Extrema Selection model for the GA [23]). Apparently this was not so during experimentation, with the esPSO diffusion data suggesting that this algorithm was not significantly better than nPSO (graph 4) and perhaps actually worse (graph 2) than nPSO in terms of the rapidity of spatial exploration. The simpler heuristic proves better in optimising the neutral test bed introduced herein; a solid explanation as to why certainly requires more detailed analysis of swarm dynamics during optimisation performance.

Graphs 5 and 6 serve to demonstrate that even on non-neutral terrains, both modified heuristics introduced herein certifiably do not impair optimisation performance upon standard, explicitly non-neutral or rugged terrains. Considering the temporal requirements in processing multiple program instructions, the nPSO algorithm is however considered better than esPSO, due to its simpler methodology containing fewer program instructions; over a lengthy run, nPSO is likely to prove the faster of the modified algorithms in terms of process time.

This study then suggests that for terrains suspected or confirmed to contain some degree of neutrality, the nPSO algorithm currently offers a simple modification to the standard that a) does not detrimentally affect optimisation of non-neutral landscapes and b) offers improved performance upon terrains that include neutrality.

6. BIBLIOGRAPHY


