Steep Gradients as a Predictor of PSO Failure

Katherine M. Malan
Department of Computer Science
University of Pretoria
South Africa
kmalan@cs.up.ac.za

Andries P. Engelbrecht
Department of Computer Science
University of Pretoria
South Africa
engel@cs.up.ac.za

ABSTRACT

There are many features of optimisation problems that can influence the difficulty for search algorithms. This paper investigates the steepness of gradients in a fitness landscape as an additional feature that can be linked to difficulty for particle swarm optimisation (PSO) algorithms. The performances of different variations of PSO algorithms on a range of benchmark problems are considered against average estimations of gradients based on random walks. Results show that all variations of PSO failed to solve problems with estimated steep gradients in higher dimensions.

Categories and Subject Descriptors

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

Keywords

Fitness landscape analysis; gradient estimation

1. INTRODUCTION

For decades, researchers have tried to find an answer the question: what make an optimisation problem hard to solve? Many have recognised the infeasibility of finding a single measure of hardness and are instead focussing on using multiple features together to capture problem difficulty for a given algorithm. Multimodality clearly affects problem difficulty, but it is the relative size of basins of attraction (local vs. global) that is more directly related to difficulty for PSOs [5]. Given an unknown optimisation problem, there is no computationally cheap way of determining the relative sizes of basins of attraction. Instead, approximate measures based on samples from the search space must be used to estimate features of a landscape. Ruggedness is a common feature that is estimated and refers to whether there are variations in neighbouring fitness values or not. In contrast, the steepness of gradients takes into account the magnitude of fitness changes of neighbouring points. The motivation for considering gradients is that a landscape with steep gradients should have a higher probability of being deceptive to PSO search, since steeper gradients could lead to narrower basins of attraction. This paper describes a technique for estimating gradients in continuous fitness landscapes, based on random walks.

GECCO'13 Companion, July 6–10, 2013, Amsterdam, The Netherlands.
ACM 978-1-4503-1964-5/13/07.

2. ESTIMATING GRADIENTS

A Manhattan Progressive Random Walk with equal-sized steps is proposed as follows: a walk starts at a random position on the outer edge of the search space; each step is of a predefined step size in a single randomly chosen dimension; the sign of each step is set so that the walk progresses away from the edge where it started. If an opposite edge is reached in any dimension, the sign of that dimension is inverted, so that the walk progresses away from the edge. Given a problem with fitness function f, the gradient in fitness space between two steps t and $t{+}1$ of a walk $\mathbf{x}(t), \mathbf{x}(t{+}1), ..., \mathbf{x}(t{+}T)$ can be estimated by:

$$g(t) = \frac{(f(\mathbf{x}(t+1)) - f(\mathbf{x}(t)))/(f_{max} - f_{min})}{(\mathbf{x}(t+1) - \mathbf{x}(t))/(\mathbf{x}_{max} - \mathbf{x}_{min})}, \qquad (1)$$

where f_{max} and f_{min} are the maximum and minimum fitness values, respectively, as encountered on the walk and $\mathbf{x}_{max}$ and $\mathbf{x}_{min}$ are the position vectors defining the bounds of the search space. Normalising the fitness and distance values allows for comparison of gradient estimations between problems with different fitness ranges and solution domains. A walk of T steps, gives T gradients $g(t), g(t+1), ..., g(t+(T-1))$ and the average estimated gradient within the walk can then be defined as:

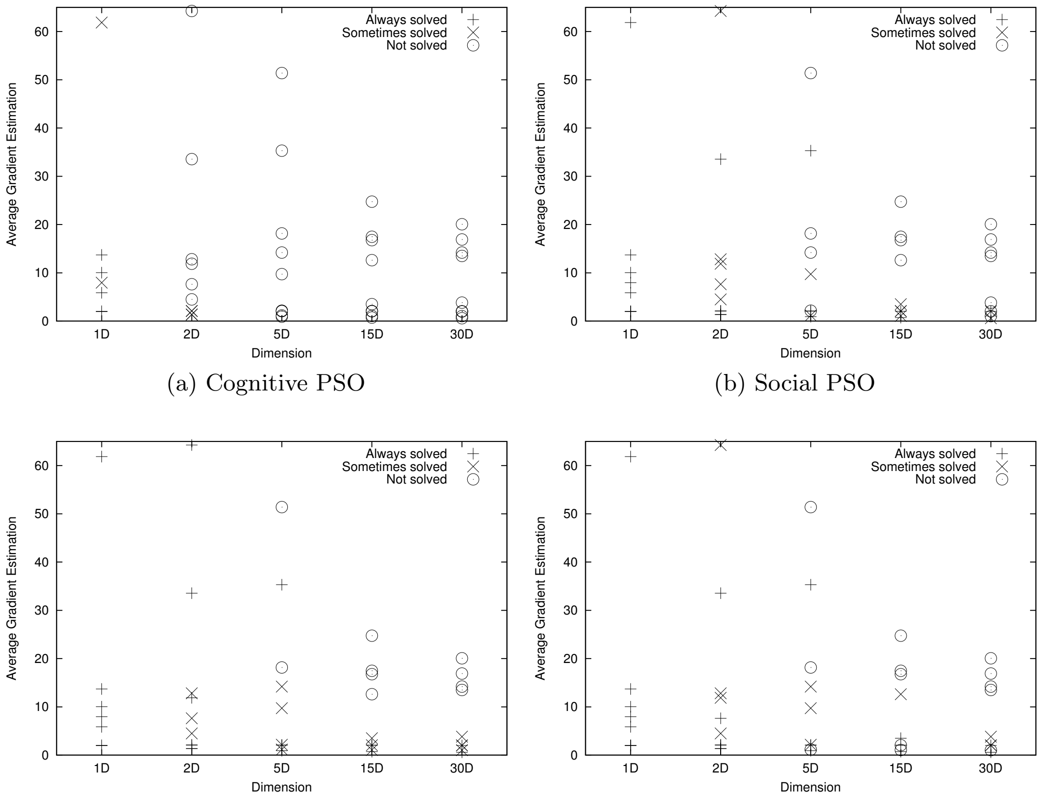
$$G_{avg} = \frac{\sum_{t=0}^{T-1} |g(t)|}{T}. \qquad (2)$$

The absolute value of each $g(t)$ is used, since the purpose is to quantify steepness, regardless of the direction of the slope. If the absolute values were not used, then negative slopes would cancel out positive slopes.

3. EXPERIMENTATION

The G_{avg} measure was calculated for each of the following functions for dimensions 1, 2, 5, 15 and 30: Ackley, Griewank, Quadric, Rana, Rastrigin, Rosenbrock, Salomon, Schwefel 2.26, Spherical and Step (defined in Table 2 of [4]). Thirty independent runs of the algorithm for calculating the G_{avg} measure were performed and mean G_{avg} measures were determined. Each run was based on samples of D Manhattan Progressive Random Walks of 1000 steps each with step size $\frac{(x_{max}-x_{min})*D}{1000}$.

Four different PSO algorithms: Cognitive PSO [2], Social PSO [2], Traditional gbest PSO [1] and Barebones PSO [3] were executed on each function/dimension combination for 30 independent runs. Parameters were as follows: 50 particles, 0.72 inertia weight and 1.496 for the cognitive and social acceleration constants. The SRate metric based on



(a) Cognitive PSO

(b) Social PSO

(c) Traditional PSO

(d) Barebones PSO

Figure 1: Performance of different PSO algorithms on benchmark problems plotted against the G_{avg} measure.

predetermined fixed accuracy levels for each problem and dimension combination (proposed in [4]) was used to measure PSO performance. For all problems, the maximum number of iterations was set to $200 \times D$, where D is the dimension of the problem. SRate is a value in the range $[0, 1]$ where the value indicates the proportion of runs that found the solution. The results are plotted by dimension in Figure 1 with performance discretised into one of the following three groups: *always solved* (SRate = 1), *sometimes solved* ($0 <$ SRate < 1), or *not solved* (SRate = 0).

Figure 1a shows that the Cognitive PSO algorithm only solved 1D and some 2D problems. For the 2D case, the problems with the higher G_{avg} values were not solved. For the other algorithms in Figures 1b to 1d it can be seen that some problems were solved in all dimensions and that for most cases (particularly in 15 and 30 dimensions), the problems with the higher G_{avg} values were not solved. Note that low gradients were not always associated with algorithm success, but high gradients in higher dimensions (15D and 30D) were in most cases associated with algorithm failure.

4. CONCLUSION

This paper has shown that the steepness of gradients is a feature of fitness landscapes that has an influence on problem difficulty for PSO algorithms. Any attempt at predicting PSO performance for unknown optimisation problems, should therefore consider gradients as one of the features with other features that have been linked to problem hardness for PSOs. More techniques are needed for measuring other features of continuous fitness landscapes, so that problems can be more fully characterised and understood.

5. REFERENCES

[1] R. Eberhart and J. Kennedy. A New Optimizer using Particle Swarm Theory. In *Proceedings of the Sixth International Symposium on Micromachine and Human Science*, pages 39–43, 1995.

[2] J. Kennedy. The Particle Swarm: Social Adaptation of Knowledge. In *Proceedings of the IEEE International Conference on Evolutionary Computation*, pages 303–308, April 1997.

[3] J. Kennedy. Bare bones particle swarms. In *Swarm Intelligence Symposium, 2003. SIS '03. Proceedings of the 2003 IEEE*, pages 80–87, 2003.

[4] K. M. Malan and A. P. Engelbrecht. Fitness Landscape Analysis for Metaheuristic Performance Prediction. In H. Richter and A. P. Engelbrecht, editors, *Recent advances in the theory and application of fitness landscapes*, Emergence, Complexity and Computation, page To appear. Springer, 2013.

[5] B. Xin, J. Chen, and F. Pan. Problem difficulty analysis for particle swarm optimization: deception and modality. In *GEC '09: Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*, pages 623–630, New York, NY, USA, 2009. ACM.