Invited paper: A Review of Thresheld Convergence

Stephen Chen
sychen@yorku.ca
York University
James Montgomery
james.montgomery@utas.edu.au
University of Tasmania
Antonio Bolufé-Röhler
bolufe@matcom.uh.cu
Universidad de La Habana
Yasser Gonzalez-Fernandez
ygf@yorku.ca
York University

ABSTRACT

A multi-modal search space can be defined as having multiple attraction basins – each basin has a single local optimum which is reached from all points in that basin when greedy local search is used. Optimization in multi-modal search spaces can then be viewed as a two-phase process. The first phase is exploration in which the most promising attraction basin is identified. The second phase is exploitation in which the best solution (i.e. the local optimum) within the previously identified attraction basin is attained. The goal of thresheld convergence is to improve the performance of search techniques during the first phase of exploration. The effectiveness of thresheld convergence has been demonstrated through applications to existing metaheuristics such as particle swarm optimization and differential evolution, and through the development of novel metaheuristics such as minimum population search and leaders and followers.

KEY WORDS: Exploration, Exploitation, Heuristic Algorithms, Optimization, Multi-modality.

INTRODUCTION

The goal of optimization is to find the best solution in a search space. In a unimodal search space, this task is relatively simple. First, search can be concentrated around the best-known solution since all solutions have a monotonic path to the global optimum, and the best solution should have the shortest path. Second, local exploitative search steps are sufficient since there are no local optima in which the search process can become stalled. Optimization in multi-modal search spaces is more complex, but the foundation of many metaheuristics is still derived from concepts initially developed for unimodal search spaces.

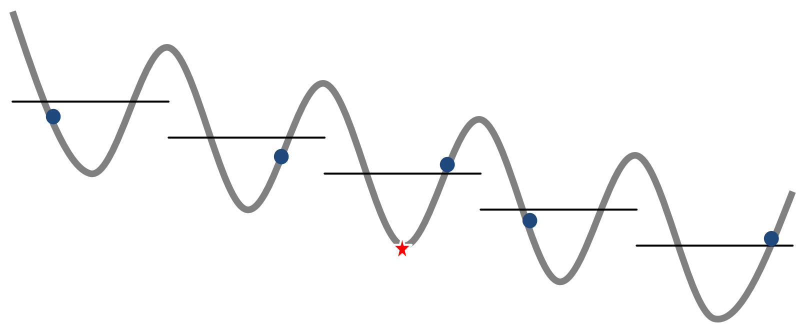
For continuous domain search spaces, particle swarm optimization (PSO) (Kennedy and Eberhart, 1995) and differential evolution (DE) (Storn and Price, 1997) are two of the most popular and well-studied metaheuristics. Both have their search mechanisms initially developed/conceptualized for unimodal search spaces: PSO begins with a cornfield vector (see Section 3.2 in (Kennedy and Eberhart, 1995)) and DE builds its foundation from a simple unimodal cost function (see Figure 1 in (Storn and Price, 1997)). Since exploration is not required in unimodal search spaces, the concept of attraction basins is not initially present in either PSO or DE. Through the consideration of attraction basins, it can be observed that local search/exploitation will only occur in an attraction basin that is represented by a stored solution. Therefore, the fitness of stored sample solutions in PSO and DE determines if the local optima will be found from these attraction basins.

Particle swarm optimization and differential evolution are just two examples of the many metaheuristics that will only perform local search around currently stored solutions. Since current solutions are stored on the basis of their fitness, an implicit assumption is made that the promise of an attraction basin can be estimated by the fitness of a sample solution. The accuracy of this estimate depends on the "relative fitness" of these sample solutions with respect to their local optima. A simple example is if the relative fitness of every sample solution is zero – if each attraction basin is represented by its local optimum, then the best attraction basin is easily identified by the fittest local optimum. Knowing all the local optima is of course a computationally infeasible task akin to exhaustive search.

Search efficiency demands that the promise of an attraction basin will be determined by something other than the fitness of its local optimum. Therefore, we make it explicit that we will use the fitness of a sample solution as an estimate for the promise of its attraction basin. The effectiveness of comparing these sample solutions to identify the most promising attraction basin (i.e. perform exploration) depends on all of the sample solutions having similar relative fitness. In general, metaheuristics do not worry about maintaining similar relative fitness among their solutions, and they instead often allow concurrent exploration and exploitation. Since exploration tends to produce solutions of low relative fitness and exploitation tends to produce solutions of higher relative fitness, concurrent exploration and exploitation will often lead to solutions of highly dissimilar relative fitness.

The negative effects of dissimilar relative fitness can be seen in the example shown in Figure 1. The simple search space used for illustrative purposes consists of attraction basins which all have the same size and shape (e.g. a sinusoid super-imposed over a linear slope). The (black) lines show the average fitness of all of the solutions in each attraction basin. If each attraction basin is represented by a random solution, it can be expected that the fittest of these sample solutions will be from the fittest attraction basin – i.e. the one with the

Figure 1: For a series of random solutions with similar relative fitness (blue circles), the fittest sample solution will often represent the fittest attraction basin. However, if one of these solutions benefits from local optimization (red star), the fittest solution will often no longer represent the fittest attraction basin.



global optimum. However, if an unpromising attraction basin is represented by a sample solution with a much better than random relative fitness (e.g. its local optimum), then this attraction basin can appear to be more promising than it is.

The comparison of sample solutions with highly dissimilar relative fitness occurs frequently and often rapidly in metaheuristics such as PSO and DE. As opposed to distinct phases of exploration and exploitation, PSO and DE both share the following search characteristic. If all of the solutions are in different attraction basins, it is still possible to create a new solution within an existing attraction basin. However, if all of the solutions are in the same attraction basin, it is largely impossible to create a new solution from a different attraction basin. There is a natural transition from the first, mostly explorative phase to the second, highly exploitative phase, and the existence of concurrent exploration and exploitation is a key part of this natural transition.

It is proposed that the effectiveness of metaheuristics in multi-modal search spaces can be improved through explicit separation of the processes for exploration and exploitation. Strict exploration is enforced by ensuring that sample solutions are not drawn from the same attraction basin. This was initially achieved by using a threshold function – in "thresheld convergence", exploitation is restricted and convergence is "held" back through the use of a threshold function. The separation of exploration and exploitation has also been achieved in leaders and followers – a new metaheuristic which uses a novel population control mechanism which supports exploration that is unbiased by previous exploitation.

Thresheld convergence has similar objectives to crowding, niching, and other diversification techniques that are described in the Background. However, thresheld convergence focuses specifically on attraction basins and their sampling, and these concepts are developed in Section III. The first implementations of thresheld convergence are presented in Section IV with current and future work presented in Section V. The alternate development of leaders and followers is presented in Section VI with the explicit integration of thresheld convergence into leaders and followers proposed as part of future work. The paper then finishes with a Summary in Section VII.

BACKGROUND

The ideal case for thresheld convergence is to have one sample solution from each attraction basin, and for each sample solution to have the same relative fitness with respect to its local optimum. The concept of having one sample solution from each attraction basin is a specific example of diversity. Diversity is a concept borrowed from biology, so the field of evolutionary computation is filled with many examples such as crowding, niching, fitness sharing, etc. (Talbi, 2009). Thresheld convergence has many similarities with these techniques, but it also has important differences.

The standard population control mechanism in evolutionary computation is that the fittest solutions survive to create new offspring solutions while the least fit solutions are removed from the population. Crowding methods will instead remove solutions from the population based on their distance to other solutions. For example, instead of comparing a new offspring solution with the least fit member of the population, it could instead be compared with the closest member of the current population. When nearby solutions are replaced, the formation of crowds is avoided and diversity is maintained. (De Jong, 1975)

Crowding has several weaknesses. First, the *n* distance measurements required to find the nearest solution in the current population represents a relatively high computational overhead. Second, crowding does not explicitly attempt to identify attraction basins, so eliminating the nearest solution could eliminate the consideration of a different and highly promising attraction basin. Thresheld convergence addresses the first weakness through the use of a threshold function against which only one distance measurement is required. This threshold function can be matched to the (measured) size of attraction basins which provides an opportunity for thresheld convergence to address the second issue.

Niching and speciation increase diversity by effectively splitting the overall population into a series of sub-populations. Each sub-population exists in its own "niche" with its subset of solutions potentially forming a (sub) species. It is possible for an attraction basin to be a niche, so many aspects of thresheld convergence and niching represent similar ideas. However, thresheld convergence is designed to eventually converge onto a single attraction basin and produce a single final solution. This design may help thresheld convergence to be more efficient than implementations of niching which produce multiple final solutions (e.g. (Brits, Engelbrecht, and Van den Bergh, 2002)) when only one is desired.

Many diversification techniques have also been developed for multi-objective optimization (MOO). Fitness sharing is one of the most popular (Talbi, 2009), and it involves reducing the fitness of similar solutions so that highly similar solutions become less likely to participate in reproduction. The distance metric which measures similarities among solutions and the sharing function which determines how fitness is reduced are both complicated aspects of fitness sharing. Further, the differences between diversification in search space and objective space mean that MOO techniques may not be suitable for single-objective, multi-modal search spaces.

Diversification is also a specific concept in tabu search (Glover and Laguna, 1997) in addition to the general concept that has been described so far. Recent moves (e.g. those used to enter a local optimum) become tabu/disallowed so that an immediate retracing of the search path is not possible (e.g. returning to the most recent local optimum). By disallowing (short) cycles, greater coverage of the search space (i.e. exploration) is achieved. Tabu steps which make more sense in combinatorial/discrete search spaces are an example of the difficulties involved with transferring concepts between continuous and discrete search spaces.

ATTRACTION BASINS

It is useful to think of a multi-modal search space as a search space with many attraction basins, each with a single (local) optimum. It is also useful to define the fitness of an attraction basin as the fitness of this optimum. Optimization in a multi-modal search space can then be thought of as a two-phase process. The first phase of exploration is tasked with finding the fittest attraction basin. The second phase of exploitation will then be tasked with reaching the optimum within the selected attraction basin.

Focusing on the first phase of exploration, the goal of finding the fittest attraction basin is complicated by the difficulty of estimating the fitness of an attraction basin. The accurate measurement of an attraction basin's fitness often requires local optimization to reach the actual local optimum. Since any precise measurement is likely to be excessively expensive computationally, many metaheuristics use the fitness of a current sample solution to estimate the potential fitness of its attraction basin. The accuracy of comparing attraction basins with these estimates depends on having a similar relative fitness for each sample solution being used to make an estimate.

The challenges faced by metaheuristics attempting to estimate the fitness of an attraction basin by the fitness of a single solution are demonstrated through an in-depth study of particle swarm optimization on the Rastrigin benchmark function. For this study, it is useful to think of PSO as a system with two populations: a population of personal best (*pbest*) solutions which represent promising attraction basins for further exploration and exploitation, and a population of current solutions which explore for better solutions. During each iteration, a new current solution is generated and evaluated. This solution is only stored if it is fitter than its corresponding *pbest* solution.

The *pbest* solutions guide the search processes of PSO. When velocities are small, the current solutions will be close to their *pbests*. Thus, the later phases of search in PSO can be described as local optimization within the attraction basins represented by the *pbest* solutions. The early phases of search in PSO (when velocities are larger) can then be thought of as an exploration for the best attraction basins. From this perspective, it should be noted that large velocities mean that current positions can make large movements, but these large movements can still create new current solutions that are close to existing *pbest* solutions. In its standard form, PSO can concurrently perform exploitation during its early exploration phase.

Concurrent exploration and exploitation can interfere with a search technique's processes for exploration. Specifically, when the potential fitness of an attraction basin is estimated by the fitness of a single solution (e.g. a *pbest* solution in PSO around which future exploitation will occur), then the accurate comparison of attraction basins depends on these solutions having similar relative fitness. However, search techniques such as PSO which allow exploitation to occur concurrently with exploitation will often compare new random sample solutions (with poor relative fitness) with locally optimized reference solutions (with better relative fitness). Comparisons of this nature are biased in favour of the stored reference solutions (e.g. *pbests* in PSO), and this bias will reduce the effectiveness of a search procedure's processes of exploration.

The existence and effects of these biased comparisons have been demonstrated by a comprehensive study with the Rastrigin function in (Gonzalez-Fernandez and Chen, in press). The Rastrigin function (in *n* dimensions) is based on a sinusoid super-imposed on a quadratic base function (1). It produces attraction basins of similar size and shape (the ideal case for thresheld convergence) with local optima at known locations (i.e. all the integer values along each coordinate axis). This second feature makes it easy to find the nearest local optimum for any solution in the search space (i.e. round each solution component to its nearest integer value) which makes it possible to quickly measure both the relative fitness of the current solution and the actual fitness of the attraction basin that it represents.

$$f(x) = 10n + \sum_{i=1}^{n} \left[x_i^2 - 10\cos(2\pi x_i) \right] \tag{1}$$

The ability to measure the fitness of each attraction basin makes it possible to observe the effectiveness of a search technique's processes for exploration. During exploration, the goal is to find the fittest attraction basin – not necessarily the fittest solution. Since search techniques are normally unable to measure the fitness of an attraction basin, they often use the fitness of a sample solution as an estimate for the potential fitness of its attraction basin. The study performed in (Gonzalez-Fernandez and Chen, in press) shows how solutions with similar relative fitness can lead to accurate comparisons of attraction basins while solutions with dissimilar relative fitness can lead to inaccurate comparisons.

The study begins by picking a random solution in the x = -3 attraction basin. If this solution is part of the initial swarm in PSO, then it would be stored as a *pbest* solution. The goal of exploration is to find a fitter attraction basin. However, when a solution from a fitter attraction basin is sampled, it does not necessarily replace the stored solution – the solution from the fitter attraction basin must also be fitter than the stored solution (e.g. a *pbest* in PSO). Successful exploration thus has two requirements: first, a sample solution must be drawn from a fitter attraction basin, and second, this sample solution must be fitter than the stored solution which represents the stored attraction basin.

The first requirement for successful exploration depends on how solutions are created. The design of metaheuristics has tended to focus on this aspect (Lones, 2014). The second requirement for successful exploration depends on how solutions are compared. Thresheld convergence focuses on this aspect by identifying the need to compare sample solutions of similar relative fitness. The study in (Gonzalez-Fernandez and Chen, in press) shows how comparing sample solutions of dissimilar relative fitness can greatly reduce the rate of successful exploration.

Starting with 10,000 sample solutions from across the entire search space, only those from fitter attraction basins are kept. Since these kept solutions satisfy the first condition for successful exploration, the study can now focus on the effects of relative fitness for achieving the second condition. The sample solution in the x = -3 attraction is initially at a random point, and then it is moved in ten equal steps towards its local optimum (i.e. x = -3). These steps simulate and demonstrate the effects of exploitation/local search which reduce the relative fitness – i.e. the difference in fitness between a solution and its local optimum. It is shown in Figure 2 how random sample solutions from fitter attraction basins have a high probability (around 75%) of achieving successful exploration when compared against the initial random solution from the x = -3 attraction basin. However, the probability of successful exploration drops rapidly as local search is performed, and it reaches near 0 after as little as 50% local optimization (i.e. moving the original random sample solution half of the distance towards its local optimum).

Figure 2: As a random sample solution is moved closer towards its local optimum, the probability of successful exploration rapidly approaches 0.

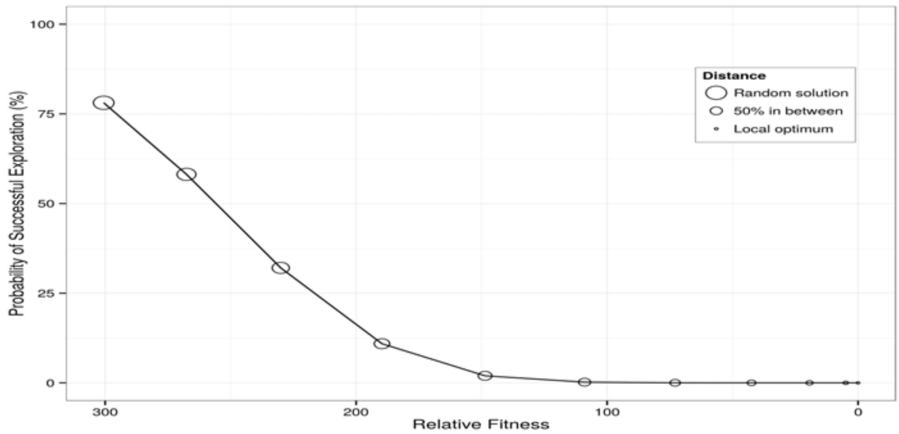
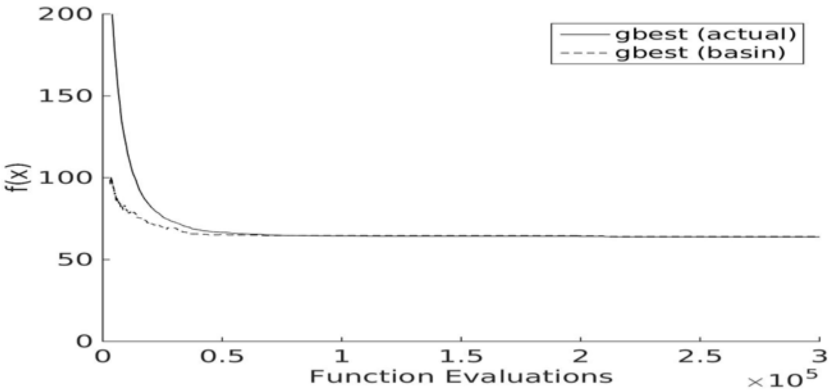


Figure 3: Concurrent exploration and exploitation is shown in PSO. During exploration, improvements to the dashed line occur. However, exploitation also occurs – as indicated by the convergence of the solid and dashed lines.



An implementation of standard PSO (Bratton and Kennedy, 2007) has also been run on the Rastrigin function. In Figure 3, the fitness of the best overall solution (i.e. *gbest*) is plotted along with the fitness of its attraction basin. The difference between the two lines represents the relative fitness. The convergence of the two lines means that the relative fitness has improved and this is indicative of exploitation/local search. Exploration is indicated by improvements to the fitness of the *gbest* attraction basin (dotted line), and the rate of these improvements vanishes rapidly as the relative fitness improves. Overall, these experiments show the negative effects of concurrent exploration and exploitation which thresheld convergence aims to address.

EARLY THRESHELD CONVERGENCE

We view optimization in a multi-modal search space as a two-phase process. The first phase of exploration is tasked with finding the fittest attraction basin, and the second phase of exploitation is tasked with attaining the local optimum within an (existing) attraction basin. Most metaheuristics do not have an explicit demarcation point which separates their processes of exploration and exploitation. Therefore, we will consider any search step which involves two solutions from two different attraction basins as exploration, and any search step which involves two solutions from the same attraction basin as exploitation. The goal of thresheld convergence is thus to achieve two distinct phases: phase one, all search steps/comparisons involve two different attraction basins, and phase two, search steps/comparisons occur within the same attraction basin.

Thresheld convergence was first developed for particle swarm optimization (Chen and Montgomery, 2011) (Chen and Montgomery, 2013). In PSO, the population of personal best (*pbest*) solutions can be viewed as representatives of the fittest attraction basins. To create an explicit first phase of (only) exploration, we initially disallow comparing/updating an existing *pbest* solution with a new solution from its current attraction basin. The determination of whether or not two solutions are in the same attraction basin is based on a threshold value – solutions closer than the threshold are deemed at risk to be in the same attraction basin. Without any specific knowledge of the search space, threshold values were initially based on a simple threshold function (2).

$$threshold = (\alpha * diagonal) * ([FEs - i] / FEs)^{\gamma} \quad (2)$$

The threshold function (2) starts with a set size (parameterized by a as a preset fraction of the search space *diagonal*), and it decays over time (γ specifies the decay rate, and $\gamma = 2$ or 3 often works well (Chen and Montgomery, 2011)) as the current iteration i reaches the maximum allowed number of function evaluations (*FEs*). At some point during this decay, the threshold size should match the size of the attraction basins in the search space. Prior to this point, exploitation/local search cannot occur (concurrently with exploration/global search), so the addition of thresheld convergence to PSO creates an explicit two-phase search process that should be more suitable for multi-modal search spaces. Experiments on benchmark functions (Hansen et al., 2009) show improvements on the targeted multi-modal functions at the cost of reduced performance on unimodal functions (Chen and Montgomery, 2011).

It should be noted that the first work with threshold functions in (Chen and Montgomery, 2011) was not actually an implementation of thresheld convergence. The goal of thresheld convergence is to support the unbiased comparison of attraction basins through the elimination of concurrent exploration and exploitation. In (Chen and Montgomery, 2011), the threshold function was used to ensure that a particle would not update its *pbest* position to form a crowd with its local best (*lbest*) attractor. This early work with threshold functions was actually an attempt to create an efficient mechanism for crowding and niching. The single distance measurement makes thresheld convergence an efficient form of crowding, and the niches created by the threshold distance support diversity. However, the true implementation of thresheld convergence in (Chen and Montgomery, 2013) eliminates all updates to *pbest* within the threshold distance, and not just those which form crowds. Compared to niching techniques which allow concurrent exploration and exploitation, thresheld convergence supports both diversity and more accurate comparisons of the potential fitness of attraction basins.

The development of thresheld convergence for PSO has been mirrored with applications to differential evolution (Montgomery and Chen, 2012) (Bolufé-Röhler et al., 2013). In DE, the population also has "pair-wise elitism" – a new solution is only stored in the population if it is better than the (paired) *target* solution which it would replace. As DE converges, shorter difference vectors will lead to more *new* solutions being created around existing *base* solutions. From this perspective, each solution in the DE population during the early stages of the search process can be viewed as a representative of a promising attraction basin for future exploitation. However, standard DE does not have distinct phases of exploration and exploitation. The existence of concurrent exploration and exploitation in DE is a key instigator of its autocatalytic convergence properties in which local search begets shorter difference vectors which begets more local search (Montgomery, 2009).

Thresheld convergence can be used in DE to halt this autocatalytic (and often premature) convergence by explicitly creating distinct and separate phases for exploration and exploitation. Through the use of the threshold function (2), we can ensure that the *new* solution is not in the same attraction basin as its paired *target* solution. The *new* solution could still be in the same attraction basin as another member of the population, but the threshold function blocks the opportunity for autocatalytic convergence by ensuring that these two solutions cannot directly cause local search which would produce more *new* solutions in attraction basins already represented in the population. Experiments on benchmark functions show improvements on the targeted multi-modal functions at the cost of reduced performance on unimodal functions (Montgomery and Chen, 2012) (Bolufé-Röhler et al., 2013).

Thresheld convergence is based on a single distance measurement, so it can be meaningfully applied to point-search techniques such as simulated annealing (SA) (Kirkpatrick, Gelatt, Jr., and Vecchi, 1983). In general, SA is thought of in terms of a high temperature phase in which worsening moves have a reasonable chance to be accepted and a low temperature phase in which mostly (or exclusively) improving moves are accepted. Equally important is that larger moves in the search space are also allowed during the high temperature phase. However, like PSO, the higher maximum step size is not accompanied by a minimum step size, so small steps which can lead to new solutions being generated in the same attraction basin can occur during the high temperature phase. Since improving moves are always accepted in SA, this can lead to concurrent exploration and exploitation

The normal operation of (simulated) annealing actually benefits from concurrent exploration and exploitation. As a memory-free, single-point (physical) system, it is possible to leave a promising attraction basin and then never return. A highly promising attraction basin that is represented by a fitter solution is less likely to be discarded, so concurrent exploration and exploitation actually plays an important role in helping SA converge upon the fittest available attraction basin. Adding thresheld convergence to a memory-free implementation of simulated annealing leads to a highly random search which does not improve performance (Chen, Xudiera, and Montgomery, 2012).

Adding a memory to simulated annealing allows thresheld convergence to be added. During the initial high temperature phase, a threshold of sufficient size can force each new solution to be in a different attraction basin. Returning to the best-known solution when the temperature/threshold size has decreased sufficiently can then lead to a more distinct exploitation phase. The ability to improve the performance of simulated annealing (a point-search technique) in multi-modal search spaces demonstrates an advantage of thresheld convergence over crowding and niching techniques which require multiple population members to perform distance measurements.

Thresheld convergence can be used to promote/maintain diversity in single-point and population-based metaheuristics. In an attempt to build a new metaheuristic which is more scalable to high-dimensional search spaces, thresheld convergence is an integral component of minimum population search (MPS) (Bolufé-Röhler and Chen, 2013). MPS derives its names from its ability to use the minimum population size of two – a population of size one makes a population-based technique indistinguishable from a point-search technique. These population members can be used to find exploitable gradients in the search space (similar to difference vectors), and thresheld convergence is used to maintain diversity. By being able to maintain diversity with a smaller population size, thresheld convergence helps improve the efficiency of MPS. The relative performance of MPS versus PSO and DE often improves in multi-modal search spaces when a more restricted limit of function evaluations is used (Bolufé-Röhler and Chen, 2013) (Bolufé-Röhler and Chen, 2014).

Overall, initial implementations with particle swarm optimization (Chen and Montgomery, 2013), differential evolution (Bolufé-Röhler et al., 2013), simulated annealing (Chen, Xudiera, and Montgomery, 2012), evolution strategies (Piad-Morffis et al., in press), and minimum population search (Bolufé-Röhler and Chen, 2014) demonstrate that thresheld convergence can lead to improved performance in multi-modal search spaces. However, these improvements are highly dependent on the form/parameters of the threshold function. The goal of the threshold function is to create two distinct phases of exploration and exploitation, but a decay function leads to relatively indistinct phases. Step-wise threshold functions should be more effective than the current (continuous) decay functions.

CURRENT AND FUTURE WORK WITH THRESHELD CONVERGENCE

A key goal of thresheld convergence is to create distinct phases of exploration and exploitation. During the exploration phase, the threshold size should be large enough to ensure that sample solutions are always drawn from different attraction basins. During the exploitation phase, the threshold size should be zero. A continuous decay function is essentially exhaustive search for the proper threshold size to support exploration. However, it will also create a less efficient exploitation phase due to a non-zero threshold size. A step-based threshold function should be more efficient, and specific measurements on the size of attraction basins in the search space should also improve the effectiveness and efficiency of thresheld convergence.

Step-based threshold functions have been introduced for PSO (Gonzalez-Fernandez and Chen, 2014) and DE (Montgomery, Chen, and Gonzalez-Fernandez, 2014). Their biggest advantage is that step-based threshold functions eliminate the a and γ parameters from the decay function (2) which have to be hand tuned for each specific problem to achieve the highest levels of performance. Conversely, step-based threshold functions are highly dependent on the accurate measurement of the size of attraction basins. The initially developed *ad hoc* methods often suffered from imprecise and inaccurate measurements, so the overall results were hindered by inconsistent performance (Gonzalez-Fernandez and Chen, 2014) (Montgomery, Chen, and Gonzalez-Fernandez, 2014).

Current work is focusing on explicitly measuring the size of attraction basins in the search space. Challenges with these measurements include attraction basins having different sizes and shapes, and attraction basins existing within "metabasins" – especially in the context of non-globally convex search spaces. An ideal search strategy with thresheld convergence might be to find the fittest metabasin, then the fittest attraction basin within this metabasin, and then the optimum within this attraction basin. The three-step threshold function would then start at the size of the metabasins, decrease to the size of the internal attraction basins after the best metabasin is found, and then go to zero to allow exploitation after the best attraction basin is found.

"THRESHELD CONVERGENCE" BY POPULATION CONTROL

The goal of thresheld convergence is to improve exploration by supporting more accurate comparisons of the fitness of attraction basins. Ideally, all sample solutions representing attraction basins will have the same relative fitness. By preventing exploitation/local search, sample solutions will be more likely to have an "average" relative fitness. However, a key observation on experiments with Rastrigin (Gonzalez-Fernandez and Chen, in press) is that relative fitness improves over time with thresheld convergence – even when the threshold size is larger than the maximum distance across an attraction basin. This is a result of random search.

The expected fitness of a random sample solution is the average fitness of all of the solutions within an attraction basin. By definition, a sample solution with this average fitness will also have a relative fitness in the 50th percentile. The first sample solution has an expected relative fitness at the 50th percentile, and the second sample solution has a 50% chance of having a better relative fitness. After 100 sample solutions, a sample solution in the 1st percentile can be expected, and it will take another 100 sample solutions to get a 50% chance to find a new solution with a better relative fitness than the best of the first 100 sample solutions. Over time, the best relative fitness will continue to improve, and the effectiveness of exploration in any metaheuristic will decrease the longer each phase of exploration lasts.

Fair, unbiased comparisons among sample solutions require that only solutions of similar relative fitness will be compared against each other. A new metaheuristic is designed to achieve these fair comparisons by creating two populations: a population of "leaders" which represents the best attraction basins and which guides the search, and a population of "followers" in which unbiased exploration can occur. Specifically, leaders affect the creation of new followers, but the survival of followers depends only on comparisons with other followers. Compared to other metaheuristics which focus on how new solutions are created (e.g. attraction vectors in PSO and difference vectors and crossover in DE), the new metaheuristic of "leaders and followers" focuses on how (new) solutions are compared.

Particle swarm optimization can also be viewed as a technique with two populations: a population of *pbests* which guide the search and a population of current solutions which perform exploration and exploitation. Current solutions become *pbest* solutions when they have better fitness, but concurrent exploration and exploitation allows *pbests* to improve their relative fitness which leads to more biased comparisons over time. Followers are only compared with followers which reduces biased comparisons, but unless followers have the opportunity to become leaders, the search trajectory cannot advance. When new leaders and selected, it is important for the existing leaders and followers to have similar fitness to ensure that neither population is favoured over the other. This is achieved by selecting new leaders when the two populations have the same median fitness (Gonzalez-Fernandez and Chen, in press).

The first implementation of leaders and followers has rapid convergence of the median fitness for each population. Each selection of new leaders causes a reset of the followers population, so the effect is a large number of restarts. An additional benefit of these rapid restarts is that the effects of random search on relative fitness have less time to accumulate, so there is more opportunity for the unbiased comparisons which support effective exploration to occur. This effect can be viewed as an implicit implementation of thresheld convergence – leaders and followers achieves the unbiased comparison of attraction basins without the use of a threshold function. The effectiveness of the new population control mechanism is demonstrated through experiments on benchmark functions (Liang et al., 2013) which show that leaders and followers has performance advantages over PSO and DE in multi-modal search spaces (Gonzalez-Fernandez and Chen, in press).

Current work is focusing on explicitly adding thresheld convergence to leaders and followers. In the first implementation (Gonzalez-Fernandez and Chen, in press), there is no minimum step size. Thus, the rapid restarts caused by the population control mechanism is the only feature which supports the unbiased comparison of attraction basins. The addition of a minimum step size during the creation of followers and/or the selection of new leaders can both improve diversity which should lead to improved performance. In general, since leaders and followers focuses on how solutions are compared, any work on how solutions are created (Lones, 2014) represents a promising opportunity for future improvements.

SUMMARY

Memory-free search techniques such as simulated annealing (Kirkpatrick, Gelatt, Jr., and Vecchi, 1983) and early genetic algorithms (with pure, non-elitist generational populations) (Holland, 1992) require concurrent exploration and exploitation. Exploration without a memory means that only the current attraction basins are known and the best-visited attraction basin can be forgotten. To reduce this risk of losing highly promising attraction basins, memory-free search techniques perform concurrent exploration and exploitation. The improved fitness through exploitation increases the odds of survival for these sample solutions and the attraction basins which they represent.

Search techniques with a memory have the opportunity to have an initial phase of pure exploration, but many metaheuristics (e.g. PSO and DE) still allow concurrent exploration and exploitation. This bias to exploitation may improve the performance of a metaheuristic in unimodal search spaces, and many benchmark function sets start with unimodal functions (e.g. (Hansen et al., 2009) (Liang et al., 2013)). However, the concept of "No Free Lunch" (Wolpert and Macready, 1997) suggests that techniques designed to perform well in unimodal search spaces will likely do so by making sacrifices to their performance in multi-modal search spaces.

Thresheld convergence is a search concept designed explicitly for multi-modal search spaces. It provides the ability to efficiently and accurately compare the fitness of attraction basins which is presented as a key factor in the performance of metaheuristics in multi-modal search spaces. In addition to the threshold function used to limit concurrent exploration and exploitation in thresheld convergence, leaders and followers is introduced as a new metaheuristic which uses a novel population control mechanism to avoid the biased comparison of attraction basins. Both techniques lead to improved performance in multi-modal search spaces.

REFERENCES

Bolufé-Röhler, A. and Chen, S. (2013). Minimum population search – Lessons from building a heuristic technique with two population members. IEEE CEC, p. 2061–2068.

Bolufé-Röhler, A. and Chen, S. (2014). Extending minimum population search towards large scale global optimization. IEEE CEC, p. 845–852.

Bolufé-Röhler, A., Estévez-Velarde, S., Piad-Morffis, A., Chen, S. and Montgomery, J. (2013). Differential evolution with thresheld convergence. IEEE CEC, p. 40–47.

Bratton, D. and Kennedy, J. (2007). Defining a standard for particle swarm optimization. IEEE SIS, p. 120–127.

Brits, R., Engelbrecht, A. P. and Van den Bergh, F. (2002). A niching particle swarm optimizer. SEAL, p. 692–696.

Chen, S. and Montgomery, J. (2011). A simple strategy to maintain diversity and reduce crowding in particle swarm optimization. Australasian AI, p. 281–290.

Chen, S. and Montgomery, J. (2013). Particle swarm optimization with thresheld convergence. IEEE CEC, p. 510–516.

Chen, S., Xudiera, C. and Montgomery, J. (2012). Simulated annealing with thresheld convergence. IEEE CEC, p. 1946–1952.

De Jong, K. A. (1975). An analysis of the behavior of a class of genetic adaptive systems, PhD thesis. Dept. of Computer and Communication Sciences, University of Michigan.

Glover, F. and Laguna, M. (1997). Tabu Search. Kluwer Academic Publishers.

Gonzalez-Fernandez, Y. and Chen, S. (2014). Identifying and exploiting the scale of a search space in particle swarm optimization. GECCO, p. 17–24.

Gonzalez-Fernandez, Y. and Chen, S. Leaders and followers – A new metaheuristic to avoid the bias of accumulated information. IEEE CEC, in press.

Hansen, N., Finck, S., Ros, R. and Auger, A. (2009). Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. INRIA Technical Report RR-6829.

Holland, J. (1992). Adaptation in Natural and Artificial Systems. MIT Press.

Kennedy, J. and Eberhart, R. C. (1995). Particle swarm optimization. IEEE ICNN, p. 1942–1948.

Kirkpatrick, S., Gelatt, Jr., C. D. and Vecchi, M. P. (1983). Optimization by simulated annealing. Science, Vol. 220, May, p. 671–680.

Liang, J. J., Qu, B. Y., Suganthan, P. N. and Hernández-Díaz, A. G. (2013). Problem definitions and evaluation criteria for the CEC 2013 special session on real-parameter optimization. Technical Report 201212, Zhengzhou University, China and Nanyang Technological University, Singapore.

Lones, A. (2014). Metaheuristics in nature-inspired algorithms. GECCO companion, p. 1419–1422.

Montgomery, J. (2009). Differential evolution: Difference vectors and movement in solution space. IEEE CEC, p. 2833–2840.

Montgomery, J. and Chen, S. (2012). A simple strategy for maintaining diversity and reducing crowding in differential evolution. IEEE CEC, p. 2692–2699.

Montgomery, J., Chen, S. and Gonzalez-Fernandez, Y. (2014). Identifying and exploiting the scale of a search space in differential evolution. IEEE CEC, p. 1427–1434.

Piad-Morffis, A., Estévez-Velarde, S., Bolufé-Röhler, A., Montgomery, J. and Chen, S. Evolution strategies with thresheld convergence. IEEE CEC, in press.

Storn, R. and Price, H. (1997). Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces. J. Global Optimization, Vol. 11. Dec., p. 341–359.

Talbi, G. (2009). Metaheuristics: From design to implementation. Wiley.

Wolpert, D. H. and Macready W. G. (1997). No free lunch theorems for optimization. IEEE TEC, Vol. 1, Apr., p. 67–82.