Leaders and Followers – A New Metaheuristic to Avoid the Bias of Accumulated Information

Yasser Gonzalez-Fernandez
School of Information Technology
York University
Toronto, Canada
ygf@yorku.ca

Stephen Chen
School of Information Technology
York University
Toronto, Canada
sychen@yorku.ca

Abstract—**Finding good solutions on multi-modal optimization problems depends mainly on the efficacy of exploration. However, many search techniques applied to multi-modal problems were initially conceptualized with unimodal functions in mind, prioritizing exploitation over exploration. In this paper, we perform a study on the efficacy of exploration under random sampling, which leads to the identification of an important comparison bias that occurs when a solution which has benefited from local search is compared to the first (random) solution in a new search area. With the goal of eliminating this bias and improving the efficacy of exploration, we have developed a new search technique explicitly designed for multi-modal search spaces. "Leaders and Followers" aims to eliminate the negative effects of information accumulation and at the same time use the information from the best solutions in a way that they have controlled influence over the newly-sampled solutions. The proposed metaheuristic outperforms both Particle Swarm Optimization and Differential Evolution across a broad range of multi-modal optimization problems.**

I. Introduction

It is well-known that finding good solutions on multi-modal optimization problems depends—not to a small extent—on the efficacy of exploration and its prevalence over exploitation [1]. Let us define an "attraction basin" as a region of the search space containing all of the solutions that lead to a particular local optimum when (greedy) local search is used. A multi-modal search space is therefore subdivided into as many attraction basins as there are local optima, while a unimodal search space has only one attraction basin. Subsequently, we define the fitness of an attraction basin as the fitness of its local optimum. With these definitions, exploration becomes the task of identifying the fittest attraction basin, and exploitation can be expressed as finding the local optimum within a given basin.

Many popular search techniques which are applied to multi-modal problems were initially conceptualized with unimodal search spaces in mind. For example, Particle Swarm Optimization (PSO) begins with a cornfield vector (see Section 3.2 in [2]) and Differential Evolution (DE) builds its foundation from a simple unimodal cost function (see Fig. 1 in [3]). As a result, their primary search mechanisms are better suited to finding the local optimum within a given attraction basin, as opposed to finding the fittest basin in the entire search space. Although these techniques have undergone many modifications to make them more suitable for multi-modal problems (e.g. DMS-PSO [4] and eADE/nrand/1 [5]), the "no free lunch" theorems for search and optimization [6] suggest that no single heuristic can be expected to perform better than all other heuristics across all problem domains. We expect that heuristics designed explicitly for multi-modal problems can perform better in these domains than metaheuristics initially developed for unimodal search spaces.

In general, obtaining a precise measurement of the fitness of an attraction basin is not possible until local search has been used to identify the actual local optimum. Therefore, search techniques often estimate the fitness of an attraction basin based on the fitness of the solutions sampled from that basin, and this estimation is inevitably affected by the consistency of the sampling process. We have performed a study on the efficacy of exploration under random sampling, which leads to the identification of an important comparison bias that occurs when a solution which has benefited from local search is compared to the first sample solution in a new attraction basin. Further experiments show that the same comparison bias may be faced by many techniques that generate new solutions using some form of random sampling, including PSO and DE. Aiming to achieve a more effective exploration through the elimination of the aforementioned comparison bias, we have developed a new search technique explicitly designed for multi-modal search spaces. The proposed metaheuristic is based on performing very frequent restarts and at the same time maintaining information about the best solutions in a way that they have controlled influence over newly-sampled solutions.

The remainder of this paper is organized as follows. Section II presents a group of experiments that illustrate in detail our motivation for this research and the introduction of the new metaheuristic. Section III then describes the proposed algorithm and presents a preliminary study of its performance on a representative optimization problem. These preliminary results are extended later in Section IV with an empirical study on a widely-used standardized benchmark set. Finally, Section V discusses the contributions of the proposed metaheuristic, and Section VI summarizes and concludes the paper.

II. Motivation

Rastrigin's function (hereafter simply Rastrigin) is a well-known multi-modal problem frequently used to assess the performance of numerical optimization algorithms. As presented

978-1-4799-7492-4/15/$31.00

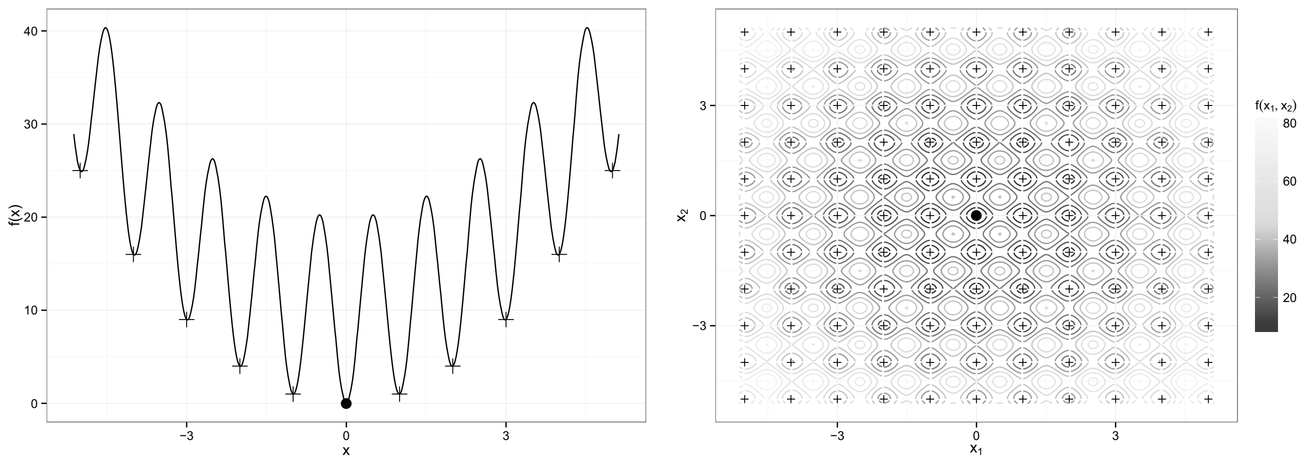

Fig. 1: Rastrigin's function as defined in Eq. 1 in one dimension (left) and two dimensions (right). The plus signs and the filled circle indicate the location of the local optima and global optimum, respectively.

in [7], this function is a minimization problem given by

$$f(\mathbf{x}) = An + \sum_{i=1}^{n} \left[x_i^2 - A\cos(2\pi x_i) \right], \qquad (1)$$

where $A = 10$ and $x_i \in [-5.12, 5.12]$. Rastrigin has a single global optimum at $\mathbf{x} = \mathbf{0}$ with $f(\mathbf{x}) = 0$, but it contains a high number of local optima evenly distributed across the entire search space (see Fig. 1 for illustrative plots in one and two dimensions).

The optima in Rastrigin are placed at the integer coordinates of a regular grid of size one, which means that the function has 11^n optima within the search boundaries defined above. Given a solution in this well-structured search space, the closest (local) optimum in the attraction basin can be easily determined by rounding each component to its nearest integer. This information allows us to present the motivation for conducting this research with the following experiment.

The difference between the fitness of a solution and its local optimum gives us a measure of the "height" of a solution in its attraction basin. We perform a simulation to show how the deeper a solution gets in its attraction basin, the more difficult it becomes to move to a better basin (by finding a better solution in it). Let us consider one particular attraction basin in the search space to demonstrate this idea: the basin centered at the local optimum $\mathbf{x} = -\mathbf{3}$. This basin has a local optimum with a medium fitness in the search space—as measured from the global optimum—and it represents an average-case scenario.

The simulation begins by selecting uniformly one random solution $\mathbf{x}'$ from the $\mathbf{x} = -\mathbf{3}$ attraction basin and recording its height $f(\mathbf{x}') - f(\mathbf{x})$. Next, we proceed to sample a total of 10,000 solutions using a uniform distribution across the entire search space. Among the samples that are in an attraction basin with a local optimum fitter than $f(\mathbf{x})$, we measure the ratio of better/worse solutions compared to the solution $\mathbf{x}'$. The focus on samples from better basins reflects our interest in measuring the efficacy of exploration, which has the goal of finding attraction basins with better local optima.

After we have computed this ratio for the initial random solution $\mathbf{x}'$, we proceed to take ten evenly spaced steps on the line between the solution and the local optimum in the same basin—i.e. simulating a progressive local optimization. After each step, we record the height of the new solution and simulate 10,000 more exploratory solutions, measuring again the ratio of better/worse solutions that are in better basins. This experiment was conducted in $n = 30$ dimensions, and the data shown in Fig. 2 corresponds to the mean of 30 repetitions.

Fig. 2 shows how the efficacy of exploration changes as a random solution approaches its local optimum (in the $\mathbf{x} = -\mathbf{3}$ basin). Initially the sampling ratio is favorable: roughly 75% of the random samples from better basins are better than the initial solution. Since many optimization algorithms employ elitism, whereby only improving solutions are allowed to replace existing solutions, this ability to find better solutions from better attraction basins is critical for effective exploration.

Exploration becomes less effective as the solution is locally optimized. When the height reaches 250 (having moved only 15% of the distance towards its local optimum), there are more worse samples than better samples even though all of these solutions are from attraction basins with fitter local optima than $\mathbf{x} = -\mathbf{3}$. This trend continues and for height values below 150 (i.e. moving a random solution about 40% of the distance towards its nearest local optimum), the ratio of better solutions is very close to zero. Under this situation, most sampled solutions are worse than the locally optimized solution and an optimization algorithm would effectively reject any possibility of moving to a better region of the search space.

This experiment identifies a problem that may be faced by many metaheuristics that generate new solutions using some form of random sampling. In the previous simulation, the initial solution under local optimization represents the best solutions found so far, and stored in one way or another by an optimization algorithm. The 10,000 sampled solutions, on the other hand, represent the newly generated solutions that explore the search space. If the new solutions are compared with the best—and likely locally optimized—solutions, the assessment of the promise of their attraction basins is biased towards the known attraction basins containing the existing best solutions.

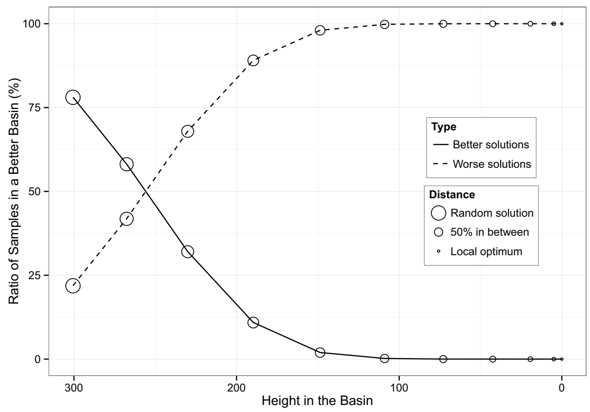

Fig. 2: Efficacy of exploration as a random solution approaches its local optimum: the deeper it gets in the basin, the less likely that it can move to a better basin by finding a better solution.

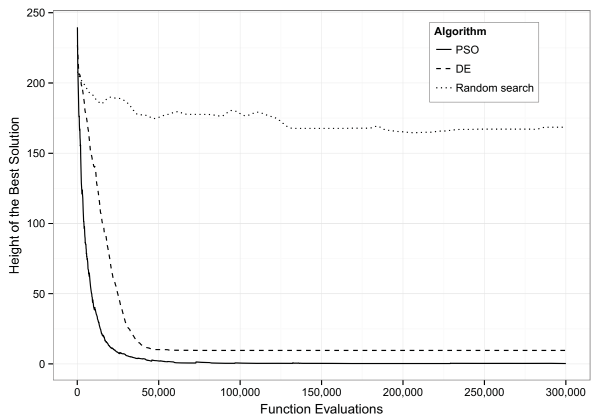

Fig. 3: Height in the basin of the best solution found so far in terms of the function evaluations on Rastrigin's function in 30 dimensions. The best solutions move deeper into their respective attraction basins as the search progresses.

In the interest of illustrating how this problem affects popular metaheuristics, we conduct an experiment with PSO and DE. PSO maintains a population of $pbests$ with the best-known position of each particle in the swarm. In order to redirect the search to an attraction basin not covered by the $pbests$, a sample from the new basin that is better than the existing $pbest$ representative must be found. Likewise, DE keeps a population of the best solutions found so far. Identifying a new more-promising basin in DE depends on generating a $trial$ in the basin that is better than the considered $target$ solution already in the population.

In this new experiment, we aim to show that the best solutions in PSO and DE move deeper into their respective attraction basins as the search progresses. Fig. 3 shows the height in its attraction basin of the best solution as the metaheuristics operate on Rastrigin in 30 dimensions. The data corresponds to the average of 30 independent simulations for PSO and DE. The implementation of standard PSO [8] is described more fully in [9], and the DE/rand/1/bin [10] implementation is described in detail in [11].

At the beginning of the search, the best solutions are the direct result of uniform random sampling and their height is close to 300 on average. Thereafter, this value declines very quickly. After the first 25,000 function evaluations, the height of the best solutions in both PSO and DE has already decayed below 50. If we look for this value on the x-axis of Fig. 2, we find that a solution which has been locally optimized to this extent practically eliminates the chances of randomly finding a better solution from a better basin. Of course, this comparison has some limitations (e.g. new solutions in PSO and DE are not generated uniformly), but it gives us an idea of the difficulties faced by the sampling/exploration processes.

The overall best solutions are generally closer to their local optima than newly sampled solutions, and this imposes an undesirable bias against the evaluation of new attraction basins. This situation is fostered by the (greedy) use of the best solutions to generate new solutions, but it also occurs simply due to the elitist selection process of keeping the best solutions found so far. In Fig. 3, we have also included the results of a pure random search algorithm (see e.g. [12]). In this case, the height in the basin of the best solution drifts to a value below 200, which according to Fig. 2 already biases the evaluation considerably. So, a fair, unbiased comparison of attraction basins is only achieved at the very early stages of the search process—even for random search.

These experiments which illustrate our research motivation have focused on Rastrigin, but we believe that the main ideas can be extrapolated to more general multi-modal search spaces. In the next section, we introduce a new search technique that attempts to avoid the shortcomings highlighted above. The primary design consideration of the new metaheuristic is to avoid the biased comparisons of attraction basins.

III. Leaders and Followers

From the results of the experiments reported in the previous section, we can identify two principles for the effective exploration of a multi-modal search space:

- First, the direct comparison of (random) exploratory sample solutions with the best-known solutions should be avoided.
- Second, an unbiased comparison of attraction basins is most easily achieved at the earliest stages of the search process.

The main argument supporting the first principle is that the current best solutions will generally be close to their local optima. Given that the first sample solution from an attraction basin will rarely be as close to its local optimum, this comparison will be biased towards the best-known solutions. As a result, better attraction basins could be rejected because of being represented by worse-quality solutions. The second principle is motivated by the observation that the height in attraction basins decays rapidly for even random search. The conditions for the accurate comparison of attraction basins will only occur at the very beginning of the search process unless the accumulated information can be isolated.

A. Design of the New Metaheuristic

The proposed technique "Leaders and Followers" uses two separate populations, one denoted as "leaders" and the other

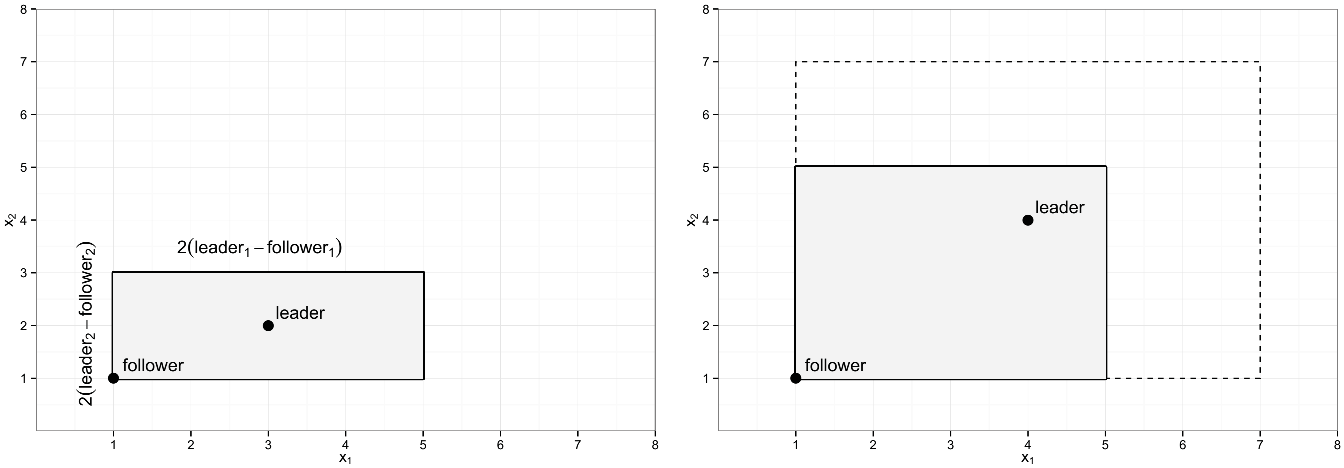

Fig. 4: Generation of a new random solution in Leaders and Followers using Eq. 2 in two dimensions. The shaded grey area represents the uniform sampling region when boundary effects are not present (left) and limited at the search space boundary in each dimension (right). Note that this limiting is different from clamping, which would map the shown area onto the boundary of the search space.

as "followers". The leaders keep track of the best solutions found by the algorithm, while the followers are tasked with the search for new solutions. This separation of roles within the two populations allows us to achieve our first design principle. The new solutions generated within the followers are compared with their peers in the followers population, and not with the best overall solutions in the leaders population. The leaders population is updated with solutions from the followers population only after enough search has been performed in the context of the followers.

We have built the Leaders and Followers algorithm around the concept of performing rapid restarts of the followers population. When the solutions are restarted, all search biases introduced by the elitist selection mechanism is reset. The very frequent restarts allow the technique to spend more time searching when the solutions have a high height in their attraction basins, which facilitates a fairer evaluation of them—satisfying the second design principle.

The following pseudo-code presents the minimization of an n-dimensional objective function f using the Leaders and Followers algorithm:

1: $L \leftarrow$ Initialize the leaders with n uniform random vectors.
2: $F \leftarrow$ Initialize the followers with n uniform random vectors.
3: **repeat**
4:     **for** $i \leftarrow 1, n$ **do**
5:         $leader \leftarrow$ Pick a leader from L.
6:         $follower \leftarrow$ Pick a follower from F.
7:         $trial \leftarrow$ create_trial$(leader, follower)$
8:         **if** $f(trial) < f(follower)$ **then**
9:             Substitute $follower$ by $trial$ in F.
10:         **end if**
11:     **end for**
12:     **if** median$(f(F)) <$ median$(f(L))$ **then**
13:         $L \leftarrow$ merge_populations(L, F)
14:         $F \leftarrow$ Reinitialize the followers uniformly.
15:     **end if**
16: **until** The termination criterion is satisfied.

Firstly, both the leaders and followers populations—denoted with L and F, respectively—are initialized with n random solutions sampled uniformly in the search space domain. The main loop of the algorithm starts at line 3 and it is executed as long as the termination criterion is not satisfied. Each iteration of the main loop begins with a round of updates performed to the followers (lines 4–11). A total of n pairs of leaders and followers are selected uniformly from the corresponding populations. For each $leader$ and $follower$ pair, a new $trial$ is generated by calling the create_trial procedure. If the new trial is better than the follower, it will replace it in the followers population. Notice that we establish a competition with the follower, not the leader (one of the best solutions found so far).

The create_trial procedure on line 7 implements the following formula for every dimension $i = 1, \ldots, n$ of the new solution $trial$:

$$trial_i = follower_i + \varepsilon_i 2\left(leader_i - follower_i\right). \qquad (2)$$

In this formula, ε_i is a uniform random number in $(0, 1)$ sampled independently for every dimension. Therefore, the new solutions are sampled in a hyper-rectangle centred around the selected leader. The leader defines the location of the hyper-rectangle in the search space, while the sampling range is determined by the distances to the follower in each dimension. If necessary, the extension of the hyper-rectangle on each dimension $2\left(leader_i - follower_i\right)$ can be limited to satisfy the box-constraints of the variables. Compared to clamping, which moves a solution beyond the search range back to the boundary, limiting the sampling range in advance avoids over-weighting the boundary of the search space. Fig. 4 illustrates the generation of a new solution.

In the second part of the main loop (lines 12–15), it is checked if enough search has been performed in the context of the followers. If the median fitness of the followers is better than the median fitness of the leaders, we say that both populations are comparable and it is safe to merge them—in terms of the risk of incurring biased comparisons.

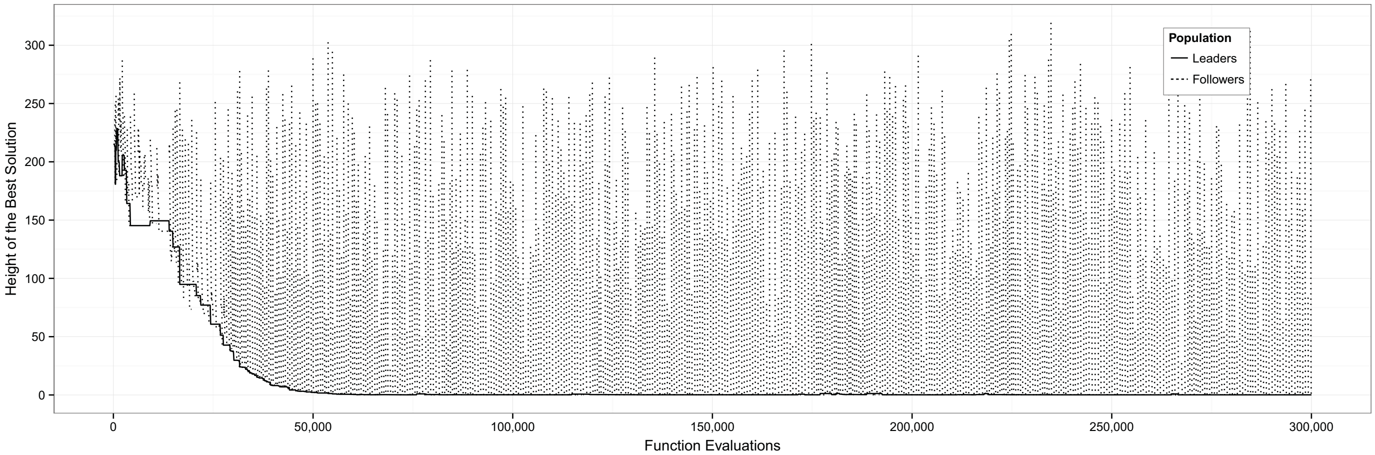

Fig. 5: Height in the basin of the best solution in the two populations of Leaders and Followers in terms of the function evaluations on Rastrigin's function in 30 dimensions. Many samples occur above the 200 height mark, where the sampling rate for finding better solutions in better basins is more favourable.

The merge_populations function on line 13 takes the two populations and returns a group of n leaders selected from the joint group of solutions. We select the new leaders using a combination of hand-picking the best solution and performing binary tournament selection without replacement for the other $n-1$ solutions. After the leaders are updated, the followers are restarted as uniformly sampled solutions. This elimination of accumulated information in the followers population reintroduces the opportunity for unbiased comparisons—see principle two.

The effect of these frequent restarts is illustrated in the following example of Leaders and Followers on Rastrigin that reports the height of the best solutions in both populations. Fig. 5 shows the height of the best solutions in terms of the number of function evaluations. We have included the data corresponding to only one run of the algorithm to make the individual restarts clearly visible. Otherwise, the variable timing of the restarts would cause the height of the followers to average into a flat line.

The key result that stands out from the plot is the large number of restarts performed by the algorithm. The criterion used to check whether the populations can be merged is effective at detecting that the solutions have similar height. At the beginning of each restart, the height of the solutions in their attraction basins is favourable for the accurate identification of the fittest attraction basins. By repeatedly isolating the accumulated information, the proposed technique can perform more unbiased comparisons. In the remainder of this section, we investigate whether these unbiased comparisons can lead to improved exploration and a better performance on Rastrigin, compared with PSO and DE.

B. Preliminary Results on Rastrigin

We present some numerical results on Rastrigin, before continuing with a more extensive computational study of the proposed technique in the next section. The performance of Leaders and Followers is compared with the results obtained by PSO and DE. The three algorithms were run on Rastrigin in 30 dimensions, recording the changes in the fitness of the best solution up to a maximum of 300,000 function evaluations. Fig. 6 illustrates the average result of 30 independent trials.

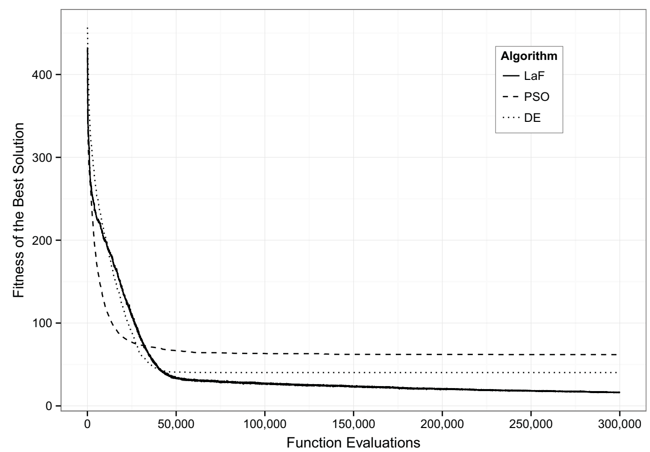

Fig. 6: Mean performance over 30 trials for Leaders and Followers (LaF), PSO, and DE on Rastrigin's function in 30 dimensions.

The results show that Leaders and Followers obtains the best overall results among the three algorithms. The average fitness values after 300,000 function evaluations are 16.2, 61.7, and 40.2 for Leaders and Followers, PSO, and DE, respectively. These values translate into a 73.6% improvement of the new technique versus PSO, and 59.5% versus DE—both statistically significant at the 5% level, according to a performed t-test.

In addition to the final results, it is also interesting to examine the search progress of each technique as represented by the three lines in Fig. 6. The solid line corresponding to Leaders and Followers shows a steep fall during the first 50,000 function evaluations and slower (but continuous) improvements thereafter. These improvements suggest that Leaders and Followers keeps improving the overall best solution until it is stopped. On the other hand, the lines corresponding to

TABLE I: Best fitness values in Leaders and Followers (LaF), PSO, and DE on Rastrigin's function in 30 dimensions.

FEs	LaF		PSO		DE	
	mean	std. dev.	mean	std. dev.	mean	std. dev.
50,000	33.6	9.5	66.7	12.4	40.8	13.4
100,000	27.4	8.5	63.2	10.5	40.2	13.3
150,000	23.4	7.5	62.2	11.6	40.2	13.3
200,000	20.4	6.6	62.0	11.7	40.2	13.3
250,000	18.1	5.0	61.9	11.8	40.2	13.3
300,000	16.2	4.7	61.7	11.9	40.2	13.3

PSO and DE exhibit a similar steep decay at the beginning, but they become mostly flat afterwards (especially the dotted line corresponding to DE). The searches in PSO and DE seem to become practically stalled after the first 50,000 function evaluations.

Table I presents more information in this regard. It reports the mean and standard deviation of the best fitness values reached by the algorithms every 50,000 function evaluations. This information allows us to quantify the performance differences observed in Fig. 6. From 50,000 to the limit of 300,000 function evaluations, the best fitness in PSO improves on average from 66.7 to 61.7 (by 5.0), and in DE from 40.8 to 40.2 (by 0.6). Conversely, in Leaders and Followers the fitness changes from 33.6 to 16.2 (by 17.4) during the same period. This result is evidence that the proposed technique systematically improves the best solution, avoiding stagnation through the opportunity for unbiased comparisons.

The improvements after 50,000 function evaluations for Leaders and Followers can be shown to be the result of continued global search. Measuring the height in the basin of the best solutions reported in Table I, the average height of the best solutions at 50,000 function evaluations is 2.9 and this value becomes 0.1 when the limit of 300,000 evaluations is reached (see Fig. 5 for reference). Although even at 50,000 function evaluations, the best solutions are very deep in the basin, Leaders and Followers is able to progress with the search and find better solutions. Most of these solutions are from better basins, since (local) search within the same basin could have provided at most 2.8 out of the 17.4 fitness improvement obtained after the first 50,000 function evaluations.

The proposed technique has shown a good performance on Rastrigin, which validates its motivation and its underlying design principles. These preliminary results suggest that Leaders and Followers is better equipped than PSO and DE to effectively explore multi-modal search spaces. In the next section, we extend the study of the proposed technique to a widely-used set of 28 functions, including 15 complex multi-modal problems.

IV. Computational Study

This section presents a computational study of the Leaders and Followers metaheuristic introduced in the previous section. As it was done during the experiments with Rastrigin, the performance of the technique is studied considering PSO and DE as baselines. The test problems of the CEC 2013 special session on real-parameter optimization [13] (CEC'13 benchmark for short) are used to evaluate the performance of the algorithms. This benchmark includes 28 minimization problems divided into three groups: unimodal functions (1–5), basic multi-modal functions (6–20), and composition functions (21–28). The second group is the most relevant to this study, given that Leaders and Followers was especially designed for multi-modal problems. All of the functions are considered in 30 dimensions.

The experimental setup follows the directions given in the CEC'13 benchmark. A total of 51 randomized trials with a maximum allocation of 300,000 function evaluations were performed on each function, reporting the mean and standard deviation of the fitness errors. In addition, the relative differences of the mean errors (%-diff) were calculated for contrasting Leaders and Followers versus PSO and DE. The %-diff of the mean error m_1 of a first algorithm (LaF) with respect to the mean m_2 of a second algorithm (PSO or DE) is given by $100(m_2-m_1)/\max(m_2, m_1)$. Hence, positive %-diff values indicate that the first algorithm outperforms the second one. A t-test between the two samples is also reported in order to make the comparison on the basis of statistically significant differences at the 5% level.

Table II summarizes the results of the comparison with PSO on the CEC'13 benchmark. The proposed heuristic obtains an average %-diff improvement of 24.7% on the targeted group of multi-modal functions. Considering the statistical significance of the results on these 15 functions, Leaders and Followers obtains better results on eight of them, worse results on five, and the remaining two are indistinguishable. Most of the (statistically significant) better results represent improvements of over 50%. Positive results were also obtained on the group of composition functions with a 16.1% average performance improvement. The results on the unimodal group are in general comparable—two being better, two worse, and one the same.

The evaluation of Leaders and Followers versus DE is summarized in Table III. The proposed technique also outperforms DE on the multi-modal problems, in this case by 8.6%. Seven of the results on this group are significantly better, three are worse, and the remaining five are equivalent. Additionally, Leaders and Followers obtains better results than DE on the composite functions by 10.9% on average, with most of the statistically significant improvements being of 10% or more. DE, on the other hand, obtains better results on the unimodal problems.

In general, the outcome of these experiments is in line with the results previously obtained on Rastrigin, where Leaders and Followers obtained the best results followed by DE and PSO, in that order. The proposed technique achieved the top overall performance on both the basic multi-modal functions, and the composition functions. The superior performance observed on the original Rastrigin studied in the previous section, was replicated on the more complex—shifted and rotated—variants of this function included in the CEC'13 benchmark (functions 11, 12 and 13). The technique was also able to deal effectively with cases of deceptive local optima and the lack of a globally convergent structure, such as Schwefel's function (functions 14 and 15). The improved exploration also resulted in very good results on the complicated composite functions, being able to handle other sources of complexity, such as disparate function properties around different local optima (functions 22 and 27).

TABLE II: Results of PSO and Leaders and Followers (LaF) on the CEC'13 benchmark.

Func.	PSO		LaF		%-diff	t-test
	mean	std. dev.	mean	std. dev.		
1	0.00e+0	0.00e+0	0.00e+0	0.00e+0	0.0%	
2	2.07e+6	7.82e+5	2.60e+6	1.17e+6	**-20.3%**	0.01
3	6.75e+7	7.06e+7	1.17e+7	2.32e+7	**82.7%**	0.00
4	1.80e+4	4.55e+3	1.32e+4	3.45e+3	**26.4%**	0.00
5	0.00e+0	0.00e+0	4.77e−9	7.07e−9	**-100.0%**	0.00
1–5					-2.2%	
6	1.67e+1	4.34e+	2.98e+1	2.41e+1	**-43.8%**	0.00
7	6.37e+1	1.74e+1	8.58e+0	5.61e+0	**86.5%**	0.00
8	2.09e+1	5.51e−2	2.09e+1	4.82e−2	**-0.2%**	0.00
9	2.86e+1	1.95e+0	1.55e+1	2.98e+0	**45.9%**	0.00
10	1.32e−1	5.53e−2	1.17e−1	5.01e−2	11.1%	0.17
11	6.47e+1	1.51e+1	1.51e+1	4.53e+0	**76.7%**	0.00
12	7.85e+1	1.72e+1	3.56e+1	1.65e+1	**54.7%**	0.00
13	1.44e+2	2.21e+1	8.12e+1	2.81e+1	**43.7%**	0.00
14	2.63e+3	3.80e+2	6.85e+2	2.29e+2	**74.0%**	0.00
15	3.95e+3	6.25e+2	6.38e+3	1.30e+3	**-38.1%**	0.00
16	1.66e+0	3.51e−1	2.47e+0	2.54e−1	**-33.0%**	0.00
17	1.01e+2	1.55e+1	4.86e+1	5.52e+0	**51.9%**	0.00
18	1.70e+2	2.68e+1	2.00e+2	1.07e+1	**-14.9%**	0.00
19	5.76e+0	1.30e+0	2.50e+0	4.34e−1	**56.6%**	0.00
20	1.17e+1	5.12e−1	1.16e+1	4.25e−1	0.5%	0.50
6–20					24.7%	
21	2.20e+2	5.30e+1	2.75e+2	8.87e+1	**-20.2%**	0.00
22	2.96e+3	5.15e+2	5.91e+2	1.81e+2	**80.0%**	0.00
23	4.66e+3	7.06e+2	5.14e+3	2.06e+3	-9.3%	0.09
24	2.76e+2	5.64e+0	2.21e+2	9.76e+0	**19.8%**	0.00
25	2.92e+2	7.12e+0	2.58e+2	1.45e+1	**11.6%**	0.00
26	2.13e+2	4.61e+1	2.09e+2	3.24e+1	1.9%	0.62
27	1.03e+3	7.43e+1	5.67e+2	1.07e+2	**45.2%**	0.00
28	3.00e+2	0.00e+0	3.00e+2	0.00e+0	0.0%	
21–28					16.1%	

TABLE III: Results of DE and Leaders and Followers (LaF) on the CEC'13 benchmark.

Func.	DE		LaF		%-diff	t-test
	mean	std. dev.	mean	std. dev.		
1	0.00e+0	0.00e+0	0.00e+0	0.00e+0	0.0%	
2	1.53e+5	6.89e+4	2.60e+6	1.17e+6	**-94.1%**	0.00
3	1.00e+7	1.28e+7	1.17e+7	2.32e+7	-13.9%	0.66
4	4.35e+2	4.07e+2	1.32e+4	3.45e+3	**-96.7%**	0.00
5	1.18e+0	8.39e+0	4.77e-9	7.07e-9	100.0%	0.32
1–5					-20.9%	
6	2.03e+1	1.57e+1	2.98e+1	2.41e+1	**-31.9%**	0.02
7	5.96e+0	5.91e+0	8.58e+0	5.61e+0	**-30.5%**	0.02
8	2.09e+1	4.89e-2	2.09e+1	4.82e-2	-0.0%	0.35
9	1.60e+1	5.93e+0	1.55e+1	2.98e+0	3.3%	0.56
10	1.31e-1	8.91e-2	1.17e-1	5.01e-2	10.7%	0.32
11	1.55e+1	4.62e+0	1.51e+1	4.53e+0	2.8%	0.62
12	7.64e+1	6.37e+1	3.56e+1	1.65e+1	**53.4%**	0.00
13	1.41e+2	4.54e+1	8.12e+1	2.81e+1	**42.4%**	0.00
14	9.72e+2	3.10e+2	6.85e+2	2.29e+2	**29.5%**	0.00
15	7.21e+3	3.07e+2	6.38e+3	1.30e+3	**11.4%**	0.00
16	2.52e+0	2.89e-1	2.47e+0	2.54e-1	1.9%	0.39
17	5.54e+1	1.04e+1	4.86e+1	5.52e+0	**12.3%**	0.00
18	1.93e+2	1.34e+1	2.00e+2	1.07e+1	**-3.3%**	0.01
19	3.35e+0	2.13e+0	2.50e+0	4.34e-1	**25.2%**	0.01
20	1.18e+1	4.24e-1	1.16e+1	4.25e-1	**1.7%**	0.01
6–20					8.6%	
21	2.84e+2	7.48e+1	2.75e+2	8.87e+1	2.8%	0.67
22	8.11e+2	3.04e+2	5.91e+2	1.81e+2	**27.1%**	0.00
23	6.96e+3	5.47e+2	5.14e+3	2.06e+3	**26.1%**	0.00
24	2.32e+2	1.01e+1	2.21e+2	9.76e+0	**4.7%**	0.00
25	2.56e+2	7.83e+0	2.58e+2	1.45e+1	-0.6%	0.48
26	2.53e+2	6.47e+1	2.09e+2	3.24e+1	**17.3%**	0.00
27	6.29e+2	7.68e+1	5.67e+2	1.07e+2	**10.0%**	0.00
28	3.00e+2	0.00e+0	3.00e+2	0.00e+0	0.0%	
21–28					10.9%	

V. Discussion

With the proliferation of new metaheuristics [14], it is relevant to determine what Leaders and Followers adds to this set. We believe the contribution is quite large. In general, the design of metaheuristics tends to focus on how new search solutions will be created—e.g. neighbourhood search, intermediate search, directional search, etc [15]. Population control and diversification measures are often addressed, but a fundamental characteristic of most metaheuristics is the comparison of new search solutions with existing (best-known) solutions. In unimodal search spaces, these comparisons are not a cause of concern.

PSO (with a star topology [2]) has a "follow the leader" behaviour. Every particle is drawn towards the best-known solution. In a unimodal search space, this greediness is rewarded—the attraction vectors represent gradients in the search space which can lead all of the particles to quickly converge around the global optimum. The use of a ring topology [8] promotes more diversity in the swarm which can improve performance in multi-modal search spaces, but each local best "leader" still uses elitism. There are no negative effects to elitism in a unimodal search space, but it can lead to comparison biases in multi-modal search spaces.

In a unimodal search space, the goal is to get to the global optimum as quickly as possible. In a multi-modal search space, it is usual to first find the most promising region(s) of the search space before switching to a greedier local search to find the best solution within this region. In PSO with a ring topology, each local best position can be seen as an attempt to identify a promising region of the search space. Its action of drawing neighbouring particles towards it will lead to local search which will eventually find the local optimum within the previously-identified attraction basin.

The ability to identify promising regions of the search space (i.e. attraction basins) is an important component of exploring a multi-modal search space. However, with the use of elitism (as in PSO and many other metaheuristics) the promise of a region is represented by the fitness of a solution representing that region—the region will only be abandoned for another region if a better solution from that region is found. Specifically, if a sample solution from a better basin happens to have a worse fitness than the (locally optimized) solution from the current basin, the (new) solution is discarded and future exploitation of its attraction basin will not occur. Therefore, the comparison bias demonstrated in Section II can negatively affect any metaheuristic which directly compares the fitness of two solutions and which uses elitism—PSO and DE as two primary examples.

The new and unique contribution of Leaders and Followers is the use of two distinct populations to reduce the comparison biases associated with elitism. Compared to PSO—which can be viewed as having a population of personal best positions and a population of current positions—Leaders and Followers primarily compares followers with followers whereas PSO directly compares current positions with personal best positions. Since the rest of the Leaders and Followers algorithm is quite trivial (e.g. drawing new solutions uniformly from a hyper-rectangle), it is reasonable to isolate the elimination of biased comparisons as the primary cause of the large performance gains on the targeted multi-modal search spaces.

The initial performance of Leaders and Followers is very promising. The chosen standard for PSO [8] is already highly refined from its original definition [2], and its use of a ring topology specifically improves its performance in multi-modal search spaces. Although PSO has also undergone many additional improvements (e.g. [4], [9]), there is reason to believe that Leaders and Followers can also benefit from similar modifications. Improvements to how Leaders and Followers creates new sample solutions (e.g. [15]) is a promising area for future research.

VI. Summary and Conclusions

Leaders and Followers has been developed considering the challenges faced by sampling-based metaheuristics while searching in multi-modal search spaces. A detailed study of the solutions generated under random sampling in a typical multi-modal problem provided us with new design principles for the effective exploration of a multi-modal search space. These principles dictated the construction of the proposed algorithm, a design process that is distinctively different than attempting to mimic a natural metaphor [14]. The proposed technique in its original definition outperformed both PSO and DE on a close examination performed on Rastrigin, and an extensive computational study performed on the CEC'13 benchmark. These findings confirm that Leaders and Followers constitutes an effective exploration method for multi-modal search spaces.

References

[1] M. Črepinšek, S.-H. Liu, and M. Mernik, "Exploration and exploitation in evolutionary algorithms: A survey," *ACM Computing Surveys*, vol. 45, no. 3, pp. 35:1–35:33, 2013.

[2] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *IEEE International Conference on Neural Networks*, 1995, pp. 1942–1948.

[3] R. Storn and K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.

[4] J.J. Liang and P.N. Suganthan, "Dynamic multi-swarm particle swarm optimizer," in *IEEE Swarm Intelligence Symposium*, 2005, pp. 124–129.

[5] M. G. Epitropakis, X. Li, and E. K. Burke, "A dynamic archive niching differential evolution algorithm for multimodal optimization," in *IEEE Congress on Evolutionary Computation*, 2013, pp. 79–86.

[6] D. Wolpert and W. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.

[7] H. Mühlenbein, M. Schomisch, and J. Born, "The parallel genetic algorithm as function optimizer," *Parallel Computing*, vol. 17, no. 6–7, pp. 619–632, 1991.

[8] D. Bratton and J. Kennedy, "Defining a standard for particle swarm optimization," in *IEEE Swarm Intelligence Symposium*, 2007, pp. 120–127.

[9] Y. Gonzalez-Fernandez and S. Chen, "Identifying and exploiting the scale of a search space in particle swarm optimization," in *Genetic and Evolutionary Computation Conference*, 2014, pp. 17–24.

[10] E. n. Mezura-Montes, J. Velázquez-Reyes, and C. A. Coello Coello, "A comparative study of differential evolution variants for global optimization," in *Genetic and Evolutionary Computation Conference*, 2006, pp. 485–492.

[11] A. Bolufé-Röhler, S. Estévez-Velarde, A. Piad-Morffis, S. Chen, and J. Montgomery, "Differential evolution with thresheld convergence," in *IEEE Congress on Evolutionary Computation*, 2013, pp. 40–47.

[12] A. Zhigljavsky and A. Žilinskas, *Stochastic Global Optimization*, ser. Optimization and Its Applications. Springer, 2008, vol. 9.

[13] J. J. Liang, B. Y. Qu, P. N. Suganthan, and A. G. Hernández-Díaz, "Problem definitions and evaluation criteria for the cec 2013 special session on real-parameter optimization," Zhengzhou University, China and Nanyang Technological University, Singapore, Tech. Rep. 201212, 2013.

[14] K. Sörensen, "Metaheuristics—the metaphor exposed," *International Transactions in Operational Research*, 2013. [Online]. Available: http://dx.doi.org/10.1111/itor.12001

[15] M. A. Lones, "Metaheuristics in nature-inspired algorithms," in *Genetic and Evolutionary Computation Conference*, 2014, pp. 1419–1422.