The **A**rtificial **L**ife and **A**daptive **R**obotics Laboratory
ALAR Technical Report Series

Dual Guidance in Evolutionary Multi-objective Optimization By Localization

Lam T. Bui, Kalyanmoy Deb, Hussein A. Abbass, and Daryl Essam

TR-ALAR-200604009

The Artificial Life and Adaptive Robotics Laboratory
School of Information Technology and Electrical Engineering
University of New South Wales
Northcott Drive, Campbell, Canberra, ACT 2600
Australia
Tel: +62 2 6268 8158 Fax:+61 2 6268 8581

Dual Guidance in Evolutionary Multi-objective Optimization By Localization

Lam T. Bui
School of ITEE.
Univ. College, Univ. of New South Wales
, Canberra ACT 2600
Australia
l.bui@adfa.edu.au

Kalyanmoy Deb
Dept. of Mechanical
Engineering
Indian Institute of
Technology,
Kanpur, PIN 208 016, India
deb@iitk.ac.in

Hussein Abbass and Daryl Essam
School of ITEE.
Univ. College, Univ. of New South Wales
, Canberra ACT 2600
Australia
{abbass,daryl}@cs.adfa.edu.au

Abstract

In this paper, we propose a framework using local models for multi-objective optimization to guide the search heuristic in both the decision and objective spaces. The localization is built using a limited number of adaptive spheres in the decision space. These spheres are usually guided, using some direction information, in the decision space towards the areas with non-dominated solutions. We use a second mechanism to adjust the spheres to specialize on different parts of the Pareto front using the guided dominance technique in the objective space. With this dual guidance, we can easily guide spheres towards different parts of the Pareto front while also exploring the decision space efficiently.

I. Introduction

Evolutionary multi-objective optimization has been applied in numerous domains [4], [5]. Researchers have been investigating theoretically as well as empirically the performance of evolutionary multi-objective optimization algorithms (EMOs) on a wide range of artificial optimization problems from combinatorial, real-valued to dynamic and noisy problems. To date, there exists a number of algorithms, such as VEGA [7], SPEA2 [9], PDE [1] and NSGA-II [5]. These algorithms are still being continuously analyzed, compared, and tested under various problems and criteria.

EMOs (or Evolutionary Algorithms in general) are usually blind search techniques in the sense that they do not usually use any auxiliary functions like derivatives (as in traditional deterministic optimization techniques). To reduce the effect of the "blindness", there are increasing number of attempts to incorporate some guidance techniques into EMOs. Basically, some guidance is employed to direct the search towards promising areas satisfying specific criteria, such as avoiding infeasible areas or of approaching particular parts of the Pareto front. Guidance can be done in either decision or objective spaces. In this paper, we hypothesize that interleaving guidance in both the decision and objective spaces can help to accelerate the search process.

Our proposed idea is to localize the search in the decision space by using the framework of local models [3] that divide the decision search space into a number of local search areas, where each area is seen as a hyper-sphere. In other words, we transform searching in EMOs on the original search space into a sphere-oriented space, where each sphere is running its own version of EMOs. These spheres move following some direction information to improve their local Pareto front. When we apply the guided dominance mechanism [6], they also tend to specialize and move towards different parts of the Pareto Optimal Front (POF).

The remainder of the paper is organized as follows: background information are the methodology and presented in Section II. An experimental study is carried out in section III. The last section is devoted to the conclusion.

II. Background

A. Guided Dominance Approach

The motivation for guided dominance [2] is that in practical problems we usually need a limited number of sample points of the POF, rather than the whole POF. With guided dominance, the dominance relation is determined from the transformed functions Ω of the original objective functions F, in which the points in the POF area of interest dominate all others in the remaining areas of POF. For the details of this approach, readers are referred to [2].

Based on this approach, Deb et al [6] proposed a technique to divide the POF into a number of parts where each part is tracked by a subpopulation. In order to do this, an equivalent number of weighted matrices is defined to transform the original objective functions. In other words, the dominance relation in each population is defined by a separate weighted matrix. With the partition of the POF in the objective space, the search process is easily guided. Note that this approach only works on problems with a convex POF; and also that is the class of problems we are addressing in this paper.

B. Local Models

In local models, the decision search space *S* is divided up into a number of non-overlapping spheres, where each sphere s_i is defined by a pair of a centroid and radius: $S = [s_0, s_1, ..s_n]$ and $s_i = (c_i, r_i)$. Initially, all r_i are set to be the same value *r*. Inside each sphere, points are generated uniformly, except for the restriction that they are kept distant from each other by a predefined distance threshold β.

More formally, let D_A^B be the Euclidean distance between two centroids of arbitrary spheres *A* and *B* and d_i^j be the Euclidean distance between two arbitrary points *i* and *j* inside any arbitrary sphere A where c_A is the centroid of that sphere *A*. The following condition must then hold:

$$D_A^B \geq 2r, d_i^{c_A} \leq r, d_j^{c_A} \leq r, d_i^j \geq \beta \quad (1)$$

To initialize a sphere s_i, we use a spherical coordinate system. Assume $x = (x_1, x_2, ...x_n)$ is a point in the Cartesian coordinate system. We can calculate *x* from the parameters of an equivalent spherical coordinate system, including a radial coordinate *r*, and *n-1* angular coordinates $\alpha_1, \alpha_2 ... \alpha_{n-1}$ as follows:

$$\begin{aligned}
&x_1 = c_1^i + r\cos(\alpha_1)\\
&x_2 = c_2^i + r\sin(\alpha_1)\cos(\alpha_2)\\
&...\\
&x_{n-1} = c_{n-1}^i + r\sin(\alpha_1)...\sin(\alpha_{n-2})\cos(\alpha_{n-1})\\
&x_n = c_n^i + r\sin(\alpha_1)...\sin(\alpha_{n-2})\sin(\alpha_{n-1})
\end{aligned} \quad (2)$$

Therefore, for a point *x* in a sphere, we generate randomly a set of *n-1* angular values (α) and then apply Eq. 2 to calculate Cartesian coordinate values for *x*. Each sphere is run by its own EMO algorithm. Over time, these spheres move and are being guided towards the global Pareto front. The general steps for the framework of local models are as follows:

- **Step 1**: *Define spheres: Number of spheres, Radius for spheres, and Minimal distance between two points*
- **Step 2**: *Calculate the initial positions of the centroids of the spheres, complying with the rules in Eq. 1*
- **Step 3**: *Initialize spheres: using an uniform distribution, while following Eq.* 2 *and complying with the rules in Eq. 1.*
- **Step 4**: *Run one evolutionary cycle with the EMO on each sphere.*
- **Step 5**: *Start the moving operator to move spheres.*
- **Step 6**: *If Stop condition is not met Goto step 4, otherwise Stop the process.*

An issue associated with the above framework is how to balance exploration and exploitation. This model might not have an advantage over the global model in the case of single-modal problems since it heavily focuses on exploration. However, in the case of multi-modal problems, the exploration ability in combination with a suitable adaptation strategy for spheres can help the system to approach the global optima quickly. Obviously, the adaptation strategy for spheres is the central point of the proposed local models, as it defines how to suitably move the spheres (including their speed and direction) and how to adjust the radius of the spheres to be suitable with the current state. We refer the readers to [3] for more details on localization.

C. Guidance of spheres

All spheres are initialized following the conditions in Equation 1. Centroids are then recalculated for all spheres after every round of evolution - a round is completed when all spheres finish one cycle of their own evolutionary process. The new and old centroids are used to determine the direction of improvement. We use PSO-V2 (a version of the local models to guide spheres in the decision space), which in simple terms, applies a weak stochastic pressure to move the spheres towards the global optima. Details of how the direction of improvement is implemented as well as PSO-V2 are given in [3] instead for space limitations.

Note that the direction of improvement in the local models exploits both local and global information. However, in problems such as the ZDTs, the use of global information might cause the spheres to quickly move closely to each other as they are approaching the POF. Thus, searching time might be wasted since the spheres search the same areas of the POF. It seems better to instead guide each sphere to occupy a different part of the POF, or at least reduce as much as possible the overlapping of the POF's parts that are discovered by the spheres.

To implement this idea, we need to divide the POF into a number of parts. Each part is then used to guide a sphere. In this way, we use *a number of global centroids* instead of only one as in PSO-V2. The number of parts, spheres, and global centroids are kept the same. However, we want to use a "*soft*" division, in the sense that the parts are allowed some overlapping, but the overlapping is kept as small as possible. For this, we select the guided dominance approach in [6].

In this approach, each sphere is associated with a POF part the one that it contributes the most non-dominated solutions to, in comparison with other parts. Each POF part is assigned only one sphere. When a sphere needs to be guided by global information, the sphere's centroid is determined from both the new local centroid, which is calculated based on all individuals of the sphere that belong to its associated POF part and the global centroid, which is considered to be the centroid of that POF part.

III. Experiments

In order to validate the proposed method, which we call GUIDED, we carried out a comparative study in which we tested the method on three ZDT problems: ZDT1, ZDT3 and ZDT4 which all have similar convex shapes of their POFs, but different types of difficulties, namely continuous, discontinuous, and multi-modal (See [11] for more details). We selected NSGA-II as the algorithm to run in each sphere of our models. Also, all the results will be analyzed and compared with two other systems, one is built on an equivalent number of sub-populations, which are defined on the global search space as well as NSGA-II itself.

We use the hypervolume ratio [4], [10] to measure the comparative performance of the different techniques. Note that hypervolume is used to indicate both the *closeness and diversity* of the obtained POFs. The reference point for each problem is seen as the worst point in the objective space obtained by all comparing methods.

We initialize the parameters as follows, and apply non-dominated sorting to update the global archive. All cases were tested on 30 separate runs with different random seeds. We used 5 spheres, total population size is 200, maximum global archive size 100, update frequency for each centroid is every 5 generations, crossover rate 0.9 and mutation rate 0.1.

We will compare the guided version with NSGA-II with the same population of 200 individuals and also with NSGA-II with five - possibly initially overlapping - populations, called 5NSGA-II. All models were run with the same number of evaluations in order to make a fair comparison.

A. Results and Discussion

Convergence is one of the most important characteristics of an optimization technique since its main use is to assess the performance of the algorithm. However, the way of looking at convergence of single objective and multiobjective optimizations are quite different [8]. If some measurements of the objective function, with regard to the number of generations, are experimentally considered as an indication for convergence in single objective optimization, it is not a suitable method for multi-objective optimizations since they do not involve the mutual assessment on all objective functions.

Further, the consideration of convergence is not only on how close the obtained POF, at the last generation, is in comparison to the true Pareto optimal front, but also the rate of convergence which is convergence over time. We consider both issues in this section.

For the closeness of the obtained POF, we use the hyper–volume ratio, denoted *H*. However, it should be noted that this measurement is not always possible in practice, since the true POF is not always known. With test problems, we calculate *H* of the last generation for all models, and report the best, median, worst, mean and standard derivation among 30 runs. All the results are reported in Table I.

Prob	**Models**	**Best**	**Median**	**Worst**	**Mean(STD)**
	GUIDED	**0.9998**	**0.9996**	**0.9988**	**0.9995 (0.0002)**
ZDT1	NSGA-II	0.9989	0.9987	0.9984	0.9987 (0.0001)
	5NSGA-II	0.9937	0.9925	0.9900	0.9923 (0.0009)
	GUIDED	**1.0000**	**0.9998**	**0.9993**	**0.9998 (0.0002)**
ZDT3	NSGA-II	0.9989	0.9980	0.9974	0.9981 (0.0003)
	5NSGA-II	0.9924	0.9890	0.9872	0.9895 (0.0013)
	GUIDED	**1.0000**	0.9984	0.9948	0.9982 (0.0015)
ZDT4	NSGA-II	**1.0000**	**0.9997**	**0.9974**	**0.9994 (0.0008)**
	5NSGA-II	0.9994	0.9951	0.9923	0.9954 (0.0016)

TABLE I

VALUES OF THE HYPER-VOLUME RATIO FOR EACH METHOD ON ZDT1, ZDT3, AND ZDT4 (BOLD INDICATES THE BEST RESULTS OBTAINED ON A PROBLEM)

It is obvious that allowing populations to run concurrently without guidance (even on the global scale) does not help to improve the performance of the optimization process. That is the reason for the inferior performance of 5NSGA-II on all problems.

For ZDT1, an easy problem, there is no contradiction between the local and global information; hence GUIDED was able to get closer to the POF and achieve the best overall performance. Moreover, the localization in the objective space resulted in some sort of a division of labor thus allowing the system to smoothly converge to the POF. ZDT3 shows similar behaviour.

However, in the case of ZDT4, the GUIDED approach was not as good as we thought it should be. Despite that one of the runs obtained the best overall hyper-volume ration, the average and overall performance is inferior to NSGA-II. We conjecture that the small population size in each sphere, is the reason for these inferior results, since ZDT4 is highly multi-modal. This point will be analyzed later in this section.

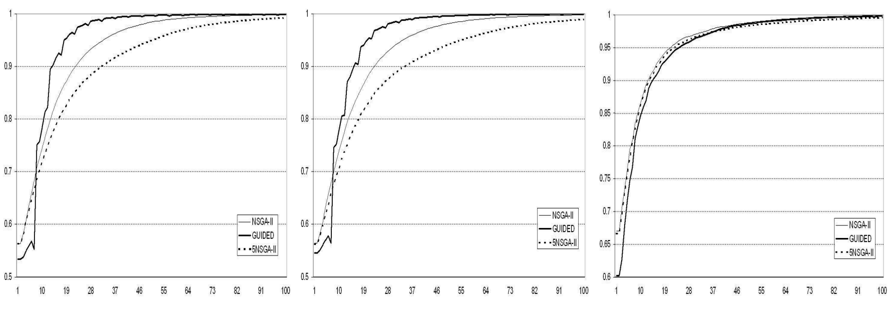

Fig. 1. The hypervolume ratio of differing techniques (up to 20000 evaluations) over time in ZDT1, ZDT3, and ZDT4.

To track the convergence of an EMO, there are several techniques in the literature. However, in this analysis, we have used another simple mechanism for tracking the convergence, by measuring the hypervolume ratio (*H*) over time, since it is consistent with the performance measure used above. We can compare *H* of all models; as an example in Figure 1, we visualize the averaged *H* of 30 runs for all models. It is very clear that in the first few generations, GUIDED is quite slow. This is because localization starts with an exploration phase which consumes time. However, the adaptive strategy of GUIDED helps the method to adjust to move faster if the search space seems smooth enough. It then becomes clearer to see that GUIDED converges very quick to the optimal. However, for ZDT4, it seems that GUIDED gets trapped in the local optima a bit longer than NSGA-II does.

As we hypothesized above, that the small sub-population sizes might have caused the poor performance of the local models, since they were unable to capture the local fitness landscape well enough when the problem is highly multi-modal such as for ZDT4. To test this hypothesis, we increased the population size for each sphere to 100 individuals (500 individuals overall). The hyper–volume ratio that each method achieved over time (up to 50000 evaluations) is visualized in Figure 2.

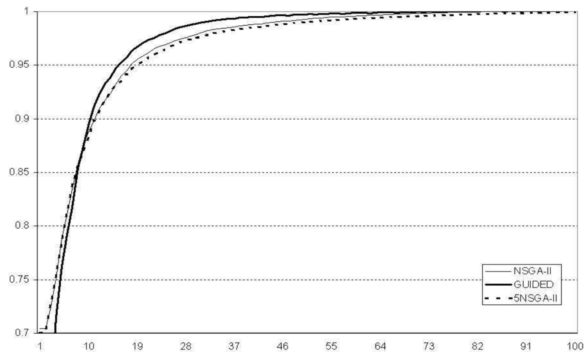

Fig. 2. The hyper–volume ratio of the three techniques over time for ZDT4 with a total population size of 500 individuals and up to 50000 objective evaluations.

It is clear from Figure 2 that GUIDED achieves faster convergence than the other two methods. Again, the pattern is repeated in which GUIDED starts slowly in the exploration phase and increases the speed over time. It is also possible to see a regular series of drops in the curve for GUIDED (also in Figure 1). These drops reflect some sort of loss of diversity in the POF. This sometimes happens when the sphere's motion changes (with the update frequency), and then the sub-population has to adjust to that change. Therefore, this is reflected in the drop-recovery cycles shown in the figure, which seems to be part of the process of first moving to a new better area in the search space, followed by an exploration phase of that area.

All in all, the dual guidance technique shows a good ability to quickly approach the true POF. With the above test problems, it was able to obtain converged and diverse POFs (see Figure 3).

IV. Conclusion

This paper proposed a novel technique (GUIDED) to guide a localized - using hyper–spheres - version of NSGA-II. Each sphere is simultaneously focusing on separate areas of the decision and objective space. The technique was tested against NSGA-II with one population, as well as with multi-populations searching on the global space. The experimental results showed the superior performance of GUIDED in comparison with NSGA-II and 5NSGA-II. For future work, we intend to validate the approach with different schemes of dividing the POF and also with the problems that have more than two objectives.

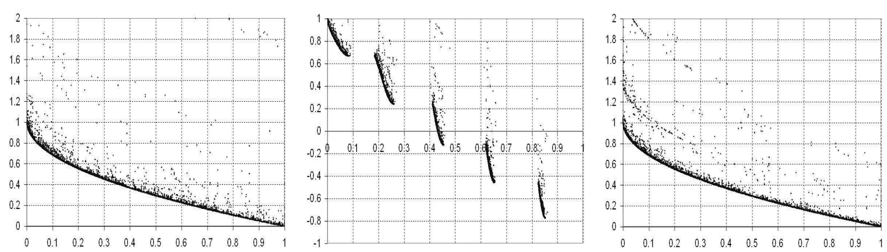

Fig. 3. POFs that GUIDED obtained over time for ZDT1, ZDT3, and ZDT4.

Acknowledgements

This work is supported by the Australian Research Council (ARC) Centre for Complex Systems grant number CEO0348249 and the Postgraduate Research Student Overseas Grant from UNSW@ADFA.

References

[1] H. A. Abbass, R. Sarker, and C. Newton. PDE: A Pareto frontier differential evolution approach for multiobjective optimization problems. In *Proceedings of CEC-2001*, volume 2, pages 971–978. IEEE Press, 2001.
[2] J. Branke, T. Kaufler, and H. Schmeck. Guiding multi-objective evolutionary algorithms towards interesting regions. technical report no. 399. Technical report, Institute AIFB, University of Karlsruhe, Germany, 2000.
[3] L.T. Bui, H.A. Abbass, and D. Essam. Local models: An approach to disibuted multi-objective optimization. technical report no. 200601002. Technical report, ALAR, ITEE,UNSW@ADFA, Australia, 2006.
[4] C. A. C. Coello, D. A. V. Veldhuizen, and G. B. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Publisher, New York, USA, 2002.
[5] K. Deb. *Multiobjective Optimization using Evolutionary Algorithms*. John Wiley and Son Ltd, New York, 2001.
[6] K. Deb, P. Zope, and A. Jain. Distributed computing of pareto optimal solutions using multi-objective evolutionary algorithms. Technical report, No. 2002008, KANGAL, IITK, India, 2002.
[7] J.D. Schaffer. Multiple objective optimization with vector evaluated genetic algorithms. In *Proceedings of the First International Conference on Genetic Algorithms*, pages 93–100, Hillsdale, New Jersey, 1985.
[8] K.C. Tan, T.H. Lee, and E.F. Khor. Evolutionary algorithms with dynamic population size and local exploration for multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 5(6):565–588, 2001.
[9] E. Zitzler, M. Laumanns, and L. Thiele. SPEA2: Improving the strength pareto evolutionary algorithm. Technical report, ETH in Zurich, Swiss, 2001.
[10] E. Zitzler and L. Thiele. Multi-objective optimization using evolutionary algorithms - a comparative case study. In *PPSN*. Springer, 1998.
[11] E. Zitzler, L. Thiele, and K. Deb. Comparision of multiobjective evolutionary algorithms: Emprical results. *Evolutionary Computation*, 8(1):173–195, 2000.