Dimensionality reduction in Many-objective problems combining PCA and Spectral Clustering

Christian von Lücken
Univ. Nacional de Asunción
Facultad Politécnica
San Lorenzo, Paraguay
clucken@pol.una.py

Hugo Monzón
Univ. Nacional de Asunción
Facultad Politécnica
San Lorenzo, Paraguay
hdmonzon@pol.una.py

Carlos Brizuela
CICESE Research Center
Ensenada, México
cbrizuel@cicese.mx

Benjamin Barán
Univ. Nacional de Asunción
Facultad Politécnica
San Lorenzo, Paraguay
bbaran@pol.una.py

ABSTRACT

In general, multi-objective optimization problems (MOPs) with up to three objectives can be solved using multi-objective evolutionary algorithms (MOEAs). However, for MOPs with four or more objectives, current algorithms show some limitations. To address these limitations, dimensionality reduction approaches try to transform the problem by eliminating not essential objectives in such a way that afterward a standard MOEA can be used. To reduce the size of the objective set, Deb and Saxena [3] proposed a method that combines Principal Component Analysis (PCA) with the NSGA-II, called PCA-NSGA-II. Using PCA-NSGA-II as a reference, this work proposes to combine PCA and a clustering procedure for improving the dimensionality reduction process. Experimental runs were conducted with test problems DTLZ2(M) and DTLZ5(I,M) obtaining better results with the proposed method than the obtained with the PCA-NSGA-II.

Categories and Subject Descriptors

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

Keywords

Multi-objective Evolutionary Algorithms; Dimensionality reduction; Spectral clustering; Principal component analysis

1. INTRODUCTION

Several approaches have been proposed to improve current MOEAs to deal with many-objective problems, i.e. MOPs with more than 4 objectives [5]. Among these approaches, dimensionality reduction techniques try to eliminate some redundant objectives, reducing the size of the set of objectives of a many-objective problem to a minimum that retains as much information as possible from the original problem. This way, it is possible to use an existing MOEA to approximate the solution set of the original problem by solving a related problem with a reduced number of objectives.

In this work, based on the PCA-NSGA-II [3], a dimensionality reduction algorithm named PCA-Cluster that combines principal component analysis with a spectral clustering procedure is proposed. Here, clustering is used with the goal of helping the PCA process with the identification and selection of the most relevant conflicting objectives by identifying populations composed of solutions with similar tendencies on the objective space.

The main goal of this work is to validate the proposed algorithm using scalable test problems for which the minimum objective set is known a priori. The algorithm will be considered valid if it is able to determine correctly the minimum set at each instance of the considered problems.

GECCO '15 July 11-15, 2015, Madrid, Spain

ACM ISBN 978-1-4503-3488-4/15/07.
DOI: http://dx.doi.org/10.1145/2739482.2764636

2. PROPOSED METHOD

A major drawback of the PCA-NSGA-II [3] is that the results of PCA strongly depend on its input data. Also, since PCA is a linear method, it may be difficult or impossible for the method to find the minimal set of objectives for some problems [6, 1]. In this work, an approach for improving the PCA-NSGA-II, called PCA-Cluster, is conceived. In this case, a spectral clustering process is executed over the population in objective space to form sub-populations by grouping individuals with similar tendencies to a subset of objectives. Thus, for the PCA it is easier to determine the most important objectives. The method presented in this work splits in two stages. First, a combination of clustering, PCA and NSGA-II is used to obtain a minimal set or a small enough set of objectives, and a population evolves towards different regions of the search space. Second, a modified PCA-NSGA-II uses the reduced objective set of the previous stage as its initial objective set.

Algorithm 1 summarizes the proposed method. For space constraints reason further explanations of the algorithm are not included.

Algorithm 1 PCA-Clustering algorithm

Let $redObjs$ be the initial set of objectives, $redObjs' = \emptyset$
First Stage:
while $\|redObjs\| > dObjs$ **and** $redObjs! = redObjs'$ **do**
  Initialize pop at random
  Evolve pop max_1 gen. using NSGA-II with $redObjs$
  Set iteration counter $t = 1$, $selObjs = \emptyset$
  while $t \leq c$ **do**
    Divide pop in $pops = \{pop_1, \ldots, pop_n\}$
    for p **in** $pops$ **do**
      Using $redObjs$ apply PCA to p to determine the reduced set of objectives $selObjs^p$
      Evolve p max_2 gen. using NSGA-II with $selObjs^p$
      $selObjs = selObjs \cup selObjs^p$
    end for
    $pop = \cup_{i=0}^{n} pop_i$
    $t = t + 1$
  end while
  $redObjs' = redObjs$, $redObjs = selObjs$
end while
Second Stage:
Apply PCA-NSGA-II over pop using $redObjs$

3. EXPERIMENTAL SETUP

Two test problems were considered: DTLZ5(I, M) [6], and DTLZ2(M) [4]. For the DTLZ5(I, M), $I = 2$, and the set composed by (f_M, f_{M-1}) is considered to be the smallest to represent the Pareto Front. Instances of DTLZ2 with 3 and 5 objectives, and DTLZ5(I, M) with 10 and 30 objectives were executed 10 times for PCA-NSGAII, PCA-Cluster and a variant of PCA-Cluster called PCA-Random. The PCA-Random form equal sized sub-populations selecting individuals at random.

The parameters for the NSGA-II were: binary tournament, simulated binary crossover with probability of 0.9 and distribution index of 5, and polynomial mutation with probability of 0.1 and distribution index of 50. To handle the constrains in DTLZ5, the Delta Penalty method [2] was implemented. In all cases, a population size of 400 is used.

For PCA-Cluster, the number of iterations in the first stage is the same that the number of objectives. In the second stage, for DTLZ2 the number of iterations is 400, whereas for DTLZ5 the number of iterations is 250. The parameters for PCA-Random in each problem are the same as those in PCA-Cluster. To execute the same number of function evaluations in both methods, the number of iterations for PCA-NSGA-II for a population size of 400 individuals was: 412, 420, 540, and 620 for DTLZ2(3), DTLZ2(5), DTLZ5(2, 10), and DTLZ5(2, 30), respectively.

4. EXPERIMENTAL RESULTS

Table 1 shows the success rate of the tested algorithms for the considered problems. In test problems DTLZ2(3) and DTLZ5(5), the three algorithms arrived at the correct reduced objective set in all 10 runs declaring all objectives as important. For the DTLZ5(2,10), it can be seen that PCA-NSGA-II arrived at the correct reduced objective set in 4 of 10 executions, while PCA-Cluster has accurately identified this set in all runs. On the other hand, PCA-Random could not find the correct set even once. Finally, for DTLZ5(2,30), PCA-NSGA-II could not identify this set at all, even having trouble in some runs to reduce the size of the set to two objectives. PCA-Cluster, on the contrary, solves this instance arriving at the correct set at each run, while the modified version, PCA-Random, again failed to solve the problem. The failure of PCA-Random to solve even once the instances of problem DTLZ5 suggests that clustering plays an important role in the whole algorithm and in its ability to arrive at the correct reduced set.

Table 1: Success rate of the tested algorithms

Problem	PCA-NSGA-II	PCA-Cluster	PCA-Random
DTLZ2(3)	10/10	10/10	10/10
DTLZ2(5)	10/10	10/10	10/10
DTLZ5(2,10)	4/10	10/10	0/10
DTLZ5(2,30)	0/10	10/10	0/10

5. CONCLUSIONS AND FUTURE WORK

A new method that is capable of correctly identifying the reduced objective set or declare that the problem has no redundant objective is proposed. Results suggest that the method is better than PCA-NSGA-II as the number of objectives increases.

Future work is intended to continue the study of the proposed algorithm with more test problems, and some real world many-objective problems to find the limits of this proposal. Also, the proposal can be further improved if the parameter selection is automated to ease its use in real world situations.

6. REFERENCES

[1] Brockhoff, D., Saxena, D., Deb, K., Zitzler, E.: On handling a large number of objectives a posteriori and during optimization. In: Multiobjective Probl. Solving from Nat. Springer (2008)

[2] Coello Coello, C.A.: Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art. Comp. Methods in Appl. Mechanics and Eng. 191(11) (2002)

[3] Deb, K., Saxena, K.: Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems. In: 2006 IEEE Congress on Evolut. Comp. (CEC'2006). pp. 3353–3360. IEEE (July 2006)

[4] Deb, K., Thiele, L., Laumanns, M., Zitzler, E.: Scalable test problems for evolutionary multiobjective optimization. In: Evolutionary Multiobjective Optim. Adv. Inf. and Knowl. Process., Springer (2005)

[5] von Lücken, C., Barán, B., Brizuela, C.: A survey on multi-objective evolutionary algorithms for many-objective problems. Comput. Optim. and Appl. 58(3), 707–756 (2014)

[6] Saxena, D.K., Deb, K.: Non-linear dimensionality reduction procedures for certain large-dimensional multi-objective optimization problems: Employing correntropy and a novel maximum variance unfolding. In: Proc. of the 4th Int. Conf. on Evol. Multi-Criterion Optim. EMO 2007. Lect. Notes in Comput. Sci., vol. 4403, pp. 772–787. Springer (2007)