DINCON 2017
CONFERÊNCIA BRASILEIRA DE DINÂMICA, CONTROLE E APLICAÇÕES
30 de outubro a 01 de novembro de 2017 – São José do Rio Preto/SP

A New Convergence Measure based on Shannon Entropy for Multi-Objective Optimization Algorithms

Thiago F. Santos[1]

Departamento de Matemtica, Instituto de Cincias Exatas e Biolgicas, UFOP, Ouro Preto, MG

Sebastio Xavier[2]

Departamento de Matemtica, Instituto de Cincias Exatas e Biolgicas, UFOP, Ouro Preto, MG

Resumo. The algorithms of multi-objective optimization had a relative growth in the last years. Thereby, it's requires some way of comparing the results of these. In this sense, performance measures play a key role. In general, it's considered some properties of these algorithms such as capacity, convergence, diversity or convergence-diversity. There are some known measures such as generational distance (GD), inverted generational distance (IGD), hypervolume (S-Metric), Spread(Δ), Averaged Hausdorff distance (Δ_p), R2-indicator, among others. In this paper, we focuses on proposing a new indicator to measure convergence based on the traditional formula for Shannon entropy. The main features about this measure are: 1) It does not require tho know the true Pareto set and 2) Medium computational cost when compared with Hypervolume.

Palavras-chave. Shannon Entropy; Performance Measure; Multi-Objective Optimization Algorithms.

1 Introduction

Nowadays, the evolutionary algorithms (EAs) are used to obtain approximate solutions of multi-objective optimization problems (MOP) and these EAs are called multi-objective evolutionary algorithms (MOEAs). Some these algorithms are very well-known among the community such that NSGA-II (See [1]), SPEA-II (See [2]), MO-PSO (See [3]) and MO-CMA-ES (See [4]). Although the most of MOEAs to use the previous criteria, when $m > 3$ the MOP is called *Many Objectives Optimization Problems* and in this case algorithms Pareto-Based is not good enough. Some papers try to explain the reason why such thing happens (See [5,6]). To avoid the phenomenon caused by Pareto relation, some researchers indicates others way of comparative the elements (See [7]) or change into Non-Pareto-based MOEAs, such as indicator-based and aggregation-based approaches (See [8, 9]).

[1]santostf@iceb.ufop.br, santostf@gmail.com
[2]sermax@iceb.ufop.br

With the growing of use these MOEAs, it's natural to know how theirs outputs are relevant. In this sense, there are metrics that measuring mainly: 1) Closeness to the theoretical Pareto set, 2) Diversity of the solutions and 3) number of Pareto-optimal solutions. This papers is addressed for the first category. Our work is divided as follow. On the section 2 we present a traditional way to multi-objective problem. Next, on section 3, we revised some indicators for performance of evolutionary algorithms. On section 4, we established our proposal for a new indicator. Finally, on section 5, we compared our idea with some indicators well-known.

2 Multi-objective Problem (MOP)

It's common to define multi-objective optimization problems (MOP) as follows:

$$\begin{array}{l} \min f(x), \quad f(x) = (f_1(x), f_2(x), \cdots, f_m(x)) \\ \text{s.t:} \;\; x \in \Omega \subset \mathbb{R}^n \end{array} \tag{1}$$

in which $x \in \mathbb{R}^n$ is the *decision variable vector*, $f(x) \in \mathbb{R}^m$ is the *objective vector*, and $\Omega \subseteq \mathbb{R}^n$ is the *feasible set* that we consider as compact and connected region. In this work, we assumed here that the functions $f_i(\cdot)$ are continuously differentiable (or $\mathcal{C}^2(\Omega)$).

The aim of multi-objective optimization is to obtain an estimate of the set of points belonging to Ω which minimize, in the certain sense that we will call by Pareto-optimality.

Definição 2.1. *Let $u, v \in \mathbb{R}^m$. We say that u dominates v, denoted by $u \preceq v$, iff $\forall i = 1, \ldots, m$*

$$u_i \leq v_i, \text{ and } u \neq v. \tag{2}$$

Definição 2.2. *A feasible solution $x^* \in \Omega$ is a Pareto-optimal solution of problem (1) if there is no $x \in \Omega$ such that $f(x) \preceq f(x^*)$. The set of all Pareto-optimal solutions of problem* (1) *is called by Pareto Set (PS) and the its image is the Pareto Front (PF). Thus,*

$$\begin{array}{rcl} PS & = & \{x \in \Omega | \nexists y \in \Omega, f(y) \leq f(x)\} \\ PF & = & \{f(x) | x \in PS\} \end{array}$$

A classical work (See [15]) established a relationship between the points of PS and gradient informations from the problem (1). That connection it is known by Karush-Kuhn-Tucker (KKT) conditions for Pareto optimality that we define as follow:

Theorem 2.1 (KKT Condition)**.** *Let $x^* \in PS$ of the problem*(1)*. Then, there exists nonnegative scalars $\lambda_i \geq 0$, with $\sum_{i=1}^m \lambda_i = 1$, such that*

$$\sum_{i=1}^{m} \lambda_i \nabla f_i(x) = 0 \tag{3}$$

This theorem will be fundamental to this paper because we will use this fact to formulate our proposal.

3 Some Convergence Indicators

In this section, we will go to relate some metrics or indicators known by scientific community and well-done established. The idea here is to compare those indicators further below with our proposal measure.

3.1 GD/IGD

The Inverted Generational Distance (IGD) indicator has been using since 1998 when it was created. The IGD measure is calculated on objective space, which can be viewed as an approximate distance from the Pareto front to the solution set in the objective space. In the order to define this metric, we assume that the set $\Lambda = \{y_1, y_2, \cdots, y_r\}$ is an approximation of the Pareto front for the problem (1). Let be the V_{MOEA} a solutions set obtained by some MOEA in the objective space as $V_{MOEA} = \{v_1, v_2, \cdots, v_k\}$ where v_i is a point in the objective space. Then, the IGD metric is calculated for the set V_{MOEA} using the reference points Λ as follows:

$$IGD(V_{MOEA}, \Lambda) = \frac{1}{r}\left(\sum_{i=1}^{r} d(y_i, V_{MOEA})^2\right)^{1/2}, \tag{4}$$

where $d(y, X)$ denotes the minimum Euclidean distance between y and the points in X. Besides, we also can to define de GD metric by

$$GD(V_{MOEA}, \Lambda) = IGD(\Lambda, V_{MOEA}) \tag{5}$$

This indicators are measure representing how "far" the approximation front is from the true Pareto front. Lower values of GD/IGD represents a better performance. The only different between GD and IGD is that in the last one you don't miss any part true Pareto set on comparison.

In [10] indicates two main advantages about it: 1) its computational efficiency even many-objective problems and 2) its generality which usually shows the overall quality of an obtained solution set. The authors in [10] studied some difficulties in specifying reference points to calculate the IGD metric.

3.2 Averaged Hausdorff Distance (Δ_p)

This metric combine generalized versions of GD and IGD that we denote by GD_p/IGD_p and defined by, with the same previous notation,

$$\begin{aligned} IGD(V_{MOEA}, \Lambda)_p &= \left(\frac{1}{r}\sum_{i=1}^{r} d(y_i, V_{MOEA})^p\right)^{1/p} && (6)\\ GD(V_{MOEA}, \Lambda)_p &= IGD(\Lambda, V_{MOEA})p && (7)\end{aligned}$$

The indicator, so called by Δ_p, was proposed in [11] defined by

$$\Delta_p(X, Y) = max\{IGD(X, Y)_p, GD(X, Y)_p\}, \tag{8}$$

In [11] the author proved that function is a semi-metric, Δ_p does not fill the triangle inequality, for $1 \leq p < \infty$. Many others properties was proved in [11].

3.3 Hypervolume

This indicator has been using by the community since 2003. Basically, the hypervolume of a set of solutions measures the size of the portion of objective space that is dominated by those solutions as a group. In general, hypervolume is favored because it captures in a single scalar both the closeness of the solutions to the optimal set and, to some extent, the spread of the solutions across objective space. There are many works on this indicator such as in [12] which the author studied how expensive to calculate this indicator was. Few years later, it was proposed a faster alternative by using Monte Carlo simulation (See [13]). In the order to get a right definition, you can look at [12, 13].

4 Proposal Measure $\mathcal{H}$

In the order to define our proposal measure, consider the quadratic optimization problem (9) associated with (1):

$$\min_{\alpha \in \mathbb{R}^n} \left\{ \left\| \sum_{i=1}^{m} \alpha_i \nabla f_i(x) \right\|^2 ;\ \alpha_i \geq 0,\ \sum_{i=1}^{m} \alpha_i = 1 \right\} \tag{9}$$

The existence and uniqueness of a global solution of the problem (9) is stabilized in [14]. The function $q : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$q(x) = \sum_{i=1}^{m} \widehat{\alpha}_i \nabla f_i(x) \tag{10}$$

where $\widehat{\alpha}$ is a solution of the (9), becomes well defined. There is a interested property about this function that was proved in [14]:

- each x^* with $\|q(x^*)\|^2 = 0$, where $|\cdot|$ represents euclidean norm, fulfills the first-order necessary conditions for Pareto optimality given by Theorem 2.1.

Thereby, these points are certainly Pareto candidates what motivates the next definition about nearness.

Definition 4.1. *A point $x \in \Omega$ is called $\epsilon-$closed to Pareto set if $\|q(x^*)\|^2 < \epsilon$.*

Let the set $X = \{x_1, x_2, \cdots, x_k\}$ the output from some evolutionary algorithms. With the feature about that function, we can to define a new measure by:

$$\mathcal{H}(X) := \frac{1}{2k} \sum_{i=1}^{k} -q_i \log_2(q_i) \tag{11}$$

where $q_i = min\{1/e^1, \|q(x_i)\|^2\}$ and put $0 \cdot log_2(0) = 0$.

The expression (11) is the traditional formula used for Shannon Entropy (See [16]). Unlike the original way that is used this as a entropy, in the our metric each q_i is not to related with a probability space.

Theorem 4.1. *About the function in* (11)*, we have that* $0 \leq \mathcal{H}(X) \leq \frac{1}{2}\frac{\log_2(e^1)}{e^1}$.

Proof. First, since $0 \leq q_i \leq 0.5$ then $\mathcal{H}(X) \geq 0$. On the other hand, it is known that the function $f(x) = -x\log_2(x)$ has a maximum at $x = 1/e^1$, therefore

$$\begin{aligned} \mathcal{H}(X) = \frac{1}{2k}\sum_{i=1}^{k} -q_i \log_2(q_i) &\leq \frac{1}{2k}\sum_{i=1}^{k} -e^{-1}\log_2(e^{-1}) \\ &\leq \frac{1}{2k}(k)(e^{-1}\log_2(e^1)) \\ &= \frac{1}{2}\frac{\log_2(e^1)}{e^1} \approx 0.26537 \end{aligned}$$

□

A reasonable interpretation of about $\mathcal{H}(\cdot)$ can be : 1) When $\mathcal{H}(X)$ closed to or equal 0 then the set X has a good convergence to the Pareto Set and 2) On the other hand, for values near to or equal 0.26537, the set X does not convergence to Pareto Set.

The main features associated with this metric are: 1) The true Pareto set doesn't require to be predefined and 2) It can be to used with many-objectives problems. The first point is important because it is almost impossible to know the true PF in general. On the other hand, the second attribute is useful since it doesn't exists many indicator to be used in this problems.

5 Simulation Results

This section presents some functions to be optimized with two objectives. The following equations are from [11] and it is known by ZDT1 and ZDT2. Here, we have decided for the two well-known algorithms: 1) MOPSO [3] and 2) NSGA-II [1].

In both algorithms, we will fix the output population at 150 individuals. The others parameters used are:

- MOPSO: inertia $w = 1/(2 \cdot log(2))$, acceleration coefficients $\phi_1 = 0.8$, $\phi_2 = 0.8$ and size of population $pop = 50$.
- NSGA-II: standard real parameter SBX and polynomial mutation operator with $n_c = 15$ and $n_m = 20$, respectively. In addition, the crossover probability of $p_c = 0.9$ and mutation probability is $p_m = 1/n$.

The algorithms are run 30 times with 30000 function evaluations running in each one.

The aim is to compare the results of problems with the indicators. From the table (1), we can conclude the same with all indicator, included our proposal. The algorithm on second column (NSGA-II) has better convergence than the other.

Table 1: Indicators results for ZDT1 function

	MOPSO-CD	NSGA-II
GD	0.005834 (0.000866)	**0.003266 (0.000768)**
IGD	0.006415 (0.000623)	**0.003490 (0.000168)**
Δ_2	0.006517 (0.000679)	**0.003705 (0.000552)**
$\mathcal{H}$	0.006429 (0.000816)	**0.000378 (0.000045)**

Table 2: Indicators results for ZDT2 function

	MOPSO-CD	NSGA-II
GD	0.005834 (0.000866)	**0.003266 (0.000768)**
IGD	0.006415 (0.000623)	**0.003490 (0.000168)**
Δ_2	0.006517 (0.000679)	**0.003705 (0.000552)**
$\mathcal{H}$	0.006429 (0.000816)	**0.000378 (0.000045)**

6 Conclusions

In this paper, we have introduced a new indicator $\mathcal{H}$ which goal is evaluate the outcome from an evolutionary algorithm on the Multi-objective optimization problems. We have tested the new approach some MOPs and make a comparative with some others indicators that is already known. By experimental results, we can to conclude the same what is indicated by the other indicators of performance. Unlike such these indicators, our proposal needs to known nothing about the true pareto of set of the MOP. This features, we consider a great to thing because most MOPs is doesn't have this information.

For the future, we will try to decrease the condition about the function of the problems, until now we required being $\mathcal{C}^2(\Omega)$.

Acknowledgment

The authors would like to be thankful for the FAPEMIG by financial support to this project.

References

[1] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA II, IEEE Trans. Evol. Comp., vol. 6 (2), 182-197, (2002).

[2] E. Zitzler, M. Laumanns and Lothar Thiele, Spea2: Improving the strength pareto evolutionary algorithm, Technical Report, (2001).

[3] C. A. Coello Coello and M.S. Lechuga. Mopso: A proposal for multiple objective particle swarm optimization, Proceedings of the Evolutionary Computation on 2002, vol. 2, 1051-1056,(2002).

[4] C. Igel, N. Hansen and S. Roth, Covariance matrix adaptation for multi-objective optimization, Evol. Comput., vol 15(1), 1-28, (2007).

[5] T. Santos and R. H. C. Takahashi, On the performance degradation of dominance-based evolutionary, algorithms in many-objective optimization, IEEE Trans. Evol. Comp., (2016), DOI: 10.1109/TEVC.2016.2623770.

[6] O. Schutze, A. Lara and C. A. C. Coello, On the influence of the number of objectives on the hardness of a multiobjective optimization problem, IEEE Trans. Evol. Comp., vol. 15(4), 444-455, (2011).

[7] B. Li, J. Li, K. Tang and X. Yao, Many-objective evolutionary algorithms: A survey, ACM Comput. Surv., vol. 48(1), 1-13, (2015).

[8] H. Ishibuchi, N. Tsukamoto, Y. Sakane and Y. Nojima, Indicator-based evolutionary algorithm with hypervolume approximation by achievement scalarizing functions, Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation GECCO '10, 527-534, (2010).

[9] T. Wagner, N. Beume and B. Naujoks, Pareto-, aggregation-, and indicator-based methods in many-objective optimization, Proceedings of the 4th International Conference on Evolutionary Multi-criterion Optimization EMO 07, 742-756, (2007).

[10] H. Ishibuchi, H. Masuda, Y. Tanigaki and Y. Nojima, Difficulties in specifying reference points to calculate the inverted generational distance for many-objective optimization problems, IEEE Symposium on Computational Intelligence in Multi-Criteria Decision-Making, 170-177, (2014).

[11] O. Schutze, X. Esquivel, A. Lara and C. A. Coello Coello, Using the averaged hausdorff distance as a performance measure in evolutionary multiobjective optimization, IEEE Trans. Evol. Comp, vol. 16(4), 504-522, (2012).

[12] L. While, P. Hingston, L. Barone and S. Huband, A faster algorithm for calculating hypervolume, IEEE Transactions on Evolutionary Computation, vol. 10(1), 29-38, (2006).

[13] J. Bader and E. Zitzler, Hype: An algorithm for fast hypervolume-based many-objective optimization, Evol. Comput., vol. 19(1), 45-76, (2011).

[14] S. Schaffler, R. Schultz and K. Weinzierl, Stochastic method for the solution of unconstrained vector optimization problems, J. Optim. Theory Appl., vol. 114(1), 209-222, (2002).

[15] H. W. Kuhn and A. W. Tucker, Nonlinear programming, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 481-492, (1951).

[16] R.M. Gray. Entropy and Information Theory. Springer-Verlag New York, USA, (1990).