A New Dynamically Reference Point Adaptation Mechanism in indicator-based EMOA based on weak convergence detection

Weiduo Liao
School of Electrical and
Computer Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332–0250
Email: http://www.michaelshell.org/contact.html

sssss
Twentieth Century Fox
Springfield, USA
Email: homer@thesimpsons.com

ssss
Starfleet Academy
San Francisco, California 96678–2391
Telephone: (800) 555–1212
Fax: (888) 555–1212

***Abstract*—The abstract goes here.**

***Keywords*—keyword 1; keyword 2**

I. Introduction

This demo file is intended to serve as a "starter file" for IEEE conference papers produced under LaTeX using IEEEtran.cls version 1.8a and later. I wish you the best of success.

mds

September 17, 2014

A. Subsection Heading Here

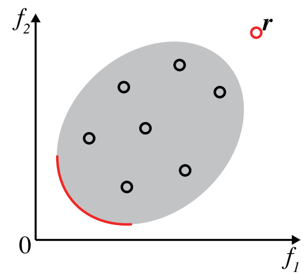
Subsection text here.

1) Subsubsection Heading Here: Subsubsection text here.

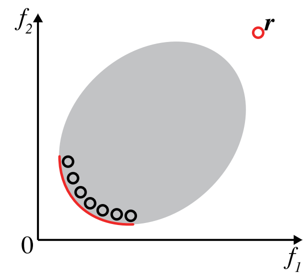
II. Reference Point Adaptation

When hypervolume(HV) is used in indicator-based algorithms, one important thing to be considered is that how to specified the reference point. Before calculating the HV values, reference point needs to be chosen in advance. However, it is not suggested that the reference point is set only once at the beginning. [**?**] This may cause a very far away reference point from solutions for those problems with a very large feasible space(As shown in Fig. 1), as the solutions set is gradually converging to the pareto front during the iteration of the algorithm process.

There is a big problem when applying this strategy to some problems with specific pareto front shape, for example, the inverted-DTLZ1 problem with a inverted-triangular pareto front in 3 dimensions, that many solutions in the final solutions set will distribute at the boundary of the pareto front (Fig. 2a comparing with Fig. 2b) [**?**], [**?**]. Although it has no effect on the distributions of solutions set in problems with triangular pareto front in 3 dimensions (Fig. 2c comparing with Fig. 2d), it is necessary using reference point adaptation during the algorithm progress. And the reason is illustrated detailedly in hisao [**?**].



(a) The initial generation.



(b) The final generation.

Fig. 1. The reference point is set with a large feasible space. pareto front can be far away from reference point. The gray region shows the feasible region and the red arc is the corresponding pareto front. The red circle r is the reference point calculated by the initial solutions in (1a) which is randomly generated. After some generations, the current solutions reach to the five black circles in (1b), which is far away from the reference point.

In many algorithms including SMS-EMOA [**?**], the reference point is adapted based on the following rules:

$$RP = r * ENP, r = 1.1. \tag{1}$$

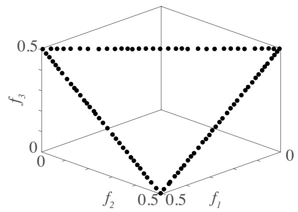
Note that the estimated nadir point(ENP) is the nadir point in current population. When the solutions in the current population is obtained, we use hypervolume as indicator to evaluate the performance of the solutions set. Then the reference point used to calculate the hypervolume is calculated by the formula above.

III. Dynamic Mechanism

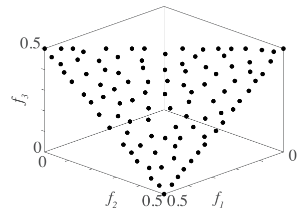
Basically the process of Evolutionary Multi-objective Optimization Algorithm can be separate into two stages:

1) Early Stage: In this stage, all the solutions are far away from pareto front. The main task is to converge the solutions to pareto front. We also call this stage convergence stage.

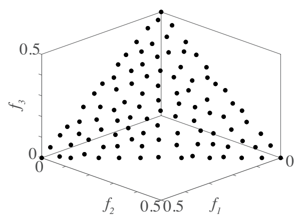
2) Final Stage: In this stage, all the solutions are in or near the pareto front. So the main task is to make the distribution of solutions more evenly in the pareto front. We also call this stage diversity stage.



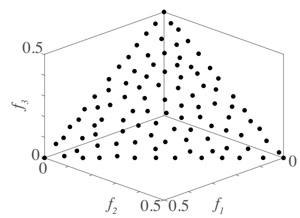
(a) The inverted-DTLZ1 problem with a fixed reference point.



(b) The inverted-DTLZ1 problem with a adapted reference point.



(c) The DTLZ1 problem with a fixed reference point.



(d) The DTLZ1 problem with a adapted reference point.

Fig. 2. The final distribution of solutions set in the inverted-DTLZ1(2a and 2b) and the DTLZ1 problem(2c and 2d). The algorithm is FV-EMOA with population size = 100, evaluation number = 20000 and $r = 1.1$. (2a and 2c): the reference point is calculated only once at the initial step; (2b and 2d): the reference point is adapted based on the formula (1). All the solutions in the final distribution are at the boundary of the pareto front in (2a), which shows the bad effect of a far away reference point on the final distribution of inverted-triangular problems. This bad effect can not be observed on triangular problems in (2c).

For different purposes in these two stages, the r should be treated differently [**?**]. Not only the reference point but also the value of r needs to be adapted in each iteration of algorithm. This is called dynamically reference point adaptation.

Unfortunately, the research on how to specified r is limited. Only a few papers [**?**], [**?**], [**?**], [**?**] did some research on the reference point. The reason is that, the effect of the location of the reference point on the pareto front is not fatal on some benchmark problems, especially triangular pareto front. But in fact, on some specific problems, the distribution of solutions on pareto front strongly depends on the location of the reference point. The sensitivity about value of r for solutions is also observerd on some real world problems, for example, distance minimization problems. This observation Potential shows the usefulness of the dynamically reference point adaptation [**?**].

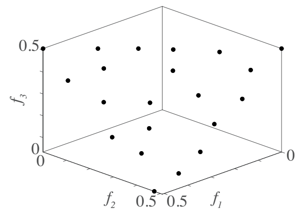
In this section, we will specify the suggested r values seperately in the early and final stages.

A. Reference Point Specification for Optimal Distribution

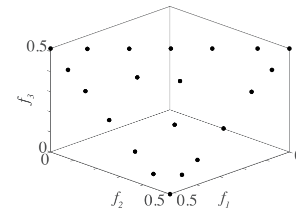
In the final stages, the major purpose is to augment the diversity of solutions set. More specifically, for inverted-triangular problems, the interval between two boundary solutions should be same as that between two inner solutions. In hisao [**?**].

$$r = 1 + \frac{1}{H}, \tag{2}$$

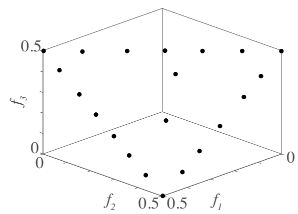
is the optimal setting for flat(not concave or convex) pareto front problem (As shown in Fig. 3). H is the number of



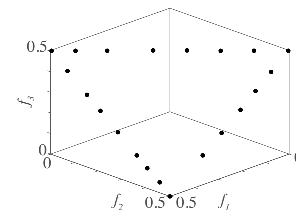
(a) $r = 1.2$(optimal).



(b) $r = 1.5$.



(c) $r = 2$.



(d) $r = 5$.

Fig. 3. The final distribution of solutions set in the inverted-DTLZ1 problem. The algorithm is FV-EMOA with population size $= 21(H = 5)$ and evaluation number $= 20000$. $r = 1.2$ is the optimal setting and we observed a evenly distribution in (3a). As the increasing of r, solutions are more likely to be at the boundary (3b-3d).

solutions intervals in 2-dimension and the number of interval at each boundary of pareto front in many-dimension. In Fig. 3a, $r = 1.2(H = 5)$ is the optimal setting for a 21 individuals inverted-DTLZ1 problem and a evenly distribution is observed. In Fig. 3b - Fig. 3d, the inner solutions are decreased and move to the boundary of pareto front.

For simply illustration, when algorithm reach to the final stage, all solutions are near the pareto front, r should be specified as $1 + 1/H$, as in equation 2.

B. Reference Point Specification for Convergence

However in the early stage, solutions set is not close to the pareto front, which makes the estimated ideal and nadir points far away from true ideal and nadir points, for the reason that they are calculated by current nondominated solutions in each generation. As a result, the search behavior will be poor for the poor normalization [**?**], [**?**].

In [**?**], a larger value of r than the suggested value $1 + 1/H$ is used in the early stage.

C. Linearly Decrease Mechanism

Based on the theory above, r is suggested to be specified dynamically at different stages of the algorithm(at early stage, a slightly larger r is chosen; at final stage, $r = 1 + 1/H$ is chosen). But unfortunately, there is no best mechanism on how to specify value of r dynamically outperforms the others in all problems and all experiments settings. One mechanism may be the best when work on some specific experiments conditions, but may be not good on other conditions.

In [**?**], a linearly decrease mechanism has been proposed:

$$r(t) = r_{Initial}\frac{(T-t)}{T} + (1 + 1/H)\frac{t}{T}, t = 0, 1, \ldots, T, \tag{3}$$

where T is the total number of generations, and $r_{Initial}$ is the initial value of r, which is larger than $1 + 1/H$. It is a simple and practical mechanism. In (3), the value of r starts from $r_{Initial}$, then gradually decrease to the suggested value in a linearly decrease process.

In next section, we will propose another good dynamic mechanism based on weak convergence detection criterion outperform simple linearly decrease mechanism on some constraint condition.

IV. New Dynamic Mechanism

In this section, we will introduce a new mechanism that uses a weak convergence detection criterion to decide whether to change the value of r from $r_{Initial}$ to $1 + 1/H$.

As we have explained before, a slightly larger r is suggested at the initial stage of the algorithms. But for well diversity at the final stage, it is needed to set r to it's optimal value $(1 + 1/H)$. For this purpose, we detect whether algorithm has convergenced or not. If solutions are all close to the pareto front, we change the value of r to $r_{Optimal}$; otherwise, we set value of r to $r_{Initial}$. The mechanism is also shown below:

$$\begin{aligned} r(t) = {} & r_{Initial}\mathbb{I}\left[t < t_{Convergent}\right] \\ & + (1 + 1/H)\mathbb{I}\left[t \geq t_{Convergent}\right], t = 0, 1, \ldots, T, \end{aligned} \tag{4}$$

where $\mathbb{I}$ is the indicator function returning 1 if argument is true and 0 otherwise. $r(t)$ equals to $r_{Initial}$ before reaching to the convergent generation $t_{Convergent}$, and changes to $1 + 1/H$ after $t_{Convergent}$. The $t_{Convergent}$ is determined by a weak convergence detection criterion.

A. Weak Convergence Detection

Consider many convergence detection paper [**?**], various indicators including convergence detection indicators are using to detect the stagnation. They focus on accuracy of convergence, which is not the purpose in our approach for the reason that after algorithm convergent, we still need some generations in order to get evenly distribution of solutions set. We summarize our weak convergence detection criterions as follow:

1) inaccuracy: It is no need to have a accurate convergence detection. The convergence can be reported if current solutions are close to the pareto front. In other words, the estimated ideal and nadir points based on the current solutions are close to the true ideal and nadir points.

2) saving time: We should not spend too much time in convergence detection for the reason that the state-of-the-art indicator-based algorithms such as SMS-EMOA and HypE, are time consuming when dimension is very high.

We are discussing the effect of reference points in calculating hypervolume. It seems to be a good idea to use progress indicator hypervolume as our convergence detection indicator, for that we have calculated hypervolume in each generation in hypervolume-based evolutionary multi-objective optimization algorithm. But during the process of algorithm, the reference point is calculated by (1) in each generation. So we can not just simply compare hypervolume calculated in algorithm among different generations.

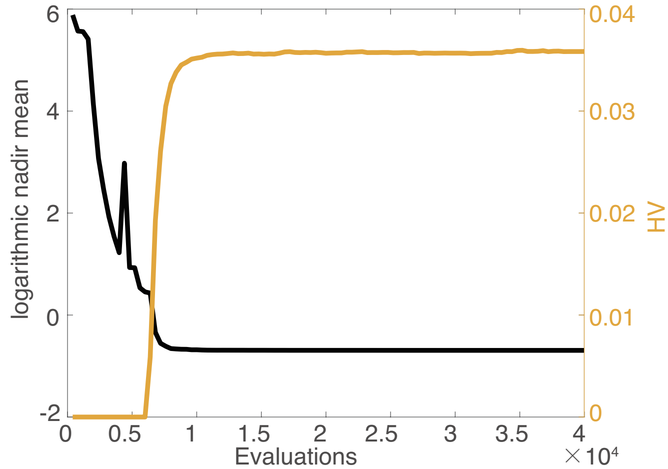


Fig. 4. Example of nadir point and hypervolume(HV) value on inverted-DTLZ1 3-dimension problem with FV-EMOA algorithm. The yellow curve is the change of hypervolume while the black curve is the change of logarithmic nadir point.

We are trying on some other good indicators satisfying our convergence detection criterions. Fig. 4 shows the change of hypervolume(HV) and nadir point on the FV-EMOA algorithm with 3-dimensional inverted-DTLZ1 problem. When current solutions are close to the pareto front, the estimated nadir point is close to the true nadir points. This imply that the estimated nadir point can be a good indicator for our purpose. But for some problems with large feasible region, the moving distance of estimated nadir point in the early generations is larger than that near convergence. So we consider the logarithm. And consider the possible bad variance of the indicator, finally we use best logarithmic nadir point so far as our indicator. more specifically, for a minimization problem, we consider the indicator as follows:

$$\begin{aligned} ENP_t &= [f_{t1}, f_{t2}, \ldots, f_{tm}]^\top \in \mathbb{R}^m, \\ I_0 &= \frac{1}{m}\sum_{i=1}^{m} lnf_{0i}, \\ I_t &= min(I_{t-1}, \frac{1}{m}\sum_{i=1}^{m} lnf_{ti}), t = 1, 2, \ldots, T, \end{aligned} \tag{5}$$

where T is the total number of generations, ENP_t is the estimated nadir point at the tth generation with m objectives $f_{t1}, f_{t2}, \ldots, f_{tm}$. I_0 is the initial indicator calculated by the initial population. And I_t is the minimum value before the tth generation (including the tth generation).

After chosen the indicator, the next step is to detect the stagnation of indicator. We use a basic linear regression method called Simple Least Squares [**?**] with a simple least squares convergence detection strategy introduced in [**?**]. If the absolute value of slope of indicator is below a threshold, the convergence is reported. Briefly speaking, for a simple linear regression $I = a + bt$, the intercept a and slope b of the tth generation can be calculated with the following matrix-based

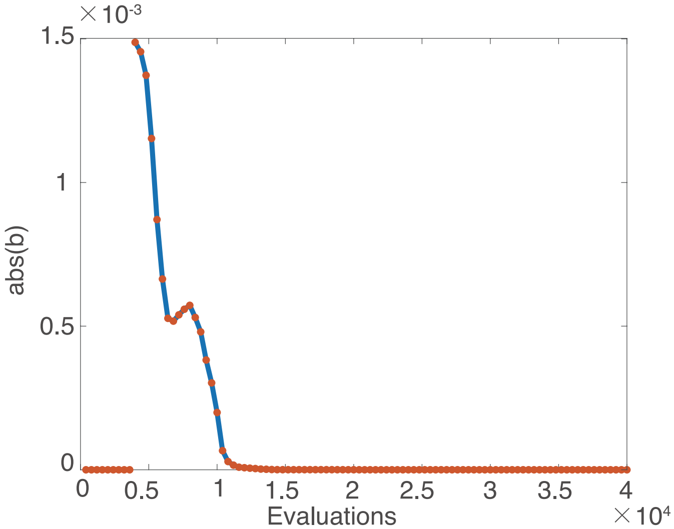


Fig. 5. Example of $|b|$ on inverted-DTLZ1 3-dimension problem (window size = 4000 evaluations with population size = 100).

formula:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum t_i^2 & \sum t_i \\ \sum t_i & w\_l \end{bmatrix}^{-1} * \begin{bmatrix} \sum t_i * I_{t_i} \\ \sum I_{t_i} \end{bmatrix} \tag{6}$$

where w_l is the length of the chosen window and t_i is the evaluated number in the chosen window. The value of slope b is shown in Fig. 5 (Note that the value in the first w_l evaluations is 0, and we should not consider the first w_l evaluations).

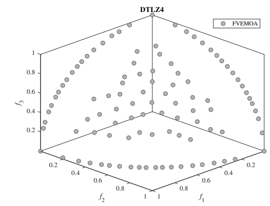
With the above formula (6), the convergence detection criterion is defined as:

$$convergence = |b| < thres \tag{7}$$

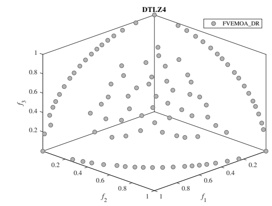
The chosen of the $thres$ value is not so important as the report ahead or delay is not fatal to the algorithm or to the final solutions set. We choose the $thres$ value as 10^{-5} after some experimental computation with the window size $w\_l = 4000$ evaluations. If we do not change the window size, this threshold can be applied to other problems or indicator-based algorithms because of a weak convergence detection purpose.

V. Computational Experiments

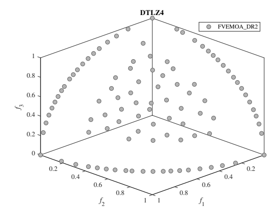
To clearly represent the superiority of dynamically reference point adaptation and to ease the comparsion process of two different dynamic mechanisms (linearly decrease mechanism and weak convergence detection mechanism), the mechanisms is going to be tested with the algorithm FV-EMOA [**?**]. The problems include two triangular pareto front problems: DTLZ1 [**?**], C1-DTLZ1 [**?**] and two inverted-triangular pareto front problems: Inverted-DTLZ1 [**?**], MaF1 [**?**]. To show the performance both on multi-objective and many-objective, we tested in 3,5,8 and 10 dimensions.



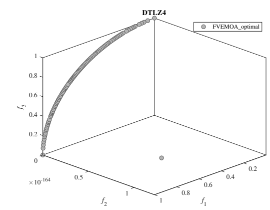
(a) FV-EMOA-2



(b) FV-EMOA-LD



(c) FV-EMOA-CD



(d) FV-EMOA-Opt

Fig. 6. The final distribution of 4 reference point strategies on DTLZ4 problems.

A. Computational Results

Each experiment has been run for twenty times independently. And the population size is 100, $r_{Initial} = 2$. The detailed algorithms are as following: FV-EMOA-2(the FV-EMOA [**?**] algorithm with $r = 2$), FV-EMOA-LD(the FV-EMOA algorithm with a linearly decrease mechanism), FV-EMOA-CD(the FV-EMOA algorithm with a weak convergence detection mechanism, proposed in this paper) and FV-EMOA-Opt(the FV-EMOA algorithm with $r = 1 + 1/H$). We obtain the hypervolume results after 40000 evaluations for each algorithm. The computational results is shown in TABLE I.In triangular pareto front problems(DTLZ1, C1-DTLZ1), we can not tell the differences among the four algorithms in TABLE I, for that all the four algorithms are more or less the best or the worst. Note that, the hypervolume results of FV-EMOA-LD with DTLZ1 problem and 10-dimension is all 0, because of the un-convergence of the algorithm. But in inverted-triangular pareto front problems(Inverted-DTLZ1, MaF1), FV-EMOA-2 performs the worst in all experiments, for that the r is 2 all the process of algorithm. What's more, we can not tell the differences among the other three algorithms, for that the r_{Final} of the three algorithms are all $1 + 1/H$.

B. The Importance of Using Dynamic Mechanism

We plot the final solutions distributions of some experiments including DTLZ4 and minus-DTLZ4 in 3-dimension. In Fig. 6d, the solutions are poor distributed comparing with Fig. 6a-6c. The only difference is that FV-EMOA-Opt applies $r_{Initial} = 1 + 1/H$ mechanism

C. Comparison of Two Dynamic Mechanisms

We want to further investigate the differences between two dynamic mechanisms.

VI. Conclusion

The conclusion goes here.

TABLE I
HV MEAN AND STANDARD DEVIATION OVER 20 INDEPENDENT RUNS

Problems	M	FV-EMOA-2	FV-EMOA-LD	FV-EMOA-CD	FV-EMOA-Opt
DTLZ1	3	1.40E-01±1.49E-04	1.40E-01±1.18E-04	1.40E-01±9.18E-05	**1.40E-01±1.04E-04**
DTLZ1	5	4.90E-02±1.74E-05	**4.90E-02±1.15E-05**	4.90E-02±1.42E-05	4.90E-02±1.38E-05
DTLZ1	8	6.26E-03±3.14E-03	7.63E-03±1.61E-03	5.78E-03±3.48E-03	**8.14E-03±4.50E-04**
DTLZ1	10	9.54E-05±2.83E-04	0.00E+00±0.00E+00	2.73E-06±8.45E-06	**5.46E-04±1.02E-03**
C1-DTLZ1	3	1.39E-01±7.22E-04	**1.40E-01±7.24E-04**	1.39E-01±7.23E-04	1.40E-01±6.10E-04
C1-DTLZ1	5	4.85E-02±3.69E-04	4.79E-02±8.35E-04	**4.85E-02±3.17E-04**	4.70E-02±2.16E-03
C1-DTLZ1	8	**8.03E-03±3.30E-04**	7.75E-03±4.98E-04	7.98E-03±3.68E-04	7.86E-03±3.93E-04
C1-DTLZ1	10	**2.40E-03±1.09E-04**	2.40E-03±8.41E-05	2.38E-03±1.21E-04	2.31E-03±2.27E-04
MaF1	3	86E-01±7.88E-04	**2.97E-01±1.57E-04**	2.97E-01±1.94E-04	2.97E-01±1.49E-04
MaF1	5	1.09E-02±1.28E-04	1.63E-02±1.38E-04	1.63E-02±1.63E-04	**1.63E-02±1.07E-04**
MaF1	8	2.24E-05±9.81E-07	4.18E-05±1.31E-06	4.23E-05±9.14E-07	**4.26E-05±7.84E-07**
MaF1	10	1.85E-07±1.70E-08	6.73E-07±1.95E-08	**6.78E-07±1.73E-08**	6.78E-07±1.34E-08
Inverted-DTLZ1	3	3.57E-02±2.97E-04	**3.69E-02±3.30E-04**	3.65E-02±1.27E-03	3.65E-02±1.20E-03
Inverted-DTLZ1	5	3.19E-04±4.11E-05	4.65E-04±7.01E-05	4.40E-04±7.11E-05	**4.90E-04±5.93E-05**
Inverted-DTLZ1	8	5.99E-08±1.32E-08	1.41E-07±3.43E-08	**1.57E-07±2.06E-08**	1.56E-07±2.12E-08
Inverted-DTLZ1	10	1.83E-10±3.06E-11	6.51E-10±5.00E-11	**6.58E-10±7.05E-11**	6.55E-10±6.20E-11

ACKNOWLEDGMENT

The authors would like to thank... [2]

REFERENCES

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.

[2] H. pka and P. W. Daly, *A Guito LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.