Comparative Evaluations of Evolutionary Computation with Elite Obtained in Reduced Dimensional Spaces

Pei, Yan
Graduate School of Design, Kyushu University

Takagi, Hideyuki
Faculty of Design, Kyushu University

出版情報：International Conference on Intelligent Networking and Collaborative Systems (INCoS). 3, pp.1-, 2011-11-30. IEEE
バージョン：
権利関係：

Comparative Evaluations of Evolutionary Computation with Elite Obtained in Reduced Dimensional Spaces

Yan Pei
Graduate School of Design, Kyushu University
Fukuoka, Japan
peiyan@ieee.org

Hideyuki Takagi
Faculty of Design, Kyushu University
Fukuoka, Japan
http://www.design.kyushu-u.ac.jp/ takagi

***Abstract*—We propose an elite synthesis optimization strategy for accelerating evolutionary computation (EC) searches using elites obtained from a lower dimensional space. The method projects individuals onto n one-dimensional spaces corresponding to each of the n searching parameter axes, approximates each landscape using Lagrange polynomial interpolation or power function least squares approximation, finds the best coordinate for the approximated shape, obtains the elite by combining the best n found coordinates, and uses the elite for the next generation of the EC. The advantage of this method is that the elite may be easily obtained thanks to their projection onto each one-dimensional space and that there is a higher possibility that the elite will be located near the global optimum. We conduct experimental tests to compare our proposed approaches with previous acceleration approaches using differential evolution and ten benchmark functions. The results demonstrate that the proposed method accelerates EC convergence significantly, especially in early generations.**

***Keywords*-evolutionary computation, fitness landscape, convergence acceleration, dimensionality reduction, approximation approach**

I. INTRODUCTION

The success of evolution in nature is not equalled by evolution-based models used in computing because resources such as time and memory, which are abundant in nature are constrained in a computer system, and even more so in the case of a system requiring human interactivity, such as interactive evolutionary computation (IEC) [1]. In recent evolutionary computation (EC) research, research has been focused in two directions. One is extending EC applicationsf, that is to design and discover effective EC search strategies to solve problems that cannot be solved or are ineffectively solved by conventional approaches. The other is making EC algorithms more efficient, that is to modify and improve existent EC algorithms such that convergence can be accelerated. Accelerating EC is necessary for many EC applications to improve the performance of the target systems. For example, as an IEC optimizes the target system based on the IEC user's subjective evaluation, user fatigue becomes a serious problem impending its practical use. Multiple trials for accelerating EC have therefore been proposed [1].

To use the EC landscape information directly is one such acceleration approach. Other approaches include landscape approximation with simpler shapes [5]. Takagi et al. have proposed to use a single peak function to approximate the EC search space [2], however, this calculation is costly because the approach requires much data to approximate the single peak function in the original higher dimensional space. If we use interpolation or approximation approaches to reduce the complexity of search space, it becomes easier to reach the global optimum in the regression space. It is not the actual global optimum, but may be a neighbor to the global optimum in the original search space. From this neighboring point, it is then easy to reach the actual global optimum. If the search space is high dimensional, nonlinear, and non-differentiable, conventional approximation approaches are difficult to apply. The best solution is to reduce the search space dimension. As the complexity of the approximated landscape is lower than that of its original searching space, it becomes easier to reach to the area neighboring the global optimum. Although the global optimum of the approximated simple space is different from that of its original space, it still provides useful information for finding the real global optimum.

We proposed a method for finding elite obtained in regression spaces, that reduces the dimension of the search space and composites each dimensional new elite together as an individual in the next EC interaction computing [3]. Here, *regression space* refers to a lower dimensional space consisting of k dimensional (k-D) axes of the original n-D axes, where we use k = 1 in this paper. In other words, individuals are projected onto the k-D space to simplify the task of finding the elites. This elitism does not destroy the original EC search space by approximation, but can accelerate the EC convergence with less computational cost by conducting the elite search in the n regression spaces of k-D (k = 1). This paper extends this study, and conducts experiments to compare its performance with previous acceleration approaches [2], then obtains some new conclusions. This is the originality contributed by this paper.

In the following section, we introduce the elite combination search approach enabled by a technique for reducing the dimensionality of the searching space and show how elite can be obtained from the regression search space. In section 3, experimental evaluations with ten benchmark functions [4][6] are conducted, and their results compared with the previous work are discussed. Finally, we discuss our proposed methods and the results obtained in section 4. In section 5, we present our conclusions on the performance of the proposed methods, and further opportunities are discussed.

II. Elite Synthesis Optimization Strategy

A. Dimensionality Reduction Approach

It is not easy for conventional interpolation or approximation approaches to find an accurate regression search space corresponding to the original multi-dimensional space. An alternative approach is to reduce the dimensionality of the original search space and find approximation curve expressions in the lower dimensional space.

Our approach for reducing the dimensionality of the searching space uses only one of the n parameter axes at a time instead of all n parameter axes, and projects individuals onto each 1-D regression space. The landscape of the n-D parameter space is given by a fitness function, $y = f(x_1, x_2, ..., x_n)$, and the fitness value of the m-th individual is given by Equation (1):

$$y_m = f(x_{1m}, x_{2m}, ..., x_{nm}) \; (m = 1, 2, ..., M) \qquad (1)$$

There are M individuals with n-D parameter variables. We project the individuals onto the n 1-D spaces as follows:

$$(x_1, y) \; (x_2, y) \; ... \; (x_n, y)$$

Each of the n 1-D regression spaces has M projected individuals (see Figure 1).

B. Approach for Simplifying Regression Space Landscape to Select Elite

We interpolate or approximate the landscape of each 1-D regression space using the projected M individuals and select the elite from the n approximated 1-D landscape shapes. In this paper, we test two approaches for approximating the 1-D regression search spaces; in the first approach we use a Lagrange two-degree polynomial interpolation and in the other linear least squares approximation. Elite are generated from the resulting approximated shapes.

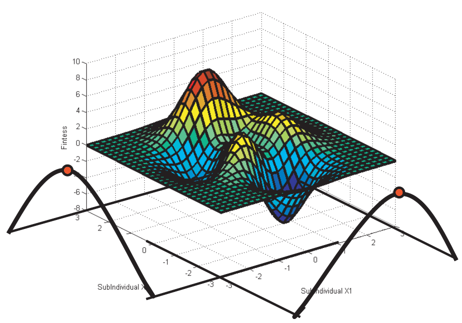

Figure 1. Original n-D space and 1-D spaces obtained by reducing dimensions of the original one.

Finding the elite corresponds to a kind of local search in the area where relatively better individuals exist in the original search space. The global optimum is expected to be near this area and we may be able to find it with a probability higher than chance [2]. So the elite selection approach is a critical step in the proposed acceleration processes. As the elite obtained by the two different elite selection approaches is different, it is expected that the acceleration performance will also differ. Further, the regression EC search space obtained by approximation or interpolation has its own characteristics and particularities, and we must use an efficient approach to obtain elite from this simplified search space after analyzing its characteristics. The actual interpolation polynomial and least square regression functions used are given by the Equations (2) and (3).

$$L(x) = \sum_{k=1}^{3} \left\{ \prod_{i=1, i \neq k}^{3} \frac{(x - x_i)}{(x_k - x_i)} \right\} y_k \qquad (2)$$

$$\begin{pmatrix} (\varphi_0 \varphi_0)(\varphi_0 \varphi_1) \\ (\varphi_1 \varphi_0)(\varphi_1 \varphi_1) \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} \qquad (3)$$

where x_i, x_k, y_i are the best projected individual and their fitness values among (x_{im}, y_m) in the i-th 1-D regression space for $(i = 1, 2, ..., n)$ and $(m = 1, 2, .., M)$, A and B are the parameters obtained via least squares method, and φ_0, and φ_1 are the vectors of the least square equation.

Lagrange two-degree polynomial interpolation simplifies a regression space with a nonlinear curve, but it is easy to obtain its stationary point from its gradient, using the stationary point as the elite. Linear least square approximation uses a linear function to approximate the regression space. Its gradient is either descent or ascent. A safer approach, taking into account both descent and ascent, is to select the middle point of the linear approximation line as the elite.

The proposed methods replace the worst individual in each generation with the selected elite. Although we cannot deny the small possibility that the global optimum is located near the worst individual, the possibility that the worst individual will become a parent in the next generation is also low; removing the worst individual therefore presents the least risk and is a reasonable choice.

C. Approach for Synthesizing Elite

The methods in section II-B select n elite points in n 1-D regression spaces, respectively: $x_{1-elite}$, $x_{2-elite}$, ..., and $x_{n-elite}$. The n-D elite used for accelerating EC convergence in the next generation is obtained as follows:

$$\text{new elite} = (x_{1-elite}, x_{2-elite}, .., x_{n-elite}).$$

It is easier to calculate the elite in a lower dimensional space than a higher dimensional space. Although we use 1-D as the lower dimensional space in this paper, the method need not be restricted to 1-D in general.

Once a new elite has been obtained, there are two approaches for how it can be handled. In one cautious approach, the fitness value of this elite is calculated to determine whether the new elite is really useful for acceleration, and in the other straightforward approach, the new elite is inserted into the next EC iteration process without any prior consideration or judgment.

Our proposed approaches is based on the hypothesis that elite calculated by interpolation or approximation from

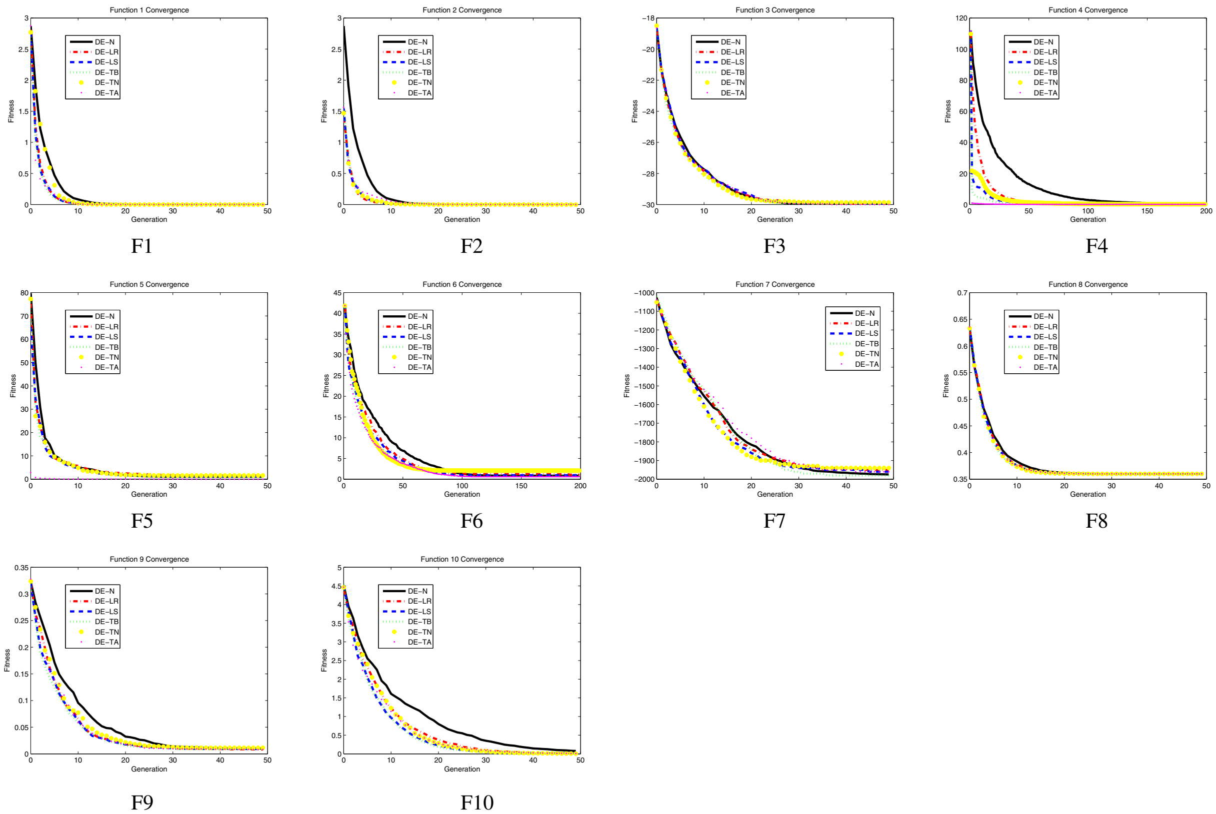

Figure 2. Average convergence curves of 50 trial runs for (F1) De Jong Function1, (F2) De Jong Function2, (F3) De Jong Function3, (F4) De Jong Function4, (F5) De Jong Function5, (F6) Rastrigin Function, (F7) Schwefel 2.26 Function, (F8) Griewank Function, (F9) Schaffer-1 Function, (F10) Schaffer-2 Function

relatively better individuals will also have good fitness; synthesizing an n-D elite from n elite points in n 1-D regression spaces will also produce good elite; the probability of the global optimum being located near to the synthesized elite is high. In general, this proposed method represents a novel local search approach for accelerating EC convergence, and it is this approach that represents this paper's contribution.

III. Experimental Evaluations

In order to verify the performance of our proposed approaches with EC, we chose 10 benchmark functions as our test suite. In our experiments, we compare our proposed acceleration approaches by Lagrange interpolation and least squares approximation with a normal EC and the previous acceleration approaches [2]. The reference [2] used a quadric function to obtain the EC search space surface used to accelerate the EC in the original search space, and the quadric function is fitted to each of: the best fitness individuals, the individuals closest to the most fit individual, or all the individuals.

We used differential evolution (DE/best/1/bin) as the optimization method to evaluate the proposed approaches. Here, we abbreviate the DE approach where the search space is regressed by a two-degree Lagrange interpolation as DE-LR, where it is regressed by a line power function least squares approximation as DE-LS, where it is fitted by a quadric function with the best individuals as DE-TB, where it is fitted by a quadric function with the individuals closest to the most fit individual as DE-TN, where it is fitted by a quadric function with all individuals as DE-TA, and normal DE as DE-N. These abbreviations are also used in Figures 2, 3, and 4.

A. Experimental Settings

We used ten benchmark functions, including the DeJong five functions (F1 - F5), Rastrigin function (F6), Schwefel function 2.26 (F7), Griewank function (F8), Schaffer-1 function (F9), and Schaffer-2 function (F10). Some of these can be found in literature [4] and [6].

The dimensions and search ranges for the parameters are 3-D F1 in [-5.12, 5.12], 2-D F2 in [-2.048, 2.048], 5-D F3 in [-5.12, 5.12], 30-D F4 in [-1.28, 1.28], 2-D F5 in [-65.536, 65.536], 5-D F6 in [-5.12, 5.12], 5-D F7 in [-512, 512], 5-D F8 in [-512, 512], 2-D F9 in [-100, 100], and 2-D F10 in [-100, 100]. All these function optimization tasks are posed as minimization problems with the optimal solution being the point with the lowest value.

The landscapes have a variety of characteristics. They include both continuous and discontinuous, convex and non-convex, unimodal and multimodal, low dimensional and high dimensional shapes, variable separable and non-separable. Classified by modal and variable separable, they can be separated into four groups, i.e. four categories problems. F8 and F9 are the multimodal and variable non-separable problems; F5, F10 and F8 are the multimodal variable separable problems; F2 is the unimodal and non-separable problem; and the remaining functions, F1, F3, F4, F6 and F7 are the the unimodal variable separable problem.

We set the population size of the DE as thirty, and tested F4 and F6 with 50 trial runs of 200 generations, and the other functions with 50 trial runs of 50 generations. We applied sign tests to evaluate the performance of our proposed approach with the previous acceleration approaches and normal DE.

B. Experimental Result

Figure 2 shows the average convergence curves of the best fitness values for 50 or 200 trial runs of (DE-LR vs. DE-N), (DE-LR vs. DE-TB), (DE-LR vs. DE-TN), (DE-LR vs. DE-TA), (DE-LS vs.DE-N), (DE-LS vs.DE-TB), (DE-LS vs.DE-TN), and (DE-LS vs.DE-TA) for ten benchmark functions, and Figure 3 and Figure 4 shows their sign test results at each generation.

From these results, we can say that:
(1) Our proposed methods could accelerate all benchmark functions well, except with F3, F5 and F7.
(2) Our proposed methods did not accelerate DE convergence well for F3 F5 and F7 though there was no significant difference between normal DE and DE with either of our proposed methods (DE-LR or DE-LS).
(3) Our proposed methods' performance look similar, and if there is any difference, the superiority depends on the task being performed. In most of cases, DE using the proposed methods are better than normal DE in initial generations, i.e. when search points approach the global optimum.
(4) From the sign test comparison of our proposed approaches with previous acceleration approaches by Takagi et al, our proposed approaches performance are as the same as DE-TB, DE-TN and DE-TA in some of functions.
(5) The DE-LS approach's performance is better than the DE-LR approach's performance, compared with previous acceleration approaches, i.e. DE-TB, DE-TN and DE-TA.

IV. Discussions

We analyze the performance of our proposed methods from the EC experimental results of Figures 2, 3, and 4. Figure 5 shows their landscape shapes in 2D. In general, our proposed acceleration approaches (DE-LR and DE-LS) can accelerate EC efficiently from the experimental results.

A. Comparison of DE-(LR and LS) with DE-N

With the exception of F3, F5 and F7, our proposed approaches could accelerate all the benchmark functions. There were no cases where the proposed methods were significantly poorer than normal DE. Although our proposed acceleration approaches use a dimensionality reduction technique to obtain the new elite in a lower dimensional search space, it seems more efficient in problems that are variable separable, and the experimental result shows that DE-LR and DE-LS also have better performance in the variable non-separable cases. For the unimodal and multimodal problems, the experimental results show our proposed approaches have the same capability to accelerate convergence.

As the global optimum of F3 is on the edge of a searching space, the elite obtained around the global optimum by function approximation may be located outside of the searching range. Our experiment did not use elite in this case, and DE

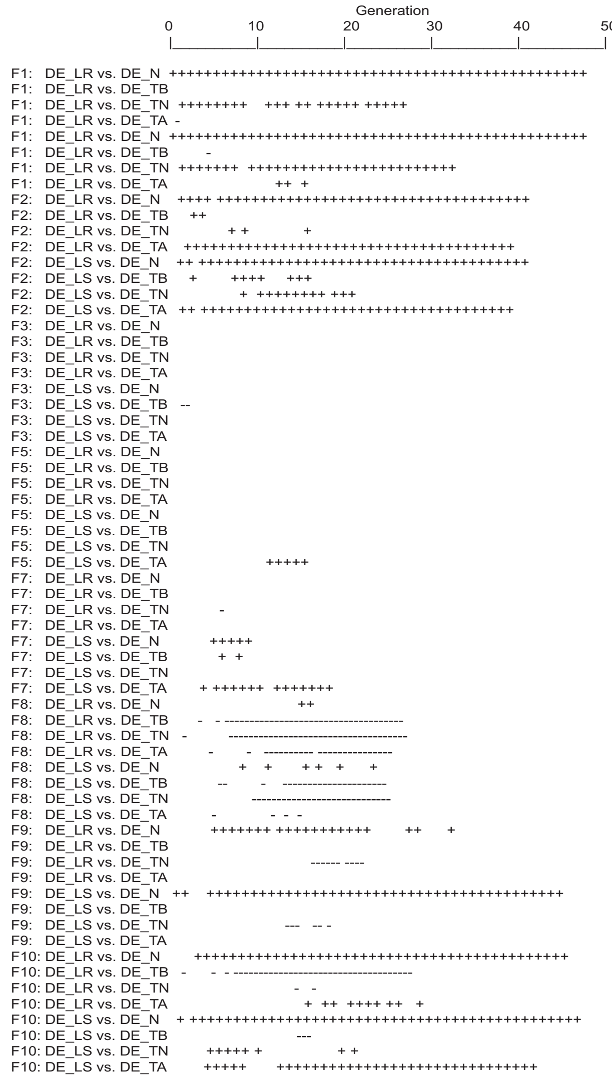

Figure 3. Sign test results for 50 trial runs of (DE-LR vs. DE-N), (DE-LR vs. DE-TB), (DE-LR vs. DE-TN), (DE-LR vs. DE-TA), (DE-LS vs.DE-N), (DE-LS vs.DE-TB),(DE-LS vs.DE-TN),and (DE-LS vs.DE-TA) per generation, where DE-N, DE-LR, DE-LS, DE-TB, DE-TN and DE-TA mean normal DE, DE with Lagrange interpolation, DE with least squares approximation, DE with best data fit, DE with nearest data fit and DE with all data fit. See F1 - F10 in Figure 2. The + mark means that a proposed method converges significantly better than the compared DE, and the - mark means that a proposed method converges significantly poorer than the compared DE ($p < 0.05$). There were no cases where the proposed methods were significantly poorer than normal DE.

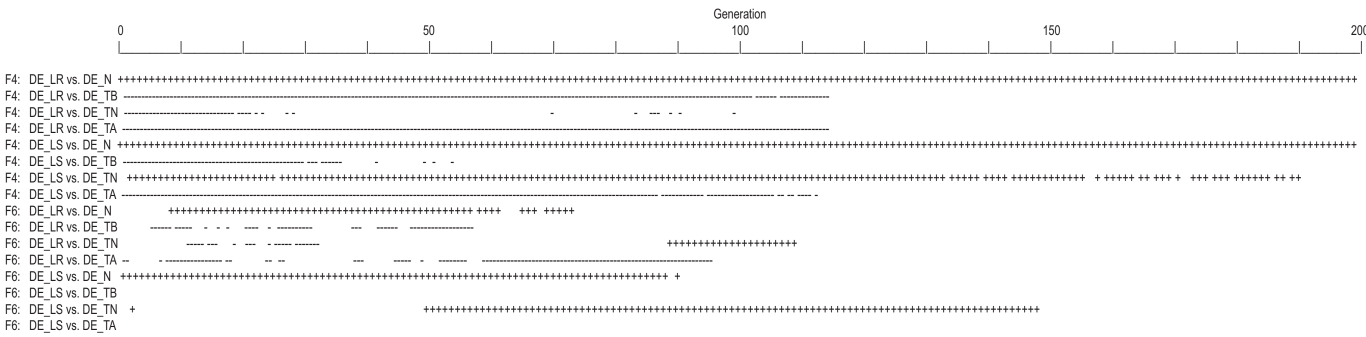

Figure 4. Sign test results of F4 and F6 for 200 trial runs of (DE-LR vs. DE-N), (DE-LR vs. DE-TB), (DE-LR vs. DE-TN), (DE-LR vs. DE-TA). See F1 - F10 in Figure 2. The + mark means that a propose method converges significantly better than the compared DE, and the - mark means that a propose method converges significantly poor than the compared DE ($p < 0.05$). There were no cases where the proposed methods were significantly poorer than normal DE.

using our proposed methods became identical to normal DE. This would explain why there was no significant difference among DE-N, DE-LR and DE-LS.

F7 is made by putting multiple sine curves on a sloped plane that is otherwise similar to F3. The difficulty with this task is that the minimum point, i.e. the global optimum, is located near the maximum point of the plane and there are local minimums around the global optimum. When local minimums and the global optimum are located close together, the elite obtained from an approximated curve may be located in a gap between these peaks and its fitness may be poor. This may be a reason for why the performance of our proposed methods was the same as with normal DE.

F4 is a quartic function with Gaussian noise and has its global optimum in the center of the searching space. Although it is a multimodal function due to the quartic function and fluctuates due to noise, their influence is relatively small and its whole shape is close to a quadratic function such as in F1. Elite from an approximated function using individuals that fluctuate slightly should be located in the center, and the performance of DE-LR and DE-LE should be better, as with F1, while that of normal DE is negatively influenced by the fluctuations. This may explain the good performance of our proposed method with F4.

B. Comparison of DE-(LR and LS) with DE-(TB, TN and TA)

From the experimental test, there is no conclusive result showing which acceleration approach is best, however the acceleration performance of DE-TA was best in most of the benchmark functions. This shows that the EC search space fitted by all individuals is a more accurate representation of the original search space, so it has a better acceleration performance. On the other hand, it is surely costly to perform the regression process for all individuals.

DE-TB has better performance than DE-LR and DE-LS in problems that were multimodal and non-separable. This indicates that when the search space is complex, especial when non-separable, DE-TB can obtain the original search space more accurately than DE-LR and DE-LS can in the reduced dimensionality space.

With the exception of the special case, the performance of DE-LR and DE-LS is better than the DE-TN. This shows that a search space fitted by those individuals that are near the best individual is not an accurate regression of the original search space. In other words, just because some individuals were near the best individual does not mean they were in the better fitness regions in the original search space.

Table I
COST TIME FOR EACH APPROACH APPLIED FOR ONE EC OPTIMIZATION (UNIT: MS), WHERE F, N, LR, LS, TB, TN AND TA MEAN FUNCTION THE APPROACHES DE-N, DE-LR, DE-LS, DE-TB, DE-TN AND DE-TA, RESPECTIVELY. THIS RESULTS DEMONSTRATES HOW OUR PROPOSED APPROACHES CAN REDUCE COMPUTATIONAL COST SIGNIFICANTLY OVER PREVIOUS APPROACHES .

F	N	LR	LS	TB	TN	TA
F1	1.88	29.06	28.44	61.88	65.32	75.00
F2	1.56	30.00	28.76	72.80	65.64	68.12
F3	2.82	30.30	33.14	42.80	67.82	87.18
F4	30.32	60.30	60.94	110.94	174.06	204.70
F5	5.62	31.88	31.88	64.68	68.12	74.70
F6	4.68	31.88	31.88	75.94	68.74	85.64
F7	5.32	32.80	31.88	79.06	69.38	86.88
F8	5.94	34.38	32.80	80	71.88	83.12
F9	3.14	30.00	29.68	62.82	63.12	68.44
F10	2.82	30.00	30.30	73.44	70.00	71.88

C. Computational Complexity

Reducing computational cost is the biggest feature of our proposed methods. This approach of approximating landscapes and using the peak of the approximated function is in the same category as the reference [2], but we extended the work by introducing the projection onto a 1-D space, synthesizing an elite from the elites on multiple 1-D functions, and thereby reducing computational cost. Let's examine the efficiency of the approaches by comparing the number of algebraic operations required.

We analyzed our approaches' computational complexity theoretically through a multi-variable regression model, i.e. $y = X\beta + \varepsilon$, where y, X, β and ε are respectively a fitness vector, a parameter vector matrix, a coefficient vector, and an error vector. The numbers of columns and rows of matrix X are decided by the search space dimension and the sampled data number that we select to obtain a regression model. These variables become scalars in 1-D space. When it is solved by least square approach to obtain the unknown coefficient vectors and by Gaussian elimination to solve system equations by M pieces of (individual, fitness) pairs of sampled data, the algebraic operations number is $O(2n^3 + 2(m-1)n^2)$ for an n-D space, while it is $O(n(2m+1))$ for our proposed method because we perform n times the number of operations required for a 1-D space. This demonstrates how our proposed methods can reduce computational cost significantly.

We also experimentally calculated the run time for 50 trials of DE-N, DE-LR, DE-LS, DE-TB, DE-TN and DE-TA, and obtained the running time for one computation (see Table I.). From the table, we can conclude that:
(1) Our proposed approaches's time (DE-LR and DE-LS) is more than the normal DE (DE-N) and less than the previous acceleration approaches (DE-TB, DE-TN and DE-TA).
(2) Our proposed approaches's time are almost the same regardless of benchmark functions.
(3) The three previous acceleration approaches (DE-TB, DE-TN and DE-TA) are costly for certain benchmark functions.

From Table I and the results above mentioned, it can be seen that our proposed approaches can significantly reduce the optimization costs posed by the previous acceleration approaches.

V. Conclusion and Future Work

We proposed that the EC landscape be approximated with a two-degree Lagrange polynomial interpolation or a linear power function least squares approximation to accelerate EC convergence. The novel feature in these acceleration methods is to use an elite synthesized from elite points in lower dimensional spaces. Our experimental evaluation of these two approaches with ten benchmark functions showed that the proposed methods can accelerate EC search especially when the landscape of the EC tasks takes on a roughly big valley structure. We also analyzed the relationship between the proposed approaches' performance with the previous acceleration approaches.

This study reveals two critical topics that should be tacled on further research of accelerating EC convergence. One is the issue of model selection; how should a proper linear or non-linear regression model be selected and constructed for a concrete application. The other issue is of dimensionality reduction information conservation; how can we reduce the search space dimension as much as possible while simultaneously preserving the original search space information? We will conduct studies on these topics in the future.

Acknowledgements

Yan Pei would like to thank Yoshida Scholarship Foundation for its support to his doctoral research.

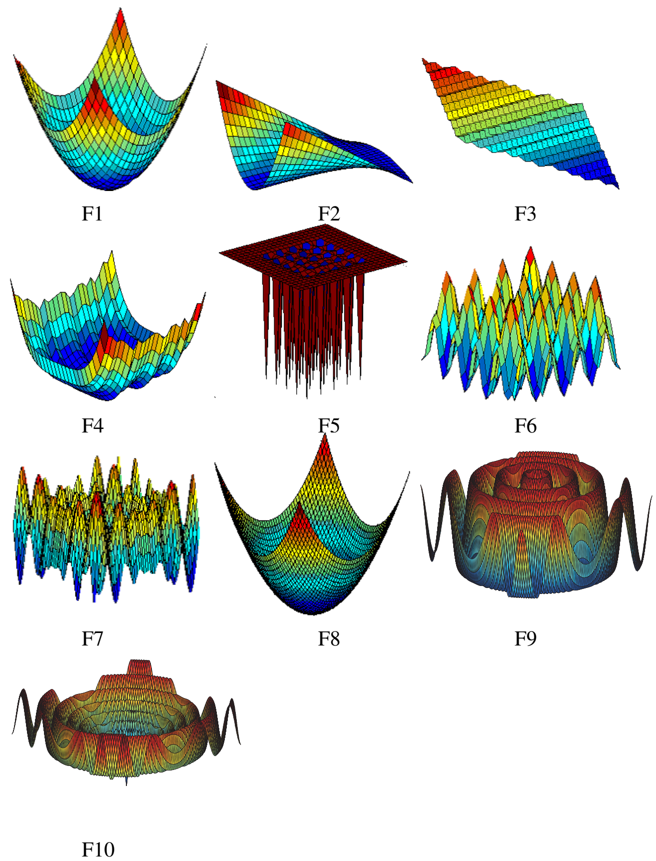

Figure 5. 2-D landscapes of (F1) De Jong Function1, (F2) De Jong Function2, (F3) De Jong Function3, (F4) De Jong Function4, (F5) De Jong Function5, (F6) Rastrigin Function, (F7) Schwefel 2.26 Function, (F8) Griewank Function, (F9) Schaffer-1 Function and (F10) Schaffer-2 Function.

References

[1] Takagi, Hideyuki, “Interactive Evolutionary Computation: Fusion of the Capabilities of EC Optimization and Human Evaluation”, Proceedings of the IEEE, Vol.89, No.9, pp.1275-1296 (2001).

[2] Takagi, H., Ingu, T., and Ohnishi, K., “Accelerating a GA Convergence by Fitting a Single-Peak Function”, J. of Japan Society for Fuzzy Theory and Intelligent Informatics, vol.15, no.2, pp.219-229 (2003) (in Japanese).

[3] Pei Y. and Takagi, H., “Accelerating Evolutionary Computation with Elite Obtained in Projected One-Dimensional Spaces”, 5th Int. Conf. on Genetic and Evolutionary Computing, Kimmen Taiwan (ICGEC2011), pp.89-92 (Aug. 29-Spt. 1, 2011).

[4] De Jong, Kenneth A., “An Analysis of the Behavior of a Class of Genetic Adaptive System”, Ph.D Dissertation, University of Michigan, USA (1975).

[5] Jin, Yaochu, “A Comprehensive Survey of Fitness Approximation in Evolutionary Computation”, Soft Computing, Springer, Vol.9, No.1, pp.3-12 (2005).

[6] Whitley, D., Mathias, K., Rana, S., and Dzubera, J., “Building Better Test Functions”, 6th Int. Conf. on Genetic Algorithms (ICGA95), Pittsburgh, PA, USA, pp.239-246 (July, 1995).