

実数値遺伝的アルゴリズムのための優性方向優先探索

田中 信* 倉田 博之** 大橋 健**

* 九州工業大学院情報工学研究科情報創成工学工学専攻

** 九州工業大学情報工学部生物化学システム工学科

現在、遺伝的アルゴリズムは、多くの分野の最適化問題で利用されている。しかし、トレードオフとして膨大な反復計算が要求される。そのため、冗長な計算負荷を軽減するために Dominant Direction Priority Searching Method(DDP Searching Method)を提案する。DDP は従来の GA 世代交代に追加されるアルゴリズムであり、勾配法の原理に基づき、母集団中の劣性個体から優性個体に向かう方向ベクトル(DD Vector)を評価値空間の勾配方向と仮定して新規探索点を設置する。この DDP Searching Method を用いて、数学的ベンチマーク関数と大腸菌熱ショック応答シミュレーションでの最適化実験を行った結果、成功確率、計算回数ともに従来手法より良好な最適化性能を示した。

Dominant Direction Priority Searching Method for Real-Coded Genetic Algorithm

Tanaka Shin* Kurata Hiroyuki** Ohashi Takeshi**

**Program of Creation Informatics Kyushu Institute of Technology

**Department of Biochemical Engineering Kyushu Institute of Technology

GA is used in optimization problems in a lot of fields. But, there is a problem that GA demands a large quantity of computational complexity. Therefore we propose DDP algorithm in order to reduce this problem. GA+DDP showed the effectiveness in optimization problems of mathematical function and gene regulatory network simulation.

1. Introduction

Real-Coded GA [1] that uses a real number value for coding in a genotype is an efficient optimization method that imitates an evolution process of a creature. And it achieves high performance for optimization for not only mathematical problems.

However, GA often demands a lot of computational complexity by repetition of generation alternation. Hybrid GA [2] uses a slant method that calculates differential values of fitness function each gene to accelerate the convergent process of the population to a local maximum. However, Hybrid GAs can be applied only to a case where fitness function is unimodal and the differential value of fitness function is able to be calculated. It cannot be employed in other cases.

In order to overcome such problems, we propose "Dominant Direction Priority Searching Method, (DDP)" that performs a slant method without calculating the differential value of fitness function, and avoids intrigued convergence in the early state of an optimization process.

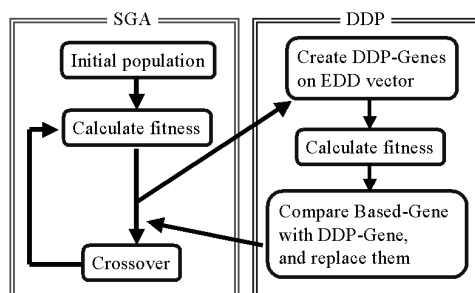


Figure 1: SGA+DDP Algorithm

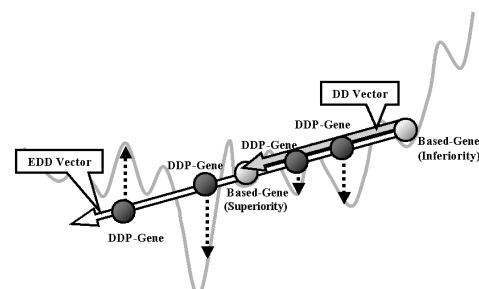


Figure 2: DDP Search

2. DDP Searching Method

DDP is added to the generation loop of a conventional GA as Figure 1. The DDP algorithm is as follows,

- Choose 2 genes (Based-Gene, B-Gene) out of the entire population.
- Calculate the vector, Dominant Direction Vector (DD Vector), which goes from one B-Gene that shows a low fitness value (inferiority) to the other B-Gene with a high fitness value

- (superiority).
- Extend t-fold (t: Extended Ratio, Ex-Ratio) the length of the DD Vector from inferior B-Gene, resulting in calculating an extended dominant direction vector (EDD Vector).
- Create m sets of new genes (DDP-Gene) at random positions on the EDD Vector.
- If the fitness of the most superior DDP-Gene is lower than the inferior B-Gene, replace them. And return the two B-Genes to population.
- Repeat the operations above k times (k: DDP Frequency) per generation.

Assuming that the DD Vector corresponds well to the inclination of a fitness function, it is likely to predict that a local maximum exists in the direction of the DD Vector. This method can be applied to optimization problems where it is hard to calculate the differential values analytically or numerically. The DDP search algorithm is expected to show high performance in unimodal functions, because it is basically derived from a slant method. In multimodal functions, however, it would be hard to find a local maximum in the direction. Thus, we create multiple DDP-genes randomly on an EDD Vector.

Combination of DDP with conventional GAs enables us to search wide areas by crossover GAs and local areas by DDP, simultaneously. Users are able to control the area searched by changing the Ex-Ratio to avoid a rapid fall into an intrigue local maximum.

In order to characterize the performance of DDP, computational complexity (CC-Value) of DDP is defined as follows,

$$\text{CC-Value of DDP} = \text{DDP-Gene} \times \text{DDP Frequency Generation}$$

3. DDP application to mathematical optimization

3-1. Benchmark function

We carried out computer simulation to optimizing five benchmark functions (Sphere, Rastrigin, Ridge, Schwefel, Rosenbrock) in order to characterize the performance of DDP. The dimension of each function is 2,5,10.

3-2. Optimization performance of SGA+DPP

We applied DDP to those functions in order to test the performance and compared it with SGA and MGG. Detailed GA condition is shown in Table 1 and Table 2.

Table 1: DDP Parameter Setting

GA Type	B-Gene Select	Ex-Ratio
(1) SGA	-	-
(2) MGG	-	-
(3) SGA+MGG	Random-Random	t=1
(4) SGA+MGG	Elite-Random	t=1
(5) SGA+MGG	Random-Random	t=2
(6) SGA+MGG	Elite-Random	t=2

Table 2: GA Parameter Setting

Number of Genes: 100 genes
Crossover: UNDX
Max CC-Value: 500000 times
DDP Frequency: 1-20 per generation
Number of DDP-Gene: 1-20 per generation
Experiment frequency: 20 trials each

In most of the cases, CC-Value and Success Ratio are taken as the trade off. Then, we define an Optimization Index (Op-Index) with the following equations. Op-Index evaluates them inclusively.

$$\text{CC-Value} = \text{Genes} \times \text{Generations}$$

$$\text{Op-Index} = \text{CC-Value} \times \text{Success Ratio}$$

The efficient optimization shows low values of both CC value and Op-Index. Result in each benchmark function is shown in Figure 3. DDP parameter setting is that the number of DDP-Gene is 5 and the DDP Frequency is 5 per generations.

In each condition of DDP, GA Type (6) showed high optimization performance. Because the direction of a DD Vector can face more precisely to the direction to a maximum. In many conditions, the addition of DPP provided higher performance than SGA and MGG+UNDX.

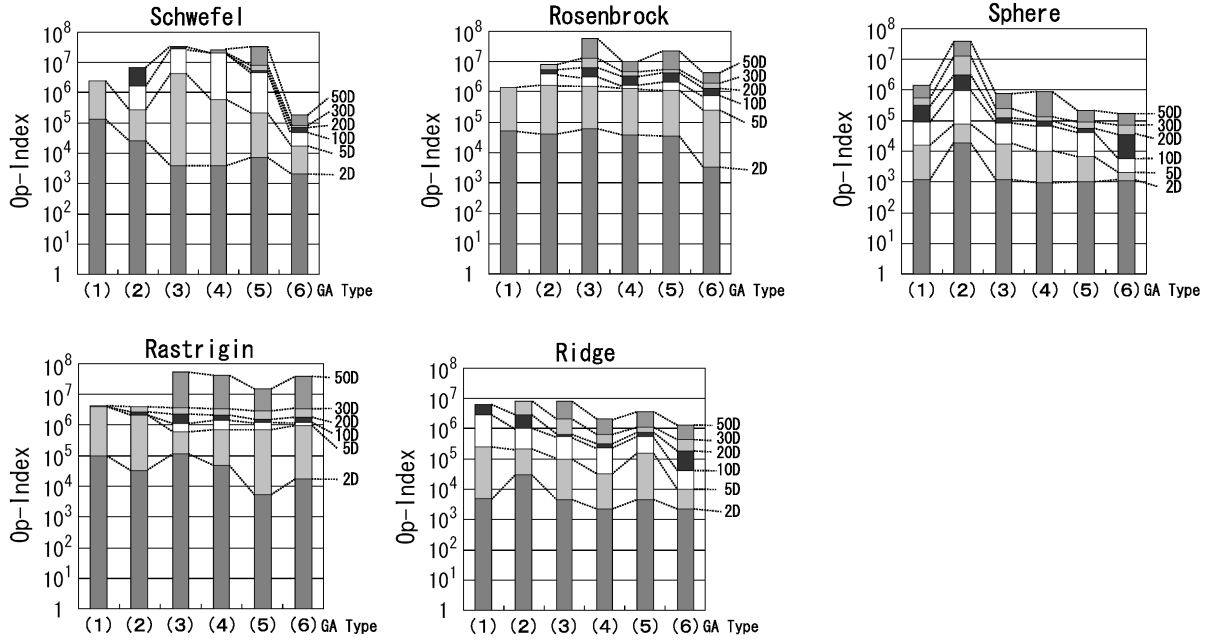


Figure 3: DDP parameter setting and Op-Index

3.3. Parameter dependence of DDP

The parameters that users determine in DDP are the DDP Frequency and the number of DDP-genes per generation. We measured the Op-Index, varying the values of these two parameters, as shown in Table 3.

When the number of DDP-genes was over 15 and the DDP Frequency was over 15, the Success Ratio decreased and the Op-Index increased, as shown in Figure 4.

On the other hand, when the DDP Frequency was from 1 to 15 and the number of DDP-genes was from 1 to 15, the optimization performance was high regardless the values of the DDP parameters, showing a high feasibility for us to determine the DDP parameters.

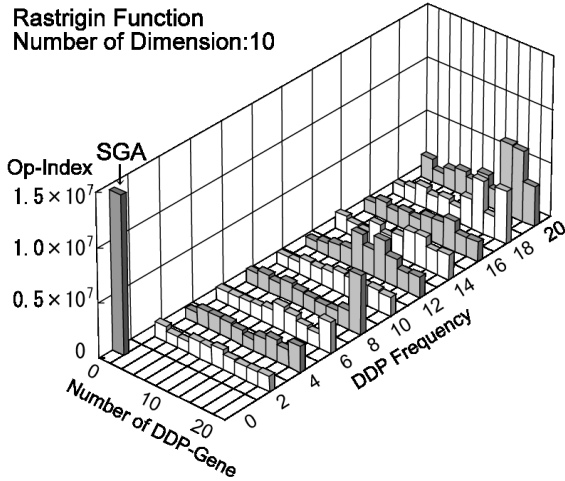


Figure 4: DDP parameter and Op-Index (Rastrigin 10D)

Table 3: GA Parameter Setting

GA Type: (6)
DDP Frequency: 1-20 per generation
Number of DDP-Gene: 1-20 per generation
Experiment frequency: 20 trials each

4. DDP application to gene regulatory network simulation

4-1. Heat shock response

We try to optimize the kinetic parameters of the E. coli heat shock response, which is a gene regulatory network, as shown in Figure 5. Heat shock response is a system to refold and degrade

heat-denatured proteins in order to prevent cell death. Heat shock response consists of multiple components and interactions among them. Time series data of some molecular components have been reported, but the values of many kinetic parameters remains to be elucidated.

Using DDP, we optimize nine kinetic parameters that express the protein binding interactions to reproduce the dynamic behavior of $\sigma 32$, DnaK, and FtsH proteins.

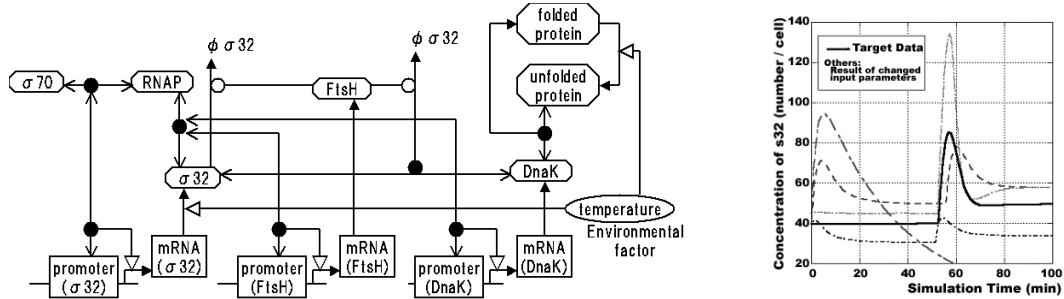


Figure 5: Heat Shock Response Reaction Model

4-2. Optimization performance

In order to demonstrate optimization performance of DDP, we compared SGA, MGG+UNDX, and SGA+DDP. Each GA condition is shown in Table 4.

Figure 6 indicates the result of the Op-Index. In both end conditions, GA Type (6) (selecting Elite - Random, $t=2$) shows higher performance than others, because DDP-Genes search the area around the EDD Vector, where a maximum may exist.

Table 4: GA Parameter Setting

Number of Genes: 100 genes
Crossover: UNDX
Max CC-Value: 500000 times
DDP Frequency: 1-10 per generation
Number of DDP-Gene: 1-10 per generation
Experiment frequency: 20 trials each
Fitness = $\{\text{Mean Squared Error}(\sigma 32)\}^{-1}$
GA END CONDITION 1: Fitness $> 10^{-5}$
GA END CONDITION 2: Fitness $> 10^{-4}$

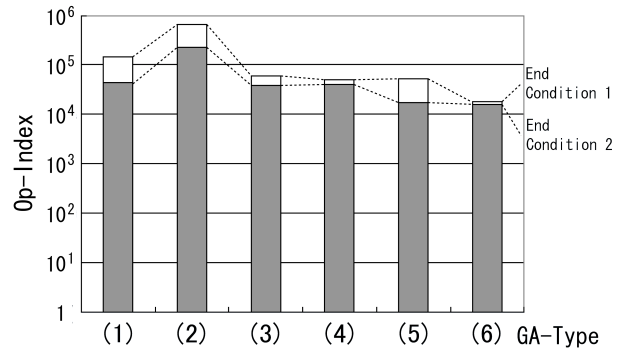


Figure 6: Op-Index in Heat Shock Simulation (DDP Frequency: 6, Number of DDP-Gene: 6)

5. Conclusions

GA+DDP algorithm greatly enhances optimizing not only several benchmark functions but also the gene regulatory network. The DDP algorithm can be applied to fitness functions where it is hard to calculate the differential values analytically or numerically. A conventional GA, which converges the population by mutation and crossover, searches for the global space, and the DDP method accelerates the search for a local maximum by generating DDP vectors. DDP is a method that accelerates the convergence to a local maximum with a high Success Ratio and with a small CC-Value. DDP has another advantage in the feasibility that it is pretty easy to determine the values of the key control parameters.

Reference

- [1] Davis, L. Genetic Algorithms in Search, Optimization, and Machine Learning. Van Nostrand Reinhold, New York, 1990.
- [2] Hiroyasu, T., Miki, M., Minami, Y., Tanimura, Y. Global Optimal-point Search of Hybrid Genetic Algorithms Using Gradient Method. Proceedings of the 5th Technical Session on Mathematical Modeling and Problem Solving, pp.57-64, 2000