Diploid GA with Real-valued vs. Binary-coded Recombination Operator in Diploid Genetic Algorithm

Bara'a Ali Attea
Department of Computer Science, College of Science, University of Baghdad, Baghdad-Iraq.

Received: 15/2/2002 Accepted: 16/6/2004

Abstract

Different evolutionary algorithms based on real coding have been presented for tracking parameter optimization problems. They include evolution strategies (ESs) and real-coded genetic algorithms (RCGAs). As the power of the GAs arises from their recombination (crossover) operator, and it may be considered to be one of the components to be borne in mind to improve the GA's behavior, researchers on evolutionary algorithms considered a new class of recombination operators- called *real-valued recombination operators.*. In this paper, we examine the eefect of real-valued recombination operator in diploid genetic algorithm, comparing its results with that of diploid genetic algorithm using bit-coding recombination operators.

المستخلص

لقد استعملت الخوارزميات التطورية ذو عامل المزج حقيقي التمثيل وعلى اختلاف انواعها في حل مشاكل امثلية المعاملات، ومن بين هذه الخوارزميات: الاستراتيجيات التطورية والخوارزميات الجينية ذات التمثيل الحقيقي. بالنسبة الى الخوارزميات الجينية فان قوة الخوارزمية تكمن في عامل الجمع او ما يسمى عامل المزج والذي من الممكن ان يعتبر جزء اساسي لتحسين سلوك الخوارزمية الجينية، وفي هذا المجال قام الباحثون بايجاد طبقة جديدة من عوامل التوحيد والتي تدعى بـ"عوامل المزج ذات القيم حقيقية التمثيل ". هذا البحث يقدم مقارنة بين كفاءة عامل المزج ذو التمثيل الثنائي في الخوارزميات الجينية ثنائية المجموعة الكروموسومية.

1. Real-Coded Genetic Algorithm: An Introduction

Genetic search operates by processing problem parameter coding rather than the parameters themselves. In natural genetics, chromosomes and the genes they contain act as a code for the features of each individual organism. In a Genetic Algorithm, GA, a finite-length string coding describes the necessary parameters for each candidate solution to the search or optimization problem at hand.

Traditionally, fixed length and binary coded strings of 1s and 0s (bit string) for the representation solution have dominated GA research since there are theoretical results that show them to be most appropriate ones, and as they are amenable to simple implementation[1]. But the GA's good properties do not stem from the use of such low cardinality bit string representation. For this reason, the path has been lain toward the use of non-binary representations more adequate for each particular application problem. One of the most important ones is the *real number representation,* which would seem particularly natural when optimization problems with variables in continuous search spaces are tackled. So a chromosome in this representation is a vector of floating point number whose size is kept the same as the length of the vector, which is the candidate solution to the problem. Genetic algorithms based on real number representation are called *real-coded GAs (RCGA)* [2].

The main difference in the implementation of binary-coded GAs and real-coded GAs is in their recombination and mutation operators [3][4][5]. On the behalf of recombination, literatures proposed several binary-coded and real-coded recombination operators. A common binary-coded recombination method is two-point crossover. In real-coded GA, on other hand, there are a number of recombination methods all of which assume that the individuals' variables are either exchanged or averaged [6][7][8].

While from the mutation viewpoint, literatures proposed one main form of mutation. In binary-coded GAs, mutation works by occasionally inverting single bits of individuals, with small probability [8]. But in real-coded GAs, mutation consists of the addition of a normally distributed random number to each variable of the individual, corresponding to a step in the search space [6].

2. Diploid Genetic Algorithms: An Overview

The typical GA (haploid GA) assumed a *haploid* representation of potential problem solutions. That is, each population member was a bit string that contained sufficient information to specify a complete solution design (detailed description for mechanism of the typical GA can be found in [9]). The analogous haploid chromosomes in natural genetics are found primarily in very simple organisms. However, more complex organisms often have diploid chromosomes, which contain *twice* the information necessary for specifying the organism's structure. Decoding this double set of genetic data into a single set of features can lead to conflicts between the homolgous chromosome strands. Conflicts or contradictions that occur between the two halves of a diploid chromosomes are resolved by a *dominance relationship*, which (in its simplest form) decides on one of the conflicting genes that is eventually expressed in the origanism itself.

In a GA, a diploid individual has two bit strings, each of which is sufficient to specify a complete solution design. A dominance relationship must be applied to specifies how these two strings are decoded into a single *expressed* string whose fitness is evaluated [9]. For instance, if we consider a dominance relationship where 1 always dominants 0 (1 is called the *dominant allele* and 0 is called the *recessive allele*) the following diploid individual,

1 0 1 1 0 1 0 1 1 0
0 0 1 0 1 0 1 1 0 0

decodes to the following *expressed string*

1 0 1 1 1 1 1 1 1 0

in the above example, there are contradictions between the two strings at loci one, four, five, six, seven, and nine. The values in the expressed string are said to be the *expressed* or *phenotypic* values associated with the diploid chromosome. This example shows that dominant allele is expressed when *heterozygous* (i.e., mixed, 10 → 1 or 01→1) or *homozygous* (i.e., pure, 11→1) and the recessive allele is expressed only when homozygous (00→0) [9]. This type of domain relation is called *Mendelian* or *complete domainance*.

Diploid GAs not only differes from haploid GAs in its methods of representing and translating genetic traits, but it also uses a different process for combining genetic information. In diploid GAs, the simple crossover operation like that found in haploid GA is replaced by a more complex, *two-stage process* [10]. In the first stage of diploid information combination, crossover occurs between the homologous chromosomes within a single individual to form single *stranded gametes*. This stage is called *gametogenesis*. In the example above, if a crossover point of four is selected:

Example1:

1 0 1 1 | 0 1 0 1 1 0
1 0 1 0 | 1 0 1 1 0 0
→crossover point

Hence, gametogenesis process yields the gametes:

1 0 1 1 1 0 1 1 0 0
and
0 0 1 0 0 1 0 1 1 0

After gametogenesis, a haploid gamete from one parent is combined with a haploid gamete from another parent to form the homologous pair for a diploid child's chromosome. This stage is called *fertilization*. Consider the following example:

Parent1	Parent2
Gamete1: 10110000	Gamete1: 01001111
Gamete2: 11101010	Gamete2: 00010101

⇩
fertilization

Offspring1	Offspring2
10110000	01001111
00010101	11101010

Diploid genetic structures with dominance have an evolutionary advantage in nonstationary environments because they can hold alternate alleles in abeyance as recessives. To illustrate how abeyant recessive can be advantageous in nonstationary environments, consider an allele μ that is related to some useful feature of a given species in a given environment. If the conditions of the environment change, and the particular feature associated with μ is no longer useful, selection tends to remove organisms with allele μ expressed. However, if μ is recessive, it usually will be shielded from selection by a dominant allele.

If the environment returns to its original state, allele μ need not be rediscovered by a lucky mutation, it will be available as an abeyant alternative in the genetic structure. In this way the abeyant allele acts as a form of probabilistic memory of past environmental conditions [10].

Probabilistic memory is not the only advantage of holding recessive alleles in abeyance. When an evolving species is suddenly faced with new environmental conditions, abeyant recessive alleles can also act as a source of diversity in the gene pool. Since recessives are only expressed in the homozygous case, the diversity they introduce is less disruptive to species development than that introduced by mutation in haploid chromosomes. If mutation is the only source of genetic diversity, the mutation rate necessary to cope with environmental changes might be so high as to stifle genetic development. Also, recessives introduced by mutations in diploid genetic structures are expressed less often than in haploid structures, thus limiting the amount that a given mutation rate disrupts a diploid species' development [9]. Figure 1 outlines a pseudo-code odiploid GA with two stages recombination scheme.

```
t=0;   // Set the generation number to zero.
Initialize (Pop(t));   //Initialize the population at random.
Evaluate (Pop(t));   //Evaluate the fitness values.
Repeat
Selection (Pop(t));   //Select better chromosomes
Two-Stage-Recombination (Pop(t))
                    // Perform gametetogenesis
                    // and fertilization
If (Mutation (Pop(t)))   // Mutate functional chromosomes
Else
Dominance-shift (Pop(t)); // Mutate modifier chromosomes
Evaluate (Pop(t));  // Evaluate the new population
  t=t+1;      // Increment generation counter
Until (t > tmax or (termination criterion TRUE))
```

Figure 1 A pseudo-code of diploid GAs with two-stage recombination scheme

Previous studies on diploid GAs have focused mainly on bit recombination or bit-coding recombination and did not present how to apply real-valued recombination on diploid GA. Hence, in the next section we will present a mapping method of diploid structure and dominance map with two-stage real-valued recombination mechanism [11].

3. Real-valued Two-stage Recombination

As presented earlier, there are two general forms of real-valued recombination that either averaged individuals variables (i.e., intermediate recombination) or exchanged individuals variables (i.e., discrete recombination). By imitating the real-valued intermediate recombination process in haploid GA [2], the intermediate recombination is applied between the two homologous chromosomes of an individual during gametogenesis process so as to produce two new gametes. Gametes for the two new offspring are produced according to the rules:

$$G_{11} = u\,P_{11} + (1-u)\,P_{12}$$
$$G_{12} = u\,P_{12} + (1-u)\,P_{11}$$
$$G_{21} = u\,P_{21} + (1-u)\,P_{22}$$
$$G_{22} = u\,P_{22} + (1-u)\,P_{21}$$

where u is a uniform random number belonging to the interval (0,1). G_{11} and G_{12} are gamete1 and gamete2 of parent1 respectively, while G_{21} and G_{22} are gamete1 and gamete2 of parent2 respectively. P_{11} and P_{12} are respectively chromosome1 and chromosome2 of parent1 while P_{21} and P_{22} are respectively chromosome1 and chromosome2 of parent2. Each variable in G_{11} and G_{12} are the results of combining the variables in the chromosome1 and chromosome2 of parent1 according to the above expressions. While each variable in G_{21} and G_{22} are the results of combining the variables in the chromosome1 and chromosome2 of parent2 according to the above expressions. After creating new gametes from two parents, fertilization occurs between two haploid gametes of the two parents to form two new diploid offspring chromosomes. As an example, consider the following two diploid

Example2:

Parent1: 2.2 4.0 1.4 3.8
1.5 3.4 2.8 2.5
Parent2: 4.8 3.1 1.8 2.0
3.6 2.3 4.6 2.2

When $u = 0.75$, gametogenesis yields the gametes:

$Gamete_{11}$: 2.03 3.85 3.15 3.48
$Gamete_{12}$: 1.68 3.55 2.45 2.83
$Gamete_{21}$: 4.5 2.9 2.5 2.05
$Gamete_{22}$: 3.9 2.5 3.9 2.15

Then fertilization yields the offspring:

Offspring1: 2.03 3.85 3.15 3.48
3.9 2.5 3.9 2.15
Offspring2: 4.5 2.9 2.5 2.05
1.68 3.55 2.45 2.83

Also for discrete recombination, the two-stage recombination process is modified. When discrete recombination is applied in gametogenesis process, each variable (composed of functional gene and modifier gene) of the created gamete is produced by choosing either the variable of the first or the second mate of the same parent with equal probability. The fertilization process occurs after gametogenesis process where a haploid gamete from one parent is combined with a haploid gamete from other parent to produce homologous pair of a new diploid offspring chromosome. Consider the following example:

Example3:

Parent1: 4.4 6.2 1.0
2.0 8.6 4.0
Parent2: 6.0 10.8 2.6
8.1 4.8 1.2

gametogenesis yields the gametes:

$Gamete_{11}$: 4.4 8.6 1.0
$Gamete_{12}$: 2.0 6.2 4.0
$Gamete_{21}$: 6.0 4.8 2.6
$Gamete_{22}$: 8.1 4.8 1.2

Then fertilization yields the offspring:

Offspring1: 4.4 8.6 1.0
8.1 10.8 1.2
Offspring2: 6.0 4.8 2.6
2.0 6.2 4.0

Compare the stages involved in real-valued recombination of both Example2 and Example3 with that of binary-coded diploid recombination scheme presented in Example1.

4. Results

A sample of four minimization functions (Sphere function (F1), Rastrigin function (F2), Ackely function (F3), and Griewangk function (F4)) taken from the litretures shown in table 1 is used as the test bed to carry out the performance results of applying real-valued intermediate recombination against binary-coded crossover but with the diploid GA using two-stage recombination in an attempt to confirm the results of the previous studies [4][6]. All these functions are tested with the medium size of function dimension i.e., $n = 10$. Range of x_i together with precision Δx_i are also presented in tab1.

Table 1 test Suite

Function formulation	**Range of x_i**	Δx_i
$F1 = \sum_{i=1}^{n} x_i^2$	[-5.12, 5.11]	0.01
$F2 = 10*n + \sum_{i=1}^{n} (x_i^2 - 10*\cos(2\pi * x_i))$	[-5.12, 5.11]	0.01
$F3 = -20*\exp(-0.2*\sqrt{\frac{1}{n}*\sum_{i=1}^{n} x_i^2} - \exp(\frac{1}{n}*\sum_{i=1}^{n}\cos(2\pi * x_i)) + 20 + e$	[-32.768, 32.768]	0.001
$F4 = 1 + \sum_{i=1}^{n} x_i^2 / 4000 - \prod_{i=1}^{n} \cos(x_i / \sqrt{i})$	[-600.0, 600.0]	1.0

Simulation conditions were as follows: population size was 75 individuals. The standard tournament selection procedure was used. For all functions, each variable is represented by thirty bits gray encoding. Hollestien's dominance relationship was used in the two algorithms [10]. Recombination occurs with recombination probablity, $p_c = 0.75$. Mutation operates at the bit-level occurs with mutation probability, $p_m = 0.001$. In addition, the diploid GAs simulate dominance shift with a specified dominance shift probabilitiy, $p_s = 0.01$.

Tables 2 to 5 compare the performance of diploid GA with two-stage recombination using intermediate recombination scheme with $u = 0.5$ against that using two-point recombination scheme. In the figures below, the diploid GA using two- point crossover and intermediate recombination are denoted respectively as DGA/2pointX, and DGA/intermediateX.

Table 2 Test comparison of diploid GAs on F1

Gen. No.	DGA/2point X	DGA/ intermediateX
25	7.1428E+00	8.2007E-02
50	3.5173E+00	2.7436E-02
75	2.7785E+00	1.8479E-02
100	2.3041E+00	1.2316E-02
125	1.9593E+00	7.3647E-03
150	1.7081E+00	4.8910E-03
175	1.4860E+00	4.1246E-03
200	1.3137E+00	1.5942E-03
225	1.2525E+00	7.0479E-04
250	1.0602E+00	5.6384E-04
275	1.0419E+00	4.7534E-04
300	8.9227E-01	4.2919E-04

Table 3 Test comparison of diploid GAs on F2

Gen. No.	DGA/2pointX	DGA/ intermediateX
25	6.2796E+01	2.1207E+01
50	3.6411E+01	1.3837E+01
75	2.6468E+01	1.0655E+01
100	2.1309E+01	8.8663E+00
125	1.8298E+01	7.7389E+00
150	1.6628E+01	7.0946E+00
175	1.5316E+01	6.4687E+00
200	1.4545E+01	6.3562E+00
225	1.4075E+01	6.0687E+00
250	1.3905E+01	6.2231E+00
275	1.3619E+01	5.8052E+00
300	1.3554E+01	5.9723E+00

Table 4 Test comparison of diploid GAs on F3

Gen. No.	DGA/2pointX	DGA/ intermediateX
25	1.4062E+01	3.3942E+00
50	1.1264E+01	1.9983E+00
75	1.0185E+01	1.5195E+00
100	9.5892E+00	1.4245E+00
125	9.2602E+00	1.3134E+00
150	9.0717E+00	1.2129E+00
175	8.5846E+00	1.1254E+00
200	8.2251E+00	1.0445E+00
225	7.8661E+00	1.0018E+00
250	7.4555E+00	9.7368E-01
275	7.0419E+00	9.4263E-01
300	6.8551E+00	9.2670E-01

Table 5 Test comparison of diploid GAs on F4

Gen. No.	DGA/2pointX	DGA/ intermediateX
25	2.5241E+01	1.1733E+00
50	1.3980E+01	7.8035E-01
75	1.1352E+01	6.1675E-01
100	9.5147E+00	4.8948E-01
125	8.0025E+00	4.2160E-01
150	6.8492E+00	3.5437E-01
175	6.2777E+00	3.1467E-01
200	5.5149E+00	2.8459E-01
225	5.0895E+00	2.4476E-01
250	4.6677E+00	2.0809E-01
275	4.5506E+00	1.9620E-01

5. Conclusion

One may expect that DGA/intermediateX should outperform that of DGA/2pointX. The above results demonstrate the success of DGA/intermediateX in finding the global minimum in both F1 and F4. While DGA/2pointX does not success in finding the global minimum in all these functions. From this we can add to our arsenal the following. Similar to the conclusion drawn from the previous researches results that real-valued recombination overcomes binary-coded recombination schemes in numerical optimization problems, there is a difference in performance here between the two schemes. In other words, the diversity provided by the diversity prolonging mechanisms of both diploidy and dominance operators is not sufficient enough to give reliable results for optimization problems with real-valued parameters. And it is then approperiate to adjust other genetic operators other than diversity-prolonging operators to handle real-valued parameter optimization problem. However, the failing in the other two functions (i.e., F2 and F3) does not indicate that real-valued recombination is not with any apparent advantages in this domain but it most probably turns this unpleasant phenomenon to the mechanism of combining information used in the diploid GA with two-stage recombination mechanism through gametogenisis and fertilization.

References

1.Goldberg, D. E. **(1991)** *Real-Coded Genetic Algorithms, Virtual Alphabets, and Blocking, In Complex Systems,* Vol. 5, No. 2, pp.139-168.

2. Herrera, F. and Lozano, M. **(1997)** Gradual Distributed Real-Coded Genetic Algorithms, Technical Report, University of Granada, #DECSAI-97-01-03, February.
3. Wright, A. **(1991)** *Genetic Algorithms for Real Parameter Optimization, Proceedings of the Foundations of Genetic Algorithms Workshop*, ed G. Rawlins (San Morgan Kaufmann), pp.205-218.
4. Deb, K. and Kumar, A. **(1995):** *Real-coded Genetic Algorithms with Simulated Binary Crossover: Studies on Multimodal and Multiobjective Problems Complex Systems,* Vol.9, pp. 431 – 454.
5. Vavak, F., Fogarty, T. C. and Jukes, K. **(1996):** *A Genetic Algorithm with Variable Range of Local Search for Tracking Changing Environments, Parallel Problem Solving from Nature-PPSN III,* eds. H. Voigt, W. Ebeling, I. Rechenberg And H. Schwefel, Berlin, Springer-Verlage, September, pp.376-387.
6. Bäck, T. and Schwefel, H.-P. **(1993):** *An Overview of Evolutionary Algorithms for Parameter Optimization. Evolutionary Computation*, Vol.1, No.1, Spring 1993.
7. Sullivan, J. and Pipe A. **(1996):** *An Evolution Strategy for On-line Optimization of Dynamic Objective Functions, Parallel Problem Solving from Nature-PPSN III,* eds. H. Voigt, W. Ebeling, I. Rechenberg And H. Schwefel, Berlin, Springer-Verlage, September, pp.781-790.
8. Goldberg, D. E. **(1989):** Genetic Algorithms in Search, Optimization, and Machine Learning, Addison Wesley.
9. Smith, R. E. and Goldberg, D. E. **(1992):** *Diploidy and Dominance in Artificial Genetic Search, Complex Systems,* Vol. 6, pp.251-285.
10. Smith, R. E. **(1988):** *An Investigation of Diploid Genetic Algorithms for Adaptive Search of Nonstationary Functions, M.Sc.Thesis,* Alabama, Tuscaloosa.
11. Abdul Halim, M. F. **(2002):** *An Alternative Recombination Scheme for Diploid Genetic Algorithms, Ph.D. Thesis, University of Technology.*