Another Look at Diploidy in Genetic Algorithms

Alden H. Wright
Computer Science Dept.
The University of Montana
Missoula, MT 59812-1008
wright@cs.umt.edu

Garrett L. Bidwell
Bolt Beranack and Neuman
Boston, MA
gbidwell@bbn.com

ABSTRACT

1 Introduction

The cannonical GA is based on a single-stranded haploid chromosome. In this simple model, a single-stranded string contains all of the problem-related information in a binary encoding. However, most of the haploid organisms in the natural world tend to be rather uncomplicated lifeforms. Most animal and plant species rely on a diploid chromosome, which consists of one or more pairs of homologous chromosomes, each containing information for the same functions. When chromosome pairs have different values (or alleles) at the same locus, dominance usually resolves the conflict by allowing the dominant allele to take precedence over the recessive allele. Although this seems redundant, there are distinct advantages to a diploid scheme.

One of the advantages of diploidy is that it allows a wider diversity of alleles to be kept in the population over time. Currently harmful, but potentially useful genetic information can be maintained in a recessive position, shielded by the dominance operator. A persistent problem with haploid GAs is premature convergence to a suboptimal solution. Once a population loses its diversity, the crossover operator loses its effectiveness, since crossover of almost identical chromosomes produces little in the way of new genetic material. Only mutation remains to explore new search-space, and this performs an unsatisfactorily slow random search. In terms of a binary string, convergence is manifested as a particular bit or bits being essentially fixed. Biologists refer to this as permanent allele loss. The diploid chromosome offers a means of retaining recessive alleles. These alleles can again become useful if dominance is allowed to evolve. Goldberg and Smith [5] hypothesized that a diploid chromosome with evolving dominance can be used to infuse a form of "long-term distributed memory" into the GA by permitting old solutions to be carried along, (but not expressed), and rapidly reinstated if it becomes desirable in the context of the current environment to do so. Biological studies such as Fisher's [3] have indicated that dominance evolves in diploid and polyploid plant and animal species, giving them the ability to adapt more readily to changing environments. The intriguing implication is that a dominance shift can produce a rapid, directed change in an organism's phenotype—its external manifestation or behavior—not possible through simple mutation. Applied to GAs, this could provide a mechanism for enhancing exploration or, in the case of a non-stationary problem, reintroducing once useful alleles that have again become useful.

2 A Diploid Viability Model

To see how diploidy differs from haploidy, it is useful to compare their respective viability models.

For the diploid case, population geneticists such as Hartl and Clark [7] have presented a simple viability selection model that conveniently explores selection-based behavior of a population despite the many complexities introduced by fitness. The model makes the following assumptions:

1. a diploid organism
2. non-overlapping generations
3. infinite population size
4. viability selection only
5. random mating
6. no mutation

It is informative to describe the steps of the model in terms of the stages in the lifecycle of a diploid organism. We begin with the gamete phase, a biological example of which is sperm or egg. This is a haploid phase, because the gametes each contain only half of the genetic information of a diploid individual. The remaining phases are all diploid phases, and they are much more conspicuous and are of greater duration than the gamete phase. Upon fertilization, we reach the zygote phase. The organism then undergoes growth and development to reach the adult phase. It is during the transition from zygote to adult that proportional selection acts, based on the differential viabilities of the genotypes. The stages of the diploid lifecycle are summarized pictorially in Figure 1.1.

Perhaps the simplest example is the one-locus, two-allele viability selection model. Let 0 and 1 denote the alleles. Let x denote the frequency of 0, $1-x$ the frequency of 1. The random mating assumption gives x^2 as the frequency of the zygote 00, $2x(1-x)$ as the frequency of 01, and $(1-x)^2$ as the frequency of 11. Note that a genotype of 10 is equivalent to 01. Let the relative fitnesses (or viabilities) of 00, 01, and 11 be f_{00}, f_{01}, and f_{11} respectively, so that the zygotes survive in the ratio f_{00}:f_{01}:f_{11}. The resulting ratio of 00:01:11 among adults is

$$f_{00}x^2 : 2f_{01}x(1-x) : f_{11}(1-x)^2.$$

The sum of these terms represents the average fitness of the population and is denoted by

$$\bar{f} = f_{00}x^2 + 2f_{01}x(1-x) + f_{11}(1-x)^2.$$

To obtain the gametic frequencies for the next generation, each of the terms in the above ratio must be normalized so that the frequencies sum to 1. This is accomplished by dividing by the average fitness. Thus, the frequency x' of the gamete 0 in the next generation is given by

$$x' = \frac{f_{00}x^2 + f_{01}x(1-x)}{\bar{f}} \tag{1}$$

Note that the coefficient 2 associated with 01 frequencies until this point has been lost, because 01 heterozygotes produce half 0 and half 1 gametes due to Mendelian segregation. From the above recurrence,

Figure 1: A generalized diploid lifecycle

it is clear that there are equilibrium points at $\hat{x} = 1$ and $\hat{x} = 0$. These fixed points are of little interest, however, since they exhibit directional selection, where one of the alleles has been completely eliminated.

When the heterozygote fitness is superior to that of both of the homozygote fitnesses, i.e. $f_{00} < f_{01} > f_{11}$, we have a condition known as overdominance. For this case, there is a third equilibrium point, which is polymorphic (because there is some non-zero fraction of each allele) and globally stable. Nagylaki [17] gives a formal proof of this. Figure 1.2 is a plot of x' versus x for fitness values of $f_{00} = 0.6$, $f_{01} = 1$, and $f_{11} = 0.3$.

To contrast this with the haploid case, a single step of the single-locus Simple Genetic Algorithm as described in Vose [19] with zero mutation (and no crossover) is outlined. Let the initial population vector be $\mathrm{x} = [x\ (1-x)]^T$, and the fitness vector be $[f_0\ f_1]^T$. Begin by performing a proportional selection step according to the fitness function defined in [19]. This yields

$$\mathcal{F}(\mathrm{x}) = \begin{bmatrix} \frac{f_0 x}{f_0 x + f_1(1-x)} \\ \frac{f_1(1-x)}{f_0 x + f_1(1-x)} \end{bmatrix}$$

Next, this vector is subjected to the recombination function $\mathcal{M}$. This gives

$$\mathrm{x}' = \mathcal{M}(\mathcal{F}(\mathrm{x})) = \begin{bmatrix} \frac{f_0^2 x^2 + f_0 f_1 x(1-x)}{[f_0 x + f_1(1-x)]^2} \\ \frac{f_1^2 (1-x)^2 + f_0 f_1 x(1-x)}{[f_0 x + f_1(1-x)]^2} \end{bmatrix}$$

The next generation frequency x' of 0 is

$$x' = \frac{f_0 x[f_0 x + f_1(1-x)]}{\bar{f}} = \frac{f_0 x}{f_0 x + f_1(1-x)}, \tag{2}$$

Figure 2: A plot of x' vs. x for the diploid model

where $\bar{f} = [f_0 x + f_1(1-x)]^2$. It is not hard to see that this recurrence can only have fixed points at 0 and 1.

Mathematically, the recurrence for the diploid model (equation 1.1) is the ratio of degree 2 polynomials, whereas the recurrence for the haploid model (equation 1.2) is the ratio of linear polynomials, which gives the diploid model inherently greater complexity. Biologically, the overdominant polymorphism of the diploid model is one of the basic mechanisms for maintaining genetic diversity in a population, and it has no analogue in the haploid model.

3 Diploidy With Varying Heterozygote Fitness

In an abstract sense, dominance is a function that maps genotypes to phenotypes. Consider a simple example, where A denotes a dominant allele and a a recessive allele. Then the following dominance map shows how to map a diploid genotype to a phenotype.

	A	a
A	A	A
a	A	a

At a given locus of the diploid chromosome, $AA \to A$, both Aa and $aA \to A$, and $aa \to a$. The dominance operator can be used in the context of GAs as a means of mapping a diploid chromosome to

a haploid chromosome, which in turn can be subjected to a haploid fitness function. However, we do modify the fitness function to introduce a "heterozygote advantage". By endowing heterozygotes with a higher fitness than homozygotes, we hope to simulate the overdominant case of the single-locus diploid viability model. As the diploid genome is mapped to a haploid genome, the number of heterozygous loci is recorded. The resulting haploid genome is subjected to the fitness function, just as is done in the cannonical GA. However, for each heterozygous locus, we add a small fitness bonus to the value obtained from the haploid fitness function.

We use a scheme with two alleles, 0 and 1. Let p be the frequency of allele 0, and q the frequency of allele 1, where necessarily $p + q = 1$. We construct the following dominance map.

	0	1
0	0	0 : p 1 : 1-p
1	0 : p 1 : 1-p	1

Here, the genotypes 01 and 10 map to 0 with probability p and 1 with probability $1 - p$, where p is defined as the frequency of allele 0 at the locus under consideration in the current generation.

A diploid organism goes through a life cycle that includes both a haploid stage and a diploid stage. For our purposes, these stages can be summarized as follows:

1. Adult: Diploid genome.
2. Gamete: Gametes are sperm and egg cells which have a haploid genome.
3. Zygote: The zygote is formed by the union of two gametes, and the two haploid genomes are merged to create the diploid genome.

It should be emphasized that crossover (recombination) occurs during the transition from the adult stage to the gamete stage. If there was no crossover, each gamete would receive an intact chromosome from the adult parent. With crossover, each chromosome of the gamete may be a mixture of the homologous chromosomes from the adult parent.

In biological terminology, the term locus refers to the position of a gene on its chromosome, where the genes can be assumed to be in linear order on the chromosome. The term allele refers to the configuration of a gene. For a binary-coded genetic algorithm, a locus corresponds to a bit-position, and an allele at that locus corresponds to a value (0 or 1) of the corresponding bit.

We will restrict ourselves to what the biologists call viability selection. Under this assumption, selection occurs entirely during the transition from the zygote stage to the adult stage.

To illustrate the difference between the haploid and diploid models, we compare the standard 1-locus diploid model of population genetics (infinite population and no mutation) to the corresponding 1-bit genetic algorithm model.

The population genetics model is taken from [17], but the same material is covered in most books on population genetics. Let the alleles be denoted by A_i; the diploid individuals (the zygote to adult stages) will have genotypes A_iA_j. Mendel's law of Segregation tells us that a gamete from an A_iA_j individual is equally likely to carry A_i and A_j. Let P_{ij} denote the relative frequency of ordered A_iA_j adult genotypes, i. e., P_{ii} is the frequency of A_iA_i homozygotes, and $2P_{ij}$, where $i \neq j$, is the frequency of unordered A_iA_j heterozygotes. We assume that $P_{ij} = P_{ji}$, and that $\sum_{i,j} P_{i,j} = 1$. Then the allele frequencies are given by

$$p_i = \sum_j P_{ij} \quad .$$

Mendel's law of segregation says that p_i is the freuquency of A_i in the gamete stage. If we assume random mating, (i. e., mating occurs without regard to genotype at this locus), and if we temporarily assume no selection, then the random union of gamates gives the genotypic proportions

$$P'_{ij} = p_i p_j$$

in the next generation. Therefore gene frequencies do not change,

$$p'_i = \sum_j P'_{ij} = p_i \quad ,$$

and the Hardy-Weinberg law holds after a single generation:

$$P'_{ij} = p'_i p'_j \quad .$$

Now we assume selection and only two alleles 0 and 1 at the locus in question. Then there are three possible unordered geneotypes, 00, 01, and 11, and we assume that these have fitnesses $1-r$, 1, and $1-s$ respectively. We start with the gamete stage with with a frequency of p for allele 0 and $1-p$ for allele 1. Then the zygote frequencies are p^2 for genotype 00, $2p(1-p)$ for unordered genotype 01, and $(1-p)^2$ for genotype 11. The adult frequencies are $P_{00} = (1-r)p^2/\overline{P}$ for genotype 00, $P_{01} = p(1-p)/\overline{P}$ for ordered genotype 01, $P_{10} = p(1-p)/\overline{P}$ for ordered genotype 10, and $P_{11} = (1-s)(1-p)^2/\overline{P}$ for genotype 11, where $\overline{P} = (1-r)p^2 + 2p(1-p) + (1-s)(1-p)^2$. Thus, the next generation gamete frequency of the 0 allele is given by

$$p' = \mathcal{G}(p) = \frac{(1-r)p^2 + p(1-p)}{(1-r)p^2 + 2p(1-p) + (1-s)(1-p)^2} = \frac{p - rp^2}{1 - s + 2sp - (r+s)p^2} \tag{3}$$

It is not hard to check that this transformation has fixed points at $p = 0$, $p = 1$, and $p = \frac{s}{r+s}$. Clearly, the fixed point is $p = \frac{s}{r+s}$ is meaningful when $r \geq 0, s \geq 0$ and when $r \leq 0, s \leq 0$. It is not hard to check that the fixed point is stable when $r > 0$ and $s > 0$ (since the derivative $\mathcal{G}'(p) < 1$ at the fixed point), and unstable when $r < 0$ and $s < 0$.

The case when $r > 0$ and $s > 0$ is called overdominance, since the heterozygote has higher fitness than either of the homozygotes. In this case, the fixed point is globally stable (see [[17]]). The equilibrium is called polymorphic since there are at least two alleles present. Thus, modulo finite population effects, variation is maintained at the locus indefinitely.

In contrast, in the haploid genetic algorithm model (again with no mutation), there are two genotypes, 0 and 1, and the relative frequency of the higher fitness genotype converges to 1. To be more specific, let p be the frequency of 0 in the current generation, and r and 1 be the fitnesses of 0 and 1 respectively. The recurrence that corresponds to equation (3) is

$$p' = \frac{rp}{1 + (r-1)p} \tag{4}$$

This recurrence clearly has fixed points only at 0 and 1, the fixed point at $p = 0$ is globally stable if $r < 1$, and the fixed point at $p = 1$ is globally stable if $r > 1$.

For more than one locus, the situation gets considerably more complicated. For example, Hastings has exhibited (numerically) cycling behavior for the 2-locus 2-allele diploid model, whereas cycling behavior has never been observed for the haploid model for any number of loci (even with mutation). Karlin and Liberman [?] have shown that the Hardy-Weinberg equilibrium is globally stable under additive selection and positive multilocus recombination. In other words, if no loci are "completely linked" under recombination, and if there is no epistatic fitness interaction between loci, then the equilibrium constructed by taking the equilibria given above in each coordinate is globally stable. For the hapliod case used in genetic algorithms, it is conjectured that with zero mutation, there are no polymorphic equilibria.

4 Using Diploidy to Maintain Polymorphism

References

[1] Beasley, David, Bull, David R., and Martin, Ralph R. "An Overview of Genetic Algorithms: Part 1, Fundamentals", University Computing, 15(2) pp. 58-69, 1993.

[2] Beasley, David, Bull, David R., and Martin, Ralph R. "An Overview of Genetic Algorithms: Part 2, Research Topics", University Computing, 15(4) pp. 170-181, 1993.

[3] Fisher, R.A. The Genetical Theory of Natural Selection, second edition. New York: Dover Press, 1958.

[4] Gillett, P. Calculus and Analytic Geometry, second edition. Lexington, MA: D.C. Heath and Company, 1984.

[5] Goldberg, David E. and Smith, Robert E. " Nonstationary Function Optimization Using Genetic Algorithms with Dominance and Diploidy", Proceedings of the Second International Congress on Genetic Algorithms, pp. 59-68, July, 1987.

[6] Golberg, David E. Genetic Algorithms in Search, Optimization, and Machine Learning. Reading, MA: Addison-Wesley, 1989

[7] Hartl, D. L. & Clark, A. G. Principles of Population Genetics, second edition. Sinauer Associates, Inc., 1989.

[8] Hollstien, R.B. "Artificial Genetic Adaptation in Computer Control Systems", Dissertation Abstracts International, 32(3), 1510B, 1971.

[9] Hunter, A. "Introduction to Genetic Algorithms", Sugal V2.0 User Manual, Section 2, University of Sunderland, England, 1995.

[10] Karlin, S. and Liberman, U. "The Two-Locus Multi-Allele Additive Viability Model", Journal of Mathematical Biology, 5, 201-211, 1978.

[11] Karlin, S. and Liberman, U. "Representation of Nonepistatic Selection Models and Analysis of Multilocus Hardy-Weinberg Equilibrium Configurations", Journal of Mathematical Biology, 7, 353-374, 1979.

[12] Karlin, S. and Liberman, U. "Global Convergence Properties in Multilocus Viability Selection Models: The Additive Model and the Hardy-Weinberg Law", Journal of Mathematical Biology, 29, 161-176, 1990.

[13] Kingman, J.F.C. "A Mathematical Problem in Population Genetics", Proc. Cambridge Phil. Soc., 57, 574-582, 1961.

[14] Lancaster, P. and Tismenetsky, M. The Theory of Matrices, second edition. Academic Press, 1985.

[15] Lewontin, R.C., Ginzburg, L.R., and Tuljapurkar, S.D. "Heterosis as an explanation for large amounts of genic polymorphism", Genetics, 88, 149-170, 1978.

[16] Mitchell, M. An Introduction to Genetic Algorithms. MIT Press, 1996.

[17] Nagylaki, T. Introduction to Theoretical Population Genetics. Springer-Verlag, 1992.

[18] Ng, Khim Peow and Wong, Kok Cheong "A New Diploid Scheme and Dominance Change Mechanism for Nonstationary Function Optimization", Proceedings of the Sixth International Congress on Genetic Algorithms, pp. 159-167, July, 1995.

[19] Vose, M. D. The Simple Genetic Algorithm: foundations and theory. MIT Press, (to appear).

[20] Vose, M. D. "Formalizing Genetic Algorithms", Proc. IEEE wksp. on G.A.s, N.N.s, & S.A. applied to problems in Signal & Image Processing, Glasgow, U.K., May 1990.