

Population Diversity and Fitness Measures Based on Genomic Distances

University of Virginia Computer Science Department

Technical Report TR-99-24

Allen L. Barker and Worthy N. Martin

July 31, 1999

Abstract

We define a class of genetic algorithms where, at each time step, two parents are selected to produce a child which then replaces one member of the population at the next time step. We consider the finite-population case. We define a general crossover and mutation operation, as well as a genomic distance between individuals. We require a specific property to hold for such operations and distance functions, and present examples of crossover operations, mutation operations, and distance functions which meet the requirements. We then define the sum over all pairwise population distances as a measure of the diversity of a population and consider the time evolution of the expected diversity of a population. We show that under uniform, independent selection of parents and the individual to be replaced the expected diversity is strictly decreasing. For this case we calculate an explicit formula for the diversity at each time step, based only on the initial population diversity. We then consider and discuss the case where independent, fitness-based selection is used and show that the expected diversity is strictly decreasing whenever the same probability function is used to select both the parents and the individual to be replaced. We qualitatively discuss conditions where expected diversity will increase rather than decrease with fitness-based selection. Finally, we discuss fitness measures based on a distance function.

1 Introduction

In this paper we consider the diversity of finite populations of individuals under a genetic algorithm. The type of genetic algorithm we consider selects two parents at each time step, as well as an individual to be replaced. The population at the next time step consists of the same population except that the individual selected

for replacement is replaced by a child of the two parents. Similar techniques should also be applicable to other definitions of genetic algorithms.

The diversity measure we define is based on a general, additive genomic distance between individuals. We analyze the dynamics of this diversity measure under the specific class of genetic algorithms we consider. In certain cases, like uniform selection without mutation, the expected diversity measure is strictly decreasing, causing a population to converge toward a state where all individuals have identical genome sequences. The intuitive idea is that when parents mate and produce a child, the child is “close” in some sense to the parent. If the child then replaces a different individual the diversity of the population will tend to decrease on average. When mutation of individuals is included in the uniform selection case the expected diversity decreases to a level where the diversity introduced by mutation counters the decrease due to the crossover operation. We calculate explicit expressions for the expected diversity measure under uniform selection which show this trend.

When fitness-based selection is used, the expected diversity is no longer strictly decreasing in general. We derive an expression for the expected diversity when independent, fitness-based selection is used and show the expected diversity is still strictly decreasing if the parent selection and individual removal probability functions are identical. We qualitatively consider the conditions when the expected diversity will decrease and when it will increase.

Finally, we discuss the relationship between the fitness value of an individual and its genomic representation. We consider the effects of assuming that the fitness measure is, at least approximately, related to the genomic distance from an individual to an individual with maximum fitness.

In Section 2 we describe the class of genetic algorithms we consider. In Section 3 we define the general form of the crossover operation and give some examples. In Section 4 we define a class of additive distance functions between the genomic representations of individuals and give some examples. In Section 5 we define the mutation operation and a basic assumption about its interaction with the distance function in terms of expected distances. We give some examples where this assumption holds and present the basic lemma used in later results. In Section 6 we define the diversity measure as the sum of all pairwise distances within a population, and in Section 7 we calculate a recursive expression for the expected diversity of a population at time step k . In Section 8 we consider the dynamics of the diversity measure in the case of independent, uniform selection. We present some theorems for this case which explicitly characterize the expected diversity of a population over time. In Section 9 we consider the diversity measure when selection is fitness-based and where the selection probabilities for parents are independent and identical and the selection probability for the individual to be replaced is independent. In Section 10 we discuss cases where the fitness function is defined by a relationship between

genomic distances and fitness values. In Section 11 we give a general discussion of the results. Appendices A and B present a derivation similar to that in Sections 7 and 8 but with a slightly different genetic algorithm definition.

2 Algorithm Definition

In this section we define the specific class of algorithms to be analyzed. In general terms, the algorithm works on a population of individuals. An individual is represented by some string or vector, often representing a genome sequence or a problem instance. Each individual is also associated a positive, scalar fitness value.

The algorithm begins with some initial population of individuals. At each successive discrete time instance two distinct individuals are chosen from the population to be mated, and one individual is chosen for replacement. This selection is random, with probabilities determined by the fitness values of the individuals. A crossover operator is applied to the two individuals to be mated, producing a new individual (the offspring or child). The population at the following time step is created by replacing the individual chosen for removal with the child.

We denote the population at time k by

$$S^{(k)} = (s_1^{(k)}, \dots, s_n^{(k)}),$$

where the population size $n \geq 1$ is fixed for all k . The initial population is defined as $S^{(0)}$. The fitness values associated with individuals at time k are written as

$$F^{(k)} = (f_1^{(k)}, \dots, f_n^{(k)}).$$

The algorithm moves from time k to time $k + 1$ by first selecting three integer indices between 1 and n , representing the parents to be mated and an individual to be replaced. The indices of the parents chosen at time k will be written as $m^{(k)}$ and $w^{(k)}$, and the index of the individual to be replaced will be written $v^{(k)}$. These indices are chosen according to a known probability function $p(m^{(k)}, w^{(k)}, v^{(k)} | F^{(k)})$.

We assume a generalized crossover operator Φ which takes two individuals and produces two new individuals. This function is typically random, but we write it as a deterministic function by having it take an additional random vector as an argument (possibly a random vector with both discrete and continuous elements). We typically write this additional random vector as ϕ . We discuss the specific properties we require for the crossover operation in the next section. Only one child produced by the crossover operation is selected, and we assume the choice is random with equal probabilities for each child. A 0-1 random variable $b^{(k)}$ will be used to denote this choice. After an individual is selected it is mutated and replaces individual s_v .

To represent the mutation operation we define a mutation operator Ψ which takes an individual and returns a modified individual. Mutation is also a random operation, and so takes a random vector ϕ as an additional argument. We discuss the specific properties we require for the mutation operation in a later section.

For now we have

$$(c_1^{(k)}, c_2^{(k)}) = \Phi(s_m^{(k)}, s_w^{(k)}, \phi^{(k)}) \quad (1)$$

$$s_v^{(k+1)} = \begin{cases} \Psi(c_1^{(k)}, \psi_1^{(k)}) & \text{if } b^{(k)} = 0 \\ \Psi(c_2^{(k)}, \psi_2^{(k)}) & \text{if } b^{(k)} = 1 \end{cases}, \quad (2)$$

where $c_1^{(k)}$ and $c_2^{(k)}$ are the individuals produced as the children of $s_m^{(k)}$ and $s_w^{(k)}$. We assume all the random vectors ϕ and ψ are independent from all other random vectors and variables. Defining $s_i^{(k+1)} = s_i^{(k)}$ for $i \neq v$ completes the definition of the algorithm. Note that there is one crossover at each time step.

3 The Generalized Crossover Operator

We require the crossover operator Φ to have a certain form. As mentioned earlier, we take individuals to be strings or vectors of some sort of element, having a fixed length h . Letting a and b be individuals, we write $a = (a_1, \dots, a_h)$ and $b = (b_1, \dots, b_h)$. Note that here we are using subscripts for elements of individuals rather than for individuals in a population.

The mutation operator on parents a and b , producing children x and y , is written as

$$\Phi(a, b, \phi) = (x, y), \quad (3)$$

where ϕ is a random vector as described earlier. We require that either $x_i = a_i$ and $y_i = b_i$ or else that $x_i = b_i$ and $y_i = a_i$, for each i . Thus the operator chooses elements of the parents and assigns them to the children, though mixing them up.

Example 1: One-point crossover. In this case the random vector ϕ passed in specifies some crossover point q between 1 and n . We then take

$$x = (a_1, \dots, a_q, b_{q+1}, \dots, b_h)$$

and

$$y = (b_1, \dots, b_q, a_{q+1}, \dots, a_h).$$

Example 2: Uniform crossover. In this case the random vector ϕ passed in contains h binary random variables with an arbitrary distribution function. Often

each will be independent with equal probabilities for 0 and 1. We then choose $x_i = a_i$ and $y_i = b_i$ if $\phi_i = 0$ and $x_i = b_i$ and $y_i = a_i$ if $\phi_i = 1$.

We note that other forms of crossover operators will also work, provided they interact correctly with the distance functions to be introduced later. That is, provided distance relations to be defined in Section 5 still hold.

4 An Additive Distance Function Between Individuals

We assume there is a distance function $d(a, b)$ defined between any individuals a and b . This distance function is required to have the following properties:

- Symmetric: $d(a, b) = d(b, a)$,
- Non-negative: $d(a, b) \geq 0$,
- Zero only for identical individuals: $d(a, b) = 0$ if and only if $a = b$,
- Additive in the sense that $d(a, b) = \sum_{i=1}^h g_i(a_i, b_i)$, where the g_i are arbitrary functions (which, with some restrictions, may also depend on k) such that the previous properties hold for them (i.e. they are symmetric, non-negative, and zero only for identical elements).

Some examples of common distance functions with these properties follow.

Example 1: Hamming distance. In this case we treat the individuals as strings. For simplicity we restrict the strings to be binary strings in this example, i.e., strings over the alphabet $\{0, 1\}$. We define $g_i(a_i, b_i) = 0$ if $a_i = b_i$ and $g_i(a_i, b_i) = 1$ otherwise.

Example 2: Squared Euclidean distance. In this case we have individuals which can be considered to be vectors of integers, rationals, or real numbers. We define $d(a, b) = \sum_{i=1}^h (a_i - b_i)^2$. That is, $g_i(a_i, b_i) = (a_i - b_i)^2$. Note that the Euclidean distance (based on the Euclidean or l_2 norm) does not meet the additive requirement. Distance based on the l_1 norm, i.e., the sum of the absolute differences, does meet the requirements and provides another example.

Example 3: Combined forms. Since the function g_i can depend on i one can treat part of an individual as a string and part as made up of real numbers, as in the two earlier examples.

Combining a distance function of this form with the generalized crossover operation defined above we obtain the following lemma. In words, the lemma states that the sum of the distances from an arbitrary individual to a pair of parents equals the sum of the distances from that arbitrary individual to the two children produced by a crossover operation on those parents.

Lemma 1 *Let*

$$\Phi(a, b, \phi) = (x, y)$$

be a crossover operator of the type defined in Section 3, and let d be a distance function with the properties above. Then

$$d(s, a) + d(s, b) = d(s, x) + d(s, y)$$

for any individual s . Also, $d(a, b) = d(x, y)$.

Proof:

$$\begin{aligned} d(s, a) + d(s, b) &= \sum_{i=1}^h g_i(s_i, a_i) + \sum_{i=1}^h g_i(s_i, b_i) \\ &= \sum_{i=1}^h [g_i(s_i, a_i) + g_i(s_i, b_i)] \\ &= \sum_{i=1}^h [g_i(s_i, x_i) + g_i(s_i, y_i)] \\ &= d(s, x) + d(s, y). \end{aligned}$$

Note that this lemma holds regardless of the properties of the random variable ϕ and allows us to write the summed distances from any individual to a pair of children in terms of the summed distances to the parents.

5 The Mutation Operator

As defined previously, the mutation operator takes an individual and a random vector and produces another, modified, individual. That is, $b = \Psi(a, \psi)$ applies the mutation operation to an individual a and produces a new individual b . The mutation operator is also required to meet certain requirements defined with respect to the expected value of the distance function d . We define those requirements later in this section.

Before proceeding, we discuss the notation we use for expected values. We use the symbol E with a superscript indicating the probability distribution with respect to which the expectation is taken. Thus, for example,

$$E^{p(x,y|z)}[g(x, y, z)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy g(x, y, z) p(x, y|z),$$

where the integrals are replaced by summation for discrete random variables (which we deal with in this paper). This allows us to manipulate expectation operators

in some cases as if we are multiplying densities, e.g., $E^{p(x,y)} = E^{p(y)}E^{p(x|y)}$ and so forth.¹ We will write the operator E without the superscript to denote the full expectation with respect to all random variables of its argument (resulting in a number rather than a random quantity).

We assume that any random argument ψ to the mutation operator is independent from all other random variables. The condition we require for the mutation operator is that

$$E^{p(\psi)}[d(\Psi(a, \psi), b)] = \epsilon_1 d(a, b) + \epsilon_2, \quad (4) \quad \epsilon_1 \text{ and } \epsilon_2 \text{ defined.}$$

where we assume ϵ_1 and ϵ_2 are nonnegative. The constants ϵ_1 and ϵ_2 are fixed for a given mutation operation Ψ and probability function $p(\psi)$. Thus we also have the relations

$$\begin{aligned} E^{p(\psi_1, \psi_2)}[d(\Psi(a, \psi_1), \Psi(b, \psi_2))] &= \epsilon_1^2 d(a, b) + \epsilon_1 \epsilon_2 + \epsilon_2 \\ E^{p(\psi_1, \psi_2)}[d(\Psi(\Psi(a, \psi_1), \psi_2), b)] &= \epsilon_1^2 d(a, b) + \epsilon_1 \epsilon_2 + \epsilon_2 \end{aligned}$$

and so forth, making use of the independence of the two (or more) mutation operations.

Some examples of mutation operators which meet this condition follow, corresponding to the examples of distances given in Section 4.

Example 1: Random changes in a string. Suppose we are using the Hamming distance between binary strings as our d function, as in Example 1 of Section 4. A mutation is assumed to change exactly one string element, chosen uniformly. The index of the element to be chosen is passed in as random variable ψ . We write this index as x to indicate that, in general, other random variables could be passed in as well. A change of a string character is assumed to always flip the bit value b_x from its previous value to $1 - b_x$. Then we have

$$\begin{aligned} E^{p(\psi)}[d(\Psi(a, \psi), b)] &= E^{p(x)}\left[\left(\sum_{i=1}^h g_i(a_i, b_i)\right) - g_x(a_x, b_x) + g_x(1 - a_x, b_x)\right] \\ &= d(a, b) + E^{p(x)}[g_x(1 - a_x, b_x) - g_x(a_x, b_x)] \\ &= d(a, b) + E^{p(x)}[(1 - g_x(a_x, b_x)) - g_x(a_x, b_x)] \\ &= d(a, b) + \frac{1}{h} \sum_{x=1}^h (1 - 2g_x(a_x, b_x)) \end{aligned}$$

¹Note that the expectation operators do not commute in general, though they do for expectations with respect to independent variables. Note also that for conditional distributions like $E^{p(x,y,z)} = E^{p(y,z)}E^{p(x|y,z)}$ the $p(y)$ expectation must be to the left. One way to view this is just as an ordering of the integration or summation operations for evaluating the total expectation (of a function of a random vector), making use of conditional probability rules. One can also view this as successive expectations with certain variables “given,” though independent variables are usually dropped from the conditioning on a probability function or density.

$$= (1 - \frac{2}{h})d(a, b) + 1.$$

The case where more than one element is mutated can be obtained by repeatedly applying the operator, or else derived similarly if some number of distinct elements are required to be mutated.

Example 2: Additive, zero-mean random noise. Suppose we are using the squared Euclidean distance, as in Example 2 of Section 4. Recall that in this case we treat the elements as numbers. The mutation operator for this example is defined to add independent, zero-mean random noise to exactly one element of the individual, i.e., to one component of the vector. The additive noise w is assumed to have variance σ^2 . Thus the random vector ψ contains two random variables: an integral index x into the string and the real-valued, zero-mean random noise w . We assume x and w are independent. Then we have

$$\begin{aligned} E^{p(\psi)}[d(\Psi(a, \psi), b)] &= E^{p(x, w)}[\sum_{i=1}^h g_i(a_i, b_i) - g_x(a_x, b_x) + g_x(\Psi(a_x, \psi), b_x)] \\ &= d(a, b) + E^{p(x, w)}[g_x(\Psi(a_x, \psi), b_x) - g_x(a_x, b_x)] \\ &= d(a, b) + E^{p(x, w)}[(a_x - b_x + w)^2 - (a_x - b_x)^2] \\ &= d(a, b) + E^{p(x)}E^{p(w)}[2w(a_x - b_x) + w^2] \\ &= d(a, b) + \sigma^2. \end{aligned}$$

Similar expressions can be obtained for the cases where more than one element is perturbed.

Using property (4) of the mutation operator and Lemma 1 we obtain the following lemma.

Lemma 2 *Let*

$$\Phi(a, b, \phi) = (x, y)$$

be a crossover operator of the type defined in Section 3, and let d be a distance function with the properties defined in Section 4. Then, with Ψ a mutation operator of the type defined in Section 5,

$$\begin{aligned} E^{p(\psi_1, \psi_2)}[d(s, \Psi(x, \psi_1)) + d(s, \Psi(y, \psi_2))] &= \epsilon_1(d(s, x) + d(s, y)) + 2\epsilon_2 \\ &= \epsilon_1(d(s, a) + d(s, b)) + 2\epsilon_2 \end{aligned}$$

for any individual s . Also,

$$\begin{aligned} E^{p(\psi_1, \psi_2)}[d(\Psi(x, \psi_1), \Psi(y, \psi_2))] &= \epsilon_1^2 d(x, y) + \epsilon_1 \epsilon_2 + \epsilon_2 \\ &= \epsilon_1^2 d(a, b) + \epsilon_1 \epsilon_2 + \epsilon_2 \end{aligned}$$

Proof: The expected value is linear, so the expected value of the sum is the sum of the expected values. Applying requirement (4) to both terms in the sum, along with Lemma 1, gives the first two equations. For the third equation, set $c = \Psi(y, \psi_2)$ and take the expectation with respect to ψ_1 (using the independence of the ψ_i and requirement (4)). Then substitute back in for c and take the expectation with respect to ψ_2 . Apply Lemma 1 to get the final line.

Note that, again, the random vector ϕ from the crossover operation does not enter into the result. In words, the sum of the distances from an individual to the two mutated children equals a constant times the sum of the distances to the parents plus another constant.

Lemma 2 is the basic property that the crossover operator, mutation operator, and distance measure between individuals must satisfy for the results of this paper to hold. Alternate definitions of these operations are possible provided this condition still holds (and provided the distance properties of Section 4 still hold).

6 A Measure of Population Diversity

We now construct a measure of population diversity from the previously defined distance function between individuals. We assume the distance function d has all the properties previously specified.

We define a diversity measure $L^{(k)}$ of a population at time k as the sum over all pairwise distances between individuals in the population. That is,

$L^{(k)}$ and $\bar{d}^{(k)}$ defined.

$$L^{(k)} \equiv \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n d(s_i^{(k)}, s_j^{(k)}) \quad (5)$$

$$= n^2 \bar{d}^{(k)}, \quad (6)$$

where $s_i^{(k)}$ and $s_j^{(k)}$ are individuals in population $S^{(k)}$ and $\bar{d}^{(k)}$ is the average of all distances (including zero self-distances) at time k . The following lemma follows directly from the properties of the distance function.

Lemma 3 $L^{(k)} = 0$ if and only if all the individuals in population $S^{(k)}$ are identical, and $L^{(k)} > 0$ otherwise.

Proof: From the requirements of the distance function d we know $d(s_i, s_j) = 0$ if and only if $s_i = s_j$, and is positive otherwise. The lemma then follows directly from the definition of $L^{(k)}$.

Actually, a weaker condition will guarantee convergence for finite populations. Here we write the j th element of individual s_i as s_{ij} .

Lemma 4 $L^{(k)} < (n-1) \min_{i,j,h} [g_i(s_{ih}^{(k)}, s_{jh}^{(k)}) \neq 0]$ if and only if all the individuals in population $S^{(k)}$ are identical.

Proof: The “if” part follows directly from Lemma 3. To show the “only if” part, assume the contrary. That is, assume that not all individuals are identical but $L^{(k)} < (n-1)\beta^{(k)}$, where $\beta^{(k)} = \min_{i,j,h} [g_i(s_{ih}^{(k)}, s_{jh}^{(k)}) \neq 0]$. Thus there is at least one individual s which is different from at least one other individual. Because of the finite number of individuals in $S^{(k)}$ we can calculate a minimum possible distance between unequal individuals, which is the value β . Because equality of individuals is transitive we partition S into two sets of individuals A and B , where all individuals in A are identical to s and all individuals in B are different from s . Thus we have

$$\begin{aligned} L^{(k)} &\geq \sum_{s_i \in A} \sum_{s_j \in B} d(s_i, s_j) \\ &\geq |A| |B| \beta^{(k)} \\ &= r(n-r) \beta^{(k)} \\ &\geq (n-1) \beta^{(k)}, \end{aligned}$$

where we have defined $|A| = r$. The expression is minimized when $r = 1$, yielding a contradiction.

Note that for the discrete-valued Hamming distance we know the minimum difference between unequal individuals is 1, independent of k , so in this case $L^{(k)} < n-1$ guarantees that all the individuals are equal.

7 The Expected Diversity

We now write an expression for $L^{(k+1)}$ in terms of $L^{(k)}$. From the definition of the algorithm and of $L^{(k)}$ it is clear that to get $L^{(k+1)}$ we need to subtract off all distances to the individual that was removed and then add in all the distances to the new child. Exactly one child is produced at each time step. Recall that m and w are the indices for the individuals chosen as parents and that v is the index for the individual selected for removal. Where the superscript is left off it will be assumed to be k .

We define the new (potential) children as c_1 and c_2 . That is, letting $(a, b) = \Phi(s_m, s_w)$ then $(c_1, c_2) = (\Psi(a, \psi_1), \Psi(b, \psi_2))$. One of the two children is chosen, randomly with equal probabilities, to go in slot v in the next generation. We write

the 0-1 random variable representing this choice as b . Writing the expression out we obtain

$$\begin{aligned} L^{(k+1)} &= L^{(k)} \\ &- \sum_{i \neq v} d(s_v, s_i) \\ &+ b \sum_{i \neq v} d(c_1, s_i) + (1-b) \sum_{i \neq v} d(c_2, s_i). \end{aligned}$$

Note that we can choose to sum over the zero self-distances or not, but leaving out other indices is significant. For example, leaving out the $i = v$ terms in the sums of the third line corresponds to not including the distance from the new child to the individual chosen to be removed.

We next take the expected value of $L^{(k+1)}$. Note that functionally, in terms of random variables,

$$L^{(k+1)} = g(S^{(k)}, F^{(k)}, m^{(k)}, w^{(k)}, v^{(k)}, \psi_1^{(k)}, \psi_2^{(k)}, b^{(k)}),$$

for some function g , since we assume that n and all other variables are known constants. Thus to find the full expectation $E[L^{(k+1)}]$ we take the expectation

$$\begin{aligned} E^{p(S, F, m, w, v, \psi_1, \psi_2, b)} &= E^{p(S, F)} E^{p(m, w, v | S, F)} E^{p(\psi_1, \psi_2)} E^{p(b)} \\ &= E^{p(S, F)} E^{p(m, w, v | F)} E^{p(\psi_1, \psi_2)} E^{p(b)}, \end{aligned}$$

where we have left off the superscripts k . We will evaluate these expectations in left-to-right order, applying the operator sequence to $L^{(k+1)}$. Notationally, we use the forms where the expectation argument is in brackets and where the argument is the expression to the right of the operator symbol interchangeably. We first take the expectation with respect to b , and obtain

$$\begin{aligned} E^{p(b)} L^{(k+1)} &= L^{(k)} \\ &- \sum_{i \neq v} d(s_v, s_i) \\ &+ \frac{1}{2} \sum_{i \neq v} d(c_1, s_i) + \frac{1}{2} \sum_{i \neq v} d(c_2, s_i). \end{aligned}$$

We now take the expected value with respect to the mutation noise by applying Lemma 2 to the distances to the children in the third line of the equation above. We also obtain an expression entirely in terms of distances between population members at time k :

$$\begin{aligned}
E^{p(\psi_1, \psi_2)} E^{p(b)} L^{(k+1)} &= L^{(k)} \\
&- \sum_{i \neq v} d(s_v, s_i) \\
&+ \frac{1}{2} \sum_{i \neq v} [\epsilon_1 d(s_m, s_i) + \epsilon_1 d(s_w, s_i) + 2\epsilon_2].
\end{aligned}$$

It follows after some simplification that

$$\begin{aligned}
E^{p(\psi_1, \psi_2)} E^{p(b)} L^{(k+1)} &= L^{(k)} + (n-1)\epsilon_2 \\
&- \sum_{i \neq v} d(s_v, s_i) \\
&+ \frac{\epsilon_1}{2} \sum_{i \neq v} d(s_m, s_i) \\
&+ \frac{\epsilon_1}{2} \sum_{i \neq v} d(s_w, s_i).
\end{aligned}$$

The expected value with respect to the mutation operations has now been evaluated. We next take the expected values with respect to the selection of parents and individual to be replaced.

Taking the expectation and rearranging the summations we obtain,

$$\begin{aligned}
E^{p(m, w, v|F)} E^{p(\psi_1, \psi_2)} E^{p(b)} L^{(k+1)} &= L^{(k)} + (n-1)\epsilon_2 \\
&+ E^{p(m, w, v|F)} [\\
&- \sum_{i \neq v} d(s_v, s_i) \\
&- \frac{\epsilon_1}{2} d(s_m, s_v) \\
&+ \frac{\epsilon_1}{2} \sum_i d(s_m, s_i) \\
&- \frac{\epsilon_1}{2} d(s_w, s_v) \\
&+ \frac{\epsilon_1}{2} \sum_i d(s_w, s_i)].
\end{aligned}$$

We now use the linearity of the expected value operation and write

$$\begin{aligned}
E^{p(m, w, v|F)} E^{p(\psi_1, \psi_2)} E^{p(b)} L^{(k+1)} &= L^{(k)} + (n-1)\epsilon_2 \\
&- E^{p(v|F)} \sum_{i \neq v} d(s_v, s_i)
\end{aligned}$$

$$\begin{aligned}
& - \frac{\epsilon_1}{2} E^{p(m,v|F)} d(s_m, s_v) \\
& + \frac{\epsilon_1}{2} E^{p(m|F)} \sum_i d(s_m, s_i) \\
& - \frac{\epsilon_1}{2} E^{p(w,v|F)} d(s_w, s_v) \\
& + \frac{\epsilon_1}{2} E^{p(w|F)} \sum_i d(s_w, s_i).
\end{aligned} \tag{7}$$

Note that the expected values sum over the variables not present in a term, resulting in expectations with respect to the marginal probability functions. Further analysis will require making some assumptions about the distribution of m , w , and v .

Note that nowhere have we assumed m , w , and v to be distinct. Thus expression (7) holds for any of the cases where $m = w$, $m = v$, $w = v$, and $m = w = v$. This is important because it allows for — though does not require — various assumptions of independence between m , w , and v (conditional on F). For example, one might assume that $p(m, w, v|F) = p(m|F)p(w|F)p(v|F)$. We use this assumption in the following section.

8 Independent, Uniform Selection

In this section we assume that m , w , and v are independent and uniformly selected. That is, we take

$$p(m, w, v) = p(m)p(w)p(v) = \frac{1}{n^3}. \tag{8}$$

The algorithm is independent of the fitness values (or they are assumed to all be equal at all times k) and so dependence on $F^{(k)}$ is removed from the expected value.

The marginal distributions, e.g. $p(m)$ and $p(m, v)$, are equal to $1/n$ for single arguments and $1/n^2$ for two arguments. We can therefore evaluate expression (7) with these distributions and obtain

$$\begin{aligned}
E^{p(m,w,v|F)} E^{p(\psi_1, \psi_2)} E^{p(b)} L^{(k+1)} &= L^{(k)} + (n-1)\epsilon_2 \\
&- \sum_v \frac{1}{n} \sum_i d(s_v, s_i) \\
&- \frac{\epsilon_1}{2} \sum_m \sum_v \frac{1}{n^2} d(s_m, s_v) \\
&+ \frac{\epsilon_1}{2} \sum_m \frac{1}{n} \sum_i d(s_m, s_i) \\
&- \frac{\epsilon_1}{2} \sum_w \sum_v \frac{1}{n^2} d(s_w, s_v)
\end{aligned}$$

$$+ \frac{\epsilon_1}{2} \sum_w \frac{1}{n} \sum_i d(s_w, s_i)].$$

Simplifying and writing in terms of $L^{(k)}$ we obtain

$$\begin{aligned} E^{p(m,w,v|F)} E^{p(\psi_1,\psi_2)} E^{p(b)} L^{(k+1)} &= L^{(k)} + (n-1)\epsilon_2 \\ &- \frac{2L^{(k)}}{n} \\ &- \frac{L^{(k)}\epsilon_1}{n^2} \\ &+ \frac{L^{(k)}\epsilon_1}{n} \\ &- \frac{L^{(k)}\epsilon_1}{n^2} \\ &+ \frac{L^{(k)}\epsilon_1}{n}. \end{aligned}$$

Simplifying again we have

$$\begin{aligned} E^{p(m,w,v|F)} E^{p(\psi_1,\psi_2)} E^{p(b)} L^{(k+1)} &= (1 - \frac{2}{n} + \frac{2\epsilon_1}{n} - \frac{2\epsilon_1}{n^2}) L^{(k)} + (n-1)\epsilon_2 \\ &= [1 - \frac{2}{n^2}(n - n\epsilon_1 + \epsilon_1)] L^{(k)} + (n-1)\epsilon_2. \end{aligned}$$

We now take the final expectation. Since $L^{(k)} = g(S^{(k)})$, where function g is the sum over pairwise distances, we know $E^{p(L^{(k)})}[L^{(k)}] = E^{p(S^{(k)})}[g(S^{(k)})]$. So we only need to take the expectation of the r.h.s. with respect to $p(L^{(k)})$, resulting in

$$E[L^{(k+1)}] = E^{p(S^{(k)})} E^{p(m,w,v|F)} E^{p(\psi_1,\psi_2)} L^{(k+1)} \quad (9)$$

$$= [1 - \frac{2}{n^2}(n - n\epsilon_1 + \epsilon_1)] E[L^{(k)}] + (n-1)\epsilon_2. \quad (10)$$

To simplify the equations we define

δ_1 and δ_2 defined.

$$\delta_2 = (n-1)\epsilon_2 \quad (11)$$

$$\delta_1 = 1 - \frac{2}{n^2}(n - n\epsilon_1 + \epsilon_1). \quad (12)$$

Summarizing in a theorem, we have

Theorem 1 *If m , w , and v are selected uniformly and independently (with replacement) from $\{1, \dots, n\}$ then*

$$E[L^{(k+1)}] = \delta_1 E[L^{(k)}] + \delta_2 \quad (13)$$

$$E[\bar{d}^{(k+1)}] = \delta_1 E[\bar{d}^{(k)}] + \frac{\delta_2}{n^2} \quad (14)$$

Using Theorem 1 we obtain another theorem on the conditions for $E[L^{(k)}]$ to be decreasing for increasing k . Note that the next three theorems follow from Theorem 1, and so similar results apply whenever the expected value of the diversity measure obeys the formulas in Theorem 1 (and $\delta_2 \geq 0$). That is, the same formulas may hold for different models with a redefinition of δ_1 and δ_2 .

Theorem 2 *If m , w , and v are selected uniformly and independently (with replacement) from $\{1, \dots, n\}$ then $E[L^{(k+1)}]$ is strictly less than $E[L^{(k)}]$ if and only if $\delta_1 < 1$ and*

$$E[L^{(k)}] > \frac{\delta_2}{(1 - \delta_1)}.$$

Proof: Apply Theorem 1 and require that $\delta_1 E[L^{(k)}] + \delta_2 < E[L^{(k)}]$. Then $\delta_2 < (1 - \delta_1)E[L^{(k)}]$. If $\delta_1 < 1$ we have $\frac{\delta_2}{1 - \delta_1} < E[L^{(k)}]$. If $\delta_1 = 1$ we must have $\delta_2 < 0$, and similarly for the case where $\delta_1 > 1$ since $L^{(k)}$ is strictly nonnegative. But we assumed $\delta_2 \geq 0$ so we know that $\delta_2 \geq 0$.

This theorem shows that, if the diversity starts out larger than $\delta_2/(1 - \delta_1)$ it will decrease monotonically until it reaches that level. The delta values are based on the mutation constants and the expression will typically be small relative to an initial population. Considering the population of individuals as a cluster, and the diversity measure as a clustering measure, the expected “size” of the initial cluster decreases until it reaches a level where the decrease in diversity is countered by the increase due to the mutation operation. With no mutation (i.e., with $\epsilon_1 = 1$ and $\epsilon_2 = 0$) the expected diversity converges toward zero.

The next theorem gives a formula for computing the expected diversity at any time step, based only on the initial average or expected population diversity. It can also be applied to calculate future expected values at any time step when the expected diversity is known or can be estimated.

Theorem 3 *If m , w , and v are selected uniformly and independently (with replacement) from $\{1, \dots, n\}$ and $k \geq 1$ then*

$$E[L^{(k)}] = \delta_1^k E[L^{(0)}] + \delta_2 \frac{1 - \delta_1^k}{1 - \delta_1} \quad (15)$$

Proof: Solve equation (13) for $E[L^{(k)}]$. A recurrence relation of the form $a_k = ca_{k-1} + f(k)$, has a general solution for $k \geq 1$ given by $a_k = c^k a_0 + \sum_{j=1}^k c^{k-j} f(j)$. [MKB83] Also, $\sum_{k=0}^n a^k = \frac{a^{n+1}-1}{a-1}$.

We now present another theorem which will allow us to use the expected value $E[L^{(k)}]$ to obtain a lower bound on the probability that $L^{(k)}$ is less than some given constant value.

Theorem 4 *Let m , w , and v be selected uniformly and independently (with replacement) from $\{1, \dots, n\}$. Let $\alpha > 0$ be a given constant. Then*

$$\begin{aligned} p(L^{(k)} < \alpha) &\geq 1 - \frac{E[L^{(k)}]}{\alpha} \\ &= 1 - \frac{\delta_1^k}{\alpha} E[L^{(0)}] - \frac{\delta_2(1 - \delta_1^k)}{\alpha(1 - \delta_1)} \end{aligned}$$

Proof: From a generalization of the Tchebycheff inequality, for a random variable $L^{(k)}$ taking only positive values, we know [Pap65]

$$\begin{aligned} p(L^{(k)} \geq \alpha) &\leq \frac{E[L^{(k)}]}{\alpha} \\ 1 - p(L^{(k)} \geq \alpha) &\geq 1 - \frac{E[L^{(k)}]}{\alpha} \\ p(L^{(k)} < \alpha) &\geq 1 - \frac{E[L^{(k)}]}{\alpha}. \end{aligned}$$

We can then use Theorem 3 to write in terms of $L^{(0)}$.

This theorem can be used to show the convergence in probability of the zero mutation case to zero expected diversity. In a practical sense it allows for computing a bound on the the number of time steps until the expected diversity is below a certain arbitrary value α with some given probability. For example, one could use Lemma 4 and the Hamming distance along with this theorem to compute an upper bound on the number of time steps k until all individuals in a population are identical with probability, say, 0.99.

9 Independent, Fitness-Based Selection

In this section we consider the dynamics of population diversity (with respect to the previously defined diversity measure) when the selection of parents and of the individual to replace is based on fitness. Including fitness leads to a much more complicated analysis in general. Our results are not as strong as for independent uniform selection, where the expected diversity can be predicted at any time k based on the initial diversity. When selection is by fitness the expected diversity is no longer strictly decreasing in general, even without mutation. We characterize some of these situations where expected diversity will still decrease and qualitatively describe when it will tend to decrease and when it will tend to increase. We obtain these results while making few assumptions about the fitness values (which can be quite complicated in general since the probability distributions change at each k , and fitness evaluation may be a complicated function).

To carry out our analysis we assume the parents are selected independently according to a common parent selection distribution, and that the individual selected for removal is selected independently according to another probability distribution. That is, we assume

$$p_{m|F}(i|F) = p_{w|F}(i|F) \equiv p_m^{(k)}(i) \quad (16)$$

and that

$$p_{v|F}(i|F) \equiv p_v^{(k)}(i), \quad (17)$$

where we have left off the superscripts k on m , w , v , and F . By the independence assumption, the joint distribution is the product of these probability functions. Note that we now use subscripts on the probability function p to indicate what distribution is being referred to, since the convention that the argument provides this information is no longer sufficient here. It is important to remember that these probability distributions are generally functions of the fitness values $f_i^{(k)}$ in $F^{(k)}$, since to obtain an expected value with respect to all random variables it is necessary to take an expectation with respect to fitness.²

Substituting these distributions into equation (7) and relabeling the dummy variables we have

$$\begin{aligned} E^{p(m,w,v|F)} E^{p(\psi_1,\psi_2)} E^{p(b)} L^{(k+1)} &= L^{(k)} + (n-1)\epsilon_2 \\ &- \sum_i p_v^{(k)}(i) \sum_j d(s_i, s_j) \end{aligned}$$

²Often fitness at time k is a deterministic function of the population $S^{(k)}$, in which case we can eliminate the expectation with respect to F and only deal with S . We have left fitness in as $p(S, F)$ because the fitness might in general have a random component as in $f_i^{(k)} = f(S^{(k)}, \mu^{(k)})$ with $\mu^{(k)}$ a random vector.

$$\begin{aligned}
& - \frac{\epsilon_1}{2} \sum_i \sum_j p_m^{(k)}(i) p_v^{(k)}(j) d(s_i, s_j) \\
& + \frac{\epsilon_1}{2} \sum_i p_m^{(k)}(i) \sum_j d(s_i, s_j) \\
& - \frac{\epsilon_1}{2} \sum_i \sum_j p_m^{(k)}(i) p_v^{(k)}(j) d(s_i, s_j) \\
& + \frac{\epsilon_1}{2} \sum_i p_m^{(k)}(i) \sum_j d(s_i, s_j).
\end{aligned}$$

Simplifying, we obtain

$$\begin{aligned}
E^{p(m,w,v|F)} E^{p(\psi_1, \psi_2)} E^{p(b)} L^{(k+1)} &= L^{(k)} + (n-1)\epsilon_2 \\
&- \sum_i p_v^{(k)}(i) \sum_j d(s_i, s_j) \\
&- \epsilon_1 \sum_i \sum_j p_m^{(k)}(i) p_v^{(k)}(j) d(s_i, s_j) \\
&+ \epsilon_1 \sum_i p_m^{(k)}(i) \sum_j d(s_i, s_j).
\end{aligned} \tag{18}$$

We now apply the final expectation operator (symbolically) and group the distance terms into a single pair of summations:

$$\begin{aligned}
E[L^{(k+1)}] &= E[L^{(k)}] + (n-1)\epsilon_2 + E^{p(S^{(k)}, F^{(k)})} [\\
&\quad \sum_i \sum_j [\epsilon_1 p_m^{(k)}(i) - p_v^{(k)}(i) - \epsilon_1 p_m^{(k)}(i) p_v^{(k)}(j)] d(s_i, s_j)].
\end{aligned}$$

To obtain a symmetric form for the summation, we replace the summation with the sum of two identical versions of it, each weighted by 1/2. Then we then rename the dummy indices on one of the summations, swapping i and j and making use of the symmetry of d .

$$\begin{aligned}
E[L^{(k+1)}] &= E[L^{(k)}] + (n-1)\epsilon_2 + E^{p(S^{(k)}, F^{(k)})} [\\
&\quad \frac{1}{2} \sum_i \sum_j [\epsilon_1 p_m^{(k)}(i) - p_v^{(k)}(i) - \epsilon_1 p_m^{(k)}(i) p_v^{(k)}(j)] d(s_i, s_j)] \\
&+ \frac{1}{2} \sum_i \sum_j [\epsilon_1 p_m^{(k)}(j) - p_v^{(k)}(j) - \epsilon_1 p_m^{(k)}(j) p_v^{(k)}(i)] d(s_i, s_j)].
\end{aligned}$$

Combining the summations we have

$$\begin{aligned}
E[L^{(k+1)}] &= E[L^{(k)}] + (n-1)\epsilon_2 \\
&+ \frac{1}{2} E^{p(S^{(k)}, F^{(k)})} \sum_i \sum_j q^{(k)}(i, j) d(s_i, s_j)
\end{aligned} \tag{19}$$

where

$$q^{(k)}(i, j) = \epsilon_1[p_m^{(k)}(i) + p_m^{(k)}(j)] - [p_v^{(k)}(i) + p_v^{(k)}(j)] \quad (20)$$

$$- \epsilon_1[p_m^{(k)}(i)p_v^{(k)}(j) + p_m^{(k)}(j)p_v^{(k)}(i)]$$

$$= \epsilon_1 p_m^{(k)}(i)[1 - p_v^{(k)}(j)] + \epsilon_1 p_m^{(k)}(j)[1 - p_v^{(k)}(i)] \quad (21)$$

$$- p_v^{(k)}(i) - p_v^{(k)}(j). \quad (22)$$

We first point out a case where the diversity can be shown to be strictly decreasing even with fitness-based selection. For simplicity we will assume that $\epsilon_2 = 0$. The basic idea is that the expectation over the summation term must be negative for the diversity to be decreasing in the $\epsilon_2 = 0$ case. (Its absolute value must be sufficiently large to also overcome the $(n - 1)\epsilon_2$ term if $\epsilon_2 \neq 0$.) We saw earlier that if the selection is uniform then (mutation-free) diversity is strictly decreasing. In fact, this is the case whenever $p_m = p_v$ and $\epsilon_1 < 1$, since then the first two terms of (20) cancel each other out (and a negative term results if $\epsilon_1 \neq 1$).

For the more general case we must consider the products $[p_m(i)[1 - p_v(j)] + p_m(j)[1 - p_v(i)]]d_{ij}$ and $[p_v(i) + p_v(j)]d_{ij}$, where we have abbreviated the notation. Basically, the sum will tend to be positive when the distances between high fitness parents tend to be large, and will tend to be negative when the distances between low fitness individuals tend to be large (assuming parents are selected for high fitness and the individual to be replaced is selected for low fitness). There is a bias toward decreasing diversity due to the $1 - p_v^{(k)}$ terms, since the probability distributions are all normalized.

To get much farther we would need to make some assumptions about the form of the fitness function. Examining the forms of these equations may lead to insights into other useful statistics of a population to track, either in addition to L or as an enhancement, as well as possibly useful assumptions to make. An amortized analysis over some number of time steps greater than one might also yield stronger results.

10 Fitness Measures Based on Genomic Distance

To effectively analyze the dynamics of some statistic of a population under fitness-based selection, such as the diversity or average fitness, it is generally necessary to make some assumptions about the particular fitness function and its interaction with the genomic representation, crossover operation, etc. For example, in the most general sense a fitness function could be a deterministic hash function such that there is essentially no relationship between fitness values and the genomic sequence. In a nondeterministic sense, a fitness could simply be an independent random variable

associated with an individual. In these cases the dynamics of a genetic algorithm amount to an undirected random search which generates individuals of arbitrary fitness and tends to keep those with the higher fitness values. The crossover operation acts somewhat like a pseudo-random generator of individuals, which may not be an especially good generator at that.

One can still trivially obtain a monotonically increasing average fitness for an undirected random search, assuming that individuals are generated uniformly over the universe of individuals. (This assumption is reasonable when fitness is independent of genomic representation, but does assume that all individuals can be generated from a population.) Define the probability function for selection such that the lowest fitness individual is always the one replaced. In this case, considering the individuals to be arranged by fitness, the lowest fitness “slot” at time k will have an expected fitness at time step $k+1$ given by the expected fitness over the universe of individuals. The other slots will slowly increase in expected fitness since there is a small chance that a better individual will be found on each time step. Thus the expected average fitness of the population, the sum of the expected fitnesses over all slots divided by n , will monotonically increase. Nonetheless, this is the probabilistic equivalent of a brute force search.

In this section we begin to discuss assumptions on the fitness landscape based on genomic distances between individuals. One assumption which is explicitly or implicitly present in many genetic algorithm optimization techniques is that, on average, the children will “inherit” fitness values which are on average near those of their parents. One way to “derive” this condition is to make a fundamental assumption that, at least approximately, the fitness values are directly related to the genomic distance from an individual to a maximum fitness individual. For example, assume that³

$$f(s_i^{(k)}) = T - d(s_i^{(k)}, z) \quad (23)$$

for some distance function d , where z is the maximum fitness individual and T is a constant larger than the largest distance value between any individuals. Note that under this definition there can be only one maximum fitness individual (i.e., with zero distance to itself). We write fitness either as a function $f(s_i^{(k)})$ or abbreviated as $f_i^{(k)}$. The constant $T = f(z)$ might be used, for example, in a theoretical sense or if the fitness of z is known. The individual z is generally unknown but is assumed to exist and the fitness values are assumed to obey this relationship.

This assumption on the fitness landscape (and hence on the problem structure in an optimization sense) is, by the additivity property of d , essentially a “divide and conquer” assumption that pieces of the optimal solution representation provide a proportionate fitness contribution even when inserted in a suboptimal solution rep-

³Another possible definition is $f(s_i^{(k)}) = 1/(1 + d(s_i^{(k)}, z))$, which leads to a similar analysis.

resentation. This is not an uncommon assumption, but we have written it explicitly in terms of a genomic distance function.⁴

Assumption (22) is a strong assumption on the problem structure. From an algorithmic viewpoint, assuming the problem is to be solved by a traditional computer, it allows a problem to be solved exactly with $h\gamma$ fitness evaluations. Here we assume the elements of individuals are discrete, each with γ possible values, and that h is the number of elements in an individual's representation. One simply tries out all possible elements at a position while keeping the other positions fixed to arbitrary but constant values. In this way, the maximum for each element can be found, and they can then be combined to form the global maximum. Nonetheless, this assumption can serve as a starting point for analysis.

Under assumption (22) the expected sum of fitnesses of two children from the mutation operation equals the sum of the parent's fitnesses, plus some terms due to mutation (assuming that Lemma 2 holds for the crossover, mutation, and distance functions). That is, letting c_1 and c_2 be children of individuals x and y ,

$$f(c_1) + f(c_2) = 2T - d(c_1, z) - d(c_2, z),$$

and so

$$\begin{aligned} E^{p(\psi_1, \psi_2)}[f(c_1) + f(c_2)] &= 2T - [\epsilon_1 d(x, z) + \epsilon_1 d(y, z) + 2\epsilon_2] \\ &= T - d(x, z) + (1 - \epsilon_1)d(x, y) \\ &+ T - d(x, y) + (1 - \epsilon_1)d(y, z) - 2\epsilon_2 \\ &= f(x) + f(y) + (1 - \epsilon_1)[d(x, y) + d(y, z)] - 2\epsilon_2. \end{aligned}$$

When the expectation with respect to selection of one of the two children is taken ($E^{p(b)}$) the result is similar, given by one half times the expression above. Thus the expected fitness of a child with respect to mutation and choosing one child of the two is the average of the parent's fitnesses plus some terms due to mutation. Such an averaging process, without mutation, depends on the variance of the distribution to create children whose fitness values are greater than either parent.

⁴Another way to reach (22) is to take as the fundamental assumption that fitness is additive in the sense of

$$f(a) = \sum_{i=1}^h f_i(a_i), \tag{24}$$

where here the f_i are functions on the components of an individual. Then we can define a distance $d(a, b) = \sum_{i=1}^h |f_i(a_i) - f_i(b_i)|$ which meets all the conditions necessary for Lemma 1, although we would have to assume unique fitness value $f_i(a_i)$ for each possible a_i in order for " $d(a, b) = 0$ if and only if a is identical to b " to hold. (Lemma 2 still requires an assumption on the form of the mutation operation and in this case its interaction with the componentwise fitness function). Then, since z has the maximum fitness of any individual and hence each component is maximum, it is easy to show that $d(a, z) = f(z) - f(a)$. One can then define $T = f(z)$.

Given this condition is is not surprising that the expected fitness of a population without mutation is strictly increasing under assumption (22) when selection is based on fitness and initial diversity is greater than zero. We assume parents are selected proportional to their fitness values and that the individual selected for replacement is chosen uniformly. We next illustrate this result, first defining the sum of fitness values over a population as

$$U^{(k)} = \sum_{i=1}^n f_i^{(k)}.$$

Then, leaving out some initial steps which are similar to those carried out earlier,

$$\begin{aligned} E^{p(\psi_1, \psi_2)} E^{(b)} U^{(k+1)} &= U^{(k)} - f_v^{(k)} + \frac{1}{2} f_m^{(k)} + \frac{1}{2} f_w^{(k)} \\ E^{p(m, w, v)} E^{p(\psi_1, \psi_2)} E^{(b)} U^{(k+1)} &= U^{(k)} - \sum_v \frac{1}{n} f_v^{(k)} \\ &+ \frac{1}{2} \sum_m \frac{f_m^{(k)}}{U^{(k)}} f_m^{(k)} + \frac{1}{2} \sum_w \frac{f_w^{(k)}}{U^{(k)}} f_w^{(k)} \\ &= U^{(k)} - \frac{1}{n} U^{(k)} + \sum_i \frac{f_i^{(k)}}{U^{(k)}} f_i^{(k)} \\ &= U^{(k)} - \frac{1}{n} U^{(k)} + U^{(k)} \sum_i \left(\frac{f_i^{(k)}}{U^{(k)}} \right)^2. \end{aligned}$$

The bracketed term in the final line is the sum of squares of a set of numbers normalized to sum to one. Its minimum value therefore occurs when all the terms in the bracket are equal to $1/n$. This occurs only when all the fitness values are equal, and results in canceling the $-(1/n)U^{(k)}$ term. Therefore $E[U^{(k+1)}] \geq E[U^{(k)}]$ since this happens for any possible fitness assignment and thus for any weighted sum of such assignments. In all other cases the summed fitness must be increasing. The expected fitness of the next generation given the current fitness values is constant if all the fitness values are equal. In this case the expected fitness given the current fitness values will again be positive at future times if $L^{(k)} \neq 0$, provided the differing individuals mate to produce a child with a different fitness value (though the given population fitness, rather than the expected fitness, may decrease).

By writing the assumption (22) explicitly we can consider both the implications of the assumption as well as when it might not provide a good approximation for certain problem spaces. The use of a single z value assumes in a sense a unimodal problem structure. A more general assumption might be constructed from multiple, weighted distances to a set of z_i optima, similar to a kernel estimate of a function. One might also consider assumptions that larger “chunks” of a genome string contribute fitness

values proportionately when inserted into any individual, but that the property does not necessarily hold for single elements.

11 Discussion

We have defined the properties for a class of distance functions and defined a diversity measure in terms of it. Mathematically, of course, we have defined a term proportional to the average distance between individuals in a population and referred to it as the diversity of a population. There are many possible measures of a population which might reasonably be called the diversity, depending on the various senses in which the term is used.

The results we presented for the uniform selection case allow for explicit predictions of the expected diversity at a given time step. How these results relate to the behavior of typical genetic algorithms using fitness-based selection is still an open question for analytical or empirical research. The uniform selection case might provide useful heuristic values to apply in choosing mutation rates and time schedules for algorithms using fitness-based selection.

It is interesting to consider the effects of separate subpopulations within a population. Suppose there are r subpopulations of a population. This can be specified by setting the probability density $p(m, w, v)$ appropriately. When there is no interaction between the subpopulations then each changes in its own $L_r^{(k)}$ diversity measure, but any tendency toward commonality across populations is driven only by the fitness function. On generations where all the subpopulations do interact there is again a common tendency toward sameness. Analysis of the diversity of populations and subpopulations may help to characterize the effects of isolated subpopulations interacting with other populations either occasionally at certain times k or with low probabilities at each time k .

The basic computational model introduced in Section 2 (and similar versions by various researchers over the years) has its roots largely in evolutionary computation, or computing algorithms based on an analogy to the process of evolution. The models may well have some applicability or lead to insights relating to biological evolutionary processes, though we do not attempt to make such claims here. It is worth pointing out some of the simplifications of the model with regard to the fitness function — besides such basic assumptions as a fixed population size, a genome represented by a string and combined with a crossover and mutation function, and so forth.

The fitness function is typically assumed to be constant over time and independent of the population and any other factors. There is no notion of a changing environment. The constant fitness function also eliminates any interaction between individuals except as an isolated competition to maximize a fixed function. In particular, there are no cooperative interactions between individuals. In a more general

sense such interactions could modify the fitness function. There are also various independence assumptions that limit the amount of feedback in the model, making it more tractable analytically though less representative of real biological and certainly sociological systems. So while such models can lead to insights into physical processes, some care must be taken to not to draw false conclusions from a purposely simplified model. A very rough physical analogy to the simplified model might be a gardener growing exactly n rose bushes and externally applying his own, approximately fixed, fitness evaluations to the rose bushes to determine which to breed. The roses would all grow independently of each other, under exactly the same conditions.

On a more limited level, the distance functions we have defined depend on comparing genome sequences element by element. In many sequence comparison applications a more general distance function such as the Levenshtein distance is used [SK83] which allows for mutations that include insertions and deletions of new elements between existing elements in a genome sequence rather than just the change of an existing string element. While the distance definition we have used does not allow for this case, it might be possible to obtain an inequality analogous to Lemma 2 for such a distance.

References

- [MKB83] Joe Mott, Abraham Kandel, and Theodore Baker. *Discrete Mathematics for Computer Scientists*. Reston Publishing Co., 1983.
- [Pap65] Athanasios Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1965.
- [SK83] David Sankoff and Joseph Kruskal. *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Addison-Wesley, 1983.

A Analysis for Four Distinct Parent and Removal Indices

This appendix contains some analysis of the case where both children produced by the crossover operation are mutated and placed in the next population. The analysis is a bit more complicated than that in the main paper and does not lend itself as readily to independence assumptions on the selection probabilities. Nonetheless, it illustrates the analysis of a slightly different model yielding similar results for the case of independent selection.

In this case four indices are selected: two for the parents and two for the individuals to be replaced. These are denoted by m , w , t , and v (where we have left off the superscript k .) We assume here that these four indices are have density $p(m, w, t, v|F)$ and that they are restricted by the definition of this density to be four distinct integers.

We use the same measure of population diversity as in the main paper, equation (5), to analyze convergence. We also make the same assumptions about the crossover operator, the mutation operator, and the distance function.

We now proceed in the same manner to write an expression for $L^{(k+1)}$ in terms of $L^{(k)}$. From the definition of the algorithm and of $L^{(k)}$ it is clear that to get $L^{(k+1)}$ we need to subtract off all distances to the individuals which were removed and then add in all the distances to the new children. Recall that m and w are the indices for the individuals chosen as parents and that t and v are the indices for the individuals selected for removal. Where the superscript is left off it will be assumed to be k . We define the new children as c_1 and c_2 , i.e., letting $(a, b) = \Phi(s_m, s_w)$ then $(c_1, c_2) = (\Psi(a, \psi_1), \Psi(b, \psi_2))$. Writing the expression out we obtain

$$\begin{aligned} L^{(k+1)} &= L^{(k)} \\ &- \sum_{i \neq t, v} d(s_t, s_i) - \sum_{i \neq t, v} d(s_v, s_i) \\ &- d(s_t, s_v) \\ &+ \sum_{i \neq t, v} d(c_1, s_i) + \sum_{i \neq t, v} d(c_2, s_i) \\ &+ d(c_1, c_2). \end{aligned}$$

Note that we can choose to sum over the zero self-distances or not, but leaving out other indices is significant. For example, leaving out the $i = t$ and $i = v$ terms in the sums of the fourth line corresponds to not including the distances from the new children to the individuals chosen to be removed.

We next take the expected value of $L^{(k+1)}$. Note that, functionally,

$$L^{(k+1)} = L^{(k+1)}(S^{(k)}, F^{(k)}, m^{(k)}, w^{(k)}, t^{(k)}, v^{(k)}, \phi_1^{(k)}, \phi_2^{(k)}),$$

since we assume that n and all other variables are known constants. Thus to find the full expectation $E[L^{(k+1)}]$ we take the expectation

$$\begin{aligned} E^{p(S,F,m,w,t,v,\psi_1,\psi_2)} &= E^{p(S,F)} E^{p(m,w,t,v|S,F)} E^{p(\psi_1,\psi_2)} \\ &= E^{p(S,F)} E^{p(m,w,t,v|F)} E^{p(\psi_1,\psi_2)}, \end{aligned}$$

where we have left off the superscript k . We will evaluate these expectations in left-to-right order, applying the operator sequence to $L^{(k+1)}$.

We now take the expected value with respect to the mutation noise by applying Lemma 2 to the distances to the children. We also obtain an expression entirely in terms of distances between population members at time k .

$$\begin{aligned} E^{p(\psi_1,\psi_2)} L^{(k+1)} &= L^{(k)} \\ &- \sum_{i \neq t,v} d(s_t, s_i) - \sum_{i \neq t,v} d(s_v, s_i) \\ &- d(s_t, s_v) \\ &+ \sum_{i \neq t,v} [\epsilon_1 d(s_m, s_i) + \epsilon_1 d(s_w, s_i) + 2\epsilon_2] \\ &+ \epsilon_1^2 d(s_m, s_w) + \epsilon_1 \epsilon_2 + \epsilon_2. \end{aligned}$$

It follows after some simplification that

$$\begin{aligned} E^{p(\psi_1,\psi_2)} L^{(k+1)} &= L^{(k)} \\ &- \sum_{i \neq t,v} d(s_t, s_i) - \sum_{i \neq t,v} d(s_v, s_i) \\ &- d(s_t, s_v) \\ &+ \sum_{i \neq t,v,m} \epsilon_1 d(s_m, s_i) + \sum_{i \neq t,v,w} \epsilon_1 d(s_w, s_i). \\ &+ \epsilon_1^2 d(s_m, s_w) \\ &+ \epsilon_1 \epsilon_2 + \epsilon_2 + 2(n-2)\epsilon_2. \end{aligned}$$

The expected value with respect to the mutation operations have now been evaluated. We next take the expected values with respect to the selection of parents and individuals to be replaced.

Taking the expectation and rearranging the summations we obtain,

$$\begin{aligned} E^{p(m,w,t,v|F)} E^{p(\psi_1,\psi_2)} L^{(k+1)} &= L^{(k)} + (\epsilon_1 + 2n - 3)\epsilon_2 \\ &+ E^{p(m,w,t,v|F)} [\\ &+ d(s_t, s_v) - \sum_{i \neq t} d(s_t, s_i) \end{aligned}$$

$$\begin{aligned}
& + d(s_v, s_t) - \sum_{i \neq v} d(s_v, s_i) \\
& - d(s_t, s_v) \\
& - \epsilon_1 d(s_m, s_t) - \epsilon_1 d(s_m, s_v) + \sum_{i \neq m} \epsilon_1 d(s_m, s_i) \\
& - \epsilon_1 d(s_w, s_t) - \epsilon_1 d(s_w, s_v) + \sum_{i \neq w} \epsilon_1 d(s_w, s_i) \\
& + \epsilon_1^2 d(s_m, s_w)].
\end{aligned}$$

We now write

$$\begin{aligned}
E^{p(m,w,t,v|F)} E^{p(\psi_1, \psi_2)} L^{(k+1)} &= L^{(k)} + (\epsilon_1 + 2n - 3)\epsilon_2 \\
&- E^{p(t|F)} \sum_{i \neq t} d(s_t, s_i) \\
&- E^{p(v|F)} \sum_{i \neq v} d(s_v, s_i) \\
&+ E^{p(t,v|F)} d(s_t, s_v) \\
&- \epsilon_1 E^{p(m,t|F)} d(s_m, s_t) \\
&- \epsilon_1 E^{p(m,v|F)} d(s_m, s_v) \\
&+ \epsilon_1 E^{p(m|F)} \sum_{i \neq m} d(s_m, s_i) \\
&- \epsilon_1 E^{p(w,t|F)} d(s_w, s_t) \\
&- \epsilon_1 E^{p(w,v|F)} d(s_w, s_v) \\
&+ \epsilon_1 E^{p(w|F)} \sum_{i \neq w} d(s_w, s_i) \\
&+ \epsilon_1^2 E^{p(m,w|F)} d(s_m, s_w)].
\end{aligned}$$

Further analysis will require making some assumptions about the distribution of m , w , t , and v .

B Four Indices Continued (Independent, Uniform Selection)

In this section we assume that m , w , t , and v are sequentially selected, uniformly from all remaining slots. That is, we take

$$p(m, w, t, v) = \frac{1}{n(n-1)(n-2)(n-3)},$$

with $p(m, n, t, v) = 0$ if any of the arguments are equal. Note that in this case the order of selection does not matter, and the sums for expected values can be taken in any order. The algorithm is independent of the fitness values (or they are assumed to all be equal at all times k) and so dependence on $F^{(k)}$ can be removed from the expected value.

The marginal distributions, e.g. $p(m)$ and $p(m, v)$, are equal to $1/n$ for single arguments and $1/(n(n-1))$ for two unequal arguments. We can therefore evaluate the expected diversity expression with these distributions and obtain

$$\begin{aligned}
E^{p(m,w,t,v|F)} E^{p(\psi_1,\psi_2)} L^{(k+1)} &= L^{(k)} + (\epsilon_1 + 2n - 3)\epsilon_2 \\
&- \sum_{t=1}^n \frac{1}{n} \sum_{i \neq t} d(s_t, s_i) \\
&- \sum_{v=1}^n \frac{1}{n} \sum_{i \neq v} d(s_v, s_i) \\
&+ \sum_{t=1}^n \sum_{v \neq t} \frac{1}{n(n-1)} d(s_t, s_v) \\
&- \epsilon_1 \sum_{m=1}^n \sum_{t \neq m} \frac{1}{n(n-1)} d(s_m, s_t) \\
&- \epsilon_1 \sum_{m=1}^n \sum_{v \neq m} \frac{1}{n(n-1)} d(s_m, s_v) \\
&+ \epsilon_1 \sum_{m=1}^n \frac{1}{n} \sum_{i \neq m} d(s_m, s_i) \\
&- \epsilon_1 \sum_{w=1}^n \sum_{t \neq w} \frac{1}{n(n-1)} d(s_w, s_t) \\
&- \epsilon_1 \sum_{w=1}^n \sum_{v \neq w} \frac{1}{n(n-1)} d(s_w, s_v) \\
&+ \epsilon_1 \sum_{w=1}^n \frac{1}{n} \sum_{i \neq w} d(s_w, s_i) \\
&+ \epsilon_1^2 \sum_{m=1}^n \sum_{w \neq m} \frac{1}{n(n-1)} d(s_m, s_w)].
\end{aligned}$$

Simplifying and writing in terms of $L^{(k)}$ we obtain

$$E^{p(m,w,t,v|F)} E^{p(\psi_1,\psi_2)} L^{(k+1)} = L^{(k)} + (\epsilon_1 + 2n - 3)\epsilon_2$$

$$\begin{aligned}
& - \frac{2L^{(k)}}{n} \\
& - \frac{2L^{(k)}}{n} \\
& + \frac{2L^{(k)}}{n(n-1)} \\
& - \epsilon_1 \frac{2L^{(k)}}{n(n-1)} - \epsilon_1 \frac{2L^{(k)}}{n(n-1)} + \epsilon_1 \frac{2L^{(k)}}{n} \\
& - \epsilon_1 \frac{2L^{(k)}}{n(n-1)} - \epsilon_1 \frac{2L^{(k)}}{n(n-1)} + \epsilon_1 \frac{2L^{(k)}}{n} \\
& + \epsilon_1^2 \frac{2L^{(k)}}{n(n-1)}.
\end{aligned}$$

Simplifying,

$$\begin{aligned}
E^{p(m,w,t,v|F)} E^{p(\psi_1,\psi_2)} L^{(k+1)} &= (\epsilon_1 + 2n - 3)\epsilon_2 + 2L^{(k)} \left[\frac{1}{2} \right. \\
& - \frac{2}{n} \\
& + \frac{1}{n(n-1)} \\
& - \epsilon_1 \frac{4}{n(n-1)} + \epsilon_1 \frac{2}{n} \\
& \left. + \epsilon_1^2 \frac{1}{n(n-1)} \right].
\end{aligned}$$

$$\begin{aligned}
E^{p(m,w,t,v|F)} E^{p(\psi_1,\psi_2)} L^{(k+1)} &= (\epsilon_1 + 2n - 3)\epsilon_2 + 2L^{(k)} \left[\frac{1}{2} \right. \\
& - \frac{2n+1}{n(n-1)} \\
& + \epsilon_1 \frac{2n-6}{n(n-1)} \\
& \left. + \epsilon_1^2 \frac{1}{n(n-1)} \right].
\end{aligned}$$

We now take the final expectation, using $E^{p(S^{(k)})} g(L(S^{(k)})) = E^{p(L)} g(L)$.

$$\begin{aligned}
E^{p(S^{(k)})} E^{p(m,w,t,v|F)} E^{p(\psi_1,\psi_2)} L^{(k+1)} &= E[L^{(k+1)}] \\
&= (\epsilon_1 + 2n - 3)\epsilon_2 + 2E[L^{(k)}] \left[\frac{1}{2} \right.
\end{aligned}$$

$$\begin{aligned}
& - \frac{2n+1}{n(n-1)} \\
& + \epsilon_1 \frac{2n-6}{n(n-1)} \\
& + \epsilon_1^2 \frac{1}{n(n-1)}].
\end{aligned}$$

To simplify the equations we define

$$\begin{aligned}
\delta_2^* &= (\epsilon_1 + 2n - 3)\epsilon_2 \\
\delta_1^* &= 2\left(\frac{1}{2} - \frac{2n+1}{n(n-1)}\right) + \epsilon_1 \frac{2n-6}{n(n-1)} + \epsilon_1^2 \frac{1}{n(n-1)}
\end{aligned}$$

Summarizing in a theorem we have

Theorem 5 *For the case of uniform selection of four distinct indices, independent of fitness,*

$$\begin{aligned}
E[L^{(k+1)}] &= \delta_1^* E[L^{(k)}] + \delta_2^* \\
E[\bar{d}^{(k+1)}] &= \delta_1^* E[\bar{d}^{(k)}] + \frac{\delta_2^*}{n^2}
\end{aligned}$$

Note that these form of the equation is the same as in Section 8, so theorems similar to the other theorems in that Section also hold. In this case it is not possible to assume the probability distribution factors into independent components, though, because we assumed distinct indices. (Two distinct indices are required for the removal position in the analysis above since otherwise it is unspecified where the two children are to be placed.)