Evolving Better Representations through Selective Genome Growth

Lee Altenberg
Institute of Statistics and Decision Sciences, Duke University,
Durham, NC 27708-0251 U.S.A. Internet: altenber@acpub.duke.edu

Abstract—

The choice of how to represent the search space for a genetic algorithm (GA) is critical to the GA's performance. Representations are usually engineered by hand and fixed for the duration of the GA run. Here a new method is described in which the degrees of freedom of the representation — i.e. the genes – are increased incrementally. The phenotypic effects of the new genes are randomly drawn from a space of different functional effects. Only those genes that initially increase fitness are kept. The genotype-phenotype map that results from this selection during the construction of the genome allows better adaptation. This effect is illustrated with the NK landscape model. The resulting genotype-phenotype maps are much less epistatic than unselected maps would be, having extremely low values of "K" — the number of fitness components affected by each gene. Moreover, these maps are exquisitely tuned to the specifics of the epistatic fitness function, creating adaptive landscapes that are much smoother than generic NK landscapes with the same genotype-phenotype maps, with fitness peaks many standard deviations higher. Thus a caveat should be made when making arguments about the applicability of generic properties of complex systems to evolved systems. This method may help to solve the problem of choice of representations in genetic algorithms.

I. Introduction

Proper representations of the search space are crucial to the performance of genetic algorithms (GAs). For a genetic algorithm to perform better than random search, the representation and genetic operators combined must contain "knowledge" about the fitness function, in the form of correlations between parental fitnesses and offspring fitness distributions, which allows the genetic operator to take fitter individuals and produce still fitter offspring with nonvanishing probability [1]. This knowledge may be implicit, fortuitous, or by design in the choice of representation. A representation that facilitates the production of fitter variants can be said to yield *evolvability*.

In traditional fixed-length GAs, the representation is created in its entirety at the outset, Athena-like from the head of the designer. Its evolvability is predetermined. One hope in genetic algorithm research has been that good representations themselves could be produced through an evolutionary approach [2, 3, 4, 5]. Here I discuss a method modeled after biological evolution for evolving representations with high evolvability.

In biological evolution, the genome has been built up incrementally by the acquisition of new genes. Many random gene additions occur through various genetic mechanisms. However, only those genes that produce an increase in fitness become stably incorporated in the genome. The nature of the gene's effect on the phenotype determines the chance that it will produce a fitness increase. Genes that disturb highly adapted organismal functions are most likely deleterious. Genes that preserve highly adapted functions while exploring novel functions have the best chance of producing a fitness increase, and thence being incorporated in the genome. Thus newly incorporated genes would tend to be modular in effect, with less deleterious side-effects (i.e. less pleiotropy). Modularity in the genotype-phenotype map would increase its evolvability[6].

In this paper, I develop an algorithm modeled after biological genome evolution as a strategy for evolving representations with high evolvability. The method is illustrated using Kauffman's NK adaptive landscape model, in which the structure of the genotype-phenotype map can be seen explicitly. The resulting adaptive landscapes are highly non-generic in their statistical properties, being much smoother than can be accounted for by their structure, allowing higher fitnesses to be reached that would be obtained through unselective generation of representations.

II. Constructional Selection

The method of evolving representations is to expand the genome one gene at a time by randomly altering the genotype-phenotype map function to include one additional argument. If that alteration reduces fitness, it is rejected, and another attempt made until an alteration that increases fitness is found. Once this is done, a conventional GA is run with this expanded genome to adapt it to a new fitness peak. Then, another round of genome expansion is run. This is illustrated in Figure 1.

This method is expressed formally as follows. Let there be n genes in the genome. The allelic value, x_i, for each gene i is an element of the gene's own domain space, G_i (for binary coding, $G_i = \{0, 1\}$; for real valued GAs, $G_i = \Re$; in genetic programming, G_i could be the function or terminal sets or both). A genotype is an array, $\boldsymbol{x} =$

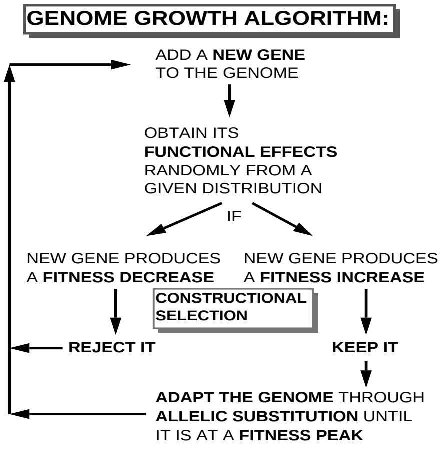
GENOME GROWTH ALGORITHM:
ADD A NEW GENE TO THE GENOME
OBTAIN ITS FUNCTIONAL EFFECTS RANDOMLY FROM A GIVEN DISTRIBUTION
IF
NEW GENE PRODUCES A FITNESS DECREASE
NEW GENE PRODUCES A FITNESS INCREASE
CONSTRUCTIONAL SELECTION
REJECT IT
KEEP IT
ADAPT THE GENOME THROUGH ALLELIC SUBSTITUTION UNTIL IT IS AT A FITNESS PEAK

Figure 1: The genome growth algorithm using constructional selection to evolve the representation.

$||x_i||_{i=1}^n \in \mathcal{G}_n$, of the allelic values, where $\mathcal{G}_n = G_1 \times G_2 \times \ldots \times G_n$ is the genotype space (in typical GAs, the domain spaces G_i are identical).

A genotype-phenotype map function, $\Psi_n : \mathcal{G}_n \mapsto \mathcal{S}$, maps the genotype $\boldsymbol{x}$ to the search space $\mathcal{S}$. A fitness function $w(s)$ is defined on each point $s \in \mathcal{S}$ in the search space.

As the genome grows, one obtains a sequence of maps:

$$\begin{array}{lcc}
\Psi_1 : & G_1 & \longmapsto \mathcal{S} \\
\Psi_2 : & G_1 \times G_2 & \longmapsto \mathcal{S} \\
\vdots & & \\
\Psi_n : & G_1 \times G_2 \ldots \times G_n & \longmapsto \mathcal{S}.
\end{array}$$

When a new gene is added to the genome, the genotype-phenotype map is modified from Ψ_n to Ψ_{n+1}. The modification is made by a probabilistic operator with probabilities $T(\Psi_n \rightarrow \Psi_{n+1})$. It is here that heuristic knowledge of the search space must be incorporated by the designer, since random search through the space of all possible genotype-phenotype maps is no better than random search through $\mathcal{S}$. One must have an *a priori* expectation that some of the modifications will produce new maps that are only incrementally changed from old maps (as in [3]). This allows one to relocate the problem of knowledge incorporation to a level where it may be more naturally accommodated.

For example, this would be the case with series representation $s = \Psi_n(\boldsymbol{x}) = \sum_{i=1}^n [a_i \cos(\omega_i) + b_i \sin(\omega_i)]$, where $x_i = (a_i, b_i, \omega_i) \in G_i$. Another example is a genetic programming approach, where new functions and terminals can be appended to the parse tree [7].

Let us return to the idea that an evolved point in the search space has some aspects that are highly adapted, and others that are not yet adapted. If this partition into subgoals is possible (and for some search spaces it may not be), then one would like to augment the genotype-phenotype map in a way that leaves the highly adapted traits unaffected while exploring the unadapted traits. Yet these "traits" (e.g. subgoals) are usually not directly observable, but are abstract, emergent properties of the fitness function. If one has *a priori* reason to believe that at least some of the modifications $\Psi_n \rightarrow \Psi_{n+1}$ will be able to correctly partition the adapted traits from the unadapted ones, then constructional selection is a means by which to incorporate those modifications into the genotype-phenotype map. This can be illustrated in a model of genome growth using Kauffman's "NK" adaptive landscape model.

III. The "NK" Adaptive Landscape Model

Kauffman's "NK" adaptive landscape model [8] will be used to illustrate the effects of constructional selection because it explicitly shows the epistatic structure of the genotype-phenotype map. The following is a generalized version of the NK model, a map between a set of genes and a set of fitness components. This is illustrated in Figure 2.

1. The genome consists of n binary-valued genes, that exert control over f phenotypic functions, each of which contributes a component to the total fitness.
2. Each gene controls a subset of the f fitness components, and in turn, each fitness component is controlled by a subset of the n genes. This genotype-phenotype map can be represented by a matrix,
$$\boldsymbol{M} = ||m_{ij}||, i = 1 \ldots n, j = 1 \ldots f,$$
of indices $m_{ij} \in \{0, 1\}$, where $m_{ij} = 1$ indicates that gene i affects fitness component j;
3. The columns of $\boldsymbol{M}$, called the *polygeny vectors*, $\boldsymbol{g}_j = ||m_{ij}||$, $i = 1 \ldots n$, give the genes controlling each fitness component j;
4. The rows of $\boldsymbol{M}$, called the *pleiotropy vectors*, $\boldsymbol{p}_i = ||m_{ij}||$, $j = 1 \ldots f$, give the fitness components controlled by each gene i;
5. If any of the genes controlling a given fitness component mutates, the new value of the fitness component will be uncorrelated with the old. Each fitness component ϕ_i is a uniform pseudo-random function[1] of the genotype, $\boldsymbol{x} \in \{0, 1\}^n$:
$$\phi_i(\boldsymbol{x}) = \Phi(\boldsymbol{x} \circ \boldsymbol{g}_i; i, \boldsymbol{g}_i) \sim \text{uniform on } [0, 1],$$
where $\Phi : \{0, 1\}^n \times \{1, \ldots, n\} \times \{0, 1\}^n \mapsto [0, 1]$, $\circ$ is the Schur product ($\boldsymbol{x} \circ \boldsymbol{g}_j = ||x_i m_{ij}||$, $i = 1 \ldots n$). Any change in i, $\boldsymbol{g}_i$, or $\boldsymbol{x} \circ \boldsymbol{g}_i$ gives a new value for $\Phi(\boldsymbol{x} \circ \boldsymbol{g}_i; i, \boldsymbol{g}_i)$ that is uncorrelated with the old;
6. If a fitness component is affected by no genes, it is assumed to be zero:
$$\Phi(\boldsymbol{x} \circ \boldsymbol{g}_i; i, \boldsymbol{g}_i) = 0 \text{ for all } \boldsymbol{x}, \text{ if } \boldsymbol{g}_i = ||0 \ldots 0||;$$
7. The total fitness is the normalized sum of the fitness components:
$$w(\boldsymbol{x}) = \frac{1}{f} \sum_{i=1}^{f} \phi_i(\boldsymbol{x}).$$

[1] Park-Miller is unsuitable. The encryption-like algorithm *ran4* [9] was used.

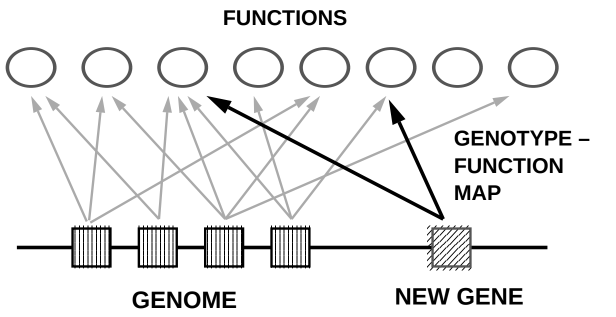

Figure 2: Kauffman's NK model recast as a map between the genotype and a set of fitness components. Arrows indicate that the gene affects the fitness component. A new gene with effects on two fitness components is shown being introduced to the genome.

A. Constructional Selection on Pleiotropy

Suppose a gene newly added to the genome has pleiotropy vector $\boldsymbol{p}_{n+1}$, and affects $k_{n+1} = \sum_{j=1}^{f} m_{n+1\,j}$ fitness components, which become resampled from the interval [0,1]. Let y be the sum, before the new gene is added, of the fitness components the new gene is going to alter. The probability that the new sum will be less than y is:

$$\begin{aligned} F_k(y) &= \Pr[S_k < y] \qquad (1) \\ &= \frac{1}{k!}\sum_{i=0}^{k}(-1)^i \binom{k}{i}\left(\frac{y-i+|y-i|}{2}\right)^k, \end{aligned}$$

where S_k is the sum of k independent uniform random variables on [0,1], from [10].

Then, from equation (1), the probability that the new gene will produce a fitness increase is $1 - F_{k_{n+1}}(y)$. When the average of the fitness components to be altered by the new gene is above $1/2$, the greater k_{n+1} is, the less the chance that the new gene will produce a fitness increase, precipitously less so for highly adapted fitness components. Since the new gene is kept only if it produces a fitness increase, constructional selection will filter out genes with high k.

Suppose that there is an underlying probability density $s(k)$ of pleiotropy values k for genes newly added to the genome. Then the density $s^*(k)$ of pleiotropy values among genes that are kept by the genome (i.e. which improve fitness) will be

$$s^*(k) = s(k) \sum_{\boldsymbol{p}\in\{0,1\}^f} \Pr[\boldsymbol{p}|k]\left[1 - F_k(\boldsymbol{p}^{\mathrm{T}}\boldsymbol{\phi})\right]/N \qquad (2)$$

where $\boldsymbol{\phi}$ is the vector of fitness components before the gene was added, $\Pr[\boldsymbol{p}|k]$ is the probability of sampling pleiotropy vector $\boldsymbol{p}$ given that the new gene's pleiotropy value is k, and N is the normalizer so that $\sum_k s^*(k) = 1$.

B. Numerical Results

A numerical simulation of constructional selection in the NK model was performed using the genome growth algorithm illustrated in Figure 1:

1. Add a new gene to the genome:
   (a) create a new pleiotropy vector $\boldsymbol{p}_{n+1}$, choosing uniformly (from $\{1,\ldots,31\}$) the number, k_{n+1}, of fitness components to be affected by the new gene, and then selecting randomly which fitness components these are, from a set of $f = 31$ possible;
   (b) pick the allelic value, x_{n+1}, of the new gene with probability $1/2$ being either 0 or 1.
2. If the new gene decreases fitness, reject it and repeat step 1. Otherwise, keep it.
3. Adapt $\boldsymbol{x}$ to the new (local) optimum $\hat{\boldsymbol{x}}$ by allelic substitution through a "greedy" 1-mutant adaptive walk.
4. Repeat step 1 until the genome has 31 genes.

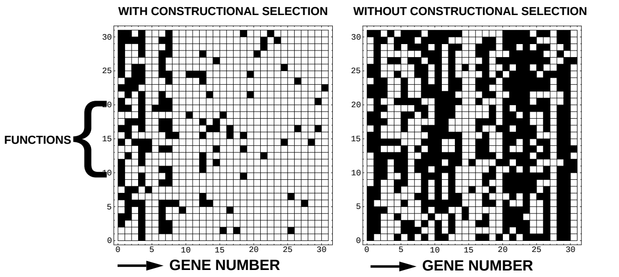

Figure 3: Two genotype-phenotype maps evolved through genome growth, with (left) and without (right) constructional selection. Dark squares indicate that fitness component j depends on gene i. The columns in the right map reflect the sampling distribution of the pleiotropy vectors, in which the number of fitness components affected is uniform on $[1, f]$. The left map shows how under constructional selection, later genes have lower pleiotropy as the genome grows and becomes more adapted.

The pleiotropy vectors, $\boldsymbol{p}_{n+1}$, are chosen from the same uniform distribution throughout the run. As a basis for comparison, the genome growth algorithm is also run without step 2, giving the result of choosing representations *a priori*.

B.1. Evolved Genotype-Phenotype Maps

Figure 3 shows typical genotype-phenotype maps produced during runs with and without constructional selection. The run without constructional selection reflects the underlying distribution of pleiotropy vectors sampled for each new gene. In the run with constructional selection, during the evolution of the first few genes, the discovery of new fitness components selects for high pleiotropy, but as these fitness components evolve toward their optima, selection becomes strong against new genes affecting them.

This increasing selection for low pleiotropy can be seen in Figure 4, which shows the distribution of pleiotropies k_n as the genome grows, over repeated runs of genome growth. The mode for k_n is always 1 after the first few genes, but as shown in Figure 5, the mean k_n tends toward 1 from initial values of around 16, or half of the maximum possible, $f = 31$.

The progress in adaptation can be compared between runs with and without constructional selection. Figure 6 shows plots for a number of runs. Without constructional

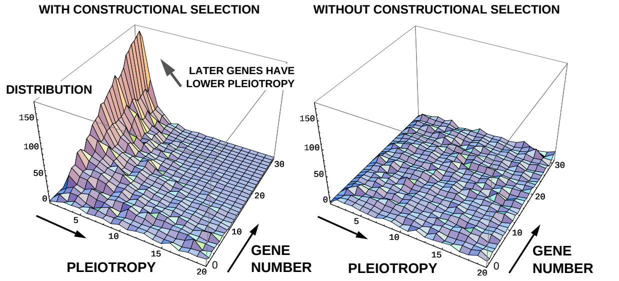

Figure 4: The distribution, from repeated runs of the genome growth algorithm, of pleiotropy values k_n, from each gene's pleiotropy vector p_n, as the genome grows; with (left) and without (right) constructional selection.

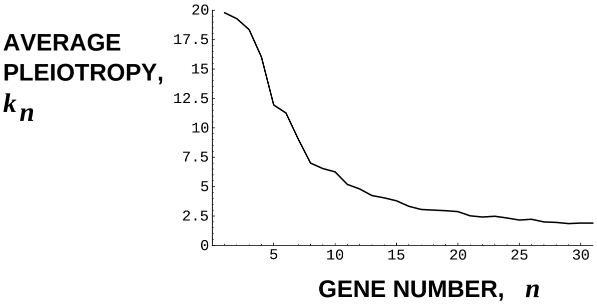

Figure 5: The average pleiotropy values k_n for each gene as the genome grows, from the runs in Figure 4, with constructional selection.

selection, disruptive new genes are not filtered out, and adaptation shows little progress once the fitness components are saturated with genes that affect them. With constructional selection, however, fitness continues to increase with each new gene throughout the genome growth.

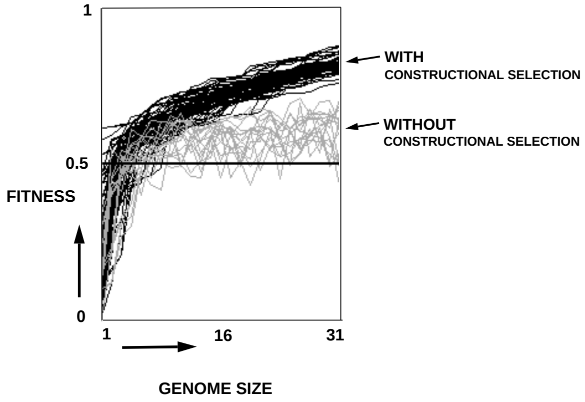

Figure 6: Fitness as a function of genome size for several runs of the genome growth algorithm. Dark lines are with, and light lines without constructional selection.

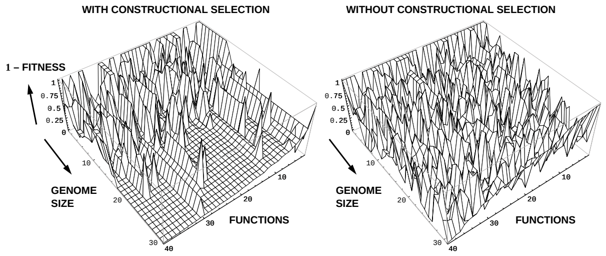

Figure 7: Fitness components during genome growth, for one genome evolved with (left) and one without (right) constructional selection. Fitness components are sorted according to their value at the end of the run.

As the genome grows, the trajectories of individual fitness components can be seen in Figure 7. With constructional selection, once a fitness component has reached a high value (low points in graph), only new genes that leave it alone are likely to be incorporated in the genome. Occasionally, however, one component is sacrificed for the improvement in another, which show up as spikes in the graph. By the time the genome has reached a size of 31 genes, most of the components have reached values well above their expected value of 1/2. Without constructional selection, the jumble of spikes represents the continuing randomization of the fitness components as genes with random pleiotropy are incorporated into the genome.

Here, most of the adaptation under constructional selection occurs during the incorporation of new genes, rather than during the adaptive walks (through allelic substitution) between gene additions. This is because there is a much larger pool of new pleiotropy vectors to sample from than the pool of genotypes in the 1-mutant neighborhood of an existing genotype (2^f vs. n). The evolutionary process under constructional selection is figuratively the "building" of a fitness peak, gene by gene, rather than the climbing of a fitness peak.

B.2. Non-Generic Properties of Evolved Landscapes

Existing theory for adaptive walks on NK landscapes [11, 8, 12, 13] is derived for generic landscapes, i.e. landscapes that one would typically obtain from a random sampling of landscapes with given values of n and k. The applicability of these results to biological examples assumes that evolutionary processes produce such generic adaptive landscapes. However, the distribution of fitness peaks in the NK landscapes grown here under constructional selection are nowhere near the distributions for generic NK landscapes with identical genotype-phenotype maps.

Constructional selection produces genotype-phenotype maps that are much more finely tuned to the fitness function under which they evolved. To illustrate this, the distribution of fitness peaks for several landscapes evolved under constructional selection are plotted in figure 8. For comparison, distributions are plotted for landscapes using the same genotype-phenotype map, but with fitness functions, Φ, chosen *a priori*. Each point represents the fitness peak obtained by starting an adaptive walk from a randomly sampled genotype. The distributions are plotted by sorting the fitness peaks by size (the transpose of the figure therefore represents the cumulative probability distribution for fitness peaks). The width of horizontal plateaus represents the size of the domain of attraction for a particular fitness peak.

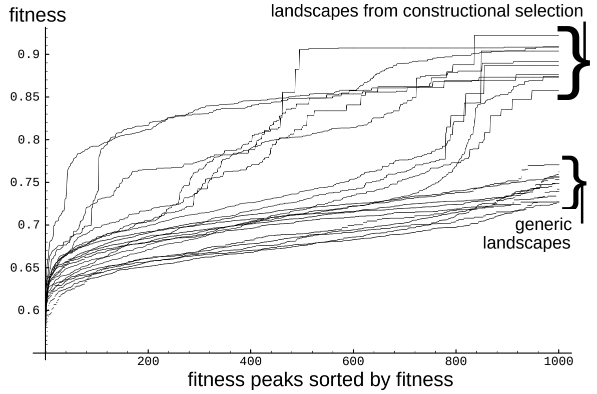

Figure 8: Distributions of fitness peaks of NK landscapes: upper 10 plots are for adaptive landscapes evolved under constructional selection; lower 10 plots are with the same genotype-phenotype maps but randomized fitness functions. In each plot, the peaks attained from 1000 random starting genotypes are sorted by fitness. Plateaus indicate large domains of attraction for the peak.

The plateaus, and discontinuities between them, indicate fewer and larger domains of attraction for the evolved landscapes, i.e. they are smoother than the generic landscapes. The distributions for the generic landscapes follow roughly the Gaussian approximation derived by Weinberger [13]. While the least fit peaks are approximately the same for both evolved and generic landscapes, at various points in the ranking, the fitness of the evolved landscapes grows much higher. Interestingly, the jumps in the distribution are highly variable.

An additional beneficial outcome of constructional selection is that the genotypes resulting at the end of the run are usually the apparent global fitness peak. In 77% of adaptive landscapes evolved under constructional selection (304 sampled), the genotypes attained at the end of genome growth were fitter than any other adaptive peak found (from 250 other starting genotypes). Of the remaining landscapes, only 19% of the peaks arrived at from random initial genotypes were fitter than the genotype attained at the end of genome growth.

C. Computational Complexity

The usefulness of constructional selection for genetic algorithms should be measured by the number of points in the search space that need to be evaluated to reach a certain fitness. There are several strategies for spending computational resources: One can take an *a priori* representation, run numerous adaptive walks from different initial genotypes, and take the fittest peak attained after expenditure of the set amount of computations. Alternately, one can partition one's computations sampling among different genotype-phenotype maps to find that with the fittest peak. Lastly, the approach examined here is to progressively modify simple genotype-phenotype maps to more complex ones with the genome growth algorithm described. The payoff in the level of optimization obtained through constructional selection shows it to be a much more efficient.

As a simple example, consider the class of NK landscapes with the highest expected fitness peaks, the $k = 1$ landscapes (which is $K = 0$ in Kauffman's original definition), in which there is a one-to-one map from each gene to each fitness component, and $f = n$. Each gene can be optimized individually, so it takes the evaluation of $2n$ genotypes to find the global peak for the landscape. Each fitness component is i.i.d., where the optimal $\hat{\phi}_i$ is distributed as the maximum of two independent uniform random variables on [0,1]. The probability density of each maximum $\hat{\phi}_i$ is $f(\hat{\phi}) = 2\hat{\phi}$. So $\mathrm{E}(\hat{\phi}) = 2/3$ and $\mathrm{Var}(\hat{\phi}) = 1/18$. With $n = f = 31$ genes and fitness components, one obtains:

$$\mathrm{E}[w(\hat{\boldsymbol{x}})] = 2/3, \text{ and } \mathrm{Var}[w(\hat{\boldsymbol{x}})] = \frac{1}{18f} \approx 0.00179.$$

The average fitness attained under constructional selection in the numerical simulations (where the average k is much larger than 1) is about 0.89, which is therefore some 5 standard deviations (5×0.0423) above the expected value of peaks obtained from generic $k = 1$ landscapes. The probability of a $k = 1$ landscape having a peak with fitness at least 0.89, under a normal approximation (an overestimate), is 3×10^{-7}. So a strategy of sampling different random $k = 1$ landscapes would require over 2×10^6 samples to be likely to obtain fitness peaks as high as those obtained through constructional selection, which in the runs here took some 3000 evaluations of the fitness function.

The ability to select on the genotype-phenotype map as it is constructed is the key to finding higher fitness values. To carry out constructional selection in the $k = 1$ case, one could add one gene at a time and for each choice of the fitness component it maps to, examine the fitnesses for both alleles, and keep the allele and map that give the fittest value. The first gene would be evaluated for the f possible genotype-phenotype maps, for a total of $2f$ evaluations. The second gene would be sampled with the remaining $f-1$ unmapped fitness components, and so forth, giving a total of $f(f+1)$ evaluations. Each resulting fitness component would be the maximum of $2(f-i+1)$ uniform i.i.d. values, for the ith gene/gene map pair. So the ith fitness component would be distributed as $F_i(\hat{\phi}) = \hat{\phi}^{2(f-i+1)}$. The fitness components would have expectation

$$\mathrm{E}(\hat{\phi}_i) = \frac{2(f+1-i)}{2(f+1-i)+1}.$$

The expected value for the fitness peaks obtained through this constructional selection process would be

$$\mathrm{E}[w(\hat{\boldsymbol{x}})] = \frac{1}{f}\sum_{i=1}^{f}\frac{2i}{2i+1}.$$

For $n = f = 31$, the expected fitness peak would be $\mathrm{E}[w(\hat{\boldsymbol{x}})] \approx 0.945$, and would take $31 \times 32 = 992$ evaluations to find. This is a much more efficient use of computational resources than randomly searching generic $k = 1$ landscapes for those with the highest peaks.

D. *Constructional Selection is a Novel Evolutionary Mechanism*

Mechanisms that have classically been proposed for the evolution of evolvability all invoke allele substitution at pre-existing loci (including Kauffman's suggestions as to how K could evolve [14]). These include:

1. smooth landscapes as a side-effect of alleles that produce advantageous phenotypic stability (e.g. stable proteins, stable developmental pathways);
2. neutral alleles that modify the landscape and hitchhike with later advantageous mutations they facilitate.

The evolution of smooth landscapes through the filtering of loci as they come into being is distinct from allelic substitution and does not require the particular effects mentioned above. Although models of gene duplication and evolution have been analyzed in evolutionary population genetics [15], the systematic effect of producing a more evolvable genotype-phenotype map has been mentioned in few sources [6, 16, 17].

Experiments of evolving computational genotype-phenotype maps using constructional selection should be explored for what insights they may provide about the structure of natural genotype-phenotype maps. The results here show that it may be wrong to assume that evolved adaptive landscapes follow patterns of mathematically generic landscapes.

IV. Conclusions

I have described a method for evolving representations for GAs with the goal of improving their performance. The method is modeled after the process of biological genome evolution, in which newly created genes become stably incorporated in the genome only when they produce a fitness increase. This selection in the construction of the genome is expected to filter out genes that disrupt highly adapted traits, biasing the evolution of the genotype-phenotype map toward a modular structure. Such a modular structure in the representation used for a GA would reduce epistatic interactions between genes and confer greater evolvability. This predicted outcome is tested using Kauffman's NK landscape model.

In the NK model, by simply adding new genes with random effects to the genome and rejecting those that reduce fitness, genotype-phenotype maps evolve very little epistasis between genes. But the resulting adaptive landscapes are smoother, with higher peaks, than even low epistasis can account for.

These results suggest that much more efficient algorithms may be obtained if there is an opportunity to build up representations incrementally, gene by gene, and keep only those additions to the representation that produce a fitness increase.

A caveat should be made to the application of adaptive landscape theory to real organisms. The evolutionary processes that constructed their genomes may result in genotype-phenotype maps that are not well described by generic models. Constructional selection provides a novel mechanism to produce the low values of K hypothesized by Kauffman [14].

Acknowledgements

This research was supported in part by the Santa Fe Institute, The Center for Nonlinear and Complex Systems at Duke University, Provost's Common Fund, and NSF Grant EAR-89-15983. It grew from Stuart Kauffman's discussion of the NK landscape model at the Santa Fe Institute's 1991 Complex Systems Summer School. I thank: Peter Haff, Marcy Uyenoyama and Richard Palmer for computational support; Joe Felsenstein, Frank Eeckman, Günter Wagner and Eric Mjolsness for the invitations to speak which facilitated this work.

References

[1] L. Altenberg, "The evolution of evolvability in genetic programming," in *Advances in Genetic Programming* (K. E. Kinnear, ed.), (Cambridge, MA), MIT Press, 1994.

[2] C. G. Shaefer, "The ARGOT strategy: adaptive representation genetic optimizer technique," in *Genetic Algorithms and their Applications: Proceedings of the Second International Conference on Genetic Algorithms* (J. J. Grefenstette, ed.), (Hillsdale, NJ), pp. 50–58, Lawrence Erlbaum Associates, 1987.

[3] I. Harvey, "The basis for a continuing SAGA," in *Toward a Practice of Autonomous Systems. Proceedings of the First European Conference on Artificial Life* (F. J. Varela and P. Bourgine, eds.), (Cambridge, MA), pp. 346–354, M.I.T. Press, 1992.

[4] N. Schraudolph and R. Belew, "Dynamic parameter encoding for genetic algorithms," *Machine Learning*, vol. 9, no. 1, pp. 9–21, 1992.

[5] D. Szarkowicz, "A multi-stage adaptive-coding genetic algorithm for design applications," in *Proceedings of the 1991 Summer Computer Simulation Conference* (D. Pace, ed.), (San Diego, CA), pp. 138–144, 1991.

[6] L. Altenberg, "Knowledge representation in the genome: new genes, exons, and pleiotropy," *Genetics*, vol. 110, supplement, p. s41, 1985. Abstract of paper presented at the 1985 Meeting of the Genetics Society of America.

[7] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press, 1992.

[8] S. A. Kauffman, "Adaptation on rugged fitness landscapes," in *Lectures in the Sciences of Complexity* (D. Stein, ed.), pp. 527–618, Redwood City: Addison-Wesley, 1989. SFI Studies in the Sciences of Complexity, Lecture Volume I.

[9] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing. Second Edition*. Cambridge University Press, 1992.

[10] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: John Wiley and Sons, 1971.

[11] S. A. Kauffman and S. Levin, "Towards a general theory of adaptive walks on rugged landscapes," *Journal of Theoretical Biology*, vol. 128, pp. 11–45, 1987.

[12] C. A. Macken and A. S. Perelson, "Protein evolution on rugged landscapes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, pp. 6191–6195, 1989.

[13] E. D. Weinberger, "Local properties of kauffman's N-k model, a tuneably rugged energy landscape," *Physical Review A*, vol. 44, no. 10, pp. 6399–6413, 1991.

[14] S. A. Kauffman, "Principles of adaptation in complex systems," in *Lectures in the Sciences of Complexity* (D. Stein, ed.), pp. 619–712, Redwood City: Addison-Wesley, 1989. SFI Studies in the Sciences of Complexity, Lecture Volume I.

[15] T. Ohta, "Further simulation studies on evolution by gene duplication," *Evolution*, vol. 42, pp. 375–386, 1988.

[16] R. J. Riedl, "A systems-analytical approach to macroevolutionary phenomena," *Quarterly Review of Biology*, vol. 52, pp. 351–370, 1977.

[17] W. F. Doolittle, "The origin and function of intervening sequences in DNA: A review," *American Naturalist*, vol. 130, pp. 915–928, 1987.