A General Coarse-Graining Framework for Studying Simultaneous Inter-Population Constraints Induced by Evolutionary Operations

Keki Burjorjee
kekib@cs.brandeis.edu

Jordan B. Pollack
pollack@cs.brandeis.edu

DEMO Lab
Computer Science Department
Brandeis University
Waltham, MA 02454

ABSTRACT

The use of genotypic *populations* is necessary for adaptation in Evolutionary Algorithms. We use a technique called *form-invariant commutation* to study the *immediate* effect of evolutionary operations on *populations* of genotypes. This technique allows us to understand compositional changes induced by evolutionary operations in terms of *constraints* between populations. Deep insight into the population-level effect of some evolutionary operation is possible when multiple constraints can be derived for all pairs of pre and post operative populations; for each such pair of populations the constraints between them are then said to hold *simultaneously*. When selection is fitness proportional we show that *any* coarse-graining of the genotype set can be used to systematically derive single constraints between between all pairs of pre and post selection populations. Matters are not so simple in the case of variation. We develop an abstract condition called *ambivalence* and show that when a coarse-graining and a variation operation satisfy this condition then a systematic derivation of single constraints between all pairs of pre and post variation populations is possible. Finally we show that it is possible to use schema partitions to systematically derive *simultaneous* constraints for any combination of variation operations that are commonly used in GAs.

Categories and Subject Descriptors

I.2.8 [**Computing Methodologies**]: Artificial Intelligence—*Problem Solving, Control Methods, and Search*; F.2 [**Theory of Computation**]: Analysis of Algorithms And Problem ComplexityMiscellaneous

General Terms

Algorithms, Theory

Keywords

Genetic Algorithms, Schema Theory, Coarse-Graining, Constraints

1. INTRODUCTION

The use of genotypic *populations* is necessary for adaptation in Evolutionary Algorithms (EAs)— this is evident from the degradation in the performance of a typical evolutionary algorithm as its genotypic population size tends to one. So, in order to understand how evolutionary algorithms perform adaptation, it is important to understand the effect of the evolutionary operations — selection and variation — on *populations* of genotypes.

One might well question whether there is anything to really understand about the effect of evolutionary operations on genotypic populations; after all, the subroutines that implement these operations are quite simple and readily understandable. What is missed by this question is that the simplicity of the subroutines only allows one to grasp their effect on any individual genotype, or on any pair of parent genotypes, *not* on some *population* of genotypes. One might then contend that the effect of evolutionary operators on populations of genotypes is just the "sum of parts", i.e. that one will not learn anything new about adaptation in EAs by focusing one's attention on populations as opposed to individual genotypes. As will be seen, this contention is false. In this paper we derive adaptation relevant theoretical results about the population level effects of evolutionary operations; these results are not immediately apparent from the action of evolutionary operations on individual genotypes or pairs of genotypes.

Vose pioneered the use of mathematical operators to express the effects of evolutionary operations on genotype *populations*. He used these operators to study the long term behavior of the Simple Genetic Algorithm (SGA). The results of this study, which is presented in [11] and summarized in [12], showed that, given the fixed points of an infinite population model of an SGA, in the long-term the finite population of some SGA 'hovers in the vicinity' of one of the fixed points, making an occasional (i.e. low probability) transition to the 'vicinity' of some other fixed point, and so on, ad infinitum.

We have found Vose's technique of using mathematical operators to express the effect of evolutionary operations on *populations* of genotypes to be quite useful (the use of this technique is essential to the analysis in this paper), however, because the analytical approach taken in [11] focuses on the *long-term* behavior of a GA, we find it improbable that it will yield insight into how GAs adapt phenotypes. Holland has criticized the approach taken in [11] for similar reasons. In [5] he says, "Standard mathematical approaches, such as

Markov processes and statistical mechanics, typically offer little insight into the dynamics of processes that involve the nonlinearities inherent in coadaptations".

Adaptation occurs during the *transient* behavior of an EA [6]. Hence, to understand how this adaptation occurs, it is important to understand the *immediate* changes in composition that evolutionary operations induce in genotypic populations. We show that we can understand these changes in terms of *simultaneous constraints* on the composition of populations.

Studies of the inter-generational effects of evolutionary operations on the frequencies of mathematical objects called schemata are described in [4, 3, 7]. In these studies schemata are said to be "processed in parallel" by the evolutionary operations of a GA. This language gives the impression that evolutionary operations implicitly implement some form of parallel computation in the platonic realm of schemata. From an adaptation standpoint what ultimately matters is the evolving population of phenotypes and, indirectly, the composition of the evolving population of genotypes. Thus the "schemata processing" view posited by these studies begs the (almost philosophical) question "How does 'processing' in the platonic realm of schemata impact changes in the composition of an evolving population of genotypes?". We avoid this question entirely by keeping our focus firmly fixed on the compositional changes occurring in the evolving population of genotypes — we interpret each theoretical result about some evolutionary operation as a statement about how a population, after the operation (post-operative population), is related to the population before the operation (pre-operative population). These statements describe inter-population *constraints* that the evolutionary operations induce on the composition of these populations.

We call the analytical technique that we use to obtain these constraints *form-invariant commutation*. This type of technique (which has been previously used in [11, 10, 13]) uses a 'tool' that is commonly used in Physics, especially statistical mechanics, called coarse-graining. Coarse-graining is widely applied in Physics to "solve for" certain statistical properties of a system. The idea there is that by projecting the microscopic level equations of a system over the original high dimensional state space to an appropriate lower dimensional state-space it becomes possible — by solving the coarse-grained (i.e. macrospic level) equations — to obtain some statistical property of the system (e.g. temperature). Coarse-graining has been applied in the same spirit to genetic algorithms in [11, 10, 13, 8, 2]. In these works the statistical property sought (but only obtained for a very specific fitness function — unitation — in [2]) is a transition matrix specifying the exact or approximate evolutionary dynamics in the coarse-grained space. In this paper we will be using coarse-graining in a *different* spirit; our aim is not so much to "solve" for some statistical property of an evolutionary system, as to understand how, for each evolutionary operation, the composition of some post-operative population is constrained by the composition of the pre-operative population.

The rest of this paper is organized as follows: in Section 2 we introduce mathematical operators that model the effects of evolutionary operations on populations of genotypes. The technique of form-invariant commutation that we use to study these effects is introduced in Section 3.

Specifically we discuss how coarse-grainings of the genotype set can be used derive inter-population constraints that are induced by some evolutionary operation — one constraint per coarse-graining. We argue that deep insight into the effect of some operation is possible only when one can systematically derive multiple *simultaneous* constraints over all pairs of pre and post-operative populations. In Section 4 we show that *any* coarse-graining of the genotype space can be used to systematically derive single constraints over all pairs of pre and post selection populations. Thus deriving simultaneous constraints for any such pair is easy. Matters are not so simple in the case of variation. An important contribution of this paper is the development, in Section 5, of an abstract condition called *ambivalence*. We show that when some coarse-graining meets this condition with respect to some specific kind of variation, then it can be used to systematically derive single constraints between all pairs of pre and post variation populations. Finally in Sections 6 and 7 we show that when genotypes are bitstrings of fixed length and variation is performed by some combination of common GA variation operations — 'canonical' mutation, n-point crossover, and uniform crossover — then we can use coarse-grainings that induce schema partitions over the genotype set to obtain multiple *simultaneous* constraints over all pairs of pre and post variation populations.

2. MATHEMATICAL PRELIMINARIES

For some sets X, Y, and some function $\beta : X \to Y$, we use the notation $\langle y \rangle_\beta$ to denote the preimage of y, i.e. the set $\{x \in X \mid \beta(x) = y\}$.

As in [9], for any set X we use the notation Λ^X to denote the set of all distributions over X, i.e. Λ^X denotes set $\{f : X \to [0, 1] \mid \sum_{x \in X} f(x) = 1\}$. For any set X, let $0^X : X \to \{0\}$ be the constant zero function over X. Even though 0^X is not strictly speaking a distribution, we will find it very useful in this paper to call it the *null* distribution over X. We use the notation Ω^X to denote the set of all null and non-null distributions over X, i.e. $\Omega^X = \Lambda^X \cup \{0^X\}$.

2.1 Transmission Functions

The idea of a transmission function was used in [1] to model the behavior of a recombination operator that takes two parents as input and produces one child as output. We make a minor extension of this idea in order to be able to model variation operators with arbitrary numbers of parents. For any set X, an m-parent transmission function over X is an element of the set

$$\left\{ T : \prod_1^{m+1} X \to [0, 1] \;\middle|\; \forall x_1, \ldots, x_m \in X, \right.$$

$$\left. \sum_{x \in X} T(x, x_1', \ldots, x_m') = 1 \right\}$$

Extending the notation introduced above, we denote this set by Λ_m^X. A transmission function is simply a conditional distribution, so following [9], we use conditional probability notation to denote transmission functions. Thus an m-parent transmission function $T(x, x_1, \ldots, x_m)$ is denoted $T(x | x_1, \ldots, x_m)$. Suppose a variation operator takes m 'parent' genotypes as arguments and stochastically outputs a single 'child' genotype, then this composition is modelled by an m-parent transmission function T such that for any m

genotypes $g_1, \ldots g_m$, $T(g|g_1, \ldots, g_m)$ is the probability that variation will produce the child g when acting on parents $g_1, \ldots, g_m$.

Let $0_m^X : \prod_1^{m+1} X \to \{0\}$ be the constant zero function over $\prod_1^{m+1} X$. Though 0_m^X is not strictly speaking a transmission function, we find it useful to call it the *null m-parent, transmission function* over X. We use the notation Ω_m^X to denote the set of all null and non-null m-parent, transmission functions over X, i.e. $\Omega_m^X = \Lambda_m^X \cup \{0_m^X\}$. For any $x_1, \ldots, x_m \in X$ and for any transmission function $T \in \Omega_m^X$, $T(\cdot|x_1, \ldots, x_m) \in \Omega^X$ denotes the distribution over X obtained by conditioning T with $x_1, \ldots, x_m$.

Given this definition of a transmission function, it can be used to model a mutation operation (which operates on one parent and produces one child), and indeed any variation operation which operates on any numbers of parents and produces one child.

For some genotype set G, let $T_1 \in \Lambda_m^G$ and $T_2 \in \Lambda_n^G$ be transmission functions that model m and n-parent variation operations respectively. Then the effect of applying these two operations one after the other can be modelled by a third transmission function: suppose the m-parent variation operation is applied to the output of the n-parent operation, then the transmission function that models the effect of the composite variation is given by $T_1 \circ T_2$, where composition of transmission functions is defined as follows:

DEFINITION 1. (COMPOSITION OF TRANSMISSION FUNCTIONS) *For any $T_1 \in \Omega_m^X, T_2 \in \Omega_n^X$, we define the composition of T_1 with T_2, denoted $T_1 \circ T_2$, to be a transmission function in Ω_n^X given by*

$$(T_1 \circ T_2)(x|y_1, \ldots, y_n) =$$

$$\sum_{\substack{(x_1, \ldots, x_m) \in \\ \prod_1^m X}} T_1(x|x_1, \ldots, x_m) \prod_{i=1}^m T_2(x_i|y_1, \ldots, y_n)$$

2.2 Dynamics Related Operators

Genotypic populations in an EA are modelled as non-null distributions over the genotype set. The effect of evolutionary operations on populations is modeled by the application of mathematical operators to distributions.

DEFINITION 2. (EXPECTATION OPERATOR) *Let X be some countable set, and let $f : X \to \mathbb{R}^+$ be some bounded function. We define the expectation operator $\mathcal{E}_f : \Omega^X \to \mathbb{R}^+ \cup \{0\}$ as follows:*

$$\mathcal{E}_f(p) = \sum_{x \in X} f(x)p(x)$$

Allowing the domain of the expectation operator to include a null-distribution in this definition, and also in the following definitions of the selection operator, transmission operator, evolution epoch operator, and projection operator is a mathematical trick which makes the proofs of the theorems in this paper simpler.

The selection operator, defined below, can be used to model the effect of fitness proportional selection on a population of genotypes.

DEFINITION 3. (SELECTION OPERATOR) *Let X be some set and let $f : X \to \mathbb{R}^+$ be some bounded function. We*

define the Selection Operator $\mathcal{S}_f : \Omega^X \to \Omega^X$ as follows: if $\mathcal{E}_f(p) = 0$, then $\mathcal{S}_f(p) = 0^X$. Otherwise,

$$(\mathcal{S}_f p)(x) = \frac{f(x)p(x)}{\mathcal{E}_f(p)}$$

Observe that that for any non-null distribution $p \in \Lambda^X$, $(\mathcal{S}_f p)(x) > p(x)$ if and only if $f(x) > \mathcal{E}_f(p)$; also observe that $(\mathcal{S}_f p)(x) < p(x)$ if and only if $f(x) < \mathcal{E}_f(p)$. Thus, selection rewards (i.e. increases the distribution mass of) any instantiated element of X whose fitness is higher than the weighted average fitness of all elements of X, and penalizes (i.e. decreases the distribution mass of) any instantiated element of X whose fitness is lower than the weighted average fitness of all elements of X.

DEFINITION 4. (VARIATION OPERATOR[1]) *Let X be a set, and for any $m \in \mathbb{N}^+$, let $T \in \Omega_m^X$ be a transmission function over X. We define the variation operator $\mathcal{V}_T : \Omega^X \to \Omega^X$ as follows:*

$$(\mathcal{V}_T p)(x) = T(x|x_1, \ldots, x_m) \prod_{i=1}^m p(x_i)$$

The variation operator can be used to model the effect of the application of the variation operators of an EA to some population of genotypes.

2.3 Projection Related Definitions

The next definition describes the projection operator (previously used in [11] and [9]). A projection operator 'projects' a distribution that is defined over the domain of some function, 'through' that function, to the co-domain. It does this by assigning to each element in the co-domain the sum of the distribution mass of all the members of the preimage of that element. As the function that the projection operator projects 'through' is not always the same, we make this function explicit in our notation by parameterizing the operator with it.

DEFINITION 5. (PROJECTION OPERATOR) *Let X, Y be some sets and let $\beta : X \to Y$ be a function. We define the projection operator, $\Xi_\beta : \Omega^X \to \Omega^Y$ as follows:*

$$(\Xi_\beta p)(y) = \sum_{x \in \langle y \rangle_\beta} p(x)$$

and call $\Xi_\beta p$ the β-projection of p.

Let $p \in \Lambda^X$ be some distribution and let $\beta : X \to K$ be some function. For any $k \in K$, the next definition allows us to create a new distribution over $\langle k \rangle_\beta$ by normalizing the probability masses of all elements in $\langle k \rangle_\beta$. We call this new distribution the (β, k)-conditional of p. Formally,

---

[1]also called the Mixing Operator in [11] and [9]

DEFINITION 6. $((\beta, k)$-CONDITIONAL OF A DISTRIBU-
TION) *Let X, K be sets and let $\beta : X \to K$ be a function. Let $p \in \Omega^X$ be some distribution. For any $k \in K$, we define the (β, k)- conditional of p to be a distribution $q \in \Omega^{\langle k \rangle_\beta}$ s.t. if $(\Xi_\beta p)(k) = 0$, then $q = 0^{\langle k \rangle_\beta}$. Otherwise,*

$$q(x) = \frac{p(x)}{(\Xi_\beta p)(k)}$$

2.4 Some Function Related Terminology

A coarse-graining is just a function from a genotype set to some other set. For any coarse-graining the following definition introduces terminology that makes it convenient to talk about the coarse-graining's co-domain, the elements in it's codomain, the partition that it induces over the domain, and the equivalence classes of this partition.

DEFINITION 7. (β-THEME SET, β-THEMES β-CLASS, β-PARTITION) *Let X, K be sets and let $\beta : X \to K$ be a coarse-graining. We call the co-domain K the β-theme set, call an element in K a β-theme, call the preimage $\langle k \rangle_\beta$ of some $k \in K$, the β-class of k, and call the set of (clearly non-intersecting) β-classes of k, for all $k \in K$, the β-partition.*

For any $x \in X$ and $k \in K$ when $\beta(x) = k$, we say that *x β-instantiates k.*

3. INTER-POPULATION CONSTRAINTS

In this section we introduce the technique that we use to derive the inter-population constraints that are induced by evolutionary operators. We call this technique Form-Invariant Commutation. It was first used in [11] and has since been used in [10] and [13]. In all of these works however, the goal with which this technique was applied — to solve for the *asymptotic* behavior of GAs — differs from our goal, which is, to understand the *immediate* compositional changes in genotypic populations that are induced by evolutionary operations.

The gist of the technique is as follows: consider some sets G, K, and some function $\beta : G \to K$. Suppose $\mathcal{W}$ is a well understood parameterizable operator (like say the selection operator or the variation operator from the previous section) such that when parameterized by the object x, $\mathcal{W}$ models some algorithmic operation over populations of objects in G. For any population $p_G \in \Lambda^G$, let $p'_G = \mathcal{W}_x p_G$, let $p_K = \Xi_\beta p_G$ be the population of β-themes that is obtained by projecting p_G through β, and let $p'_K = \Xi_\beta p'_G$ be a population of β-themes that is obtained by projecting p'_G through β. If we derive an "understandable" object y such that the following diagram commutes

then we have succeeded in deriving a single "understandable" constraint that illuminates how the composition of population p'_G is constrained by the composition of p_G. This constraint is depicted by the dashed arrows in the diagram above. In line with the terminology used in chapter 17 of [11] we call y the *quotient parameter* of the constraint. A single constraint by itself is of limited use in understanding the

effect of $\mathcal{W}_x$. Deeper insight into how an evolutionary operator changes the composition of some population is possible if one can find multiple coarse-grainings $\beta_1 \ldots \beta_n$ such that each coarse-graining in this set induces a different partition over the genotype set and the following diagram commutes:

If one is able to find such a set of coarse-grainings then one can derive several constraints each of which provides different information about how the composition of p'_G is constrained by the composition of p_G. We say that these constraints hold *simultaneous*.

4. SELECTIONAL CONSTRAINTS

For any bounded fitness function $f : G \to \mathbb{R}^+$, and any pair of populations p_G, and p'_G such that $p'_G = \mathcal{S}_f p_G$, the following theorem gives us *simultaneous* constraints between p_G and p'_G. Remarkably it says that *any* coarse-graining of G can be used to derive a selectional constraint between p_G and p'_G.[2]

THEOREM 1 (SELECTIONAL CONSTRAINTS THEOREM). *Let G and K be countable sets, let $\beta : G \to K$ be any coarse-graining of G, and let $f : G \to \mathbb{R}^+$ be some bounded function, let $p_G \in \Omega^G$ be a distribution. For all $k \in K$, let $p_{\langle k \rangle_\beta} \in \Omega^{\langle k \rangle_\beta}$ be the (β, k)-conditional of p_G. Let $F_{(\beta, p_G)} : K \to \mathbb{R}^+$ be defined as $F_{(\beta, p_G)}(k) = \mathcal{E}_f(p_{\langle k \rangle_\beta})$. Then for all $k \in K$,*

$$(\Xi_\beta \circ \mathcal{S}_f p_G)(k) = (\mathcal{S}_{F_{(\beta, p_G)}} \circ \Xi_\beta p_G)(k)$$

i.e. letting $p'_G = \mathcal{S}_f p_G$, $p_K = \Xi_\beta p_G$ and $p'_K = \Xi_\beta p_K$, the following diagram commutes

As $F_{(\beta, p_G)}$ assigns to each β-theme $k \in K$ the expected fitness of genotypes that β-instantiate k, we call $F_{(\beta, p_G)}$ the *β-theme fitness function given p_G*. For any β-theme $k \in K$ we call $F_{(\beta, p_G)}(k)$ the β-theme fitness of k given p_G

For any pair of populations p_G, p'_G such that $p'_G = \mathcal{S}_f p_G$ the constraint corresponding to some coarsegraining $\beta : G \to K$ is as follows: if any β-theme k has β-theme

---

[2]Due to space restrictions proofs have been omitted. They can be found in the full version of this paper which is posted on the first author's website

fitness given p_G that is greater than (or less than) the expected β-theme fitness given p_G then there will be 'more' (or correspondingly, 'fewer') genotypes that β-instantiate k in p'_G.

Form-invariant commutation of the selection operator has been studied in [11, 10, 13] using coarse-grainings that map fixed length strings to schemata. In these studies the authors succeeded in 'forcing' form-invariant commutation to obtain a quotient parameter for the selection operator. What we call theme fitness in this paper was called the 'utility' of a schema in [10]. The authors of these works do not however show that the utility of a schema relative to some population is simply the expected fitness of the genotypes of the population that instantiate the schema.

Moreover upon finding that the utility of each schema is always dependent on a population the authors deemed the results of their analyses to be unuseful because population-independent transition matrices that describe the effect of the selection operator on schemata are unobtainable, and because in the absence of such transition matrices one is unable to perform a Markov chain analysis to solve for the asymptotic coarse-grained behavior of an EA.

Our goal is to understand the *transient* behavior of an EA by studying the constraints on the composition of populations that are induced by evolutionary operations. Given this purpose, the dependence of the quotient parameter on the pre-selection population is not problematic provided that the nature of this dependence is well understood.

5. VARIATIONAL CONSTRAINTS

We have shown that we can use *any* coarse-graining to systematically derive constraints on the composition of any pair of pre and post selection populations. A systematic derivation of constraints for all pairs of pre and post variation populations is not as simple. In the following we introduce a condition on coarse-grainings and transmission functions called ambivalence. We show that if this condition is met by some coarse-graining β and some transmission function T, then, for all pairs of populations p_G and p'_G such that $p'_G = \mathcal{V}_T p_G$, we can use β to systematically derive *single* constraints between the populations in these pairs.

To illustrate the idea of ambivalence consider a coarse-graining β which partitions a genotype set into three subsets. Fig 5 depicts the behavior of a two-parent variation operation that is ambivalent under β^3. Given some child that is produced by this variation operation, the probability that the child will belong to some theme class depends *only* on the theme classes of the parents and *not* on the specific parent genotypes. Hence the name 'ambivalent' – when the behavior of the variation operaion is viewed from the coarse-grained level of the theme classes, the variation operation 'does not care' about the specific genotypes of the parents or the child.

DEFINITION 8 (AMBIVALENCE). *A transmission function $T \in \Omega_m^X$ is ambivalent under some coarse-graining $\beta : X \to K$ if for all $k, k_1, \ldots, k_m \in K$ and for all*

Figure 1: Let $\beta : G \to K$ be a coarse-graining which partitions the genotype set G into three theme classes. This figure depicts the behavior of a two-parent variation operator that is ambivalent under β. The small dots denote specific genotypes and the solid arrows denote the recombination of these genotypes. A dashed arrow denotes that a child from a recombination may be produced 'somewhere' within the theme class that it points to, and the label of a dashed arrow denotes the probability with which this might occur. As the diagram shows the probability that the child of a variation operation will belong to a particular theme class depends *only* on the theme classes of the parents and *not* on their specific genotypes

$x_1, y_1 \in \langle k_1 \rangle, \ldots, x_m, y_m \in \langle k_m \rangle$,

$$\sum_{x \in \langle k \rangle_\beta} T(x|x_1, \ldots, x_m) = \sum_{y \in \langle k \rangle_\beta} T(y|y_1, \ldots, y_m)$$

Observe that the concept of ambivalence can also be defined in terms of the projection operator as follows: $T \in \Omega_m^X$ is ambivalent under $\beta : X \to K$ if for all $k_1, \ldots, k_m \in K$ and for all $x_1, y_1 \in \langle k_1 \rangle_\beta, \ldots, x_m, y_m \in \langle k_m \rangle_\beta$, $\Xi_\beta(T(\cdot | x_1, \ldots, x_m)) = \Xi_\beta(T(\cdot | y_1, \ldots, y_m))$.[4]

As the following definition shows, given a transmission function with domain X that is ambivalent under some coarse-graining $\beta : X \to K$, there is a 'natural' way to define a transmission function over the theme set K by 'projecting' the transmission function through β onto the theme set K.

DEFINITION 9. (PROJECTION OF AN AMBIVALENT TRANSMISSION FUNCTION) *Let $T \in \Omega_m^X$ be a transmission function that is ambivalent under some coarse-graining $\beta : X \to K$. We define the β-projection of T, denoted $T^{\overrightarrow{\beta}} \in \Omega_m^K$ to be a transmission function in Ω_m^K such that for all $k, k_1, \ldots, k_m \in K$ and for any*

---

[3]Formally the notion of ambivalence is defined for transmission functions, however we will also use it to qualify the variation operations of an EA; a variation operation is ambivalent under some coarse-graining if it can be modeled by a transmission function that is ambivalent under that coarse-graining.
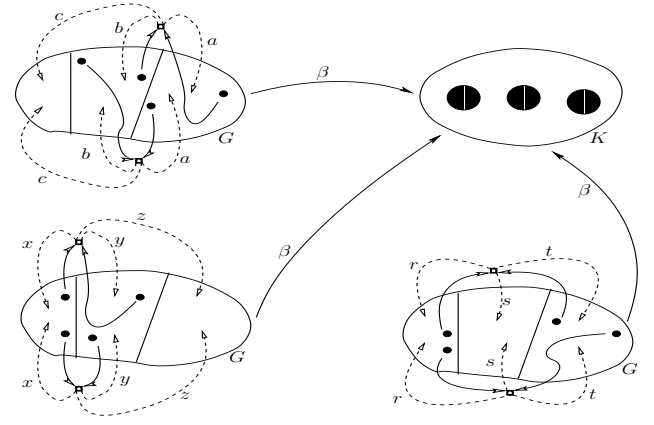
[4]Ambivalence is related to the concept of compatibility defined in chapter 17 of [11]. Theorem 17.5 in [11] and Theorem 2 of this paper can be used to show that if some transmission function T is ambivalent under some coarse-graining $\beta : G \to K$ then $\mathcal{V}_T$ is compatible with the equivalence class relation over G that is induced by β.

$x_1 \in \langle k_1 \rangle_\beta, \ldots, x_m \in \langle k_m \rangle_\beta,$

$$T^{\overrightarrow{\beta}}(k|k_1, \ldots, k_m) = \sum_{x \in \langle k \rangle_\beta} T(x|x_1, \ldots, x_m)$$

As T is ambivalent, $T^{\overrightarrow{\beta}}$ is well-defined.

Once again the projection of an ambivalent transmission function can be defined in terms of the projection operator: given that $T \in \Omega_m^X$ is ambivalent under some $\beta : X \to K$, for all $k, k_1, \ldots, k_m \in K$, and any $x_1 \in \langle k_1 \rangle_\beta, \ldots, x_m \in \langle k_m \rangle_\beta$, $T^{\overrightarrow{\beta}}(k|k_1, \ldots k_m)$ is given by $(\Xi_\beta(T(\cdot|x_1, \ldots, x_m)))(k)$.

From the definitions given above, it is clear that for any transmission function T and any function β, T is ambivalent under β if and only if $T^{\overrightarrow{\beta}}$ is well defined.

An Example from Genetic Programming: Let β be a coarse-graining that maps each S-expression to it's shape, i.e. its tree structure. Consider a crossover operator (modeled by some transmission function T) that randomly picks a cut point in each of two parent S-expressions and creates a child by splicing the part of the first S-expression that lies above it's cut point onto the part of the second S-expression that lies below it's cut point. This operator is ambivalent under β because, given the shapes of two-parent S-expressions, the distribution over the shapes of child S-expressions is well defined, i.e. $T^{\overrightarrow{\beta}}$ is well defined.

For any transmission function T and any coarse-graining β, if T is ambivalent under β, then, for each pair of populations p_G, p'_G such that $p'_G = \mathcal{V}_T p_G$, the following theorem gives us a *single* constraint between p_G and p'_G.

THEOREM 2 (VARIATIONAL CONSTRAINTS THEOREM). *Let G be some set, let $T \in \Omega_m^G$ be a transmission function and let $\beta : G \to K$ be some function such that T is ambivalent under β. Then T is ambivalent under β if and only if for all distributions $p_G \in \Omega^G$, and all $k \in K$,*

$$(\Xi_\beta \circ \mathcal{V}_T p_G)(k) = (\mathcal{V}_{T^{\overrightarrow{\beta}}} \circ \Xi_\beta p_G)(k)$$

i.e. if and only if for any $p_G \in \Omega^G$, letting $p'_G = \mathcal{V}_T p_G$, $p_K = \Xi_\beta p_G$, and $p'_K = \Xi_\beta p'_G$, the following diagram commutes:

$$
\begin{array}{ccc}
p_G & \xrightarrow{\quad \mathcal{V}_T \quad} & p'_G \\
\Xi_\beta \searrow & & \nearrow \Xi_\beta \\
p_K & \dashrightarrow{\; \mathcal{V}_{T^{\overrightarrow{\beta}}} \;} & p'_K
\end{array}
$$

Connecting back with the example introduced above, the coarse-graining that maps each S-expression to it's shape gives us single constraints between all pairs of pre and post crossover populations. As mentioned earlier, a single constraint is of limited use in understanding compositional changes induced by some variation operation. What is needed is a systematic way of deriving *simultaneous* constraints between all pairs of pre and post variation populations. In Section 7 we show that this is possible when the genotypes are fixed length bitstrings and the variation operators are the common variation operators of a GA - 'canonical mutation', n-point crossover and uniform crossover. In preparation for this we derive some additional results about ambivalence in the following section .

6. THE ALGEBRA OF AMBIVALENCE

Given several transmission functions $T_1, \ldots T_n$ that are all ambivalent under some coarse-graining β we show two ways of combining $T_1, \ldots T_n$ to create a new transmission function that is also ambivalent under β. Also, given a single transmission function T and several coarse-grainings $\beta_1, \ldots, \beta_n$ such that T is ambivalent under each of these coarse-grainings, we show how, if a certain condition is met, the coarse-grainings can be combined to create a new coarse-graining such that T will be ambivalent under this new coarse-graining.

6.1 Combining Transmission functions

Given several ambivalent transmission functions over some set, the following two lemmas give us two ways to create new ambivalent transmission functions — firstly by taking a weighted sum, and secondly by composition.

LEMMA 1. , (THE WEIGHTED SUM OF AMBIVALENT TRANSMISSION FUNCTIONS IS AMBIVALENT) *For any set X, any function $\beta : X \to K$, and any $n \in \mathbb{N}^+$, let $T_1, \ldots, T_n \in \Omega_m^X$ be ambivalent transmission functions. Let $p \in \Omega^{\{1, \ldots, n\}}$ and let $T \in \Omega_m^X$ be defined as follows:*

$$T(x|x_1 \ldots, x_m) = \sum_{i=1}^{n} p(i) T_i(x|x_1, \ldots, x_m)$$

Then T is ambivalent under β with β-projection $T^{\overrightarrow{\beta}}$ given as follows:

$$T^{\overrightarrow{\beta}}(k|k_1, \ldots, k_m) = \sum_{i=1}^{n} p(i) T_i^{\overrightarrow{\beta}}(k|k_1, \ldots, k_m)$$

LEMMA 2. (THE COMPOSITION OF AMBIVALENT TRANSMISSION FUNCTIONS IS AMBIVALENT) *For any $T_1 \in \Omega_m^X, T_2 \in \Omega_n^X$, if T_1 and T_2 are both ambivalent under some coarse-graining $\beta : X \to K$. Then $T_1 \circ T_2$ is ambivalent under β with β-projection given as follows:*

$$(T_1 \circ T_2)^{\overrightarrow{\beta}}(k|j_1, \ldots, j_n) =$$
$$\sum_{\substack{(k_1, \ldots, k_m) \\ \in \Pi_1^m K}} T_1^{\overrightarrow{\beta}}(k|k_1, \ldots, k_m) \prod_{i=1}^{m} T_2^{\overrightarrow{\beta}}(k_i|j_1, \ldots, j_n)$$

6.2 Combining Coarse-Grainings

Given several coarse-grainings with the same domain, the following definition allows us to create a new coarse-graining which induces a finer partition over the domain.

DEFINITION 10. (CARTESIAN PRODUCT OF COARSE-GRAININGS) *For any $n \in \mathbb{N}^+$ and any functions $\beta_1 : X \to K_1, \ldots, \beta_n : X \to K_n$, which share the same domain we define the cartesian product of $\beta_1, \ldots, \beta_n$ to be the function $\beta_1 \times \ldots \times \beta_n : X \to \prod_{i=1}^{n} K_i$ as follows:*

$$(\beta_1 \times \ldots \times \beta_n)(x) = (\beta_1(x), \ldots, \beta_n(x))$$

For notational convenience we will denote some cartesian product $\beta_1 \times \ldots \times \beta_n$ as $\prod_{i=1}^{n} \beta_i$.

Given two coarse-grainings with the same domain $\beta_1 : X \to K_1$ and $\beta_2 : X \to K_2$ and some transmission function $T \in \Omega_m^X$, we say that β_1 and β_2 are independent with respect to T if for any choice of m parents $x_1, x_2, \ldots, x_m,$

the mutual information between $\Xi_{\beta_1}(T(\cdot|x_1,\ldots,x_m))$ and $\Xi_{\beta_2}(T(\cdot|x_1,\ldots,x_m))$ is zero. In other words knowing something about the distribution $\Xi_{\beta_1}(T(\cdot|x_1,\ldots,x_m))$ gives no information about the distribution $\Xi_{\beta_2}(T(\cdot|x_1,\ldots,x_m))$. Formally, for all $k_1 \in K_1$, $k_2 \in K_2$

$$\Xi_{\beta_1 \times \beta_2}(T(\cdot|x_1,\ldots,x_m))(k_1,k_2) =$$
$$\Xi_{\beta_1}(T(\cdot|x_1,\ldots,x_m))(k_1)\Xi_{\beta_2}(T(\cdot|x_1,\ldots,x_m))(k_2)$$

The following definition extends this idea of independence to multiple coarse-grainings. Unfortunately it defines independence in a form that is more suited for use in proofs than for understandability. The proposition that follows shows that this form is equivalent to the more intuitive form that we used above.

DEFINITION 11. *For any set X and any functions $\beta_1 : X \to K_1,\ldots$, $\beta_n : X \to K_n$, and any $T \in \Omega_m^X$, we say that $\beta_1,\ldots,\beta_n$ are independent with respect to T if for all $x_1,\ldots,x_m \in X$ and all $k_1 \in K_1$, $\ldots$, $k_n \in K_n$,*

$$\sum_{\substack{x \in \\ \langle k_1 \rangle_{\beta_1} \cap \ldots \cap \langle k_n \rangle_{\beta_n}}} T(x|x_1,\ldots,x_m) =$$
$$\left(\sum_{x \in \langle k_1 \rangle_{\beta_1}} T(x|x_1,\ldots,x_m) \right) \ldots \left(\sum_{x \in \langle k_n \rangle_{\beta_n}} T(x|x_1,\ldots,x_m) \right)$$

PROPOSITION 1. *For any set X and any functions $\beta_1 : X \to K_1,\ldots$, $\beta_n : X \to K_n$, and any $T \in \Omega_m^X$, $\beta_1,\ldots,\beta_n$ are independent with respect to T if and only if for all $x_1,\ldots,x_m \in X$ and all $k_1 \in K_1$, $\ldots$, $k_n \in K_n$,*

$$\Xi_{\prod_{i=1}^n \beta_i}(T(\cdot|x_1,\ldots,x_m))((k_1,\ldots,k_n)) =$$
$$\Xi_{\beta_1}(T(\cdot|x_1,\ldots,x_m))(k_1) \ldots \Xi_{\beta_n}(T(\cdot|x_1,\ldots,x_m))(k_n)$$

The next lemma shows that given some transmission function T and some coarse-grainings $\beta_1,\ldots,\beta_n$ that are independent with respect to T, if T is ambivalent under each of the coarse-grainings, then T will also be ambivalent under the cartesian product of the coarse-grainings.

LEMMA 3. *For any $n \in \mathbb{N}^+$ let $\beta_1 : X \to K_1,\ldots$, $\beta_n : X \to K_n$ be functions which share the same domain. For any $m \in \mathbb{N}^+$ let $T \in \Omega_m^X$ be a transmission function such that $\beta_1,\ldots,\beta_n$ are independent with respect to T and such that for all $i \in \{1,\ldots,n\}$, T is ambivalent under β_i. Then T is ambivalent under $\prod_{i=1}^n \beta_i$ with $\prod_{i=1}^n \beta_i$-projection $T^{\overrightarrow{\prod_{i=1}^n \beta_i}}$ given as follows,*

$$T^{\overrightarrow{\prod_{i=1}^n \beta_i}}((k_1,\ldots k_n)|(k_1^1,\ldots,k_m^1),\ldots,(k_1^n,\ldots,k_m^n)) =$$
$$T^{\overrightarrow{\beta_1}}(k_1|k_1^1,\ldots,k_m^1) \ldots T^{\overrightarrow{\beta_n}}(k_n|k_1^n,\ldots,k_m^n)$$

7. SIMULTANEOUS VARIATIONAL CONSTRAINTS IN GENETIC ALGORITHMS

Recall that a schema partition over some set of fixed length bitstrings partitions the set into schemata which all have 'wildcards' (denoted by $*$) at the same locii. For example, if the genotype set is the set of all bitstrings of length 8, then the schema partition $***\#\#***$ partitions the genotype set into the schemata $***00***$, $***01***$, $***10***$, and $***11***$. We call any coarse-graining that induces a

schema partition over some a set of fixed-length bitstrings a *schema partitioning*.

The cartesian product structure of the genotype set and the nature of the variation operations commonly used in GAs makes it possible to systematically derive simultaneous variational constraints for all pairs of pre and post variation populations. We show, that a transmission function that models any combination of common GA variation operations — 'canonical' mutation, n-point crossover, and uniform crossover — is ambivalent under any schema partitioning.

Let $\mathfrak{B}_\ell$ be the set of all bitstrings of length ℓ. For any $x \in \mathfrak{B}_\ell$, and any $i \in \{1,\ldots,\ell\}$, let x_i denote the ith locus of x.

Let $n \in \mathbb{N}^+$. For any $i \in \mathbb{N}^+$ such that $i \leq n$ let $\beta_i : \mathfrak{B}_n \to \{0,1\}$ be a coarse-graining such that $\beta_i(x) = x_i$, i.e. β_i maps a bitstring to the value of its ith locus.

Let $\mathbb{I}$ be the power set of $\{1,\ldots,n\}$, let $I \in \mathbb{I}$ be some index set $\{a_1,\ldots,a_{|I|}\}$ where $a_1 < \ldots < a_{|I|}$ and let $\beta_I : \mathfrak{B}_n \to \mathfrak{B}_{|I|}$ be the coarse-graining $\beta_{a_1} \times \ldots \times \beta_{a_{|I|}}$. For each $I \in \mathbb{I}$, β_I is a schema partitioning. To see this formally, for any index set $I \in \mathbb{I}$ observe that the schema partition induced by β_I is given by the concatenation $\xi_I(1)\ldots\xi_I(n)$, where for any $j \in \{1,\ldots,n\}$, $\xi_I(j)$ is given by

$$\xi_I(j) = \begin{cases} \# & \text{if } j \in I \\ * & \text{otherwise} \end{cases}$$

7.1 Ambivalence of Canonical Mutation

Let us call a mutation operation that flips each bit independently with some fixed probability α *canonical mutation* and let $M \in \Lambda_1^{\mathfrak{B}_n}$ be a transmission function that models this operation. Observe that for any $i \in \{1,\ldots,n\}$, M is ambivalent under β_i with $M^{\overrightarrow{\beta_i}} \in \Lambda_1^{\{0,1\}}$ given as follows:

$$M^{\overrightarrow{\beta_i}}(k|l) = \begin{cases} \alpha & \text{if } k \neq l \\ 1-\alpha & \text{otherwise} \end{cases}$$

Observe that for any $I \in \mathbb{I}$, such that $I = \{a_1,\ldots,a_{|I|}\}$ and $a_1 < \ldots < a_{|I|}$, $\beta_{a_1},\ldots,\beta_{a_{|I|}}$ are independent with respect to M as M models a mutation operator that flips bits *independently* of each other. Hence, by lemma 3, M is ambivalent under β_I with $M^{\overrightarrow{\beta_I}} \in \Lambda_1^{\mathfrak{B}_{|I|}}$ given by

$$M^{\overrightarrow{\beta_I}}(k|l) = M^{\overrightarrow{\beta_{a_1}}}(k_1|l_1) \ldots M^{\overrightarrow{\beta_{a_{|I|}}}}(k_{|I|}|l_{|I|}).1$$

Multiplication by 1 at the end of the right hand side of this expression is necessary to account for the case when I is the empty set.

7.2 Ambivalence of Uniform and n-point Crossover

Let Ψ be the set of all masks for length n bitstrings, i.e. Ψ is itself a set of bitstrings in $\mathfrak{B}_n$. For any mask $\psi \in \Psi$ let the 2-parent transmission function $T_\psi \in \Lambda_2^{\mathfrak{B}_n}$ be defined as follows:

$$T_\psi(x|y,z) = \begin{cases} 1 & \text{if } \forall i \in \{1,\ldots,n,\}, \\ & \quad (x_i = y_i \wedge \psi_i = 0) \vee \\ & \quad (x_i = z_i \wedge \psi_i = 1) \\ 0 & \text{otherwise} \end{cases}$$

Note that for any parents y,z $T_\psi(\cdot|y,z)$ is a discrete

delta function which concentrates all its distribution mass on one child. In other words, T_ψ is 'deterministic'.

For all ψ and for all $i \in \{1, \ldots, n\}$, see that T_ψ is ambivalent under β_i with $T_\psi^{\overrightarrow{\beta_i}} \in \Lambda_2^{\{0,1\}}$ given as follows:

$$T_\psi^{\overrightarrow{\beta_i}}(k \,|\, l, m) = \begin{cases} 1 & \text{if } (k = l \wedge \psi_i = 0) \ \vee \\ & \qquad (k = m \wedge \psi_i = 1) \\ 0 & \text{otherwise} \end{cases}$$

Observe that for any $I \in \mathbb{I}$, such that $I = \{a_1, \ldots, a_{|I|}\}$ and $a_1 < \ldots < a_{|I|}$, $\beta_{a_1}, \ldots, \beta_{a_{|I|}}$ are independent with respect to T_ψ because for any two parents y, z the mutual information between the distributions $\Xi_{\beta_{a_1}}(T_\psi(\cdot|y, z)), \ldots, \Xi_{\beta_{a_{|I|}}}(T_\psi(\cdot|y, z))$ is zero. (This is because the distribution $\Xi_{\beta_i}(T_\psi(\cdot|y, z))$ depends only on the values y_i, z_i, and ψ_i). Hence, by lemma 3, T_ψ is ambivalent under β_I with $T_\psi^{\overrightarrow{\beta_I}}$ given as follows:

$$T_\psi^{\overrightarrow{\beta_I}}(k \,|\, l, m) = T_\psi^{\overrightarrow{\beta_{a_1}}}(k_1 \,|\, l_1, m_1) \ldots T_\psi^{\overrightarrow{\beta_{a_{|I|}}}}(k_{|I|} \,|\, l_{|I|}, m_{|I|}).1$$

Multiplication by 1 at the end of the right hand side of this expression is once again necessary to account for the case when I is the empty set.

For some choice of distribution over the set of all masks $q \in \Lambda^\Psi$, let $T \in \Lambda_2^{\mathfrak{B}^n}$ be given as follows:

$$T(x|y, z) = \sum_{\psi \in \Psi} q(\psi) T_\psi(x|y, z)$$

As stated above, for all $\psi \in \Psi$, T_ψ is ambivalent under β_I. Therefore by lemma 1, T is ambivalent under β_I with $T^{\overrightarrow{\beta_I}} \in \Lambda_2^{\mathfrak{B}^{|I|}}$ given by

$$T^{\overrightarrow{\beta_I}}(k \,|\, l, m) = \sum_{\psi \in \Psi} q(\psi) T_\psi^{\overrightarrow{\beta_I}}(k \,|\, l, m)$$

By appropriately choosing q, T can be made to model any n-point or uniform crossover operation [13], so any n-point or uniform crossover operation is ambivalent under β_I.

Finally, by lemma 2 any composition of n-point or uniform crossover operations with the canonical mutation operation, in any order, is ambivalent under β_I.

Given this result one can use schema partitionings and the variational constraints theorem to systematically derive simultaneous constraints for all pairs of pre and post variation populations for any GA with common variation operations.

8. CONCLUSION

In order to understand how EAs adapt phenotypes it is important to study the *immediate* effect of evolutionary operations on *populations* of genotypes. This paper has made the following contributions towards such a study.

1) We described how the technique of form-invariant commutation allows us to understand the immediate effect of evolutionary operations on genotypic populations in terms of *inter-population constraints*. We argued that deep insight into this population-level effect is possible when one can systematically derive *simultaneous* constraints over all pairs of pre and post operative populations.

2) In section 4 we showed that the fitness proportional selection operator is quite remarkable because *any* coarse-graining induces single constraints over all pairs of pre and post selection populations. Moreover we succeeded in deriving these constraints in a form that is easy to understand. The results obtained are clearly important to an understanding of how EAs adapt phenotypes.

3) We developed an abstract condition called *ambivalence* and showed that when a transmission function (that models some variation operation) is ambivalent under some coarse-graining then it is possible to systematically derive *single* constraints over all pairs of pre and post variation populations.

4) Finally we showed that when the evolutionary system is a GA that uses some combination of 'common' variation operations, it is possible to use schema partitionings to derive *simultaneous* constraints over all pairs of pre and post variation populations. In future work we will examine how knowledge about these constraints contributes to an understanding of adaptation in genetic algorithms.

9. REFERENCES

[1] ALTENBERG, L. The evolution of evolvability in genetic programming. In *Advances in Genetic Programming*, K. E. Kinnear, Jr., Ed. MIT Press, 1994.

[2] CONTRERAS, A. A., ROWE, J. E., AND STEPHENS, C. R. Coarse-graining in genetic algorithms: Some issues and examples. In *GECCO* (2003), pp. 874–885.

[3] GOLDBERG, D. E. *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison-Wesley, Reading, MA, 1989.

[4] HOLLAND, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan, 1975.

[5] HOLLAND, J. H. Building blocks, cohort genetic algorithms, and hyperplane-defined functions. *Evolutionary Computation 8*, 4 (2000), 373–391.

[6] HOLLAND, J. H. A derived markov process for modeling reaction networks. *Evolutionary Computation 11*, 4 (2003), 339–362.

[7] MITCHELL, M. *An Introduction to Genetic Algorithms*. The MIT Press, Cambridge, MA, 1996.

[8] STEPHENS, C., AND WAELBROECK, H. Schemata evolution and building blocks. *Evolutionary Computation 7*, 2 (1999), 109–124.

[9] TOUSSAINT, M. *The Evolution of Genetic Representations and Modular Neural Adaptation*. PhD thesis, Institut fr Neuroinformatik, Ruhr-Universiät-Bochum, Germany, 2003.

[10] VOSE, M., AND WRIGHT, A. Form invariance and implicit parallelism, 2001.

[11] VOSE, M. D. *The simple genetic algorithm: foundations and theory*. MIT Press, 1999.

[12] VOSE, M. D. Infinite population ga tutorial. Tech. rep., The University of Tennessee, Knoxville, 2004.

[13] WRIGHT, A. H., VOSE, M. D., AND ROWE, J. E. Implicit parallelism. In *GECCO* (2003), pp. 1505–1517.