Coarse Graining Selection and Mutation

Jonathan E. Rowe[1], Michael D. Vose[2], and Alden H. Wright[3]

[1] School of Computer Science, University of Birmingham
J.E.Rowe@cs.bham.ac.uk
[2] Computer Science Department, University of Tennessee
vose@cs.utk.edu
[3] Dept. of Computer Science, University of Montana
alden.wright@umontana.edu

Abstract. Coarse graining is defined in terms of a commutative diagram. Necessary and sufficient conditions are given in the continuously differentiable case. The theory is applied to linear coarse grainings arising from partitioning the population space of a simple Genetic Algorithm (GA). Cases considered include proportional selection, binary tournament selection, and mutation. A nonlinear coarse graining for ranking selection is also presented. Within the context of GAs, the primary contribution made is the introduction and illustration of a technique by which the possibility for coarse grainings may be analyzed. A secondary contribution is that a number of new coarse graining results are obtained.

1 Introduction

Managing complexity involves quotients (or some generalization thereof) if by "managing complexity" one intends to reduce complexity while simultaneously maintaining important aspects of fidelity. The following diagram is an abstraction of the general scheme being considered. In that illustration, $x \in X$ represents state and $h : X \to X$ transforms state. Complexity is managed by Ξ, which maps state into a simpler form, and by $\tilde{h}$ which has reduced complexity by virtue of transforming simplified state.

$$
\begin{array}{ccc}
x & \xrightarrow{\;h\;} & h(x) \\
\Xi \downarrow & & \downarrow \Xi \\
\Xi x & \xrightarrow{\;\tilde{h}\;} & \Xi h(x)
\end{array}
$$

Maintaining important aspects of fidelity is interpreted to mean the diagram commutes; both paths from x to $\Xi h(x)$ yield identical results. Thus Ξ can be regarded as defining what aspects of fidelity *are maintained* – if leeway exists in choosing it – or what aspects of fidelity are *capable of preservation* – if there is virtually no leeway. The reduced complexity model $\tilde{h}$ is the *quotient* of h corresponding to the *coarse graining* Ξ.

Whereas modeling h in an approximate fashion (by relaxing commutativity of the diagram) is interesting, the central question this paper is concerned with

is whether one can do better than approximation, and if so, then how? Moreover, knowledge of what it is that can be exact may identify a useful starting point for what it is that later will be approximated or perturbed from.

This abstract framework may provide a useful context in which to consider systems comprised of large collections of components interacting with each other (and with possibly some background environment). Assuming practical limitations to exact computation of the dynamics $x, h(x), h \circ h(x), \ldots$, approximation may be the best one can do. One would like to know, however, if that was the case or whether useful quotients did exist. It is natural to ask whether the underlying components could somehow be partitioned into a collection of disjoint subsets which could be considered as units in their own right. If obtaining a description of the dynamics of the subsets – *in terms of the subsets alone* – is possible, then the original system might be coarse grained into higher level units (the subsets) having dynamics compatible with the dynamics of the original system.

This scenario will be made concrete by taking the system to be a Genetic Algorithm (GA). In that case the underlying components comprise the search space, the environment is modeled by the fitness function (which determines competition between population members), and the state space is the set of possible populations. The primary contribution made by this paper is to introduce and illustrate a technique by which the possibility for coarse grainings may be analyzed. We are concerned with analytical tools rather than establishing particular results about any specific fitness function. The potential utility of those tools is demonstrated by obtaining a number of new coarse graining results.

This paper is organized as follows. First, some conceptual examples of quotients are discussed. Second, a necessary and sufficient condition characterizing quotients is described (assuming h is continuously differentiable, X is an open subset of a finite dimensional Euclidean space, . . .), followed by a reduction to special cases. Third, aspects of the theory of the Simple Genetic Algorithm (Vose, 1999) are reviewed, followed by an application of the necessary and sufficient condition (characterizing quotients) to investigate coarse grainings of selection and mutation. The paper concludes with a summary of results.

2   Conceptual Overview

A few real world examples are briefly mentioned to make the framework introduced above less abstract and to illustrate that in practice complex systems are frequently managed and understood with the aid of coarse grainings. It should be kept in mind that we must necessarily coarse grain some *model* of the real world, because the state space X and the transformation h are mathematical abstractions.

1. Modeling the motion of a body by assuming it is rigid leads to the typical coarse-graining (of that rigid model) where $\Xi(x)$ is the center of gravity.

   Examples of this sort employ coarse grainings to *transfer* the domain of analysis to a simplified setting (namely, $\tilde{h}$ acting on ΞX).

2. Conservation laws assert that the dynamics h (of some model of a physical system) is compatible with a coarse graining under which the quotient $\tilde{h}$ is the identity map. For instance, $E = mc^2$ corresponds to the coarse graining $\Xi(x) = E(x) - m(x)c^2$.

   Examples of this sort show the existence of coarse grainings may be used to *constrain* the analysis (in the original setting X) by invariants.

3. The quantum mechanics describing the hardware of a computer is usually modeled by digital logic. A familiar coarse-graining (of that gate-level digital model) is the high level gnu/linux interface seen by the C programmer.

   Examples of this sort suggest that the quotient $\tilde{h}$ may be the primary object of concern; commutativity of the coarse graining ($\Xi \circ h = \tilde{h} \circ \Xi$) may serve as a proof of correctness for the implementation h.

The quotient in the last example above is obtained only if the state transition $x \mapsto h(x)$ corresponds to a number of micro cycles which depends on x (namely, that number required for completion of the high level service/command corresponding to x). This point is made to clarify the general phenomenon that even though a desirable quotient of a system's trajectory

$$x \mapsto h(x) \mapsto h^2(x) \mapsto \cdots \tag{1}$$

might not exist (think of h as being analogous to a single micro cycle), it nevertheless could be the case that the trajectory

$$x \mapsto h^{p(x)}(x) \mapsto h^{p(h^{p(x)}(x))}(x) \mapsto \cdots \tag{2}$$

does admit useful quotients. The applications to genetic algorithms presented in sections 4 through 6, however, are limited to scenario (1) – where h corresponds to a single generation – rather than the more general situation (2).

Because models are coarse grained, an exact coarse graining (of a model) can be an *approximation* (to reality) if the model itself is an approximate one. This points to another reason why quotients are significant; they may aid in identifying tractable approximate models (i.e., models which have useful quotients). The applications to genetic algorithms presented in sections 4 through 7, however, are not concerned with approximation since the models being coarse grained are themselves exact.

3   Differentiable Coarse Graining

The following expands upon the account given at Dagstuhl (Vose, 2004). Rather than immediately beginning with Ξ, a coarse graining is instead obtained as a byproduct of a continuously differentiable map Ψ. A reason for this is to constrain the context of the general framework for coarse graining to a more specific setting wherein differential calculus may be brought to bear (most coarse graings appearing in the Evolutionary Computation literature correspond to

equivalence relations obtainable as a byproduct of linear – and thus trivially differentiable – maps). The hope is that this may facilitate the *computation* of coarse grainings in some circumstances. That possibility is in fact achieved and is demonstrated in subsequent sections.

Let $\Psi : V \longrightarrow W$ be a continuously differentiable function between open subsets of finite dimensional Euclidean spaces. A *path* with respect to Ψ (called simply a path, when Ψ is understood) is a smooth function[1] $\rho : [0,1] \longrightarrow V$ such that $\Psi \circ \rho$ is constant. The path ρ is said to be *from u to v* provided $\rho(0) = u$ and $\rho(1) = v$. Let the equivalence relation $\equiv$ on V be defined by

$$u \equiv v \iff \text{there exists a path } \rho \text{ from } u \text{ to } v$$

and let $\Xi : V \longrightarrow V/\equiv$ map element v to its equivalence class $\tilde{v}$. Equivalence classes are path connected components of level sets of Ψ. It follows that the image of any path is contained in some equivalence class.

A continuously differentiable function $h : V \longrightarrow V$ is said to be *compatible with* $\equiv$ provided there exists a function $\tilde{h}$ for which the following diagram commutes,

$$
\begin{array}{ccc}
V & \xrightarrow{\ h\ } & V \\
{\scriptstyle \Xi}\downarrow & & \downarrow{\scriptstyle \Xi} \\
V/\equiv & \xrightarrow{\ \tilde{h}\ } & V/\equiv
\end{array}
$$

In that case $\tilde{h}$ is called the quotient of h (with respect to Ξ),

$$\Xi(u) = \Xi(v) \implies \Xi \circ h(u) = \Xi \circ h(v) \tag{3}$$

and $\tilde{h}$ is well defined by $\tilde{h} \circ \Xi(u) = \Xi \circ h(u)$.

By restricting h to the equivalence class $V_w = \Xi^{-1}(\tilde{w})$, compatibility of h with $\equiv$ implies that for each $w \in V$,

$$V_w \xrightarrow{\ h\ } h(V_w) \xrightarrow{\ \Psi\ } \{\Psi(h(w))\}$$

since $\Xi(u) = \Xi(v) = \tilde{w}$ for all $u, v \in V_w$, and so by (3)

$$\Xi \circ h(u) = \Xi \circ h(v) \implies h(u) \equiv h(v) \implies \Psi \circ h(u) = \Psi \circ h(v)$$

Thus on V_w, the composition of Ψ with h is the constant function

$$\widehat{h} \ = \ \Psi \circ h \tag{4}$$

Let T_w be the *tangent space* of the equivalence class V_w at w defined by

$$T_w = \mathcal{L}\{ d\rho_0(1) \ : \ \rho \text{ is a path from } w \text{ to } v, \text{ for some } v\}$$

---

[1] We call a continuous function *smooth* if it's differential (over the interior of the domain) has a continuous extension to the entire domain

where $\mathcal{L}\{\cdots\}$ denotes the linear span of $\{\cdots\}$ and, for any function f differentiable at x, the differential df_x of f at x is the linear function of z

$$df_x(z) \; = \; \lim_{t \to 0} \frac{f(x+tz) - f(x)}{t}$$

Note that $d\rho_0(1)$ occurring in the definition of T_w is the tangent at w of the path ρ from w to v,

$$d\rho_0(1) \; = \; \lim_{t \downarrow 0} \frac{\rho(t) - w}{t}$$

Assuming compatibility (of h with $\equiv$), $\widehat{h}$ is constant (on equivalence classes) and so by equation (4)

$$\Psi \circ h \circ \rho$$

is also constant, for every path ρ (since the image of a path is comprised of equivalent elements). Moreover, it follows from the chain rule that

$$d\Psi_{h \circ \rho(t)} dh_{\rho(t)} d\rho_t(1) \; = \; 0 \tag{5}$$

for $0 < t < 1$.

Therefore, choosing $x = \rho(0)$ and $t \downarrow 0$ in (5),

$$dh_x : T_x \longrightarrow K_{d\Psi_{h(x)}} \tag{6}$$

where, for any linear function L, the kernel of L is denoted by K_L.

Condition (6) is therefore *necessary* (for all x) in order that h be compatible with $\equiv$. It will also be shown sufficient. Note first, however, that because $\Psi \circ \rho$ is constant (for any path ρ), the chain rule yields

$$d\Psi_{\rho(t)} d\rho_t(1) = 0$$

Hence T_x is a subspace of $K_{d\Psi_x}$ [2]. A condition which implies (6) is therefore

$$dh_x : K_{d\Psi_x} \longrightarrow K_{d\Psi_{h(x)}}$$

Conversely, suppose (6) holds for all $x \in V$. Since the compatibility of h with $\equiv$ is implied by condition (3), suppose $u, v \in V_w$ and let ρ be a path from u to v (i.e., assume $\Xi(u) = \Xi(v)$). Define the function f by

$$f(t) = \Psi \circ h \circ \rho(t)$$

and note that

$$\Psi \circ h \circ \rho(y) - \Psi \circ h(u) \; = \; \int_{0^+}^{y^-} df_t(1)\, dt \; = \; \int_{0^+}^{y^-} d\Psi_{h \circ \rho(t)} dh_{\rho(t)} d\rho_t(1)\, dt$$

---

[2] They need not coincide; at an extrema or saddle point (for instance) there may exist dimensions orthogonal to the level set along which $d\Psi_x$ vanishes

If (5) held (for $0 < t < 1$), then the integral would be zero, establishing that $h \circ \rho$ is a path from $h(u)$ to $h(v)$, hence $\Xi \circ h(u) = \Xi \circ h(v)$, and thus condition (3) would hold, implying compatibility of h with $\equiv$. Choosing $x = \rho(t)$ shows that condition (5) is a consequence of (6), provided $d\rho_t(1) \in T_x$. Let ϱ be the path from x to v defined by

$$\varrho(s) \;=\; \rho(t(1-s)+s)$$

Note that $d\varrho_0(1) \in T_x$ (by definition of tangent space), and

$$
\begin{aligned}
d\rho_t(1-t) &= \lim_{s\downarrow 0} \frac{\rho(t+(1-t)s) - \rho(t)}{s} \\
&= \lim_{s\downarrow 0} \frac{\rho(t(1-s)+s) - x}{s} \\
&= \lim_{s\downarrow 0} \frac{\varrho(s) - x}{s} \\
&= d\varrho_0(1)
\end{aligned}
$$

Hence $d\rho_t(1) = d\varrho_0(1)/(1-t) \in T_x$. The following theorem has been established.

Theorem 1. *A necessary and sufficient condition for h to be compatible with $\equiv$ is that for all $x \in V$,*

$$dh_x : T_x \longrightarrow K_{d\Psi_{h(x)}}$$

Moreover, T_x is a subspace of $K_{d\Psi_x}$.

In the special case where Ψ is linear, the necessary and sufficient condition reduces to

$$dh_x : K_\Psi \longrightarrow K_\Psi$$

since a linear function is its own differential, and for $x \in K_\Psi$ the path $\rho(t) = x(1-t) + 2tx$ shows $x \in T_x$ (the tangent to ρ is x), hence $K_\Psi \subset T_x \subset K_\Psi$.

If both h and Ψ are linear, then the situation reduces to the case considered in Rowe, Vose, Wright (2004); the kernel of Ψ is an invariant subspace of h.

4   Proportional Selection + Mutation

A brief summary of relevant background (Vose, 1999) is given, followed by an application of theorem 1 to proportional selection + mutation.

Let τ denote the stochastic transition function for a finite population GA over the search space $\Omega = \{0, \dots, n-1\}$, and let $\mathcal{G}$ be the corresponding infinite population model. The transition matrix Q of the GA's Markov chain is defined by the probability that $\tau(p) = q$ and satisfies

$$Q_{p,q} = r! \prod \frac{(\mathcal{G}(p)_j)^{rq_j}}{(rq_j)!}$$

where r is the population size, and where the the population represented by the n dimensional vector p contains $r\,p_j$ instances of j.

Let $\equiv$ be an arbitrary equivalence relation over Ω, and let $\{0^*, \ldots, (k-1)^*\}$ be equivalence class representatives. The linear operator with $k \times n$ matrix Ξ defined by

$$\Xi_{i,j} \;=\; [i^* \equiv j]$$

(where $[expression]$ denotes 1 if $expression$ is true, and 0 otherwise) lifts $\equiv$ to populations by

$$p \equiv p' \iff \Xi p = \Xi p' \tag{7}$$

Compatibility in the stochastic case generalizes the definition given in the previous section; τ is said to be *compatible with* $\equiv$ iff

$$p \equiv p' \implies \forall q \,.\, \mathrm{Prob}\,\{\tau(p) \equiv q\} = \mathrm{Prob}\,\{\tau(p') \equiv q\}$$

In that case, $\tilde{\tau}$ defined by $\tilde{\tau}(\Xi x) = \Xi\tau(x)$ is referred to as the *quotient* of τ. It is known that $\tilde{\tau}$ exists if and only if a corresponding quotient $\tilde{\mathcal{G}}$ of $\mathcal{G}$ exists. Moreover, the transition matrix for $\tilde{\tau}$ is obtained from the formula for $Q_{p,q}$ above by replacing $\mathcal{G}$ by $\tilde{\mathcal{G}}$, p by Ξp, and q by Ξq (Vose, 1999).

The "selection + mutation" case refers to the simple GA with (proportional) fitness and mutation, but no crossover. There the infinite population model takes the form

$$\mathcal{G}(p) \;=\; \frac{Gp}{\mathbf{1}^T G p}$$

where $G = MF$ is a $n \times n$ matrix and $\mathbf{1}$ is the vector of all 1 s. Here M is a column stochastic mutation matrix ($\mathbf{1}^T M = \mathbf{1}^T$) where $M_{i,j} = \mathrm{Prob}\,\{j \text{ mutates to } i\}$, and F is a diagonal fitness matrix where $F_{i,i} = f_i$ is the fitness of i (the vector f is referred to as the fitness function). In particular, $\mathbf{1}^T G p = f^T p$. The domain of immediate interest is

$$p \in \Lambda = \{\langle x_0, \ldots, x_{n-1} \rangle \,:\, x_i \geq 0, \mathbf{1}^T x = 1\}$$

since that is the completion of the population representation space. Note that $\mathbf{1}^T G p$ does not vanish on Λ provided fitness is positive. Positive fitness will be assumed throughout the remainder of this paper. The results of the previous section will be applied with $h = \mathcal{G}$ and V a neighborhood of Λ.

The situation is particularly simple, since choosing $\Psi = \Xi$ yields the equivalence relation above; it follows from (7) that $p \equiv p'$ iff p and p' are contained in a level set of Ψ (i.e., a translate of K_Ξ). Moreover, the coarse graining (as defined in sections 1 and 3) is also Ξ (which is a fortunate happenstance for notation), since Ψp can be regarded as representing the equivalence class $\tilde{p}$ of p. As noted after theorem 1, compatibility reduces to

$$d\mathcal{G}_x : K_\Xi \longrightarrow K_\Xi$$

The differential of $\mathcal{G}$ at x is

$$d\mathcal{G}_x \;=\; \frac{\mathbf{1}^T G x G - G x \mathbf{1}^T G}{(\mathbf{1}^T G x)^2}$$

Compatibility is therefore the condition that for all $x \in V$, and for all $v \in K_\Xi$,

$$\Xi \mathbf{1}^T G x G v = \Xi G x \mathbf{1}^T G v \tag{8}$$

Assume compatibility, and let $v \in K_\Xi$. Choosing $x = e_i$, where e_i is the ith column of the $n \times n$ identity matrix (indices begin with zero) and solving (8) for $\Xi G v$ yields

$$\Xi G v = \mathbf{1}^T G v \Xi \frac{G e_i}{\mathbf{1}^T G e_i} \tag{9}$$

Replacing e_i with e_j in the right hand side of (9) – the left hand side is invariant under that replacement – and equating the right hand side before replacement with after yields

$$\mathbf{1}^T G v \Xi \left(\frac{G e_i}{\mathbf{1}^T G e_i} - \frac{G e_j}{\mathbf{1}^T G e_j} \right) = 0 \tag{10}$$

If $\mathbf{1}^T G v = 0$, then it follows from (9) that $\Xi G v = 0$. Otherwise, it follows from (10) that

$$\frac{G e_i}{\mathbf{1}^T G e_i} \equiv \frac{G e_j}{\mathbf{1}^T G e_j}$$

Taking into account $G = MF$, these alternatives simplify to

$$f^T v = 0 \tag{11}$$

$$M e_i \equiv M e_j \tag{12}$$

First suppose there exist $y, z \in \Omega$ such that $y \equiv z$ and $f_y \neq f_z$. Note that $v = e_y - e_z \in K_\Xi$ and $f^T v \neq 0$. Since condition (11) does not hold, condition (12) must.

Next suppose no such y, z exist. Then all equivalent population members have identical fitness. Note that $v \in K_\Xi$ is equivalent to the condition that for all equivalence class representatives c^*,

$$\sum_{i \equiv c^*} v_i = 0 \tag{13}$$

(in particular, $v \in K_\Xi \implies \mathbf{1}^T v = 0$, i.e., $v \in \mathbf{1}^\perp$). Since fitness is constant over equivalence classes,

$$\sum_{i \equiv c^*} f_i v_i = 0$$

It follows that $F : K_\Xi \to K_\Xi$ (F is compatible with $\equiv$) and $\mathbf{1}^T G v = f^T v = 0$ for all $v \in K_\Xi$. As observed after (10),

$$\Xi G v = 0 \tag{14}$$

Moreover, since F is invertible (fitness is positive) quantification over $v \in K_\Xi$ is equivalent to quantification over $v \in \{F^{-1} w : w \in K_\Xi\}$ (an injective linear map on a finite dimensional space is surjective). Replacing v by $F^{-1} w$ in (14) yields

$$w \in K_\Xi \implies \Xi M w = 0$$

Hence M is compatible with $\equiv$.

Theorem 2. *Suppose positive fitness and zero crossover. Equivalent population members have identical fitness if and only if F is compatible with Ξ. When F is compatible with Ξ, a necessary and sufficient condition for τ to be compatible with $\equiv$ is that M is. If F is not compatible with Ξ, then a necessary and sufficient condition for τ to be compatible with $\equiv$ is that the columns of M are equivalent.*

Proof. It was shown above that F is compatible (with Ξ) when all equivalent population members have identical fitness. Conversely,

$$\Xi F v = 0 \Longrightarrow \mathbf{1}^T \Xi F v = \mathbf{1}^T F v = f^T v = 0$$

so F cannot be compatible if there exists $v \in K_\Xi$ such that $f^T v \neq 0$. But if $y \equiv z$ and $f_y \neq f_z$, then $v = e_y - e_z \in K_\Xi$ and $f^T v \neq 0$.

It was shown above that if all equivalent population members have identical fitness (i.e., F is compatible), then compatibility of $\mathcal{G}$ – which is equivalent to the compatibility of τ (Vose, 1999) – implies M is compatible. Conversely, if K_Ξ is invariant under both M and F, then it is invariant under $G = MF$, and $\mathbf{1}^T G v = \mathbf{1}^T \Xi G v = 0$ for all $v \in K_\Xi$. Therefore, the compatibility condition (8) reduces to the identity

$$\mathbf{1}^T G x \Xi G v \;=\; 0$$

It was shown above that if F is not compatible with Ξ (all equivalent population members do not have identical fitness), then compatibility of τ – which is equivalent to the compatibility of τ (Vose, 1999) – implies the columns of M are equivalent (12). Conversely, if ΞM has the form $c\mathbf{1}^T$ (columns of M are equivalent), then $\Xi G = c f^T$ and the compatibility condition (8) reduces to the identity

$$f^T x c f^T v \;=\; c f^T x f^T v \qquad\qquad \square$$

Theorem 2 is put into sharper focus by the following result (recall that e_i is the ith column of the $n \times n$ identity matrix, indices begin with zero).

Theorem 3. *A necessary and sufficient condition for M to be compatible with $\equiv$ is that for all i, j,*

$$i \equiv j \;\Longrightarrow\; M e_i \equiv M e_j$$

Proof. If $i \equiv j$, then $v = e_i - e_j \in K_\Xi$. Therefore if M is compatible, then

$$\Xi M (e_i - e_j) = 0$$

Hence $M e_i \equiv M e_j$. Conversely, if $i \equiv j \Longrightarrow M e_i \equiv M e_j$, then ΞM has the form

$$\sum_{h=0}^{k-1} C_h \sum_{l \equiv h^*} e_l^T$$

where $C_h = \Xi M e_i$ for $i \equiv h^*$ (the choice of i does not matter; $\Xi M e_i = \Xi M e_j$ when $i \equiv j$). If $v \in K_\Xi$, then

$$\Xi M v \;=\; \sum_{h=0}^{k-1} C_h \sum_{l \equiv h^*} e_l^T v \;=\; \sum_{h=0}^{k-1} C_h \sum_{l \equiv h^*} v_l \;=\; 0 \qquad\qquad \square$$

Theorem 3 provides a method by which a mutation operator can be constructed compatible with a given equivalence relation; whenever $i \equiv j$, choose columns i and j of M to differ by an element of K_Ξ. Moreover, since $K_\Xi \subset \mathbf{1}^\perp$, obtaining column i by adding an element $v \in K_\Xi$ to the jth column will not disturb the column stochasticity of M, provided $v + Me_j$ is non negative.

5 Binary Tournament Selection

A zero mutation, zero crossover, tournament selection GA with tournament size t and fitness function f has corresponding infinite population model (Vose, 1999)

$$\mathcal{G}(p)_i \;=\; t! \sum_{v \in X_n^t} \int_{\sum [f_j < f_i](v/t)_j}^{\sum [f_j \le f_i](v/t)_j} \varrho(y)\, dy \prod_{j<n} \frac{p_j^{v_j}}{v_j!}$$

where

$$X_n^t \;=\; \{\langle x_0, \ldots x_{n-1}\rangle \;:\; x_i \in \mathcal{Z}^{\ge 0},\; \mathbf{1}^T x = t\}$$

and ϱ is any continuous increasing probability density over $[0,1]$. *Binary tournament selection* refers to the result of choosing $t = 2$ and taking the limit as ϱ tends to point mass at 1. Assuming injective fitness (which will be assumed for the remainder of this paper), the result is

$$\mathcal{G}(p)_i \;=\; p_i^2 + 2p_i \sum_j p_j[f_j < f_i]$$

It follows that

$$(d\mathcal{G}_x v)_i \;=\; 2v_i x_i + 2 \sum_l [f_l < f_i](v_i x_l + x_i v_l) \tag{15}$$

Note that (15) is a symmetric expression in x and v, and therefore $d\mathcal{G}_x v = d\mathcal{G}_v x$ is linear in both x and v. In view of this, the compatibility condition is that for all $x \in V$, and for all $v \in K_\Xi$,

$$d\mathcal{G}_x v \;=\; \sum_h x_h d\mathcal{G}_{e_h} v \;\in\; K_\Xi$$

Since K_Ξ is a subspace, compatibility is therefore equivalent to the condition that for all h,

$$v \in K_\Xi \;\Longrightarrow\; d\mathcal{G}_{e_h} v \in K_\Xi$$

Moreover, the ith component of the differential above simplifies (from 15) to

$$(d\mathcal{G}_{e_h} v)_i \;=\; 2v_i[f_h < f_i] + 2[h = i] \sum_l [f_l \le f_i]v_l \tag{16}$$

This equality constrains what equivalence relations are possible. Consider the case where there exist nonequivalent elements a and b. Choosing $h = a$ in (16)

and applying condition (13) for membership in K_Ξ, compatibility requires the following implication

$$0 = \sum_{i \equiv b} v_i$$

$$\implies 0 = \sum_{i \equiv b} \left(v_i[f_a < f_i] + [a = i] \sum_l [f_l \le f_i]v_l \right) = \sum_{i \equiv b} v_i[f_a < f_i]$$

Therefore (by suitable choice of v), either every i equivalent to b must satisfy $[f_a < f_i]$ or else no i equivalent to b can satisfy $[f_a < f_i]$. In other words, equivalence classes are "fitness-contiguous" as defined below.

Let θ be a permutation of $\{0, \ldots, n-1\}$ such that $i < j \iff f_{\theta(i)} < f_{\theta(j)}$ and let $\equiv$ be any equivalence relation on Ω for which the equivalence classes are fitness-contiguous, meaning they are

$$\{\theta(0), \ldots, \theta(z_0)\}, \; \{\theta(z_0 + 1), \ldots, \theta(z_1)\}, \; \ldots, \; \{\theta(z_{k-2} + 1), \ldots, \theta(z_{k-1})\}$$

for some $0 \le z_0 < \cdots < z_{k-1} = n - 1$. Let the equivalence class representative of the cth class be $c^* = \theta(z_c)$. It follows that if $b < c$ then everything equivalent to b^* has fitness less than everything equivalent to c^*.

Lemma 1. *If the equivalence classes of $\equiv$ are fitness-contiguous and $v \in K_\Xi$, then for all l,*

$$\sum_{i \equiv j} [f_l \le f_i]v_i = [l \equiv j] \sum_i [f_l \le f_i]v_i$$

Proof. If $l \not\equiv j$, then f_l is either less than everything equivalent to j or else it is greater than everything equivalent to j. In the latter case, both sides above are zero. In the former case, both sides are also zero since then the left hand size vanishes due to (13). By what has been shown so far,

$$\sum_i [f_l \le f_i]v_i = \sum_{c=1}^{k-1} \sum_{i \equiv c^*} [f_l \le f_i]v_i = \sum_{c=1}^{k-1} [l \equiv c^*] \sum_{i \equiv c^*} [f_l \le f_i]v_i = \sum_{i \equiv l} [f_l \le f_i]v_i$$

which completes the proof for the remaining case $l \equiv j$. $\square$

Theorem 4. *Binary tournament selection is compatible with $\equiv$ if and only if the equivalence relation is fitness-contiguous.*

Proof. The "only if" part has already been established. Let $v \in K_\Xi$. By (16) and what has been established above, the "if" part follows from

$$\frac{1}{2}\sum_{i\equiv j}(d\mathcal{G}_{e_h}v)_i = \sum_{i\equiv j}\left(v_i[f_h < f_i] + [h = i]\sum_l [f_l \le f_i]v_l\right)$$

$$= \sum_{i\equiv j}v_i[f_h < f_i] + [h \equiv j]\sum_l [f_l \le f_h]v_l$$

$$= [h \equiv j]\Big(\sum_i v_i[f_h < f_i] + \sum_l [f_l \le f_h]v_l\Big)$$

$$= [h \equiv j]\sum_i v_i$$

$$= 0$$

$\square$

For "binary tournament selection + mutation" to be made compatible with a fitness-contiguous equivalence relation, mutation may be chosen as in theorem 3.

6   Ranking Selection

A zero mutation, zero crossover, ranking selection GA with parameter ϱ and fitness function f has corresponding infinite population model

$$\mathcal{G}(x)_i = \int_{\sum [f_j < f_i]x_j}^{\sum [f_j \le f_i]x_j} \varrho(y)\, dy$$

where ϱ is any continuous increasing probability density over $[0, 1]$ (Vose, 1999). Define η by

$$\eta_{\theta(0)} = 0$$

$$\eta_{\theta(i+1)} = \eta_{\theta(i)} + x_{\theta(i)}$$

(recall that $i < j \iff f_{\theta(i)} < f_{\theta(j)}$). It follows that

$$\mathcal{G}(x)_i = \varphi(x_i + \eta_i) - \varphi(\eta_i) \tag{17}$$

$$d\mathcal{G}_x v = \sum_i e_i \sum_k \left(\varrho(x_i + \eta_i)[f_k \le f_i] - \varrho(\eta_i)[f_k < f_i]\right)v_k \tag{18}$$

where φ is an anti-derivative of ϱ (Vose, 1999). Choosing $x = e_h$, the last expression above simplifies to yield

$$(d\mathcal{G}_{e_h}v)_i = \varrho([f_i \ge f_h])v_i + [i = h](\varrho(1) - \varrho(0))\sum_k [f_k < f_i]v_k$$

Compatibility requires that for all c^*, and all v satisfying (13),

$$0 = \sum_{i\equiv c^*}\left(\varrho([f_i \ge f_h])v_i + [i = h](\varrho(1) - \varrho(0))\sum_k [f_k < f_i]v_k\right)$$

$$= \sum_{i\equiv c^*}\varrho([f_i \ge f_h])v_i + [h \equiv c^*](\varrho(1) - \varrho(0))\sum_k [f_k < f_h]v_k \tag{19}$$

Assuming the equivalence relation is nontrivial, choose $h \not\equiv c^*$ to obtain

$$0 = \sum_{i \equiv c^*} \varrho([f_i \geq f_h])v_i = \varrho(0) \sum_{i \equiv c^*} [f_i < f_h]v_i + \varrho(1) \sum_{j \equiv c^*} [f_j \geq f_h]v_j \quad (20)$$

As seen in the previous section, this implies $\equiv$ must be fitness-contiguous (it follows from $\varrho(1) > \varrho(0)$ and choosing v to have exactly two nonzero components; either every j equivalent to c must satisfy $[f_j < f_h]$ or else every i equivalent to c must satisfy $[f_i \geq f_h]$ since otherwise $v_i = -v_j \neq 0$ contradicts 20).

Theorem 5. *Ranking selection is compatible with $\equiv$ if and only if the equivalence relation is fitness-contiguous.*

Proof. The "only if" part has already been established. Let $v \in K_{\Xi}$. Appealing to (18) and the fact that $\equiv$ is fitness-contiguous, the "if" part follows from

$$\sum_{i \equiv c^*} \sum_k \left(\varrho(x_i + \eta_i)[f_k \leq f_i] - \varrho(\eta_i)[f_k < f_i] \right) v_k$$

$$= \sum_{k \equiv i \equiv c^*} \left(\varrho(x_i + \eta_i)[f_k \leq f_i] - \varrho(\eta_i)[f_k < f_i] \right) v_k + \sum_{k \not\equiv c^*} v_k \sum_{i \equiv c^*} \varrho(x_i + \eta_i) - \varrho(\eta_i)$$

$$= \sum_{z_{c-1} < u,v \leq z_c} \left(\varrho(x_{\theta(v)} + \eta_{\theta(v)})[f_{\theta(u)} \leq f_{\theta(v)}] - \varrho(\eta_{\theta(v)})[f_{\theta(u)} < f_{\theta(v)}] \right) v_{\theta(u)}$$

$$= \sum_{z_{c-1} < u \leq z_c} v_{\theta(u)} \sum_{z_{c-1} < v \leq z_c} \varrho(\eta_{\theta(v+1)})[u \leq v] - \varrho(\eta_{\theta(v)})[u < v]$$

$$= \sum_{z_{c-1} < u \leq z_c} v_{\theta(u)} \left(\sum_{u \leq v \leq z_c} \varrho(\eta_{\theta(v+1)}) - \sum_{u < v \leq z_c} \varrho(\eta_{\theta(v)}) \right)$$

$$= \varrho(\eta_{\theta(z_c+1)}) \sum_{z_{c-1} < u \leq z_c} v_{\theta(u)}$$

$$= 0$$

□

7   Nonlinear Coarse Graining

Applications have so far involved linear coarse grainings corresponding to an equivalence relation over Ω. A nonlinear coarse graining is derived below for ranking selection. To simplify analysis, let $\varphi(x) = x^\gamma$ (where γ is a parameter), and let m and M denote the minimal fitness and maximal fitness elements of Ω, respectively. We seek a coarse graining where Ψ is real valued, independent of γ, and depends on x_m and x_M.

The derivation of Ψ is simplified by exploiting the invariant $1 = x_M + \eta_M$, so we choose to work with $\Psi(x) = \psi(x_m, \eta_M)$ for some function ψ. Let ψ_1 and ψ_2 denote the partial derivative of ψ with respect to its first and second argument, respectively. It follows that

$$\frac{\partial \Psi}{\partial x_j} = \begin{cases} \psi_1 + \psi_2 & \text{if } j = m \\ \psi_2 & \text{if } j \neq m \text{ and } j \neq M \\ 0 & \text{if } j = M \end{cases}$$

The condition $v \in K_{d\Psi_x}$ can therefore be expressed as

$$0 = \sum_j v_j \frac{\partial \Psi}{\partial x_j} = v_m \psi_1(x_m, \eta_M) + \psi_2(x_m, \eta_M) \sum_{j \neq M} v_j \tag{21}$$

Hence

$$\sum_{j \neq M} v_j = -v_m \psi_1(x_m, \eta_M)/\psi_2(x_m, \eta_M) \tag{22}$$

In view of (17), and using the form of (21) with $v \leftarrow d\mathcal{G}_x v$ and $x \leftarrow \mathcal{G}(x)$, the sufficient condition for compatibility (namely, $v \in K_{d\Psi_x} \implies d\mathcal{G}_x v \in K_{d\Psi_{\mathcal{G}(x)}}$) requires

$$0 = (d\mathcal{G}_x v)_m \psi_1(z_0, z_1) + \psi_2(z_0, z_1) \sum_{j \neq M} (d\mathcal{G}_x v)_j \tag{23}$$

where

$$z_0 = x_m^\gamma, \quad z_1 = \sum_{i \neq M} (x_i + \eta_i)^\gamma - \eta_i^\gamma$$

According to (18),

$$\frac{1}{\gamma} \sum_{j \neq M} (d\mathcal{G}_x v)_j$$

$$= \sum_{j \neq M} \sum_k \left((x_j + \eta_j)^{\gamma-1}[f_k \leq f_j] - \eta_j^{\gamma-1}[f_k < f_j] \right) v_k$$

$$= \sum_{j \neq M} \left((x_j + \eta_j)^{\gamma-1} \sum_k [f_k \leq f_j] v_k - \eta_j^{\gamma-1} \sum_k [f_k < f_j] v_k \right)$$

$$= \eta_M^{\gamma-1} \sum_{k \neq M} v_k$$

since the sum telescopes. Combining the last expression above with (22), the compatibility condition (23) becomes

$$0 = x_m^{\gamma-1} v_m \psi_1(z_0, z_1) - \psi_2(z_0, z_1) \eta_M^{\gamma-1} v_m \psi_1(x_m, \eta_M)/\psi_2(x_m, \eta_M)$$

Since Ψ is to be independent of γ, let $\gamma \downarrow 0$ and note that $z_0 \to 1$ and $z_1 \to 1$. After simplifying and rearranging the equation above, the result as $\gamma \downarrow 0$ is

$$\frac{\psi_2(x_m, \eta_M)}{\psi_1(x_m, \eta_M)} = \frac{x_m}{\eta_M} \frac{\psi_2(1, 1)}{\psi_1(1, 1)} \tag{24}$$

Focusing attention on an equivalence class – which makes x_m a function of η_M – consider the relation

$$\psi(x_m, \eta_M) = c$$

(for some constant c). Applying the implicit function theorem,

$$\frac{\psi_2}{\psi_1} = -\frac{d}{d\eta_M} x_m$$

and therefore (24) becomes the differential equation

$$\frac{d}{d\eta_M} x_m = \beta \frac{x_m}{\eta_M}$$

(for some constant β). Solving the differential equation yields

$$x_m = c\eta_M^{\beta}$$

and

$$\psi(x_m, \eta_M) = \frac{x_m}{\eta_M^{\beta}}$$

Using the invariant $1 = x_M + \eta_M$, this may be rephrased in terms of x_M by redefining ψ as

$$\psi(x_m, x_M) = \frac{x_m}{(1 - x_M)^{\beta}}$$

8   Conclusion

Coarse graining is a pervasive concept in science, but has so far not been systematically investigated within the field of Genetic Algorithms. Whereas the phrase "coarse graining" has previously been used by other researchers in connection with GAs (most notably by Chris Stephens) that use typically ascribes a different meaning to the phrase than considered here.

Previous examples of coarse grainings (in the sense used here) include the papers by Rabinovich and Wigderson, and by Muhlenbein and Voigt. Rather than considering specific fitness functions or operators (as they do), our intent is to develop methods which may discover, characterize, and elucidate general invariants of the mathematical objects by which genetic search is formalized.

The principal contribution made by this paper is the introduction and illustration of techniques which facilitate the analysis of coarse graining within the context of GAs. Most remarkable is the manner in which coarse gainings are dealt with. They are not guessed or noticed, to be pointed out and subsequently verified. Instead, they are *derived*.

The potential utility of the methods presented has been demonstrated by obtaining a number of new coarse graining results. In several cases, the coarse grainings derived were characterized as being the only ones possible (within the class of linear coarse grainings corresponding to partitions of the search space). In one case (section 7), a non linear coarse graining was computed by solving a differential equation.

Acknowledgment

References

H. Muhlenbein and H. M. Voigt: "Gene pool recombination in genetic algorithms" In *Proceedings of the Metaheuristics Inter. Conf., Norwell*, I. H. Osman and J. P. Kelly (eds). Kluwer Academic Publishers, c1995.

Y. Rabinovich and A. Wigderson: "An analysis of a simple generic algorithm" In *Proceedings of the 4th International Conference on Genetic Algorithms*, p 215–221. R. K. Belew and L. B. Booker (eds.). Morgan Kaufmann Publishers, c1991.

J. E. Rowe, M. D. Vose, A. H. Wright: "State aggregation and population dynamics in linear systems" *Artificial Life*, in press
`http://www.cs.umt.edu/u/wright/pubs.htm`

C. R. Stephens and H. Waelbroeck: "Schemata Evolution and Building Blocks". In *Evolutionary Computation Vol. 7 n. 2*, p 109–124. 1999

M. D. Vose: "Coarse Graining". Dagstuhl Seminar No 04081, 15.02.-20.02.04
`http://www.dagstuhl.de/04081/Materials/`

M. D. Vose: *The Simple Genetic Algorithm: Foundations and Theory*. Cambridge, Mass. MIT Press, c1999. ISBN: 026222058X