Available online at www.sciencedirect.com
ScienceDirect
Theoretical Computer Science 361 (2006) 111 – 129
Theoretical Computer Science
www.elsevier.com/locate/tcs

Differentiable coarse graining

Jonathan E. Rowe[a], Michael D. Vose[b,*], Alden H. Wright[c]

[a] School of Computer Science, University of Birmingham, England
[b] Computer Science Department, University of Tennessee, USA
[c] Department of Computer Science, University of Montana, USA

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, ranking selection, and mutation. A nonlinear coarse graining for ranking selection is also presented. A number of results concerning "form invariance" are given. Within the context of GAs, the primary contribution made is the illustration of a technique by which coarse grainings may be analyzed. It is applied to obtain a number of new coarse graining results.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Differentiable; Coarse graining; Selection; Mutation; Form invariance

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

$$\begin{array}{ccc} x & \xrightarrow{\;h\;} & h(x) \\ {\scriptstyle \Xi}\big\downarrow & & \big\downarrow{\scriptstyle \Xi} \\ \Xi x & \xrightarrow{\;\tilde{h}\;} & \Xi h(x) \end{array} \tag{1}$$

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. If the diagram commutes, the reduced complexity model $\tilde{h}$ is the *quotient* of h corresponding to the *coarse graining* Ξ. The quotient $\tilde{h}$ is referred to as a coarse graining of h (with respect to Ξ), and h is said to be *compatible* with Ξ.

* Corresponding author.
E-mail address: vose@cs.utk.edu (M.D. Vose).

0304-3975/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2006.04.007

Whereas modeling h in an approximate fashion (by relaxing commutativity of the diagram) is interesting, the central question of 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. Whereas reading this paper should acquaint one with what it is primarily concerned with, a few remarks will be made—in the context of GAs—to help clarify what is not a primary concern.

GA dynamics may equivalently be described with respect to various bases, typically either a string basis (elements of which correspond to particular strings) or a schema basis (elements of which correspond to particular schemata) [2]. Schemata are widely thought of as "coarse grained", by virtue of being defined in terms of collections of strings. [1] That notion, however, is to some extent arbitrary; one might likewise regard strings as "coarse grained", by virtue of the fact that they are definable in terms of collections of schemata. In this paper, the coarse graining results concerning GAs do not coincide with a change of basis (as, for instance, moving from strings to schemata); in that case the quotient $\tilde{h} = \Xi \circ h \circ \Xi^{-1}$ always exists, since a change of basis (i.e., Ξ) is invertible. [2]

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 the application of analytical tools rather than establishing particular results about any specific fitness function. Rather than addressing the general situation, however, those tools speak to a special kind of coarse graining (*differentiable coarse graining*) which is introduced in Section 3. The potential utility of those tools is demonstrated by obtaining a number of new coarse graining results.

Roughly speaking, this paper is organized into four parts. The first is this introduction and the next section where a few conceptual examples of general 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, etc.), followed by a reduction to special cases. Third, aspects of the theory of the Simple Genetic Algorithm [12] are reviewed in preparation for applying the necessary and sufficient condition, and to relate coarse graining to the more general stochastic setting in which GAs are defined. Fourth, the theory is applied to GAs in an investigation of quotients for selection and mutation, followed by a summary of results.

2. Conceptual overview

A few examples are briefly mentioned to make the general framework introduced above less abstract, to illustrate that in practice complex systems are frequently managed and understood with the aid of coarse grainings, and to provide some idea of where our applications fit within a more general context.

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 a simple 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. Invariants 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 $\Xi(x) =$

[1] The term "coarse grained" is put in quotes to distinguish it from coarse graining as used in this paper.

[2] In the invertible case, Ξ is called the *conjugacy map* between h and $\tilde{h}$, and they are said to be *conjugate*.

$E(x) - m(x)c^2$. 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 microcycles 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 single-step trajectory

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

might not exist (think of h as being analogous to a single microcycle), it nevertheless could be the case that a multi-step trajectory

$$x \mapsto h^{p(x)}(x) \mapsto h^{p(h^{p(x)}(x))}(x) \mapsto \cdots \tag{3}$$

does admit useful quotients. The applications to genetic algorithms presented in Sections 5–8, however, are limited to the single-step scenario (2) rather than the more general multi-step situation (3).

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 may be significant; they may aid in identifying tractable approximate models (i.e., models which have useful quotients).

If an "approximate coarse graining" of a model is desired (meaning that commutativity of diagram 1 is not strictly enforced), one might take that to be a strict coarse graining $\tilde{h}$ of some h which approximates the model. In the situation where such h is not given, but a candidate Ξ and $\widetilde{h} : \Xi X \to \Xi X$ are known, a relevant observation is that a compatible $h : X \to X$ such that commutativity holds is trivial to construct from Ξ and $\widetilde{h}$;[3] if the constructed h is deemed to approximate the intended model, then $\widetilde{h}$ could be regarded as an approximate coarse graining of the model.

The applications to genetic algorithms presented in Sections 5–8, however, are not concerned with approximation since the models being coarse grained are themselves exact.

3. Differentiable coarse graining

The following summarizes from [10]. Rather than beginning with a coarse graining, one will be obtained as a byproduct of a continuously differentiable map. Constraining the framework for coarse graining in this way facilitates the application of differential calculus (most coarse grainings 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 provide a useful vantage point from which to consider coarse grainings, and, in some circumstances, to enable their computation.

Let $\Psi : V \longrightarrow W$ be a continuously differentiable function between open subsets of finite-dimensional Euclidean spaces.[4] A *path* (with respect to Ψ) is a smooth function[5] $\rho : [0, 1] \longrightarrow V$ 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, in particular, path-connected components of level sets of Ψ. It follows that the image of any path is contained in some equivalence class.

---

[3] Let h map elements of $\Xi^{-1} \circ \Xi(x)$ to elements of $\Xi^{-1} \circ \tilde{h} \circ \Xi(x)$, where Ξ^{-1} denotes inverse image (under Ξ).

[4] Such spaces suffice for our purposes; a more general development is possible.

[5] By *smooth* we mean a differentiable function whose differential (over the interior of the domain) has a continuous extension to the entire domain.

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:

$$\begin{array}{ccc} V & \xrightarrow{h} & V \\ \Xi\downarrow & & \downarrow\Xi \\ V/\equiv & \xrightarrow{\tilde{h}} & V/\equiv \end{array} \tag{4}$$

In that case $\tilde{h}$ is called the quotient of h (corresponding to the coarse graining Ξ); the quotient $\tilde{h}$ is referred to as a coarse graining of h (with respect to Ξ), and h is said to be compatible with Ξ.

Let T_v be the *tangent space* of the equivalence class $\tilde{v}$ at v, defined by

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

where $\mathcal{L}\{\cdots\}$ denotes the linear span of $\{\cdots\}$, and for any function f differentiable at x, the differential of f at x is denoted by $\mathrm{d}f_x$. For any linear function L, denote the kernel of L by K_L. The proof of the following theorem can be found in [10].

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

$$\mathrm{d}h_x : T_x \longrightarrow K_{\mathrm{d}\Psi_{h(x)}}.$$

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

The special case where Ψ is linear is referred to as *linear coarse graining*, and the necessary and sufficient condition reduces to

$$\mathrm{d}h_x : K_\Psi \longrightarrow K_\Psi. \tag{5}$$

If both h and Ψ are linear, then the situation reduces to the case considered in [9],

$$h : K_\Psi \longrightarrow K_\Psi. \tag{6}$$

It should be noted that the sort of coarse graining presented in this section, which we call *differentiable coarse graining*, is not without loss of generality. The connected set

$$C = \{(x, y) : y = \sin(x^{-1}),\ x \neq 0\} \cup \{(x, y) : x = 0\}$$

is a level set of the continuously differentiable function

$$\Psi(x, y) = \mathrm{e}^{-x^{-2}}(y - \sin(x^{-1}))$$

but it cannot be an equivalence class (by our definition) since it is not path-connected [7]. Our requirement that paths be smooth is also restrictive; the Koch snowflake curve is arc-connected, but nowhere differentiable [4], and, any closed set (e.g., the Koch snowflake) can be a level set of a continuously differentiable function [3].

As mentioned in the Introduction, commutativity of diagram 1 is trivial when Ξ is invertible and $\tilde{h} = \Xi \circ h \circ \Xi^{-1}$. More generally, if $\tilde{h}$ is a any coarse graining of h with respect to Ξ, and if Θ is invertible, then $\Theta \circ \tilde{h} \circ \Theta^{-1}$ is a coarse graining of h with respect to $\Theta \circ \Xi$.[6]

[6] Any map conjugate to a quotient is also a quotient, when quotient and coarse graining are general (i.e., if the coarse graining is not required to be differentiable).

4. GAs and stochastic compatibility

This section presents a brief summary of relevant background from [12] to introduce the mathematical framework in which Theorem 1 will be applied.

Let τ denote the stochastic transition function for a finite population GA [7] over the search space $\Omega = \{0, \ldots, n-1\}$, [8] and let $\mathcal{G}$ be the corresponding infinite population model. [9] The transition matrix Q of the GA's Markov chain is defined by the probability that $\tau(p) = q$ and satisfies

$$Q_{p,q} = r! \prod \frac{(\mathcal{G}(p)_j)^{rq_j}}{(rq_j)!}, \tag{7}$$

where r is the population size, and where the population represented by the n-dimensional real vector p contains rp_j instances of j.

The (completion of the) population representation space is the simplex

$$\Lambda_n = \{\langle x_0, \ldots, x_{n-1}\rangle : x_i \geqslant 0,\ \mathbf{1}^{\mathrm{T}} x = 1\},$$

where $\langle \cdots \rangle$ denotes a column vector, $\mathbf{1}$ is the vector of all 1s, [10] and $\cdot^{\mathrm{T}}$ denotes transpose (of $\cdot$). Results of the previous section will be applied with $h = \mathcal{G}$ and V a path-connected (by smooth paths) neighborhood of Λ_n.

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 an equivalence relation between elements $p,\ p' \in V$ by

$$p \equiv p' \iff \Xi p = \Xi p'. \tag{8}$$

Note that

$$\mathbf{1}^{\mathrm{T}} \Xi = \mathbf{1}^{\mathrm{T}}. \tag{9}$$

The set of equivalence classes (of $\equiv$ on V) is

$$V/\equiv\, = \{V \cap (\Xi^{-1} \circ \Xi v) : v \in V\},$$

where Ξ^{-1} denotes inverse image. Therefore, $V/\equiv$ is set-isomorphic to ΞV (by Ξ) and the coarse graining (of the previous section) may without loss of generality be taken to coincide with the linear operator Ξ (of this section). This makes sense if Ψ is also chosen to coincide with Ξ, since then the level sets (of Ψ) are precisely the elements of $V/\equiv$ (they are path-connected by smooth paths since the inverse image of a point under a linear map is a subspace).

The observations above may be summarized as follows. An arbitrary equivalence relation over Ω gives rise to a linear operator Ξ, which is naturally a linear coarse graining. Under that coarse graining, populations p and p' are equivalent provided the populations they represent coincide when equivalent members of Ω are regarded as indistinguishable (in view of the definition of Ξ, that is what $\Xi p = \Xi p'$ asserts).

Compatibility in the stochastic case generalizes the definition given in the previous section; τ is said to be *compatible with* $\equiv$ (also said to be compatible with Ξ) if and only if

$$p \equiv p' \implies \forall q. \mathrm{Prob}\{\tau(p) \equiv q\} = \mathrm{Prob}\{\tau(p') \equiv q\}. \tag{10}$$

---

[7] τ maps the current population to the next generation.

[8] Whatever finite search space is intended, its elements may (in principle) be enumerated and referred to by integers.

[9] $\mathcal{G}$ maps the current population to the expected next generation.

[10] The dimension of the (column) vector $\mathbf{1}$ is intentionally ambiguous, to be inferred from context.

In that case, $\tilde{\tau}$ defined by $\tilde{\tau}(\Xi x) = \Xi\tau(x)$ is referred to as the *quotient* of τ (with respect to Ξ). It is known that the quotient $\tilde{\tau}$ exists if and only if a quotient $\tilde{\mathcal{G}}$ of $\mathcal{G}$ exists (with respect to Ξ), and the transition matrix $\tilde{Q}$ for the Markov chain corresponding to $\tilde{\tau}$ can be obtained from the formula for $Q_{p,q}$ (7) by replacing $\mathcal{G}$ by $\tilde{\mathcal{G}}$, p by Ξp, and q by Ξq. Moreover, the image under Ξ of an evolutionary trajectory beginning from any population p and generated by the transition matrix Q is statistically indistinguishable from an evolutionary trajectory beginning from Ξp and generated by $\tilde{Q}$ [12].

Therefore, the stochastic case has been reduced to a deterministic setting, and commutativity (of diagram 4) is of particular interest—for the theory of GAs—when h coincides with $\mathcal{G}$. Applications of linear coarse grainings in the following sections rely upon condition (5) to establish compatibility, where $\Psi = \Xi$ and V is a neighborhood of Λ_n. A relevant observation is therefore

$$K_\Xi = \left\{ v : \forall c^*. \sum_{i \equiv c^*} v_i = 0 \right\}. \tag{11}$$

Whereas the discussion above relates the behavior of a GA (i.e., τ or equivalently Q) to that of its infinite population model ($\mathcal{G}$), that discussion is in the context of a linear coarse graining. Independent of that context, connections between their respective evolutionary trajectories are extensive; progress made with coarse graining $\mathcal{G}$—by any means, whether the coarse graining is linear or not—reflects on τ (though in a less direct and more qualitative manner [12]).

5. Proportional selection + mutation

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

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

where $G = MF$ is a $n \times n$ matrix and $\mathbf{1}$ is the vector of all 1s. Here M is a column-stochastic [11] mutation matrix, 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) [12]. In particular,

$$\mathbf{1}^{\mathrm{T}} Gp = f^{\mathrm{T}} p. \tag{12}$$

The domain of immediate interest is Λ_n (the completion of the population representation space). Note that (12) implies $\mathbf{1}^{\mathrm{T}} Gp$ does not vanish in a neighborhood of Λ_n, provided fitness is positive (i.e., f has positive components). Positive fitness will be assumed throughout the remainder of this paper.

The following was established in [10]:

Theorem 2. *Let τ be the stochastic transition function for a simple GA with (proportional) fitness matrix F and mutation matrix M. Suppose positive fitness and zero crossover, and let coarse graining Ξ correspond (as in Section 4) to any equivalence relation $\equiv$ over Ω. 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 Ξ is that M is. If F is not compatible with Ξ, then a necessary and sufficient condition for τ to be compatible with Ξ is that the columns of M are equivalent.*

Theorem 2 speaks to the no-mutation case if $M = I$, and then the condition that columns of M be equivalent reduces to the requirement that $\equiv$ has only one equivalence class (in Ω), all populations (in Λ_n) are equivalent, and then $\Xi = \mathbf{1}^{\mathrm{T}}$.

Theorem 2 is put into sharper focus by the following result (established in [11]), where e_i refers to the ith column of the $n \times n$ identity matrix (indices begin with zero).

[11] M is nonnegative, and $\mathbf{1}^{\mathrm{T}} M = \mathbf{1}^{\mathrm{T}}$.

Theorem 3. *Let coarse graining Ξ correspond (as in Section 4) to any equivalence relation $\equiv$ over Ω. A necessary and sufficient condition for a mutation matrix M to be compatible with $\equiv$ is that for all i, j,*

$$i \equiv j \implies Me_i \equiv Me_j.$$

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 nonnegative.

A mutation operator whose matrix M has equivalent columns (with respect to an equivalence relation $\equiv$) is called *restorative* (with respect to $\equiv$). This is equivalent to the existence of a vector $w \in \Lambda_k$ (where k is the number of equivalence classes) such that

$$\Xi M = w\mathbf{1}^{\mathrm{T}}.$$

Whereas τ is compatible with $\equiv$ for every fitness function when mutation is restorative (by Theorems 2 and 3), restorative mutation has a more remarkable property. Note that if $x \in \Lambda_n$, then $\mathbf{1}^{\mathrm{T}}x = 1$. Given restorative mutation,

$$\Xi Mx = w\mathbf{1}^{\mathrm{T}}x = w$$

is independent of $x \in \Lambda_n$. That observation implies the following:

Theorem 4. *Let mutation having corresponding matrix M be restorative with respect to $\equiv$. For all $u, v \in \Lambda_n$, $Mu \equiv Mv$.*

A mutation operator is called *universal* when it is compatible with every equivalence relation over Ω.

Theorem 5. *A mutation operator with matrix M is universal (i.e., compatible with every equivalence relation $\equiv$ over Ω) if and only if there exists $w \in \Lambda_n$ and $0 \leqslant \alpha$ such that*

$$M = (1-\alpha)I + \alpha w\mathbf{1}^{\mathrm{T}}$$

is nonnegative.

Proof. Suppose that the displayed matrix above is nonnegative. Since

$$\mathbf{1}^{\mathrm{T}}M = (1-\alpha)\mathbf{1}^{\mathrm{T}} + \alpha\mathbf{1}^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}}$$

the matrix M is both nonnegative and column stochastic, and therefore corresponds to a mutation operator. Let $i \equiv j$, so that $e_i - e_j \in K_\Xi$ (where the coarse graining Ξ corresponds to an equivalence relation $\equiv$ over Ω). Note that

$$Me_i = (1-\alpha)e_i + \alpha w.$$

Hence

$$Me_i - Me_j = (1-\alpha)(e_i - e_j) \in K_\Xi.$$

Conversely, suppose M is universal. First, consider the special case $n = 2$ (the case $n = 1$ is trivial).

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a-b & 0 \\ 0 & d-c \end{pmatrix} + \begin{pmatrix} b \\ c \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix},$$

where $a + c = 1 = b + d$. Therefore $a - b = d - c$. If $b = c = 0$ then $M = I$, so let $\alpha = 0$ and choose w arbitrarily. Otherwise, let $\alpha = b + c$ and choose $w = \alpha^{-1}\langle b, c\rangle$ to obtain

$$M = \beta I + \alpha w\mathbf{1}^{\mathrm{T}}.$$

Since $\mathbf{1}^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}}M = (\beta + \alpha)\mathbf{1}^{\mathrm{T}}$, it follows that $\beta = 1 - \alpha$.

Next, consider the general case $n > 2$. Let $0^* \equiv h$ and $k \neq h \Rightarrow 1^* \equiv k$. Let $i \neq j$ both be equivalent to 1^*. Since M is compatible (with $\equiv$),

$$\Xi M e_i = \Xi M e_j .$$

Multiplying through by e_1^{T} leads to

$$\sum_{k \equiv 1^*} M_{k,i} = \sum_{k \equiv 1^*} M_{k,j} .$$

Since M is column stochastic (and $k \neq h \Leftrightarrow 1^* \equiv k$), the above is equivalent to

$$1 - M_{h,i} = 1 - M_{h,j} .$$

It follows that with the possible exception of the diagonal elements, M has identical rows (h, i, j are arbitrary, subject to being distinct) and can therefore be expressed as

$$M = D + \alpha w \mathbf{1}^{\mathrm{T}},$$

where D is diagonal, $w \in \Lambda_n$, and $0 \leqslant \alpha$. Multiplying through by $\mathbf{1}^{\mathrm{T}}$ yields

$$\mathbf{1}^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}} D + \alpha \mathbf{1}^{\mathrm{T}} .$$

It follows that $D = (1 - \alpha) I$. □

Unlike universal mutation, restorative mutation cannot (for general $\equiv$) be arbitrarily close to the identity, since otherwise

$$\lim_{M \to I} \Xi M e_j = \Xi e_j$$

is independent of j, and consequently $[i^* \equiv j]$ is independent of j (i.e., there can be only one equivalence class). Hence, when $\equiv$ has more than one class and if there exist equivalent population members with unequal fitness, it is impossible for τ to remain compatible (in the proportional selection + mutation case) as mutation vanishes.

5.1. Form invariance

If h is compatible with a coarse graining Ξ corresponding (as in Section 4) to any equivalence relation $\equiv$ over Ω, the corresponding quotient is given by

$$\tilde{h} = \Xi \circ h \circ D \Xi^{\mathrm{T}},$$

where D is the diagonal matrix

$$D_{i,i} = 1 / \sum_{j \equiv i} 1 .$$

Moreover, $\Xi D \Xi^{\mathrm{T}} = I$, which implies $x \equiv D \Xi^{\mathrm{T}} \Xi x$, and therefore when h is compatible, $h(x) \equiv h \circ D \Xi^{\mathrm{T}} \Xi x$ and so $\Xi h(x) = \Xi \circ h \circ D \Xi^{\mathrm{T}} \Xi x$ (see [12]).

To say h is *form invariant under coarse graining* is to assert that $\tilde{h}$ has the same form as h (when h is compatible with $\equiv$).

Theorem 6. *Proportional selection + mutation is form invariant.*

Proof. Theorems 2 and 3 imply M is compatible when $\mathcal{G}$ is, and therefore

$$\begin{aligned}\Xi \circ \mathcal{G} \circ D\Xi^{\mathrm{T}} x &= \frac{(\Xi M) F D \Xi^{\mathrm{T}} x}{\mathbf{1}^{\mathrm{T}}(\Xi M) F D \Xi^{\mathrm{T}} x} \\ &= \frac{(\Xi M D \Xi^{\mathrm{T}} \Xi) F D \Xi^{\mathrm{T}} x}{\mathbf{1}^{\mathrm{T}}(\Xi M D \Xi^{\mathrm{T}} \Xi) F D \Xi^{\mathrm{T}} x} \\ &= \frac{(\Xi M D \Xi^{\mathrm{T}})(\Xi F D \Xi^{\mathrm{T}}) x}{\mathbf{1}^{\mathrm{T}}(\Xi M D \Xi^{\mathrm{T}})(\Xi F D \Xi^{\mathrm{T}}) x} \\ &= \frac{\widetilde{M}\widetilde{F} x}{\mathbf{1}^{\mathrm{T}}\widetilde{M}\widetilde{F} x}.\end{aligned}$$

Note that

$$\mathbf{1}^{\mathrm{T}}\widetilde{M} = \mathbf{1}^{\mathrm{T}}\Xi M D \Xi^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}} M D \Xi^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}} D \Xi^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}} \Xi D \Xi^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}} I = \mathbf{1}^{\mathrm{T}}$$

hence $\widetilde{M}$ is column stochastic (i.e., a mutation matrix). Moreover, $\widetilde{F}$ is a diagonal fitness matrix,

$$\begin{aligned}\widetilde{F}_{i,j} &= \sum_{a,b,c} \Xi_{i,a} F_{a,b} D_{b,c} \Xi_{j,c} \\ &= \sum_{a,b,c} [a \equiv i^*][a = b] f_a \frac{[b = c]}{\sum_{\ell \equiv c} 1} [c \equiv j^*] \\ &= [i^* \equiv j^*] \sum_{c \equiv i^*} \frac{f_c}{\sum_{\ell \equiv c} 1}. \qquad \square\end{aligned}$$

6. Binary tournament selection + mutation

A zero mutation, zero crossover, tournament selection GA with tournament size t and fitness function f has corresponding infinite population model [12]

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

where

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

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

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

It follows that

$$(\mathrm{d}\mathcal{F}_x v)_i = 2 v_i x_i + 2 \sum_l [f_l < f_i](v_i x_l + x_i v_l). \tag{13}$$

Note that (13) is a symmetric expression in x and v, and therefore $\mathrm{d}\mathcal{F}_x v = \mathrm{d}\mathcal{F}_v x$ is linear in both x and v. In view of this, the compatibility condition is that for all $x \in V$, and for all $v \in K_\Xi$,

$$\mathrm{d}\mathcal{F}_x v = \sum_h x_h \mathrm{d}\mathcal{F}_{e_h} v \in K_\Xi.$$

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

$$v \in K_\Xi \implies \mathrm{d}\mathcal{F}_{e_h} v \in K_\Xi. \tag{14}$$

Moreover, the ith component of the differential above simplifies (from (13)) to

$$(\mathrm{d}\mathcal{F}_{e_h} v)_i = 2v_i[f_h < f_i] + 2[h = i]\sum_l [f_l \leqslant f_i]v_l. \tag{15}$$

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 $-1 = z_{-1} < z_0 < \cdots < z_{k-1} = n-1$. 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^*. An equivalence relation is referred to as fitness-contiguous when its equivalence classes are. The following was established in [10]:

Theorem 7. *Binary tournament selection is compatible with $\equiv$ if and only if the equivalence relation is fitness-contiguous.*

Incorporating mutation complicates matters. Let mutation have corresponding matrix M, and consider the GA as above *with mutation included* (i.e., "binary tournament selection + mutation"); its infinite population model is $\mathcal{G} = M\mathcal{F}$ with differential $\mathrm{d}\mathcal{G}_x = M\,\mathrm{d}\mathcal{F}_x$. Choosing $h = \theta(0)$ in (15) yields

$$\mathrm{d}\mathcal{G}_{e_{\theta(0)}} = M\,\mathrm{d}\mathcal{F}_{e_{\theta(0)}} = 2M.$$

Therefore, a necessary condition for $\mathcal{G}$ to be compatible with $\equiv$ is that M is. Just as $\mathrm{d}\mathcal{F}_x v = \mathrm{d}\mathcal{F}_v x$ is linear in both x and v, the same is true of $\mathrm{d}\mathcal{G}_x$, and consequently (as for (14)) compatibility of $\mathcal{G}$ is equivalent (via (11) and (15)) to the condition that for all h and c^*,

$$v \in K_\Xi \implies 0 = \sum_j v_j[f_h < f_j]\sum_{i\equiv c^*} M_{i,j} + \sum_l [f_l \leqslant f_h]v_l \sum_{i\equiv c^*} M_{i,h}. \tag{16}$$

Since fitness is injective, and $v \in K_\Xi$ implies $\mathbf{1}^{\mathrm{T}} v = 0$, this can be rewritten as

$$v \in K_\Xi \implies 0 = \sum_j v_j[f_h < f_j]\sum_{i\equiv c^*} (M_{i,j} - M_{i,h}). \tag{17}$$

This condition is linear in v, and so attention may be restricted to a basis for K_Ξ, which (by (11)) can be taken to be

$$\mathcal{B} = \bigcup_{r^*} \left\{ v : \sum v_i = 0,\ v_i \neq 0 \Rightarrow i \equiv r^* \right\}.$$

The *extent* of representative r^* (with respect to f and $\equiv$) is defined as

$$E_{r^*} = \left\{ i : \min_{k\equiv r^*} f_k \leqslant f_i \leqslant \max_{k\equiv r^*} f_k \right\}.$$

A mutation operator with matrix M is called *contiguous* (with respect to f and $\equiv$) if it is compatible with $\equiv$ and for all r^*,

$$i, j \in E_{r^*} \implies Me_i \equiv Me_j.$$

Theorem 8. *Let τ be the stochastic transition function for a simple GA with fitness f, binary tournament selection $\mathcal{F}$, and mutation matrix M. Suppose injective fitness and zero crossover, and let coarse graining Ξ correspond* *(as in Section 4) to any equivalence relation $\equiv$ over Ω. A necessary and sufficient condition for τ to be compatible with Ξ is that M is contiguous with respect to f and $\equiv$.*

Proof. Suppose M is not contiguous. If M is not compatible, then neither is τ (it was previously observed that a necessary condition for $\mathcal{G}$ to be compatible with $\equiv$ is that M is). Suppose, therefore, that M is compatible, and there exist $j, h \in E_{r^*}$ such that $Me_j \not\equiv Me_h$. In particular, $Me_j - Me_h \notin K_\Xi$ and therefore (by (11)) there exists c^* such that

$$0 \neq \sum_{i \equiv c^*} (M_{i,j} - M_{i,h}).$$

Without loss of generality $f_h < f_j$ (relabel if necessary). Let α and β be minimally and maximally fit elements of E_{r^*} (since f is injective, $\alpha \equiv \beta \equiv r^*$). Without loss of generality $j \equiv r^*$ since if that is not the case then either $Me_j \equiv Me_\beta$ in which case redefine j to be β, or else $Me_j \not\equiv Me_\beta$ in which case redefine h and j to be j and β, (respectively). Let $v \in \mathcal{B}$ be zero except for the two components $v_\alpha = -v_j \neq 0$. It follows that the right-hand side of (17) is

$$v_j \sum_{i \equiv c^*} (M_{i,j} - M_{i,h}) \neq 0.$$

Conversely, suppose M is contiguous. Let $v \in \mathcal{B}$, where $v_j \neq 0 \Rightarrow j \equiv r^*$, and let α and β be as above. If $f_\beta \leqslant f_h$, then the right-hand side of (17) is zero due to the factor $v_j[f_h < f_j]$. If $f_h < f_\alpha$, then the right-hand side of (17) is

$$\sum_j v_j \sum_{i \equiv c^*} M_{i,j} - \sum_{i \equiv c^*} M_{i,h} \sum_j v_j = \sum_{i \equiv c^*} M_{i,\alpha} \sum_j v_j = 0$$

since $\mathbf{1}^{\mathrm{T}} v = 0$ and $v_j \neq 0 \Longrightarrow Me_j \equiv Me_\alpha$. The remaining case is $h \in E_{r^*}$, but then the inner sum on the right-hand side of (17) is zero (since $v_j \neq 0 \Longrightarrow j \in E_{r^*}$ and so $Me_j \equiv Me_h$). $\square$

6.1. Form invariance

A simple computation verifies that binary tournament selection has the form

$$\mathcal{F}(p) = \mathrm{diag}(p)Bp,$$

where $B_{i,j} = [f_i \geqslant f_j] + [f_i > f_j]$. Abstracting out the fitness function f, the matrix B may be characterized as being similar via a permutation matrix to a lower triangular matrix F of the form

$$F_{i,j} = \begin{cases} 2 & \text{if } i > j, \\ 1 & \text{if } i = j, \\ 0 & \text{if } i < j \end{cases}$$

(since the permutation matrix then determines some corresponding f).

Theorem 9. *Binary tournament selection + mutation is form invariant if the equivalence relation is fitness-contiguous.*

Proof. Theorem 8 implies M is compatible (when $\mathcal{G}$ is), and therefore

$$\begin{aligned} \Xi \circ \mathcal{G} \circ D\Xi^{\mathrm{T}} x &= (\Xi M)\mathrm{diag}(D\Xi^{\mathrm{T}} x)BD\Xi^{\mathrm{T}} x \\ &= (\Xi M D\Xi^{\mathrm{T}} \Xi)\mathrm{diag}(D\Xi^{\mathrm{T}} x)BD\Xi^{\mathrm{T}} x \\ &= (\Xi M D\Xi^{\mathrm{T}})(\Xi\,\mathrm{diag}(D\Xi^{\mathrm{T}} x)BD\Xi^{\mathrm{T}} x) \\ &= \widetilde{M}(\mathrm{diag}(x)\mathcal{Q}x), \end{aligned}$$

where $\mathcal{Q}$ is a matrix satisfying

$$\Xi\,\mathrm{diag}(D\Xi^{\mathrm{T}} y)BD\Xi^{\mathrm{T}} x = \mathrm{diag}(y)\mathcal{Q}x. \tag{18}$$

Note that both sides (above) are bilinear, so it suffices to determine $\mathcal{Q}$ when x and y are basis vectors. Choosing $y = e_i$, $x = e_j$ and multiplying through by $\mathbf{1}^{\mathrm{T}}$ yields

$$\begin{aligned}\mathbf{1}^{\mathrm{T}}\Xi\,\mathrm{diag}(D\Xi^{\mathrm{T}}e_i)BD\Xi^{\mathrm{T}}e_j &= \mathbf{1}^{\mathrm{T}}\mathrm{diag}(D\Xi^{\mathrm{T}}e_i)BD\Xi^{\mathrm{T}}e_j\\ &= (D\Xi^{\mathrm{T}}e_i)^{\mathrm{T}}BD\Xi^{\mathrm{T}}e_j\\ &= e_i^{\mathrm{T}}(\Xi DBD\Xi^{\mathrm{T}})e_j,\\ \mathbf{1}^{\mathrm{T}}\mathrm{diag}(e_i)\mathcal{Q}e_j &= \mathcal{Q}_{i,j}.\end{aligned}$$

The matrix $\mathcal{Q} = \Xi DBD\Xi^{\mathrm{T}}$ satisfies (18) (as a simple computation verifies). Note that

$$\begin{aligned}\mathcal{Q}_{i,j} &= \sum_{a,b,c,d}\Xi_{i,a}D_{a,b}B_{b,c}D_{c,d}\Xi^{\mathrm{T}}_{\mathrm{d},j}\\ &= \sum_{a,b,c,d}[a\equiv i^*]\frac{[a=b]}{\sum_{\ell\equiv a}1}B_{b,c}\frac{[c=d]}{\sum_{\ell\equiv c}1}[d\equiv j^*]\\ &= \sum_{a\equiv i^*}\sum_{c\equiv j^*}B_{a,c}\Big/\left(\sum_{\ell\equiv i^*}1\sum_{\ell\equiv j^*}1\right)\\ &= \frac{\sum_{a\equiv\theta(z_i)}\sum_{c\equiv\theta(z_j)}[f_a\geqslant f_c]+[f_a>f_c]}{\sum_{\ell\equiv\theta(z_i)}1\sum_{\ell\equiv\theta(z_j)}1} = \begin{cases}2 & \text{if } \theta(z_i)>\theta(z_j),\\ 1 & \text{if } \theta(z_i)=\theta(z_j),\\ 0 & \text{if } \theta(z_i)<\theta(z_j).\end{cases}\end{aligned}$$ □

7. Ranking selection + mutation

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

$$\mathcal{F}(x)_i = \int_{\sum[f_j<f_i]x_j}^{\sum[f_j\leqslant f_i]x_j}\varrho(y)\,\mathrm{d}y,$$

where ϱ is any continuous increasing probability density over [0, 1] (see [12]). Define η by

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

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

(recall that $i<j \iff f_{\theta(i)} < f_{\theta(j)}$), and let $\eta_{\theta(n)}$ denote 1. It follows that

$$\mathcal{F}(x)_i = \varphi(x_i+\eta_i) - \varphi(\eta_i), \tag{19}$$

$$\mathrm{d}\mathcal{F}_x v = \sum_i e_i\sum_k(\varrho(x_i+\eta_i)[f_k\leqslant f_i] - \varrho(\eta_i)[f_k<f_i])v_k, \tag{20}$$

where φ is an anti-derivative of ϱ (see [12]). Choosing $x = e_h$,

$$\eta_{\theta(\ell)} = \eta_{\theta(\ell-1)} + (e_h)_{\theta(\ell-1)} = \begin{cases}0 & \text{for } \ell\leqslant q \text{ such that } \theta(q)=h,\\ 1 & \text{for } \ell>q \text{ such that } \theta(q)=h.\end{cases}$$

Therefore, let $i = \theta(\ell)$ to conclude

$$\begin{aligned}\varrho(\eta_i) &= \varrho([\ell > q \text{ such that } \theta(q) = h])\\ &= \varrho([f_i > f_h]),\end{aligned}$$

$$\begin{aligned}\varrho((e_h)_i + \eta_i) &= \varrho(\eta_{\theta(\ell+1)})\\ &= \varrho([\ell \geqslant q \text{ such that } \theta(q) = h])\\ &= \varrho([f_i \geqslant f_h]).\end{aligned}$$

It follows that $(\mathrm{d}\mathcal{F}_{e_h} v)_i$ simplifies to yield

$$\begin{aligned}&\varrho([f_i \geqslant f_h])\sum_k [f_k \leqslant f_i] v_k - \varrho([f_i > f_h])\sum_k [f_k < f_i] v_k\\ &= [i = h]\left(\varrho([f_i \geqslant f_h]) v_i + (\varrho(1) - \varrho(0))\sum_k [f_k < f_i] v_k\right) + [i \neq h](\varrho([f_i \geqslant f_h]) v_i)\\ &= \varrho([f_i \geqslant f_h]) v_i + [i = h](\varrho(1) - \varrho(0))\sum_k [f_k < f_h] v_k. \qquad (21)\end{aligned}$$

Compatibility requires (via (11)) that for all c^*, and all $v \in K_\Xi$,

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

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

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

This implies $\equiv$ is fitness-contiguous (let v have exactly two nonzero components $v_i = -v_j \neq 0$ where $i \equiv j \equiv c^*$; since $\varrho(1) > \varrho(0)$, the expression above can only be zero if every j equivalent to c^* satisfies $[f_j < f_h]$ or else every i equivalent to c^* satisfies $[f_i > f_h]$). The following was established in [10]:

Theorem 10. *Ranking selection is compatible with $\equiv$ if and only if the equivalence relation is fitness-contiguous.*

"Ranking selection + mutation" has differential

$$\mathrm{d}\mathcal{G}_x v = M\, \mathrm{d}\mathcal{F}_x v.$$

Note that $(\mathrm{d}\mathcal{F}_{e_{\theta(0)}} v)_i = \varrho(1) v_i$ (by (21)), thus $\mathrm{d}\mathcal{G}_{e_{\theta(0)}} v = \varrho(1) M v$. Therefore, compatibility of $\mathcal{G}$ requires compatibility of M. It follows from (20) that

$$\begin{aligned}(\mathrm{d}\mathcal{G}_x v)_i &= \sum_{\ell,k} M_{i,\ell}(\varrho(x_\ell + \eta_\ell)[f_k \leqslant f_\ell] - \varrho(\eta_\ell)[f_k < f_\ell]) v_k\\ &= \sum_{\ell,k} M_{i,\theta(\ell)}\varrho(x_{\theta(\ell)} + \eta_{\theta(\ell)})[f_{\theta(k)} \leqslant f_{\theta(\ell)}] v_{\theta(k)} - \sum_{\ell,k} M_{i,\theta(\ell)}\varrho(\eta_{\theta(\ell)})[f_{\theta(k)} < f_{\theta(\ell)}]\, v_{\theta(k)}\\ &= \sum_k (M_{i,\theta(n-1)}\varrho(\eta_{\theta(n)}) + \sum_{\ell=0}^{n-2} M_{i,\theta(\ell)}\varrho(\eta_{\theta(\ell+1)})[f_{\theta(k)} \leqslant f_{\theta(\ell)}]) v_{\theta(k)}\\ &\quad -\sum_k \sum_{\ell=1}^{n-1} M_{i,\theta(\ell)}\, \varrho(\eta_{\theta(\ell)})[f_{\theta(k)} < f_{\theta(\ell)}] v_{\theta(k)}\end{aligned}$$

$$= M_{i,\theta(n-1)}\, \varrho(1) \sum_k v_{\theta(k)} + \sum_k \sum_{\ell=1}^{n-1} M_{i,\theta(\ell-1)} \varrho(\eta_{\theta(\ell)})[k \leqslant \ell-1]\, v_{\theta(k)} - \sum_k \sum_{\ell=1}^{n-1} M_{i,\theta(\ell)} \varrho(\eta_{\theta(\ell)})[k < \ell\,]\, v_{\theta(k)}$$

$$= \sum_k \sum_{\ell=1}^{n-1} M_{i,\theta(\ell-1)} \varrho(\eta_{\theta(\ell)})[k < \ell]\, v_{\theta(k)} - M_{i,\theta(\ell)} \varrho(\eta_{\theta(\ell)})[k < \ell\,] v_{\theta(k)}$$

$$= \sum_\ell (M_{i,\theta(\ell-1)} - M_{i,\theta(\ell)}) \varrho(\eta_{\theta(\ell)}) \sum_{k<\ell} v_{\theta(k)}.$$

Therefore, the compatibility condition is that for all $v \in K_\Xi$ and all c^*,

$$0 = \sum_\ell \varrho(\eta_{\theta(\ell)}) \sum_{i \equiv c^*} (M_{i,\theta(\ell-1)} - M_{i,\theta(\ell)}) \sum_{k<\ell} v_{\theta(k)}. \tag{22}$$

The analogue of Theorem 8 holds for ranking selection + mutation:

Theorem 11. *Let τ be the stochastic transition function for a simple GA with fitness f, ranking selection $\mathcal{F}$, and mutation matrix M. Suppose injective fitness and zero crossover, and let coarse graining Ξ correspond* (*as in Section* 4) *to any equivalence relation $\equiv$ over Ω. A necessary and sufficient condition for τ to be compatible with Ξ is that M is contiguous with respect to f and $\equiv$.*

Proof. Suppose M is not contiguous. If M is not compatible, then neither is τ (it was previously observed that a necessary condition for $\mathcal{G}$ to be compatible with $\equiv$ is that M is). Suppose, therefore, that M is compatible, and there exist $\theta(u-1), \theta(u) \in E_{r^*}$ such that $Me_{\theta(u-1)} \not\equiv Me_{\theta(u)}$, where u is chosen as large as possible. In particular, there exists c^* such that

$$0 \neq \sum_{i \equiv c^*} (M_{i,\theta(u-1)} - M_{i,\theta(u)}).$$

Let α and β be minimally and maximally fit elements of E_{r^*}. Let $v \in \mathcal{B}$ be zero except for the two components $v_\alpha = -v_\beta \neq 0$. Note that (22) is zero if either $\ell \leqslant \theta^{-1}(\alpha)$ or $\ell > \theta^{-1}(\beta)$ (because of the innermost sum). Choosing $x = \mathbf{1}/n + \varepsilon e_{\theta(u-1)} - \varepsilon e_{\theta(u)}$, condition (22) reduces to

$$0 = \sum_{\theta^{-1}(\alpha) < \ell \leqslant \theta^{-1}(\beta)} \varrho(\eta_{\theta(\ell)}) \sum_{i \equiv c^*} (M_{i,\theta(\ell-1)} - M_{i,\theta(\ell)})\, v_\alpha$$

$$= \varrho(\varepsilon + u/n) \sum_{i \equiv c^*} (M_{i,\theta(u-1)} - M_{i,\theta(u)}) + \sum_{\theta^{-1}(\alpha) < \ell < u} \varrho(\ell/n) \sum_{i \equiv c^*} (M_{i,\theta(\ell-1)} - M_{i,\theta(\ell)}).$$

A contradiction is obtained by varying ε.

Conversely, suppose M is contiguous. Let $v \in \mathcal{B}$, where $v_j \neq 0 \Rightarrow j \equiv r^*$. As observed above, condition (22) reduces to

$$0 = \sum_{\theta^{-1}(\alpha) < \ell \leqslant \theta^{-1}(\beta)} \varrho(\eta_{\theta(\ell)}) \sum_{i \equiv c^*} (M_{i,\theta(\ell-1)} - M_{i,\theta(\ell)}) \sum_{k<\ell} v_{\theta(k)}$$

$$= \sum_{\theta^{-1}(\alpha) \leqslant k < \theta^{-1}(\beta)} v_{\theta(k)} \sum_{k < \ell \leqslant \theta^{-1}(\beta)} \varrho(\eta_{\theta(\ell)}) \sum_{i \equiv c^*} (M_{i,\theta(\ell-1)} - M_{i,\theta(\ell)}).$$

The inner sum above is zero (since $\theta(\ell-1), \theta(\ell) \in E_{r^*}$). □

7.1. Form invariance

By (19), ranking selection has the form

$$\mathcal{F}(x)_i = \varphi(x_i + \eta_i) - \varphi(\eta_i).$$

Here $\eta = P^{\mathrm{T}}(T - I)Px$, T is the triangular matrix $T_{i,j} = [i \geqslant j]$, and P is some permutation matrix (which corresponds to f via $P_{i,j} = [\theta(i) = j]$),

$$\eta_{\theta(i+1)} = (P\eta)_{i+1} = \sum_j (T - I)_{i+1,j}(Px)_j = \sum_{j<i+1} x_{\theta(j)} = \eta_{\theta(i)} + x_{\theta(i)}.$$

Therefore

$$\mathcal{F}(x)_{\theta(i)} = \varphi(\eta_{\theta(i+1)}) - \varphi(\eta_{\theta(i)}). \tag{23}$$

If $\equiv$ is fitness-contiguous,

$$\begin{aligned}(T\Xi x)_i &= \sum_{j,k} [i \geqslant j][k \equiv j^*]x_k \\ &= \sum_{j \leqslant i} \sum_h [\theta(h) \equiv \theta(z_j)]x_{\theta(h)} \\ &= \sum_{j \leqslant i} \sum_{z_{j-1} < h \leqslant z_j} x_{\theta(h)} \\ &= \eta_{\theta(z_i+1)}.\end{aligned}$$

Assume $\equiv$ is fitness-contiguous, let $x = D\Xi^{\mathrm{T}}y$, and recall $\Xi D \Xi^{\mathrm{T}} = I$. It follows from the above that

$$\begin{aligned}\widetilde{\mathcal{F}}(y)_i &= \sum_j \Xi_{i,j}\mathcal{F}(D\Xi^{\mathrm{T}}y)_j \\ &= \sum_j [j \equiv i^*]\mathcal{F}(x)_j \\ &= \sum_h [\theta(h) \equiv \theta(z_i)]\mathcal{F}(x)_{\theta(h)} \\ &= \sum_{z_{i-1} < h \leqslant z_i} \varphi(\eta_{\theta(h+1)}) - \varphi(\eta_{\theta(h)}) \\ &= \varphi(\eta_{\theta(z_i+1)}) - \varphi(\eta_{\theta(z_{i-1}+1)}) \\ &= \varphi((T\Xi D\Xi^{\mathrm{T}}y)_i) - \varphi((T\Xi D\Xi^{\mathrm{T}}y)_{i-1}) \\ &= \varphi((Ty)_i) - \varphi((Ty)_{i-1}).\end{aligned} \tag{24}$$

Theorem 12. *Ranking selection + mutation is form invariant if the equivalence relation is fitness-contiguous.*

Proof. Theorem 11 implies M is compatible (when $\mathcal{G}$ is), and therefore

$$\begin{aligned}\Xi \circ \mathcal{G} \circ D\Xi^{\mathrm{T}}y &= (\Xi M)\mathcal{F}(D\Xi^{\mathrm{T}}y) \\ &= (\Xi M D\Xi^{\mathrm{T}}\Xi)\mathcal{F}(D\Xi^{\mathrm{T}}y) \\ &= (\Xi M D\Xi^{\mathrm{T}})(\Xi\mathcal{F}(D\Xi^{\mathrm{T}}y)) \\ &= \widetilde{M}\widetilde{\mathcal{F}}(y).\end{aligned}$$

In view of (24), it remains to show there exists some permutation matrix P such that

$$\begin{aligned}(Ty)_i &= (P^{\mathrm{T}}(T - I)Py)_i + y_i, \\ (Ty)_{i-1} &= (P^{\mathrm{T}}(T - I)Py)_i.\end{aligned}$$

Note that $P = I$. □

8. A nonlinear coarse graining

Previous applications have involved linear coarse grainings corresponding to an equivalence relation over Ω. A nonlinear coarse graining is derived below for ranking selection $\mathcal{F}$.

To simplify analysis choose $\varphi(x) = x^\gamma$ in (19) (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 $1 = x_M + \eta_M$ (see Eq. (19) for the definition of η), so we choose to work with $\Psi(x) = \psi(x_m, \eta_M)$ for some function ψ.

The "invariant parameter" (γ) is introduced above to provide the analysis with additional leverage by enabling simplifications. The motivation for restricting to a function ψ of two arguments is to provide a simple setting for the application of the implicit function theorem to obtain a differential equation which may yield a coarse graining.

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

$$0 = \sum_j v_j \frac{\partial \Psi}{\partial x_j} = v_m \psi_1(x_m, \eta_M) + \psi_2(x_m, \eta_M) \sum_{j \neq M} v_j. \tag{25}$$

Hence

$$\sum_{j \neq M} v_j = -v_m \psi_1(x_m, \eta_M)/\psi_2(x_m, \eta_M). \tag{26}$$

Using the form of (25) with $v \leftarrow \mathrm{d}\mathcal{F}_x v$ and $x \leftarrow \mathcal{F}(x)$, a sufficient condition for compatibility (from Theorem 1: $v \in K_{\mathrm{d}\Psi_x} \Longrightarrow \mathrm{d}\mathcal{F}_x v \in K_{\mathrm{d}\Psi_{\mathcal{F}(x)}}$) is

$$0 = (\mathrm{d}\mathcal{F}_x v)_m \psi_1(x'_m, \eta'_M) + \psi_2(x'_m, \eta'_M) \sum_{j \neq M} (\mathrm{d}\mathcal{F}_x v)_j, \tag{27}$$

where (via (23))

$$x'_m = \mathcal{F}(x)_{\theta(0)} = \varphi(x_m), \tag{28}$$

$$\eta'_M = \sum_{i<n-1} \mathcal{F}(x)_{\theta(i)} = \sum_{i<n-1} (\varphi(\eta_{\theta(i+1)}) - \varphi(\eta_{\theta(i)})) = \varphi(\eta_M), \tag{29}$$

$$x'_M = 1 - \eta'_M = 1 - \varphi(1 - x_M) \tag{30}$$

and the anti-derivative of ϱ is chosen to satisfy $\varphi(0) = 0$. According to (20),

$$\begin{aligned} \sum_{j \neq M} (\mathrm{d}\mathcal{F}_x v)_j &= \sum_{j \neq M} \sum_k (\varrho(x_j + \eta_j)[f_k \leqslant f_j] - \varrho(\eta_j)[f_k < f_j])\, v_k \\ &= \sum_{h<n-1} \left(\varrho(\eta_{\theta(h+1)}) \sum_{\ell<h+1} v_{\theta(\ell)} - \varrho(\eta_{\theta(h)}) \sum_{\ell<h} v_{\theta(\ell)} \right) \\ &= \varrho(\eta_M) \sum_{\ell \neq M} v_\ell. \end{aligned}$$

Combining the above with (26) gives

$$\sum_{j \neq M} (\mathrm{d}\mathcal{F}_x v)_j = -\varrho(\eta_M) v_m \psi_1(x_m, \eta_M)/\psi_2(x_m, \eta_M). \tag{31}$$

Note also that

$$(\mathrm{d}\mathcal{F}_x v)_m = \sum_k \varrho(\eta_{\theta(1)})[f_k \leqslant f_{\theta(0)}]v_k = \varrho(x_m)v_m. \tag{32}$$

Therefore, the (sufficient) compatibility condition is implied by the following (using (28), (29), (31) and (32) with (27) and simplifying)

$$\frac{\psi_1(x_m, \eta_M)}{\psi_2(x_m, \eta_M)} = \frac{\varrho(x_m)}{\varrho(\eta_M)} \frac{\psi_1(\varphi(x_m), \varphi(\eta_M))}{\psi_2(\varphi(x_m), \varphi(\eta_M))}. \tag{33}$$

Focusing attention on an equivalence class—which makes x_m a function of η_M—consider a level set determined by

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

for some constant c. Applying the implicit function theorem (see [5])

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

Combining this with (33) yields

$$\frac{\mathrm{d}}{\mathrm{d}\eta_M} x_m = -\frac{\varrho(\eta_M)}{\varrho(x_m)} \frac{\psi_2(\varphi(x_m), \varphi(\eta_M))}{\psi_1(\varphi(x_m), \varphi(\eta_M))}. \tag{34}$$

Note that by our choice of φ,

$$\frac{\mathrm{d}}{\mathrm{d}x}\varphi(x) = \varrho(x) = \gamma x^{\gamma-1}.$$

Since Ψ is to be independent of γ, let $\gamma \downarrow 0$ and note that $\varphi(x_m) \to 1$ and $\varphi(\eta_M) \to 1$. After simplifying (34) in view of the specific form of φ and ϱ above, the limit produces the differential equation

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

for some constant β. Solving this differential equation yields

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

for some constant c'. Let $\alpha = c/c'$ to obtain

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

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

$$\Psi(x) = \frac{\alpha\, x_m}{(1 - x_M)^{\beta}}.$$

If $\Xi = \Psi$, the commutative diagram would be

$$\begin{array}{ccccc} x & \xrightarrow{\mathcal{F}} & x' & & \\ \Psi\big\downarrow & & \big\downarrow\Psi & & \\ \frac{\alpha x_m}{(1-x_M)^{\beta}} & \xrightarrow{\widetilde{\mathcal{F}}} & \frac{\alpha x'_m}{(1-x'_M)^{\beta}} & = & \frac{\alpha x_m^{\gamma}}{(1-x_M)^{\gamma\beta}} \end{array}$$

since $\mathcal{F}(x)_m = x_m^{\gamma}$ and $\mathcal{F}(x)_M = 1 - \eta_M^{\gamma}$. Commutativity of the diagram is validated by the quotient

$$\widetilde{\mathcal{F}}(x) = \alpha^{1-\gamma} x^{\gamma}.$$

9. Conclusion

Coarse graining is a pervasive concept in science, but has so far not been systematically investigated within the field of GAs. Whereas the phrase "coarse graining" has previously been used by other researchers in connection with GAs (most notably by Chris Stephens [1]) 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 [8], and by Muhlenbein and Voigt [6]. 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 grainings are dealt with. They are not guessed or noticed, to be pointed out and subsequently verified in an ad hoc manner. Instead, they are derived within a systematic general framework.

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 (Ranking Selection: see Section 8), a nonlinear coarse graining was computed by solving a differential equation.

We also advocate an alternative to ill advised notions of "coarse graining", [12] and propose that an "approximate coarse graining" of $H : X \to X$ should mean a strict coarse graining $\tilde{h}$ of some h which approximates H. In the situation where such h is not given, but a candidate Ξ and $\widetilde{h} : \Xi X \to \Xi X$ are known, defining h to map elements of $\Xi^{-1} \circ \Xi(x)$ to elements of $\Xi^{-1} \circ \tilde{h} \circ \Xi(x)$ (where Ξ^{-1} denotes inverse image under Ξ) yields a compatible $h : X \to X$ with coarse graining Ξ and quotient $\tilde{h}$ (i.e., diagram 1 commutes). If such h is deemed to approximate H, then $\widetilde{h}$ can be regarded as an approximate coarse graining of H.

In particular, injective fitness was assumed in Section 6 and throughout the remainder of the paper. It might be argued the assumption is both uncommon and a strong restriction. Let $H : X \to X$ be free of that restriction, and let h result from H by perturbing fitness values—by amounts less than δ—to ensure injective fitness. Given any N and any $\varepsilon > 0$, there exists some δ such that h approximates H within ε for every generation less than N. Therefore, quotients of h—which this paper is concerned with—are approximate coarse grainings of H. [13]

Acknowledgments

We would like to recognize the Dagstuhl Theory of Evolutionary Algorithms seminars and the Foundations of Genetic Algorithms workshops. Both have been a source of inspiration for this work. We would also like to thank Adam Prügel-Bennett for making comments leading toward a deeper understanding of the role played by paths in the general definition of equivalence (Section 3).

References

[1] A.A. Contreras, J.E. Rowe, C.R. Stephens, Coarse-graining in genetic algorithms: some issues and examples, in: E. Cantu-Paz (Ed.), Proc. GECCO 2003, Lecture Notes in Computer Science, Vols. 2723–2724, Springer, Berlin, 2003.

[2] D.E. Goldberg, Genetic Algorithms in Search Optimization and Machine Learning, Addison-Wesley, Reading, MA, 1989.

[3] A.A. Kirillov, Theorems and Problems in Functional Analysis, Springer, Berlin, 1982.

[4] H. von Koch: Une méthode géométrique élémentaire pour l'étude de certaines questions de la théorie des courbes plane, 1906.

[5] L.H. Loomis, S. Sternberg, Advanced Calculus, Addison-Wesley, Reading, 1968.

[6] H. Muhlenbein, H.M. Voigt, Gene pool recombination in genetic algorithms, in: I.H. Osman, J.P. Kelly (Eds.), Proc. Metaheuristics Internat. Conf., Norwell, Kluwer Academic Publishers, Dordrecht, 1995.

[7] C.W. Patty, Foundations of Topology, PWS Publishing Company, Boston, 1993.

[8] Y. Rabinovich, A. Wigderson, An analysis of a simple generic algorithm, in: R.K. Belew, L.B. Booker (Eds.), Proc. Fourth Internat. Conf. on Genetic Algorithms, Morgan Kaufmann Publishers, Los Altos, CA, 1991, pp. 215–221.

[9] J.E. Rowe, M.D. Vose, A.H. Wright, State aggregation and population dynamics in linear systems, Artificial Life 11 (4) (2005) 473–492.

[12] . . .if schemata are "coarse grained" by virtue of being definable in terms of subsets of Ω, then elements of Ω are "coarse grained" by virtue of being definable in terms of subsets of schemata. . .

[13] Approximation is not the only alternative; future work will discuss how exact results translate to the case of noninjective fitness.

[10] J.E. Rowe, M.D. Vose, A.H. Wright, Coarse graining selection and mutation, Foundations of Genetic Algorithms, Lecture Notes in Computer Science, Vol. 3469, Springer, Berlin, 2005, pp. 176–191.
[11] M. Toussaint, On the evolution of phenotypic exploration distributions, in: Cotta, DeJong, Poli, Rowe (Eds.), Foundations of Genetic Algorithms, Vol. 7, Morgan Kaufmann Publishers, Los Altos, CA, 2003.
[12] M.D. Vose, The Simple Genetic Algorithm: Foundations and Theory, MIT Press, Cambridge, 1999.