GENERICITY OF THE FIXED POINT SET FOR THE INFINITE POPULATION GENETIC ALGORITHM

TOMÁŠ GEDEON, CHRISTINA HAYES, AND RICHARD SWANSON

Abstract. The infinite population model for the genetic algorithm, where the iteration of the genetic algorithm corresponds to an iteration of a map G, is a discrete dynamical system. The map G is a composition of a selection operator and a mixing operator, where the latter models the effects of both mutation and crossover. This paper shows that for a typical mixing operator, the fixed point set of G is finite. That is, an arbitrarily small perturbation of the mixing operator will result in a map G with finitely many fixed points. Further, any sufficiently small perturbation of the mixing operator preserves the finiteness of the fixed point set.

1. Introduction

In this paper we study a dynamical systems model of the genetic algorithm (GA). This model was introduced by Vose [12] and is further extended in [6] and [9]. The dynamical systems model of the genetic algorithm provides an attractive mathematical framework for investigating the properties of GAs.

A practical implementation of the genetic algorithm seeks solutions in a finite search space which we denote $\Omega = \{1, 2, \ldots, n\}$. Each element of Ω can be thought of as a "species" with a given fitness value; the goal of the algorithm is to maximize the fitness. Usually there are multiple species with high fitness value and n is large. In order to avoid local maxima the GA algorithm uses mutation and crossover operations to maintain diversity in the pool of r individuals, representing the n species. The infinite population model considers an infinite population of individuals represented by the probability distribution over Ω,

$$P = (P_1, P_2, ..., P_n)$$

where P_i is the proportion of the i-th species in the population. An update of the genetic algorithm consists of mutation, selection and

1991 *Mathematics Subject Classification.* 37N99,68W40,68J05.

Key words and phrases. Genetic algorithm, generic property, fixed point, transversality.

crossover steps and is represented in the infinite population model as an iteration of a fixed function G.

Although the precise correspondence between behavior of such infinite population genetic algorithm and the behavior of the GA for finite population sizes has not been established in detail, the infinite population model has the advantage of being a well defined dynamical system. Therefore, the techniques of dynamical systems theory can be used to formulate and answer some fundamental questions about the GA.

The fixed points are fundamental objects of interest in our study, because in realistic implementations of genetic algorithms one usually observes convergence to a fixed point. Convergence to a unique fixed point is automatic when the quadratic map is a contraction. However, this paper is about a more subtle situation when the genetic algortithm is not a contraction. Wright and Bidwell [10] found examples of genetic algorithms with stable period two points, that is, examples of GAs for which not all solutions converge to a fixed point. These examples, however, were constructed using mutation and crossover operators which do not correspond to operators used in practice. In spite of this discovery, it is clear that one cannot expect to prove convergence to a fixed point for all genetic algorithms. Thus, the important problem is to carefully define the largest possible class of realistic genetic algorithms for which all solutions do converge to a fixed point. This problem is still open.

Wright and Vose [11] considered a class of mappings G that were a composition of a fixed mutation and crossover maps, and a proportional selection map. The set of fitness functions that correspond to the proportional selection was parameterized by the positive orthant. They have shown that for an open and dense set of such fitness functions, the corresponding map G has finitely many fixed points.

In this contribution we will take a different path. We consider a class of mappings $G = M \circ F$ where F is an arbitrary, but fixed, selection map and M is a mixing map from a class $\mathcal{M}$ described in Definition 2.2. This class is broad enough to include all mixing maps formed by a composition of the mutation and crossover maps described in monographs by Reeves and Rowe [6] and Vose [9]. We show that for a typical (i.e. open and dense) set of mixing maps, the corresponding map G has finitely many fixed points.

Theorem 1.1. *Let $G = M \circ F$ be a composition of a selection map F and a mixing operator M. Then for a typical mixing operator $M \in \mathcal{M}$, G has finitely many fixed points.*

The main tool in the proof of this result is the powerful notion of transversality. This idea has been successfully used in differential topology and dynamical systems for over forty years. Before we go into mathematical details, we wish to illustrate this notion on some simple examples. Consider a scalar function $f(x,a) := x^2 + a$, that depends on a real parameter a, and let W be the $x-$axis in the plane $\mathbb{R}^2$, see Figure 1. Notice that for all parameters $a \neq 0$ the graph of $f(x,a)$ either does not intersect W at all (for $a > 0$) or it intersects W in two points (for $a < 0$). Both of these situations are stable under small change in the parameter a. That is, if there are two intersections for $a = a_0$, then for all a_1 sufficiently close to a_0 the graph of $f(x, a_1)$ also intersects W in two points. A similar statement is true for no intersection. We observe that the set of these "stable" values of a (i.e. $a \neq 0$) is open and dense in the set of all a. The value $a = 0$ is exceptional since arbitrary small change in a changes the number of intersections with W.

FIGURE 1. An exceptional value of a corresponds to non-transverse intersection of the graph of f and W (left figure); typical value of a corresponds to transverse intersection.

We can characterize geometrically the "stable" values of a by observing that the sum of the tangent space to the graph of f and the tangent space to W at all points x of intersection generates the tangent space to $\mathbb{R}^2$ at x. This tangent space is again $\mathbb{R}^2$. In this case we will say that *f is transversal to W*. For the exceptional value $a = 0$ these two tangent spaces at the point of intersection $x = 0$ coincide and therefore their sum is a strict subspace of the tangent space of $\mathbb{R}^2$ at 0. This is a just a sophisticated way of saying that the graph of f and W "cross" for the "stable" values of a and only "touch" for the exceptional values of a. Notice that we have to define "crossing" loosely since empty intersection of the graph of f and W counts as "crossing".

The fundamental mathematical result we will use states that, under certain assumptions, the set of parameter values for which a general parameterized function intersects a given manifold W transversally, is both open and dense in the set of all parameters.

In proving our result, the essential step is to transform the problem of finiteness of the fixed point set to a transversality problem. We will set $f(x, \mathcal{M}) := M(F(x)) - x = G(x) - x$, where instead of a scalar parameter a we have multi-dimensional parameter space $\mathcal{M}$ that characterizes the set of all admissible mixing operators. If we let $W = \{0\}$, then the intersections of the graph of f and W correspond precisely to the fixed points of G. To prove finiteness of the set of fixed points, we evoke another general transversality theorem (Theorem 3.3) which states that this set has finitely many components for every M for which we have transversality.

The paper is organized as follows. In section 2 we carefully define the infinite population model, GA map and the set of admissible mixing operators $\mathcal{M}$. In section 3 we review transversality and provide the necessary background. In section 4 we prove the main result and conclude in section 5.

2. The Infinite Population Genetic Algorithm

The genetic algorithm searches for solutions in the search space $\Omega = \{1, 2, \ldots n\}$; each element of Ω can be thought of as a type of individual. We consider a total population of size r with $r << n$. We represent such a population as an *incidence vector*:

$$v = (v_1, v_2, ..., v_n)^T$$

where v_i is the number of times the individual of type i appears in the population. It follows that $\sum_i v_i = r$. We also identify a population with the *population incidence vector*

$$p = (p_1, p_2, ..., p_n)^T$$

where $p_i = \frac{v_i}{r}$ is the proportion of the i-th individual in the population. The vector p can be viewed as a probability distribution over Ω. In this representation, the iterations of the genetic algorithm yield a sequence of vectors $p \in \Lambda^r$ where

$$\Lambda^r := \{(p_1, p_2, ..., p_n)^T \in \mathbb{R}^n \mid p_i = \frac{v_i}{r} \text{ and } v_i \in \{0, \ldots, r\} \text{ for all } i \in \{1, \ldots, n\} \quad \text{with} \sum_i v_i = r\}. \tag{2.1}$$

We define

$$\Lambda := \{x \in \mathbb{R}^n | \sum x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i \in \{1, \ldots, n\}\}.$$

Note that $\Lambda^r \subset \Lambda \subset \mathbb{R}^n$, where Λ is the unit simplex in $\mathbb{R}^n$. Not every point $x \in \Lambda$ corresponds to a population incidence vector $p \in \Lambda^r$, with fixed population size r, since p has non-negative rational entries with denominator r. However, as the population size r gets arbitrarily large, Λ^r "becomes dense" in Λ, that is, $\cup_{r \geq N} \Lambda^r$ is dense in Λ for all N. Thus Λ may be viewed as a set of admissible states for infinite populations. We will use p to denote an arbitrary point in Λ^r and x to denote an arbitrary point in Λ. Thus p always represents a population incidence vector in a finite population and x the corresponding quantity in infinite population, which is the probability distribution over Ω. Unless otherwise indicated, $x \in \Lambda$ is a column vector.

Let $G(x)$ represent the action of the genetic algorithm on $x \in \Lambda$, and assume that $G : \Lambda \rightarrow \Lambda$ is a differentiable map ([9]). The map G is a composition of three maps: selection, mutation, and crossover. We will now describe each of these in turn.

We let $F : \Lambda \rightarrow \Lambda$ represent the selection operator. The i-th component, $F_i(x)$, represents the probability that an individual of type i will result if selection is applied to $x \in \Lambda$. As an example, consider proportional selection where the probability of an individual $k \in \Omega$ being selected is

$$Pr[k|x] = \frac{x_k f_k}{\sum_{j \in \Omega} x_j f_j},$$

where $x \in \Lambda$ is the population incidence vector, and f_k, the k-th entry of the vector f, is the fitness of $k \in \Omega$. Define $diag(f)$ as the diagonal matrix with entries from f along the diagonal and zeros elsewhere. Then, for $F : \Lambda \rightarrow \Lambda$, proportional selection is defined as

$$F(x) = \frac{diag(f)x}{f^T x}.$$

We restrict our choice of selection operators, F, to those which are $\mathcal{C}^1$, that is, selection operators with continuous derivatives.

We let $U : \Lambda \rightarrow \Lambda$ represent mutation. Here U is an $n \times n$ real valued matrix with ij-th entry $u_{ij} > 0$ for all i, j, and where U_{ij} represents the probability that item $j \in \Omega$ mutates into $i \in \Omega$. That is, $(Ux)_k := \sum_i u_{ki} x_i$ is the probability an individual of type k will result after applying mutation to population x.

Let crossover, $C : \Lambda \rightarrow \Lambda$, be defined by

$$C(x) = (x^T C_1 x, \ldots, x^T C_n x)$$

for $x \in \Lambda$, where $C_1, \ldots, C_n$ is a sequence of symmetric non-negative $n \times n$ real valued matrices. Here $C_k(x)$ represents the probability that an individual k is created by applying crossover to population x.

Definition 2.1. Let $Mat_n(\mathbb{R})$ represent the set of $n \times n$ matrices with real valued entries. An operator $A : \mathbb{R}^n \to \mathbb{R}^n$ is *quadratic* if there exist matrices $A_1, A_2, \ldots, A_n \in Mat_n(\mathbb{R})$ such that $A(x) = (x^T A_1 x, \ldots, x^T A_n x)$. We denote a quadratic operator with its corresponding matrices as $A = (A_1, \ldots, A_n)$.

Thus, the crossover operator, $C = (C_1, \ldots, C_n)$, is a quadratic operator ([8]).

We combine mutation and crossover to obtain the mixing operator $M := C \circ U$. The k-th component of the mixing operator

$$M_k(x) = x^T (U^T C_k U) x$$

represents the probability that an individual of type k will result after applying mutation and crossover to population x. Since C_k is symmetric, M_k is symmetric. Further, since C_k is non-negative and U is positive for all k, M_k is also positive for all k. Additionally, it is easy to see check that since $\sum_{k=1}^{n} [M_k]_{ij} = 1$, $M : \Lambda \to \Lambda$, and mixing is also a quadratic operator ([8]). Here $[M_k]_{ij}$ denotes the ij-th entry of the matrix M_k. This motivates the following general definition of a mixing operator.

Definition 2.2. Let $Mat_n(\mathbb{R})$ represent the set of $n \times n$ matrices with real valued entries. We call a quadratic operator, $M = (M_1, \ldots, M_n)$, a *mixing operator* if the following properties hold:

(1) $M_k \in Mat_n(\mathbb{R})$ is symmetric for all $k = 1, \ldots, n$;
(2) $(M_k)_{ij} > 0$ for all $i, j \in \{1, \ldots, n\}$, and for all $k = 1, \ldots, n$;
(3) $\sum_{k=1}^{n} [M_k]_{ij} = 1$ for all $j = 1, \ldots, n$ and $i = 1, \ldots, n$.

Let $\mathcal{M}$ be the set of quadratic operators M satisfying (1)-(3). It is easy to see that (3) implies that $M \in \mathcal{M}$ maps Λ to Λ. We define a norm, $||\cdot||$, on $\mathcal{M}$ by considering for $M \in \mathcal{M}$, $M \in \mathbb{R}^{n^3}$, and using the Euclidean norm. For an alternative norm on the set of quadratic operators, see [7].

Definition 2.3. We set

$$G := M \circ F, \text{ for } M \in \mathcal{M} \tag{2.2}$$

to be the complete operator for the genetic algorithm, or a *GA map*.

We extend the domain of definition of F to the positive cone in $\mathbb{R}^n$, denoted $\mathbb{R}^{n+}$. The extension of F is denoted $\tilde{F}$ and is defined by

$$\tilde{F}(u) := F\left(\frac{u}{\sum_i u_i}\right).$$

Thus $\tilde{F}|_\Lambda = F$, and for $x \in \Lambda$, $D\tilde{F}(x)|_\Lambda = DF(x)$, the Jacobian of F. Let $\mathbb{R}^n_0 := \{x \in \mathbb{R}^n | \sum_i x_i = 0\}$. Since $\mathbb{R}^n_0$ is the tangent space to Λ at x and $F(\Lambda) \subset \Lambda$, $D\tilde{F}(\mathbb{R}^n_0) \subseteq \mathbb{R}^n_0$. Because $\tilde{F} : \mathbb{R}^{n+} \to \Lambda$, it is also clear that the map G is extended to a map $\tilde{G} : \mathbb{R}^{n+} \to \Lambda$ and the preceding remarks apply to $\tilde{G}$ as well. In order to simplify the notation we will use symbols F and G for these extended functions.

Definition 2.4. If $f(x) = x$, a point x is called a fixed point of f.

Definition 2.5. A property is *typical*, or *generic*, in a set S, if it holds for an open and dense set of parameter values in S.

3. Transversality: Background and Terminology

This section provides the reader with the necessary background in differential topology. The material provided below follows [1]. Other references include [2], [3] and [5].

Let X, Y be n and m dimensional manifolds, respectively. For $x \in X$, we let T_xX denote the tangent space to X at x. For a differentiable map $f : X \to Y$, we let $df_x : T_xX \to T_{f(x)}Y$ denote the derivative of the map f at the point x. In the special case that $T_xX = \mathbb{R}^n$ and $T_{f(x)}Y = \mathbb{R}^m$, we note that $df_x : \mathbb{R}^n \to \mathbb{R}^m$ is given by the Jacobian matrix $Df(x)$ ([4]).

The following notation is adopted from [1]. Let $\mathcal{A}, X, Y$ be $\mathcal{C}^r$ manifolds. Let $\mathcal{C}^r(X, Y)$ be the set of $\mathcal{C}^r$ maps from X to Y.

Definition 3.1 (Restricted definition from [1] for finite vector spaces)**.** Let X and Y be $\mathcal{C}^1$ manifolds, $f : X \to Y$ a $\mathcal{C}^1$ map, and $W \subset Y$ a submanifold. We say that f *is transversal to* W *at a point* $x \in X$, in symbols: $f \pitchfork_x W$, if, where $y = f(x)$, either $y \notin W$ or $y \in W$ and

$$(T_xf)(T_xX) + T_yW = T_yY.$$

We say f *is transversal to* W, in symbols: $f \pitchfork W$, if and only $f \pitchfork_x W$ for every $x \in X$.

We reformulate the example in Figure 1 in terms of Definition 3.1. Let $X := [-1, 1], Y := \mathbb{R}^2$ and let $W := \{(x, y) \in \mathbb{R}^2 \mid y = 0\}$ be the x-axis. We define the the family of maps $f_a : [-1, 1] \to \mathbb{R}^2$ by $f_a(t) := (t, t^2 + a)$. Then the map f_{-1} is transversal to W (Figure 1 right), but f_0 is not transversal to W (Figure 1 left).

Theorem 3.2 (Transversal Density Theorem, [1]). *Let $\mathcal{A}, X, Y$ be C^r manifolds and $W \subset Y$ a submanifold. Let $\rho_a : X \to Y$ be a family of maps such that the correspondence $ev(a,x) := \rho_a(x)$ is C^r. Define $\mathcal{A}_W \subset \mathcal{A}$ by*

$$\mathcal{A}_W = \{a \in \mathcal{A} | \rho_a \pitchfork W\}.$$

Assume that

(1) X has finite dimension m and W has finite codimension q in Y;
(2) $r > \max(0, m-q)$;
(3) $ev(a,x) \pitchfork W$.

Then $\mathcal{A}_W$ is residual (and hence dense) in $\mathcal{A}$.

In our example, the set $\mathcal{A}_W := \{a \neq 0\}$ and the set $\mathcal{A} = \mathbb{R}$. Density of $\mathcal{A}_W$ in $\mathcal{A}$ means that in an arbitrary neighborhood of the value $a = 0$, there is a value of the set $\mathcal{A}_W$.

Theorem 3.3 (Corollary 17.2, [1]). *Let X, Y be C^r manifolds ($r \geq 1$), $f : X \to Y$ a C^r map, $W \subset Y$ a C^r submanifold. Then if $f \pitchfork W$:*

(1) W and $f^{-1}(W)$ have the same codimension;
(2) If W is closed and X is compact, $f^{-1}(\{W\})$ has only finitely many connected components.

As an example, let $f_a : [-1,1] \to \mathbb{R}^2$ be as defined above with $a \neq 0$. Since $W = \{(x,y) \in \mathbb{R}^2 | y = 0\}$, the set $f_a^{-1}(W)$ consists of finitely many points. In fact, this set consists of two points when $a < 0$ and no points when $a > 0$. Each of these points is a zero-dimensional submanifold of the dimension 1 interval $[-1,1]$. Thus, $f_a^{-1}(W)$ has codimension 1. Since W is a 1-dimensional submanifold of $\mathbb{R}^2$, it also has codimension 1. Note that in our example, $f_0^{-1}(W) = 0$ and is a point. However, we can modify the family f_a in such a way that $f_0(-\epsilon, \epsilon) = 0$ and 0 is still the only value of a where we lack hyperbolicity, see Figure 2. This example shows that hyperbolicity is necessary for the conclusion of the Theorem 3.3.

Observe that the compactness of X is also necessary. Consider the function $g : \mathbb{R} \to \mathbb{R}^2$ defined by $g(t) = (t, \sin t)$. Then $g^{-1}(0) = k\pi$ has infinitely many components in $X = \mathbb{R}$.

Theorem 3.4 (Openness of Transversal Intersection, [1]). *Let $\mathcal{A}, X, Y$ be C^1 manifolds with X finite dimensional, $W \subset Y$ a closed C^1 submanifold, $K \subset X$ a compact subset of X, and $\rho_a : X \to Y$ be a family of maps such that the correspondence $ev(a,x) = \rho_a(x)$ is C^r. Then the subset $\mathcal{A}_{KW} \subset \mathcal{A}$ defined by $\mathcal{A}_{KW} = \{a \in \mathcal{A} | \rho_a \pitchfork_x W \text{ for } x \in K\}$ is open.*

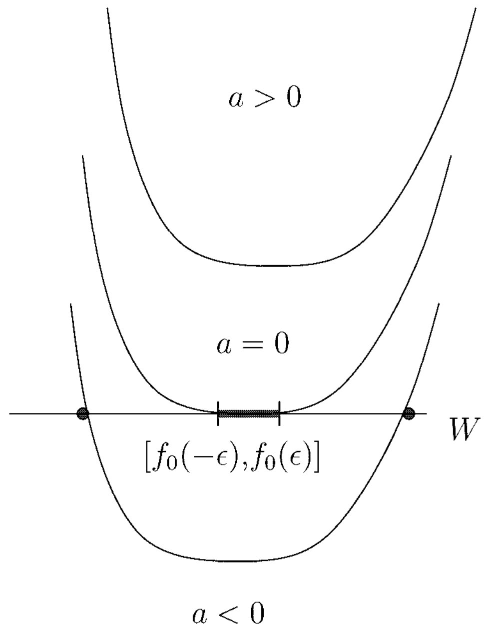

FIGURE 2. A modified example from Figure 1. The map f_0 maps a subinterval $(-\epsilon, \epsilon)$ into W.

In our example, $K = X = [-1, 1]$ is compact and $\mathcal{A}_{KW} = \mathcal{A}_W := \{a \neq 0\}$. The set $\mathcal{A}_{KW}$ is open in $\mathcal{A} = \mathbb{R}$ if a small enough neighborhood, $N_\epsilon(a)$, of the value $a \in \mathcal{A}_{KW}$, has $N_\epsilon(a) \subset \mathcal{A}$. That is, for any $a \neq 0$, there is a small interval $(a - \epsilon, a + \epsilon) \subset \mathcal{A}$.

4. Proof of Main Results

Let $\mathcal{A} = \mathcal{M}$, where $\mathcal{M}$ denotes the set of mixing operators given by Definition 2.2, and let $X = \Lambda \subset \mathbb{R}^n$ and $Y = \mathbb{R}^n_0$. For $M \in \mathcal{M}$, we define a family of maps $\rho_M(x) : \Lambda \to \mathbb{R}^n_0$ by $\rho_M(x) := (M \circ F - I)x$. Recall that $\mathbb{R}^n_0 := \{x \in \mathbb{R}^n | \sum_i x_i = 0\}$. Note that because $F, M \in \mathcal{C}^1(\Lambda, \Lambda)$,we also have $\rho_M \in \mathcal{C}^1(\Lambda, \mathbb{R}^n_0)$.

Finally, we define $\text{ev}_\rho(M, x) : \mathcal{M} \times \Lambda \to \mathbb{R}^n_0$ by $\text{ev}_\rho(M, x) := \rho_M(x)$ for $M \in \mathcal{M}$ and $x \in \Lambda$. That is,

$$\text{ev}_\rho(M, x) := \rho_M(x) = M(F(x)) - x.$$

Finally, note that since G, F are $\mathcal{C}^1$, the function ev_ρ is also $\mathcal{C}^1$.

Lemma 4.1. *For* $\text{ev}_\rho(M, x) := M(F(x)) - x$, $\text{rank}(d(\text{ev}_\rho)) = n - 1$.

Proof. Note first that since $\Lambda \subset \mathbb{R}^{n+}$, the derivative $d\ \text{ev}_\rho$ of ev_ρ, $d\ \text{ev}_\rho = D\ \text{ev}_\rho$, is a Jacobian of ev_ρ. Similarly, we note that

$$D(\text{ev}_\rho|_{(\mathcal{M}\times\Lambda)}) = D\text{ev}_\rho|_{T_{(P,y)}(\mathcal{M}\times\Lambda)}.$$

Because $T(\mathbb{R}^n_0) = \mathbb{R}^n_0$, and

$$T(\mathcal{M} \times \Lambda) = \{(P, y) | P = (P_1, \ldots, P_n) \text{ with } \sum_i P_i = 0 \text{ and } y \in \mathbb{R}^n\},$$

it suffices to show that the Jacobian $D\ \mathrm{ev}_\rho : T(\mathcal{M} \times \Lambda) \to \mathbb{R}^n_0$ is onto. Thus, for $(P, y) \in T(\mathcal{M} \times \Lambda)$, we calculate

$$D\ \mathrm{ev}_{(M,x)}(P, y) = \left[\frac{\partial \mathrm{ev}_\rho}{\partial M}, \frac{\partial \mathrm{ev}_\rho}{\partial x}\right] \begin{pmatrix} P \\ y \end{pmatrix} = \frac{\partial \mathrm{ev}_\rho}{\partial M} P + \frac{\partial \mathrm{ev}_\rho}{\partial x} y$$

By a short computation we get

$$\frac{\partial \mathrm{ev}_\rho}{\partial M} P = PF(x),$$

and

$$\frac{\partial \mathrm{ev}_\rho}{\partial x} y = 2(MF(x))DF(x)y - y.$$

Finally, for any $z \in \mathbb{R}^n_0$, and $(M, x) \in \mathcal{M} \times \Lambda$, we show there exists $(P, y) \in T(\mathcal{M} \times \Lambda)$ such that

$$D\mathrm{ev}_{(M,x)}(P, y) = PF(x) + 2(MF(x))DF(x)y - y = z. \tag{4.1}$$

We start by choosing $y = 0 \in \mathbb{R}^n$. Now, by equation (4.1), it suffices to find $P = (P_1, \ldots, P_n)$ such that

$$D\mathrm{ev}_{(M,x)}(P, y) = PF(x) + 0 - 0 = z. \tag{4.2}$$

Because $F : \mathbb{R}^{n+} \to \Lambda$, we let $u = F(x) \in \Lambda$. By equation (4.2), we see that for fixed $u \in \Lambda$, we want $P = (P_1, \ldots, P_n)$ such that $Pu = (u^T P_1 u, \ldots, u^T P_n u) = z$. Clearly, for $i = 1, \ldots, n-1$, we can choose P_i such that $u^T P_i u = z_i$. Finally, because $z \in \mathbb{R}^n_0$, that is $\sum_i z_i = 0$, we see that $z_n = -\sum_i z_i$. Thus, for our choice of $P_1, \ldots, P_{n-1}$,

$$z_n = -\sum_i z_i = -\sum_i u^T P_i u = u^T(\sum_i P_i)u,$$

and $P_n = -\sum_{i=1}^{n-1} P_i$. Because $P_n = -\sum_{i=1}^{n-1} P_i$, it is clear that $\sum_{i=1}^n P_i = 0$, and for this choice of $P = (P_1, \ldots, P_n)$ with $y = 0$ we have $D\mathrm{ev}_{(M,x)}(P, y) = z$. Because z, M, x were arbitrary, we have shown that $D\mathrm{ev}_\rho : T(\mathcal{M} \times \Lambda) \to \mathbb{R}^n_0$ is onto. That is, $\mathrm{rank}(D\mathrm{ev}_\rho) = n - 1$.

□

Lemma 4.2. *Let* $\mathrm{ev}_\rho : \mathcal{M} \times \Lambda \to \mathbb{R}^n_0$, $x \in \Lambda$, *and* $M \in \mathcal{M}$. *Then* $\mathrm{ev}_\rho \pitchfork \{0\}$.

Proof. Choose $W := \{0\}$. Since $T_0 W = \{0\}$, to prove transversality using Definition 3.1, we need to show that the image $(T_x \mathrm{ev}_\rho)(T_x X) =$

$T_0Y = \mathbb{R}^{n-1}$. In other words, we need to show that $Dev_\rho(x)$ is surjective. By Lemma 4.1, $\text{rank}(Dev_\rho(x)) = n-1$, and therefore $Dev_\rho(x)$ is surjective and $\text{ev}_\rho \pitchfork \{0\}$. □

Proposition 4.3. *Let $\mathcal{M}_{\{0\}} := \{M \in \mathcal{M} | \rho_M \pitchfork \{0\}\}$. Then $\mathcal{M}_{\{0\}}$ is dense in $\mathcal{M}$. That is, the set of parameter values for which ρ_M is transversal to $\{0\}$ is dense in $\mathcal{M}$.*

Proof. We apply the Transversal Density Theorem: Theorem 3.2. We first note that by Lemma 4.2, $\text{ev}_\rho \pitchfork \{0\}$, and therefore condition (3) of Theorem 3.2 holds. We now verify the remaining conditions (1)-(2).

(1) $X = \Lambda$ has finite dimension m and $W = \{0\}$ has finite codimension q in $Y = \mathbb{R}^n_0$. Because $X = \Lambda$, $m = n-1 < \infty$. Clearly, the codimension of $\{0\}$ in $\mathbb{R}^n_0$ is $n-1$. That is, $q = n-1$.
(2) $r > \max(0, m-q)$. Since $r = 1$, clearly $r > \max(0,0) = 0$.

□

Proposition 4.4. *The set $\mathcal{M}_{\{0\}}$ is open in $\mathcal{M}$.*

Proof. We apply Theorem 3.4, and therefore start by verifying its hypothesis conditions. Clearly $\mathcal{M}, X, Y$ are $\mathcal{C}^1$ manifolds. We take $K = X = \Lambda$, and thus K is a compact subset of the finite dimensional manifold X. Similarly $W = \{0\} \subset Y$ is closed. By the previous argument, the maps $\rho_M(x)$ are C^1. Thus, all hypothesis requirements have been met and by Theorem 3.4,

$$\mathcal{M}_{K\{0\}} = \{M \in \mathcal{M} | \rho_M \pitchfork_x \{0\} \text{ for } x \in K = X\}$$

is open in $\mathcal{M}$. □

Proposition 4.5. *For generic $M \in \mathcal{M}$,*

(1) $\rho_M \pitchfork \{0\}$. That is, the set of parameter values for which ρ_M is transversal to $\{0\}$ is open and dense in $\mathcal{M}$.
(2) The set of parameter values for which $\rho_M^{-1}(\{0\})$ has finitely many solutions is open and dense in $\mathcal{M}$.

Proof. The proof of part (1) follows directly from Lemmas 4.3 and 4.4.

We now prove part (2). By part (1), the set of parameter values for which ρ_M is transversal to $\{0\}$ is open and dense in $\mathcal{M}$. Thus by Theorem 3.3, for this open and dense set, $\rho_M^{-1}(\{0\})$ has only finitely many connected components. We now show by contradiction that there are finitely many solutions to $\rho_M(x) = 0$ in Λ.

For $x \in \rho_M^{-1}(\{0\})$, let $M_x \subset \rho_M^{-1}(\{0\})$ denote the connected component with $x \in M_x$. Assume x is not isolated in $\rho_M^{-1}(\{0\})$. Then, there exists a sequence $\{x_n\} \subset M_x$ such that $x_n \to x$, and by choosing a subsequence $\{x_{n_k}\}$,

$$\lim_{n_k \to \infty} \frac{x_{n_k} - x}{||x_{n_k} - x||} = v, \tag{4.3}$$

where $v \in T_x(\rho_M^{-1}(\{0\}))$. Here $v \neq 0$ because the terms in the quotient are on the unit sphere. Since

$$M \circ F(x_{n_k}) - M \circ F(x) = D(M \circ F)(x) \cdot (x_{n_k} - x) + R \tag{4.4}$$

where R is a remainder, then

$$\lim_{n_k \to \infty} \frac{M \circ F(x_{n_k}) - M \circ F(x)}{||x_{n_k} - x||} = \lim_{n_k \to \infty} \frac{D(M \circ F)(x) \cdot (x_{n_k} - x) + R}{||x_{n_k} - x||}. \tag{4.5}$$

By equations (4.3) and (4.5), and because x_{n_k}, x are fixed points,

$$v = D(M \circ F)(x) \cdot \left(\lim_{n_k \to \infty} \frac{x_{n_k} - x}{||x_{n_k} - x||} \right) + \lim_{n_k \to \infty} \frac{R}{||x_{n_k} - x||}. \tag{4.6}$$

That is,

$$v = D(M \circ F)(x) \cdot v + 0, \tag{4.7}$$

and $v \neq 0$ is an eigenvector of $D(M \circ F)(x)$ with eigenvalue 1. However, since $x \in \rho_M^{-1}(\{0\}$, and ρ_M is transversal to $\{0\}$ at x, $D(M \circ F)(x) - I$ is a linear isomorphism on a finite vector space. Thus, for all $v \neq 0$, $(D(M \circ F)(x) - I)v \neq 0$ which is a contradiction. Thus, all the components of $\rho_M^{-1}(\{0\})$ only contain isolated points and each connected component of $\rho_M^{-1}(\{0\})$ is itself an isolated point. Since there are finitely many connected components of $\rho_M^{-1}(\{0\})$, there are finitely many solutions to $\rho_M(x) = 0$ for $x \in \Lambda$. □

Proof of Theorem 1.1. By Lemma 4.5, for $M \in \mathcal{M}_{\{0\}} \subset \mathcal{M}$, $\rho_M(x) = 0$ has finitely many solutions in Λ. That is, for generic $M \in \mathcal{M}$,

$$\rho_M(x) = M(F(x)) - x$$

has finitely many solutions in Λ. Thus solutions to $\rho_M(x) = 0$ correspond to fixed points of $G = M \circ F$.

5. Conclusions

In this paper we have shown that given an arbitrary selection function and a typical mixing function, their composition has finitely many fixed points. This composition represents an infinite population model of a GA. Even though the correspondence between the infinite population model of a GA and the finite population models that are used by practitioners is not straightforward and likely depends on the details of that implementation, our result adds to the increasing body of evidence that the infinite population model can give qualitative insights into the functioning of the GA. Genericity of the finiteness of the fixed point set is expected for any reasonably rich model, consisting of iterations of a map on a compact space. Our results can be interpreted as showing that the GA map and the class of mixing operators constitute such model. That is, as analogously stated in [11], for a given mixing operator, unless proven otherwise, it is reasonable to assume that G has finitely many fixed points. We note that the perturbation from an infinite to a large finite population model can be viewed as a small perturbation of the infinite population model. Therefore, for any finite population model which in the infinite population limit falls into the class we study in this paper, one can reasonably assume that each large population model in this class has finitely many fixed points. There is, however, a caveat behind the words "reasonably assume". It is generally more difficult to prove density results in a smaller parameter set as opposed to in a larger parameter set. This difficulty arises for the simple reason that in the smaller parameter set, there are fewer available perturbations to perturb off of "bad" parameter values (as is $a = 0$ in our example). Therefore, it is possible that there may be a specific class of mixing operators in which genetic algorithms with a finite number of fixed points will not be dense. Whether, for instance, the widely used crossover bit string operator for finite population gives rise to such a class, remains on open question and will be addressed in our future work.

The additional contribution of our work is to once again (after [11]) bring the attention of the GA community to a set of powerful ideas from differential topology, centered around the notion of transversality. We believe that these ideas can be applied in different contexts to many problems in the study of genetic algorithms.

References

1. Ralph Abraham and Joel Robbin, *Transversal mappings and flows*, W. A. Benjamin, Inc., 1967.
2. Victor Guillemin and Alan Pollack, *Differential topology*, Prentice-Hall, Inc., 1974.
3. Morris Hirsch, *Differential topology*, Springer-Verlag, 1976.
4. Klaus Jänich, *Vector analysis*, Springer-Verlag, 2001.
5. Jacob Palis Jr. and Welington de Melo, *Geometric theory of dynamical systems*, Springer-Verlag, 1982.
6. Colin R. Reeves and Jonathan E. Rowe, *Genetic algorithms - principles and perspectives: A guide to* GA *theory*, Kluwer Academic Publishers, 2003.

7. Jonathan E. Rowe, *A normed space of genetic operators with applications to scalability issues.*, Evolutionary Computation **9** (2001), no. 1.
8. Jonathan E. Rowe, Michael D. Vose, and Alden H. Wright, *Group properties of crossover and mutation*, Evolutionary Computation **10** (2002), no. 2, 151–184.
9. Michael D. Vose, *The simple genetic algorithm: Foundations and theory*, The MIT Press, 1999.
10. Alden H. Wright and Garrett Bidwell, *A search for counterexamples to two conjecture on the simple genetic algorithm*, Foundations of Genetic Algorithms 4, Morgan Kaufman Publishers **4** (1997).
11. Alden H. Wright and Michael D. Vose, *Finiteness of the fixed point set for the simple genetic algorithm*, Evolutionary Computation **3** (1995), no. 4.
12. ———, *Stability of vertex fixed points and applications*, Foundations of Genetic Algorithms 3, Morgan Kaufman Publishers **3** (1995).

Department of Mathematical Sciences, Montana State University, Bozeman, MT 59715
E-mail address: gedeon@math.montana.edu

Department of Mathematics, Gettysburg College, Gettysburg, PA 17325
E-mail address: chayes@gettysburg.edu

Department of Mathematical Sciences, Montana State University, Bozeman, MT 59715
E-mail address: swanson@math.montana.edu