

Hyperbolic Fixed Points are Typical in the Space of Mixing Operators for the Infinite Population Genetic Algorithm

Christina Hayes
Department of Mathematical Sciences
Montana State University
Bozeman, Montana 59717-2400
hayes@math.montana.edu

Tomáš Gedeon^{*}
Department of Mathematical Sciences
Montana State University
Bozeman, Montana 59717-2400
gedeon@math.montana.edu

ABSTRACT

We study an infinite population model for the genetic algorithm, where the iteration of the algorithm corresponds to an iteration of a map G . The map G is a composition of a selection operator and a mixing operator, where the latter models effects of both mutation and crossover. We examine the hyperbolicity of fixed points of this model. We show that for a typical mixing operator all the fixed points are hyperbolic.

Categories and Subject Descriptors

I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, And Search - Genetic Algorithms

General Terms

Algorithms, Artificial Intelligence

Keywords

Genetic algorithm, mixing, generic, typical, hyperbolic fixed point.

1. INTRODUCTION

In this paper we consider a dynamical systems model of the genetic algorithm (GA). This model was introduced by Vose (see [10]) by replacing finite population by population densities modelling an infinite population. The model is further extended in [3], [6], [7], and [9]. Although the precise correspondence between behavior of such infinite population Genetic Algorithm and the behavior of the GA for finite population has not been established in detail, the infinite population model has the advantage of being a well

^{*}This work was partially supported by NSF EIA-BITS grant 426411.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'05, June 25–29, 2005, Washington, DC, USA.
Copyright 2005 ACM 1-59593-097-3/05/0006 ...\$5.00.

defined dynamical system. Therefore, the techniques of dynamical systems theory can be used to formulate and hopefully answer some fundamental questions about the GA. One such question is the question of convergence. For plausible crossover, mutation and selection¹, does the algorithm always converge to a unique solution for all initial states? In the infinite population model, the iterations of the GA are represented as iterations of a fixed map G on a space of admissible population densities p . Thus, the question of convergence can be reformulated in this setting as existence of a globally attracting stable fixed point, that is, a population p_0 such that $G(p_0) = p_0$.

The fixed points, p_0 such that $G(p_0) = p_0$, are fundamental objects of interest in our study. The behavior of the map G in the neighborhood of p_0 is determined by the eigenvalues of the linearization $DG(p_0)$. If all the eigenvalues have absolute value less than one, then all iterates starting near p_0 converge to p_0 . If there is at least one eigenvalue with absolute value greater than one, then almost all iterates will diverge from p_0 [4]. Such classification is, however, possible only if no eigenvalues lie on the unit circle in the complex plane. Fixed points p_0 , for which $DG(p_0)$ has this property, are called *hyperbolic*. If at least one eigenvalue of $DG(p_0)$ has modulus 1, the fixed point is *non-hyperbolic*.

It is easy to see that hyperbolicity is an open condition, i.e. if a fixed point is hyperbolic, then all small perturbations of the map G will still admit a fixed point with eigenvalues off the unit circle. It follows that for sufficiently large finite population, the GA will also admit a fixed point. Thus, hyperbolic fixed points under G predict behavior for finite population GA.

On the other hand, non-hyperbolic fixed points can disappear under arbitrarily small perturbations. If the infinite population model wants to be a viable model of the behavior of the finite population GA, non-hyperbolic fixed points should be rare. It is clear that they must be present for some admissible maps G , since they occur when a fixed point bifurcates. Vose and Eberlein [6] considered a class of mappings G that were a composition of a mutation and crossover operator, with proportional selection scheme. The set of fitness functions was parameterized by the positive orthant. They have shown that for an open and dense set of such fitness functions, the corresponding operator, G , has

¹Wright and Bidwell [8] found counterexamples for mutation and selection corresponding to distributions that are not used in practice.

hyperbolic fixed points.

In this contribution we consider a class of mappings $G = M \circ F$ where F is arbitrary, but fixed, selection operator and M is a mixing operator from a class described in section 2. The class of mixing operators we consider include all mixing operators that are a composition of the mutation a selection operators as described in Reeves and Rowe [3] and Vose [6]. We show that for an open and dense set of mixing operators, the corresponding operator G has hyperbolic fixed points.

Due to length limitations we will not provide complete proofs. Instead, we give outlines of proofs to introduce the reader to the main structure of the arguments involved. For more details please see [1]. We start with a preliminary section introducing notation and the specifics of the map G .

2. PRELIMINARIES

The dynamical systems model of the genetic algorithm provides an attractive mathematical framework for investigating the properties of GAs. In this paper we study the model introduced by Vose [6].

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

$$v = (v_1, v_2, \dots, v_n)$$

where v_i is the number of times the species i appears in the population. It follows that $\sum_i v_i = r$. We associate each population with a probability distribution over Ω . That is, a population is identified with the population incidence vector

$$p = (p_1, p_2, \dots, p_n)$$

where $p_i = \frac{v_i}{r}$ is the proportion of the i -th species in the population. In this representation, the iterations of the genetic algorithm yield a sequence of vectors $p \in \Lambda$ where

$$\Lambda = \{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mid \sum x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i = 1, \dots, n\}.$$

Note that $\Lambda \subset \mathbb{R}^n$ is the unit simplex in \mathbb{R}^n . Not every point $x \in \Lambda$ corresponds to a population incidence vector p , since these have non-negative rational entries with denominator r . However, as the population size r gets arbitrarily large, population incidence vectors become dense in the simplex. Thus Λ may be viewed as a set of admissible states for infinite populations.

Let $G(p)$ represent the action of the genetic algorithm on $p \in \Lambda$, where $G : \Lambda \mapsto \Lambda$ is a differentiable map ([6]). The map G is a composition of three operators: selection, mutation, and crossover. We will now describe each of these in turn.

We let $F : \Lambda \mapsto \Lambda$ represent the selection operator. There are many possible models for the selection operator and our results do not depend on this choice of selection operator. Possible choices include proportional, tournament, or rank selection (see [6]). The i -th component, $F_i(p)$ represents the probability that an individual of type i will result if selection is applied to $p \in \Lambda$. We extend the domain of definition of F to the positive orthant in \mathbb{R}^n , \mathbb{R}^{n+} and define $F : \mathbb{R}^{n+} \mapsto \Lambda$ by

$$F(x) := \frac{F(x)}{\sum_i F_i(x)}.$$

We let $U : \Lambda \mapsto \Lambda$ represent mutation with a positive mutation rate. Here U is a matrix with $U_{ij} > 0$ for all i, j , where U_{ij} represents the probability that item $j \in \Omega$ mutates into $i \in \Omega$. That is, $U_k(p)$ is the probability an individual of type k will result after applying mutation to population p .

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

$$C(p) = (p^T C_1 p, \dots, p^T C_n p)$$

for $p \in \Lambda$, where C_1, \dots, C_n is a sequence of symmetric non-negative matrices. Here $C_k(p)$ represents the probability that an individual k is created by applying crossover to population p . Recall that an operator $A : \mathbb{R}^n \mapsto \mathbb{R}^n$ is quadratic if there exist matrices A_1, A_2, \dots, A_n such that $A(x) = (x^T A_1 x, \dots, x^T A_n x)$. We denote a quadratic operator with its corresponding matrices as $A = (A_1, \dots, A_n)$. Thus $C = (C_1, \dots, C_n)$ is a quadratic operator ([5]).

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

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

represents the probability that an individual of type k will result after applying mutation and crossover to population p . Observe that mixing is also a quadratic operator from Λ to Λ ([5]). This motivates the definition of a mixing operator. Let $\mathcal{A}^{n \times n}$ represent the set of $n \times n$ matrices with real valued entries. We call a quadratic operator, $M = (M_1, \dots, M_n)$, a *mixing operator* if the following properties hold:

1. $M_k \in \mathcal{A}^{n \times n}$ is symmetric for all $k = 1, \dots, n$;
2. $M_k > 0$ for all $k = 1, \dots, n$;
3. $\sum_{k=1}^n [M_k]_{ij} = 1$ for all $j = 1, \dots, n$ and $i = 1, \dots, n$.

Let \mathcal{M} be the set of quadratic operators M satisfying (1)-(3). Observe that $M \in \mathcal{M}$ maps Λ to Λ . This is easily seen since, for $x \in \Lambda$, $M(x) = (x^T M_1 x, \dots, x^T M_n x)$, and

$$\begin{aligned} \sum_k [M(x)]_k &= x^T \left(\sum_k M_k \right) x = x^T \cdot \left(\sum x_i, \sum x_i, \dots, \sum x_i \right) \\ &= x^T \cdot (1, \dots, 1) = 1. \end{aligned}$$

Finally, we define

$$G = M \circ F, \text{ for } M \in \mathcal{M} \quad (1)$$

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

In addition to the above model, the following notation and terminology will be used. For an $n \times n$ matrix A , let $\det(A)$ denote the determinant of the matrix A . The characteristic polynomial for the matrix A is denoted $\det(A - \lambda I)$, where I is the $n \times n$ identity matrix. The eigenvalues of a matrix are the roots of $\det(A - \lambda I)$. We call an eigenvalue simple if it is a root of $\det(A - \lambda I)$ with multiplicity one. Let $\text{spec}(A)$ denote the set of eigenvalues of A . A matrix A is symmetric if $A_{ij} = A_{ji}$ for all i, j . The transpose of a matrix A is denoted A^T . We use the notation $A > 0$ to indicate that $A_{ij} > 0$ for all ij . For a matrix A , $\text{rank}(A)$ denotes the dimension of the range of A , or the number of linearly independent columns of the matrix. Let $\|A\|$ denote the norm of the matrix A . Let \mathcal{S}^1 denote the unit circle in \mathbb{C} .

3. MAIN RESULTS

Before we present the main result, we introduce key definitions.

DEFINITION 1. If $f(p) = p$, a point p is called a fixed point of f .

DEFINITION 2. A fixed point p for $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ is called hyperbolic if the Jacobian $Df(p)$ has no eigenvalues on the unit circle. A fixed point p is non-hyperbolic if $\text{spec}(Df(p)) \cap S^1 \neq \emptyset$.

DEFINITION 3. A map G is hyperbolic if all fixed points are hyperbolic.

DEFINITION 4. A property is typical, or generic, if it holds for an open and dense set of parameter values.

We now present our main result.

THEOREM 1. Let $G = M \circ F$ be a GA map (1). For a typical mixing operator, G is hyperbolic.

To prove the above theorem, we will need the following two propositions.

PROPOSITION 1. Let $G = M \circ F$ be a GA map (1). The set of mixing operators M , for which the fixed points of G are hyperbolic, forms an open set in \mathcal{M} .

The proof of this proposition is based on the fact that

$$\det(DG(p) - \lambda I) = \det([DM \circ F(p)]DF(p) - \lambda I)$$

is a continuous function of M and, therefore, if $\lambda_i \notin S^1$, then small perturbations do not change this fact. The proof of proposition 1 is relatively easy. The proof of the following proposition, 2, is considerably more difficult.

PROPOSITION 2. Let $G = M \circ F$ be a GA map (1). The set of mixing operators for which the fixed points of G are hyperbolic, forms a dense set in \mathcal{M} .

To prove this proposition, we will assume we have a fixed point p of G with one or more eigenvalues on S^1 . We first characterize perturbations, $M_\epsilon \in \mathcal{M}$, that preserve the fixed point.

We construct $\mathcal{P}(p)$ to simplify the characterization of the perturbations of M with the fixed point preserving characteristic. Let $\mathcal{P}(p)$ represent quadratic operators

$$P = (P_1, \dots, P_n)$$

for which the following properties hold:

1. $P_i \in \mathcal{A}^{n \times n}$ is symmetric for all $i = 1, \dots, n$;
2. $M_i \pm P_i > 0$;
3. $\sum_i P_i = \mathbf{0}$;
4. $[F(p)]^T P_i F(p) = 0$ where p is the fixed point.

It is easy to see that $\mathcal{P}(p) \neq \emptyset$.

For $P \in \mathcal{P}(p)$, let $M_\epsilon := M + \epsilon P$. In Lemma 1 we show that M_ϵ is a quadratic operator on Λ and $M_\epsilon \in \mathcal{M}$. Therefore, $G_\epsilon = M_\epsilon \circ F$ is a GA map.

LEMMA 1. Let $G = M \circ F$ be a GA map (1). Assume $p \in \Lambda$ has $G(p) = p$. If $P \in \mathcal{P}(p)$, then for sufficiently small $\epsilon > 0$, $M_\epsilon = M + \epsilon P$ satisfies

1. $M_\epsilon \in \mathcal{M}$
2. $G_\epsilon(p) = M_\epsilon \circ F(p) = p$.

That is, $G(p) = p = G_\epsilon(p)$.

We observe that

$$\begin{aligned} G_\epsilon &= M_\epsilon \circ F \\ &= (M + \epsilon P) \circ F \\ &= (M \circ F) + \epsilon(P \circ F) \\ &= G + \epsilon(P \circ F). \end{aligned}$$

Thus,

$$\begin{aligned} DG_\epsilon(p) &= D[G + (\epsilon P \circ F)](p) \\ &= DG(p) + H \end{aligned}$$

where $H \in \mathcal{A}^{n \times n}$. In order to trace the effects of perturbations of M on the derivative DG_ϵ , we define

$$\mathcal{H} = \{H \in \mathcal{A}^{n \times n} | H = D(P \circ F)(p) \text{ for } P \in \mathcal{P}(p)\}$$

Observe that conditions (3) and (4) for defining the class $\mathcal{P}(p)$ restrict the admissible set of perturbations P . It can be shown that this restriction implies that each $H \in \mathcal{H}$ has rank at most $n - 1$. Therefore, we can only perturb $\text{spec}(DG(p))$ in $n - 1$ directions. This makes the proof of density non-trivial.

LEMMA 2. Let $G = M \circ F$ be a GA map (1). Assume $G(p) = p$ for $p \in \Lambda$, and that $DG(p)$ has at most one simple eigenvalue λ_0 of norm one. Then, there exists $M_\epsilon \in \mathcal{M}$ such that $G_\epsilon(p) = p$ and $\text{spec}[DG_\epsilon(p)] \cap S^1 = \emptyset$.

To find this perturbation, we consider $DG(p)$ in the Jordan normal form, denoted $[DG(p)]_J$. The class of matrices \mathcal{H} becomes the class \mathcal{H}_J in the new basis. We find $H_J \in \mathcal{H}_J$, such that $\text{spec}(DG(p)_J + H_J) \cap S^1 = \emptyset$. Since $\text{spec}(DG(p)_J + H_J) = \text{spec}(DG(p) + H)$, we then calculate H corresponding to H_J to determine the appropriate $M_\epsilon \in \mathcal{M}$.

For repeated eigenvalues, the argument showing H_J exists becomes very complicated. In this case we address collections of eigenvalues on S^1 of multiplicity greater than one through use of the lemmas below.

LEMMA 3. Let $G = M \circ F$ be a GA map (1) with fixed point p . If $DG(p)$ has eigenvalue $\lambda_0 \in S^1$ and multiplicity $k > 1$, then there exists $P \in \mathcal{P}(p)$ such that $DG_\epsilon(p)$ has eigenvalue $\lambda_0 \in S^1$ with multiplicity at most 1.

The rank of the perturbation matrix H plays a critical role in the proof of Lemma 3.

LEMMA 4. Let $G = M \circ F$ be a GA map (1). There exists a perturbation $M_\epsilon \in \mathcal{M}$ of M with $G_\epsilon = M_\epsilon \circ F$ and $DG_\epsilon(p) = DG(p) + H$ such that H is of rank $n - 1$.

That such an H exists can be shown by explicitly forming an operator $P \in \mathcal{P}$ so that the corresponding $H \in \mathcal{H}$ has $\text{rank}(H) = n - 1$.

LEMMA 5. Let $G = M \circ F$ be a GA map (1). For each $H \in \mathcal{H}$ there exists an interval $[0, \delta_H]$, $\delta_H > 0$, such that for all $\delta \in [0, \delta_H]$, $\delta H \in \mathcal{H}$.

We finally prove Lemma 3 using the analytic function $g(c) = \det(DG(p) - \lambda_0 I + cH)$ in combination with the results of Lemma 5.

Recall that $\lambda_0 \in \mathcal{S}^1$ is the eigenvalue of $DG(p)$ with multiplicity $k > 1$. Since the polynomial $g : \mathbb{R} \mapsto \mathbb{C}$, $g(c) = \det(DG(p) - \lambda_0 I + cH)$, defines an analytic function in c , either

1. $g(c) \equiv 0$;
2. $g(c)$ has isolated zeros ([2]).

By Lemma 4, we can choose H to have rank $n - 1$. Thus 0 is a simple eigenvalue of H . For large values of c , we have $0 \in \text{spec}(cH)$ but for $\mu \in [\text{spec}(cH) \setminus \{0\}]$, $|\mu| > K$ for some large K . If $\|DG(p)\| \ll K$, then we can view $DG(p)$ as a small perturbation of cH . Two possibilities arise:

- (a) There exists $c \in \mathbb{R}$ such that $g(c) = \det(cH + DG(p) - \lambda_0 I) \neq 0$.
- (b) For all $c \in \mathbb{R}$, $g(c) = \det(cH + DG(p) - \lambda_0 I) = 0$.

Case (a) implies (2), i.e. g has isolated zeros. Since $g(0) = 0$, there is δ arbitrarily close to 0 such that $g(\delta) \neq 0$. The proof now follows from Lemma 5. In case (b), we note that since H had the simple eigenvalue 0, λ_0 must be a simple eigenvalue of $(cH + DG(p))$ for large c . This proves Lemma 3 in case (b).

4. CONCLUSIONS

This paper investigates the hyperbolicity of fixed points for the infinite population genetic algorithm as represented by the GA map(1). We show that for an open and dense set of mixing operators in \mathcal{M} , the fixed points of the GA map are hyperbolic. This implies that for most mixing operators the behavior of the infinite population model in the neighborhood of fixed points is a good predictor of the behavior of finite, but large, population models. In particular, for most mixing operators these fixed points perturb into fixed points of the finite population model, and furthermore, the stability properties of the fixed points in infinite and finite population models are the same.

With the exceptional set of those mixing operators for which these statements are not true is nowhere dense. This means, in particular, that an arbitrarily small perturbation of such exceptional mixing operators leads to a regular operator.

Since the local dynamics around fixed points perturbs from infinite population model to a finite population model, we conclude that the GA map can serve as a good approximation of the finite population model.

5. REFERENCES

- [1] C. Hayes and T. Gedeon. Hyperbolic fixed points are typical in the space of mixing operators for the infinite population genetic algorithm. *Technical Report*, http://www.math.montana.edu/~gedeon/gen_alg.html.
- [2] S. Lang. *Complex Analysis*. Springer, 4th edition, 1999.

- [3] C. R. Reeves and J. E. Rowe. *Genetic Algorithms - Principles and Perspectives: A Guide to GA Theory*. Kluwer Academic Publishers, 2003.
- [4] C. Robinson. *Dynamical Systems: Stability, Symbolic Dynamics and Chaos*. CRC Press, 1995.
- [5] J. E. Rowe, M. D. Vose, and A. H. Wright. Group properties of crossover and mutation. *Evolutionary Computation*, 10(2):151–184, 2002.
- [6] M. D. Vose. *The Simple Genetic Algorithm: Foundations and Theory*. The MIT Press, 1999.
- [7] M. D. Vose and A. H. Wright. Simple genetic algorithms with linear fitness. *Evolutionary Computation*, 2(4):347–368, 1994.
- [8] A. H. Wright and G. Bidwell. A search for counterexamples to two conjectures on the simple genetic algorithm. *Foundations of Genetic Algorithms 4*, Morgan Kaufman Publishers, 1997.
- [9] A. H. Wright and J. E. Rowe. Continuous dynamical systems models of steady-state genetic algorithms. *Foundations of Genetic Algorithms, 6*, Morgan Kaufmann, 2001.
- [10] A. H. Wright and M. D. Vose. Stability of vertex fixed points and applications. *Foundations of Genetic Algorithms 3*, Morgan Kaufman Publishers, 1995.