

# The Computational Complexity of Genetic Diversity

Ruta Mehta<sup>1</sup>, Ioannis Panageas<sup>\*2</sup>, Georgios Piliouras<sup>†3</sup>, and Sadra Yazdanbod<sup>4</sup>

- 1 University of Illinois at Urbana-Champaign, USA  
mehta.ruta@gmail.com
- 2 Georgia Institute of Technology, Atlanta, USA  
ioannis@gatech.edu
- 3 Singapore University of Technology and Design, Singapore  
georgios.piliouras@gmail.com
- 4 Georgia Institute of Technology, Atlanta, USA  
yazdanbod@gatech.edu

---

## Abstract

A key question in biological systems is whether genetic diversity persists in the long run under evolutionary competition, or whether a single dominant genotype emerges. Classic work by Kalmus in 1945 [14] has established that even in simple diploid species (species with chromosome pairs) diversity can be guaranteed as long as the heterozygous<sup>1</sup> individuals enjoy a selective advantage. Despite the classic nature of the problem, as we move towards increasingly polymorphic traits (e.g., human blood types) predicting diversity (and its implications) is still not fully understood. Our key contribution is to establish complexity theoretic hardness results implying that even in the textbook case of single locus (gene) diploid models, predicting whether diversity survives or not given its fitness landscape is algorithmically intractable.

Our hardness results are structurally robust along several dimensions, e.g., choice of parameter distribution, different definitions of stability/persistence, restriction to typical subclasses of fitness landscapes. Technically, our results exploit connections between game theory, nonlinear dynamical systems, and complexity theory and establish hardness results for predicting the evolution of a deterministic variant of the well known multiplicative weights update algorithm in symmetric coordination games; finding one Nash equilibrium is easy in these games. In the process we characterize stable fixed points of these dynamics using the notions of Nash equilibrium and negative semidefiniteness. This as well as hardness results for decision problems in coordination games may be of independent interest. Finally, we complement our results by establishing that under randomly chosen fitness landscapes diversity survives with significant probability. The full version of this paper is available at <http://arxiv.org/abs/1411.6322>.

**1998 ACM Subject Classification** F.2 Analysis of Algorithms and Problem Complexity

**Keywords and phrases** Dynamical Systems, Stability, Complexity, Optimization, Equilibrium

**Digital Object Identifier** 10.4230/LIPIcs.ESA.2016.65

---

\* Ioannis Panageas was supported by NSF EAGER award grants CCF-1415496 and CCF-1415498.

† Georgios Piliouras was supported by SUTD grant SRG ESD 2015 097.

<sup>1</sup> Having different alleles for a gene on two chromosomes.



## 1 Introduction

The beauty and complexity of natural ecosystems have always been a source of fascination and inspiration for the human mind. The exquisite biodiversity of Galapagos' ecosystem, in fact, inspired Darwin to propose his theory of natural selection as an explanatory mechanism for the origin and evolution of species. This revolutionary idea can be encapsulated in the catch phrase “survival of the fittest”<sup>2</sup>. Natural selection promotes the survival of those genetic traits that provide to their carriers an evolutionary advantage.

A marker of the true genius behind any paradigm is its longevity and in this respect the success of natural selection is without precedent. The formalized study of evolution, dating back to the work of Fisher, Haldane, and Wright in the beginning of the twentieth century, still and to a large extent focuses on simple, almost toy-like, alas concrete, models of this famous aphorism and the experimental and theoretical analysis of them. Arguably, the most influential result in the area of mathematical biology is Fisher's fundamental theorem of natural selection (1930) [9]. It states that the rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time. In the classical model of population genetics (Fisher-Wright-Haldane, discrete or continuous version) of single locus (one gene) multi-allele diploid models<sup>3</sup> it implies that the average fitness of the species populations is always strictly increasing unless we are at an equilibrium. In fact, convergence to equilibrium is point-wise even if there exist continuum of equilibria (See [21] and references therein). From a dynamical systems perspective, this is a rather strong characterization, since it establishes that the average fitness acts as a Lyapunov function for the system and that every trajectory converges to an equilibrium.

Besides the purely dynamical systems interpretation, an alternative, more palpable, game theoretic interpretation of these genetic systems is possible. Specifically, these systems can be interpreted as symmetric coordination/partnership two-agent games<sup>4</sup> where both agents, starting with the same mixed initial strategy, apply (discrete) replicator dynamics<sup>5</sup>. The analogies are as follows: The two players correspond each to a locus (or gene) on a pair of homologous chromosomes<sup>6</sup> and the alleles are their strategies. When both players choose a strategy, say  $i$  and  $j$ , an individual  $(i, j)$  is defined whose fitness, say  $A_{ij}$ , is the payoff to both players, hence we have a coordination game. Furthermore, allele pairs are unordered so we have  $A_{ij} = A_{ji}$ , *i.e.*,  $A$  is symmetric and so is the game. The frequencies of the alleles in the initial population, namely  $\mathbf{x} := (x_1, \dots, x_n) \in \Delta_n$ <sup>7</sup> of  $n$  different alleles, corresponds to the initial common mixed strategy of both players. In each generation, every individual from the population mates with another individual picked at random from the population, and the updates of the mixed strategies/allele frequencies are captured by replicator/MWUA dynamics, *i.e.*,

$$x'_i = x_i \frac{\sum_j A_{ij} x_j}{\mathbf{x}^T A \mathbf{x}} \quad (1)$$

<sup>2</sup> The phrase “survival of the fittest” was coined by Herbert Spencer.

<sup>3</sup> We present information related to biology in Section A.1.

<sup>4</sup> A coordination/partnership game is a game where at each outcome all agents receive the same utility.

<sup>5</sup> Replicator dynamics (as well as their discrete variants) are close analogues to the well known multiplicative updates (MWUA) family of dynamics [19].

<sup>6</sup> Most multicellular organisms have two sets of chromosomes; *i.e.*, they are diploid. These chromosomes are referred to as homologous chromosomes. If both alleles at a locus (or gene) on the homologous chromosomes are the same, they and the organism are homozygous with respect to that gene. If the alleles are different, they and the organism are heterozygous with respect to that gene. See section A.

<sup>7</sup>  $\Delta_n$  denotes the simplex of dimension  $n$ .

where  $x'_i$  is the proportion of allele  $i$  in the next generation (for details see Section 4). In game theoretic language, the fundamental theorem of natural selection implies that the social welfare  $\mathbf{x}^T A \mathbf{x}$  (average fitness in biology terms) of the game acts as potential for the game dynamics. This implies convergence to fixed points. Fixed points are superset of Nash equilibria where each strategy played with positive probability has the same average payoff.

We say that population is genetically diverse if at least two alleles have non-zero proportion in the population, *i.e.*, allele frequencies form a *mixed (polymorphic) strategy*. The game theoretic results do not provide insight on the survival of genetic diversity. One way to formalize this question is whether there exists a mixed fixed point that the dynamics converges to with positive probability, given a uniformly random starting point in  $\Delta_n$ . The answer to this question for the minimal case of  $n = 2$  alleles (alleles  $b/B$ , individuals  $bb/bB/BB$ ) is textbook knowledge and can be traced back to the classic work of Kalmus (1945) [14]. The intuitive answer here is that diversity can survive when the heterozygote individuals,  $bB$ , have a fitness advantage. Intuitively, this can be explained by the fact that even if evolution tries to dominate the genetic landscape by  $bB$  individuals, the random genetic mixing during reproduction will always produce some  $bb, BB$  individuals, so the equilibrium that this process is bound to reach will be mixed. On the other hand, it is trivial to create instances where homozygote individuals are the dominant species regardless of the initial condition.

As we increase the size/complexity of the fitness landscape, not only is not clear that a tight characterization of the diversity-inducing fitness landscape exists (a question about global stability of nonlinear dynamical systems), but also, it is even less clear whether one can decide efficiently whether such conditions are satisfied by a given fitness landscape (a computational complexity consideration). How can one address this challenge and moreover, how can one account for the apparent genetic diversity of the ecosystems around us?

In a nutshell, we establish that the decision version of the problem is computationally hard, by sandwiching limit points of the dynamics between various stability notions. This core result is shown to be robust across a number of directions. Deciding the existence of stable (mixed) polymorphic equilibria remains hard under a host of different definitions of stability examined in the dynamical systems literature. The hardness results persist even if we restrict the set of allowable landscape instances to reflect typical instance characteristics. Despite the hardness of the decision problems, randomly chosen fitness landscapes are shown to support polymorphism with significant probability (at least  $1/3$ ). The game theoretic interpretation of our results allow for proving hardness results for understanding standard game theoretic dynamics in symmetric coordination games. We believe that this is an important result of independent interest as it points out at a different source of complexity in understanding social dynamics.

## 2 Technical Overview

To study survival of diversity in diploidy, we need to characterize limiting population under evolutionary pressure. We focus on the simplest case of single locus (one gene) species. For this case, evolution under natural selection has been shown to follow replicator dynamics in symmetric two-player coordination games [21], where the genes on two chromosomes are players and alleles are their strategies as described in the introduction. Losert and Akin established point-wise convergence for this dynamics through a potential function argument [21]; here average fitness  $\mathbf{x}^T A \mathbf{x}$  is the potential. The limiting population corresponds to fixed points (FP), and so to make predictions about diversity (if the limiting population has support size at least 2) we need to characterize and compute these limiting FPs.

Let  $\mathcal{L}$  denote the set of fixed points (FP) with region of attraction<sup>8</sup> of positive (Lebesgue) measure. Hence, given a random starting point replicator dynamics converges to such a FP with positive probability, *i.e.*, the set of FP with region of attraction of positive measure. It seems that an exact characterization of  $\mathcal{L}$  is unlikely. Establishing necessary and sufficient conditions so that a FP has a region of attraction of positive measure in classes of nonlinear systems is a rather formidable task. Instead we try to capture this property as closely as possible through different stability notions. First we consider two standard notions defined by Lyapunov, the *stable* and *asymptotically stable* FP. Informally, if we start close to a *stable* FP then we stay close to it forever, while in case of *asymptotically stable* FP furthermore the dynamics converges to it (see Section 4.3). Thus *asymptotically stable*  $\subseteq \mathcal{L}$  follows, *i.e.*, an asymptotically stable FP has region of attraction of positive measure (*e.g.*, a small ball around the FP). Note that  $\mathcal{L}$  may have points that are not (asymptotically) stable.

There exist some well known connections between stability notions and properties of the (absolute) eigenvalues (EVal) of the Jacobian of the update rule (function) of the dynamics are well known: if the Jacobian at a FP has an EVal  $> 1$  then the FP is *un-stable* (not *stable*), and if all EVal  $< 1$  then it is asymptotically stable. The case when all EVal  $\leq 1$  with equality holding for some, is the ambiguous one. In that case we can say nothing about the stability because the Jacobian does not suffice. We will call these FPs *linearly-stable*. At a FP, say  $\mathbf{x}$ , if some EVal  $> 1$  then the direction of corresponding eigenvector is repelling, and therefore any starting vector with a component of this vector can never converge to  $\mathbf{x}$ . Thus points converging to  $\mathbf{x}$  can not have positive measure. Using this as an intuition we show that in our dynamical systems  $\mathcal{L} \subseteq$  *linearly-stable* FPs. In other words the set of initial points so that the dynamics converges to linearly-*un-stable* FPs has zero measure (Theorem 13). This theorem is heavily utilized to understand the (non-)existence of diversity.

Efficient computation requires efficient verification. However, note that whether a given FP is (asymptotically) stable or not does not seem easy to verify. To achieve this, one of the contributions of this paper is the definition of two more notions: *Nash stable* and *strict Nash stable*.<sup>9</sup> It is straightforward to check that Nash equilibria (NE) of the corresponding coordination game described in introduction are FPs of the replicator dynamics (Equations 1,2) but not vice-versa. Keeping this in mind we define *Nash stable* FP, which is a NE with the additional property that the sub-matrix corresponding to its support satisfies certain negative semi-definiteness. The latter condition is derived from the fact that *stability* is related to local optima of  $\mathbf{x}^T \mathbf{A} \mathbf{x}$  and also from Sylvester's law of inertia [38] (see Section 5 and proofs). For *strict Nash stable* both conditions are strict, namely strict NE and negative definiteness. Combining all of these notions we show the following:

► **Theorem 1.**

$$\begin{array}{c} \textit{Strict Nash stable} \subseteq \textit{Asymptotically stable} \subseteq \mathcal{L} \subseteq \textit{linearly-stable} = \textit{Nash stable} \\ \cap \\ \textit{stable} \subseteq \textit{linearly-stable} = \textit{Nash stable} \end{array}$$

We note that the sets of asymptotically stable, stable,  $\mathcal{L}$  and linearly stable FPs of Theorem 1 do not coincide in general.<sup>10</sup> For example, let  $x_{t+1}$  be the next step for the

<sup>8</sup> Region of attraction of an FP is the set of all initial points so that dynamics converges to it.

<sup>9</sup> These two notions are not the same as evolutionary stable strategies/states.

<sup>10</sup> We also note that generically these sets coincide. It can be shown [22] that given a fitness matrix, its entries can be perturbed to ensure no eigenvalue on the unit circle for the Jacobian at any fixed point. Such fixed points are called *hyperbolic*. Formally, if we consider the dynamics as an operator (called *Fisher operator*) then the set of hyperbolic operators is dense in the space of Fisher operators.

following update rules:

$$f(x_t) = \frac{1}{2}x_t, g(x_t) = x_t - \frac{1}{2}x_t^2, h(x_t) = x_t + \frac{x_t^3}{2}, d(x_t) = x_t.$$

Then for dynamics governed by  $f$ , 0 is asymptotically stable, stable and linearly stable, and hence is also in  $\mathcal{L}$ . While for  $g$  it is linearly stable and is in  $\mathcal{L}$ , but is not stable or asymptotically stable. For  $d$  it is linearly stable and stable, but not asymptotically stable and is not in  $\mathcal{L}$ . Finally, for  $h$  it is only linearly-stable, and does not belong to any other class. Our primary goal was to see if diversity is going to survive. We formalize this by checking whether set  $\mathcal{L}$  contains a mixed point, *i.e.*, where more than one alleles have non-zero proportion, implying that *diversity survives with some positive probability*, where the randomness is w.r.t the random initial  $\mathbf{x} \in \Delta_m$ . In Section 7 we show that for all five notions of stability, checking existence of mixed stable FP is NP-hard.

Our reductions are from  $k$ -**clique** - given an undirected graph check if it has a clique of size  $k$ ; a well known NP-hard problem. Given an instance  $G$  of  $k$ -*clique*, we will construct a symmetric matrix  $A$  as shown in Figure 1, and consider coordination game  $(A, A)$ . We show game  $(A, A)$  has the following two properties.

- (i) If  $G$  has a clique of size  $k$  then  $(A, A)$  has a mixed *strict* Nash stable equilibrium. Therefore, it has a mixed stable FP for any notions of the (strict) Nash stability, (asymptotic) stability, linearly-stability, and  $\mathcal{L}$  because of Theorem 1.
- (ii) If  $(A, A)$  has a mixed Nash stable equilibrium then  $G$  has a clique of size  $k$ . Therefore, if  $(A, A)$  has a mixed stable FP for all notions of the (strict) Nash stability, (asymptotic) stability, linearly-stability, and  $\mathcal{L}$  then  $G$  has a clique of size  $k$  because of Theorem 1.

Note that the above two properties imply NP-hardness for checking existence of mixed stable FP for all five notions of stability as well as set  $\mathcal{L}$ . The latter implies NP-hardness for checking survival of diversity in diploid species even with single gene.

► **Theorem 2 (Informal).** *Given a symmetric matrix  $A$ , it is NP-hard to check if replicator dynamics with payoff  $A$  has mixed (asymptotically) stable, linear-stable, or (strict) Nash stable fixed points. A common reduction for all together with Theorem 1 will imply that it is NP-hard to check whether diversity survives for a given fitness matrix.*

The main idea in the construction of matrix  $A$  (Figure 1) is to use a modified version of the adjacency matrix  $E$  of the graph as one of the blocks in the payoff matrix such that the existence of a clique of size  $k$  or more implies a stable Nash equilibrium in that block, and all stable mixed equilibria are only in that block. Here  $E'$  is the modification of  $E$  where off-diagonal zeros are replaced with  $-h$  where  $h$  is a large (polynomial-size number). Note that  $A$  is  $2n \times 2n$ , where the first  $n$  strategies correspond to the  $n$  nodes of the graph. To argue (i), given a  $k$ -clique we construct a maximal clique containing it, say of size  $m$ , and then show that the distribution that assigns probability  $\frac{1}{m}$  to the strategies corresponding to nodes in the clique, and zero otherwise, is a (strict Nash) stable FP. To show (ii), we use the negative semi-definiteness property of Nash stable (which contains all other notions), which implies  $A_{ii} \leq 2A_{ij}$  for all  $i, j$  in the support a Nash stable FP. Therefore if  $h > 2(k-1)$ , then none of  $\{n+1, \dots, 2n\}$  can be in the support. Next we concentrate only on strategies that have big enough probability, call it set  $S$ . The Nash equilibrium property of Nash stable states can be shown to force each non-zero probability to be  $\leq \frac{1}{k}$ . Using these facts we show that  $|S| \geq k$  and the corresponding vertices in  $G$  form a clique. Thus, we prove the existence of a  $k$ -clique.

The fitness matrix constructed for the hardness results is rather specific. Do these hardness results carry over to “typical” (at least not completely worst case) instances of

fitness landscapes? What is the complexity of checking if a given allele survives? We show that our computational intractability results still apply to these settings.

There has been a lot of work on NP-hardness for decision versions of Nash equilibrium in general games [11, 5, 34, 10], where finding one equilibrium is also PPAD-hard [3]. Whereas to the best of our knowledge these are the first NP-hardness results for coordination games, where finding one Nash equilibrium is easy, and therefore may be of independent interest. Finally, in Section 6 we show that even though checking the survival of diversity is hard, on average diversity persists.

► **Theorem 3 (Informal).** *If the entries of a fitness matrix are i.i.d. on a continuous distribution then with significantly high probability, at least  $\frac{1}{3}$ , diversity will surely survive.*

Survival is ensured if every fixed point in  $\mathcal{L}$  is mixed. This itself is guaranteed as long as every diagonal entry  $(i, i)$  of the fitness matrix is dominated by some entry in its row or column. We can lower bound the probability of the latter *by a constant* for a random symmetric matrix (from continuous distribution) of *any size*. The tricky part is to avoid correlation arising due to symmetry and we achieve this using inclusion-exclusion arguments.

### 3 Related Work

In the last few years we have witnessed a rapid cascade of theoretical results on the intersection of computer science and evolution. Livnat *et al.* [20] introduced the notion of mixability, the ability of an allele to combine itself successfully with other alleles within a specific population. In [2, 1] connections were established between haploid evolution and game theoretic dynamics in coordination games. Even more recently Meir and Parkes [24] provided a more detailed examination of these connections. These dynamics are close variants of the standard (discrete) replicator dynamics [12]. Replicator dynamics is closely connected to the multiplicative weights update algorithm [19]. Analogous game theoretic interpretations are known for diploids [21].

Analyzing limit sets of dynamical systems is a critical step towards understanding the behavior of processes that are inherently dynamic, like evolution. There has been an upsurge in studying the complexity of computing these sets. Quite few works study such questions for dynamical systems governed by arbitrary continuous functions or ODEs [15, 16, 35]. Limit cycles are inherently connected to dynamical systems and recent works by Papadimitriou and Vishnoi [29] showed that computing a point on an approximate limit cycle is PSPACE-complete. Cyclic limit sets also arise in game theoretic dynamics [18, 32, 31, 28]. On the positive side, in [27], it was shown that a class of evolutionary Markov chains mix rapidly, where techniques from dynamical systems were used.

The complexity of checking if a game has an evolutionary stable strategy (ESS) has been studied first by Nissan and then by Etessami and Lochbihler [26, 8] and has been nailed down to be  $\Sigma_2^P$ -complete by Conitzer [4]. Unlike our setting here the game is between different species to survive in a common environment. These problems are orthogonal to understanding issues of genetic diversity, and thus not directly comparable to our work.

Other connections between computational complexity and ecology/evolution examine the complexity of finding local/global minima of structured fitness landscapes [37, 17, 39], as well as, complexity questions in regards to the probability that a new invader (or a new mutant) will take over a resident population [33]. The sexual dynamics and the questions about diversity considered here are not captured in any of the settings above.

In [23] Mehta, Panageas and Piliouras examine the question of diversity for haploid species. Despite the systems' superficial similarities the two analyses come to starkly different

conclusions. In haploids systems all mixed (polymorphic) equilibria are unstable and evolution converges to monomorphic states. In the case of diploid systems the answer to whether diversity survives or not depends crucially on the geometry of the fitness landscape.

## 4 Preliminaries

**Notations:** We use boldface letters, like  $\mathbf{x}$ , to denote column vectors, and  $x_i$  is its  $i^{\text{th}}$  coordinate. Vectors  $\mathbf{0}_n$  and  $\mathbf{1}_n$  are  $n$ -D vectors with all zeros and ones. For matrix  $A$ ,  $A_{ij}$  is the entry in  $i^{\text{th}}$  row and  $j$ -th column and we denote  $(A\mathbf{x})_i = \sum_j A_{ij}x_j$ . By norm  $\|\cdot\|$  we mean  $\|\cdot\|_\infty$ .  $[n]$  denotes set  $\{1, \dots, n\}$ , and  $\Delta_n$  denotes  $n$ -D simplex, and let  $SP(\mathbf{x})$  be  $\{i \mid x_i > 0\}$ .

### 4.1 Evolutionary dynamics

Consider a diploid single locus species, *i.e.*, a species with a chromosome pair and single gene. Every gene has a set of representative alleles, e.g., gene for eye color has different alleles for brown, black and blue eyes. Let  $n$  be the number of alleles for the single gene of our species, and let these be numbered  $1, \dots, n$ . An individual is represented by an unordered pair of alleles  $(i, j)$  one in each chromosome, and we denote its fitness by  $A_{ij}$ ; clearly  $A$  is symmetric. Here fitness represents its ability to reproduce during a mating. In every generation two individuals are picked uniformly at random from the population, say  $(i, j)$  and  $(i', j')$ , and they mate. The allele pair of the offspring can be any of the four possible combinations, namely  $(i, i')$ ,  $(i, j')$ ,  $(i', j)$ ,  $(j, j')$ , with equal probability. Let  $x_i$  be a random variable that denotes the proportion of the population with allele  $i$ . After one generation, the expected number of offsprings with allele  $i$  is proportional to  $x_i \cdot x_i \cdot (A\mathbf{x})_i + 2\frac{1}{2}(1 - x_i)x_i \cdot (A\mathbf{x})_i = x_i(A\mathbf{x})_i$  ( $x_i^2$  stands for the probability that first individual has both his alleles  $i$ , *i.e.*, is represented by  $(i, i)$  - and thus the offspring will inherit allele  $i$  - and  $2\frac{1}{2}(1 - x_i)x_i$  stands for the probability that the first individual has allele  $i$  exactly once in his representation and the offspring will inherit). Hence, if  $\mathbf{x}$  denote the frequencies of the alleles in the population in the next generation (random variables)

$$E[x'_i | \mathbf{x}] = \frac{x_i(A\mathbf{x})_i}{\mathbf{x}^T A \mathbf{x}}.$$

We focus on the *deterministic* version of the equations above, which captures the infinite population model. Thus if  $\mathbf{x} \in \Delta_n$  represents the proportions of alleles in the current population. Under the evolutionary process of natural-selection (the reproduction happens as described) this proportion changes as per the following multi-variate function  $f: \Delta_n \rightarrow \Delta_n$  under the infinite population model [21]; in the literature this is often called *Discrete Replicator Dynamics*.

$$\mathbf{x}' = f(\mathbf{x}) \quad \text{where} \quad x'_i = f_i(\mathbf{x}) = x_i \frac{(A\mathbf{x})_i}{\mathbf{x}^T A \mathbf{x}}, \quad \forall i \in [n] \quad (2)$$

where  $\mathbf{x}'$  are the proportions of the next generation.  $f$  is a continuous function with convex, compact domain (= range), and therefore always has a fixed point [13]. Furthermore, limit points of  $f$  have to be fixed points, *i.e.*,  $\mathbf{x}$  such that  $f(\mathbf{x}) = \mathbf{x}$ .

► **Fact 4.** Profile  $\mathbf{x}$  is a fixed point of  $f$  iff  $\forall i \in [n], x_i > 0 \Rightarrow (A\mathbf{x})_i = \mathbf{x}^T A \mathbf{x}$ .

One of the fundamental questions is: starting from arbitrary population of alleles how does the population look like in the limit under evolutionary pressures? Does the system

converge? If so, then what are the properties of these limit points? Fisher's Fundamental Theorem of Natural Selection [21, 22] says that mean fitness function  $\pi(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$  (potential function in game theory terms) satisfies the inequality  $\pi(f(\mathbf{x})) \geq \pi(\mathbf{x})$  and the equality holds iff  $\mathbf{x}$  is a fixed point. Losert and Akin [21, 22] showed that the dynamics above converge point-wise to fixed points and that  $f$  is a diffeomorphism in  $\Delta_n$ .

## 4.2 Games, Nash equilibria and symmetries

In this paper we consider two-player (locus) games, where each player has finitely many pure strategies (alleles). Let  $S_i$ ,  $i = 1, 2$  be the set of strategies for player  $i$ , and let  $m \stackrel{\text{def}}{=} |S_1|$  and  $n \stackrel{\text{def}}{=} |S_2|$ , then such a game can be represented by two payoff matrices  $A$  and  $B$  of dimension  $m \times n$ . Players may randomize amongst their strategies. The set of mixed strategies are  $\Delta_m$  and  $\Delta_n$  respectively.

► **Definition 5.** Nash Equilibrium [36] A strategy profile is said to be a Nash equilibrium (NE) if no player can achieve a better payoff by a unilateral deviation [25]. Formally,  $(\mathbf{x}, \mathbf{y}) \in \Delta_m \times \Delta_n$  is a NE iff  $\forall \mathbf{x}' \in \Delta_m$ ,  $\mathbf{x}'^T A \mathbf{y} \geq \mathbf{x}'^T A \mathbf{y}$  and  $\forall \mathbf{y}' \in \Delta_n$ ,  $\mathbf{x}^T B \mathbf{y}' \geq \mathbf{x}^T B \mathbf{y}'$ .

Game  $(A, B)$  is said to be symmetric if  $B = A^T$ . In a symmetric game the strategy sets of both players are identical, *i.e.*,  $m = n$ , and  $S_1 = S_2$ . We will use  $n$ ,  $S$  and  $\Delta_n$  to denote the strategy related quantities. Strategy  $\mathbf{x} \in \Delta_n$  is a symmetric NE, both play  $\mathbf{x}$ , iff

$$\forall i \in S, x_i > 0 \Rightarrow (A\mathbf{x})_i = \max_k (A\mathbf{x})_k \quad (3)$$

► **Definition 6.** NE  $\mathbf{x}$  is **strict** if  $\forall k \notin SP(\mathbf{x})$ ,  $(A\mathbf{x})_k < (A\mathbf{x})_i$ , where  $i \in SP(\mathbf{x})$ .

**Symmetric Coordination Game.** In a coordination/partnership game  $B = A$ , *i.e.*, both the players get the same payoff always. Thus if  $A$  is symmetric then  $(A, A)$  is a symmetric coordination game. The next lemma follows using (3) and Fact 4

► **Lemma 7.** If  $\mathbf{x}$  is a symmetric NE of game  $(A, A)$ , it is a fixed point of dynamics (2).

From now on, by *mixed* (fixed point) strategy we mean strictly mixed (fixed point) strategy, *i.e.*,  $\mathbf{x}$  such that  $|SP(\mathbf{x})| > 1$ , and non-mixed are called *pure*.

## 4.3 Basics in Dynamical Systems

Next we describe well-known stability notions in dynamical systems.

► **Definition 8.** A fixed point  $\mathbf{r}$  of  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is called **stable** if, for every  $\epsilon > 0$ , there exists a  $\delta = \delta(\epsilon) > 0$  such that, for all  $\mathbf{p} \in \mathbb{R}^n$  with  $\|\mathbf{p} - \mathbf{r}\| < \delta$  we have that  $\|f^n(\mathbf{p}) - \mathbf{r}\| < \epsilon$  for every  $n \geq 0$ , otherwise it is called **unstable**.

► **Definition 9.** A fixed point  $\mathbf{r}$  of  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is called **asymptotically stable** if it is stable and there exists a (neighborhood)  $\delta > 0$  such that, for all  $\mathbf{p} \in \mathbb{R}^n$  with  $\|\mathbf{p} - \mathbf{r}\| < \delta$  we have that  $\|f^n(\mathbf{p}) - \mathbf{r}\| \rightarrow 0$  as  $n \rightarrow \infty$ .

By definition it follows that if  $\mathbf{x}$  is asymptotically stable w.r.t. dynamics  $f$  (2), then the set of initial conditions in  $\Delta$  so that the dynamics converge to  $\mathbf{x}$  has positive measure. Using that under  $f$  the potential function  $\pi(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$  strictly decreases unless  $\mathbf{x}$  is a fixed point, the next theorem was derived in [22].

► **Theorem 10** ([22], § 9.4.7). *A fixed point  $\mathbf{r}$  of dynamics (2) is stable if and only if it is a local maximum of  $\pi$ , and is asymptotically stable if and only if it is a strict local maximum.*

As the domain of  $\pi$  is closed and bounded, there exists a global maximum of  $\pi$  in  $\Delta_n$ , which by Theorem 10 is a stable fixed point, and therefore its existence follows. However, existence of asymptotically stable fixed point is not guaranteed, for example if  $A = [1]_{m \times n}$  then no  $\mathbf{x} \in \Delta_n$  is attracting under  $f$ .

### 4.3.1 Stability and Eigenvalues

To analyze limiting points of  $f$  with respect to the notion of stability in terms of perturbation resistant, we need to use the eigenvalues of the Jacobian of  $f$  at fixed points. Let  $J^{\mathbf{r}}$  denote the Jacobian at  $\mathbf{r} \in \Delta_n$ . The following theorem in dynamics/control theory relates (asymptotically) stable fixed points with the eigenvalue of its Jacobian.

► **Theorem 11** ([30]). *At fixed point  $\mathbf{x}$  if  $J^{\mathbf{x}}$  has at least one eigenvalue with absolute value  $> 1$ , then  $\mathbf{x}$  is unstable. If all the eigenvalues have absolute value  $< 1$  then it is asymptotically stable.*

► **Definition 12.** A fixed point  $\mathbf{r}$  is called linearly stable, if the eigenvalues  $J^{\mathbf{r}}$  are at most 1 in absolute value. Otherwise, it is called linearly unstable.

Theorem 11 implies that eigenvalues of the Jacobian at a stable fixed point have absolute value at most 1, however the converse may not hold. Using properties of  $J^{\mathbf{x}}$  and [21], we prove next: (see the full version for Jacobian equations).

► **Theorem 13.** *The set of initial conditions in  $\Delta_n$  so that the dynamics (2) converge to linearly unstable fixed points has measure zero.*

In Theorem 13 we manage to discard only those fixed points whose Jacobian has eigenvalue with absolute value  $> 1$ , while characterizing limiting points of  $f$ ; the latter is finally used to argue about the survival of diversity.

## 5 Convergence, Stability, and Characterization

As established in Section 4.1, evolution in single locus diploid species is governed by dynamics  $f$  of (2). Understanding survival of diversity requires to analyze the following set,

$$\mathcal{L} = \{\mathbf{x} \in \Delta_n \mid \text{positive measure of starting points converge to } \mathbf{x} \text{ under } f\} \quad (4)$$

The next lemma follows from Definition 9 and Theorem 13.

► **Lemma 14.** *asymptotically stable  $\subseteq \mathcal{L} \subseteq$  linearly stable.*

In this section we try to characterize  $\mathcal{L}$  using various notions of stability, which have game theoretic and combinatorial interpretation. These notions sandwich set-wise the classical notions of stability given in Section 4.3, and thereby give us a partial characterization of  $\mathcal{L}$ . This characterization turns out to be crucial for our hardness results as well as results on survival in random instances.

Given a symmetric matrix  $A$ , a two-player game  $(A, A)$  forms a symmetric coordination game. In this section we identify special symmetric NE of this game to characterize stable

## 65:10 The Computational Complexity of Genetic Diversity

fixed points of  $f$ . Given a profile  $\mathbf{x} \in \Delta_n$ , define a transformed matrix  $T(A, \mathbf{x})$  of dimension  $(k-1) \times (k-1)$ , where  $k = |SP(\mathbf{x})|$ , as follows.

$$\text{Let } SP(\mathbf{x}) = \{i_1, \dots, i_k\}, B = T(A, \mathbf{x}). \forall a, b < k, B_{ab} = A_{i_a i_b} + A_{i_k i_k} - A_{i_a i_k} - A_{i_k i_b} \quad (5)$$

Since  $A$  is symmetric it is easy to check that  $B$  is also symmetric, and therefore has real eigenvalues. Recall the Definition 6 of strict symmetric NE.

► **Definition 15.** A strategy  $\mathbf{x}$  is called (strict) Nash stable if it is a (strict) symmetric NE of the game  $(A, A)$ , and  $T(A, \mathbf{x})$  is negative (definite) semi-definite.

► **Lemma 16.** For any given  $\mathbf{x} \in \Delta_n$ ,  $T(A, \mathbf{x})$  is negative (definite) semi-definite iff  $(\mathbf{y}^T A \mathbf{y} < 0) \mathbf{y}^T A \mathbf{y} \leq 0, \forall \mathbf{y} \in \mathbb{R}^n$  such that  $\sum_i y_i = 0$  and  $x_i = 0 \Rightarrow y_i = 0$ .

Using that stable fixed point are local optima, we map them to Nash stable strategies.

► **Lemma 17.** Every stable fixed point  $\mathbf{r}$  of  $f$  is a Nash stable of game  $(A, A)$ .

Since stable fixed points always exist, so do Nash stable strategies (Lemma 17). Next we map strict Nash stable strategies to asymptotically stable fixed points, as the negative definiteness and strict symmetric Nash of the former implies strict local optima, and the next lemma follows.

► **Lemma 18** ([22] § 9.2.5). Every strict Nash stable is asymptotically stable.

The above two lemmas show that strict Nash stable  $\subseteq$  asymptotically stable (by definition)  $\subseteq$  stable (by definition)  $\subseteq$  Nash stable. Further, by Theorem 11 and the definition of *linearly stable* fixed points we know that stable  $\subseteq$  linearly-stable. What remains is the relation between Nash stable and linearly stable. The next lemma answers this.

► **Lemma 19.** Strategy  $\mathbf{r}$  is Nash stable iff it is a linearly stable fixed point.

Using Theorems 10 and 13, and Lemmas 14, 17, 18 and 19 we get the following characterization among all the notions of stability that we have discussed so far.

► **Theorem 20.** Given a symmetric matrix  $A$ , we have

$$\begin{aligned} \text{Strict Nash stable} \subseteq \text{Asymptotically stable} \subseteq \mathcal{L} \subseteq \text{linearly-stable} = \text{Nash stable} \\ \cap \\ \text{stable} \subseteq \text{linearly-stable} = \text{Nash stable} \end{aligned}$$

As stated before, generically (random fitness matrix) we have hyperbolic fixed points and all the previous notions coincide.

## 6 Survival of Diversity

In this section we characterize two extreme cases of fitness matrix for the survival of diversity, namely where diversity always survives and where diversity disappears regardless of the starting population. Using this characterization we analyze the chances of survival of diversity when fitness matrix and starting populations are picked uniformly at random from continuous distributions.

Given a fitness (positive) matrix  $A$ , let  $\mathbf{x}$  be a limit point of dynamics  $f$  governed by (2) If it is not pure, *i.e.*,  $|SP(\mathbf{x})| > 1$  then at least two alleles survive among the population, and we say the population is diverse in the limit.

► **Definition 21.** We say that *diversity survives* in the limit if there exists  $\mathbf{x} \in \mathcal{L}$  such that  $\mathbf{x}$  is not pure. And *diversity survives surely* if no  $\mathbf{x} \in \mathcal{L}$  is pure.

Since  $\mathcal{L} \subseteq \text{Nash stable} = \text{linearly stable}$  (Theorem 20), there has to be at least one mixed Nash (or linearly) stable strategy for diversity to survive.

Next we give a definition that captures the homozygote/heterozygote advantage and a lemma which uses it to identify instances that lack mixed Nash stable strategies.

► **Definition 22.** Diagonal entry  $A_{ii}$  is called *dominated* if and only if  $\exists j$ , such that  $A_{ij} > A_{ii}$ . And it is called *dominating* if and only if  $A_{ii} > A_{ij}$  for all  $j \neq i$ .

In full version we show that diversity dies for matrices with all *dominating* diagonals. Next we show sure survival of diversity when diagonals are *dominated*.

► **Lemma 23.** Let  $\mathbf{r}$  be a fixed point of  $f$  with  $r_t = 1$ . If  $A_{tt}$  is dominated, then  $\mathbf{r}$  is linearly unstable.

If all pure fixed points are linearly unstable, then all linearly stable fixed points are mixed, and thus the next theorem follows using Theorem 20 and Lemma 23.

► **Theorem 24.** If every diagonal of  $A$  is dominated then no  $\mathbf{x} \in \mathcal{L}$  is pure, i.e., diversity survives almost surely.

The following lemma shows that when the entries of a fitness matrix are picked uniformly independently from a continuous distribution, there is a positive probability (bounded away from zero for all  $n$ ) so that every diagonal in  $A$  is dominated. This essentially means that generically, diversity survives with positive probability, bounded away from zero, where the randomness is taken with respect to both the payoff matrix and initial conditions.

► **Lemma 25.** Let entries of  $A$  be chosen i.i.d from a continuous distribution. The probability that all diagonals of  $A$  are dominated is at least  $\frac{1}{3} - o(1)$ .

The next theorem follows using Theorem 24 and Lemma 25.

► **Theorem 26.** Assume that the fitness matrix has entries picked independently from a continuous distribution then with probability, at least  $\frac{1}{3}$ , diversity will survive almost surely.

► **Remark 27 (Typical instance).** Observe that letting  $X_i$  be the indicator random variable that  $A_{ii}$  is dominating and  $X = \sum_i X_i$  we get that  $E[X] = \sum_i E[X_i] = \sum_i \Pr[E_i] = n \times \frac{1}{n} = 1$  so in expectation we will have one dominating element. Also from the above proof of Lemma 25 we get that  $E[X^2] = \sum_i E[X_i] + 2 \sum_{i < j} E[X_i X_j] = 1 + n(n-1)\Pr[E_i \cap E_j] \approx 2 - o(1)$  (namely  $\text{Var}[X] \approx 1 - o(1)$ ) so by Chebyshev's inequality  $\Pr[|X - 1| > k]$  is  $O(\frac{1}{k^2})$ .

## 7 NP-Hardness Results

Positive chance of survival of phenotypic (allele) diversity in the limit under the evolutionary pressure of selection (dynamics (2)), implies existence of a mixed linearly stable fixed point (Theorem 13). This notion encompasses all the other notions of stability (Theorem 20), and may contain points that are not attracting. Whereas, strict Nash stable and asymptotically stable are attracting.

In this section we show that checking if there exists a mixed stable profile, for any of the five notions of stability (Definitions 8, 9, 12 and 15), may not be easy. In particular, we show that the problem of checking if there exists a mixed profile that satisfies any of the stability

$E'$		$k-1$	$-\epsilon$
		$-\epsilon$	$k-1$
$k-1$	$-\epsilon$	$h$	$-\epsilon$
$-\epsilon$	$k-1$	$-\epsilon$	$h$

■ **Figure 1** Matrix  $A$  as defined in (6), where  $E'$  is modification of  $E$  where off-diagonal zeros are replaced with  $-h$  where  $h$  is a large (polynomial-size number).

conditions is NP-hard. In order to obtain hardness for checking survival of diversity, in other words checking if set  $\mathcal{L}$  has a mixed strategy, we need to obtain a unifying reduction.

Our reduction also gives NP-hardness for checking if a given pure strategy is played with non-zero probability (subset) at these. In other words, it is NP-hard to check if a particular allele is going to survive in the limit under the evolution. Finally we extend all the results to the typical class of matrices, where exactly one diagonal entry is dominating (see Remark 27). All the reductions are from **k-Clique**, a well known NP-complete problem [6].

► **Definition 28. (k-Clique)** Given an undirected graph  $G = (V, E)$ , with  $V$  vertices and  $E$  edges, and an integer  $0 < k < |V| - 1 = n - 1$ , decide if  $G$  has a clique of size  $k$ .

**Properties of G.** Given a simple graph  $G = (V, E)$  if we create a new graph  $\tilde{G}$  by adding a vertex  $u$  and connecting it to all the vertices  $v \in V$ , then it is easy to see that graph  $G$  has a clique of size  $k$  if and only if  $\tilde{G}$  has a clique of size  $k + 1$ . Therefore, w.l.o.g we can assume that there exists a vertex in  $G$  which is connected to all the other vertices. Further, if  $n = |V|$ , then for us such a vertex is the  $n^{\text{th}}$  vertex. By abuse of notation we will use  $E$  an adjacency matrix of  $\tilde{G}$  too,  $E_{ij} = 1$  if edge  $(i, j)$  present in  $\tilde{G}$  else it is zero.

## 7.1 Hardness for checking stability

In this section we show NP-hardness (completeness for some) results for decision versions on (strict) Nash stable strategies and (asymptotically) stable fixed points. Given graph  $G = (V, E)$  and integer  $k < n$ , we construct the following symmetric  $2n \times 2n$  matrix  $A$  (see Figure 1), where  $E'$  is modification of  $E$  where off-diagonal zeros are replaced with  $-h$  where  $h > 2n^2 + 5$ .

$$\forall i \leq j, A_{ij} = A_{ji} = \begin{cases} E'_{ij} & \text{if } i, j \leq n \\ k - 1 & \text{if } |i - j| = n \\ h & \text{if } i, j > n \text{ and } i = j, \text{ where } h > 2n^2 + 5 \\ -\epsilon & \text{otherwise, where } 0 < \epsilon \leq \frac{1}{10n^3} \end{cases} \quad (6)$$

$A$  is a symmetric but is not non-negative. The next lemma maps a  $k$ -clique to a mixed-strategy that is also strict Nash stable fixed point (FP). Note that such a FP satisfies all other stability notions as well, and hence implies existence of mixed limit point in  $\mathcal{L}$ .

► **Lemma 29.** *If there exists a clique of size at least  $k$  in graph  $G$ , then the game  $(A, A)$  has a mixed strategy  $\mathbf{p}$  that is strict Nash stable.*

Since strict Nash stable is contained in all other sets, the above lemma implies existence of mixed-strategy for all of them if there is a clique in  $G$ . Next we want to show the converse for all notions of stability. That is if mixed-strategy exists for any notion of the five notions of stability then there is a clique of size at least  $k$  in the graph  $G$ . Since each of the five stability implies Nash stability, it suffices to map mixed Nash stable strategy to clique of size  $k$ . For this, and reductions that follow, we will use the following property due to negative semi-definiteness of Nash stability extensively.

► **Lemma 30.** *Given a fixed point  $\mathbf{x}$ , if  $T(A, \mathbf{x})$  is negative semi-definite, then  $\forall i \in SP(\mathbf{x}), A_{ii} \leq 2A_{ij}, \forall j \neq i \in SP(\mathbf{x})$ . Moreover if  $\mathbf{x}$  is a mixed Nash stable then it has in its support at most one strategy  $t$  with  $A_{tt}$  is dominating.*

Nash stable also implies symmetric Nash equilibrium. The next lemma maps (special) symmetric NE to  $k$ -clique.

► **Lemma 31.** *Let  $\mathbf{p}$  be a symmetric NE of game  $(A, A)$ . If  $SP(\mathbf{p}) \subset [n]$  and  $|SP(\mathbf{p})| > 1$ , then there exists a clique of size  $k$  in graph  $G$ .*

We obtain the following lemma essentially using Lemmas 30 and 31.

► **Lemma 32.** *If game  $(A, A)$  has a mixed Nash stable strategy, then graph  $G$  has a clique of size  $k$ .*

Note that adding a constant to  $A$  does not change its set of strict Nash stable and Nash stable.

► **Lemma 33.** *Let  $A$  be a symmetric matrix, and  $B = A + c$  for a  $c \in \mathbb{R}$ , then the set of (strict) Nash stable strategies of  $B$  are identical to that of  $A$ .*

The next theorem follows using Theorem 20, Lemmas 29 and 32, and the property observed in Lemma 33. Since there is no polynomial-time checkable condition for (asymptotically) stable fixed points<sup>11</sup> its containment in NP is not clear, while for (strict) Nash stable strategies containment in NP follows from the Definition 15.

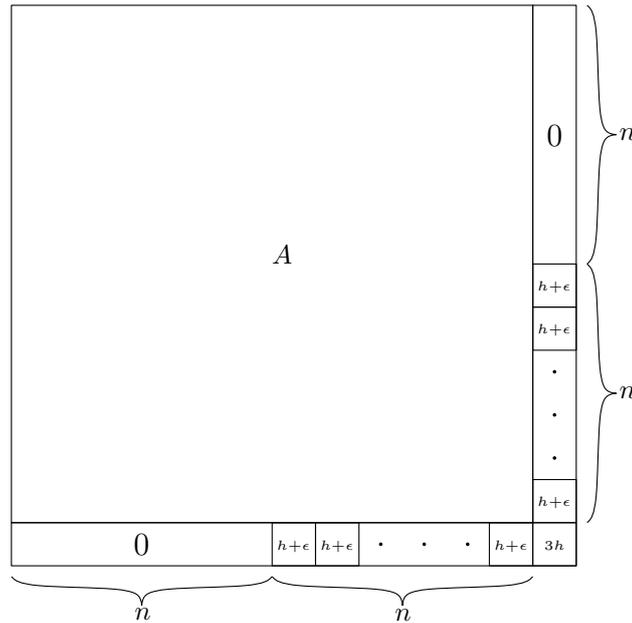
► **Theorem 34.** *Given a symmetric matrix  $A$ ,*

- *it is NP-complete to check if game  $(A, A)$  has a mixed Nash stable (or linearly stable) strategy.*
- *it is NP-complete to check if game  $(A, A)$  has a mixed strict Nash stable strategy.*
- *it is NP-hard to check if dynamics (2) applied on  $A$  has a mixed stable fixed-point.*
- *it is NP-hard to check if dynamics (2) applied on  $A$  has a mixed asymptotically stable fixed-point.*

*even if  $A$  is assumed to be positive.*

As we note in Remark 27, matrix with i.i.d entries from any continuous distribution has in expectation exactly one row with dominating diagonal. One could ask does the problem become easier for this typical case. We answer negatively by extending all the NP-hardness results of Theorem 34 to this case as well. Consider the following modification of matrix  $A$

<sup>11</sup>These are same as (strict) local optima of function  $\pi(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$ , and checking if a given  $\mathbf{p}$  is a local optima can be inconclusive if hessian at  $\mathbf{p}$  is (negative) semi-definite.



■ **Figure 2** Matrix  $M$  as defined in (7).

from (6), where we add an extra row and column. Matrix  $M$  is of dimension  $(2n+1) \times (2n+1)$  (See Figure (2)). Recall that  $h > 2n^2 + 5$  and  $k$  is the given integer.

$$\begin{aligned}
 M_{ij} &= A_{ij} && \text{if } i, j \leq 2n \\
 M_{(2n+1)i} &= M_{i(2n+1)} = 0 && \text{if } i \leq n \\
 M_{(2n+1)i} &= M_{i(2n+1)} = h + \epsilon && \text{if } n < i \leq 2n, \text{ where } 0 < \epsilon < 1 \\
 M_{(2n+1)(2n+1)} &= 3h &&
 \end{aligned}
 \tag{7}$$

Note that  $M$  has exactly one row whose diagonal entry dominates all other entries of the row, *i.e.*,  $\exists i : M_{ii} > M_{ij}, \forall j \neq i$ . See the full version of paper for details, and thus the next theorem holds.

► **Theorem 35.** *Given a symmetric matrix  $M$  such that exactly one row/column in  $M$  has a dominating diagonal,*

- *it is NP-complete to check if game  $(M, M)$  has a mixed Nash stable (or linearly stable) strategy.*
- *it is NP-complete to check if game  $(M, M)$  has a mixed strict Nash stable strategy.*
- *it is NP-hard to check if dynamics (2) applied on  $M$  has a mixed stable fixed-point.*
- *it is NP-hard to check if dynamics (2) applied on  $M$  has a mixed asymptotically stable fixed-point.*

*even if  $M$  is assumed to be positive.*

**Hardness for Subset.** Another natural question is whether a particular allele is going to survive with positive probability in the limit for a given fitness matrix. In full version we show that this may not be easy either, by proving hardness for checking if there exists a stable strategy  $\mathbf{p}$  such that  $i \in SP(\mathbf{p})$  for a given  $i$ . In general, given a subset  $S$  of pure strategies it is hard to check if  $\exists$  a stable profile  $\mathbf{p}$  with  $S \subseteq SP(\mathbf{p})$ .

**Survival of Diversity and Hardness.** As discussed in Section 5 checking if diversity survives in the limiting population of single locus diploid organism reduces to checking “if  $f$  converges to a *mixed* fixed point with positive probability”. In absence of clear characterization of the *mixed* limit points of  $f$  in terms of any of the stability definition, the hardness does not follow directly from the above result. In full version we explain how above results can be combined to obtain the following theorem. Also see the following remark on complexity of decision problem for general Nash equilibrium in coordination games.

► **Theorem 36.** *Given a fitness matrix  $A$  for a diploid organism with single locus, it is NP-hard to decide if, under evolution, diversity will survive (by converging to a specific mixed equilibrium with positive probability) when starting allele frequencies are picked i.i.d from a continuous distribution. Also, deciding if a given allele will survive is NP-hard.*

► **Remark 37.** As noted in Section 4.2, coordination games are very special and they always have a pure Nash equilibrium which is easy to find; NE computation in general game is PPAD-complete [7]. Thus, it is natural to wonder if decision versions on coordination games are also easy to answer.

In the process of obtaining the above hardness results, we stumbled upon NP-hardness for checking if a symmetric coordination game has a NE (not necessarily symmetric) where each player randomizes among at least  $k$  strategies. Again the reduction is from  $k$ -clique. Thus, it seems highly probable that other decision version on (symmetric) coordination games are also NP-complete.

**Acknowledgements.** Part of this work was completed while Ruta Mehta, Ioannis Panageas and Georgios Piliouras were hosted by the Simons Institute for the Theory of Computing.

---

## References

- 1 E. Chastain, A. Livnat, C. H. Papadimitriou, and U. Vazirani. Algorithms, games, and evolution. *PNAS*, 2014. doi:10.1073/pnas.1406556111.
- 2 E. Chastain, A. Livnat, C. H. Papadimitriou, and U. V. Vazirani. Multiplicative updates in coordination games and the theory of evolution. In *ITCS*, pages 57–58, 2013. doi:10.1145/2422436.2422444.
- 3 X. Chen, X. Deng, and S.-H. Teng. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM*, 56(3), 2009.
- 4 V. Conitzer. The exact computational complexity of evolutionarily stable strategies. In *The 9th Conference on Web and Internet Economics (WINE)*, 2013.
- 5 V. Conitzer and T. Sandholm. New complexity results about Nash equilibria. *Games and Economic Behavior*, 63(2):621–641, 2008.
- 6 T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson. *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd edition, 2001.
- 7 C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.
- 8 K. Etessami and A. Lochbihler. The computational complexity of evolutionarily stable strategies. *International Journal of Game Theory*, 37(1):93–113, 2008.
- 9 R. A. Fisher. *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford, 1930.
- 10 J. Garg, R. Mehta, V. V. Vazirani, and S. Yazdanbod. Etr-completeness for decision versions of multi-player (symmetric) nash equilibria. In *ICALP*, 2015.
- 11 I. Gilboa and E. Zemel. Nash and correlated equilibria: Some complexity considerations. *Games Econ. Behav.*, 1:80–93, 1989.
- 12 J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, 1998.

- 13 V. I. Istratescu. *Fixed Point Theory: An Introduction*. Mathematics and Its Applications. Springer Netherlands, 2001. URL: <http://books.google.com/books?id=2bsDnwEACAAJ>.
- 14 H. Kalmus. Adaptive and selective responses of a population of drosophila melanogaster containing e and e+ to differences in temperature, humidity, and to selection for development speed. *Journal of Genetics*, 47:58–63, 1945.
- 15 A. Kawamura. Lipschitz continuous ordinary differential equations are polynomial-space complete. *Computational Complexity*, 19(2):305–332, 2010.
- 16 A. Kawamura, H. Ota, C. Rosnick, and M. Ziegler. Computational complexity of smooth differential equations. In *In Mathematical Foundations of Computer Science*, pages 578–589, 2012.
- 17 A. Kaznatcheev. Complexity of evolutionary equilibria in static fitness landscapes. *arXiv preprint arXiv:1308.5094*, 2013.
- 18 R. Kleinberg, K. Ligett, G. Piliouras, and É. Tardos. Beyond the Nash equilibrium barrier. In *Symposium on Innovations in Computer Science (ICS)*, 2011.
- 19 R. Kleinberg, G. Piliouras, and É. Tardos. Multiplicative updates outperform generic no-regret learning in congestion games. In *STOC*, 2009.
- 20 A. Livnat, C. H. Papadimitriou, J. Dushoff, and M. W. Feldman. A mixability theory for the role of sex in evolution. *PNAS*, 2008. doi:10.1073/pnas.0803596105.
- 21 V. Losert and E. Akin. Dynamics of games and genes: Discrete versus continuous time. *Journal of Mathematical Biology*, 1983.
- 22 Y. Lyubich. *Mathematical Structures in Population Genetics*. Springer-Verlag, 1992.
- 23 R. Mehta, I. Panageas, and G. Piliouras. Natural selection as an inhibitor of genetic diversity: Multiplicative weights updates algorithm and a conjecture of haploid genetics. In *ITCS*, 2015.
- 24 R. Meir and D. Parkes. A note on sex, evolution, and the multiplicative updates algorithm. In *Proceedings of the 12th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 15)*, 2015.
- 25 J. Nash. Equilibrium points in n-person games. *PNAS*, pages 48–49, 1950.
- 26 N. Nisan. A note on the computational hardness of evolutionary stable strategies. *Electronic Colloquium on Computational Complexity (ECCC)*, 13(076), 2006.
- 27 I. Panageas, P. Srivastava, and N. K. Vishnoi. Evolutionary dynamics in finite populations mix rapidly. In *Proc. of the 27th Annual ACM-SIAM Symp. on Discrete Algorithms (SODA '16)*, pages 480–497, 2016. doi:10.1137/1.9781611974331.ch36.
- 28 C. Papadimitriou and G. Piliouras. From nash equilibria to chain recurrent sets: Solution concepts and topology. In *Proc. of the 2016 ACM Conf. on Innovations in Theoretical Computer Science*, ITCS'16, pages 227–235, New York, NY, USA, 2016. ACM. doi:10.1145/2840728.2840757.
- 29 C. H. Papadimitriou and N. K. Vishnoi. On the computational complexity of limit cycles in dynamical systems. In *ITCS*, 2016.
- 30 L. Perko. *Differential Equations and Dynamical Systems*. Springer, 1991.
- 31 G. Piliouras, C. Nieto-Granda, H. I. Christensen, and J. S. Shamma. Persistent patterns: Multi-agent learning beyond equilibrium and utility. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems*, AAMAS'14, pages 181–188, Richland, SC, 2014. International Foundation for Autonomous Agents and Multiagent Systems.
- 32 G. Piliouras and J. S. Shamma. Optimization despite chaos: Convex relaxations to complex limit sets via Poincaré recurrence. In *SODA*, 2014.
- 33 K. C. Rasmus Ibsen-Jensena and M. A. Nowak. Computational complexity of ecological and evolutionary spatial dynamics. In *PNAS*, 2015.

- 34 M. Schaefer and D. Štefankovič. Fixed points, Nash equilibria, and the existential theory of the reals. Manuscript, 2011.
- 35 S.-M. Sun and N. Zhong. Computability aspects for 1st-order partial differential equations via characteristics. *Theoretical Computer Science*, 583:27–39, 2015.
- 36 B. von Stengel. Equilibrium computation for two-player games in strategic and extensive form. *Algorithmic Game Theory*, eds. Nisan, Roughgarden, Tardos, and Vazirani, 2007.
- 37 E. D. Weinberger. Np completeness of kauffman’s nk model, a tuneable rugged fitness landscape. Technical report, Santa Fe Institute, 1996.
- 38 Wikipedia. Sylvester’s law of inertia. URL: [https://en.wikipedia.org/wiki/Sylvester’s\\_law\\_of\\_inertia](https://en.wikipedia.org/wiki/Sylvester’s_law_of_inertia).
- 39 A. H. Wright, R. K. Thompson, and J. Zhang. The computational complexity of nk fitness functions. *IEEE Transactions on Evolutionary Computation*, 4(4):373–379, 2000.

## **A** Terms Used in Biology

### **A.1** Terms in Biology

We provide brief non-technical definitions of a few biological terms that we use in this paper.

**Gene.** A unit that determines some characteristic of the organism, and passes traits to offsprings. All organisms have genes corresponding to various biological traits, some of which are instantly visible, such as eye color or number of limbs, and some of which are not, such as blood type.

**Allele.** Allele is one of a number of alternative forms of the same gene, found at the same place on a chromosome. Different alleles can result in different observable traits, such as different pigmentation.

**Genotype.** The genetic constitution of an individual organism.

**Phenotype.** The set of observable characteristics of an individual resulting from the interaction of its genotype with the environment.

**Diploid.** Diploid means having two copies of each chromosome. Almost all of the cells in the human body are diploid.

**Haploid.** A cell or nucleus having a single set of unpaired chromosomes. Our sex cells (sperm and eggs) are haploid cells that are produced by meiosis. When sex cells unite during fertilization, the haploid cells become a diploid cell.

### **A.2** Heterozygote Advantage (Overdominance)

Cases of heterozygote advantage have been demonstrated in several organisms. The first confirmation of heterozygote advantage was with a fruit fly, *Drosophila melanogaster*. Kalmus demonstrated in a classic paper [14] how polymorphism can persist in a population through heterozygote advantage. In humans, sickle-cell anemia is a genetic disorder caused by the presence of two recessive alleles. Where malaria is common, carrying a single sickle-cell allele (trait) confers a selective advantage, *i.e.*, being a heterozygote is advantageous. Specifically, humans with one of the two alleles of sickle-cell disease exhibit less severe symptoms when infected with malaria.