

Evolving Satisfying Many-Allele Assignments in Richer Fitness Landscapes

Sitan Chen

1 Introduction

One of the central questions in the study of evolution is the role of genetic recombination in facilitating the emergence of complex traits depending on multiple genes. In [4], Livnat et al. showed that evolution which selects for mixability rather than absolute fitness favors sex over asex, while in [5], they showed moreover that recombination gives rise to a hierarchy of evolutionary modules. In this work, we focus on extending another framework for genetic recombination due to Livnat et al., introduced in [6]. There, they explore a simple model of evolution with sexual reproduction and prove that if fit genotypes, represented by bitstrings which satisfy some Boolean “fitness function,” are weakly selected for over unfit genotypes, then the entire population will consist of fit individuals after only polynomially many generations with high probability, provided that the initial average fitness of the population is sufficiently high.

While this result represents an intriguing step towards understanding both the role of sexual reproduction in evolution as well as the emergence of complex traits controlled by multiple genes, their model makes several simplifying assumptions, and in the conclusion of [6], they leave as open the problem of removing some of these assumptions. One of the most obviously unrealistic is the modeling of the fitness function as a Boolean function, which assumes that 1) every gene has exactly one genotype, and 2) every individual is only regarded as “fit” or “unfit.” In this paper, we remove these two assumptions by demonstrating a method of generalizing Livnat et al.’s results to genotypes with many alleles per gene and fitness functions with non-discrete image.

1.1 Livnat et al.’s Model for Weak Selection

We first summarize Livnat et al.’s model for weak selection, which was motivated by Waddington’s famous 1953 experiment on genetic assimilation [3]. They regard every gene as having two alleles so that genotypes are regarded as strings inside $\{\pm 1\}^n$. To model evolution under weak selection, they fix a Boolean fitness function $f : \{\pm 1\}^n \rightarrow \{1, 1 + \epsilon\}$ on the set of possible genotypes which sends the satisfying assignments in $\{\pm 1\}^n$ to $1 + \epsilon$ and everything else to 1, with the ϵ capturing how strongly fit individuals are selected for.

Livnat et al. argue that by Naglyaki’s Theorem [8], under weak selection the population of all possible genotypes quickly takes the form of a *product distribution*, that is, sampling from the population amounts to sampling each gene independently.

As such, define p_i^t to be the population-wide expected value of the i th gene at time t , and let p^t denote the vector $\{p_1^t, \dots, p_n^t\}$. By abuse of notation, we say that x is “sampled from p^t ” to mean that x is sampled from the product distribution whose i th coordinate has expected value p_i^t . Under their model of recombination in which mating occurs randomly, there is no distinction in genders, and generations are non-overlapping, genes “recombine freely” in the sense that for two parents described by bitstrings x and y , the offspring z is such that for any i, j , z_i and z_j come from the same parent with probability $1/2$.

Under this recombination, the distribution of alleles in the next generation is a product distribution where the expected value of the i th allele is given by

$$p_i^{t+1} = \frac{\mathbb{E}_{x \sim p^t}[f(x) \cdot x_i]}{\mathbb{E}_{x \sim p(x)^t}[f]}.$$

If f is monotone, it turns out that we can assume the population is infinite, but otherwise, it turns out that evolution may converge on non-satisfying assignments under this assumption. Indeed, consider $f(x) = x_1 \oplus x_2$ and a population that’s initially just the uniform distribution over $\{-1, 1\}^n$. Under the above dynamics, the distribution of alleles for each gene remains unchanged.

So instead, for a particular true distribution $p^t := \{p_i^t\}_i$, we’ll say that at time t , nature generates a population of N individuals by sampling N bitstrings y^1, \dots, y^N from p^t . Consider the empirical approximation $p^{*t} := \{p_i^{*t}\}_i$ to p^t defined by $p_i^{*t} = \frac{1}{N} \sum_{m=1}^N y_i^m$. We call the step of producing p^{*t} from p^t the **sampling step** in generation t , and we denote the distribution of such p^{*t} by $B(p^t)$. The actual transformation rule is then

$$p_i^{t+1} = \frac{\mathbb{E}_{x \sim p^{*t}}[f(x) \cdot x_i]}{\mathbb{E}_{x \sim p^{*t}}[f(x)]} \tag{1}$$

The step of producing p^{t+1} from p^{*t} is called the **fitness step** in generation t .

Let $\mu^t(f)$ denote the probability that an x drawn from p^t satisfies f , and let $\tilde{f} : [-1, 1]^n \rightarrow [1, 1 + \epsilon]$ denote the multilinear extension given by $\tilde{f}[p] = \mathbb{E}_{x \sim p}[f(x)]$, i.e. the average fitness of the population described by p . Livnat et al.’s key finding was that under weak selection, $\mu^t(f)$ efficiently converges towards 1, or equivalently, $\tilde{f}(p^t)$ quickly converges to $1 + \epsilon$. Specifically, they obtained the following two results.

Theorem 1.1 ([6], Theorem 3.1). *Let $f : \{-1, 1\}^n \rightarrow \{1, 1 + \epsilon\}$ be monotone. Then*

$$\mu^t(f) \geq 1 - \frac{n(1 + \epsilon)}{\epsilon t \mu^0(f)}.$$

Theorem 1.2 ([6], Theorem 6). *Let $f : \{-1, 1\}^n \rightarrow \{1, 1 + \epsilon\}$ and let $\beta = \sqrt{\frac{\epsilon}{N(1 - n\epsilon)}}$. If*

$$\tilde{f}(p^0) > 1 + \sqrt{2\beta \ln(2/\beta)},$$

then there is some constant C such that for any $T \geq C \cdot \frac{\epsilon n^8 \cdot N^4}{1 - n\epsilon}$,

$$\Pr[\tilde{f}(p^T) = 1 + \epsilon] \geq 1 - 2\beta - 2/n.$$

1.2 Our Generalization

We wish to generalize Theorems 1.1 and 1.2 by accounting for 1) richer fitness landscapes and 2) genes with more than just two alleles. In other words, we will study fitness functions with images in the union of continuous intervals $[1 - \delta, 1] \cup [1 + \epsilon, 1 + \epsilon + \delta]$ such that genotypes that are selected against, i.e. which do not satisfy f , take on values in $[1 - \delta]$ rather than just $\{1\}$, while genotypes that are selected for, i.e. which satisfy f , take on values in $[1 + \epsilon, 1 + \epsilon + \delta]$ rather than just $\{1 + \epsilon\}$, and with domains in G_m^n , where $G_m = \{\gamma_1, \dots, \gamma_{m-1}, \gamma_m = 1\}$ is the group of m th roots of unity.

Because the average value for a gene no longer holds enough information to capture the entire distribution of alleles for that gene, to specify how this distribution changes we need to specify multiple transformation rules for simulating genetic recombination. If $D^t = D_1^t \cdots D_n^t$ denotes the distribution over G_m^n of all genotypes in generation t , then for all $i, \ell \in [m]$, define

$$p_{i,\ell}^t = \mathbf{E}_{x_i \sim D_i^t}[x_i^\ell], \quad (2)$$

i.e. the ℓ th moment of the i th gene. If p^t is the matrix of all moments $p_{i,\ell}^t$ at generation t , then when we say that x is “sampled from p^t ” we mean that each x_i is sampled from the distribution over G_m whose ℓ th moment is $p_{i,\ell}^t$ for all ℓ . This distribution is uniquely defined because the $m - 1$ moments, together with the condition that the probabilities for each of the m alleles add to 1, provide m independent linear conditions in m unknowns.

The sampling step is defined completely analogously to that in Section 1.1, i.e. given a true distribution D^t , nature generates a population of N individuals by sampling N bitstrings y^1, \dots, y^N from D^t . The new distribution is described by a new moment matrix p^{*t} given by $p_{i,\ell}^{*t} = \frac{1}{N} \sum_{m=1}^N (y_i^m)^\ell$.

Likewise, the fitness step takes an empirical distribution D^{*t} and sends it to one given by moment matrix p^{t+1} given by

$$p_{i,\ell}^{t+1} = \frac{\mathbf{E}_{x \sim D^{*t}}[f(x) \cdot x_i^\ell]}{\mathbf{E}_{x \sim D^{*t}}[f(x)]}. \quad (3)$$

1.3 Our Results

Henceforth, for the sake of transparency of notation and argument, we will prove our results for the case of $m = 3$, though the approach for higher m appears to be entirely analogous.

Our key technical contribution is as follows: as we discuss more in depth later in the paper, Livnat et al. make crucial use of *biased Fourier analysis* in their proof of Theorems 1.1 and 1.2, but this technique breaks down for $m > 2$. We demonstrate a method of bypassing this by directly considering the *discrete derivative operator* d_i given by

$$d_i f = \frac{1}{m} \cdot \left((m - 1) f^{x_i \leftarrow 1} - \sum_{j=1}^{n-1} f^{x_i \leftarrow \gamma_j} \right).$$

Our other contribution is the neat observation that Livnat et al.’s arguments generalize with very little modification to account for functions with non-discrete image.

Concretely, we obtain the following generalizations of Theorems 1.1 and 1.2.

Theorem 1.3. *Let $f : G_3^n \rightarrow [1 - \delta, 1] \cup [1 + \epsilon, 1 + \epsilon + \delta]$ be monotone. Then*

$$\mu^t(f) \geq 1 - \frac{(2\epsilon + 4\delta)(1 + \epsilon + \delta)n}{\epsilon^{2t}\mu^0(f)}.$$

Theorem 1.4 (informal statement, see Theorem 4.1). *If the variance introduced from sampling at each step exceeds both $\text{poly}(\epsilon)$ and the imaginary part of the change in expectation for each allele during the fitness step, then after T generations of N individuals, the population consists solely of satisfying assignments with probability arbitrarily close to 1, where T and N are polynomial in n , $1/(\epsilon + \delta)$, and $1/\mu^0(f)$.*

Regarding notation, in place of $p_{i,1}^t$ and $p_{i,2}^t$, we will write p_i^t and q_i^t (note that these are conjugates of each other). We will occasionally omit the superscript t in our notation for these quantities in the cases where the index of the generation is irrelevant or clear from context, e.g. writing p'_i and p_i for p_i^{t+1} and p_i^{*t} respectively.

1.4 Organization

In Section 2, we provide a brief summary of discrete Fourier analysis, explain why the biased Fourier analysis used by Livnat et al. breaks down for $m > 2$, and prove some preliminary results about how the transformation law (3) relates to the Fourier expansion of f . In Section 3, we prove Theorem 1.3 to demonstrate our technique of bypassing Livnat et al.’s biased Fourier analysis. In Section 4, we prove Theorem 1.4, restated more precisely in the form of Theorem 4.1 below.

2 Preliminaries

2.1 Discrete Fourier Analysis

The primary technical tool which Livnat et al. used and which we will use is **discrete Fourier analysis**, a technique which has proven highly effective in theoretical computer science, finding applications in pseudorandomness, arithmetic combinatorics, circuit complexity, communication complexity, learning theory, and quantum computing. We begin with a review of Fourier analysis of the Boolean hypercube. The point is to “diagonalize” the space of all Boolean functions, that is, to study Boolean functions by expressing them in terms of some orthonormal basis of functions, the so-called **Fourier basis**.

Definition 2.1. The **characters** of $\{-1, 1\}^n$ are the functions $\chi_S(x) = (-1)^{x \cdot s}$ for $s \in \{-1, 1\}^n$, where $x \cdot s$ denotes the bitwise dot product.

For any function $f : \{-1, 1\}^n \rightarrow \mathbb{C}$, the **Fourier transform** of f is the function $\hat{f} : \{-1, 1\}^n \rightarrow \mathbb{C}$ given by

$$\hat{f}[S] := \mathbb{E}_{x \sim U} [f(x) \cdot \chi_S(x)].$$

We call $\hat{f}[S]$ the *s th Fourier coefficient* of f , and its **weight** is defined to be $|s|$, the number of nonzero bits in s .

The characters form an orthonormal basis for the space of all $f : \{-1, 1\}^n \rightarrow \mathbb{C}$ endowed with the inner product given by $\mathbb{E}_{x \sim U}[f(x)\overline{g(x)}]$. In particular, the **Fourier expansion of f** is

$$f(x) = \sum_S \hat{f}[S] \cdot \chi_S(x).$$

The expectation of f under any distribution X can then be written as

$$\mathbb{E}_{x \sim X}[f(x)] = \sum_S \hat{f}[S] \cdot \mathbb{E}_{x \sim X}[\chi_S(x)].$$

There are two relevant directions in which this can be generalized. Firstly, we can ask what happens if we define our inner product with respect not to the uniform distribution U on $\{-1, 1\}^n$, but to some product distribution where each bit is biased. Let X_p denote the distribution over $\{-1, 1\}$ with expectation p .

Definition 2.2. Let $X = X_{p_1} \times X_{p_2} \times \cdots \times X_{p_n}$. The **biased Fourier characters of $\{-1, 1\}^n$** with respect to the distribution X are the functions $\chi_S^{p_i}(x) = \prod_{i \in S} \frac{x_i - p_i}{\sigma_i}$, where σ_i^2 is the variance $1 - p_i$ (note that for $p_i = 0$, i.e. when the i th gene's distribution is unbiased, $\chi_S^0 = \chi_S$). One can check that this basis is still orthonormal, so it still makes sense to define biased Fourier coefficients, i.e. by $\hat{f}[S, p] = \mathbb{E}_X[f \cdot \phi_S^p]$.

The second direction in which the terms in Definition 2.1 is to larger domains, for example G_m^n for $m > 2$. It is easy to define a Fourier basis with respect to the uniform distribution. For a string $S \in G_m^n$ and $i \in [m]$, let $i(S)$ denote the set of $j \in [n]$ for which $S_j = i$.

Definition 2.3. Let U be the uniform distribution over G_m^n , and let $f : G_m^n \rightarrow \mathbb{C}$. Then the **Fourier characters of G_m^n** with respect to U are the functions $\chi_S(x) = \prod_{\ell \in [m]} \left(\prod_{i \in i(S)} x_i^\ell \right)$. Again, it is easy to check that this basis is orthonormal, so it still makes sense to define Fourier coefficients, i.e. by $\hat{f}[S] = \mathbb{E}_U[f \cdot \phi_S]$.

One useful fact of which we will make extensive use is that if a function $f : G_m^n \rightarrow \mathbb{C}$ is actually real-valued, then its Fourier coefficients come in conjugate pairs.

Observation 2.1. If $f : G_m^n \rightarrow \mathbb{C}$ takes values solely in \mathbb{R} , then for all $S \in G_m^n$, $\hat{f}[S] = \overline{\hat{f}[-S]}$, where $-S$ denotes the string for which $(-S)_i \cdot S_i = 1$.

Proof. Because f is real-valued, $f - \bar{f} \equiv 0$. Furthermore, its Fourier expansion is

$$0 = (f - \bar{f})(x) = \sum_S \left(\hat{f}[S] - \overline{\hat{f}[-S]} \right) \chi_S(x).$$

But by definition of basis, each of the coefficients $\hat{f}[S] - \overline{\hat{f}[-S]}$ must be zero, as desired. \square

2.2 Transformation of Allele Distributions

One of the analytical conveniences (and perhaps motivations) for Livnat's transformation rule, equation (1) is that the change in expected value for each gene during the fitness step can be quantified in terms of the weight-1 biased Fourier coefficients.

One of the first technical obstacles one encounters in passing to more than two alleles per gene is that we no longer have a simple-looking p -biased Fourier product basis to work with, i.e. there is no straightforward modification of the Fourier characters of G_m^n with respect to U to characters with respect to other product distributions.

As an informal demonstration, we could try defining a basis consisting of products of the weight-1 Fourier characters $\phi_{\ell_i}^{p_i}(x) = \frac{x_i - p_i}{\sigma_{\ell_i}}$ for some normalization factor σ_{ℓ_i} , where ℓ_i denotes the string supported at index i such that $(\ell_i)_i = \ell$. The issue is that while $\mathbb{E}[\phi_{\ell_i}^{p_i}] = 0$, $\langle \phi_{\ell_i}^{p_i}, \phi_{\ell'_i}^{p'_i} \rangle \neq 0$. One could also try defining $\phi_{\ell_i}^{p_i}(x) = \frac{(x_i - C)^\ell}{\sigma_{\ell_i}}$ for some C for which $\mathbb{E}[\phi_{\ell_i}^{p_i}] = 0$, but here again, we run into the same issue that these basis elements are not orthogonal. As we discuss later, replacing the biased Fourier analytic arguments of Livnat et al. with arguments which work around such tools forms the core technical difficulty of this work.

The good news is that we can still quantify the change in the moments of each gene during the fitness step in a manner amenable to analysis. Define the functions $\partial_i^t, \bar{\partial}_i^t : G_m^n \rightarrow \mathbb{C}$ by

$$\begin{aligned}\partial_i^t(x) &= \sum_{S:i \in 1(S)} \hat{f}[S] \prod_{j \in 1(S) \setminus \{i\}} x_j \prod_{k \in 2(S)} x_k^2 \\ \bar{\partial}_i^t(x) &= \sum_{S:i \in 2(S)} \hat{f}[S] \prod_{j \in 1(S)} x_j \prod_{k \in 2(S) \setminus \{i\}} x_k^2.\end{aligned}$$

Our substitute for the weight-1 p -biased Fourier coefficients are the quantities $X_i^t = \mathbb{E}_D[\partial_i^t]$ and $Y_i^t = \mathbb{E}_D[\bar{\partial}_i^t]$. Note that by Observation 2.1, $X_i^t = \bar{Y}_i^t$.

Now for convenience, respectively denote p_i^{t+1} and p_i^t by p'_i and p_i , q_i^{t+1} and q_i^t by q'_i and q_i , and D_i^t by D_i .

Lemma 2.1.

$$\begin{aligned}p'_i - p_i &= \frac{1}{\mathbb{E}_D[f]} \left((q_i - p_i^2) \cdot X_i + (1 - p_i q_i) \cdot Y_i \right) \\ q'_i - q_i &= \frac{1}{\mathbb{E}_D[f]} \left((p_i - q_i^2) \cdot Y_i + (1 - p_i q_i) \cdot X_i \right)\end{aligned}$$

Proof. From (3), we know that

$$p'_i - p_i = \frac{\mathbb{E}_{x \sim D}[f(x) \cdot (x_i - p_i)]}{\mathbb{E}_{x \sim D}[f(x)]} = \frac{1}{\mathbb{E}_{x \sim D}[f(x)]} \mathbb{E}_{x \sim D} \left(\left(\sum_S \hat{f}[S] \prod_{j \in 1(S)} x_j \prod_{k \in 2(S)} x_k^2 \right) \cdot (x_i - p_i) \right),$$

and the first half of the claim follows from the observation that $\mathbb{E}_{x_i \sim D_i}[x_i(x_i - p_i)] = q_i - p_i^2$ and $\mathbb{E}_{x_i \sim D_i}[x_i^2(x_i - p_i)] = 1 - p_i q_i$. Taking conjugates on both sides, we immediately get the second half of the claim. \square

These quantities $q_i - p_i^2$ and $1 - p_i q_i$ should be regarded as the variances of x_i and x_i^2 respectively. For convenience, we'll denote them by a_i and b_i . One can quickly verify the following from Lemma 2.1.

Corollary 2.1.

$$\frac{1}{2}(p'_i - p_i + q'_i - q_i) = \frac{\operatorname{Re}((a_i + b_i) \cdot X_i)}{\mathbf{E}_D[f]} \quad \text{and} \quad \frac{1}{2}(p'_i - p_i - q'_i + q_i) = i \cdot \frac{\operatorname{Im}((a_i - b_i) \cdot X_i)}{\mathbf{E}_D[f]}.$$

3 Monotone Functions

As a preliminary result to demonstrate how to bypass the fact that there is no convenient biased Fourier basis to work with by using the quantities X_i^t and Y_i^t defined above, we generalize Livnat et al.'s result on the evolution of populations of satisfying assignments to monotone functions to the case of genes with many alleles and a richer fitness landscape as described previously. Firstly, as we are no longer working in the Boolean case, we need to clarify what we mean by “monotone”:

Definition 3.1. Order the elements in G_m by $\gamma_i < \gamma_j$ iff $i < j$, and order G_m^n by $x \leq y$ iff $x_i \leq y_i$ for all $i \in [n]$. A function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ is **monotone** if $f(x) \leq f(y)$ for all $x \leq y$.

We will now prove the following result on the efficient evolution under weak selection of satisfying assignments to monotone functions (restated from above).

Theorem 3.1. *Let $f : G_3^n \rightarrow [1 - \delta, 1] \cup [1 + \epsilon, 1 + \epsilon + \delta]$ be monotone. Then*

$$\mu^t(f) \geq 1 - \frac{(2\epsilon + 4\delta)(1 + \epsilon + \delta)n}{\epsilon^2 t \mu^0(f)}.$$

Proof. We will show that with each generation, the quantity $\sum_{i \in [n]} \operatorname{Re}(p'_i - p_i) = \frac{1}{2} \sum_{i \in [n]} (p'_i - p_i + q'_i - q_i)$ increases appreciably relative to the variance of f with respect to D . For fixed i , note that

$$p'_i - p_i + q'_i - q_i = \frac{2}{\mathbf{E}_D[f]} \operatorname{Re}((a_i + b_i)X_i). \quad (4)$$

Define the operator d_i by

$$d_i f(x) = \frac{2f^{x_i \leftarrow 1} - f^{x_i \leftarrow \gamma_1} - f^{x_i \leftarrow \gamma_2}}{3}.$$

One should regard this as the higher-dimensional analogue of the discrete derivative operator in the analysis of Boolean functions. By construction, we find that $\partial_i + \bar{\partial}_i = d_i f$, and by monotonicity, $d_i f$ is nonnegative. In particular, $d_i f$ takes on values in the range $[0, \lambda]$ for $\lambda = (2\epsilon + 4\delta)/3$, so

$$\operatorname{Re}((a_i + b_i)X_i) = \frac{1}{2} \mathbf{E}_D[(a_i + b_i)\partial_i + (\bar{a}_i + \bar{b}_i)\bar{\partial}_i] \geq \frac{1}{2\lambda} \mathbf{E}_D[((a_i + b_i)\partial_i + (\bar{a}_i + \bar{b}_i)\bar{\partial}_i) \cdot d_i f]. \quad (5)$$

We would like to relate this to $\text{Var}(f) = \mathbb{E}_D[f^2] - \mathbb{E}_D[f]^2$, and if we take the Fourier expansion of f , we find that

$$\begin{aligned} \text{Var}(f) &= \sum_{S,T} \hat{f}[S] \hat{f}[T] \left(\prod_{S_j+T_j=1,4} p_j \prod_{S_k+T_k=2} q_k - \prod_{S_j+T_j=1,4} q_j^2 \prod_{S_k+T_k=2} p_k^2 \prod_{S_\ell+T_\ell=3} p_\ell q_\ell \right) \\ &\leq \sum_{i \in [n]} \left(\sum_{S: S_i=1} (a_i + b_i) \cdot \left(\sum_{T_i \neq 0} \hat{f}[S] \hat{f}[T] \prod_{\substack{S_j+T_j=1,4 \\ j \neq i}} p_j \prod_{\substack{S_k+T_k=2 \\ k \neq i}} q_k \right) + \right. \\ &\quad \left. \sum_{S: S_i=2} (\bar{a}_i + \bar{b}_i) \cdot \left(\sum_{T_i \neq 0} \hat{f}[S] \hat{f}[T] \prod_{\substack{S_j+T_j=1,4 \\ j \neq i}} p_j \prod_{\substack{S_k+T_k=2 \\ k \neq i}} q_k \right) \right) \end{aligned}$$

But we can rewrite $\mathbb{E}_D[((a_i + b_i)\partial_i + (\bar{a}_i + \bar{b}_i)\bar{\partial}_i) \cdot d_i f]$ using the Fourier expansion of $d_i f$ and taking expectation over D to obtain

$$\begin{aligned} \mathbb{E}_D[((a_i + b_i)\partial_i) \cdot d_i f] &= \mathbb{E}_D \left[\sum_{S: S_i=1} (a_i + b_i) \cdot \left(\sum_{T: T_i \neq 0} \hat{f}[S] \hat{f}[T] \prod_{\ell=1}^4 \left(\prod_{j: S_j+T_j=\ell, j \neq i} x_j^\ell \right) \right) \right] \\ &= \sum_{S: S_i=1} (a_i + b_i) \cdot \left(\sum_{T_i \neq 0} \hat{f}[S] \hat{f}[T] \prod_{\substack{S_j+T_j=1,4 \\ j \neq i}} p_j \prod_{\substack{S_k+T_k=2 \\ k \neq i}} q_k \right), \end{aligned}$$

and we get a similar identity for $\mathbb{E}_D[((\bar{a}_i + \bar{b}_i)\bar{\partial}_i) \cdot d_i f]$, and we conclude that

$$\text{Var}(f) \leq \sum_{i \in [n]} \mathbb{E}_D[((a_i + b_i)\partial_i + (\bar{a}_i + \bar{b}_i)\bar{\partial}_i) \cdot d_i f]$$

and thus, after applying (4) and (5), that

$$\sum_{i \in [n]} (p'_i - p_i + q'_i - q_i) \geq \frac{1}{\lambda \mathbb{E}_D[f]} \text{Var}(f) = \frac{\epsilon^2 \mu(f)(1 - \mu(f))}{\lambda \mathbb{E}_D[f]}$$

Say that at time t , $\mu^t(f) = 1 - \gamma$. Then for all $t' \leq t$,

$$\begin{aligned} \sum_{i \in [n]} (p_i^{t'+1} + q_i^{t'+1}) - \sum_{i \in [n]} (p_i^{t'} + q_i^{t'}) &\geq \frac{\epsilon^2 \gamma \mu^{t'}(f)}{\lambda(1 + \epsilon + \delta)} \\ &\geq \frac{3\epsilon^2}{2\epsilon + 4\delta} \cdot \frac{\gamma \mu^0(f)}{1 + \epsilon + \delta}, \end{aligned} \tag{6}$$

where the second inequality follows from the inductive hypothesis that $\mu^{t'}(f)$ is nondecreasing in t' . Now because the left hand side of (6) is the difference of two sums taking values in $[-n, 2n]$ and thus at most $3n$, so we see that

$$t \leq \frac{(2\epsilon + 4\delta)(1 + \epsilon + \delta)n}{\epsilon^2 \gamma \mu^0(f)} \Rightarrow \mu^t(f) \geq 1 - \frac{(2\epsilon + 4\delta)(1 + \epsilon + \delta)n}{\epsilon^2 t \mu^0(f)}$$

□

Remark 3.1. *This method of bounding variance appears again in the proof of Lemma 4.1 in Section 4.3 where we try to control not the variance of f itself but rather the variance introduced by the sampling step.*

4 General Functions

Having demonstrated how to compensate for our inability to use biased Fourier analysis, we now present our main result which draws upon these techniques. Rather than consider the probability that a genotype drawn from the population satisfies f , we will analyze the population's average fitness, $\mathbb{E}_D[f]$. Specifically, consider the multilinear extension $\tilde{f} : \Delta_3^n \rightarrow [1 - \delta, 1] \cup [1 + \epsilon, 1 + \epsilon + \delta]$, where Δ denotes the set of all $z \in \mathbb{C}$ given by $z = a + b\gamma_1 + c\gamma_2$ for $a + b + c = 1$. Roughly speaking, the following result says that if the population starts off sufficiently fit, the entire population will with high probability consist of genotypes satisfying f after a moderate number of generations. In fact, as we will show, the population converges upon a single satisfying assignment.

4.1 Assumptions and Theorem

The only caveat is that while our abovementioned approach to bypassing any mention of biased Fourier analysis is effective for most of the argument, there are certain technical hurdles that currently offer no immediate solution. For this reason, we will make some relatively reasonable but rather artificial assumptions to make the analysis in Section 4.3 more tractable.

Assumption 1 is that the variance introduced by the sampling step is not too small at each step. Specifically, we will assume that $\mathbb{E}_{p^{*t} \sim B(p^t)}[(\tilde{f}(p^{*t}) - \tilde{f}(p^t))^2] > (\epsilon + 2\delta)^3$. The motivation is that in the original proof by Livnat et al. for the Boolean case, variance from sampling is key for pushing populations close to the corners of the Boolean hypercube into a corner representing a satisfying assignment.

Assumption 2 is that the imaginary part of the change $p_i^{t+1} - p_i^{*t}$ in expected value for any allele over any fitness step is bounded by the variance introduced in the preceding variance step. The motivation for this is a bit more contrived, but as we shall see, the quantity by which we lower-bound the change in average fitness over the fitness step consists of both the real and imaginary parts of this change in expected allele value, and while the former can be handled in a fashion similar to that of Section 3, this approach fails to control the latter. In any case, we note that these two assumptions are only vital for proving Lemma 4.1 below.

Theorem 4.1. *Say that Assumptions 1 and 2 hold. Let $f : G_3^n \rightarrow [1 - \delta, 1] \cup [1 + \epsilon, 1 + \epsilon + \delta]$, and let $\beta = \sqrt{\frac{2\epsilon + 4\delta}{N/2 - 4n^2}}$. If*

$$\tilde{f}(p^0) > 1 + \sqrt{2\beta \ln(2/\beta)},$$

then there is some constant C such that for any $T \geq C \cdot \frac{(\epsilon + 2\delta)N^4 n^8}{1 - n(8\epsilon + 16\delta)/9}$:

$$\Pr[\mu^T(f) = 1] \geq 1 - 2\beta - 2/n.$$

Following the general outline of Livnat et al.'s argument, we argue that the variance introduced from the sampling step is bounded above by the increase in average fitness from the fitness step. We then show that the total loss in fitness from all the sampling steps is negligible by noting that the total effect of the sampling steps carried out prior to any point in time behaves as a martingale and can thus be bounded by an inequality proven in [2]. The conclusion is that average fitness increases appreciably at each step. Unfortunately, this increase is insufficient when the population is near a vertex of Δ_3^n to argue that fitness increases from the fitness steps will make every member of the population a satisfying assignment. Instead, as a last step, we show that genetic drift from the sampling step is responsible for this final push. More precisely, we argue that the sampling steps force populations to eventually get stuck at the vertices of Δ_3^n , so because the probability that the sampling steps decrease fitness significantly is low, with high probability the vertex that the population gets stuck at is a satisfying assignment if the population already starts out as sufficiently fit on average.

4.2 Upper-Bounding Fitness Increase By Formal Derivative

In this and the following subsection, we show that the increase in average fitness from each fitness step is significantly greater than the variance introduced from each sampling step.

As we are concerned with happens over the course of a single generation, we will omit mentions of the index t . We will call the initial distribution in the fitness step D and the resulting distribution D' . Say that the vectors of first and second moments for D are respectively denoted by $\{p_1, \dots, p_n\}$ and $\{q_1, \dots, q_n\}$, and the vectors for D' by $\{p'_1, \dots, p'_n\}$ and $\{q'_1, \dots, q'_n\}$. Our goal is to show the following:

Lemma 4.1. *If Assumptions 1 and 2 above hold, then*

$$\mathbb{E}_{p^{*t} \sim B(p^t)}[(\tilde{f}(p^{*t}) - \tilde{f}(p^t))^2] \leq \mathbb{E}_{p^{*t} \sim B(p^t)}[\tilde{f}(p^{t+1}) - \tilde{f}(p^{*t})]/(N/2 - 4n^2)$$

To that end, in this subsection, we prove the following bound:

Lemma 4.2. *For some $c, c' \in [-(8\epsilon + 16\delta)/9, (8\epsilon + 16\delta)/9]$,*

$$\begin{aligned} \tilde{f}(p') - \tilde{f}(p) \geq & [2\operatorname{Re}(X_i^0) \cdot \operatorname{Re}((a_i + b_i) \cdot X_i^0) - 2\operatorname{Im}(X_i^0) \cdot \operatorname{Im}((a_i - b_i) \cdot X_i^0) + \\ & (i - 1) (\operatorname{Re}((a_i + b_i) \cdot X_i^0)^2 \cdot c - \operatorname{Im}((a_i - b_i) \cdot X_i^0)^2 \cdot c')] \cdot \frac{1}{1 + \epsilon + \delta} \end{aligned}$$

Livnat et al.'s first step is to express the change in average fitness between two such intermediate distributions in terms of the partial derivative of average fitness with respect to average value of the particular gene whose distribution changed. While it no longer makes sense to differentiate average fitness with respect to (complex-valued) average allele value p_i in the real analytic sense, we can still *formally* differentiate \tilde{f} with respect to this quantity, giving us the quantities X_i, Y_i defined in the previous section.

Proof. We will proceed by a hybrid argument. Define p^i to be the vector of first moments $\{p'_1, \dots, p'_i, p_{i+1}, \dots, p_n\}$ so that the first moments for D and D' are respectively given by p^0 and p^n .

First note that $\tilde{f}(p) = \sum_S \hat{f}[S] \prod_{j \in 1(S)} p_j \prod_{k \in 2(S)} q_k$ so that

$$\begin{aligned}
\tilde{f}(p^i) - \tilde{f}(p^{i-1}) &= (p'_i - p_i) \sum_{S:i \in 1(S)} \hat{f}[S] \prod_{j \in 1(S) \setminus \{i\}} p_j \prod_{k \in 2(S)} q_k + (q'_i - q_i) \sum_{S:i \in 2(S)} \hat{f}[S] \prod_{j \in 1(S)} p_j \prod_{k \in 2(S) \setminus \{i\}} q_k \\
&= (p'_i - p_i) X_i^{i-1} + (q'_i - q_i) Y_i^{i-1} \\
&= \frac{1}{2} (p'_i - p_i + q'_i - q_i) (X_i^{i-1} + Y_i^{i-1}) + \frac{1}{2} (p'_i - p_i - q'_i + q_i) (X_i^{i-1} - Y_i^{i-1}) \\
&= \frac{\operatorname{Re}((a_i + b_i) \cdot X_i^0)}{\mathbb{E}_D[f]} \cdot (X_i^{i-1} + Y_i^{i-1}) + i \cdot \frac{\operatorname{Im}((a_i - b_i) \cdot X_i^0)}{\mathbb{E}_D[f]} \cdot (X_i^{i-1} - Y_i^{i-1})
\end{aligned}$$

As clarification, the superscripts in X_i^{i-1} and Y_i^{i-1} denote not the index of the generation but the index of the distribution inside the collection of hybrid distributions between D and D' .

Write X_i^{i-1} and Y_i^{i-1} respectively as $X_i^0 + \sum_{\ell=1}^{i-1} (X_i^\ell - X_i^{\ell-1})$ and $Y_i^0 + \sum_{\ell=1}^{i-1} (Y_i^\ell - Y_i^{\ell-1})$. Our goal now is to bound $B_i^\ell := (X_i^\ell - X_i^{\ell-1}) + (Y_i^\ell - Y_i^{\ell-1})$ and $C_i^\ell := (X_i^\ell - X_i^{\ell-1}) - (Y_i^\ell - Y_i^{\ell-1})$. Indeed,

$$X_i^\ell - X_i^{\ell-1} = (p'_\ell - p_\ell) \cdot A_{11}^\ell + (q'_\ell - q_\ell) \cdot A_{12}^\ell$$

where

$$\begin{aligned}
A_{11}^\ell &= \left(\sum_{S:S_i=S_\ell=1} \hat{f}[S] \prod_{j \in 1(S) \setminus \{i, \ell\}} p_j^\ell \prod_{k \in 2(S)} q_k^\ell \right) \\
A_{12}^\ell &= \left(\sum_{S:S_i=1, S_\ell=2} \hat{f}[S] \prod_{j \in 1(S)} p_j^\ell \prod_{k \in 2(S) \setminus \{\ell\}} q_k^\ell \right).
\end{aligned}$$

We can define A_{21}^ℓ and A_{22}^ℓ analogously and get that

$$Y_i^\ell - Y_i^{\ell-1} = (p'_\ell - p_\ell) \cdot A_{21}^\ell + (q'_\ell - q_\ell) \cdot A_{22}^\ell.$$

Then

$$\begin{aligned}
B_i^\ell &= (p'_\ell - p_\ell + q'_\ell - q_\ell) \cdot (A_{11}^\ell + A_{21}^\ell + A_{12}^\ell + A_{22}^\ell)/2 + \\
&\quad (p'_\ell - p_\ell - q'_\ell + q_\ell) \cdot (A_{11}^\ell + A_{21}^\ell - A_{12}^\ell - A_{22}^\ell)/2 \\
&\in \frac{1}{\mathbb{E}_D[f]} \left(\operatorname{Re}((a_\ell + b_\ell) \cdot X_\ell^0) \cdot [-\lambda_1, \lambda_1] - \operatorname{Im}((a_\ell - b_\ell) \cdot X_\ell^0) \cdot [-\lambda_2, \lambda_2] \right),
\end{aligned}$$

for $\lambda_1 = (8\epsilon + 16\delta)/9$ and $\lambda_2 = (4\epsilon + 8\delta)/(3\sqrt{3})$. The last inclusion follows from both the observation that

$$\begin{aligned}
A_{11}^\ell + A_{21}^\ell - A_{12}^\ell - A_{22}^\ell &= E_{p^\ell}[\delta_i \delta_\ell f] \in [-\lambda_1, \lambda_1] \\
A_{11}^\ell + A_{21}^\ell - A_{12}^\ell - A_{22}^\ell &= \frac{1}{\lambda_1 - \lambda_2} E_{p^\ell}[(\delta_i f)^{x_\ell \leftarrow \gamma_1} - (\delta_i f)^{x_\ell \leftarrow \gamma_2}] \in i \cdot [-\lambda_2, \lambda_2]
\end{aligned}$$

and Corollary 2.1. Similarly, we can conclude that

$$C_i^\ell \in \frac{1}{\mathbb{E}_D[f]} \left(\operatorname{Im}((a_\ell - b_\ell) X_\ell^0) \cdot [-\lambda_1, \lambda_1] - \operatorname{Re}((a_\ell + b_\ell) \cdot X_\ell^0) \cdot [-\lambda_2, \lambda_2] \right).$$

Without loss of generality, we can assume that $\frac{\operatorname{Re}((a_i+b_i) \cdot X_i^0)}{\mathbb{E}_D[f]} \cdot B_i^\ell + i \cdot \frac{\operatorname{Im}((a_i-b_i) \cdot X_i^0)}{\mathbb{E}_D[f]} \cdot C_i^\ell$ is minimized over $\ell \leq i$ by $\ell = i$ so that

$$\begin{aligned} \tilde{f}(p^i) - \tilde{f}(p^{i-1}) &\geq \left[\frac{\operatorname{Re}((a_i+b_i) \cdot X_i^0)}{\mathbb{E}_D[f]} \cdot (2\operatorname{Re}(X_i^0) + (i-1) \cdot B_i^i) - \right. \\ &\quad \left. \frac{\operatorname{Im}((a_i-b_i) \cdot X_i^0)}{\mathbb{E}_D[f]} \cdot (2\operatorname{Im}(X_i^0) + (i-1) \cdot C_i^i) \right] \cdot \frac{1}{\mathbb{E}_D[f]} \\ &\geq [2\operatorname{Re}(X_i^0) \cdot \operatorname{Re}((a_i+b_i) \cdot X_i^0) - 2\operatorname{Im}(X_i^0) \cdot \operatorname{Im}((a_i-b_i) \cdot X_i^0) + \\ &\quad \operatorname{Re}((a_i+b_i) \cdot X_i^0)^2 \cdot [-\lambda_1, \lambda_1] \cdot (i-1) - \\ &\quad \operatorname{Im}((a_i-b_i) \cdot X_i^0)^2 \cdot [-\lambda_1, \lambda_1] \cdot (i-1)] \cdot \frac{1}{1+\epsilon+\delta} \end{aligned} \quad (7)$$

□

4.3 Lower-Bounding Formal Derivative by Sampling Variance

We complete the proof of Lemma 4.1 that the increase in average fitness can be bounded from below in terms of the variance introduced in the sampling step. Denote the distribution after the fitness step by D , and let its vectors of first and second moments be $p = \{p_1, \dots, p_n\}$ and $q = \{q_1, \dots, q_n\}$ respectively, and from the distribution B of approximations of D , we will draw some E described by vectors of first and second moments of $p^* = \{p_1^*, \dots, p_n^*\}$ and $q^* = \{q_1^*, \dots, q_n^*\}$. We will first compute the variance introduced by sampling, that is, the quantity $V := \mathbb{E}_{p^* \sim B}[(\tilde{f}(p^*) - \tilde{f}(p))^2]$.

Similar to our computation of the variance in the case of f monotone, we find that

$$\begin{aligned} \mathbb{E}_{p^* \sim B}[\tilde{f}(p^*)^2] &= \mathbb{E}_{p^* \sim B} \left[\sum_{S,T} \hat{f}[S] \hat{f}[T] \prod_{j \in 1(S\Delta T)} p_j^* \prod_{k \in 2(S\Delta T)} q_k^* \prod_{S_j=1=T_j} p_j^{*2} \prod_{S_k=2=T_k} q_k^{*2} \prod_{S_j+T_j=3} p_j^* q_j^* \right] \\ &= \sum_{S,T} \hat{f}[S] \hat{f}[T] \prod_{j \in 1(S\Delta T)} p_j \prod_{k \in 2(S\Delta T)} q_k \prod_{S_j=1=T_j} \left(p_j^2 + \frac{q_j - p_j^2}{N} \right) \\ &\quad \prod_{S_k=2=T_k} \left(q_k^2 + \frac{p_k - q_k^2}{N} \right) \prod_{S_j+T_j=3} \left(p_j q_j + \frac{1 - p_j q_j}{N} \right) \end{aligned}$$

so that the variance $\mathbb{E}_{p^* \sim B}[\tilde{f}(p^*)^2] - 2\mathbb{E}_{p^* \sim B}[\tilde{f}(p^*)] \cdot \tilde{f}(p) + \tilde{f}(p)^2 = \mathbb{E}_{p^* \sim B}[\tilde{f}(p^*)^2] - \tilde{f}(p)^2$ is bounded above, using the argument of section 3, by

$$\sum_{i \in [n]} 4\operatorname{Re} \left(\left(\frac{a_i + b_i}{N} \right) X_i^0 \right) \operatorname{Re}(X_i^0).$$

We now consider the other components of (7).

By our assumption that $|\operatorname{Im}(p'_i - p_i)| = \operatorname{Im}((a_i - b_i) \cdot X_i) / \mathbb{E}_D[f] < V$ and the fact that

$$|\operatorname{Im}(X_i^0)| = \left| \frac{X_i^0 - Y_i^0}{2i} \right| = \left| \frac{f^{x_i \leftarrow \gamma_1} - f^{x_i \leftarrow \gamma_2}}{2i(\gamma_1 - \gamma_2)} \right| \leq \frac{\epsilon + 2\delta}{2\sqrt{3}} := \lambda_3,$$

the quantity

$$(-2 \operatorname{Im}(X_i^0) \cdot \operatorname{Im}((a_i - b_i) \cdot X_i^0) - \operatorname{Im}((a_i - b_i) \cdot X_i^0)^2 \cdot [-\lambda_1, \lambda_1] \cdot (i - 1)) \cdot \frac{1}{\mathbf{E}_D[f]}$$

in (7) is at least $-V(n\lambda_1(1 + \epsilon + \delta) + 2\lambda_3)$.

Lastly,

$$|\operatorname{Re}(X_i^0)| = \left| \frac{X_i^0 + Y_i^0}{2} \right| = \left| \frac{2f^{x_i \leftarrow -1} - f^{x_i \leftarrow \gamma_1} - f^{x_i \leftarrow \gamma_2}}{3} \right| \leq \frac{2\epsilon + 4\delta}{3}$$

so that $|\operatorname{Re}((a_i + b_i) \cdot X_i^0)| \leq 2|\operatorname{Re}(X_i^0)| + 2|\operatorname{Im}(X_i^0)| < 2\epsilon + 4\delta$; by our assumption that $V > (\epsilon + 2\delta)^3$, we see that

$$\operatorname{Re}((a_i + b_i) \cdot X_i^0)^2 \cdot [-\lambda_1, \lambda_1] \cdot (i - 1) \geq -32nV/9.$$

The upshot of all of this is that

$$\sum_{i \in [n]} \tilde{f}(p^i) - \tilde{f}(p^{i-1}) \geq N/2 - n(n\lambda_1(1 + \epsilon + \delta) + 2\lambda_3 + 32n/9)V > (N/2 - 4n^2)V,$$

i.e. the variance introduced from sampling is low relative to the increase in average fitness.

4.4 Total Loss from Sampling

The result in the previous subsection will allow us to show that the total loss in fitness caused by the sampling step is low. The proof of this result is essentially identical to the one in Livnat et al., but we include it for the sake of completeness.

They make use of the following martingale inequality:

Theorem 4.2 ([2], Theorem 3.3). *For ζ_1, ζ_2, \dots a martingale difference sequence with respect to a filtration $\{\mathcal{F}_t\}$, define $S_T = \sum_{t=1}^T \zeta_t$ and $H_T = \sum \zeta_t^2 + \sum \mathbf{E}[\zeta_t^2 \mid \mathcal{F}_{t-1}]$. Then for all $\tau \in \mathbb{N}$,*

$$\Pr \left[\max_{T \leq \tau} |S_T| > z, H_\tau \leq L \right] \leq 2 \cdot \exp \left(\frac{-z^2}{2L} \right).$$

The main result of this subsection is the following:

Lemma 4.3. *Let $\beta = \sqrt{\frac{2\epsilon + 4\delta}{N/2 - 4n^2}}$ and $\alpha = \sqrt{2\beta \ln(2/\beta)}$. Then*

$$\Pr \left[\left| \sum_{t=1}^T \tilde{f}(p^{*t}) - \tilde{f}(p^{t-1}) \right| \geq \alpha \right] \leq 2\beta.$$

Proof. To define our filtration $\{\mathcal{F}_t\}$, for each sequence $(p^t)_{t=0}^T$ of populations up to generation T , define its congruence class to be the collection of all its infinite continuations $(p^{*t})_{t=0}^\infty$, i.e. all sequences for which $p^{*t} = p^t$ for all $t \leq T$. If we let \mathcal{F}_T denote the collection of all such congruence classes at time T , we get a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots$.

Define

$$S_T = \sum_{t=0}^T \zeta_t := \sum_{t=0}^T \tilde{f}(p^{*t}) - \tilde{f}(p^{t-1}).$$

This is certainly a martingale, because $\tilde{f}(p^{*T} - p^{T-1})$ does not depend on the sequence of populations leading up to time $T - 1$, meaning

$$\mathbb{E}[S_T \mid \mathcal{F}_{T-1}] = S_{T-1} + \mathbb{E}[\tilde{f}(p^{*T}) - \tilde{f}(p^{T-1}) \mid \mathcal{F}_{T-1}] = S_{T-1}.$$

We want to apply Theorem 4.2 to this martingale, so the corresponding H_T will be

$$H_T = \sum_{t=0}^T \left(\tilde{f}(p^{*t}) - \tilde{f}(p^{t-1}) \right)^2 + \sum_{t=0}^T \mathbb{E} \left[\left(\tilde{f}(p^{*t}) - \tilde{f}(p^{t-1}) \right)^2 \mid \mathcal{F}_{t-1} \right] := M_T + V_T.$$

But by Theorem 4.2,

$$\Pr \left[\max_{T \leq \tau} |S_T| \geq \alpha, H_T \leq \beta \right] \leq 2 \exp(-\alpha^2/2\beta) \leq \beta,$$

so we're done by union bound if we can show that $\Pr[H_T \geq \beta] \leq \beta$. To do this, we will show that H_T is small in expectation and then apply Markov's.

From Lemma 4.1, we know that $\mathbb{E}[M_T]$ is bounded above by $\frac{1}{N/2-4n^2}$ times the expected total change in average fitness over all fitness steps. But because the expected total change over all fitness steps and sampling steps is merely $\mathbb{E}_T \left[\tilde{f}(p^{T+1}) - \tilde{f}(p^{*T}) \right] \leq \epsilon + 2\delta$, and the expected total change over all sampling steps is obviously 0,

$$\mathbb{E}[M_T] \leq \frac{\epsilon + 2\delta}{N/2 - 4n^2}.$$

The same use of Lemma 4.1 gives us that $\mathbb{E}[S_T] \leq (\epsilon + 2\delta)/(N/2 - 4n^2)$, so it follows that $\mathbb{E}[H_T] \leq (2\epsilon + 4\delta)/(N/2 - 4n^2) \leq \beta^2$ and thus that $\Pr[H_T \geq \beta] \leq \beta$. \square

4.5 Ending Up at Corners

To conclude the proof of our main theorem, we need to show that after sufficiently many generations, the population D^T lies on a vertex of G_m^n with high probability, implying that either $\tilde{f}(p^T) \in [1 - \delta, 1]$ or $\tilde{f}(p^T) \in [1 + \epsilon, 1 + \epsilon + \delta]$. Lemma 4.3 then tells us that with high probability the total drop in fitness over all the sampling steps cannot be low enough that the former holds. The outline of this proof is also essentially identical to that of Livnat et al., but we present the full argument here again for the sake of completeness.

Lemma 4.4. *There is a constant $C > 0$ such that for any $T \geq C \cdot \frac{(\epsilon+2\delta)N^4n^8}{1-n\lambda_1}$, we have $\Pr[D^T \notin G_m^n] < 2/n$.*

Proof. We first prove rigorously our earlier informal claim that the genetic drift due to sampling is key to proving convergence to a population of satisfying assignments. Certainly if $|p_i^t| = 1$ for some t , then $|p_i^{*t'}| = 1$ and $|p_i^{t'}| = 1$ for all $t' \geq t$. But in fact, if $|p_i^t| > 1 - 1/(n^2N)$, then

$\Pr[|p_i^{*t}| < 1] \leq 1/n^2$ by Markov's inequality; in this case, we say that i is “ α -determined” for $\alpha = 1/(n^2N)$.

Let $A_{i;t}$ denote the event that i is not α -determined for any of the first t generations. The goal is to show that $\Pr[\bigvee_{i=1}^n A_{i;T}] \leq 1/n$ for T set to be the value in the hypothesis of this lemma. To this end, we show that $\Pr[A_{i;T}] \leq 1/n^2$ for all i .

We first make the following two elementary observations, roughly speaking that 1) $\Pr[A_{i;T}]$ is small relative to the total change in p_i due to the sampling steps, and 2) the total change in p_i due to the sampling steps is small relative to the total change due to the fitness steps.

Proposition 4.1. *For any time t_0 and any length of time T_1 for which $t_0 + T_1 \leq T$, we have that*

$$\mathbb{E} \left[\left| \sum_{t=t_0}^{t_0+T_1} p_i^{*t} - p_i^t \right|^2 \right] \geq \frac{\alpha \cdot \Pr[A_{i;T}] T_1}{2N}.$$

Proof. Because $p_i^{*t'} - p_i^{t'}$ has expectation 0 for all $t' > t$ regardless of any information about generation t , we can rewrite the left-hand side of the desired inequality as

$$\sum_{t=t_0}^{t_0+T_1} \mathbb{E}_{p^{*t} \sim B(p^t)} [|p_i^t - p_i^{*t}|^2] \geq \sum_{t=t_0}^{t_0+T_1} \Pr[A_{i;t}] \mathbb{E}_{p^{*t} \sim B(p^t)} [|p_i^{*t} - p_i^t|^2 \mid A_{i;t}],$$

but because $\mathbb{E}_{p^{*t} \sim B(p^t)} [|p_i^{*t} - p_i^t|^2 \mid A_{i;t}] = \frac{|q_i^t| - |p_i^t|^2}{N} \geq \frac{\alpha}{2N}$ and $\Pr[A_{i;t}] \leq \Pr[A_{i;t'}]$ for $t \geq t'$, we get the desired inequality. \square

Proposition 4.2. *With t_0, T_1 as above,*

$$\left(\sum_{t=t_0}^{t_0+T_1} p_i^{t+1} - p_i^{*t} \right)^2 + 3 \geq \frac{1}{2} \left(\sum_{t=t_0}^{t_0+T_1} p_i^{*t} - p_i^t \right)^2.$$

Proof. Because

$$p_i^{t_0+T_1+1} - p_i^{t_0} = \left(\sum_{t=t_0}^{t_0+T_1} p_i^{*t} - p_i^t \right) + \left(\sum_{t=t_0}^{t_0+T_1} p_i^{t+1} - p_i^{*t} \right)$$

and the left-hand side has modulus bounded by $\sqrt{3}$, it follows by the triangle inequality that

$$\left| \sum_{t=t_0}^{t_0+T_1} p_i^{*t} - p_i^t \right| \leq \left| \sum_{t=t_0}^{t_0+T_1} p_i^{t+1} - p_i^{*t} \right| + \sqrt{3}.$$

Squaring both sides, we get the desired inequality from the fact that $2|z|^2 + 6 \geq (|z| + \sqrt{3})^2$. \square

Combining the above two propositions, we obtain that

$$\mathbb{E} \left[\left| \sum_{t=t_0}^{t_0+T_1} p_i^{t+1} - p_i^{*t} \right|^2 \right] \geq \frac{\alpha \Pr[A_{i;T}] T_1}{4N} - 3.$$

We want to show that the quantity on the left-hand side is small, so we will relate it to change in fitness during the fitness step. By Cauchy-Schwarz, we have that

$$\left| \sum_{t=t_0}^{t_0+T_1} p_i^{t+1} - p_i^{*t} \right|^2 \leq T_1 \sum_{t=t_0}^{t_0+T_1} |p_i^{t+1} - p_i^{*t}|^2.$$

But from our analysis in Section 4.2, recall that

$$\begin{aligned} \tilde{f}(p^{t+1}) - \tilde{f}(p^{*t}) &\geq (p_i^{t+1} - p_i^{*t}) (X_i^0 + (i-1)(A_{11}^i(p_i^{t+1} - p_i^{*t}) + A_{12}^i(q_i^{t+1} - q_i^{*t}))) + \\ &\quad (q_i^{t+1} - q_i^{*t}) (Y_i^0 + (i-1)(A_{21}^i(p_i^{t+1} - p_i^{*t}) + A_{22}^i(q_i^{t+1} - q_i^{*t}))), \end{aligned}$$

and from our bound on $A_{11}^i + A_{12}^i + A_{21}^i + A_{22}^i$ one can check that this is at least

$$(p_i^{t+1} - p_i^{*t})X_i^0 + (q_i^{t+1} - q_i^{*t})Y_i^0 - n\lambda_1|p_i^{t+1} - p_i^{*t}|^2 \geq (1 - n\lambda_1)|p_i^{t+1} - p_i^{*t}|^2,$$

where the inequality follows from Lemma 2.1. We deduce that

$$\left| \sum_{t=t_0}^{t_0+T_1} p_i^{t+1} - p_i^{*t} \right|^2 \leq \frac{T_1}{1 - n\lambda_1} \sum_{t=t_0}^{t_0+T_1} \tilde{f}(p^{t+1}) - \tilde{f}(p_i^{*t}).$$

Taking expectations on both sides and using (4.5), we get that

$$\frac{\alpha \Pr[A_{i;T}]}{4N} - 4/T_1 \leq \frac{1}{1 - n\lambda_1} \mathbb{E} \left[\sum_{t=t_0}^{t_0+T_1} \tilde{f}(p^{t+1}) - \tilde{f}(p_i^{*t}) \right],$$

and applying this bound T_2 times yields

$$\begin{aligned} T_2 \left(\frac{\alpha \Pr[A_{i;T}]}{4N} - 4/T_1 \right) &\leq \sum_{\ell=0}^{T_2-1} \mathbb{E} \left[\sum_{t=\ell T_1}^{(\ell+1)T_1} \tilde{f}(p_i^{t+1}) - \tilde{f}(p_i^{*t}) \right] \cdot \frac{1}{1 - n\lambda_1} \\ &= \mathbb{E}[\tilde{f}(p_i^{T_1 T_2 + 1}) - \tilde{f}(p_i^{*0})] \cdot \frac{1}{1 - n\lambda_1} \\ &\leq \frac{\epsilon + 2\delta}{1 - n\lambda_1}. \end{aligned}$$

Finally, we have

$$\Pr[A_{i;T}] \leq \frac{4N(\epsilon + 2\delta)}{\alpha T_2(1 - n\lambda_1)} + \frac{16N}{\alpha T_1} = \frac{4n^2 N^2(\epsilon + 2\delta)}{T_2(1 - n\lambda_1)} + \frac{16n^2 N^2}{T_1},$$

and if we pick $T_1 = 32N^2 n^4$ and $T_2 = \frac{8(\epsilon+2\delta)N^2 n^4}{1-n\lambda_1}$, $\Pr[A_{i;T}] \leq 1/n^2$ as desired. \square

We can now finish off the proof of our main theorem. Lemma 4.4 tells us that the population reaches a corner of G_m^n with probability $1 - 2n$ after T steps, meaning the average fitness lies either in $[1 - \delta, 1]$ or $[1 + \epsilon, 1 + \epsilon + \delta]$. But Lemma 4.3 tells us that if the initial average fitness of the population exceeds $1 + \alpha$, then only with probability 2β will the sampling steps decrease fitness to 1 or less, so we get the desired result.

5 Discussion

Admittedly, in dealing specifically with the case of $m = 3$, keeping track of the matrix of all moments in generation t was easier than for general m , because rather than keep track of a matrix of first and second moments for each allele, it sufficed to keep track of the vector $p^t = \{p_1^t, \dots, p_n^t\}$ of first moments of each allele. That said, this seems more like a notational convenience than anything, and to the author there seems little doubt that the arguments above, given a suitable adaption of Assumptions 1 and 2, apply to all m rather than just $m = 3$. Of course, this is a worthy immediate next step to the work in this paper, as is attempting to refine the analysis to get rid of our rather artificial Assumptions 1 and 2.

In the discussion of [6], Livnat et al. also raise the question of why it seems that the fitness gap ϵ must be below a certain threshold in order for Theorem 1.2 to be meaningful, the intuition being that perhaps because the fitness steps are essentially ϵ -long steps of the gradient ascent of \tilde{f} , at least away from the boundary of Δ_m^n , making convergence of the gradient ascent harder to verify for overly large step sizes. Determining whether Theorem 4.1 still holds for moderately larger ϵ thus also represents an interesting open problem to consider.

Another direction of further pursuit is to modify this model to account for speciation, i.e. to allow for equilibria in which members of the population are not all the same satisfying assignment. To this end, it might be worthwhile to look at multiple fitness functions and a modified transformation law which takes into account an allele's performance across all fitness functions. One could then ask whether, given a sufficiently heterogeneous initial population, will the population no longer converge to a single vertex of Δ_m^n ?

References

- [1] E. Chastain, A. Livnat, C. Papadimitriou, and Umesh Vazirani (2014) Algorithms, games, and evolution. *Proc Natl Acad Sci USA* 111(29): 10620-10623.
- [2] K. Dzhaparidze and J.H. van Zanten. On Bernstein-type inequalities for martingales. *Stochastic Processes and their Applications*, 93(1):109–117, 2001.
- [3] R. Lewontin. The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics*, 49:49-67, 1964.
- [4] A. Livnat, C. Papadimitriou, J. Dushoff, M.W. Feldman (2008) A mixability theory for the role of sex in evolution. *Proc Natl Acad Sci USA* 105(50): 19803-19808.
- [5] A. Livnat, C. Papadimitriou, N. Pippenger, M.W. Feldman (2010) Sex, mixability, and modularity. *P. Natl. Acad. Sci. USA* 107(4): 1452-1457.
- [6] A. Livnat, C. Papadimitriou, A. Rubinfeld, A. Wan, and Gregory Valiant. Satisfiability and evolution. *In FOCS*, 2014.
- [7] R. O'Donnell. Some topics in analysis of boolean functions. *Proc. STOC* 2008, 569-578, 2008.
- [8] T. Naglyaki (1993) The evolution of multilocus systems under weak selection. *Genetics* 134(2): 627-647.