Optimization as a Technique for Studying Population Genetics Equations

Alan Hastings\(^1\) and Gordon A. Fox\(^2\)

\(^1\) Division of Environmental Studies, Center for Population Biology, and Institute for Theoretical Dynamics University of California Davis, CA. 95616
Email: amhastings@ucdavis.edu

\(^2\) Department of Ecology and Evolutionary Biology University of Arizona Tucson, AZ 85721

Abstract. We use methods from dynamic optimization to study the possible behavior of simple population genetic models. These methods can be used, at least conceptually, to determine limits to the behavior of optimization algorithms based on genetic equations.

1 Introduction

The primary focus of this book is to look at how to use the equations of population genetics to study and understand problems in optimization. Most of the rationale for using ideas borrowed from natural selection to solve problems in optimization comes from Fisher's fundamental theorem. Unfortunately, it is well known that Fisher's result applies only to random-mating, single-locus population genetic models with constant selection (see for example, Ewens and Hastings, 1995). Multilocus population genetic models are complicated nonlinear dynamic equations. The dynamics and the equilibrium behavior of these multilocus equations are not well understood, except for some special cases. In this chapter, we will describe approaches for trying to understand bounds to the behavior of these equations by using optimization methods. This work may, in turn, provide some insights on the performance of methods that use genetic equations to solve optimization problems.

We will summarize two primary approaches: equilibrium behavior of the two-locus models, and dynamics of two-locus models. In both cases, the approach has been to use optimization methods to find limits to the behavior of the equations (Hastings, 1981; Fox and Hastings, 1992) for fitnesses that are only known within some bounds. One reason for this is that in population genetics the fitnesses are not well specified. Thus the fitnesses are treated either as the unknowns or as parameters to be determined.

To place these approaches in a larger context, we will begin by examining the simpler one-locus population genetic equations. This will provide background
and motivation for the methods we will discuss for studying the multilocus equations. Moreover, the single-locus viewpoint, in combination with the multilocus results, will help illustrate the role played by recombination and linkage disequilibrium in the dynamics and equilibrium behavior of multilocus population genetic equations.

2 Single Locus Population Genetic Models

Here, we will start with the simplest case, a single locus with two alleles. Let the alleles be A and a, and denote the frequency of A by p and the frequency of a by q. We will begin with a description of the deterministic discrete time model with random mating and nonoverlapping generations. Let the fitness of the genotypes AA, Aa, and aa be denoted by $w_{AA}, w_{Aa},$ and $w_{aa},$ respectively. We define the average fitness of the allele A as

$$w_A = p w_{AA} + q w_{Aa}$$

(1)

and the average fitness of the population as

$$\bar{w} = p w_A + q w_a$$

(2)

The dynamics of the allele frequencies are then given by the equation

$$p' = p w_A / \bar{w},$$

(3)

where $p'$ is the allele frequency in the next generation.

The equilibrium behavior of this model is easy to analyze. The usual approach is to view the fitnesses as parameters and then solve for the equilibrium value of $p = p'$ (e.g., Ewens, 1979; Nagylaki, 1992). If one could readily estimate the fitnesses, this would make it easy to predict the evolution of gene frequencies in natural populations. However, it is much easier to measure allele frequencies than it is to estimate fitnesses in natural populations. Attractive as this approach may be, then, it is usually impossible to implement in practice.

So here we will use an alternate approach (Hastings, 1981): view the equilibrium allele frequency as the parameter, and the fitnesses as unknowns. Doing so allows us to find values for the fitnesses that may explain the observed allele frequencies.

2.1 Equilibria

In the simple one-locus, two allele case, this alternate approach leads to a single linear equation with two unknowns. To see this, note that only the relative fitnesses are important, thus reducing the three unknown fitnesses to two, if, e.g., we normalize $w_{Aa}$ to be one. Thus, a particular equilibrium allele frequency can be 'explained' by any of the fitnesses in a one dimensional set of possible fitnesses. If we add the constraint (which is easy to specify in this case) that the equilibrium be stable, this restricts the possible fitnesses to those lying along a