SGA Search Dynamics on Second Order Functions

Bart Naudts* and Alain Verschoren
Dept. of Mathematics and Computer Science
University of Antwerp, RUCA
Groenenborgerlaan 171, B-2020 Antwerpen, Belgium
e-mail: [bnaudts,aver]@ruca.ua.ac.be

Abstract. By comparing its search dynamics to that of a simple $O(n^2)$ heuristic, we are able to analyze the behavior of the simple genetic algorithm on second order functions, whose optimization is shown to be an NP-equivalent problem. It appears that both algorithms approach these fitness functions in the same, intuitive way: they start by deciding the obvious and most probable, and then they proceed to make more difficult decisions. Useful information about the optimization problem is, among others, provided by statistical physics: lattice gases can be modeled as second order functions.

1 Introduction

In this note, we analyze the behavior of the simple genetic algorithm (SGA) on instances of a particular NP-equivalent optimization problem. (See [10] for a general introduction to complexity theory.) By choosing such a problem, we are ensured of instances of real world difficulty level, while still having a single and easily definable fitness function class to focus on.

A consequence of this choice is that we have to face the structural difficulties typical to NP-completeness. Firstly, the optimum of a function is something unreachable: exhaustive search is too expensive, and no efficient algorithm can tell a local optimum from a global one. A second problem is that it is very difficult to give a general characterization of difficult instances. E.g., recently a special volume [2] was dedicated to locating difficult instances of the 3SAT problem. Finally, one often has no idea of the characteristics of the fitness landscape, while still knowing that it contains many local optima which are fairly close to the global optimum. Only one path could lead to the global optimum, and there is no information to tell a good path from a bad one.

Throughout this note, we consider fitness functions of the form $f : S \rightarrow \mathbb{Z}$, where $S$ is the set of bit strings indexed by $A = \{0, 1, \ldots, \ell-1\}$, i.e., $S = \{0, 1\}^A$. Our very basic function class is the class of first order functions

$$f(s) = \sum_{i \in A} g_i(s_i), \quad g_i : \{0, 1\} \rightarrow \mathbb{Z}, \quad s \in S,$$

* Research assistant of the Fund for Scientific Research – Flanders (F.W.O.) (Belgium)
also called linear or fully separable functions. Typical is that each locus contributes to the function value independently of the other loci, which makes these simple functions almost perfect for SGA optimization because the uniform crossover operator literally “exchanges properties”.

First order functions occur all over the place. The epistasis measure (e.g., [1],[6]) computes the least squares distance between a fitness function and the class of first order functions, whereas the bit decidability metric [6] is based on the concept of independent decidability of loci. Integer Programming (e.g., [11]) requires a first order function to be optimized, under the restriction of a number of linear relations. In Sect. 2.2, we see that the well known problem from statistical physics of finding the ground state of a lattice gas can be modeled into a fitness function maximization problem. There, the loci represent the sites of the lattice: each site can be occupied by a gas molecule or can be empty, or, alternatively, can be associated with a magnetic spin with either north (e.g., bit value 0) or south pole (e.g., bit value 1). In the latter case, the first order function represents an external magnetic field.

One adds second order components $g_{ij} : \{0,1\}^2 \rightarrow \mathbb{Z}$ to Eq. 1 to obtain the second order functions

$$f(s) = \sum_{i \in A} g_i(s_i) + \sum_{i<j \in A} g_{ij}(s_i, s_j). \quad (2)$$

The optimization problem associated with the second order functions will be denoted as 2ORD. It is a restriction of this class (hence a subproblem of 2ORD) which will be the subject of our analysis.

As we will see in the next section, a fair number of real world optimization problems can be embedded into 2ORD. This connection with extensively studied problems is important as a source of algorithms to compare with, a source of theoretic results which provides insight in the structure of the class, and, if nothing else, as a source of difficult instances. Two additional motivations for the choice of 2ORD are that it is probably the simplest yet very rich function class beyond the first order functions, and that there is no encoding from an abstract problem space to the space of bit strings involved. Also, note that Kauffman’s NK-landscapes [4] with $K = 1$ and $N = \ell$ can essentially be embedded in the substantially richer class 2ORD.

It has been proved in [7] that 2ORD is an NP-equivalent problem. In view of this complexity, one may wonder whether any metric based on first order functions can grasp the richness of the second order functions and render it using one single value. Simple as it is, though, the bit decidability measure provides valuable (but by no means exhaustive!) information on the GA-hardness of the instances.

The principal aim of this note is to show that the search dynamics of the SGA can be predicted for many functions in U-2ORD-01, an NP-equivalent subclass of 2ORD. This is done by comparing the dynamics of the SGA to those of a very intuitive $O(\ell^2)$ heuristic, which tries to optimize U-2ORD-01 functions by deciding one locus at a time. One observes that the SGA and the heuristic