UCL DEPARTMENT OF
COMPUTER SCIENCE

# Research Note

RN/16/05

# Fitness Landscape of the Triangle Program

10 June 2016

*To be presented at*
Workshop on Landscape-Aware Heuristic Search (PPSN 2016)
Nadarajen Veerapen and Gabriela Ochoa (editors), 17-21 September

*William B. Langdon and Mark Harman*

## Abstract

Trying all hopeful high order mutations to source code shows none of the first order schema of triangle software engineering benchmark are deceptive. Indeed these unit building blocks lead to all global optima. Suggesting program improvements may not be as hard to find as is often assumed.

## Keywords

theory, genetic improvement, genetic algorithms, genetic programming, software engineering, SBSE, search, heuristic methods,

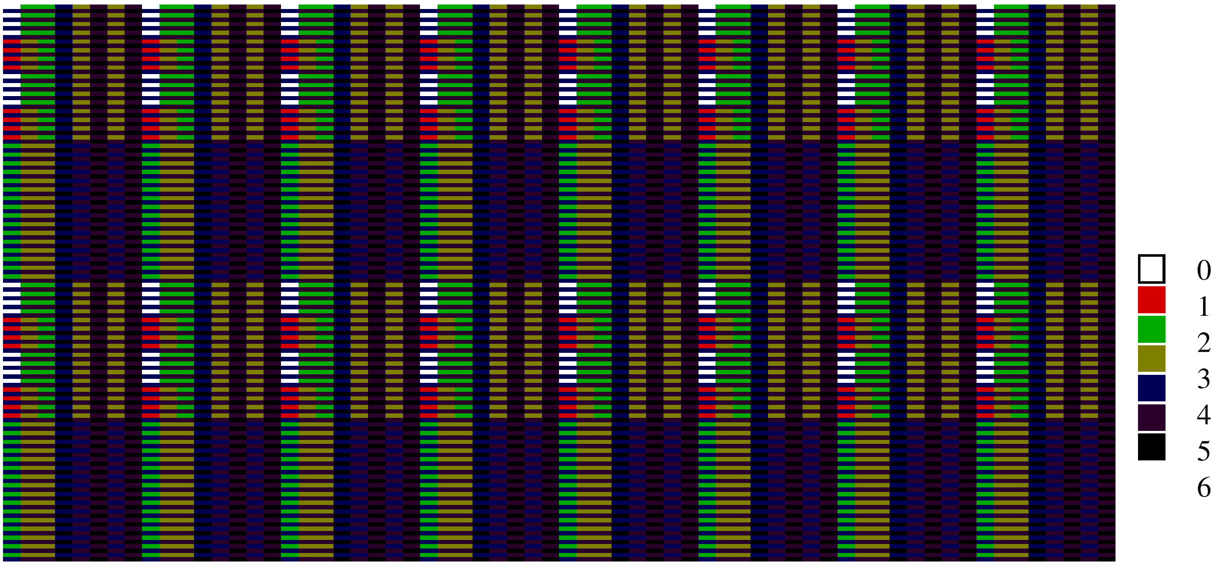

Figure 1: Fitness landscape of binary comparison improvement of triangle program. 2048 test equivalent mutants in white. The regular pattern of individuals with the same fitness indicates short building blocks. E.g. the vertical strips 8 pixels wide indicates the first three bits do not impact fitness. In contrast the last but one bit divides the figure into four horizontal stripes, two contain 50 176 mutants which fail 4 or more tests (dark pixels) whilst the others hold all the solutions (white).

Table 1: Hardest to detect mutations of triangle program [1, Fig. 3]. Column 1 number of individual changes in test equivalent mutations up to 4th order.

| | |
|---|---|
| 354 | == replaced by >= |
| 576 | <= replaced by < |
| 708 | == replaced by <= |
| 1062 | > replaced by != |
| 1992 | <= replaced by == |

## 1 Introduction

The triangle program is well studied software engineering benchmark. It can be thought of as a model of unit testing. It classifies triangles as scalene, isosceles, equilateral or not a triangle. We have previously used it to study high order mutation, concentrating particularly on injecting faults which change the numeric comparison operators (<, <=, ==, !=, => and >) [1]. We now consider mutation of all the comparison operators in the triangle program as a genetic algorithm fitness landscape. Taking as our fitness the number of tests [1, Tab. 2] which the modified code fails. A test equivalent mutant is one that passes all the tests and so has the best fitness value, which is zero. In [1] we considered higher order mutations, i.e. making multiple changes. For the triangle program we considered up to four simultaneous changes. Up to fourth order, there are 6 211 885 mutations, of which 1313 are test equivalent [1, Tab. 3].

Table 1 shows the individual changes and their frequency in all (up to 4th order) test equivalent mutants. Instead of allowing all possible combination, we study all of the most hopeful combinations. I.e. only the three most common mutations in Table 1 are used. Since there are six comparison operators and 17 potential mutation sites, this reduces the search space from $6^{17}$ to $2^{17}$. We evaluate all possible mutants.

## 2 The Whole Space, Zeroth Order Schema

There are 2048 global optima. (shown in white in Figure 1). On average each mutant fails only 4.344 ±1.360 tests. The worst mutants fail six of the 14 tests.

Table 2: Mean and standard deviation of number of tests failed for first order schema (excluding 22 with average means). Last column is estimated population size needed for a random sample to distinguish between competing pairs of first order schema.

| | | |
|---|---|---|
| -4 | 3.719 ±1.328 | 1.9 |
| 4 | 4.969 ±1.075 | |
| -5 | 4.062 ±1.478 | 4.7 |
| 5 | 4.625 ±1.166 | |
| -6 | 3.812 ±1.509 | 2.4 |
| 6 | 4.875 ±0.927 | |
| -11 | 3.438 ±1.273 | 1.1 |
| 11 | 5.250 ±0.661 | |
| -14 | 4.312 ±1.424 | 43.5 |
| 14 | 4.375 ±1.293 | |
| -16 | 4.188 ±1.550 | 8.6 |
| 16 | 4.500 ±1.118 | |

## 3 First Order Schema are Not Deceptive

Of the 34 first order schema, 22 have exactly average fitness and contain exactly half the global optima.

The other 12 schema either contains no solutions or all of them. In the six schema which contain solutions, on average individuals are better than the average of the whole space. In the other six, the schema average is worse than the average of the whole space.

That is, 22 schema have no signal and the remaining 12 are not deceptive. In the best schema mutants pass on average 1.813 ±1.015 more tests than its opposite (see also Table 2).

## 4 Schema Predict All Test Equivalent Mutants

As the previous section showed, there are 22 first order schema that have exactly average fitness. These correspond to 11 gene locations ($2^{11} = 2048$). If we treat each as though it did not care and fix the remainder to the value corresponding to the better than average first order schema, we get 2048 combinations, each of which is one of the solutions!

That is, once we fix the six mutation sites in the C source code corresponding to the schema in Table 2 we are free to mutate all the others (using our restricted mutation operator, last three rows of Table 1) and the new program will return the correct answer for all of the tests.

## 5 Conclusions

Although the triangle program is small, the number of possible triangle programs is huge. We have fully explored a regular subset of it. We reduced the size of its search space by considering only potential improvements to the existing code made by replacing its comparisons and by restricting the comparator mutations. This enabled us to analyse a systematic subset of the whole improvement fitness landscape. Solutions in the subset will still be solutions in the full problem. There are many solutions all of which are readily found by first order schema analysis. Suggesting the program fitness landscape is not as difficult to search as is often assumed.

## References

[1] Langdon, W.B., Harman, M., Jia, Yue: Efficient multi-objective higher order mutation testing with genetic programming. Journal of Systems and Software **83**(12) (2010) 2416–2430