Fitness Landscaping for 1-0 Knapsack Instances.

Ashwin Padmanabhan

***Abstract*— This report explores the Fitness Landscape of different Knapsack instances. In the first section, we perform random walks to determine the difficulty of the knapsack instance.**

I. Section 1

a) Random walk setting: : For each given instance, the global optima is provided with data file, which is taken as a distance benchmark. Distance between two genotype is calculated using a simple euclidean distance. For each walk, we have a walk length of about 30 steps, and have around 70 total walks. Furthermore, I make sure that the starting point of the very first walk is from the optimal genotype. I did this to make sure that the area around the optimal is mapped. The starting points of the other 69 walks are random. Also, at each point in the walk, I calculate all possible (feasible) steps we can take from the given point and choose one at random, this takes care of the problem of ending up with an infeasible genotype. That is, for a given point during a walk, a list of feasible bit flips are evaluated along the length of the genotype; if n such paths exist, the probability of taking one of these n paths occurs with probability $\frac{1}{n}$. The function and distance is mapped and plots are shown in Figure 1. Other fitness evaluations are done with this basis of the random walk. Note that all the fitness landscape measures are compiled in a single table shown below.

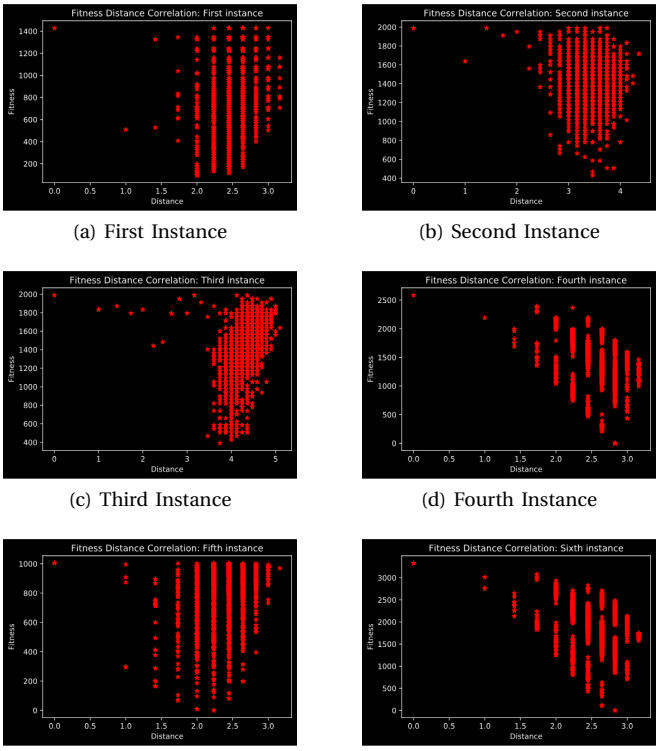
b) Fitness Distance Correlation: : The FDC scores were calculated for each instance and are shown in the table. For a given random walk, FDC is calculated using (same notation as used in class);

$$\text{FDC} = \frac{G_{FD}}{G_F * G_D} \tag{1}$$

c) Average Fitness v Walk length: The same experiment was performed, walk of length 30 steps, restarted 70 times each, starting, the first walk encoded to start from the optimal. Refer fig 2 for plots. Note that we usually start high due to the design of the experiment in which we always start the first walk from the optimal point.

d) Auto-correlation Length, IC, PIC, Density Basic: : The setting is the same as the previous case. No special points come to mind, except for PIC (Partial Information content). The measures consistent with class notation are used instead of the one used in **Write paper name** and the number of modes are used for μ in the formulation of $M(\epsilon)$ and not 2μ.

e) Comments about Problem Difficulty:: **First, Third and Fifth instance:**



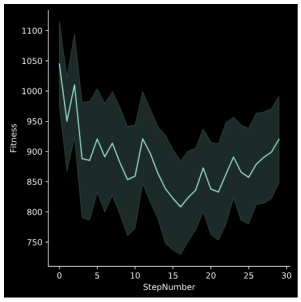
(a) First Instance (b) Second Instance

(c) Third Instance (d) Fourth Instance

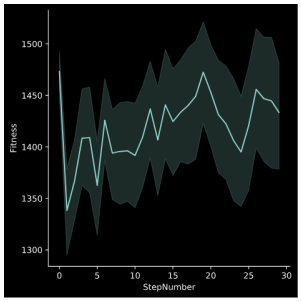
(e) Fifth Instance (f) Sixth Instance

Fig. 1: Average Fitness - Step Number for each instance

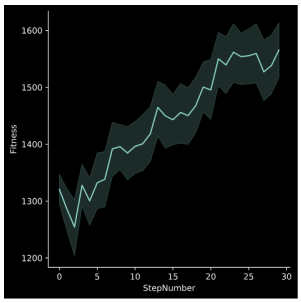
The first/third/fifth instance have a high FDC score, implying that the landscape is difficult/misleading. The fitness seems to increase with increasing distance. Especially note fitness at distances of 1 are lower than the fitness at distances of 2.5 units in Figure 1. It is also noteworthy that the auto-correlation length for the first and third instances are similar (small hence more smooth) and hence we can conclude that they are of similar ruggedness (or smoothness), a inference supported by the similar IC values, suggesting the instances to be of similar ruggedness, and contribution from "types" of ruggedness. The fifth one however, has a lower ACL (more ruggedness) and a high IC value (more "type" of ruggedness) which might mean a more (possibly) challenging search-space compared to the first and third instance (fifth has a high ruggedness and more contribution from "types" or ruggedness). However, before we say fifth is tougher, we inspect the Density Basin information we see that the third has more isolated peaks than the other two. **Second Instance:** For the second instance, the fitness at distance of 1 is almost similar to



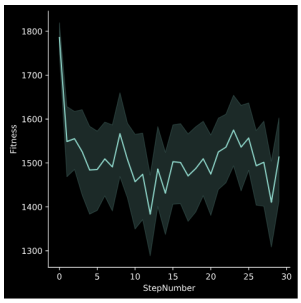
(a) First Instance



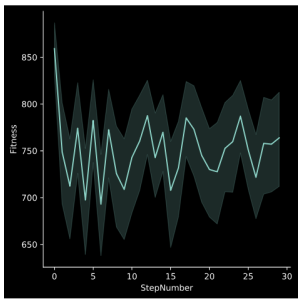
(b) Second Instance



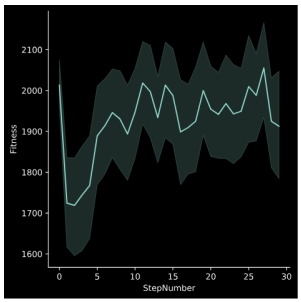
(c) Third Instance



(d) Fourth Instance



(e) Fifth Instance



(f) Sixth Instance

Fig. 2: Fitness - Distance Correlation for each instance

fitness we get at distances of 2 or more, justifying lower FDC value. This problem instance FDC value implies a difficult problem. We see a high auto-correlation (implying smooth) length and a comparatively lower Density Basin, suggesting (relatively) more isolated peaks, which might generally suggest a tougher problem. **Fourth and Sixth Instance:** For the fourth and sixth instance, going off the FDC values, one would conclude instances are easy since fitness decreases as the distance increases. This is also evidenced in Figure 1d and Figure 1f. We also see a lower ACL for the two instances compared to most of the other instances, meaning a more rugged landscape and a lower IC meaning lesser "types" of ruggedness. They have moderate-low Density Basin information, implying perhaps isolated (if less so than the third) peaks.