GAGP Tutorial 5 (week 8)
Fitness Distance Correlation

You will need calculators for this practical.

Fitness distance correlation (Jones and Mitchell) is a way of quantifying the relationship between a fitness function and distance to the global optimum. The fitness function can be seen as an estimate of how far it is to the global optimum and, for it to be a "good" fitness function one would like it to give a good estimate. This is not always possible.

Assume you have a set of fitnesses $F = f1, f2, \ldots$ and a set of known distances to the global optimum $D = d1, d2, \ldots$.

$$\text{FDC} = \frac{C}{SF \times SD}$$

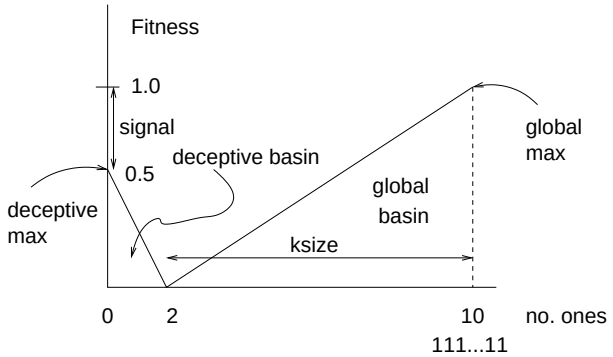
SF and SD are the standard deviation of F and D respectively and

$$C = \frac{1}{n}\sum_{i=1}^{n}(f_i - \bar{f})(d_i - \bar{d})$$

where $\bar{f}$ and $\bar{d}$ are the means of F and D. C is the covariance of F and D.

1. What is the FDC in the max ones problem? How do you interpret the value you get? What do you need to define before you can do this calculation?

2. What is the FDC for the fitness function shown in the figure? The global maximum is at the right side, so the global solution is a binary string of P 1s where P is "problem length". How does the FDC vary as "ksize" and "signal" are varied?



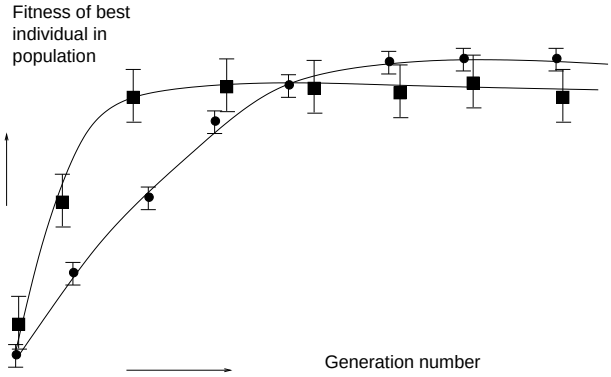
3. What happens if we introduce a "plateau"? How would you expect the GA's performance to be affected?

4. What about "needle in haystack" problems? Consider the landscape, the distance measure and the FDC.

5. What does FDC tell you about the difficulty the GA will find in solving a problem?

6. In the homework you investigate a GA running on MAX-1s with standard binary encoding. You plot the results of one run: average fitness of population vs. generation number. What can you say about the performance of your GA from this plot?

7. Look at the plot of best fitness vs. generation number for two different experiments with a GA (maybe you tried two different values of the mutation probability).



How might you calculate the error bars shown? Which of the two plots shows better performance?

8. What are the parameters you'll need to consider in the practical? How many experiments will you have to do? How will you decide which experiments to do? How will you evaluate your results?