Advanced Genetic Algorithms

GAs are less likely to get stuck than hill climbers because:

- They maintain and evolve a population of solutions
- Genetic operators may generate a wider range of neighbours
- These reasons exploit and rely on diversity in the evolving population

Premature Convergence:

If the population converges on a single basin of attraction they can find it difficult to escape, this is called premature convergence.

Neutrality:

Neighbouring genotypes have equal fitness they are 'selectively neutral' even if they have different phenotypes. This can result in evolutionary drift.

'Neutral networks' that percolate the search space may allow converged populations to escape 'local optima'

Invisible Needle:

If the population starts on a needle they might not ever find it because its highly unlikely for offspring to be an exact copy of its parents. So a perfect solution will tend to have unfit offspring. Elitism can help to solve this.

Quasi-Species:

- High GA mutation rates mean that the evolving population is a cloud of points on the landscape this is a quasi species. If the selection pressure on the population to reward good solutions is overcome by the mutation pressure that constantly corrupts them then the population cannot see the needle.
- Conversely: if an evolving population is able to find a good solution and stay there we can infer that the solution must be surrounded by other pretty good solutions there it is robust.

Epistasis:

- If fitness is a linear function of gene alleles, the contribution of each gene to fitness is independent of other genes. These are trivial to optimise as we can solve each gene independently.
- Most of the time this is not the case, fitness is a non-linear function of the gene alleles, the contribution of genes to fitness is interdependent to some extent, the best allele at one gene depends on the alleles at other genes. This interdependency is known as **Epistasis.**
- The presence of epistasis is equivalent to the existence of local optima, they are non-linearities in the mapping of fitness onto the search space

- Epistasis means we cannot optimize genes independently instead our algorithm must co-optimize the alleles of multiple genes.
- The size of the set of interacting genes is the order of interaction. More generally what is the **structured modularity** of the problem

Partial Decomposability:

- Partial modularity means the solution has some epistasis.

No Free Lunch:

When algorithm performance is averaged across all possible problems. Any two optimization algorithms will exhibit equivalent performance across the space of all possible problems. Meaning no optimization algorithm can do better than blind random search.

The insight is that every search algorithm other than a random search has a search bias. For every time the bias pays off there is a problem that it doesn't

- Satisficing - Finding a good solution, not necessarily the best.
- The algorithm quality depends on the problem structure that it faces.

Learning How to Search:

The structure of the solution space is influenced by our choice of representation and genetic operators, so we can restructure the space by making different choices. Can the algorithm evolve itself as it searches the landscape.

A poor GA has low evolvability

Genetic algorithms can analyse how fitness varies in the current population, building a model of the problem, we can use this model to predict what type of genotypic variation will be best, effectively learning the structure of the problem during evolution.