NAT-DL Self-study and answers: MOO & HH Set 6 (week 8)

1. Portfolio selection is a typical application for multi-objective MHO algorithms. For the portfolio a selection of assets has to be made, so that the task consists in finding an optimal distribution of a budget over these assets. In addition to maximal return, also risk minimisation and asset preferences by the clients play a role. Specify fitness functions for a metaheuristic MOO algorithm, and discuss your approach to solving the problem.

   Answer: The problem is similar to the knapsack problem that was discussed in an earlier tutorial. There is a total amount of money that is to be spent over stocks, and --- like the items in the knapsack having weight and usefulness --- each stock has a utility and a cost. The (expected) utility of a stock can be defined in various ways, just like the utility of an item in the knapsack depends on where the hike goes, how long it will take, and who's joining. For the portfolio selection it is possible to find out as much as possible about the customers in order to use a scalarisation approach. If this is not possible, e.g., if there are many different customers, then a variety of stocks with various ratios between expected revenue and risk need to be found within the part of the search space that suits the customers preferences. It is also possible to use a fitness for the customer preference. For example, if the customer state "Invest in green energy, perhaps in natural gas, but not in coal", then gas-related stock would get lower fitness and coal would get fitness minus infinity. As this is a separate fitness function, the customer can decide later, whether they want to buy any low-risk high-gain gas stocks, while coal would remain ruled out, of the customer has agreed before that the eco-fitness cannot be ignored.
   Note that here the number of trial is limited in a different way than before, because any gain can be used to pay for more fitness evaluations, although also models or data-based evaluation schemes exist. It is not really known whether metaheuristic methods are used in finance in practice, because revealing their usefulness would change the type of the optimisation problem. There are many studies and, now and then, announcements in the news that is has been tried, which is neither a guarantee that they are being used.
   It may be also interesting to note that Jianming Xia (2004)] has shown that for a coordinated group of investors the expected share of the return per investor is higher than what can be expected by a single investor. This could be related to the co-evolutionary free lunch, but it is also recommended considering the social and ethical implications.

2. Consider the following variant of the All-Ones problem over a discrete search space with two objectives:

   $$f_1(x) = |2\, x_1 - x_2| \text{ if } x \neq \text{"11 . . . 11", and otherwise } f_1(\text{"11 . . . 11"}) = x_1\, x_2$$

   $$f_2(x) = |2\, x_2 - x_1| \text{ if } x \neq \text{"11 . . . 11", and otherwise } f_2(\text{"11 . . . 11"}) = x_1\, x_2$$

   where x is a string of an even number of bits, with x_1 represents the number of bits equal to 1 in the first half of the string, and x_2 the number of bits equal to 1 in the second half of the chain. What is the optimal Pareto front? What approximations of the Pareto front are likely to be found? Show that using a genetic algorithm one can reach quickly the global optimum, whereas the local search is likely to get trapped. (adapted from Exercise 4.21 in Talbi)

   Answer: The problem is obviously that the first objective implies to have many 1s in the first part and very few in the second part, while the second objective suppressed 1s in the first part and accumulates 1s in the second part. Note that the factor 2 in both objectives helps that the gain of 1s in one half by one criterion is preferable to loosing any 1s in this part by the other criterion.
   The optimal Pareto front is by definition the point "11 . . . 11", but as long as this point has not yet been found, the Pareto front is approximated by the best individuals in the population which are initially any vectors that are biased to either side. Later, the individuals

that most strongly biased towards one side are the approximation of the Pareto front which is likely to converge to the two points "1 . . . 10 . . . 0" and "0 . . . 01 . . . 1". You can visualise the problem by drawing a two-dimensional array spanned by the x_1 axis and the x_2 axis. For random vectors the population is initially near the centre $(x_1, x_2)=(N/4, N/4)$. A local algorithm would move the points toward the corners $(x_1, x_2)=(N/2, 0)$ or $(x_1, x_2)=(0, N/2)$, i.e. not towards the global optimum at $(x_1, x_2)=(N/2, N/2)$. So, by following the Pareto front, an algorithm that works only locally would be drawn away from the global optimum. This is a characteristic of a deceptive problem which we see here to be possible in multi-objective optimisation as well. A GA would be able to crossover the two subsolutions, once they are found, and to jump right to the global optimum $(x_1, x_2)=(N/2, N/2)$. This means that there are some optimisation problems that are easy for one optimisation algorithm and difficult for other algorithms.
For completeness, if the length of the strings is only 2, then the two points "10" and "01" are global optima, while "11" has only fitness 1 and, because it is neighbouring to a better state, is not even a local optimum. If the length is 4 (not odd string lengths here), then "1100", "0011" and "1111" are all global optima, so we should have said that the length of the string is at least 6 to make the problem straightforward. We did this at best implicitly by writing out 4 digits and added some dots.

3. Tracking objects in a video in conditions can be difficult for changes of lighting or if the object is get frequently occluded or if it rotates. Design a population-based metaheuristic based on particle swarms for such a tracking task. It can use a number of interest points that characterise the object, although not all of these points can be identified all the time.

   Answer: Assume we are given a library of interest point (or features) such as corners, centres of symmetry, branching points etc. Some are more likely to be miss-detected or lost in sample task over which a set of indicators can be found. The idea would be that not just the individual indicators are to be evaluated and then "selected", but their joint usefulness so that for all relevant objects such an optimal set can be found. We can use here a hierarchical hybrid which first detects what object it is and then employs the best set (or continues to evolve the set). This should be done in a population so that several object hypotheses can be maintained. Alternatively, a relay-type hybrid can be considered, where the object identification and feature matching are used alternately, so that a wrong object identification can be corrected when the feature set has turned out to work poorly. Note that for this application ist can make sense to check the Reynolds rules in particular velocity alignment.

4. In multi-objective metaheuristic optimisation, various strategies have been proposed to maintain diversity. For instance, the NSGA-II algorithm is based on a crowding distance measure. Propose a modification of the NSGA-II algorithm in which the crowding operator is replaced by a *k*-means clustering algorithm for some given value of *k*.
   (adapted from Exercise 5.18 in Talbi)

   Answer: Hybrid algorithms can include methods from outside MHO, and mostly will. As usual in *k*-means, it is not easy to find a good value for *k*, but let us assume *k* is fixed. We can now consider all solutions that have made it into the first Pareto front, and apply the *k*-means algorithm to them. After *k*-means has nicely converged (which is not always the case), we can calculate the mean distance of the solutions from their nearest cluster centre. While *k*-means aims at minimising this mean distance, we can reward solutions that contribute much to this mean distance, which can help to avoid the data from clustering at particular positions. To take to methods or mechanisms and let them fight, is an often used approach not only in MHO but generally in machine learning. It is also possible to assign higher fitness solutions that join a centroid that represents fewer points than other centroids or that has a large standard deviation.

5. In many problems some solution components are discrete and some are continuous. How can a hyperheurstic algorithm be applied to this problem? This problem is occurs in most GP applications. Can you think of any interesting or more specific cases?

   Answer: A hyperheuristic algorithm could try to find out which components are discrete and which continuous. It does not need to state this explicitly, but would simply choose a different optimisation algorithms for either. E.g. certain numerical constants in GP may be discrete other continuous, without this being clear from the beginning, so the higher order algorithm can decide which optimisation method to apply.
   In a more simplified setting a hybrid approach can be taken where a continuous metaheuristic in the inner loop optimises the problem for any configuration of the discrete decision variables that are fixed for the moment. In the outer loop, a discrete metaheuristic will optimise the problem while the continuous variables are fixed to the best known values found by in the inner loop. The search then alternate between the space of continuous variables and the space of discrete variables. The drawback of such a nested optimisation strategy is that it takes very long and there is no guarantee that it converges. The aforementioned hyperheuristic approach could work better if its outer loop can be optimised over a certain set of problems.

6. It is easy to produce toy examples with a non-connected Pareto front. Can you think of an example of a real-world problem where the Pareto front is non-connected?

   Answer: Obviously in discrete cases the Pareto front is not connected, which follows from the very idea of discreteness. For a practical example, we could consider the task to buy as much as possible for least cost. The criterion of minimal cost will lead to by a small package, while the goal of buying a large amount will imply choosing the large package. Obviously, there cannot be continuously many package sizes, but package sizing is clearly a problem that is solved by optimisation. In other cases, we have one (quasi-)continuous and one discrete variable, as for example in tax categories. A fully continuous case could be driving with the goals of being fast with minimal energy usage and minimal risk. And we can consider the case of decision-making whether to overtake a car in front.