Stochastic optimization methods (FFR 105), 2018
Solutions to the exam (2018-10-31)

1. (a) i. Two reasons for swarming: (i) Protection against predators (safety in numbers) and (ii) efficient food search (for example ants and some bird species)

   ii. The equation for the velocity update takes the form

$$v_{ij} \leftarrow wv_{ij} + c_1 q \left(\frac{x_{ij}^{\text{pb}} - x_{ij}}{\Delta t} \right) + c_1 r \left(\frac{x_j^{\text{sb}} - x_{ij}}{\Delta t} \right), \tag{1}$$

where $i = 1, \ldots, N$ enumerates the particles and $j = 1, \ldots, n$ enumerates the variables (dimensions). x_{ij}, $j = 1, \ldots, n$ are the position components for particle i, and v_{ij} are the velocity components for the same particle. x_{ij}^{pb} are the components of the best position found by particle i, whereas x_j^{pb} are the components of the best position found by any particle in the swarm (either best-in-current-swarm or best-ever). c_1 and c_2 are constants, usually set to 2. Δt is another constant (dimension: time), typically set to 1. q and r are random numbers, one for each particle. w is the inertia term (see also (c) below). The c_1-term is called the *cognitive component* and the c_2-term is called the *social component*. These components can be seen as a particle's level trust in itself and the swarm, respectively, regarding the ability to find the optimum. The trade-off is handled via the variation in the inertia weight w. This parameter is initially set to a value of around 1.4, and is then allowed to decrease in each iteration (by multiplying by with a factor $\beta < 1$, usually set to 0.99 or so), until it reaches 0.3-0.4. After that, w is kept constant. When $w > 1$, exploration is favoured, since the particle will then pay less attention (relatively speaking) to the best positions found so far, and instead mostly continue in its current direction. By contrast, when $w < 1$, exploitation is favoured, since the particle will then pay more attention to the best positions found so far.

(b) The hessian H takes the form

$$H = \begin{pmatrix} 8 & -2 \\ -2 & 4 \end{pmatrix} \tag{2}$$

The eigenvalues are obtained from the determinant equation $\det(H - \lambda I) = 0$ that, in this case, becomes

$$(8 - \lambda)(4 - \lambda) - 4 = \lambda^2 - 12\lambda + 28 = 0, \tag{3}$$

with the solutions $\lambda_{1,2} = 6 \pm 2\sqrt{2} > 0$. Since H is positive definite, f is convex.

(c) There are 25 possible pairs of individuals, of which 9 contain individual 4. Individual 4 is the better individual (of the pair) in six cases, namely (1,4), (2,4), (3,4), (4,1), (4,2), (4,3). In those six cases, individual 4 is selected with probability p_{tour} (once the pair has been formed) In two cases, namely (4,5) and (5,4), Individual 4 is the worse individual of the pair, and it is then selected with probability $1 - p_{\text{tour}}$. Finally, for the pair (4,4), Individual 4 is of course selected with probability 1. Taking into account that all pairs are formed with equal probability (1/25), the probability of selecting Individual 4 in a single step of tournament selection thus becomes

$$p_4 = \frac{1}{25}\left(1 + 6p_{\text{tour}} + 2(1 - p_{\text{tour}})\right) \approx 0.248. \tag{4}$$

2. (a) The general expression for Newton-Raphson's method is

$$x_{j+1} = x_j - \frac{f'(x_j)}{f''(x_j)}. \tag{5}$$

With $f(x) = 1 + x^6 - x^2$ one obtains

$$f'(x) = 6x^5 - 2x, \tag{6}$$

and

$$f''(x) = 30x^4 - 2, \tag{7}$$

so that

$$x_{j+1} = x_j - \frac{6x_j^5 - 2x_j}{30x_j^4 - 2}. \tag{8}$$

Starting from $x = x_0 = 1$, one then obtains

j	x_j	difference
0	1.000000000	-
1	0.857142857	0.142857143
2	0.782339675	0.074803183
3	0.761365962	0.020973712
4	0.759843344	0.001522618
5	0.759835686	$0.000007658 < 10^{-5}$

Thus, $x^* \approx 0.75984$ is a stationary point. (It is easy to check that $f'(x^*) \approx 0$). From the expression for $f''(x)$ one gets $f''(x^*) \approx 8 > 0$, showing that the stationary point is indeed a minimum.

(b) The function $L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda h(x_1, x_2)$ takes the form

$$L(x_1, x_2, \lambda) = 3x_1x_2 + 2 + \lambda(x_1^2 + x_2^2 - 1). \tag{9}$$

Taking partial derivatives and setting them to zero, one obtains the equations

$$\frac{\partial L}{\partial x_1} = 3x_2 + 2\lambda x_1 = 0, \tag{10}$$

$$\frac{\partial L}{\partial x_2} = 3x_1 + 2\lambda x_2 = 0, \tag{11}$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 + x_2^2 - 1 = 0. \tag{12}$$

Solving for λ one gets

$$\lambda = -\frac{3x_2}{2x_1} = -\frac{3x_1}{2x_2}, \tag{13}$$

provided that neither x_1 or x_2 is equal to zero (those cases can be checked separately, see below). From this equation one then obtains

$$6x_2^2 = 6x_1^2, \tag{14}$$

so that

$$x_2 = \pm x_1. \tag{15}$$

Thus, there are two cases. With $x_2 = x_1$, Equation (12) gives $2x_1^2 = 1$, so that $x_1 = \pm 1/\sqrt{2}$ (and, therefore, $x_2 = \pm 1/\sqrt{2}$). Thus, two points are found: $(1/\sqrt{2}, 1/\sqrt{2})$ and $(-1/\sqrt{2}, -1/\sqrt{2})$. If instead $x_2 = -x_1$, one obtains the same equation for x_1 as before. Thus, two additional points are found, namely $(1/\sqrt{2}, -1/\sqrt{2})$ and $(-1/\sqrt{2}, 1/\sqrt{2})$. Inserting numerical values one finds that $f = 7/2$ for $\pm(1/\sqrt{2}, 1/\sqrt{2})$ and $f = 1/2$ for $\pm(1/\sqrt{2}, -1/\sqrt{2})$. Thus, the maximum value is 7/2 and the minimum value 1/2. (For $x_1 = 0$ one finds $x_2 = 0$, which does not fulfil the constraint. Similarly, $x_2 = 0$ gives $x_1 = 0$, again violating the constraint).

3. (a) In order for an improvement to occur, two conditions should be fulfilled (according to the problem formulation). First of all, no 1s should mutate. The probability of mutation is p_{mut}. Thus, the probability of *not* mutating a gene is thus $1 - p_{\text{mut}}$. Since the mutations (of different genes) are independent of each other, the probability of not mutating any gene with the allele 1 will be $(1-p_{\text{mut}})^{m-l}$, where $m-l$ is the number of 1s in the chromosome. Similarly, the

probability of not mutating any of the 0s will thus be $(1-p_{\text{mut}})^l$. Therefore, the probability of mutating *at least* one 0 will be $1-(1-p_{\text{mut}})^l$. Combining the two expressions by multiplication (since, again, genes mutate independently of each other) one obtains the following expression for the improvement probability P

$$P(l, p_{\text{mut}}) = (1 - p_{\text{mut}})^{m-l}(1 - (1 - p_{\text{mut}})^l). \tag{16}$$

(b) The proof is given in Appendix B2.4, pp. 181-182. The first step is to note that the expected time to an improvement can be approximated as $1/P(l, p_{\text{mut}})$, where P was defined in part (a) above. Next, one must note that, initially the number of 1s (or 0s) will be $m/2$ on average. Thus, $m/2$ improvement steps are required, giving a sum to be computed. The sum can then be simplified by using a series expansion and a known mathematical limit. Then, a final approximation (allowing the simplified sum to be computed) leads to the expression

$$E(L) \approx \mathrm{e}^k \frac{m}{k} \ln \frac{m}{2}. \tag{17}$$

Thus, the expected computation time varies with m as $m \ln \frac{m}{2}$.

4. (a) By examining the traversal times, one can easily establish that the fastest route is A→B→D→F→H, with a duration of 10 time units.

The equation for node selection takes the form

$$p(e_{ij}|S) = \frac{\tau_{ij}^{\alpha}\eta_{ij}^{\beta}}{\sum_{\nu_l \notin L_T(S)} \tau_{lj}^{\alpha}\eta_{lj}^{\beta}}, \tag{18}$$

where, in this part of the problem, the pheromone levels can be ignored since they are all equal. At node A, there are two possible moves, either to node B or to node C. The probability of moving along $e_{B\leftarrow A}$ then becomes

$$p_{B\leftarrow A} = \frac{\eta_{B\leftarrow A}^2}{\eta_{B\leftarrow A}^2 + \eta_{C\leftarrow A}^2} = \frac{1/4}{1/4 + 1/9} = \frac{9}{13} \approx 0.6923077 \tag{19}$$

At node B, there are three possible moves, to nodes D, E, and G. The probability of moving to node D becomes

$$p_{D\leftarrow B} = \frac{\eta_{D\leftarrow B}^2}{\eta_{D\leftarrow B}^2 + \eta_{E\leftarrow B}^2 + \eta_{G\leftarrow B}^2} = \frac{1/9}{1/9 + 1/16 + 1/64} = \frac{64}{109} \approx 0.5871560. \tag{20}$$

At node D, there are two possible moves, to nodes F and G. The probability of moving to node F equals

$$p_{F \leftarrow D} = \frac{\eta^2_{F \leftarrow D}}{\eta^2_{G \leftarrow D}} = \frac{1/4}{1/4 + 1/16} = \frac{4}{5} = 0.8. \tag{21}$$

At node F, there is only one possible move (to node H), and it therefore occurs with probability 1. The probability of selecting the fastest path thus equals

$$p_{\text{fast}} = p_{B \leftarrow A} \times p_{D \leftarrow B} \times p_{F \leftarrow D} \approx 0.32519. \tag{22}$$

(b) Pheromones are updated as usual in AS (except that, as mentioned in the problem formulation, the update occurs directly after each vehicle has completed its path), i.e. as $\tau_{ij} \leftarrow (1-\rho)\tau_{ij} + \Delta\tau_{ij}$, where $\Delta\tau_{ij} = 1/T$ (where T is the traversal time) if the edge e_{ij} was traversed and 0 otherwise.

In this case, with $T = 10$ time units, $\rho = 0.5$, and $\tau_{ij} = \tau_0 = 0.1$, the pheromone level becomes $\tau = 0.5 \times 0.1 + 0.1 = 0.15$ for edges $e_{B \leftarrow A}, e_{D \leftarrow B}, e_{F \leftarrow D}, e_{H \leftarrow F}$ and $\tau = 0.5 \times 0.1 = 0.05$ for all other edges.

(c) The solution is rather similar to the solution of part (a), except that, now, the pheromone levels must also be taken into account. More specifically, for the edges along the fastest route, i.e. the edges traversed by the first vehicle, the pheromone levels are 3 times higher than on the non-traversed edges. Proceeding as in part (a), the probabilities for the second vehicle thus become

$$p_{B \leftarrow A} = \frac{3/4}{3/4 + 1/9} = \frac{27}{31} \approx 0.8709677, \tag{23}$$

$$p_{D \leftarrow B} = \frac{3/9}{3/9 + 1/16 + 1/64} = \frac{64}{79} \approx 0.8101266, \tag{24}$$

$$p_{F \leftarrow D} = \frac{3/4}{3/4 + 1/16} = \frac{12}{13} \approx 0.9230769. \tag{25}$$

The move from node F to node H still occurs with probability 1. Thus, the probability of the second vehicle to take the fastest route becomes

$$p_{\text{fast}} = p_{B \leftarrow A} \times p_{D \leftarrow B} \times p_{F \leftarrow D} \approx 0.65132. \tag{26}$$

The probability is thus much higher than for the first vehicle, as expected.