Chalmers University of Technology, Department of Mechanics and Maritime Sciences
Teacher: Mattias Wahde, tel. 772 3727

Exam in FFR 105 (Stochastic optimization algorithms), 2018-10-31, 14.00-18.00, M.

The examiner will visit the exam rooms twice, around 15.00 and around 17.00. It will be possible to review your results (exam and home problems) after Nov. 13.

In the exam, it is allowed to use a calculator, as long as it cannot store any text. Furthermore, mathematical tables (such as Beta, Standard Math etc.) are allowed, provided that no notes have been added. However, it is *not* allowed to use the course book, or any lecture notes from the course, during the exam.

Note! In problems involving computation, show *clearly* how you arrived at your answer, i.e. include intermediate steps etc. Only giving the answer will result in zero points on the problem in question. There are four problems in the exam, and the maximum number of points is 25.

1. (a) Particle swarm optimization (PSO) is a stochastic optimization based on swarming in biological systems, and somewhat similar to genetic algorithms.
   i. Swarming is a frequent phenomenon in nature. Describe at least two reasons why this phenomenon occurs. (1p)
   ii. Write down the equation for the velocity update in the standard PSO algorithm, and explain all parts of the equation in detail. Pay particular attention to the variable indices (where applicable). In the equation for the velocity update, one parameter is responsible for handling the trade-off between exploration and exploitation. Your description should include a clear explanation (with equations and parameter values) of how this trade-off is handled in PSO. (4p)

   (b) In optimization, convexity of the objective function is a desirable property. Determine whether or not the function

$$f(x_1, x_2) = 4x_1^2 + 2x_2^2 - 2x_1x_2 \tag{1}$$

   is convex. (1p)

   (c) Tournament selection is a common selection operator in evolutionary algorithms (EAs). Consider a case with a population size of five where, in a given generation, the fitness values are $F_1 = 3$, $F_2 = 6$, $F_3 = 7$, $F_4 = 9$ and $F_5 = 15$. Assuming a tournament size of two, what is the probability of selecting individual 4 (in a single selection step), assuming that the tournament selection parameter p_{tour} is equal to 0.8. Show clearly how you arrive at your answer. (2p)

2. (a) Newton-Raphson's method is an iterative method for finding local optima of a twice differentiable function. Use this method to find the minimum of the function $f(x) = 1 + x^6 - x^2$, starting from $x \equiv x_0 = 1$. First, write down (for this particular function) the expression for x_{j+1} as a function of x_j. Then iterate until the difference between two consecutive iterates (i.e. $|x_{j+1} - x_j|$) drops below 10^{-5}. Next, make a table with three columns: (1) the index $j+1$ of the iterate, (2) the corresponding value x_{j+1} and (3) the difference $|x_{j+1} - x_j|$, and insert the values obtained for $j = 0, 1, 2, \ldots$ Finally, *prove* that the point found is really a minimum of $f(x)$. (3p)

   (b) The Lagrange multiplier method is applicable to optimization problems involving equality constraints. Using this method, find the minimum and maximum values of the function

$$f(x_1, x_2) = 3x_1x_2 + 2 \tag{2}$$

   subject to the constraint

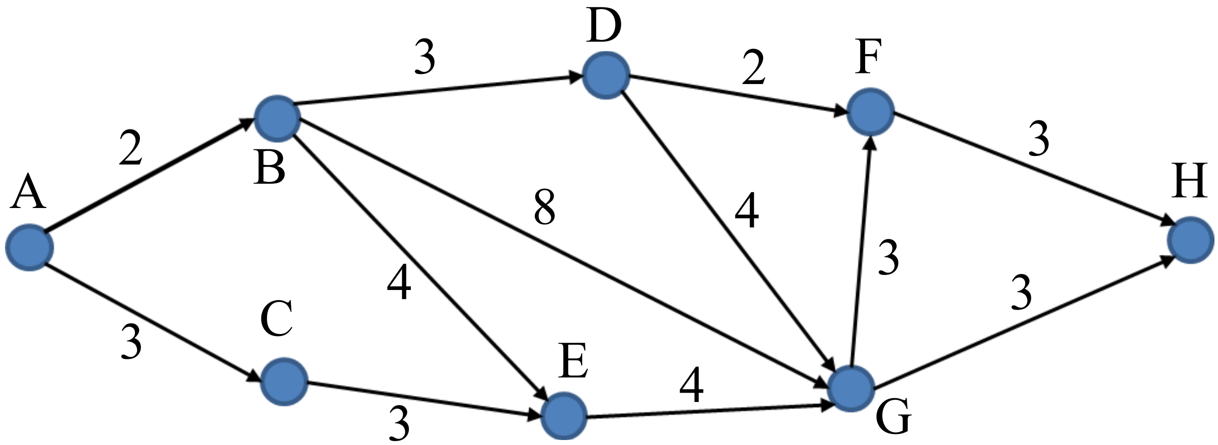
$$x_1^2 + x_2^2 = 1. \tag{3}$$

   In your answer, in addition to giving the minimum and maximum values, provide also the corresponding values of (x_1, x_2) for all points where the function takes either the minimum or the maximum value. (4p)

3. In analytical studies of genetic algorithms (GAs), it is common to use the Onemax problem, for which the value of the fitness function for a given (binary) chromosome equals the number of 1s in the chromosome. For this simple problem, one can derive an expression for the runtime and the optimal mutation rate for a GA with a single individual, which is modified using mutations only. In this GA, a mutated individual is kept if and only if it is better (i.e. its chromosome contains more 1s) than the previous individual.

   (a) Consider a chromosome of length m with l 0s (and, therefore, $m-l$ 1s). Let the mutation rate be p_{mut}. Derive an approximate expression for the probability of improving this chromosome (i.e. increasing the number of 1s). The expression should summarize a case in which none of the 1s mutate, and at least one of the 0s does. (1p)

   (b) Setting the mutation rate p_{mut} to k/m, where m is the chromosome length, and $k \ll m$ is a positive integer, *derive* an estimate for the runtime of this GA, i.e. the number of iterations required to reach the global optimum. Motivate clearly any approximations made in the derivation. (4p)



4. Ant colony optimization (ACO) can be used in vehicle routing problems. Consider the road network shown in the figure above. The network, which is not fully connected, consists of eight nodes and a number of directed edges. For simplicity, the edges are one-way in this problem. Thus, for example, there is an edge $e_{\mathrm{B}\leftarrow\mathrm{A}}$ from node A to node B, but no edge $e_{\mathrm{A}\leftarrow\mathrm{B}}$ from node B to node A etc. The numbers given on the edges are the measured traversal times (in some suitable time unit) between the two nodes connected by the edge in question. The traversal time depends both on the distance (not shown) between the two points and the amount of traffic.

   In this problem it is assumed that all vehicles start at node A and should move to node H, using probabilistic path generation as in the standard Ant system (AS) algorithm. Thus, it is possible for a vehicle to select a nominally slower route, perhaps to explore whether the traffic situation might have changed. (The traversal times are based on estimates from preceding days).

   In this case, the visibility of an edge is equal to the inverse of the traversal time for that edge. The objective function is taken as the inverse of the *total* traversal time as a vehicle moves from its start node to its end node. It is assumed that pheromone updates take place as soon as any vehicle reaches its end node.

   The parameters are as follows, using standard AS notation: $\alpha = 1$, $\beta = 2$, and $\rho = 0.5$, and the initial pheromone levels are set to $\tau_0 = 0.1$ for all edges.

   (a) What is the probability that the first vehicle will follow the fastest path from node A to node H (assuming that the estimated travel times are accurate for all edges)? (2p)

   (b) Assuming that the first vehicle actually *does* follow the fastest path, compute the updated pheromone levels on all edges (using the standard AS method for updating pheromones). (1p)

   (c) Assuming that the traversal times have not changed, what is the probability that the *next* vehicle will follow the fastest path from node A to node H, taking into account the changed pheromone levels resulting from the traversal of the first vehicle? (2p)

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

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

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

$$H = \begin{pmatrix} 8 & -2 \\ -2 & 4 \end{pmatrix} \quad (2)$$

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

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

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.