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

Exam in FFR 105 (Stochastic optimization algorithms), 2019-10-30, 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) any day after Nov. 17.

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) Premature convergence is a common problem in evolutionary algorithms. There are different ways of avoiding this problem, by modifying or extending the various operators used (or their parameters). Consider a genetic algorithm that (before modification) uses tournament selection (TS) and standard operators for crossover and mutation, with typical parameter settings. How can the algorithm be modified in order to prevent premature convergence (assuming that TS is used even after the modification)? Describe *two* ways for doing so. (2p)

   (b) In genetic algorithms, the concept of *genes* is central. In the biological counterpart, genes serve the purpose of providing the necessary information for generating proteins by means of a process that involves two major steps. Name and *describe* the two steps. (2p)

   (c) Tournament selection (TS) and roulette-wheel selection (RWS) are two commonly used selection operators in evolutionary algorithms.

      i. Consider a population consisting of five individuals with the fitness values $F_1 = 3$, $F_2 = 6$, $F_3 = 7$, $F_4 = 10$, and $F_5 = 12$. Using tournament selection with a tournament size of two and a tournament selection parameter of 0.8, what is the probability (in a single selection step) of selecting individual 4? (1p)

      ii. Roulette-wheel selection relies on the cumulative, normalized fitness sum, denoted ϕ_j. Write down the expression for ϕ_j, and explain clearly how it is used in RWS. (2p)

      iii. Consider again the population from part (i). If the random number $r = 0.3$ is drawn, which individual will be selected assuming that now RWS is used? (1p)

2. (a) In the gradient descent method, starting from a given point $\mathbf{x}_j$ (where $\mathbf{x}$ is a vector and the index enumerates the iterations) iterates are computed such that, once the search direction has been determined, the next iterate $\mathbf{x}_{j+1}$ depends only on the step length η, so that the *function value* at that point can be expressed as some function $\phi(\eta)$. Consider now the problem of minimizing the function $f(x_1, x_2) = 2x_1^2 + 3x_1x_2 + x_2^2 - 4$ using gradient descent, starting from the point $(x_1, x_2)^{\mathrm{T}} = (1, 1)^{\mathrm{T}}$. Find and write down the (non-normalized) search direction (i.e. a vector with two components) and the expression for the next iterate, inserting numerical values. Then derive (and simplify as much as possible) the expression for $\phi(\eta)$, again with numerical values inserted. Give your answer in the form $a_2\eta^2 + a_1\eta + a_0$, where a_0, a_1, and a_2 are constants. Note: You do *not* have to carry out the line search. It is sufficient that you find the search direction, the next iterate, and $\phi(\eta)$! (2p)

   (b) The Lagrange multiplier method is often used in problems with equality constraints. Use the Lagrange multiplier method to find the point on the sphere $x_1^2 + x_2^2 + x_3^2 = 4$ that is closest to the point $\mathbf{p} = (3, 1, -1)^{\mathrm{T}}$. (3p)

3. In ant colony optimization (ACO), a population of artificial ants cooperate to find the solution of a problem expressed in the form of a graph search. A common special case is the travelling salesman problem (TSP).

   (a) ACO is based on the cooperative behavior of ants and, in particular, a special form of communication used by ants. Name and *describe* this form of communication. (1p)

   (b) The (probabilistic) method for generating paths is a central feature of ACO in general, and ant system (AS) in particular. Write down the general equation for the probability $p(e_{ij}|S)$ for taking a step from a node j to another node i for the special case of TSP, given a path fragment S. Describe carefully all variables and parameters in the equation, and give typical numerical values for the parameters. (2p)

   (c) For TSP, AS generally finds the nearest-neighbour path (from a given start node) very quickly. Assuming that the pheromone level on all edges is equal to some value $\tau_0 > 0$, explain clearly *why* AS easily finds the nearest-neighbour path. (1p)

   (d) Consider a two-dimensional TSP problem with four nodes located at $(1, 0)$, $(0, 1)$, $(-1, 0)$, and $(0, -2)$, and where the ant starts from the first node, i.e. at $(1, 0)$. Using AS, what is the probability that this ant will follow the nearest-neighbour path, assuming that the pheromone levels are equal to $\tau_0 > 0$ on all edges, and the parameters α and β are equal to 1 and 2, respectively? (2p)

4. (a) In particle swarm optimization (PSO) there is a specific mechanism that handles the tradeoff between exploration and exploitation. Write down the general equation for the velocity updates in PSO and describe, in detail, the mechanism just mentioned. (2p)

(b) Consider now a simple one-dimensional application of PSO, in which one is trying to minimize the function $f(x) = (x - \frac{1}{4})^2$, using a swarm size of three. Initially the three particles are located at $x = -1/3$ (particle 1), $x = 0$ (particle 2), and $x = 3/4$ (particle 3), and their speeds are $v = 3$ (particle 1), $v = 1/4$ (particle 2) and $v = -1$ (particle 3). The parameters α and Δt are both equal to 1, w is (here) kept *constant* at the value 1, and $c_1 = c_2 = 2$. Moreover, assume (somewhat unrealistically) that the random numbers q and r are always equal to 1. The initial range $[x_{\min}, x_{\max}]$ is equal to $[-2, 2]$, and the particle speeds are thus restricted to a maximum of 4. Given these parameters, determine, under the PSO algorithm

i. ... the velocities and positions of all particles after one iteration (i.e. one updating step for both velocities and positions). (2p)

ii. ... the velocities and positions of all particles after two iterations (2p)

Stochastic optimization methods (FFR 105), 2019
Solutions to the exam (2019-10-30)

1. (a) Premature convergence can be prevented either by introducing a varying mutation rate (based on the degree of diversity in the population) or by reducing the crossover probability p_c. Another alternative is to introduce some form of mating restriction (as is done in diffusion models). Using fitness ranking would have been an alternative if roulette-wheel selection were used, but in this case it was stated that tournament solution would be used, thus excluding fitness ranking as an option.

   (b) The two steps are called transcription and translation. In transcription, the information in a gene (in the form of a sequence of bases, from the alphabet A, C, G, and T) is read by RNA polymerase, resulting in an mRNA molecule, containing the same information (albeit coded slightly differently) as the gene. In translation, the mRNA molecule is used as a template when forming a chain of amino acids (i.e. a protein). Each codon, i.e. a sequence of three bases in the mRNA molecule, e.g. CAA, encode a particular amino acid. Some codons encode the start and stop command. Once the stop command has been reached the amino acid chain is complete.

   (c) i. The number of possible pairs of individuals equals $5 \times 5 = 25$. Each of these pairs have equal probability of occurring, namely 1/25. Of those pairs, nine involved individual 4, namely (1,4), (2,4), (3,4), (4,4), (5,4), (4,1),(4,2),(4,3),(4,5). Individual 4 is the better individual in 6 cases, namely (4,1),(4,2),(4,3),(1,4),(2,4), and (3,4) and the worse individual in two cases, namely (4,5) and (5,4). In the remaining case, (4,4), individual 4 is selected with probability 1. Thus, with a tournament selection parameter of 0.8, one finds

$$p_4 = \frac{1}{25}\left(6 \times 0.8 + 2 \times (1 - 0.8) + 1\right) = 0.248. \tag{1}$$

   ii. The expression for the cumulative normalized fitness sum is

$$\phi_j = \frac{\sum_{i=1}^{j} F_i}{\sum_{i=1}^{N} F_i}, \tag{2}$$

   where F_i denotes the fitness of individual i and N is the population size. In RWS, a random number $r \in [0, 1[$ is drawn, and the selected individual is taken as the one with the smallest j that satisfies $\phi_j > r$.

iii. The fitness sum equals 38. Using the equation for ϕ_j one finds $\phi_1 = 3/38 \approx 0.0789$, $\phi_2 = 9/38 \approx 0.2368$, $\phi_3 = 16/38 \approx 0.4211 > 0.3$. Thus, the individual with the smallest j that satisfies ϕ_j is individual 3, which will thus be selected.

2. (a) The search direction is the negative gradient $(-\nabla f)$. Here, the gradient takes the form

$$\nabla f(x_1, x_2) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right)^T = (4x_1 + 3x_2, 3x_1 + 2x_2)^T, \tag{3}$$

where T denotes the transpose of the vector. Thus,

$$-\nabla f(x_1, x_2)|_{x_1=1, x_2=1} = -(7, 5)^T. \tag{4}$$

In gradient descent, the next iterate $\mathbf{x}_{i+1}$ is given by

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \eta \nabla f(\mathbf{x}_i). \tag{5}$$

At the point $(1, 1)^T$, one then obtains

$$\mathbf{x}_{i+1} = (1, 1)^T - \eta(7, 5)^T = (1 - 7\eta, 1 - 5\eta)^T. \tag{6}$$

At this point, the function thus becomes

$$\begin{aligned} \phi(\eta) &\equiv f(1 - 7\eta, 1 - 5\eta) = 2(1 - 7\eta)^2 + 3(1 - 7\eta)(1 - 5\eta) + (1 - 5\eta)^2 - 4 \\ &= 2 - 28\eta + 98\eta^2 + 3 - 15\eta - 21\eta + 105\eta^2 + 1 - 10\eta + 25\eta^2 - 4 = \\ &\quad 228\eta^2 - 74\eta + 2. \end{aligned} \tag{7}$$

(b) The function to be minimized is the distance (squared) from a point $(x_1, x_2, x_3)^{\mathrm{T}}$ to the point $\mathbf{P}$, given by

$$f(x_1, x_2, x_3) = (x_1 - 3)^2 + (x_2 - 1)^2 + (x_3 + 1)^2 \tag{8}$$

The constraint h takes the form

$$h(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 4 = 0. \tag{9}$$

In order to use the Lagrange multiplier method we set up $L(x_1, x_2, x_3, \lambda)$ as

$$L(x_1, x_2, x_3, \lambda) = f(x_1, x_2, x_3) + \lambda h(x_1, x_2, x_3). \tag{10}$$

Taking the partial derivatives of L with respect to x_1, x_2, x_3 and λ one obtains, in order, the equations

$$2(x_1 - 3) + 2\lambda x_1 = 0, \tag{11}$$

$$2(x_2 - 1) + 2\lambda x_2 = 0, \tag{12}$$

$$2(x_3 + 1) + 2\lambda x_3 = 0, \tag{13}$$

and

$$x_1^2 + x_2^2 + x_3^2 - 4 = 0. \tag{14}$$

From the first of those equations, one gets $x_1(1+\lambda) = 3$, so that

$$x_1 = \frac{3}{1+\lambda}. \tag{15}$$

From the second and third equation, one obtains, in a similar fashion,

$$x_2 = \frac{1}{1+\lambda}. \tag{16}$$

$$x_3 = -\frac{1}{1+\lambda}. \tag{17}$$

Using the constraint, one thus finds

$$\frac{9}{(1+\lambda)^2} + \frac{1}{(1+\lambda)^2} + \frac{1}{(1+\lambda)^2} = 4, \tag{18}$$

so that

$$(1+\lambda)^2 = \frac{11}{4}. \tag{19}$$

Therefore $\lambda = -1 \pm \frac{\sqrt{11}}{2}$. With these values of λ one finally obtains the two points $\mathbf{Q}_1 = (6, 2, -2)/\sqrt{11}$ and $\mathbf{Q}_2 = (-6, -2, 2)/\sqrt{11}$. It is then easy to check that the smallest value of f occurs at $\mathbf{Q}_1$ (and the largest value occurs at $\mathbf{Q}_2$).

3. (a) Cooperative behavior in ants depends on *stigmergy*, which is a form of communication relying on (local) modification of the environment: As the ants move, they deposit pheromones (a form of volatile hydrocarbon) that other ants can (and often will) follow.

(b) The probability $p(e_{ij}|S)$ takes the form

$$p(e_{ij}|S) = \frac{\tau_{ij}^{\alpha}\eta_{ij}^{\beta}}{\sum_{\nu_l \notin L_{\mathrm{T}}(S)} \tau_{lj}^{\alpha}\eta_{lj}^{\beta}}, \tag{20}$$

where τ_{ij} is the pheromone level on the edge from node j to node i, η_{ij} is the visibility (which for TSP takes the form $1/d_{ij}$, where d_{ij} is the distance from node j to node i). $L_{\mathrm{T}}(S)$ is the tabu list, i.e. the list of all nodes visited so far. α is a parameter that usually takes the value 1, whereas the parameter β usually takes values in the range 2 to 5.

(c) If the pheromones are equal on all edges (as was assumed here), the probability of following a given edge is proportional to $\eta_{ij}^{\beta} = (1/d_{ij})^{\beta}$. Now, since the value of β generally is (at least) 2, it is evident that the probability of going to the nearest node (for which $1/d_{ij}$ is maximal) will be higher than the probability of going to any other node. Thus, it is not unlikely that at least one or a few ants will follow the nearest-neighbour path in the first iteration (or one of the first iterations).

(d) Starting from Node 1, at $(1, 0)$, the nearest node is clearly Node 2 (at $(0, 1)$), which is at a distance $d_{21} = \sqrt{2}$. The distances to the other nodes equal $d_{31} = 2$ and $d_{41} = \sqrt{5}$. Since the pheromone levels are the same on all edges, one can neglect them, and the probability of going from Node 1 to Node 2 thus takes the form

$$p(e_{21}|S = \{\nu_1\}) = \frac{\left(\frac{1}{d_{21}}\right)^2}{\left(\frac{1}{d_{21}}\right)^2 + \left(\frac{1}{d_{31}}\right)^2 + \left(\frac{1}{d_{41}}\right)^2} = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{4} + \frac{1}{5}} = \frac{10}{19}. \tag{21}$$

Once the artificial ant reaches Node 2, it can either go to Node 3 (distance $d_{32} = \sqrt{2}$) or to Node 4 (distance $d_{42} = 3$). The probability of moving to the nearest node (Node 3) is given by

$$p(e_{32}|S = \{\nu_1, \nu_2\}) = \frac{\left(\frac{1}{d_{32}}\right)^2}{\left(\frac{1}{d_{32}}\right)^2 + \left(\frac{1}{d_{42}}\right)^2} = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{9}} = \frac{9}{11}. \tag{22}$$

At Node 3, the ant has no option but to go to Node 4 (with probability 1), from which it then returns to Node 1. Thus, the probability of traversing the nearest-neighbour path, starting from Node 1, equals

$$p_{1234} = \frac{10}{19} \times \frac{9}{11} = \frac{90}{209} \approx 0.431. \tag{23}$$

4. (a) The velocity update for particle i is given by

$$v_{ij} \leftarrow wv_{ij} + c_1 q\left(\frac{x_{ij}^{\text{pb}} - x_{ij}}{\Delta t}\right) + c_2 r\left(\frac{x_j^{\text{sb}} - x_{ij}}{\Delta t}\right),\ j = 1, \ldots, n, \tag{24}$$

where w, the inertia weight, handles the tradeoff between exploration and exploitation. If $w > 1$, exploration is favored. If instead $w < 1$, the particle focuses on exploitation of the results already found. Normally, one starts with a value of w of around 1.4, then reduces w by a factor $\beta \approx 0.99$ until w reaches a lower limit of around $0.3 - 0.4$, where it is then kept constant.

(b) i. Initially, the function values are 49/144 (particle 1), 1/16 (particle 2), and 1/4 (particle 3). Thus, the swarm best position is equal to the position of particle 2 (i.e. $x = 0$). With the simplifications, the velocity update takes the form

$$v_i \leftarrow v_i + 2(x_i^{\text{pb}} - x_i) + 2(x^{\text{sb}} - x_i),\ i = 1, 2, 3. \tag{25}$$

One then obtains:

$$v_1 = 3 + 2(-1/3 - (-1/3)) + 2(0 - (-1/3)) = 11/3, \tag{26}$$

$$v_2 = 1/4 + 2(0 - 0) + 2(1/3 - 1/3) = 1/4, \tag{27}$$

and

$$v_3 = -1 + 2(3/4 - 3/4) + 2(0 - 3/4) = -5/2. \tag{28}$$

Thus, using the equation $x \leftarrow x + v$, the new positions become

$$x_1 = -1/3 + 11/3 = 10/3, \tag{29}$$

$$x_2 = 0 + 1/4 = 1/4, \tag{30}$$

$$x_3 = 3/4 - 5/2 = -7/4. \tag{31}$$

ii. In the second iteration, the swarm best position is $x = 1/4$, i.e. the position of particle 2 (which, of course, also is the particle best position for that particle). The particle best position is unchanged for particle 1 and particle 3, since the function values at their new positions exceeds those obtained at their initial positions. Using the same equations as above, one obtains

$$v_1 = 11/3 + 2(-1/3 - 10/3) + 2(1/4 - 10/3) = -59/6. \tag{32}$$

However, this value exceeds (in magnitude) the maximum (negative) speed of -4, meaning that the actual speed of the particle will be $v_3 = -4$ instead. For particle 2 one gets

$$v_2 = 1/4 + 2(1/4 - 1/4) + 2(1/4 - 1/4) = 1/4 \tag{33}$$

and for particle 3

$$v_3 = -5/2 + 2(3/4 - (-7/4)) + 2(1/4 - (-7/4)) = 13/2. \tag{34}$$

This value is larger than the limit of 4, so that the actual speed will be $v_3 = 4$ instead. Thus, finally, one obtains

$$x_1 = 10/3 - 4 = -2/3, \tag{35}$$

$$x_2 = 1/4 + 1/4 = 1/2, \tag{36}$$

and

$$x_3 = -7/4 + 4 = 9/4. \tag{37}$$