

Chalmers University of Technology, Department of Applied Mechanics  
Teacher: Mattias Wahde, tel. 772 3727

**Exam in FFR 105 (Stochastic optimization algorithms), 2011-10-17,  
14.00-18.00, V.**

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) When generating new individuals in a genetic algorithm, several different operators are used, namely selection, crossover and mutation. Name and describe two selection methods. For each method, you should include a clearly described algorithm for selecting one individual, given a random number  $r$  in the range  $[0, 1[$ . (3p)
- (b) Particle swarm optimization (PSO) is a stochastic optimization method somewhat similar to genetic algorithms. Write down the standard PSO algorithm and describe it in detail. Include clear definitions of all variables and parameters used in your description. In particular, describe how the tradeoff between exploration and exploitation (of the results already found) is handled in PSOs. (4p)
- (c) Newton's method is an iterative method for finding local optima of an objective function  $f(\mathbf{x})$ . In the one-dimensional case, the method is referred to as the Newton-Raphson method. Starting from the Taylor (series) expansion of  $f(x)$ , *derive* Newton-Raphson's method, i.e. the equation that determines how to obtain the new iterate  $x_{j+1}$ , starting from the previous iterate  $x_j$ . Note: Provide a clear derivation of the equation. Just writing down the equation will not give any points. (2p)
- (d) In optimization, convexity of the objective function is a desirable property. Determine whether or not the function

$$f(x_1, x_2) = 4x_1^2 - 5x_1x_2 + 3x_2^2 - 7x_2 - 4 \quad (1)$$

is convex. (1p)

2. Consider a function adaptation task in which linear genetic programming (LGP) is used for finding an unknown function  $f(x)$  based on measurements taken for several different values of  $x$ . The LGP chromosomes consist of a sequence of instructions, each represented using four genes. The first gene in each instruction represents the operator, the second gene represents the destination register, and the two remaining genes are the operands. In this task, there are three variable registers (denoted  $r_1$ ,  $r_2$ , and  $r_3$ ), and three constant registers (denoted  $c_1$ ,  $c_2$ , and  $c_3$ ). There are four operators, namely  $o_1$  (addition),  $o_2$  (subtraction),  $o_3$  (multiplication) and  $o_4$  (division). Initially the constant registers are set as  $c_1 = 1$ ,  $c_2 = 2$  and  $c_3 = -1$ . The variable registers are initiated as  $r_1 = x$ ,  $r_2 = r_3 = 0$ . The output (i.e. the estimate  $\hat{f}(x)$ ) is taken as the contents of  $r_2$ . The operands are chosen from the set  $\{a_1, \dots, a_6\} = \{r_1, r_2, r_3, c_1, c_2, c_3\}$ .

- (a) Consider an LGP chromosome given by

$$c_1 = 1214\ 3315\ 3123\ 3333\ 1323\ 4213. \quad (2)$$

Which function is obtained when decoding this chromosome? (2p)

- (b) During mutation, the fourth gene in the chromosome  $c_1$  is mutated from 4 to 1. What will be the corresponding function? (1p)

3. Ant colony optimization (ACO), which is inspired by the behavior of ants, is typically used for solving routing problems, such as the traveling salesman problem (TSP).

- (a) Consider the construction graph (for TSP) shown in Fig. 1 (see next page). If the level of artificial pheromone is equal to 0.5 for all edges  $e_{ij}$ , what is the probability that an ant will follow the nearest neighbour path, starting from node 1? For the parameter values, choose  $\alpha = 1$  and  $\beta = 2$ . Make sure to include all relevant intermediate steps in your calculations! (3p)

- (b) Several ACO algorithms have been defined, one of which is the Max-min ant system (MMAS) in which (among other things) pheromone limits are imposed. Consider again the construction graph for TSP in Fig. 1. Assuming that MMAS is being used, with  $N = 4$  artificial ants, and that the initial pheromone levels  $\tau_{ij}$  in this case are equal to  $1/(\rho D_{\text{nn}})$  for all edges  $e_{ij}$ , where  $D_{\text{nn}}$  is the length of the nearest-neighbour path starting from node 1 (i.e. the path considered above), determine the pheromone levels for *all* edges  $e_{ij}$  after the first iteration, where the four ants followed the paths (1, 4, 2, 5, 3), (2, 4, 1, 3, 5), (5, 3, 1, 4, 2), and (1, 2, 3, 4, 5), respectively. (Note that, as usual, the ants also return to their start node in the final step.) For the pheromone updating rule, set the evaporation rate  $\rho$  to 0.5, and the pheromone limits  $\tau_{\text{min}}$  and  $\tau_{\text{max}}$  to 0.1 and  $1/(\rho D_{\text{nn}})$ , respectively. (4p)

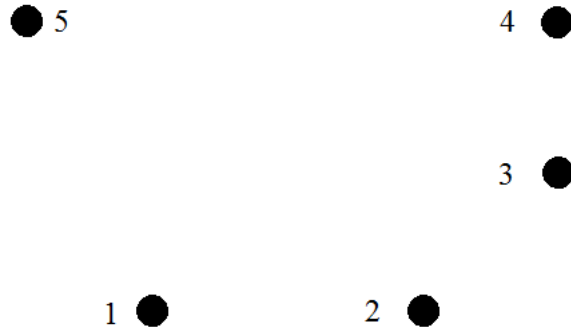


Figure 1: Construction graph for Problem 3. The nodes are located at  $(1, 0)$  (node 1),  $(3, 0)$  (node 2),  $(4, 1)$  (node 3),  $(4, 2)$  (node 4), and  $(0, 2)$  (node 5).

4. The penalty method is a classical optimization method (which, however, also can be used in connection with stochastic optimization) for solving constrained minimization problems.
  - (a) In the penalty method a penalty function is used for measuring the degree to which the constraints are violated for a given variable vector  $\mathbf{x}$ . The penalty function is then added to the objective function  $f(\mathbf{x})$ . Write down the general expression for the penalty function, carefully explaining all variables and parameters. (1p)
  - (b) Use the penalty method to find the minimum of the function

$$f(x_1, x_2) = (x_1 - 6)^2 + (x_2 - 7)^2, \quad (3)$$

subject to the constraints

$$g_1(x_1, x_2) = -3x_1 - 2x_2 + 6 \leq 0, \quad (4)$$

$$g_2(x_1, x_2) = -x_1 + x_2 - 3 \leq 0, \quad (5)$$

$$g_3(x_1, x_2) = x_1 + x_2 - 7 \leq 0, \quad (6)$$

and

$$g_4(x_1, x_2) = \frac{2}{3}x_1 - x_2 - \frac{4}{3} \leq 0, \quad (7)$$

Hint: Start at the unconstrained minimum, and examine the constraints carefully. It is also a good idea to plot the set of feasible points before starting with the actual minimization. (4p)

1. (a) Roulette-wheel selection and tournament selection, see pp. 48-50 in the course book. For roulette-wheel selection, the quantities

$$\phi_j = \frac{\sum_{i=1}^j F_i}{\sum_{i=1}^N F_i} \quad (1)$$

are generated for  $j = 1, 2, \dots, N$ . Next a random number  $r \in [0, 1[$  is drawn and the selected individual is taken as the individual with the smallest  $j$  that satisfies  $\phi_j > r$ . For tournament selection (with tournament size 2), two individuals are picked randomly from the population. Next a random number  $r \in [0, 1[$  is generated. If  $r < p_{\text{tour}}$  (typically around 0.7-0.8), the better of the two individuals is chosen, if not, the worse individual is chosen. Tournament selection can also be generalized to the case of tournaments with more than two participants. In that case, the best individual (of the  $j$  randomly picked individuals) is selected with probability  $p_{\text{tour}}$  as just described. If this individual is not selected, it is removed from the tournament, a new random number  $r$  is drawn, and the best of the *remaining* individuals is selected with probability  $p_{\text{tour}}$  etc. Note that both selection methods take place with replacement. That is, a given individual can be selected several times.

- (b) The standard PSO is given in Algorithm 5.1 in the book *but with the inertia term added*, see Eq. (5.20) in the book. For full points, the description should include the five steps of the algorithm (initialization, evaluation, best position updates (particle best and swarm best), position and velocity updates, and the return to step 2. The indices  $i$  (enumerating particles) and  $j$  (enumerating dimensions) should be introduced correctly in all parts of the algorithm; for example, the swarm best vector  $x_j^{\text{sb}}$  should have only *one* index. Furthermore, the velocity update equation should be clearly described (the cognitive and social terms, with the two constants  $c_1$  and  $c_2$  and the random numbers  $q$  and  $r$ ). The velocity restriction should be defined. The trade-off between exploration and exploitation is taken care of by the inertia term, which should vary from around 1.4 down to 0.3-0.4.
- (c) Making a Taylor expansion of  $f(x)$ , one obtains

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 \equiv f_{[2]}(x). \quad (2)$$

Taking the derivative and setting it to zero, one finds the stationary point. Thus

$$f'_{[2]}(x) = 0 \Leftrightarrow f'(x_0) + (x - x_0)f''(x_0) = 0. \quad (3)$$

Solving this equation, one obtains

$$x^* = x_0 - \frac{f'(x_0)}{f''(x_0)}. \quad (4)$$

Thus, the iteration rule takes the form

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

- (d) The convexity of a function can be investigated by considering the properties of the Hessian. For the function in question, the Hessian equals

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 8 & -5 \\ -5 & 6 \end{pmatrix}, \quad (6)$$

with eigenvalues  $7 \pm \sqrt{26}$  which are both larger than zero. Thus, the function is convex.

2. (a) Decoding the chromosome, one obtains:

$$\begin{aligned} 1214 &\Leftrightarrow r_2 := r_1 + c_1 \\ 3315 &\Leftrightarrow r_3 := r_1 \times c_2 \\ 3123 &\Leftrightarrow r_1 := r_2 \times r_3 \\ 3333 &\Leftrightarrow r_3 := r_3 \times r_3 \\ 1323 &\Leftrightarrow r_3 := r_2 + r_3 \\ 4213 &\Leftrightarrow r_2 := r_1 / r_3 \end{aligned}$$

With the initial values  $r_1 = x$ ,  $r_2 = r_3 = 0$ , and with  $c_1 = 1$ ,  $c_2 = 2$ , and  $c_3 = -1$ , one obtains:

$$\begin{aligned} \text{Step 1: } &r_1 = x, r_2 = x + 1, r_3 = 0 \\ \text{Step 2: } &r_1 = x, r_2 = x + 1, r_3 = 2x \\ \text{Step 3: } &r_1 = 2x(x + 1), r_2 = x + 1, r_3 = 2x \\ \text{Step 4: } &r_1 = 2x(x + 1), r_2 = x + 1, r_3 = 4x^2 \\ \text{Step 5: } &r_1 = 2x(x + 1), r_2 = x + 1, r_3 = 4x^2 + x + 1 \\ \text{Step 6: } &r_1 = 2x(x + 1), r_2 = 2x(x + 1) / (4x^2 + x + 1), r_3 = 4x^2 + x + 1 \end{aligned}$$

Thus, the answer is

$$\hat{f}(x) = \frac{2x^2 + 2x}{4x^2 + x + 1}. \quad (7)$$

- (b) The only difference compared to the case considered above is that the first instruction now takes the form

$$1211 \Leftrightarrow r_2 := r_1 + r_1$$

With the same initial values as above, one obtains

$$\begin{aligned} \text{Step 1: } &r_1 = x, r_2 = 2x, r_3 = 0 \\ \text{Step 2: } &r_1 = x, r_2 = 2x, r_3 = 2x \\ \text{Step 3: } &r_1 = 4x^2, r_2 = 2x, r_3 = 2x \\ \text{Step 4: } &r_1 = 4x^2, r_2 = 2x, r_3 = 4x^2 \\ \text{Step 5: } &r_1 = 4x^2, r_2 = 2x, r_3 = 4x^2 + 2x \\ \text{Step 6: } &r_1 = 4x^2, r_2 = 4x^2 / (4x^2 + 2x), r_3 = 4x^2 + 2x \end{aligned}$$

Thus, the answer is

$$\hat{f}(x) = \frac{4x^2}{4x^2 + 2x} = \frac{2x}{2x + 1} \quad (x \neq 0) \quad (8)$$

3. (a) It is easy to see that the nearest-neighbour path starting from node 1 is  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$  (with a return to node 1 as the final step). The equation for determining the probability of a move from node  $j$  to node  $i$  takes the following form

$$p(e_{ij}|S) = \frac{\tau_{ij}^\alpha \eta_{ij}^\beta}{\sum_{\nu_i \notin L_T(S)} \tau_{ij}^\alpha \eta_{ij}^\beta}, \quad (9)$$

where  $\eta_{ij} = 1/d_{ij}$ . Since the pheromone levels are equal on all edges, the  $\tau_{ij}$  terms cancel out, and one is left with the expression

$$p(e_{ij}|S) = \frac{\eta_{ij}^\beta}{\sum_{\nu_i \notin L_T(S)} \eta_{ij}^\beta} \quad (10)$$

In node 1, there are four possible moves, with distances  $d_{21} = 2$ ,  $d_{31} = \sqrt{10}$ ,  $d_{41} = \sqrt{13}$ , and  $d_{51} = \sqrt{5}$ . With  $\beta = 2$  the probability of moving to node 2 becomes

$$p(e_{21}|S = \{\nu_1\}) = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{10} + \frac{1}{13} + \frac{1}{5}} \approx 0.398773. \quad (11)$$

In node 2, there are three possible moves, with distances  $d_{32} = \sqrt{2}$ ,  $d_{42} = \sqrt{5}$ , and  $d_{52} = \sqrt{13}$ . Thus

$$p(e_{32}|S = \{\nu_1, \nu_2\}) = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{5} + \frac{1}{13}} \approx 0.643564. \quad (12)$$

In node three, there are two possible moves, with distances  $d_{43} = 1$  and  $d_{53} = \sqrt{17}$ . Thus

$$p(e_{43}|S = \{\nu_1, \nu_2, \nu_3\}) = \frac{1}{1 + \frac{1}{17}} \approx 0.944444. \quad (13)$$

The remaining steps, from node 4 to node 5 and then back to node 1, take place with probability 1. Thus, the probability of following the nearest-neighbour path, starting at node 1, becomes

$$p_{12345} = p_{21} \times p_{32} \times p_{43} \approx 0.242. \quad (14)$$

- (b) The length of the nearest-neighbour path starting from node 1 equals

$$L_{12345} = 2 + \sqrt{2} + 1 + 4 + \sqrt{5} \approx 10.65 \quad (15)$$

length units. The initial pheromone level  $\tau_{ij}$  is thus equal to

$$\tau_{ij} = \frac{1}{\rho D_{\text{nn}}} \approx 0.1878. \quad (16)$$

In MMAS, only the best ant is allowed to deposit pheromone. One can easily see that the path of the fourth ant is the shortest (i.e. the best). Since this path is the nearest-neighbour path considered above, we can write

$$\Delta\tau_{ij}^{[b]} = \frac{1}{D_{\text{nn}}} \quad (17)$$

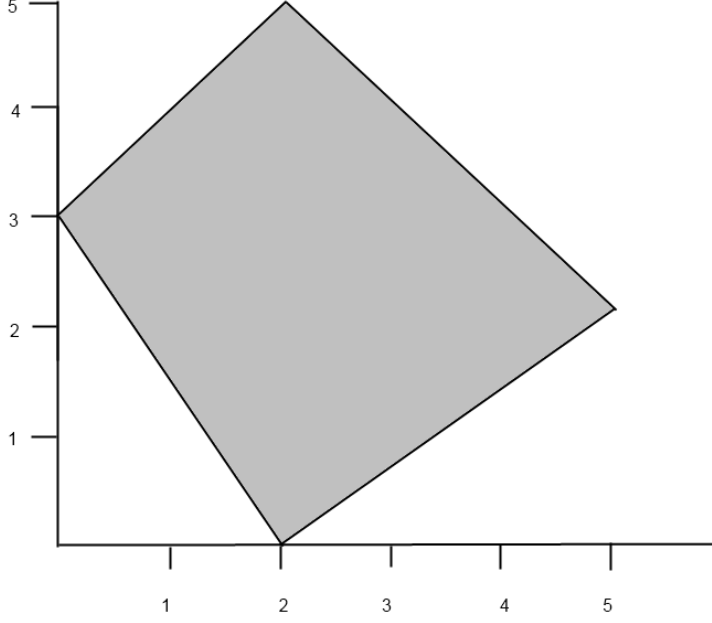


Figure 1: The feasible region for Problem 4.

for the edges in that path, namely  $e_{21}$ ,  $e_{32}$ ,  $e_{43}$ ,  $e_{54}$  and  $e_{15}$ . the pheromones are updated as

$$\tau_{ij} \leftarrow \tau_{ij}(1 - \rho) + \Delta\tau_{ij}^{[b]} = \frac{1}{\rho D_{nn}}(1 - \rho) + \frac{1}{D_{nn}} = \frac{1}{\rho D_{nn}}. \quad (18)$$

That is, the pheromone levels remain unchanged on these edges. For all other edges, the pheromones change as

$$\tau_{ij} \leftarrow \tau_{ij}(1 - \rho) = \frac{1}{\rho D_{nn}}(1 - \rho) \approx 0.0939. \quad (19)$$

However, since MMAS introduces a lower pheromone limit, in this case 0.1, the pheromone levels on those edges will be equal to 0.1 rather than 0.0939.

4. (a) The penalty term takes the form

$$p(\mathbf{x}; \mu) = \mu \left( \sum_{i=1}^m (\max\{g_i(\mathbf{x}), 0\})^2 + \sum_{i=1}^k (h_i(\mathbf{x}))^2 \right), \quad (20)$$

where  $g_i(\mathbf{x})$  and  $h_i(\mathbf{x})$  are the ( $m$ ) inequality and ( $k$ ) equality constraints, respectively and  $\mu$  is a positive parameter that determines the magnitude of the penalty.

- (b) Using the four constraints, one can plot the feasible region, see the figure above. It is easy to see that the unconstrained minimum occurs at  $(x_1, x_2) = (6, 7)$ . Starting at this point, one can see that, in fact, only the third constraint is violated here. Thus, at this point, one can write objective function as

$$f_p(\mathbf{x}; \mu) = (x_1 - 6)^2 + (x_2 - 7)^2 + \mu(x_1 + x_2 - 7)^2. \quad (21)$$

Setting the gradient to zero, one obtains

$$\frac{\partial f_p}{\partial x_1} = 2x_1 - 12 + 2\mu(x_1 + x_2 - 7) = 0 \quad (22)$$

and

$$\frac{\partial f_p}{\partial x_2} = 2x_2 - 14 + 2\mu(x_1 + x_2 - 7) = 0. \quad (23)$$

From these equations, one finds a single solution, namely

$$x_1(\mu) = \frac{6(1 + \mu)}{1 + 2\mu}, \quad (24)$$

$$x_2(\mu) = 7 - \frac{6\mu}{1 + 2\mu}. \quad (25)$$

With  $\mu \rightarrow \infty$  one obtains  $(x_1, x_2) = (3, 4)$ . Moreover, for any finite value of  $\mu$ , the point  $(x_1(\mu), x_2(\mu))$  violates the third constraint (and only that constraint), as can be seen by studying the constraints and the feasible region in the figure. One can also see that  $f_p(\mathbf{x}; \mu)$  is strictly convex for any  $\mu > 0$ , so the point  $(3, 4)$  is the global minimum of  $f$  subject to the constraints.